



National Conference on Artificial Intelligence and its Applications

Nassim Dennouni, Benazzouz Mourtada, Hadjila Fethallah, Meryem Saidi, Chikhi Kamel, Mhamed Hadjhenni, Baba Hamed Amel, Rafika Benladghem, Belmokeddem Mohamed, Sarah Medjroud, et al.

► To cite this version:

Nassim Dennouni, Benazzouz Mourtada, Hadjila Fethallah, Meryem Saidi, Chikhi Kamel, et al.. National Conference on Artificial Intelligence and its Applications. National Conference on Artificial Intelligence and its Applications, figshare, 2023, 10.6084/m9.figshare.24902382.v3 . hal-04368377

HAL Id: hal-04368377

<https://univ-rennes.hal.science/hal-04368377>

Submitted on 31 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
Higher School of Management-TLEMCEN



National Conference on Artificial Intelligence and its Applications

NCAIA'2023

Tlemcen, Thursday, 7th December 2023



Proceedings



<https://sites.google.com/view/cniaa2023/home>



OVERVIEW

Artificial Intelligence (AI) plays an increasingly important role in the economies of countries worldwide. Its impact is felt at various levels, both on a macroeconomic scale and within businesses. At the macroeconomic level, AI contributes to stimulating economic growth by enhancing efficiency and productivity. AI-based technologies automate repetitive tasks and improve process efficiency, resulting in time and cost savings for businesses. AI also fosters innovation by enabling the development of new products and services based on advanced machine learning and data analysis capabilities.

In many sectors, AI creates new economic opportunities on the healthcare, tourism domain, finance, and e-commerce sectors. Companies that adopt AI can benefit from better understanding of customer needs, product and service personalization, and improved user experiences, leading to increased their revenues and market share.

NCAIA aims to promote the use of AI in order to sensitize the university community to its impact on the economies of countries. This day is also an opportunity to present research work related to all areas of IA. The program of this scientific event includes tutorials on current topics and papers on research work. This event is open to all topics on IA, but a preference is given to topics as follows:

- Recommendation of item in e-commerce
- Recommendation of POI in tourism
- Credit risk assessment
- Fraud detection and prevention
- Financial forecasting
- Intelligent tutoring systems
- Personalized learning platforms
- Customer segmentation
- Targeted advertising
- Sentiment analysis of social media content
- Trend prediction and monitoring
- Medical image analysis
- Disease diagnosis and prognosis
- Image and video analysis
- Text summarization and Language generation

ORGANIZING COMMITTEE

Honorary chair

Dr. SAIDI Tarik, Director of the higher school of management -Tlemcen-

President of the committee

Dr. DENNOUNI Nassim, Higher school of management -Tlemcen-

Coordinators:

Dr ABDELMALAK Mohamed, Higher school of management -Tlemcen- (Algeria)
Dr BENABDDALLAH Ali, Higher school of management -Tlemcen- (Algeria)
Pr BENBOUZIANE Mohamed, Higher school of management -Tlemcen- (Algeria)
Dr BELGHERBI Latifa, Higher school of management -Tlemcen- (Algeria)
Dr BEZZAR Mohammed Soufyane, Higher school of management -Tlemcen- (Algeria)
Dr BRAHIMI Assia, Higher school of management -Tlemcen- (Algeria)
Pr CHENINI Moussa, Higher school of management -Tlemcen- (Algeria)
Dr KHEDIM Amel, Higher school of management -Tlemcen- (Algeria)
Dr IZNASNI Ali, Higher school of management -Tlemcen- (Algeria)
Dr OTMANI Habib, Higher school of management -Tlemcen- (Algeria)
Dr SAIDI Meryem, Higher school of management -Tlemcen- (Algeria)

SCIENTIFIC COMMITTEE

Pr ADJOUDJ Reda, SBA University (Algeria)
Dr ARIDJ Mohamed, Chlef University (Algeria)
Dr BENABDELLA Aouicha, Higher school of management -Tlemcen- (Algeria)
Dr BANAZZOUZ Mourtada, Tlemcen University (Algeria)
Dr BENOMAR Mohammed Lamine, Ain Temouchent University
Pr BOUKLI HACENE Sofiane, SBA University (Algeria)
Pr CHIKHI Kamel, Higher school of management -Tlemcen- (Algeria)
Pr KADRI Benamer, Tlemcen University (Algeria)
Dr HADJILA Fethallah, Tlemcen University (Algeria)
Dr.HAID zahia, higher school of management -Tlemcen- (Algeria)
Dr LOUAZANI Ahmed, Blida University (Algeria)
Dr LOUKAM Mourad, Chlef University (Algeria)
Dr MAGNAFI Hicham, Higher School of Applied Science, -Tlemcen- (Algeria)
Dr MAHAMMED Bachir Nadir, Higher School of Computer Science -SBA-(Algeria)
Dr MALIKI Fouad, Higher School of Applied Science, -Tlemcen- (Algeria)
Dr MERZOUG Mohamed, Tlemcen University (Algeria)
Pr MEZIANE Tadj, Higher school of management -Tlemcen- (Algeria)
Dr NAHNOUH Chakib, Chlef University (Algeria)
Dr RAHALI Amina Souad, Higher school of management -Tlemcen- (Algeria)
Dr SAIDI Meryem, Higher school of management -Tlemcen- (Algeria)
Dr SELMI Abdeldjebbar, Higher school of management -Tlemcen- (Algeria)
Dr SLAMA Zohra, SBA University (Algeria)
Pr SOUIER Mahdi, Tlemcen University (Algeria)
Dr TAHRAOUI Mohammed Amin, Chlef University (Algeria)

CONTACTS

The Higher school of management -Tlemcen-

Adresse : 01, Rue Barka Ahmed Bouhannak Imama (près du commissariat des 400 logements de Bouhannak)
ou bien Demandez l'ex faculté de Biologie de Tlemcen (Coordonnée GPS Latitude, GPS Longitude:
34.87837569635192, -1.3619841740396654)

Tél: +213 (0)43-21-16-47

Fax: +213 (0)43-21-16-45

Email: cniaa2023@gmail.com

Web site: <https://sites.google.com/view/cniaa2023/>

Table of Contents

Artificial Intelligence: understanding an increasingly prevalent technology	5
Machine intelligence: from Turing test to generative pretrained transformers	13
EEG Signal Analysis for Automatic Detection of psychiatric diseases	20
Hardware Design for Medical Image Enhancement.....	25
Analyzing the Influence of Medical Imbalanced Data on Performance and Fairness in Differentially Private Deep Learning.....	30
Residual network (resnet-18) for laparoscopic image distortion classification.....	36
Artificial intelligence techniques in the smart grid: a review	39
Forecasting Algeria's General Industrial Index: A Comprehensive Data-Driven Analysis and Predictive Insights	44
Balanced exploration and exploitation properties of multi-objective sooty tern optimizer	51
Orchestration of POIs ubiquitous contexts: a review of recommendation systems based on matrix factorization model ...	55
Automatic Opinion Extraction from Football-Related Social Media: A Gazetteer and Rule- Based Approach.....	61
Optimizing Arabic spam filtering through unsupervised and ensemble learning approaches.....	66
Investigating explainable AI for enhanced atherosclerosis detection and decision transparency	69
Toward Mitigating Tourists' Indifference in POI recommendation systems using MAS	75
MLP-powered smart application to enhance efficiency and productivity in Algerian agriculture.....	80
Recommendation System: a review of trust techniques	84
Biometric authentication using invariant Gray Level Co-occurrence Matrices (GLCM)	89
AI-based pathogenicity classification of flea and tick-borne diseases and seasonal distribution of some species in Algeria	95
Enhancing education decision-making with deep learning for Arabic spoken digit recognition	99
Deep Learning Models for Left Atrial Segmentation in MRI	103
Recommendation Systems: techniques, challenges and applications.....	107
The impact of AI on credit risk assessment: A SWOT analysis.....	110
Financial forecasting using an ARIMA model	115
Usage of artificial intelligence in online shopping: Algerian customer attitude	118
How AI is helping marketers and advertisers?	122
Demystifying AI in consumer product logistics: A case study of « Maersk »	126

Artificial Intelligence: understanding an increasingly prevalent technology

BENAZZOUZ Mourtada

Dept. Computer Science Genie Biomedical Laboratory
University of Tlemcen, Algeria
mourtada.benazzouz@univ-tlemcen.dz

Abstract— Artificial Intelligence (AI) has rapidly evolved and integrated into various aspects of our daily lives. This paper explores the increasing prevalence of AI technologies and their multifaceted impact on society. We examine how AI-driven systems have become integral in fields such as healthcare, finance, transportation, and entertainment, fundamentally reshaping industries and work processes. Additionally, we discuss the challenges and opportunities associated with the widespread adoption of AI, including ethical considerations, privacy concerns, and the profound transformations it brings to the job market. This study provides valuable insights into the omnipresent nature of AI, shedding light on the ways in which it shapes our contemporary world and redefines the landscape of employment.

Keywords—Artificial Intelligence (AI); Machine Learning; Deep Learning; Neural Networks; Ethics; Privacy Concerns

I. INTRODUCTION

Artificial Intelligence (AI) is a concept that has garnered considerable attention in recent years, permeating various aspects of our daily lives and reshaping the way we interact with technology and data. While the term "Artificial Intelligence" has been subject to diverse interpretations, it is valuable to begin our exploration by considering two distinct definitions. According to Wikipedia, "Artificial Intelligence refers to a collection of theories and techniques aimed at developing devices capable of replicating human intelligence." This definition provides a foundational perspective, emphasizing AI's aspiration to simulate human-like intelligence through the application of theories and techniques.

In contrast, when we consult an AI system like ChatGPT, we find another intriguing perspective: "Artificial Intelligence, often abbreviated as AI, refers to the field of study and development focused on creating computer systems and software that possess the ability to perform tasks typically requiring human intelligence, such as learning from data, problem-solving, and making decisions. AI aims to replicate human-like cognitive functions, enabling machines to understand, adapt, and improve their performance in a wide range of applications." Here, we witness the synthesis of modern AI's capabilities and aspirations, which extend beyond

mere replication to encompass practical problem-solving and adaptability.

As we delve deeper into this paper, we will explore the ever-expanding presence of AI in our lives, its multifaceted impact on various sectors, and the challenges and opportunities it presents. Furthermore, we will investigate how AI has not only become omnipresent but also transformative, influencing the job market, ethical considerations, privacy concerns, and the ways in which we navigate our contemporary world [1].

The forthcoming section provides an overview of the history of AI and its significant developments. It is succeeded by the various forms of AI. The succeeding section explains the workings of AI, along with its primary functions. Lastly, our study concludes by examining the ethical concerns and future prospects.

II. BACKGROUND ON AI

The history of artificial intelligence is a tale of human ingenuity and technological advancement. It begins with the rudimentary mechanical calculating device created by the French mathematician Blaise Pascal in 1642 and evolves through a series of significant milestones, each shaping the trajectory of AI research. In this chapter, we will explore key dates and events that have paved the way for the development of artificial intelligence, including the first programmable machine by Ada Lovelace, the introduction of neural networks, Alan Turing's pioneering work, and the emergence of modern AI technologies [2].

A. Blaise Pascal's Calculating Machine

In 1642, Blaise Pascal, a French mathematician, constructed the Pascaline, a mechanical device designed for performing basic arithmetic calculations. While not an AI system by today's standards, the Pascaline marked the earliest recorded attempt to automate mathematical operations, setting the stage for the automation of more complex tasks in the centuries to come [3].

B. Ada Lovelace and the Analytical Engine

The year 1837 brought forth another significant development in the history of AI, courtesy of Ada Lovelace. Lovelace, a mathematician and writer, collaborated with Charles Babbage on his design for the Analytical Engine [2].

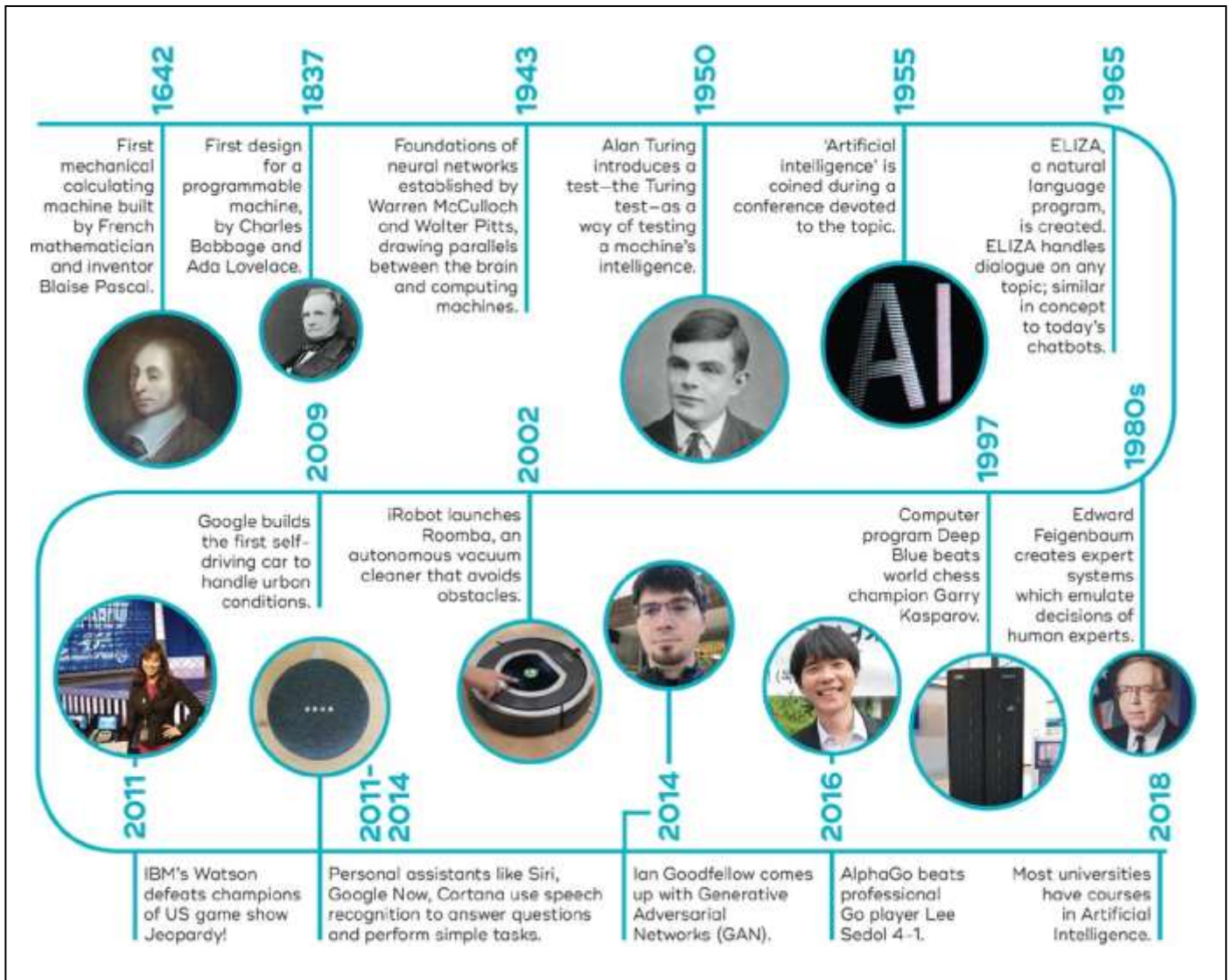


Fig. 1. History of Artificial Intelligence [4].

In her notes on the machine, Lovelace proposed the concept of using it to generate not only mathematical results but also music and art, effectively envisioning the idea of a programmable machine capable of creative output.

C. McCullough and Pitts' Logical Calculus of Ideas

The true commencement of AI research as a formal discipline is often attributed to the publication of "A Logical Calculus of Ideas Immanent in Nervous Activity" by Warren McCullough and Walter Pitts in 1943. This work introduced the first mathematical model of neural networks, laying the groundwork for the future development of artificial neural networks that underpin modern AI systems [5].

D. Alan Turing and the Turing Test

Alan Turing, a British mathematician and computer scientist, made a profound contribution to the field of AI in 1950 with the proposal of the Turing Test.

This test was designed to assess a machine's ability to exhibit intelligent behavior indistinguishable from that of a human. It remains a fundamental concept in AI and a benchmark for evaluating machine intelligence [6].

E. The Birth of "Artificial Intelligence"

The term "artificial intelligence" was coined in 1955 at the Dartmouth Conference, where a group of visionary researchers gathered to discuss the possibilities of creating machines that could simulate human intelligence. This event marked the birth of the formal field of AI and initiated a wave of research and development [7].

F. ELIZA: The Early Chatbot

In 1965, the first rudimentary chatbot, ELIZA, was created. Designed by Joseph Weizenbaum, ELIZA could engage in text-based conversations on various topics, foreshadowing the chatbots we encounter today. It was an early exploration of natural language understanding [8].

G. Edward Feigenbaum's Expert Systems

Edward Feigenbaum's work in the 1980s led to the creation of expert systems, AI programs that could mimic the decision-making processes of human experts in specific domains. These systems revolutionized fields like medicine and finance by providing expert-level guidance [9].

H. Deep Blue vs. Garry Kasparov

In 1997, IBM's Deep Blue computer defeated Garry Kasparov, the reigning world chess champion, in a historic match. This victory demonstrated the potential of AI to excel in complex tasks and set the stage for further advancements in AI gaming and decision-making [10].

I. iRobot's Roomba

In 2002, iRobot launched Roomba, an autonomous vacuum cleaner capable of navigating and cleaning a home while avoiding obstacles. Roomba exemplified the practical applications of AI in everyday life and became a precursor to modern robotics [11].

J. Google's Autonomous Vehicle

Google introduced its autonomous vehicle project in 2009, showcasing a car that could navigate urban environments with minimal human intervention. This marked a significant step towards the development of self-driving cars and furthered the integration of AI into transportation [12].

K. IBM's Watson Triumphs in "Jeopardy!"

IBM's Watson supercomputer made history in 2011 by defeating human champions in the television quiz show "Jeopardy!". Watson's success highlighted advancements in natural language processing and machine learning.

L. Ian Goodfellow's Generative Adversarial Networks (GANs)

In 2014, a watershed moment in the history of artificial intelligence occurred when Ian Goodfellow, a computer scientist and machine learning researcher, introduced Generative Adversarial Networks (GANs). GANs represent a groundbreaking concept in the field of AI and machine learning, particularly in the domain of generative modeling.

A GAN consists of two neural networks, a generator, and a discriminator, engaged in a continual process of competition. The generator attempts to create content, such as images, audio, or text, while the discriminator evaluates the generated content for authenticity. The key innovation lies in their adversarial relationship, where the generator seeks to improve its output to deceive the discriminator, and the discriminator strives to become better at distinguishing real from generated content [13].

This dynamic interaction leads to a rapid refinement of the generator's ability to produce high-quality, realistic data. GANs have found applications in image synthesis, style transfer, data augmentation, and anomaly detection, among others. They have been instrumental in creating life-like images, solving complex problems in data generation, and

have significantly contributed to the advancement of AI-driven creativity and problem-solving. The introduction of GANs has fundamentally expanded the boundaries of what AI systems can achieve and has had a profound and lasting impact on the field of artificial intelligence.

M. AlphaGo vs. Lee Sedol

In 2016, Deep Mind's Alpha Go, an AI system, defeated the world champion Go player, Lee Sedol, in a 4-1 victory. This marked a significant milestone in AI's ability to master complex strategy games, demonstrating the power of deep reinforcement learning.

N. 2018 to Present - The Rise of AI Education

In recent years, the growth of AI has led to the inclusion of AI courses in most universities, reflecting the increasing importance of AI in today's world. As of 2018 and continuing to the present day, educational institutions worldwide offer courses in artificial intelligence, fostering the next generation of AI researchers and practitioners.

The evolution of artificial intelligence bears witness to our innate human curiosity, innovation and persistence. Its origins can be traced back to Pascal's mechanical calculator, progressing to the modern age of GANs, self-driving cars and AI education, where significant advancements have been made. The progress AI has achieved is truly remarkable and continues to evolve. The legacy of these milestones will shape the future of AI.

III. THE VARIOUS FORMS OF AI

Artificial Intelligence (AI) can be categorized into different forms based on its capabilities and functions [14]. The main forms of AI include:

A. Narrow AI (Weak AI)

This type of AI is designed for a specific task or a narrow range of tasks. It operates within a limited domain and is not capable of generalizing its knowledge or skills to other areas. Examples include virtual personal assistants like Siri and Alexa, recommendation algorithms on streaming platforms, and image recognition software.

B. General AI (Strong AI)

General AI refers to a form of artificial intelligence that possesses human-like cognitive abilities. It can understand, learn, and apply knowledge across a wide range of tasks and adapt to new situations, much like a human being. General AI has not been achieved yet and remains a goal for future AI development.

C. Superintelligent AI

Super intelligent AI, also known as artificial super intelligence, represents a level of AI that surpasses human intelligence in every aspect. It is hypothetical and theoretical, as it would have the ability to outperform humans in virtually all intellectual tasks and potentially surpass human capabilities in creative thinking, problem-solving, and more.

IV. HOW AI WORKS ?

In the context of human learning, various methodologies are employed, drawing upon different paradigms. These encompass imitation, where individuals, particularly children, emulate observed behavior; association, which involves linking stimuli and responses to create learned associations; trial and error, where learning occurs through repeated experimentation; explanation, which entails understanding through elucidation; repetition, which reinforces knowledge through continued practice; immersion, as seen in language acquisition where learners immerse themselves in a language environment; and combination, which integrates multiple learning approaches to optimize the learning process.

Machines perceive learning through a process that involves gathering and analyzing data, identifying patterns, and adapting their behavior or performance based on the acquired knowledge.

The main components of machine learning are presented below.

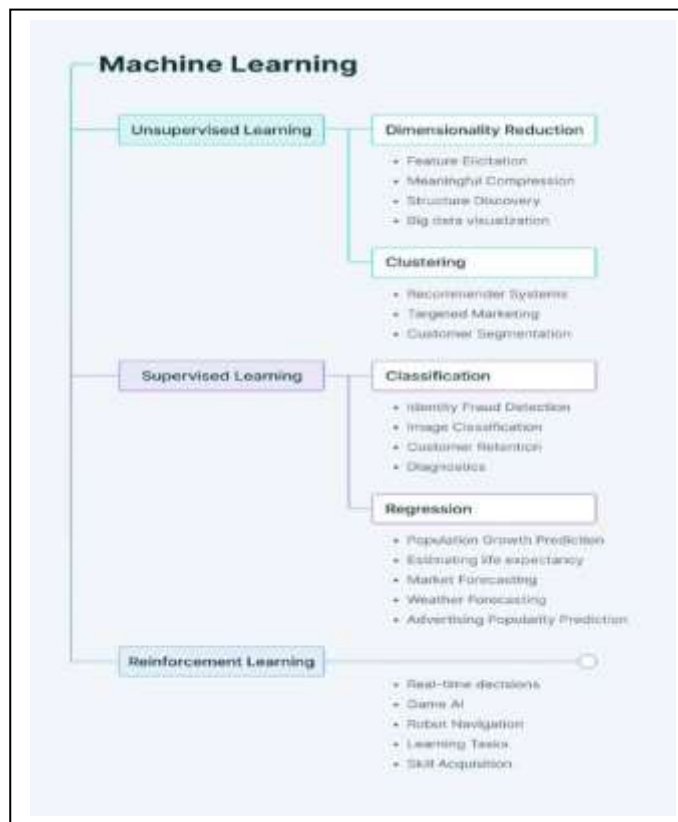


Fig. 2. Machine Learning Techniques [15].

A. Machine Learning Algorithms

Use statistical and data processing techniques to find models and rules for solving complex problems.

- Supervised learning is a machine learning approach where an algorithm is trained on a labelled dataset containing both input and corresponding target or output values. The algorithm learns how to make

predictions or classifications by generalising from the provided examples. The model is guided by the supervision of the labelled data, and its goal is to minimise the error between its predictions and the true labels. Supervised learning is a fundamental procedure in various real-world applications like classification, regression and object detection. It facilitates the discovery of relationships and insights from complex data sets, and in particular, in natural language processing and image recognition.

- Unsupervised learning, on the other hand, involves training algorithms on data sets that lack labeled output. The methodology of this learning paradigm aims to uncover structures or models within the data that are innate. Clustering and dimensionality reduction are common applications of unsupervised machine learning, wherein grouping similar data points or reducing data complexity discloses underlying relationships. This strategy is particularly useful in customer segmentation, anomaly detection and data compression where valuable insights can be extracted from unstructured data.
- Reinforcement learning constitutes a unique machine learning paradigm in which agents acquire proficiency in interacting with an environment and ascertaining the optimal decisions to enhance cumulative rewards. The agent procures feedback in the form of rewards or penalties from its actions, thus facilitating the acquisition of optimal strategies through trial and error. This approach to learning is common in fields such as game playing, robotics, and autonomous systems since it enables agents to evolve and enhance their behavior over time, resulting in effective decision-making even within changing and uncertain circumstances.

B. Neural networks

These are designed to mimic the way in which the human brain works.

An artificial neural network operates as a computational model inspired by the neural structure of the human brain. It comprises interconnected layers of nodes or artificial neurons that process and change input data. Each connection correlates with a weight, and the network's output derives from the weighted amount of inputs, proceeded by the activation function's application. While training, a significant mechanism referred to as back-propagation comes into play. This process entails iteratively adjusting the connection weights based on the computed error between the network's predictions and the actual target values. The error is then propagated backward through the network, facilitating the update of the weights in each layer to reduce the prediction error. Additionally, common optimization techniques, like gradient descent, are employed to refine these weights by identifying the optimal combination that minimizes the network's loss function. This process continues iteratively until the network reaches a satisfactory level of accuracy, rendering artificial neural networks a potent instrument for various machine learning assignments.

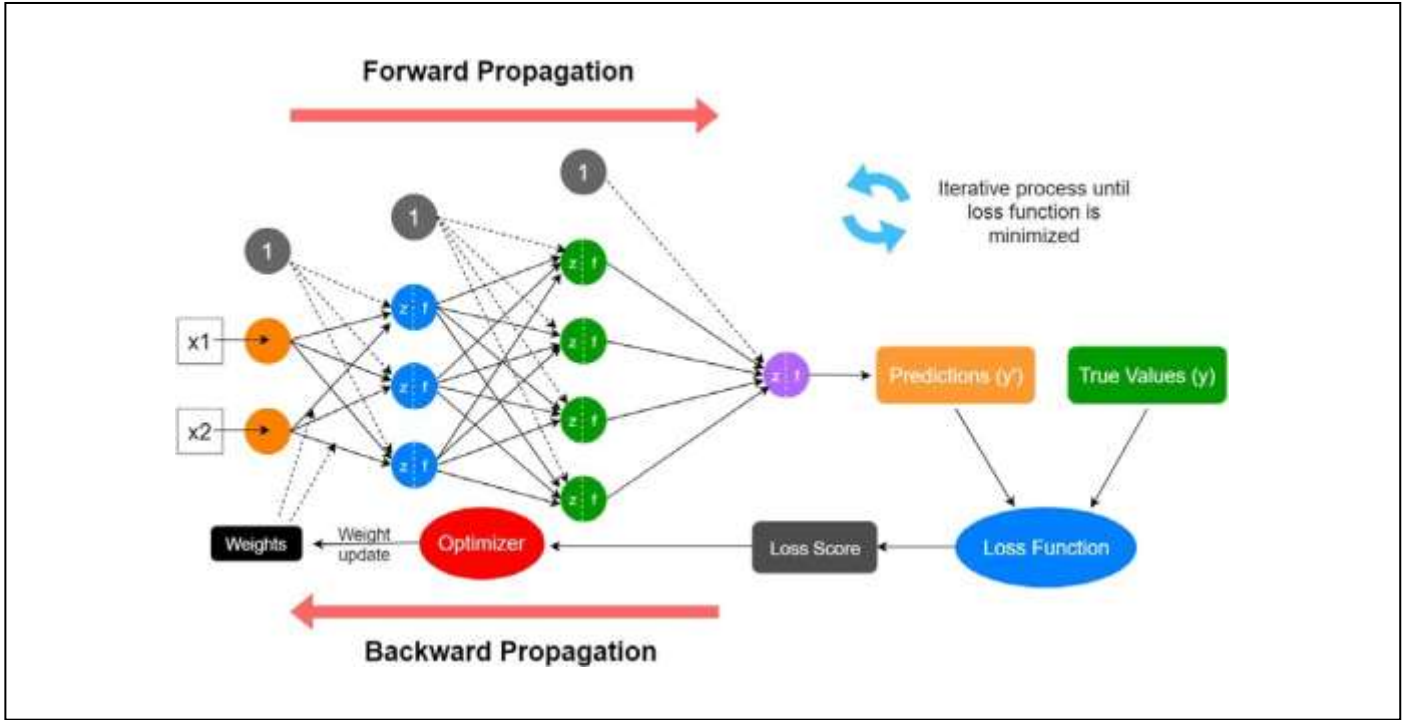


Fig. 3. Training Techniques for ANNs [16].

C. Data and its role in AI

Data is of paramount importance as it enables systems to enhance their functionality. The higher the volume of data, the more AI can learn and improve.

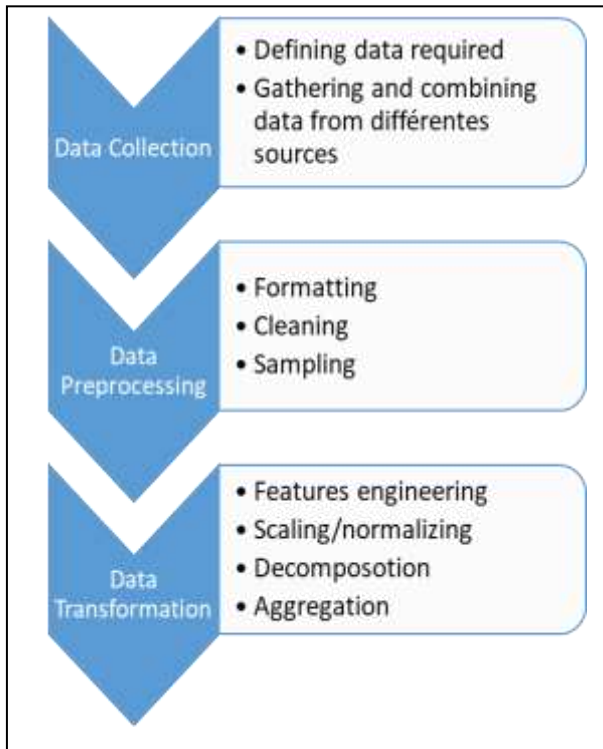


Fig. 4. Data Preparation Process [17].

Data is crucial in improving the machine learning models' efficiency. The primary key is the data collection process, which delivers the raw materials. The quantity, quality, and diversity of the data gathered have a significant impact on the model's ability to generalize and make accurate forecasts. Effective data pre-processing is crucial to guarantee high-quality data that machine learning algorithms can effectively use. This involves cleaning, handling missing values, and standardising data. Equally important is data transformation or feature engineering, which enables the extraction of relevant information from the data, allowing the model to capture intricate patterns and relationships. Well-designed features have the potential to significantly improve a model's predictive abilities. Ultimately, the complete data pipeline, encompassing everything from data collection to pre-processing and transformation, is crucial for achieving optimum machine learning performance and providing valuable insights.

V. AI APPLICATIONS

Artificial intelligence (AI) is a transformative field with diverse and far-reaching applications across numerous domains [18].

A. In daily life, individuals:

- Frequently receive suggestions for films, TV shows, and music through streaming platforms like Netflix and Spotify.

- Personalised advertisements and product recommendations are also common on online retailers such as Amazon and eBay.
- Additionally, voice assistants like Siri and Alexa employ natural language processing to answer queries and complete tasks.

B. In industry:

- Is also employed in quality control systems to detect defects in manufactured products in real time.
- AI systems are utilized to automate production lines through the use of industrial robotics.
- Additionally, predictive maintenance, which deploys AI to detect anomalies and potential breakdowns in industrial equipment before they occur, is gaining prominence.

C. In the field of medicine:

- AI has facilitated computer-aided medical diagnosis for quicker and more precise detection of diseases.
- Additionally, AI-based personalised medicine customises treatments based on the patients' genetic profile and medical background.
- Drug discovery utilizes AI to simulate numerous molecular combinations, leading to the discovery of new treatments.

D. In transportation:

- Autonomous cars, which use AI to detect and avoid obstacles, make real-time driving decisions and optimise the route.
- AI-based traffic management systems optimise traffic flow, decrease waiting periods and avert crashes.
- Autonomous drones used to monitor pipelines, critical infrastructure and hard-to-reach areas.

E. In services:

- Chatbots that use AI to answer customer questions, solve common problems and improve the customer experience.
- AI-powered product recommendation systems can identify and suggest corresponding or analogous products based on the user's purchase history and preferences.
- Personal assistants that use AI to manage schedules, send reminders and carry out everyday tasks for users.

VI. THE ETHICAL CONCERNS

A. Data Security and Privacy

Data security and privacy are ethical concerns of utmost importance in the field of artificial intelligence and

technology. Protecting privacy is crucial, as it ensures the safety of personal data and empowers individuals to exercise their autonomy. Failure to adequately safeguard data may result in serious repercussions, such as identity theft, surveillance, and exploitation.

For instance, take into account a scenario where an AI system is implemented in the healthcare sector to scrutinize patient data for diagnostic purposes. If the system fails to provide adequate protection for patient information privacy, it may result in unauthorised access to sensitive medical records. Such a privacy violation may expose private health information to malicious actors and lead to identity theft or discrimination based on medical history.

Ensuring data privacy is crucial, not only to safeguard individual rights but also to maintain trust in the AI systems and the organisations employing them. If individuals have a lack of confidence in the safety of their data, they may be disinclined to share it. This could impede the advancement and implementation of AI technologies, which could potentially provide significant advantages in healthcare, finance, and other fields. Hence, guaranteeing data security and privacy is an essential ethical consideration in AI utilisation

B. Impact on employment

Ensuring data privacy is crucial, not only to safeguard individual rights but also to maintain trust in the AI systems and the organisations employing them. If individuals have a lack of confidence in the safety of their data, they may be disinclined to share it. This could impede the advancement and implementation of AI technologies, which could potentially provide significant advantages in healthcare, finance, and other fields. Hence, guaranteeing data security and privacy is an essential ethical consideration in AI utilisation.

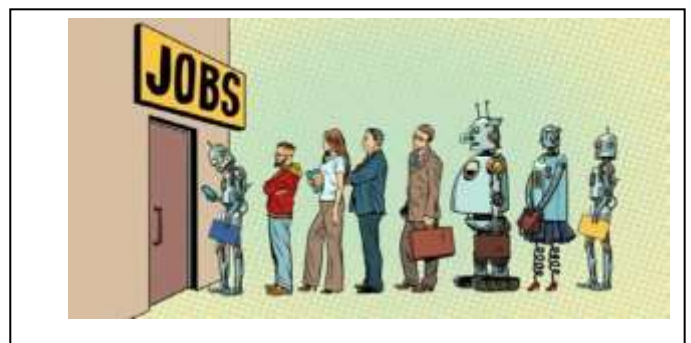


Fig. 5. Will AI one day steal our jobs and replace us ? [19]

The potential for job displacement and economic disruption is a major concern regarding the impact of AI on employment. AI technologies, such as automation and robotics, possess the ability to execute tasks that were once only accomplished by humans. This could lead to a rise in unemployment in specific industries. Algorithmic Biases (summarize as follows: "Algorithms can exhibit bias because

they are often trained on data that reflects societal prejudices and inequalities.

For instance, a recruitment algorithm trained on historical data may perpetuate gender or race biases that have existed in the industry. AI must be cognizant of this risk and take steps to mitigate biases, such as using more balanced data or employing transparent analysis methods.

TABLE I. PROFESSIONS SUSCEPTIBLE TO AI DISRUPTION

Tasks	
<i>Repetitive and routine</i>	<i>Reasoning, Interpretation</i>
<ul style="list-style-type: none"> • Data entry operators, cashiers. • Accountants. • Drivers, delivery drivers, • Estate agents. • Journalists. • Manufacturing industry workers. 	<ul style="list-style-type: none"> • General practitioners, pathologists, radiologists, surgeons, pharmacists. • Musicians, Artists. • Lawyers. • Teaching!

For example, the manufacturing industry has experienced considerable job losses as a result of the automation of assembly lines and procedures. Robots can function constantly, without interruptions or the necessity for benefits, which renders them cost-effective for businesses but also displaces human workers.

Another instance is the conceivable influence on customer service and call centre positions. AI-powered chatbots and virtual assistants are increasingly employed for handling routine customer inquiries, which leads to a decline in the need for human customer service representatives. This change in the job market presents ethical issues regarding the obligation of businesses and society to retrain and provide support to workers who are vulnerable to losing their jobs.

It's crucial to highlight that whilst AI can lead to job displacement in certain industries, it can also pave the way for new job opportunities in AI development, maintenance and oversight. Addressing ethical concerns surrounding AI's impact on employment necessitates taking proactive measures to retrain and upskill the workforce, as well as implementing policies to guarantee a just and equitable transition for affected workers.

C. Algorithmic biases

The ethical concern regarding algorithmic biases stems from the fact that algorithms, often trained on data that reflects societal prejudices and inequalities, can continue and even worsen these biases. For instance, a recruitment algorithm trained on historical data could inadvertently reinforce gender or race biases that were prevalent in the industry's past practices. It is vital that AI systems recognise the inherent risk and take action to curb biases. This can be achieved through measures such as incorporating balanced and representative data or adopting transparent analysis methods [20].

The growing utilization of big data bears potential opportunities as well as risks. To facilitate regulatory practices, it is necessary to objectively identify and classify the risks associated with AI systems.

There are four categories of risk: unacceptable risk, which includes AI applications that may be manipulative or involve social scoring (these AI systems are prohibited); high-risk, such as medical devices, which are allowed but only under strict legal requirements; limited risk, such as chatbots, which are required to have transparency obligations in place (users need to be informed that AI is being deployed); and minimal or no risk, such as video games, which are largely unregulated. The development of AI with user safety in mind is imperative. The AI Act aims to provide a comprehensive policy that covers all types of AI, including those that have yet to be developed. Criticism has arisen regarding the potential regulatory overreach that could result from its enforcement in all member states, which may hinder innovation and attractiveness to the EU. But the aim is for establishing these guidelines to offer companies and start-ups a clear plan during a period of unparalleled technological advancement, while upholding the fundamental rights of citizens.

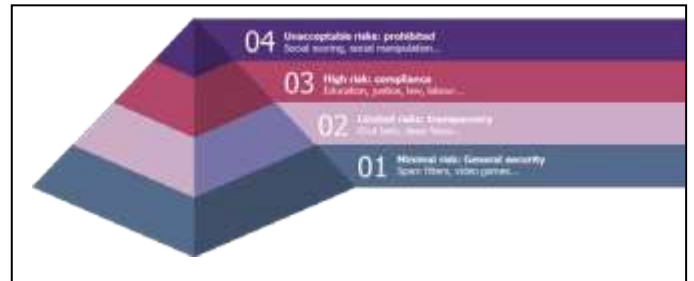


Fig. 6. Levels of Risk [21]

VII. OUTLOOK FOR THE FUTURE

A. Future prospects for AI

The promising future of AI is expected to yield substantial impact across various sectors:

- Smarter virtual assistants and sophisticated connected devices are set to integrate
- AI will be increasingly integrated into our daily lives, with smarter virtual assistants and more sophisticated connected devices.
- Moreover, self-driving cars are anticipated to become increasingly prevalent, which could decrease road accidents and provide greater mobility support for individuals with mobility impairments.
- AI could be employed to tackle intricate issues in fields such as healthcare, ecology, power and farming.

B. The challenges ahead

In addition to data confidentiality, impact on employment, bias and prejudice.

- Ethical issues related to the use of AI need to be addressed; the emergence of instruments like ChatGPT has greatly revolutionised multiple professional fields by elevating content production and optimising communication. Additionally, ChatGPT has profoundly impacted the education sector, with former conventions like homework losing their significance as pupils are now able to obtain immediate help through automated text production. This has caused a blurring of boundaries between autonomous student undertaking and external guidance. Moreover, the assessment of students has become increasingly intricate as the legitimacy of their contributions may be impacted by the widespread use of AI-based aid. As such, there is an urgent need to reassess traditional educational practices and redefine evaluation approaches to ensure just and valuable assessments in an ever-more AI-enhanced learning milieu.

C. Opportunities for research and innovation

- Opportunities for academic research and innovation arise in developing more advanced and sophisticated AI systems capable of learning from increasingly complex and varied data.
- Further applications of AI, in fields including medicine, agriculture, the environment and security, hold potential for research and development.
- Addressing issues of bias and prejudice within AI algorithms is a key concern, alongside improving the transparency and accountability of these systems.

VIII. CONCLUSION

In summary, this paper offers an all-encompassing survey of the different types of artificial intelligence, explaining the inner mechanisms of AI systems, exploring their different applications within multiple sectors, and carrying out a critical examination of the ethical quandaries associated with the widespread prevalence of AI technology. The findings presented highlight the multifariousness of AI and its ubiquitous influence on modern-day society. The potential of AI presents innovative solutions but also ethical dilemmas. Therefore, continued research, ethical considerations, and regulatory measures are essential to optimize AI whilst mitigating its risks. As AI evolves, academia, industry, and government stakeholders must collaborate to foster responsible AI development, ensuring maximum benefit and ethical concerns mitigation.

IX. References

- [1] A. Aldoseri, K.N. Al-Khalifa, Hamouda, A.M. "Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges," *Appl. Sci.* 2023, 13, 7082. <https://doi.org/10.3390/app13127082>
- [2] LC. Aiello, "The multifaceted impact of Ada Lovelace in the digital age," *Artificial Intelligence*. 2016 Jun 1;235:58-62.
- [3] J. Schäfer, *A Brief History of Computing* 2010.
- [4] "History of Artificial Intelligence," *The University of Queensland Australia*. May 2023, <https://qbi.uq.edu.au/brain/intelligent-machines/history-artificial-intelligence>.
- [5] WS. McCulloch, Pitts W. "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*. 1943 Dec;5:115-33.
- [6] B. Gonçalves, "The Turing test is a thought experiment. *Minds and Machines*," 2023 Mar;33(1):1-31.
- [7] M. Van Assen, E. Muscogiuri, G. Tessarin, CN. De Cecco, "Artificial Intelligence: A Century-Old Story," *In Artificial Intelligence in Cardiothoracic Imaging* 2022 Apr 22 (pp. 3-13). Cham: Springer International Publishing.
- [8] H. Shah, K. Warwick, J. Vallverdú, D. Wu, "Can machines talk? Comparison of Eliza with modern dialogue systems," *Computers in Human Behavior*. 2016 May 1;58:278-95.
- [9] EA. Feigenbaum, "A personal view of expert systems: Looking back and looking ahead," *Knowledge Systems Laboratory, Department of Computer Science, Stanford University*; 1992 Apr.
- [10] FH. Hsu, "Behind Deep Blue: Building the computer that defeated the world chess champion," *Princeton University Press*; 2002.
- [11] J. Forlizzi, C. DiSalvo, "Service robots in the domestic environment: a study of the roomba vacuum in the home," *In Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction* 2006 Mar 2 (pp. 258-265).
- [12] SL. Poczter, LM. Jankovic, "The google car: driving toward a better future?," *Journal of Business Case Studies (JBACS)*. 2014;10(1):7-14.
- [13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "Generative adversarial networks," *Communications of the ACM*. 2020 Oct 22;63(11):139-44.
- [14] R. Fjelland, "Why general artificial intelligence will not be realized," *Humanities and Social Sciences Communications*. 2020 Jun 17;7(1):1-9.
- [15] Pragati Baheti, "What is Machine Learning?," *The Ultimate Beginner's Guide*. ©V7Labs • June 1, 2021. <https://www.v7labs.com/blog/machine-learning-guide>
- [16] R. Pramoditha, "Overview of a Neural Network's Learning Process," *Published in Data Science* 365. Feb 1, 2022. <https://medium.com/data-science-365/overview-of-a-neural-networks-learning-process-61690a502fa>
- [17] S. García, J. Luengo, F. Herrera, "Data preprocessing in data mining. Cham, v Switzerland: Springer International Publishing; 2015.
- [18] OpenAI. (2023). ChatGPT (Mar 14 version) [Large language model]. <https://chat.openai.com/chat>
- [19] R. Woodbury, "Will AI Take Our Jobs? Exploring AI's Impact on Society," *5 avr. 2023. Will AI Take Our Jobs? Exploring AI's Impact on Society*. <https://www.digitalnative.tech/p/will-ai-take-our-jobs-exploring-ais>
- [20] A. Jobin, M. Ienca, E. Vayena, "The global landscape of AI ethics guidelines," *Nature machine intelligence*. 2019 Sep;1(9):389-99.
- [21] Loosely based on: Eve Gaumond, *Artificial Intelligence Act: "What is the European Approach for AI?"*, in *Lawfare*, June 2021, available at: <https://www.lawfareblog.com/artificial-intelligence-act-what-european-approach-ai>

Machine intelligence: from Turing test to generative pretrained transformers

Hadjila Fethallah
Computer Science Department
University of Tlemcen
fethallah.hadjila@univ-
tlemcen.dz

Merzoug Mohammed
Computer Science Department
University of Tlemcen
mohammed.merzoug@univ-
tlemcen.dz

Belabed Amine
Computer Science Department
University of Tlemcen
amine.belabed@univ-tlemcen.dz

Abstract—Artificial intelligence (A.I) has impressively evolved during the last seventy years, especially with the advances of machine learning. The objective of this paper is to present a survey on the chronological development of A.I from its early beginning up to the last breakthroughs. Namely, we will cover the Turing test ingredients, dark era of neural networks, resurgence of neural networks through back-propagation, ensemble learning, and deep learning. We also review the actual challenges and the eventual dangers that still face the research community.

1. Introduction

Understanding and simulating the human brain capabilities is one of the old objectives of the computer science community. It is worth noting that the human intelligence is a delicate term to define and it spawned much debate and ink to flow in the past decades. In 1994, a group of psychology researchers led by Linda Gottfredson proposed a definition of human intelligence that is specified as follows: “Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience.” [1]. Even if this definition is not a consensual one, we can use it to draft a first definition for the artificial intelligence (A.I) utterance. In this line of thought A.I is the simulation of the above-mentioned brain capabilities on machines. Consequently, some software systems may not be considered as intelligent (by other researchers) if they cannot fulfill all the capabilities mentioned above. The reasoning skill means the ability to think and compose inference rules and axioms to build new valid knowledge that can be considered as theorems. The planning skill means the ability to produce a sequence of actions that map an initial state of affairs into a goal

state. Thinking abstractly is a kind of reasoning or problem resolution that builds solutions or target objects using abstract input blocks, the target built objects must show the desired properties while masking the less important details. The comprehension of complex ideas means the ability to process the natural language (NLP) while performing all necessary steps in terms of acoustic/phonetic analysis, morphological analysis, syntactic analysis, semantic analysis, and pragmatic analysis. The last skill concerns the ability of training and acquiring new knowledge and competencies through the rewiring (connections’ update) of the brain, we can distinguish off-line learning (which may consume much time before the effective utilization) and real-time learning which must meet the delays imposed by the current problem.

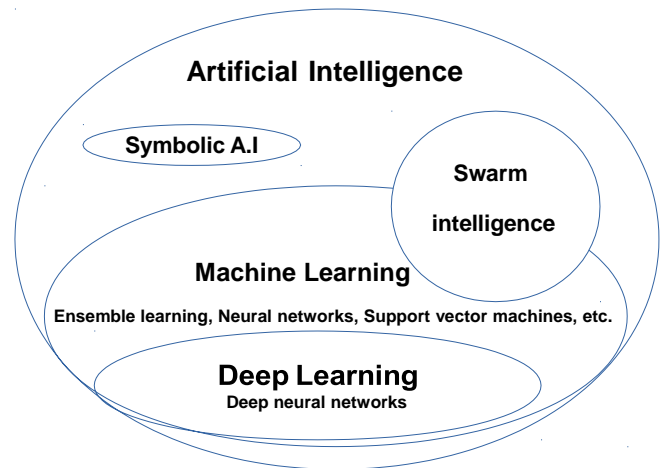


Figure 1. A.I ingredients.

In Figure 1, we point out that A.I is mainly constituted

of symbolic intelligence, machine learning, and emergent intelligence (i.e., swarm and connectionist intelligence). This entire AI may pass different levels of perfection. We distinguish narrow AI (or specialised /weak A.I), Strong A.I sometimes called Artificial general intelligence (i.e., AGI) and finally Super A.I. Narrow AI is actual state of machine intelligence. It mainly means that the developed (smart) softwares are better than a human in a specific task and a specific context. For instance, in the object recognition field, we witnessed many models (starting from 2015) that bypass the human capabilities (see Figure 4). According to Ray Kurzweil [2], the term “narrow AI” refers to the creation of systems that carry out specific “intelligent” behaviors in specific contexts. As concrete Narrow AI systems, we can cite object recommender systems, self-driving vehicles, Chatbots, and IBM Watson.

We can find other artificial systems that are way better than human being in a given task, but this cannot be generalised to all tasks. The bottom line is that narrow A.I is better than humans in a single task (or related tasks in the extreme case), but the human is better than narrow A.I when we consider all tasks. In contrast to that, AGI or strong A.I is a kind of software systems that have human like capabilities in a variety of tasks and contexts. Their generalisation capability is as good as that of humans, and they may surpass them in some tasks. Goertzel describes AGI as follows: “General intelligence involves the ability to achieve a variety of goals, and carry out a variety of tasks, in a variety of different contexts and environments” [3]. These tasks include the skills found in the Turing test and they involve perception, reasoning, planning, real-time /batch learning, and natural language processing. Until now, we have not such an existing AGI system; researchers estimate that the future versions of GPTs (Generative pre-trained Transformers) may have these capabilities or at least a percentage of them around 2050. Artificial Super Intelligence ASI is the last level of A.I. In this case, ASI is better than human in every task and every context. It also enjoys the consciousness capability and other kinds of feelings. Certainly ASI will constitute a real danger for the other species, including humanity.

To meet the aforementioned capabilities A.I has gone through a remarkable path and has witnessed many ups and downs. The A.I history begins with the works of Warren McCulloch and Walter Pitts in 1943 that created the first artificial neuron. Then, the Turing test specified the main ingredient of A.I in 1950, and in this perspective, we witnessed the effective birth of A.I with the Dartmouth Workshop in 1956. This conference laid the foundation for symbolic A.I and logic-based reasoning which flourished in the 1970s and 1980s. Almost in the same period (1960s- 1970s), the dark era of machine learning (or connectionist

A.I winter) is experienced with the book of Minsky and Papert, this latter one highlighted the main drawbacks of the Perceptron neural network and encouraged the abandonment of works in this field. The resurgence of neural networks and connectionism in the mid 1980s constituted a pivotal point in the A.I life-cycle and led to the modern era of data-driven

machine learning; more specifically, the period of the 1990s and 2000s witnessed the emergence of effective algorithms like multi layer perceptrons [4], support vector machines [5], and decision trees [6]. Nonetheless, the true renaissance of A.I unfolded after the 2010s, as A.I delved into the world of deep learning. Many remarkable frameworks of deep neural networks, such as convolutional neural networks (CNNs) [7], [8], recurrent neural networks (RNNs), [9], [10] and transformers [11], have revolutionized critical fields like computer vision, speech recognition, and natural language processing. Nowadays A.I systems are providing brilliant solutions that transform a multitude of sectors (including healthcare, industry, and cities) into a highly smart ecosystem that eases and promotes the life of human beings. In this work, we mainly discuss the chronological development of A.I while showing the most eminent successes of A.I softwares. We also review some challenges, as well as some threats that must be handled in order avoid the eventual deviation of A.I systems and other potential misuses.

The rest of the paper is structured as follows, in Section 2, we review the main stages through which A.I has gone. In Section 3, we show some prominent successes of A.I systems. In Section 4, we highlight the challenges and threats of A.I. In the last Section, we give a summary about the future applications of A.I.

1. A.I CHRONOLOGICAL PROGRESS

The AGI level of A.I can be realized using various architectures, according to Duch’s overview [12], there are three types of approaches (or paradigms) to implement machine intelligence, we have the symbolic way, emergentist way and hybrid way (this view is approximately depicted in Figure 1). Even though this categorization is not a consensual one, it brings an initial view about the progress path of A.I. In what follows, we will review the main stages through which A.I has gone.

1.1. EARLY FOUNDATIONS (1950s-1960s)

In 1950, Turing proposed one of the most accepted definitions for A.I, the main idea consists of performing a test in which there are three entities that communicate through a network (see Figure 2):

- Human interrogator: his role is to ask a series of questions written in natural language to both interlocutors (the human and the machine). The interrogator does not know the exact the location of both interlocutors and must only use the answers to perform the distinction.
- Human (interlocutor): his role is to answer the questions via the network and using written natural language.
- Machine (interlocutor): its role is to answer the questions via the network and using written natural language.

If the interrogator can not distinguish between the two interlocutors after the end of questions' session, then the machine is qualified to be intelligent. To succeed this test, the machine must be endowed with various skills including natural language processing, knowledge representation and reasoning, learning from examples. Moreover, the advanced version of the Turing test must also include computer vision in order to process multimedia data, and also robotic parts to perform real world actions and planning.

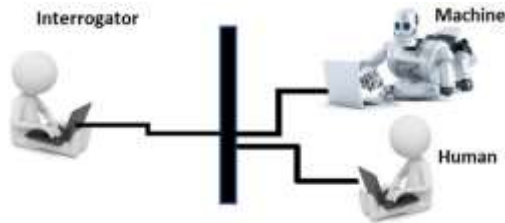


Figure 2. Turing test.

1.2. A.I DARK ERA AND EXPERT SYSTEMS (1960S- 1980S).

In 1956, a group of researchers organized the first conference on A.I in Dartmouth and this latter one opened the door for the advances in the field of symbolic A.I. In particular, many theorem provers and general problem solvers have been developed. Moreover, the formalization of the human expertise in terms of rules and other knowledge formalisms led to the creation of rule-based systems and knowledge-based systems. As an example, we cite the MYCIN system which is developed in the early 1970s. MYCIN is a rule-based system that uses backward chaining in order to deduce the bacteria that are involved in many infection diseases including septicemia and meningitis. In the same period (1960-1980), the "machine learning winter" entered the scene with the work of Minsky and Papert in the late 1960s. According to these authors, the connectionist approaches (Perceptrons) are very limited and can't even learn simple functions such as XOR. Due to this work, the research in this field had been frozen for at least fifteen years.

1.3. CONNECTIONISM, NEURAL NETWORKS, AND STATISTICAL LEARNING THEORY (1980S-1990S)

In 1986, Geoffrey Hinton and his colleagues developed a method called "back-propagation" to train multilayer neural networks, and this allows the learning of complex mathematical functions and ensures the ultimate objective of universal approximation. On the other hand, an innovative learning theory was proposed in 1984 by Leslie Valiant in order to mathematically analyse machine learning algorithms and models. This theory (also called probably-approximately correct (PAC) learning) [13] specifies the necessary conditions that help building learning models with a good generalization capability. Starting from this theory, researchers

have built many learning models with guaranteed performance (and generalization power) including support vector machines and ensemble learning methods. It is worth noting that this type of machine learning models have realized a spectacular success on small size and medium size problems such as handwritten letters and digits recognition.

1.4. THE RISE OF MACHINE LEARNING (1990S-2000S)

This era is mainly marked by the creation and development of models that respect the PAC learning paradigm and mainly the successful development of support vector machines and its upgrades in the 1990s. The genesis of ensemble learning in the late 1990s and the early 2000s added a powerful layer in machine learning and reinforced the PAC learning paradigm. Both the boosting family which essentially represented by Adaboost [14] and the bagging family which is predominantly represented by Random forest [15] are able to convert a set of simple learning models (called weak learners) into a strong learner which satisfies the PAC learning objectives.

1.5. DEEP LEARNING REVOLUTION (2010-PRESENT)

With advent of complex tasks such as perception challenges (in which tremendous and disparate types of data must be effectively handled by A.I models), the inefficiency of current machine learning methods such as SVM [5], hidden markovian models (HMM) [16], and multilayer perceptrons (MLP) [4] was clear, especially in computer vision. For instance, in object recognition, the best models, which preceded the deep learning era, obtained only an error rate equal to 25.8% on the ImageNet challenge [17]. In this perspective, Alex Krizhevsky [8] proposed an effective Convolutional neural network (CNN) on ImageNet dataset and decreased the error rate to 16%. This spectacular progress invited all IT companies and research laboratories to turn their efforts to deep neural networks in all perception tasks. In the perspective of CNN, we mainly cite the following advanced architectures: inception networks [18], residual networks [19], and efficient networks [20] for object recognition; U-nets [21] and deeplab [22] for image segmentation. In the subsequent years, many research teams proposed their own deep neural networks in all supervised and unsupervised challenges, and each time a sufficient number of examples is available for a given issue (with an adequate computing power), there will be a successful performance that opens the doors for real breakthroughs. In the context of recurrent networks, we highlight the revival of long short term memory and its competitor Gated recurrent units in crucial tasks such as auto-completion and automatic translation. In 2015, the mechanism of attention [23] was introduced in sequence to sequence models (namely LSTM) to enhance the capabilities of translators and gain better results in handling longer sentences. Thereafter (in 2016),

1. <https://web.archive.org/web/20200907212153/http://image-net.org/about-stats.php>

the self-attention [24] concept was coined for single encoder networks in order to learn the context of each word of sentences. Afterwards (in 2017), the self-attention mechanism, residual blocks, and positional encoding were combined to create an innovative sequence to sequence model that impressively succeed in automatic translation. This new architecture was termed Transformer [11] and gained a lot of success in sequential data processing (see Figure 3).

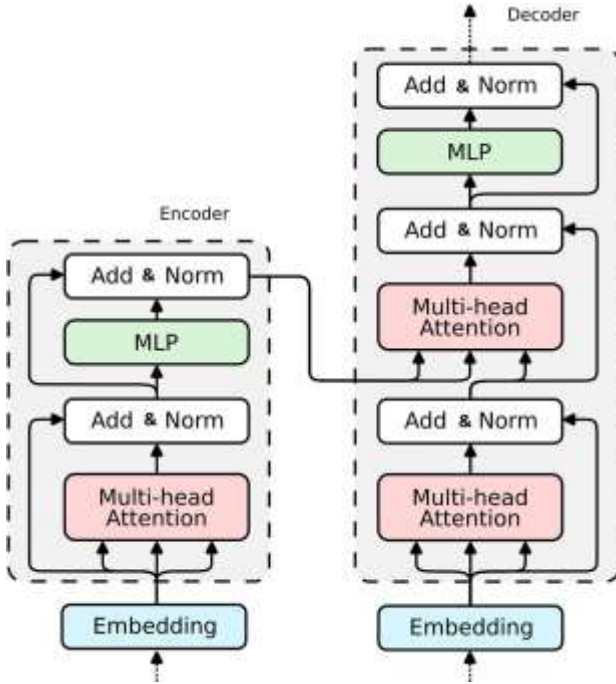


Figure 3. Transformer architecture [11].

After that, the transformer architecture is used to train a large language model or LLM (a kind of auto-regressive model that predicts the next token of a large collection of texts) and results in the first generation of generative pretrained transformers (GPT1) [25]. GPT1 is trained on the BookCorpus [26] dataset which consists of around 7000 books and a total size of 4.5 GB (the training data was not filtered). The training of GPT1 is done in two steps: first, the LLM is trained on unlabeled text (the self-supervised stage) to learn the initial version of the neural network weights; then, the network is fined-tuned to a specific target task using supervised learning. Consequently, the pretrained weights are re-adapted to the new task (which is mainly instruction following). The new versions such as GPT2 and GPT3 have a larger size of weights and a variety of data sources with respect to the first version. More specifically, GPT2 is trained on 40 G.B of human filtered text data and has a total of 1.5 billion parameters; while, GPT3 is trained on 450 G.B of human filtered data and has a total of 175 billions parameters. We notice that ChatGPT is an LLM that leverages either GPT3.5 or GPT4. It is trained in two stages: (1) an unsupervised learning for the auto-completion problem, (2) a combination of supervised learning and a

reinforcement learning for the instruction following task.

2. SOME A.I'S ACHIEVEMENTS

In this section, we present various A.I systems that created breakthroughs and bypassed the human capabilities in a specific field, we will skim over many fields such as ob- ject recognition, speech recognition, reinforcement learning (including complex games), and protein folding.

2.1. OBJECT RECOGNITION

Starting from 2015, deep neural networks including residual networks [19], efficient networks [20] (Efficient- Net), EfficientNet-L2 [27] and other models bypassed the human performance in identifying the main objects and categories present in images. We notice that the human performance is equal to an error rate of 5.1% on the Ima- geNet Large Scale Visual Recognition Challenge (ILSVRC). Figure 4 [28] shows the performance of a set of machine learning models in the ImageNet challenge. We notice that

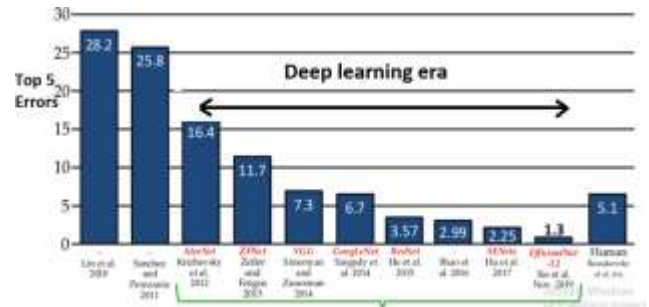


Figure 4. A.I models performance on ImageNet [28].

the first breakthrough was ensured by residual networks of Microsoft. This architecture was able to reach this success level through the use of residual blocks (RB). RB is a com- position of 02 convolutional layers (each one is a sequence of a dot product and a Relu activation function) in which we add a skip connection that connects the inputs of RB to the the Relu function of the second convolutional layer. The main motivation of RB is to allow the learning of identity functions that ensure the convergence of deeper CNN and consequently reaching a satisfactory training error.

2.2. SPEECH RECOGNITION

The automatic speech recognition (ASR) field has expe- rienced many A.I models that include connectionist temporal classification (CTC), attention-based encoder decoder, recur- rent neural networks (RNN), fully convolutional networks (FCN), transformers, and their combinations. The current models have an exceedingly increasing performance on well-known datasets such as the challenge of the Wall Street Journal dataset (WSJ) and the Librispeech dataset. We notice that WSJ is an english clean speech database having 80

hours, whereas, LibriSpeech is a dataset of reading speech from audio-books that contains 1000 hours of data. The following table shows a subset of the most effective models on ASR.

TABLE 1. PERFORMANCE OF SOME ASR MODELS.

Word-error-rate	Authors	Datasets	Clean validation set	Other set
RNN: [29]		LibriSpeech dataset	5.4%	15.6
		WSJ dataset	5.1%	8.4%
FCN: [30]		LibriSpeech dataset	3.08%	9.94 %
		WSJ dataset	3.5%	6.8 %

2.3. REINFORCEMENT LEARNING AND GAMES

The GO game (see Figure 5) is one of the most complex games in the history of humanity, its board contains 19 lines and 19 columns, and therefore there are 19×19 possible configurations to explore. In addition, the average branching factor (the number of moves) is 150. It is worth noting that the number of legal configurations of Go is close to 2.1×10^{170} , and this amount is higher than the number of atoms in the observable universe [31].



Figure 5. The board of Go.

In 2016, an A.I model based on deep reinforcement learning (AlphaGo) [32] of DeepMind company was trained on the Go game. The AlphaGo model leveraged deep convolutional networks and beat the world champion Lee Sedol with a score of 4-1.

2.4. PROTEIN FOLDING

In 2021, The deep neural network Alpha-fold of DeepMind succeeded in resolving the problem of protein folding (PF). PF is the shape prediction of protein from the amino acid sequence. Alpha-fold [33] won the standard competition of the issue known as Critical Assessment of Protein Structure Prediction (CASP 14th edition). Alpha-fold scored 244.0 compared to 90.8 which is the score gained by the next best group. The predictions made by AlphaFold had a median success rate of 92.4%.

3. A.I challenges and threats

Both narrow A.I and AGI are facing many challenges that are far from being positively handled at the current state. Likewise, some potential A.I threats are also pending and must be addressed by the both the research community and decision makers. As regards the challenges, we give an overview about the main obstacles that we face today:

- A.I bias and ethical issues: the training data used in large language models and GPT like models can be unbalanced and ethically deviated (since it reflects the convictions and hidden thoughts of materials' owners) and this lead to biases in the learned knowledge. For instance, A.I may generate negative content about specific ethnic groups or specific religions and this reinforces divisions in the same society.
- Robust A.I: the ability of learning reliable models despite the lack of labeled data, presence of noise, and contradictory data present in the web is a difficult task; in addition, the adversarial attacks may change the training examples in order to mislead the future decisions of A.I systems. In this perspective, the use of multimodal data is a way of producing reliable predictions and compensating the lack of labeled data, especially in medical fields. In the medical domain, the integration of multi-modal inputs not only enhances the accuracy of the diagnosis but also provides valuable explanations about the automated decisions and the forecastings. Moreover, theoretical frameworks such as probably-approximately-correct learning (PAC-learning) [13] constitute a solid way for dealing with the issues of robust I.A, but the theoretical conditions of the PAC framework are hard to meet in practical cases.
- Auto machine learning (AutoML): the ability of creating machines that not only learn good models for specific tasks, but also automatically fine-tune the hyperparameters including the activation functions, kernels' configurations, rearrangement of the cost function landscape. Moreover, AutoML involves the capability to perform transfer learning from previously learned problems to a wide range of new tasks. This human inspired ability allows for saving the time and resources of retraining models from scratch and therefore these past skills can be re-adapted to other problems and fields.
- Limited resources and energy: deep learning models and especially large language models (such as GPTs) are highly greedy in terms of computing power and time of training; in this regard, the requirement of resilient data centers with permanent energy of functioning and cooling is a prime necessity. This issue is exacerbated with the expensive and fluctuating prices of hydrocarbon-based energy; moreover, the green energy is still insufficient and unavailable in some periods due to environmental and atmospheric constraints.

Concerning the A.I threats, it is well known that the possible dangers may arise from the misuse of A.I technologies. In what follows, we mainly cover the prominent dangers that can affect the humanity and their life style.

- **A.I misuse and fake information:** One of the main misuse of A.I systems is the utilization of collected data about users (of mobile applications, web applications, and social networks) and the deduced knowledge to harm specific ethnic groups or minorities. In the same line of thought, some generative models may create fake data (images, audios, or videos) which are almost indistinguishable from real data, and stick them to other entities in order to defame or harm them. these fake data can also be used to falsify evidences and deviate the investigation paths in case of crimes or dereliction.
- **Job loss:** Traditionally, robotics and narrow I.A have tendencies to replace human workers by automated machines in routine tasks (e.g., manufacturing, mail filtering and ranking). But the category of target tasks is increasingly expanding, and it involves tasks such as data analysis or document review, newspaper report writing, novel writing, research paper writing, and medical imaging analysis. Consequently, job positions in domains like finance, lawyering, and journalism are at high risks. In the same regard, we observe that workers who have low-quality (and sometimes mid-quality skills) in some fields such programming are at risk, and they can be laid off by their companies and replaced by automated tools. This can lead to a growing gap between high-skilled and low-skilled workers and create an economical issue since the unemployment rate increases.
- **A.I and privacy concerns:** Nowadays, many A.I systems collect private and other sensitive data about users and institutions without their agreement. This type of data can be sold to third parties or can be used for training A.I models that can harm people and organizations in the future. Another concern is that even authorized institutions (such as hospitals) can use private data and learn vulnerable models that can be attacked by malicious entities. In this regard, model inversion attacks and membership inference attacks can be applied on vulnerable A.I models and infer sensitive information. To tackle these risks, researchers combine differential privacy [34], secure multiparty computation [35], federated learning [36], and homomorphic encryption [37] to alleviate the menaces.

4. Conclusion

In this paper, we have presented the definitions and the perfection levels of A.I. We have also reviewed the various stages of the A.I progress (From the Turing Test up to the actual GPT's). Furthermore, we have shown the impressive results of some existing A.I models and have given an

overview over the potential dangers of the future machine intelligence. The next questions must focus on the possible improvement and adaptation of future AGI Systems in our daily life and environment, especially in healthcare, smart cities (e.g., transport, smart buildings, energy generation, and government), agronomy and industry.

References

- [1] L.S.Gottfredson. Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence* 24, no. 1. 1997. 13-23.
- [2] R. Kurzweil The singularity is near. In *Ethics and emerging technologies*. London: Palgrave Macmillan UK. 2005. 393-406.
- [3] B.Goertzel. Artificial general intelligence: concept, state of the art, and future prospects. *Journal of Artificial General Intelligence*, vol. 5, no 1, 2014. p. 1
- [4] D.E. Rumelhart, G.E. Hinton and R.J. Williams. Learning internal representations by error propagation. Chapter 8 in *Parallel Distributed Processing: Foundations*. Vol. 1, MIT Press, Cambridge, MA, 1986. 318–362.
- [5] C. Corinna, and V.Vapnik. Support-vector networks. *Machine learning* 20 .1995. 273-297.
- [6] L.Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. *Classification and Regression Trees*, Wadsworth, Belmont, CA. Republished by CRC Press. 1984.
- [7] Y.LeCun, L. Bottou, Y.Bengio, and P.Haffner. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE* 86, no. 11. 1998. 2278-2324.
- [8] A. Krizhevsky, I. Sutskever, and G.E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25. 2012.
- [9] H.Sepp, and J. Schmidhuber. "Long short-term memory." *Neural computation* 9, no. 8. 1997. 1735-1780.
- [10] K.Cho, B.Van Merriënboer, C. Gulcehre, D.Bahdanau, F. Bougares, H.Schwenk, and Y. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.2014.
- [11] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and I.Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30. 2017.
- [12] D.Wlodzislaw, R.J. Oentaryo, and M.Pasquier. "Cognitive architectures: Where do we go from here?." *Agi*. Vol. 171. No. 1. 2008.
- [13] L.Valiant. Probably approximately correct: nature's algorithms for learning and prospering in a complex world. *Basic Books (AZ)*, 2013.
- [14] Y.Freund, R.E.Schapire: A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55(1), 1997. 119–139.
- [15] L.Breiman. Random forests. *Machine learning* 45. 2001. 5-32.
- [16] L. E.Baum; T.Petrie "Statistical Inference for Probabilistic Functions of Finite State Markov Chains". *The Annals of Mathematical Statistics*. 37 (6). 1966. 1554–1563. doi:10.1214/aoms/1177699147.
- [17] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.
- [18] C.Szegedy, W.Liu, Y.Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.2015. (pp. 1-9).
- [19] K.He, X. Zhang, S.REN, J.Sun. Deep residual learning for image recognition. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. p. 770-778.
- [20] M.TAN, Q.LE. Efficientnet: Rethinking model scaling for convolutional neural networks. In : *International conference on machine learning*. PMLR, 2019. p. 6105-6114.
- [21] O.Ronneberger, P.Fischer, and L.Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18 (pp. 234-241). Springer International Publishing.
- [22] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*. 2017 Apr 27;40(4):834-48.
- [23] D.Bahdanau, K.Cho, and Y.Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.2014.
- [24] J.Cheng, L.Dong, and M.Lapata. "Long short-term memory-networks for machine reading." *arXiv preprint arXiv:1601.06733*.2016.
- [25] A.Radford, K.Narasimhan, T.Salimans, I.Sutskever. Improving language understanding by generative pre-training. 2018.

- [26] Y.Zhu , R.Kiros, R.Zemel, R.Salakhutdinov, R. Urtasun, A.Torralba, S.Fidler, S. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In Proceedings of the IEEE international conference on computer vision. 2015. (pp. 19-27).
- [27] Q.XIE, M.T. LUONG, E.HOVY, et al. Self-training with noisy student improves imagenet classification. In : Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 10687-10698.
- [28] L.ALZUBAIDI, J.ZHANG, A.J.HUMAIDI, et al. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, vol. 8. 2021.p. 1-74.
- [29] T.Hori, J.Cho, S.Watanabe. End-to-end speech recognition with word- based RNN language models. In 2018 IEEE Spoken Language Technology Workshop (SLT). 2018. (pp. 389-396).
- [30] N.Zeghidour, Q.Xu, V.Liptchinsky, N.Usunier, G.Synnaeve, R.Collobert. Fully convolutional speech recognition. arXiv preprint arXiv:1812.06864. 2018.
- [31] "A Brief History of Go — British Go Association". www.britgo.org. Retrieved 2023-12-16.
- [32] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, et al. Mastering the game of go without human knowledge. nature 550, 7676. 2017. 354–359.
- [33] J.Jumper, R.Evans, A. Pritzel, T.Green, M.Figurnov, O.Ronneberger, K.Tunyasuvunakool et al. "Applying and improving AlphaFold at CASP14." Proteins: Structure, Function, and Bioinformatics 89, no. 12 .2021. 1711-1721.
- [34] C.Dwork. "Differential privacy." In International colloquium on automata, languages, and programming, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.pp. 1-12.
- [35] A.C. Yao. Protocols for secure computations. In Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science, 1982.
- [36] B. Keith, V. Ivanov, B. Kreuter, A. Marcedone, H. Brendan McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth. "Practical secure aggregation for privacy-preserving machine learning." In proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017. pp. 1175-1191.
- [37] R.L. Rivest, L.Adleman, and M.L. Dertouzos. On data banks and privacy homomorphisms. Foundations of Secure Computation 4, 11. 1978, 169–180.

EEG Signal Analysis for Automatic Detection of psychiatric diseases

Meryem Saidi

Higher School of Management of Tlemcen
Biomedical Engineering Laboratory, Tlemcen University
miryem.saidi@gmail.com

Abstract—Neurological diseases such as epilepsy, Parkinson’s disease, and Alzheimer, or psychiatric diseases like depression, personality disorders, schizophrenia, or addictive behavior, among many others, affect numerous people around the world. Diagnosing these diseases is a challenge in medicine. The symptoms of neurological and psychiatric diseases can vary greatly, making it difficult for healthcare professionals to accurately diagnose and treat patients. Hence, it is based on the interpretation of symptoms by doctors and the analysis of EEG signals (the electroencephalogram). There is therefore an important need for diagnostic support systems to assist doctors in their decision-making. Particularly in Algeria, where the health system suffers from a shortage of doctors specializing in neurology and psychiatry.

In this work, we compare the performance of different methods used in creating systems to assist in the medical diagnosis of an EEG signal in the case of a mental illness. This system will be designed using intelligent algorithms on electroencephalogram signals (EEGs), which are generally non-stable and complex and whose interpretation is long and laborious. Hence, the application of artificial learning algorithms such as random forest (RF) or deep neuronal networks.

This study brings light to the importance of machine learning algorithms in significantly reducing the time and effort required for interpreting EEG signals. We also raised the critical short- come of these same algorithms under some conditions in real- world problems, such as imbalanced datasets. **Index Terms**—EEG analysis, neurological diseases, psychiatric disorder, machine learning, automatic diagnosis.

I. INTRODUCTION

Neurological and psychiatric diseases can have a profound impact on individuals and their families, affecting their quality of life and overall well-being. Early detection and diagnosis of these conditions play a crucial role in ensuring timely intervention and appropriate treatment. Some common neurological and psychiatric diseases include Alzheimer’s disease, Parkinson’s disease, multiple sclerosis, epilepsy, depression, personality disorders, and schizophrenia. Each of these diseases has its own unique characteristics and can present a variety of symptoms.

By identifying symptoms and risk factors at an early stage, healthcare professionals can implement targeted interventions and support systems, ultimately improving outcomes for patients.

EEG signal analysis refers to the process of extracting meaningful information from electroencephalogram (EEG) signals, which are electrical brain wave recordings. This analysis plays a crucial role in various fields, including neuroscience, clinical medicine, neuromarketing, and brain- computer interfaces. By analyzing EEG signals, researchers, and clinicians can gain insights into brain activity patterns, identify abnormalities or anomalies, and even decode specific mental states or intentions.

This analysis plays a crucial role in understanding brain functioning, diagnosing neurological and psychiatric disorders, and monitoring the effectiveness of treatments. By deciphering the patterns and abnormalities in EEG signals, researchers and medical professionals gain valuable insights into brain activity, cognitive processes, and overall brain health.

Diagnoses in the field of psychology are made on a phenomenological and categorical basis. Clinicians assess explicit and observable signs and symptoms and offer categorical diagnoses based on which those symptoms fit into, in accordance with the “International Classification of Disorders (ICD)” and the “Diagnostic and Statistical Manual for Mental Disorders (DSM)” [1], [2]. This descriptive nosology improves communication clarity, but it has limitations due to its reliance on clinician observation and/or presenting problems recorded by patients or informants, which may not be sufficiently objective [3].

By analyzing the patterns and abnormalities in EEG signals through machine learning algorithms, healthcare professionals can save valuable time and resources while ensuring a more objective evaluation of patients. This can greatly help the detection and monitoring of the progression of the disease and evaluate the effectiveness of different treatment strategies, enabling healthcare professionals to make informed decisions and adjustments to the treatment plan.

In this work, we focus on the use of advanced algorithms and machine learning techniques to analyze the EEG signal to identify patterns and markers of different psychiatric conditions. We briefly reviews existing methods and presents results of applying these classification algorithms to a psychiatric disorder dataset. The rest of this article is organized into six sections. In the second one we present a review of some

existing methods of EEG signal analysis techniques for the automatic detection of neurological and psychiatric diseases. In third one, we introduce the used dataset. In the fourth section, we present the results of applying some of the well-known classification algorithms on the psychiatric dataset and we discuss the limitations and challenges faced during the analysis process. Finally, we present a conclusion that summarized the observations and the possible perspectives for future work.

II. EEG SIGNAL ANALYSIS TECHNIQUES

The brain is an incredibly complex and interesting organ that is responsible for everything from our thoughts and emotions to our physical movements and sensations. Cerebral electromagnetic activity is what scientists use to refer to the electrical and magnetic fields that are generated by brain neural activity. These fields are due to the flow of ions across the membranes of neurons, which creates an electrical potential that can be measured outside the brain. These electrical and magnetic fields can be detected using techniques such as electroencephalography (EEG).

Advanced machine learning algorithms can be applied to EEG signals to classify different brain states or detect specific abnormalities. These algorithms can be trained to recognize patterns in the EEG signals that are characteristic of different brain states, such as sleep, wakefulness, or certain mental disorders. By analyzing the EEG data in real-time, these algorithms can provide valuable insights into the brain's functioning and aid in the diagnosis and treatment of various neurological conditions. Furthermore, the integration of EEG with other neuro-imaging techniques, such as fMRI or PET scans, can provide a more comprehensive understanding of brain activity and its relation to different states or abnormalities.

A. Automatic Detection of Neurological and Psychiatric Diseases

In recent years, there has been significant progress in the development of automatic detection systems based on machine learning to analyze EEG data and identify potential signs of mental disorders.

Some research was conducted to detect abnormal EEG signals using machine learning classifiers as SVM [4], ensemble learning [5] or deep learning [6], [7], [8], [9]. Other research aims to identify a specific disease, such as epileptic seizures, using deep learning [10], SVM [11], [12] or linear programming boosting [13]. Saminu & al. [14] present a rigorous review of the existing methods.

These studies have shown promising results in accurately identifying mental disorders by detecting patterns and abnormalities in EEG signals. However, further research and validation are still needed to ensure the reliability and effectiveness of these systems in real-world clinical settings.

New intelligent approaches are presented to enable an earlier and more accurate diagnosis of Alzheimer disease; we cite SVM classifier [15], deep learning [16], [17], [18] or decision tree and K-nearest neighbor [19]. Parkinson disease detection is also an important research topic; Loh & al. have provided a comprehensive review of the existing literature on Parkinson's disease detection, shedding light on various approaches and techniques that have been explored in this field. This review serves as a valuable resource for researchers and clinicians seeking to enhance the accuracy and timeliness of Parkinson's disease diagnosis. [20].

Another important topic is the development of decision support systems for automatic detection of psychiatric diseases. These decision support systems leverage various techniques to identify such as Schizophrenia [18], [21], [22], autism [23], [24], [25], depression [26], [27], [28] or addiction [29], [30]. EEG-based approaches have shown promising results in detecting disorders, highlighting the potential for using these systems in a wide range of psychiatric diagnoses.

The aforementioned works, despite their promising results are limited to a single specific disorder. However, in [3] Park & al. intended to create new classification system for identifying patients with severe psychiatric illnesses from healthy controls. They gathered EEG data from patients with schizophrenia, mood disorders, anxiety disorders, obsessive-compulsive disorders, addictive disorders, trauma or stress-related disorders using RF and SVM algorithms.

This study aims to evaluate the performances of various classifiers including classification tree, random forests, and deep neural networks in classifying psychiatric disorders based on EEG data. By comparing the performances of these classifiers, we intend to identify the most effective approach for accurately diagnosing conditions such as depression, schizophrenia, and bipolar disorder using EEG signals. This research has the potential to contribute to the development of more reliable and efficient diagnostic tools for psychiatric disorders.

III. METHOD

To study the performance of different machine learning algorithms to identify and compare multiple psychiatric disorders using electroencephalography (EEG). We employ:

Classification And Regression Tree (CART): Proposed by Breiman et al. in (1984) [31], it is a supervised learning algorithm that uses the Gini index to find the best possible variable to split the node into two child nodes. The tree is grown to their maximum size until no splits are possible.

The K-nearest neighbors (K-nn): It is a non-parametric method based on a calculation of distances between the characteristic vector of the instance to be classified and the vectors of the instances of the learning base. Then the instance to be classified is assigned the majority class among

the classes of the k closest instances [32].

the Random Forest (RF) : Introduced by Breiman in 2001 [33], it is an ensemble method formed by a collection of trees, each tree contributes to the final decision made by majority vote.

Long Short-Term Memory (LSTM): is a variation of Re- current Neural Network (RNN) that is intended to handle sequential data. This algorithm has a memory cell that is capable of long-term information storage. LSTM networks are capable of learning long-term dependencies in sequential data. They have the capability of identifying patterns in data that deviate from the norm [34].

We compare the results based on accuracy, precision, recall and F-score.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F\text{-score} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

IV. DATASET

Available EEG psychiatric dataset are rare making it difficult for researchers and clinicians to access sufficient data for their studies and diagnostic purposes. This scarcity of available EEG datasets hinders progress in understanding and treating psychiatric disorders, as EEG data plays a crucial role in understanding brain activity and identifying biomarkers. The quantitative EEG (QEEG) at resting-state assessments, medical records, psychological test batteries, and other data were gathered retrospectively from the Seoul Metropolitan Government-Seoul National University (SMG-SNU) Boramae Medical Center in Seoul, South Korea, from January 2011 to December 2018. The dataset included 945 subjects and the diagnostic decision for the patients was established from March 2019 to August 2019, by two psychiatrists and two psychologists. The EEG data included 5 min eyes-closed resting-state with 19 or 64 channels acquired with 500–1,000 Hz sampling rate and 0.1–100 on-line filters via Neuroscan. The data were down-sampled to 128 Hz, and 19 channels were selected based on the international 10–20 system. [3]. Figure 1 and table I represent the different disorder categories and their distribution.

V. RESULTS AND DISCUSSION

This section goal is to determine the best framework to perform an automatic diagnosis. This study employs different classifier CART, RF, K-nn, and DeepL.

The CART, RF, K-nn, and DeepL classifiers were chosen due to their proven effectiveness in similar studies. Each

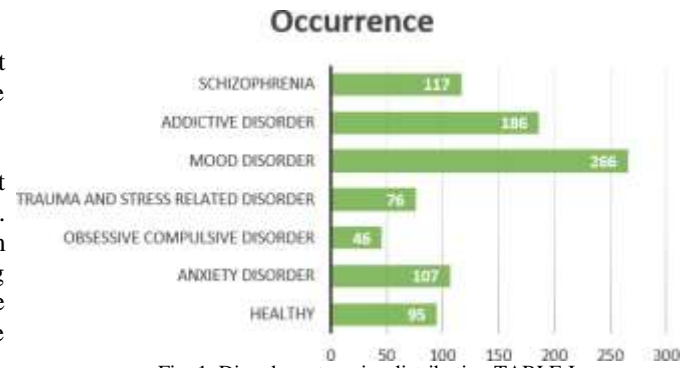


Fig. 1. Disorder categories distribution TABLE I

TABLE I: Disorder categories frequencies

Main disorder	Occurrence
Healthy	95
Anxiety disorder	107
Obsessive compulsive disorder	46
Trauma and stress related disorder	76
Mood disorder	266
Addictive disorder	186
Schizophrenia	117

classifier was trained and tested on a dataset consisting of labeled samples belonging to different classes. The accuracy of each classifier was measured by comparing its predictions to the actual class labels.

Each classifier is tested under various conditions such as number of classes. The performance of each classifier is evaluated based on metrics such as accuracy, precision, recall, and F-score.

Precision refers to the ability of a classifier to correctly identify positive cases, while recall measures its ability to find all positive cases. The F-score takes into account both precision and recall, providing a more balanced assessment of a classifier's performance. By comparing the performance of each classifier using these metrics, we can determine which framework is the most suitable for an automatic diagnosis system in this specific context.

The extremely bad scores above (table II) indicate that the dataset used in this study is not suitable for multi-class classification. Additionally, the poor results may also be attributed to the choice of classification algorithm used. Considering alternative algorithms and tuning their parameters may lead to better performance.

TABLE II: Multi-classification results

Multi-classification	Accuracy	Precision	Recall	F-score
CART	24,55	19,33	19,43	19,15
K-nn	21,48	16,79	16,07	14,97
RF	29,20	21,66	18,19	14,99
DeepL	28,16	0,08	0,28	0,12

Moreover, it is important to consider other factors such as data quality, the complexity of the problem, or the distribution of the classes. Ultimately, a comprehensive evaluation of all these factors is necessary to determine the most suitable approach for multi-class classification in this study.

After further investigation and experimentation we adjust the framework to improve the classification performance by performing a binary classification.

TABLE III: Binary classification results

Binary classification	Accuracy	Precision	Recall	F-score
CART	84,65	50,76	52,71	51,33
K-nn	88,67	63,22	52,68	53,39
RF	89,52	66,96	52,74	52,90
DeepL	90,70	61,21	51,07	49,59

Therefore, in a second experimentation, we implement a binary classification of healthy EEG signal vs. abnormal EEG signal (psychiatric disorder) to gain insights into the distinguishability of healthy controls from specific disorders and pave the way for more refined multi-class classification models in the future.

This binary classification approach allows us to accurately identify the presence of psychiatric disorders based on distinct EEG signal characteristics, which can greatly improve diagnostic accuracy. With the knowledge gained from this experimentation, we can develop more advanced and precise multi-class classification models that can differentiate between various psychiatric disorders with greater accuracy and efficiency.

Table III shows the obtained results of the binary classification task. The binary classification lead to significant improvements in the classification performance. Based on these results, the Deep learning model (90.70%) reaches the first place in term of classification accuracy followed by the RF (89.52%), k-nn (88.67%) algorithm and CART (84.64%) respectively.

However, when it comes to precision and F-score, the RF model outperforms the Deep learning model. This suggests that while DeepL may be better at overall classification accuracy, the RF model is more reliable in correctly identifying positive instances within the categories.

However, the precision and f-score remain insufficient, 51,33% for CART, 53,39% for K-nn, 52,90% for RF and 49.59 % for the Deep learning algorithm.

Fig. 2. Healthy EEG signal vs. abnormal EEG signal distribution

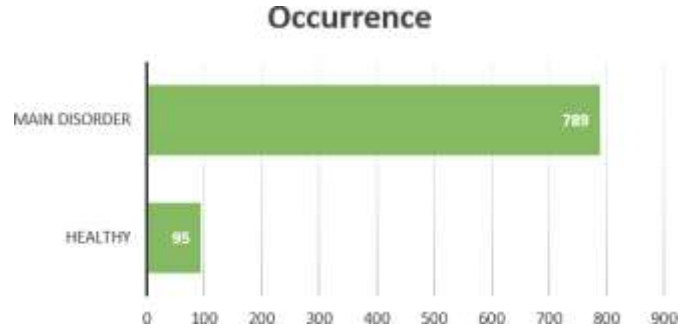


TABLE IV: Disorder categories frequencies

Diagnosis	Occurrence
Healthy	95
Abnormal	789

Further analysis to understand the reasons behind these differences suggests that the lack of data for certain classes (see figure 2, table IV), leads to imbalanced training sets and affects the performance of the model.

Additionally, the chosen algorithm may not be the most suitable for the given problem. Indeed, traditional approaches presuppose that all errors are equal and aim to reduce misclassification errors.

However, the dataset being used is very unbalanced, and traditional classifiers frequently misclassify instances belonging to minority classes. The solution is the use of resampling techniques and cost-sensitive algorithms.

These techniques aim to address the issue of imbalanced training sets by either oversampling the minority class or undersampling the majority class. Oversampling involves duplicating instances from the minority class to balance the dataset, while undersampling involves removing instances from the majority class.

Cost-sensitive algorithms, on the other hand, assign different misclassification costs to different classes, prioritizing the minority class and reducing the bias towards the majority class. By utilizing these approaches, the performance of the model can be significantly improved in scenarios with imbalanced datasets.

VI. CONCLUSION

Applications of machine learning algorithms in EEG signal analysis have shown promising results in various medical fields, including the detection and diagnosis of neurological or psychiatric disorders such as epilepsy, sleep disorders, or schizophrenia.

By leveraging the power of machine learning algorithms, researchers have been able to extract meaningful information from EEG signals, enabling accurate and efficient detection of abnormalities. These algorithms cannot only identify specific patterns associated with different neurological conditions, but, in some cases, they allow the differentiation between various stages of a disease.

In this work, we have compared the performance of different well-known machine learning algorithms. The results have shown that the classical methods perform poorly in terms of handling imbalanced datasets. Since they tend to misclassify the minority class. This limitation can have serious consequences in the context

of neurological disorder diagnosis, as early detection is crucial for effective treatment and most EEG signals datasets tend to be imbalanced.

In future works, we seek to use resampling techniques and cost-sensitive algorithms to improve performance. Since these techniques help in mitigating the bias towards the majority class and ensuring a more balanced classification.

VII. ACKNOWLEDGEMENTS

Park, S. M. (2021, August 16). EEG machine learning. Retrieved from osf.io/8bsvr

VIII. REFERENCES

- [1] W. H. Organization, *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. World Health Organization, 1992, vol. 1.
- [2] A. American Psychiatric Association, A. P. Association *et al.*, *Diagnostic and statistical manual of mental disorders: DSM-IV*. American psychiatric association Washington, DC, 1994, vol. 4.
- [3] S. M. Park, B. Jeong, D. Y. Oh, C.-H. Choi, H. Y. Jung, J.-Y. Lee, D. Lee, and J.-S. Choi, "Identification of major psychiatric disorders from resting-state electroencephalography using a machine learning approach," *Frontiers in Psychiatry*, vol. 12, p. 707581, 2021.
- [4] F. A. Alturki, K. AlSharabi, A. M. Abdurraqueeb, and M. Aljalal, "Eeg signal analysis for diagnosing neurological disorders using discrete wavelet transform and intelligent techniques," *Sensors*, vol. 20, no. 9, p. 2505, 2020.
- [5] T. Wu, X. Kong, Y. Zhong, and L. Chen, "Automatic detection of abnormal eeg signals using multiscale features with ensemble learning," *Frontiers in Human Neuroscience*, vol. 16, p. 943258, 2022.
- [6] H. Albaqami, G. M. Hassan, and A. Datta, "Automatic detection of abnormal eeg signals using wavenet and lstm," *Sensors*, vol. 23, no. 13, p. 5960, 2023.
- [7] R. T. Schirmer, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggenberger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.
- [8] S. Roy, I. Kiral-Kornek, and S. Harrer, "Chrononet: A deep recurrent neural network for abnormal eeg identification," in *Artificial Intelligence in Medicine: 17th Conference on Artificial Intelligence in Medicine, AIME 2019, Poznan, Poland, June 26–29, 2019, Proceedings 17*. Springer, 2019, pp. 47–56.
- [9] H. Albaqami, G. M. Hassan, A. Subasi, and A. Datta, "Automatic detection of abnormal eeg signals using wavelet feature extraction and gradient boosting decision tree," *Biomedical Signal Processing and Control*, vol. 70, p. 102957, 2021.
- [10] M. Varlı and H. Yılmaz, "Multiple classification of eeg signals and epileptic seizure diagnosis with combined deep learning," *Journal of Computational Science*, vol. 67, p. 101943, 2023.
- [11] A. K. Jaiswal and H. Banka, "Epileptic seizure detection in eeg signal using machine learning techniques," *Australasian physical & engineering sciences in medicine*, vol. 41, pp. 81–94, 2018.
- [12] O. Faust, U. R. Acharya, L. C. Min, and B. H. Spath, "Automatic identification of epileptic and background eeg signals using frequency domain parameters," *International journal of neural systems*, vol. 20, no. 02, pp. 159–176, 2010.
- [13] A. R. Hassan and A. Subasi, "Automatic identification of epileptic seizures from eeg signals using linear programming boosting," *computer methods and programs in biomedicine*, vol. 136, pp. 65–77, 2016.
- [14] S. Saminu, G. Xu, S. Zhang, I. Abd El Kader, H. A. Aliyu, A. H. Jabire, Y. K. Ahmed, and M. J. Adamu, "Applications of artificial intelligence in automatic detection of epileptic seizures using eeg signals: A review," 2022.
- [15] L. R. Trambaioli, A. C. Lorena, F. J. Fraga, P. A. Kanda, R. Anghinah, and R. Nitrini, "Improving alzheimer's disease diagnosis with machine learning techniques," *Clinical EEG and neuroscience*, vol. 42, no. 3, pp. 160–165, 2011.
- [16] C. L. Alves, A. M. Pineda, K. Roster, C. Thielemann, and F. A. Rodrigues, "Eeg functional connectivity and deep learning for automatic diagnosis of brain disorders: Alzheimer's disease and schizophrenia," *Journal of Physics: complexity*, vol. 3, no. 2, p. 025001, 2022.
- [17] Y. Zhao and L. He, "Deep learning in the eeg diagnosis of alzheimer's disease," in *Computer Vision-ACCV 2014 Workshops: Singapore, Singapore, November 1-2, 2014, Revised Selected Papers, Part I 12*. Springer, 2015, pp. 340–353.
- [18] Y. Zhao, "Deep learning in the eeg diagnosis of alzheimer's disease 2014 asian conf," *Computer Vision*, pp. 1–15, 2014.
- [19] D. Pirrone, E. Weitschek, P. Di Paolo, S. De Salvo, and M. C. De Cola, "Eeg signal processing and supervised machine learning to early diagnose alzheimer's disease," *Applied sciences*, vol. 12, no. 11, p. 5413, 2022.
- [20] H. W. Loh, W. Hong, C. P. Ooi, S. Chakraborty, P. D. Barua, R. C. Deo, J. Soar, E. E. Palmer, and U. R. Acharya, "Application of deep learning models for automated identification of parkinson's disease: a review (2011–2021)," *Sensors*, vol. 21, no. 21, p. 7034, 2021.
- [21] S. K. Khare and V. Bajaj, "A hybrid decision support system for automatic detection of schizophrenia using eeg signals," *Computers in Biology and Medicine*, vol. 141, p. 105028, 2022.
- [22] Z. Aslan and M. Akin, "A deep learning approach in automated detection of schizophrenia using scalogram images of eeg signals," *Physical and Engineering Sciences in Medicine*, vol. 45, no. 1, pp. 83–96, 2022.
- [23] M. N. A. Tawhid, S. Siuly, H. Wang, F. Whittaker, K. Wang, and Y. Zhang, "A spectrogram image based intelligent technique for automatic detection of autism spectrum disorder from eeg," *Plos one*, vol. 16, no. 6, p. e0253094, 2021.
- [24] M. Radhakrishnan, K. Ramamurthy, K. K. Choudhury, D. Won, and T. A. Manoharan, "Performance analysis of deep learning models for detection of autism spectrum disorder from eeg signals," *Traitement du Signal*, vol. 38, no. 3, 2021.
- [25] M. Liao, H. Duan, G. Wang *et al.*, "Application of machine learning techniques to detect the children with autism spectrum disorder," *Journal of Healthcare Engineering*, vol. 2022, 2022.
- [26] A. Rafiei, R. Zahedifar, C. Sitaula, and F. Marzbanrad, "Automated detection of major depressive disorder with eeg signals: a time series classification using deep learning," *IEEE Access*, vol. 10, pp. 73 804– 73 817, 2022.
- [27] P. P. Thoduparambil, A. Dominic, and S. M. Varghese, "Eeg-based deep learning model for the automatic detection of clinical depression," *Physical and Engineering Sciences in Medicine*, vol. 43, pp. 1349–1360, 2020.
- [28] G. Sharma, A. Parashar, and A. M. Joshi, "Dephnn: a novel hybrid neural network for electroencephalogram (eeg)-based screening of depression," *Biomedical signal processing and control*, vol. 66, p. 102393, 2021.
- [29] W. Mumtaz, N. Kamel, S. S. A. Ali, A. S. Malik *et al.*, "An eeg-based functional connectivity measure for automatic detection of alcohol use disorder," *Artificial intelligence in medicine*, vol. 84, pp. 79–89, 2018.
- [30] W. Mumtaz, P. L. Vuong, L. Xia, A. S. Malik, and R. B. A. Rashid, "An eeg-based machine learning method to screen alcohol use disorder," *Cognitive neurodynamics*, vol. 11, pp. 161–171, 2017.
- [31] L. Breiman, J. Friedman, R. Olshen, and C. Stone, "Classification and regression trees. wadsworth int," *Group*, vol. 37, no. 15, pp. 237–251, 1984.
- [32] N. S. Altman, "An introduction to kernel and nearest-neighbor non-parametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.
- [33] L. Breiman and R. Cutler, "Random forests machine learning [j]," *journal of clinical microbiology*, vol. 2, pp. 199–228, 2001.
- [34] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

Hardware Design for Medical Image Enhancement

Dr. BABA HAMED Amel
Department of Electrical Engineering,
LAAS laboratory National
Polytechnic School of OranMaurice
Audin (ENPO-MA)
Oran, Algeria
amel.bensemain@enp-oran.dz

Dr. BECHAR Hassane
Department of Biomedical
Engineering,
Faculty of technology,
University of Tlemcen
Tlemcen, Algeria
hassane.bechar@univ-tlemcen.dz

Dr. BENDOUMA Houaria
Department of Biomedical
Engineering, Faculty of
technology
University of Tlemcen
Tlemcen, Algeria
houara.bendouma@univ-tlemcen.dz

Abstract—In medicine, the images have become essential elements for establishing a good diagnostic. Therefore, the reconstruction and visualization of these medical images refers to a set of techniques, which make it possible to obtain information on the internal and external structure of an object against the background of an image. However, the large amount of data in medical image requires real-time processing. The aim of this article is to initiate a new research theme concerning the implementation of a few blocks for the reconstruction and visualization of medical images on a dedicated architecture FPGA (Field Programmable Gate array).

A high-level hardware description language VHDL (VHSIC: Hardware Description Language (Very High Speed Integrated Circuits)) describes the different blocks of the hardware architecture. For this, we used CAD tools (Computer-Aided Design).

Keywords—*Medical Image; Binarization; Dynamic Inversion; Luminance; FPGA; VHDL; CAD tools.*

I. Introduction

The beginning of medical imaging began with the discovery of X-rays in 1895 by the German Wilhelm Röntgen [1]. In recent years, the acquisition techniques have become increasingly sophisticate: X-rays, Ultrasound and Nuclear Magnetic Resonance [2].

In many cases, the raw filmed image does not provide all the details necessary for proper interpretation or analysis. In other cases, it is useful to see only certain points or specific elements of an image while neglecting the others. To do this, the image must be retouches in various ways to meet our needs (this is image processing).

For most methods of these processing tools (spot, local and global operations), the number of operations to be performed depends on the size of the images. Some methods have a high algorithmic complexity, which has the consequence of increasing the execution times on the computer (sequential calculation) [3], [4]. In order to speed up these processes, FPGA (Field Programmable Gate array) reconfigurable circuits were chose because they offer real-time calculation (parallel calculation) [5], [6]. The materialization of these operations opens another aspect in the reconstruction, visualization and analysis of medical images because it allows medical interpretation in real time [7], [8], [9].

The notion of real-time processing is define that information must be processed at the same rate as its acquisition. In this context, this article proposes a hardware implementation of some point operations tools, allowing reconstruction, transformation and visualization of medical images in real time.

This article is organize as follows. Section 2 presents a study on selected spot operations for the treatment of medical images. Section 3 is the essence of this work, as it includes the proposed hardware architecture, in order to establish the reconstruction, transformation and visualization of medical images in real time.

The fourth section is devoted to the study of performance results and the synthesis of implementation of these point operations on the reconfigurable FPGA circuit.

Finally, a conclusion given to summarize the work done in this project and provide new guidance for future work.

II. Proposed method

The purpose of processing an image is to extract the information it contains to solve a problem by eliminating unnecessary information. Processing takes place in several steps:

- Processing to transform the image (binarization, dynamic inversion ...).
- Preprocessing to improve image quality (Filtering...).
- Image decomposition to extract characteristics (Contour, segmentation).

Therefore, the equipment used should be improve to reduce the delays of reconstruction, transformation, processing and display of images. To do this, we propose the use of FPGA circuits “Fig. 1”, which are increasingly used and offer embedded solutions in image processing applications. We propose to implement, in VHDL language (VHSIC: Hardware Description Language (Very High Speed Integrated Circuits)) and using the Vivado CAD tool (Computer-Aided Design), treatments to offer quick and effective solutions.

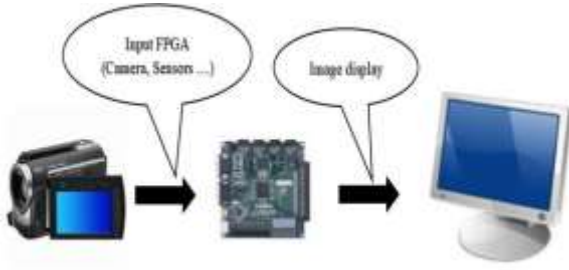


Fig.1. Embedded solutions in image processing applications.

For our work, we use the medical image presented in grayscale “Fig. 2”, whose pixels are coded on 8 bits. The size of this image is 600 x 519 pixels.



Fig.2. Medical image selected for treatment.

A. Simple thresholding

Simple thresholding is to turn a grayscale image into a binary image (0 and 255). By taking a threshold value (S), the pixel value $I(x, y)$ will be calculated as follows:

$$\begin{aligned} \text{if } I(x, y) < S \quad \text{then } T(x, y) &= 0 \\ \text{else } T(x, y) &= 255 \end{aligned} \quad (1)$$

With: $I(x, y)$ the input pixel values.

$T(x, y)$ output pixel values (processed pixels).

S the value of the threshold

B. Dynamic Image inversion

Dynamic inversion is an image transformation, which consists in reversing the black and white extremes. Sometimes we distinguish certain details better in a medical image in white on a black background than in black on a white background. the pixel value $I(x, y)$ will be calculated as follows:

$$T(x, y) = 255 - I(x, y) \quad (2)$$

With: $I(x, y)$ the input pixel values.

$T(x, y)$ output pixel values (processed pixels).

C. The luminance

The luminance of an image is defined as the average of the pixels in the original image:

$$\text{Average} = \frac{1}{N \cdot M} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} I(x, y) \quad (3)$$

With: $I(x, y)$ the input pixel values.

N the number of rows.

M the number of columns.

Adding a constant value to an image causes its overall brightness to increase and subtracting a constant value from an image causes its overall brightness to decrease.

D. Hardware Architecture

The hardware architecture of the system “Fig. 3” consists of several blocks, allowing the reconstruction, transformation and visualization of our image:

- Block of the original image (Image original),
- Block that will perform a simple thresholding of our image (Image Binarization),
- Block to perform dynamic image inversion (Dynamic Image Inversion),
- Block to adjust the luminance in the image (Image Luminance).

The pixel values of the image (Paramet_In) arrive successively to the system, whose processing at the level of the four blocks is carried out in parallel. We used a multiplexer at the output of these blocks, which has two command switches, to allow us to display the desired image (Paramet_Out). Both switches are located on the development board containing the FPGA circuit.

The process is synchronized on a clock (Clk), whose cycle number equals the number of pixels in the image (the image size).

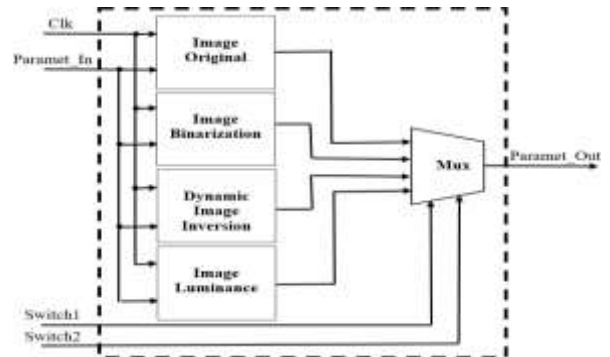


Fig.3. Hardware architecture.

III. Results and discussions

After the implementation of the hardware architecture under the Vivado environment, we perform an RTL analysis

on the system file. The system model (design) will be display in a logical view “Fig. 4”.

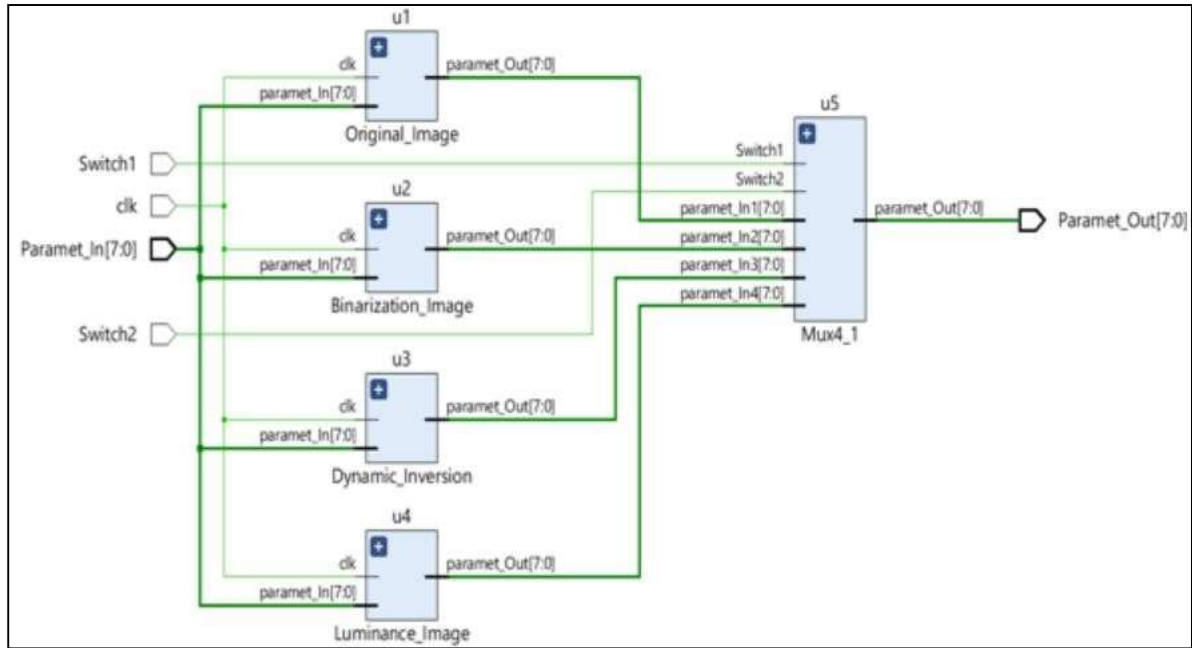


Fig.4. A logic view of the design

As we recorded the functional simulation (test bench) of each block of the system under the Vivado environment, to verify our calculations made on the input pixel values.

A. Image original block

This block allows the reconstruction and visualization of the original image “Fig. 14.a”, of which we use a register with parallel loading of 8 bits, which memorizes the values of the input pixels (Paramet_In) successively “Fig. 5”.

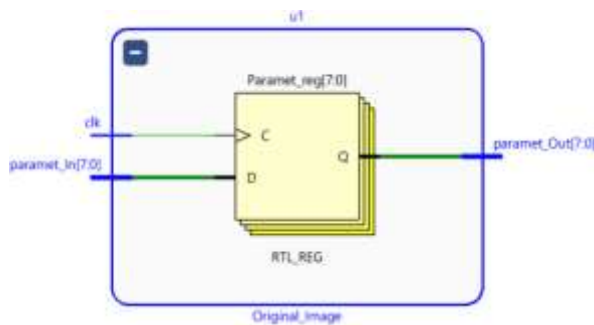


Fig.5. A logical view of the Original Image block.

The memorization of input pixel values at each clock cycle is show in the following figure “Fig. 6”:



Fig.6. The test bench of the Image Original block.

B. Image Binarization block

By choosing the threshold value, we apply the mathematical function “(1)”, which allowed us to binarize the input image “Fig. 14.b”. For this, we use in this block, a register with parallel loading for the memorization of the values of the input pixels (Paramet_In) and a logical comparator, which makes it possible to make the comparison between Paramet_In and the chosen threshold “Fig. 7”.

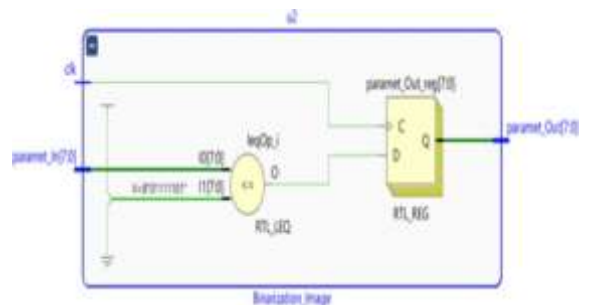


Fig.7. A logical view of the Image Binarization block.

The verification of calculations using functional simulation “Fig. 8”.

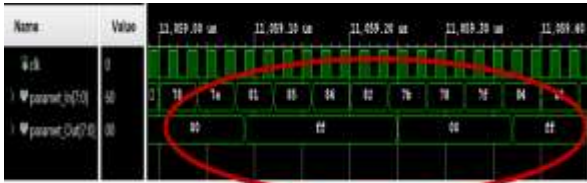


Fig.8. The test bench of the Image Binarization block.

C. Dynamic Image Inversion block

After applying the mathematical function “(2)” to all the pixel values of the image (Paramet_In), we reversed the dynamic of the input image “Fig. 14.c”. This block uses a parallel load register for storing input pixel values (Paramet_In) and a reverse logic gate for dynamic inversion “Fig. 9”.

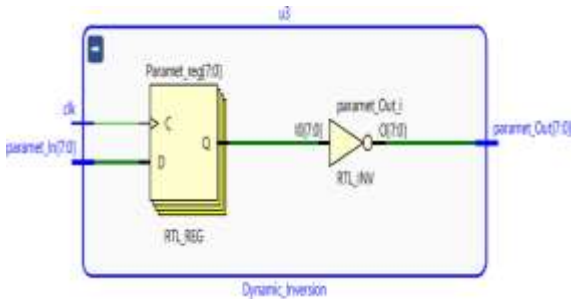


Fig.9. A logical view of the Dynamic Image Inversion block.

The inversion of the pixel values is check in the following figure “Fig. 10” after their memorization by the register.



Fig.10. The test bench of the Dynamic Image Inversion block.

D. Image Luminance block

Since the image is perfectly illuminated, we chose to adjust the luminance by subtracting from each input pixel

value (Paramet_In) a constant (the choice of the constant is arbitrary), which caused an overall decrease in luminance on the input image “Fig. 14.d”. This block uses a parallel load register that stores input pixel values and an algorithm that performs a binary subtraction between Paramet_In and the chosen constant “Fig. 11”.

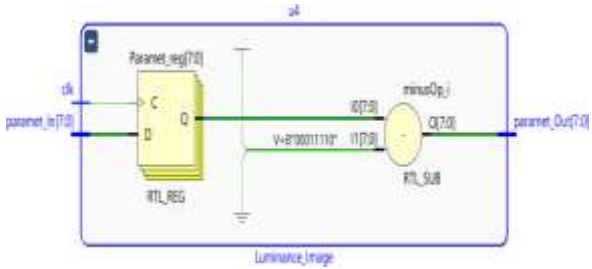


Fig.11. A logical view of the Image Luminance block.

The verification of calculations using functional simulation “Fig. 12”.



Fig.12. The test bench of the Image Luminance block.

The multiplexer used in this hardware design, for the choice of the processed image to be visualize is define by the following figure “Fig. 13”.

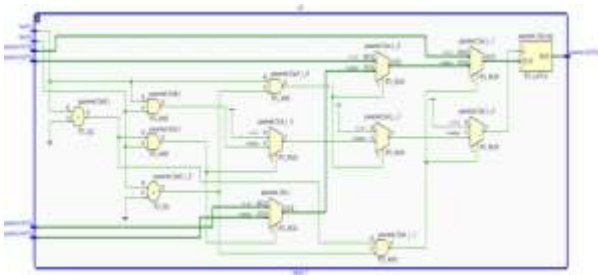


Fig.13. A logical view of the multiplexer.

The result of our treatment on the chosen medical image using the hardware design is present in the following figure “Fig. 14”.

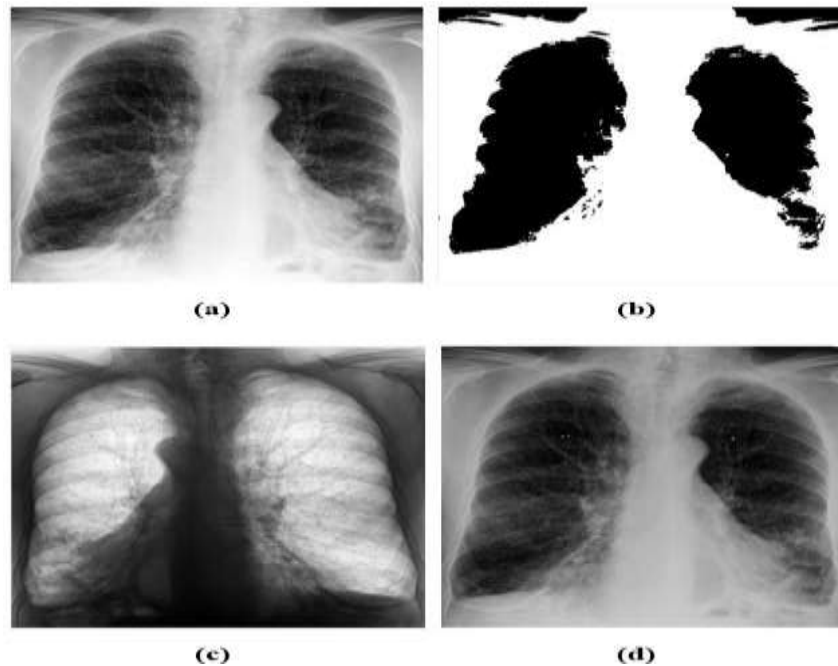


Fig.14. Images obtained after hardware implementation of the system. (a). Restitution of the original image. (b). Binarization of the original image. (c). Dynamic inversion of the original image. (d). Adjusting the luminance in the original image.

IV. Conclusion

Image processing is the new gateway for medical applications to help the physician make a decision. Our application consisted in implementing some medical image processing tools (reconstruction, transformation and visualization) so that the doctor's decision is made in real time.

Our work consisted in processing the medical image in real time (restitution, binarisation, dynamic inversion and luminance adjustment), whose material architecture of the proposed system realized these four transformations at the same time (parallel calculation); processing each pixel value of the image in a clock cycle (the clock cycle is 20 ns). So, the image processing time equal to the size of that image multiplied by the clock cycle.

Our future work will focus on:

- Make an automatic selection of the threshold for thresholding (for example, binarization).
- Realize hardware designs for other image processing tools (filtering, segmentation...).
- Use hardware classifiers for recognition of abnormalities in medical images.

V. References

- [1] "X-rays and the history of their properties", "Les rayons X et l'histoire de leurs propriétés", <https://enseignants.lumni.fr/fiche-media>
- [2] "Nuclear Magnetic Resonance (NMR) Theory and Solution NMR" - Physics LibreTexts, https://phys.libretexts.org/Courses/University_of_California_Davis
- [3] Maitine Bergounioux, "Some mathematical methods for image processing", "Quelques méthodes mathématiques pour le traitement d'image", HAL Id: cel-00125868 <https://cel.hal.science/cel-00125868v4> Submitted on 4 Jan 2009.
- [4] Dr. J. Thirumararan, S. Shylaja, "Medical Image Processing – An Introduction", International Journal of Science and Research (IJSR), ISSN (Online): 2319-7064, November 2014.
- [5] N.K. Ratha, A.K. Jain and D.T. Rover, "Convolution on Splash 2", Conference: FPGAs for Custom Computing Machines, 1995. Proceedings. IEEE Symposium on. May 1995
- [6] B.A. Draper, J. Ross Beveridge, A. P. Willem Böhm, C. Ross, M. Chawathe, Implementing Image Applications on FPGAs, Conference: Pattern Recognition, Proceedings. 16 th International Conference on, Volume 3, 2002.
- [7] M. Kraft, Adam Schmidt and Andrzej Kasinski, "HIGH-SPEED IMAGE FEATURE DETECTION USING FPGA IMPLEMENTATION OF FAST ALGORITHM", Conference: VISAPP 2008: Proceedings of the Third International Conference on Computer Vision Theory and Applications, Funchal, Madeira, Portugal, January 22-25, 2008 - Volume 1
- [8] F. HAMDAOUI, A. KHALIFA and A. SAKLY, "Real Time Implementation of Medical Images Segmentation Based on PSO", 2013 International Conference on Control, Decision and Information Technologies (CoDIT), May 2013.
- [9] H. Bendaoudi, "FLEXIBLE HARDWARE ARCHITECTURES FOR RETINAL IMAGE ANALYSIS", Thesis presented for the graduation of philosophy doctor (computer engineering) April 2017

Analyzing the Influence of Medical Imbalanced Data on Performance and Fairness in Differentially Private Deep Learning

RAFIKA BENLADGHEM

Tlemcen University

LRIT

Tlemcen, ALGERIA

rafika.benledghem@univ-tlemcen.dz

FETALLAH HADJILA

Tlemcen University

LRIT

Tlemcen, ALGERIA

fethallah.hadjila@univ-tlemcen.dz

MOHAMMED MERZOUG

Tlemcen University

LRIT

Tlemcen, ALGERIA

mohammed.merzoug@univ-tlemcen.dz

ADAM BELLOUM

University of Amsterdam Informatics

Institute Amsterdam

The Netherlands

A.S.Z.Belloum@uva.nl

Abstract— deep learning carries a significant potential for a paradigm shift in healthcare and medicine. Unfortunately, deep learning poses privacy risks, as various inference attacks have revealed. Differential Privacy offers robust guarantees and substantial defense against privacy threats, making it a prevalent approach for privacy-preserving deep learning lately. Many recent approaches to deep learning and differentially private deep learning assume identically Distributed data, which is often not the case in real-world situations. In our study, we examine the impact of imbalanced data on differentially private deep learning. We find that imbalanced data negatively affects both the model's performance and fairness. We explore the trade-off between privacy, usefulness, and fairness. Our findings underscore the challenges of using standard deep learning algorithms in a differentially private context to achieve reliable results for underrepresented groups.

Keywords— *Imbalanced medical data, privacy-utility/fairness trade-off, differentially private deep learning.*

I. INTRODUCTION

Deep learning, a potent subset of machine learning, has spearheaded remarkable progress across diverse domains, particularly in healthcare and medicine. However, it is essential to note that deep learning's voracious appetite for data is a fundamental characteristic of its success. Nevertheless, the widening scope of deep learning's capabilities also opens the door to increased privacy risks. Among the primary concerns are privacy breaches facilitated by attacks like membership Inference attacks [1], [2], [3], which aim to determine whether a specific patient's record is part of the dataset utilized to train the model. and Inversion attacks [4], seek to reconstruct the complete patient data using only access to an intermediate layer within the deep network. These privacy challenges underscore the intricate trade-off between the immense potential of deep learning and the paramount necessity of protecting sensitive data. Due to the sensitive nature of patient data in medical records, harnessing the full capabilities of deep learning in healthcare necessitates an inventive strategy for constructing and using deep neural networks while preserving patient privacy. Of the various approaches introduced to offer quantifiable assurances of privacy, Differential Privacy (DP) [5] stands out for its ability

to furnish algorithmic assurances of privacy in the face of several types of privacy threats.

While embracing DL and DP can offer privacy assurances, they come with associated drawbacks, including computational burdens [8], performance degradation [9], and fairness implications [10]. The last two issues are particularly accentuated in the context of deep learning when dealing with imbalanced class distributions within a classification dataset.

In our research, we explore the impact of imbalanced data on differentially private learning using a deep learning model designed for binary classification. Our primary focus is on evaluating the model's utility, as measured by commonly used metrics like F1-score and accuracy, along with its fairness. Initially, we establish baseline performance for non-differentially private deep learning models by assessing both utility and fairness in the presence of class imbalances. Subsequently, we extend our experiments to differentially private deep learning, quantifying the influence of imbalanced data on the model's utility and performance.

A. Contributions

In our study, our aim is to assess how various aspects of imbalanced data affect the deep, differentially private training of models used for binary classification in the context of imbalanced medical data. We aim to empirically investigate the influence of imbalanced data within a differentially private deep learning framework.

Our research makes the following key contributions:

- Establishing a performance benchmark and investigating the trade-offs between privacy, utility, and fairness in the context of imbalanced data. We demonstrate the adverse effects of differential privacy (DP) on both the fairness and utility of both non-private and DP deep learning models.
- Examining the intricate relationship involving data distribution, privacy, utility, and fairness in differentially private deep learning. We simulate varying degrees of privacy for DP deep learning and observe that an increase in the variability of data distribution tends to have a more

pronounced adverse effect on utility and fairness, especially for underrepresented classes.

II. BACKGROUND

This section offers fundamental foundational knowledge for the primary concepts and algorithms employed in our analysis.

A. Deep learning models for classification problems

Deep learning (DL) has become a pivotal methodology for addressing various classification problems, including binary classification. It leverages artificial neural networks, which are inspired by the structure and function of the human brain, to make predictions and decisions based on data. DL encompasses a class of machine learning techniques that excel in tasks involving the classification of data into one of two categories, commonly referred to as binary classification. These models are particularly suited for scenarios where discerning between two distinct outcomes is crucial, such as spam detection, disease diagnosis, sentiment analysis, and fraud detection.

Key Components of Deep Learning:

1. **Neural Networks:** Deep learning models are built upon neural networks, which consist of interconnected layers of artificial neurons. These networks learn to recognize patterns and relationships in data through a process known as backpropagation.
2. **Deep Architectures:** The "deep" in deep learning refers to the presence of multiple hidden layers within a neural network. These deep architectures allow the model to learn complex, hierarchical representations of the input data.
3. **Activation Functions:** Activation functions introduce non-linearity into neural networks, enabling them to model complex relationships. Common activation functions include the sigmoid, ReLU (Rectified Linear Unit), and softmax functions.
4. **Loss Functions:** Loss functions measure the disparity between the predicted output and the actual target. In binary classification, commonly used loss functions include binary cross-entropy and hinge loss.
5. **Optimization Techniques:** To train deep learning models effectively, optimization algorithms like stochastic gradient descent (SGD) and Adam are employed to adjust the model's parameters iteratively.

Challenges in Deep Learning for Binary Classification:

Despite their effectiveness, deep learning models can face challenges in the context of binary classification. These challenges may include data imbalance, where one class significantly outnumbers the other, and the need to strike a balance between model performance, fairness, and interpretability.

Deep learning offers a potent framework for tackling binary classification problems by learning intricate patterns and relationships within the data.

With the right architecture, optimization, and regularization techniques, deep learning models can deliver impressive results in discerning between two distinct classes, making them a valuable tool in a wide range of applications.

B. Differential Privacy

The original notion of ϵ -differential privacy (ϵ -DP) was initially proposed by [8]. Later, the same researchers introduced a revised form referred to as (ϵ, δ) -DP [6] [7]. In this adaptation, they incorporated the parameter δ as an additional element within the original definition. This adjustment was made to accommodate privacy protection in the context of the Gaussian distribution.

1. Definition 1: (ϵ, δ) -differential privacy [6] [7]:

A randomized mechanism, represented as K , achieves (ϵ, δ) -differential privacy if, for any pair of neighboring data samples B and B' , and for all possible outcome subsets Z within the set K :

$$\forall B, B' \in B^n, \forall Z \subseteq W :$$

$$\Pr[K(B) \in Z] \leq e^\epsilon \Pr[K(B') \in Z] + \delta \quad (1)$$

Where: W is the set of all possible outputs,

$$\delta \ll 1/|Z|$$

The interpretation of mechanism K as meeting (ϵ, δ) -differential privacy implies that it attains ϵ -differential privacy with a probability of $1-\delta$. However, (ϵ, δ) -DP is not suitable when the privacy-sensitive set Z consists of just one element. It's crucial to emphasize that the value of δ must be exceedingly small in comparison to the size of set Z (i.e., $\delta \ll 1/|Z|$) to prevent the unfavorable scenario where privacy is consistently compromised for a significant portion of the dataset represented by δ .

One of the prominent techniques for introducing differential privacy (DP) into the field of machine learning is differentially private stochastic gradient descent DP-SGD, which was introduced by Abadi et al. in 2016 [11]. The DP-SGD method operates by limiting the gradients to manage the sensitivity of the mechanism and incorporating precisely calibrated noise into the gradient values. To monitor the privacy budget expenditure, an accounting mechanism has been suggested. Additionally, other accounting methods have been put forward in existing literature, offering more stringent estimates of privacy costs, such as the Rényi-DP-based accountant [12].

There are two important concepts that we need to delve into: the notion of neighboring datasets and function sensitivity.

2. **Definition 2: Neighboring datasets:** Consider two datasets, B and B' , both belonging to the dataset domain B^n . These datasets, B and B' , are regarded as neighboring when they differ by a single data point. In simpler terms, B' is obtained by either adding or removing a single data point from B .

3. **Definition 3: Sensitivity [7]:** Sensitivity measures the most significant change in the output resulting from the alteration of a single data point within the database. The sensitivity of a query function, denoted as f , is expressed as:

$$\forall B, B' \in B^n, f: B^n \rightarrow \mathbb{R}^d$$

diagnosti

c measurements. Notably, all individuals in the $\Delta = \max \|f(B) - f(B')\|$

- (2) dataset are female, aged at least 21 years, and of Pima Indian heritage. The dataset comprises various medical where $\|\cdot\|$ denotes the l_2 norm.

In this paper, to attain (ϵ, δ) -differential privacy, we select the Gaussian mechanism, which utilizes L2 norm sensitivity. It involves the introduction of Gaussian noise to each dimension of the output $f(B)$:

$$f(B) + N(0, \Delta^2 \sigma^2 I) \quad (3)$$

4. Imbalanced data in deep learning

Imbalanced data in the context of deep learning refers to a situation where the distribution of classes in a dataset is highly skewed, with one class significantly outnumbering the other(s). This imbalance can pose challenges for machine learning models, including deep neural networks, as they may have difficulty learning to accurately classify the minority class. The model tends to be biased toward the majority class, resulting in poorer performance for the minority class and potential unfairness in predictions.

Consider a medical diagnosis scenario where a deep learning model is trained to classify X-ray images as either "normal" or "disease present." In this dataset, there is a significant class imbalance, with 90% of the images being "normal" and only 10% showing "disease present." In this case, the deep learning model may perform exceptionally well in correctly identifying "normal" cases due to their abundance in the dataset. However, it may struggle to accurately detect "disease present" cases, as they are underrepresented. This imbalance can lead to life-critical errors, where the model fails to identify individuals with the disease, posing serious consequences.

III. METHODOLOGY

In this study, the experiments investigate the training of two types of models: non-private models, specifically employing the original "vanilla" Stochastic Gradient Descent (SGD), and differentially private deep learning (DP-DL) models using DP-SGD. These models are trained on two distinct medical datasets, namely, the PIMA dataset and the Breast Cancer Wisconsin dataset. The experimentation encompasses varying privacy levels and employs two different network architectures and settings for each of the two datasets.

A. Experiment setup

All experiments were carried out using Python version 3.8.16. The tests were conducted on Google Colab, which offers access to the NVIDIA Tesla GPU and CUDA compilation tools version 11.2.152. The advantage of this setup is the ability to execute code at a significantly higher rate of operations per second compared to a CPU. Two open-source libraries suitable for training deep learning models with DP are Opacus for PyTorch [13] and TensorFlow Privacy for TensorFlow [14]. Both of these libraries are compatible with CUDA. For this research, the choice was made to utilize Opacus.

A. PIMA

1) Datasets:

The objective is to predict the likelihood of a patient having diabetes using the Pima database, which contains specific predictor variables and a binary target variable, where "1" represents diabetic and "0" represents non-diabetic outcomes. The Pima database is of size (768x9), and it is characterized by an imbalanced distribution of classes as depicted in Figure 1.

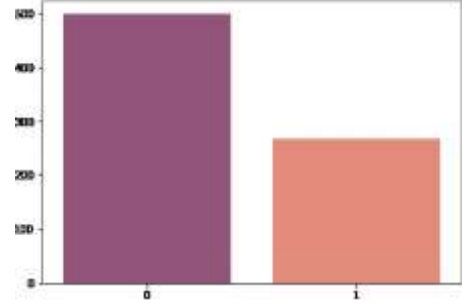


Fig. 1. Class distribution in the PIMA database.

2) Model architecture:

The model architecture can be described as follows:

- Input Layer (8 features)
- Hidden Layer 1 (20 neurons) with ReLU activation
- Hidden Layer 2 (5 neurons) with ReLU activation
- Hidden Layer 3 (5 neurons) with ReLU activation
- Output Layer (2 neurons)

This architecture is designed for binary classification tasks, such as those commonly encountered in the medical field, including the prediction of diseases like diabetes. The model takes 8 input features, passes them through the hidden layers with ReLU activations, and produces binary classification output.

B. Breast Cancer Wisconsin:

1) Datasets:

This dataset provides details regarding the attributes of cell nuclei found in images. These features are derived from digitized images of fine needle aspirates (FNA) of breast masses. For each image, the mean, standard error, and the worst or largest values (mean of the three largest) of these features were calculated, resulting in a total of 30 features. The Breast Cancer Wisconsin database is sized at (569x32), and it exhibits an imbalanced distribution, with approximately 37% of cases diagnosed as malignant (cancerous) and around 63% classified as benign (non-cancerous) as depicted in Figure 2.

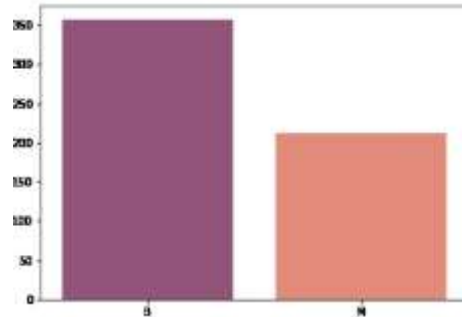


Fig. 2. Class distribution in the BREAST CANCER WISCONSIN database.

2) *Model architecture*: The model architecture can be described as follows:

- Input Layer (30 features)
- Hidden Layer 1 (20 neurons) with ReLU activation
- Hidden Layer 2 (10 neurons) with ReLU activation
- Hidden Layer 3 (10 neurons) with ReLU activation
- Output Layer (2 neurons)

This architecture is designed for binary classification tasks, such as the diagnosis of breast cancer (malignant or benign). The model takes 30 input features, processes them through the hidden layers with ReLU activations, and produces binary classification output. It's a common architecture for tasks involving medical diagnosis and classification.

C. Metrics

- Precision (pre)**: Precision measures the accuracy of positive predictions. It is the ratio of true positives to the sum of true positives and false positives. A high precision indicates a low rate of false positive predictions.
- Recall (rec)**: Recall, also known as Sensitivity or True Positive Rate, measures the ability of the model to correctly identify positive instances. It is the ratio of true positives to the sum of true positives and false negatives. A high recall indicates a low rate of false negative predictions.
- Specificity (spe)**: Specificity measures the ability of the model to correctly identify negative instances. It is the ratio of true negatives to the sum of true negatives and false positives.
- F1-Score (f1)**: The F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall and is useful when there is an imbalance between the two.
- Geometric Mean (geo)**: The geometric mean of precision and recall is used to calculate the G-mean. It is a metric that takes into account both false positives and false negatives, making it suitable for imbalanced datasets.
- Index of Balanced Accuracy (iba) [15]**: The Index of Balanced Accuracy is a metric that combines the sensitivity (recall) and specificity of a classification model to provide a balanced measure of accuracy. It's particularly useful for imbalanced datasets.

These metrics are especially important when dealing with imbalanced datasets because they provide a more comprehensive evaluation of the model's performance, taking into account both the majority and minority classes.

The hyper-parameter settings for training both the private and non-private models are provided in Table 1, for both datasets.

Dataset	Settings	Value
PIMA	Learning rate	0.01
	Loss function	CrossEntropyLoss
	Batch size	64
	Epochs	500
	Noise parameter(σ)	Variable
BREAST CANCER WISCONSIN	Learning rate	0.001
	Loss function	CrossEntropyLoss
	Batch size	64
	Epochs	500
	Noise parameter(σ)	Variable

Table.1. Hyper-parameters of the model that was trained on the Datasets.

Dataset	Settings	Value
PIMA	Learning rate	0.01
	Loss function	CrossEntropyLoss
	Batch size	64
	Noise parameter(σ)	Variable
	Gradient_clipping_norm(C)	1
	Delta(δ)	10^{-4}
	Epochs	500
BREAST CANCER WISCONSIN	Learning rate	0.001
	Loss function	CrossEntropyLoss
	Batch size	64
	Noise parameter(σ)	Variable
	Gradient_clipping_norm(C)	1
	Delta(δ)	10^{-4}
	Epochs	500

Table.2. Hyper-parameters of DP-model that was trained on the Datasets.

IV. RESULT AND DISCUSSION

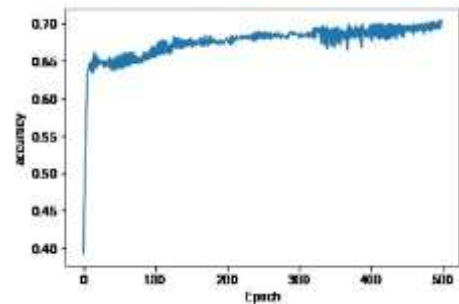


Fig. 3. Accuracy of the Non-DP Model Trained on the PIMA Dataset.

For the model trained on the PIMA dataset the model exhibits reasonable accuracy, with an overall accuracy rate of 72% as shown in Figure 3. The model trained on the Breast Cancer dataset showcases outstanding accuracy, with an overall accuracy rate of 92% as shown in Figure 4.

The difference in model performance can be traced back to the class distribution in the datasets. The PIMA dataset has a class imbalance, with a significantly larger number of samples in one class (class 0) compared to the other (class 1) – 498 samples for class 0 and 270 for class 1. In contrast, the Breast Cancer dataset features a more balanced distribution, with 357 samples for non-cancerous (class B) and 212 samples for cancerous (class M).

This class imbalance in the PIMA dataset has a notable impact on accuracy. In imbalanced datasets, models tend to perform well on the majority class (class 0 in this case) but struggle with the minority class (class 1). The model trained on the PIMA dataset is influenced by the dominance of class 0 samples, which can lead to accuracy results skewed in favor of that class. Conversely, the Breast Cancer dataset's balanced class distribution contributes to the exceptional accuracy achieved by the model trained on it.

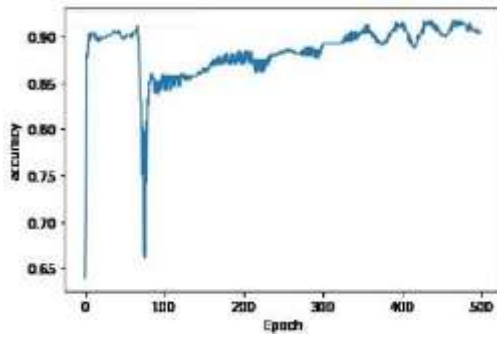


Fig. 4. Accuracy of the Non-DP Model Trained on the Breast cancer Wisconsin SET.

In our analysis, we observed distinct performance characteristics in two different datasets as shown in table 3. In the PIMA dataset, Class 0, representing non-diabetic cases, demonstrated reasonably balanced precision and recall, albeit with a lower specificity, indicating its ability to effectively classify negative cases but also raising concerns about potential false positives. Conversely, Class 1, representing diabetic cases, exhibited lower precision and recall, implying challenges in correctly identifying positive cases. While weighted metrics indicated a balanced overall performance. In contrast, the Breast Cancer dataset displayed notable performance attributes. Class 0, representing benign cases, showcased exceptional precision, recall, and specificity, signifying its proficiency in identifying negative cases. Similarly, Class 1, representing malignant cases, achieved high precision, balanced recall, and specificity, reflecting an overall commendable performance. Weighted metrics reinforced the excellence of the model on this dataset, which benefits from a more balanced class distribution. Our findings underscore the substantial impact of class distribution on model accuracy, with balanced datasets, like the Breast Cancer dataset, leading to superior performance. Achieving fairness in class representation remains essential, as it directly influences precision and recall, making the Breast Cancer model a more reliable classifier.

Results							
Non Private							
Datasets		Pre	Rec	Spe	F1	Geo	Iba
PIMA	0	0.79	0.81	0.51	0.80	0.64	0.43
	1	0.55	0.51	0.81	0.53	0.64	0.40
Average/Toral	/	0.72	0.72	0.60	0.72	0.64	0.42
Breast cancer Wisconsin	B	0.88	1.00	0.81	0.94	0.90	0.82
	M	1.00	0.81	1.00	0.89	0.90	0.79
Average/Toral	/	0.93	0.92	0.89	0.92	0.90	0.81

Table.3. Non-Private Binary Classification Model Evaluation Metrics for Pima and Breast Cancer Wisconsin Datasets.

In Table.4. The provided results illustrate how adjusting the privacy budget ϵ impacts the performance metrics of differentially private deep learning models trained on imbalanced datasets (PIMA and Breast Cancer). ϵ governs the level of noise introduced to the model updates to ensure differential privacy.

For the PIMA dataset at $\epsilon = 8$, both classes exhibit reasonable precision, recall, and F1-scores, signaling a balanced performance. The lower specificity for Class 0 compared to Class 1 indicates a potential imbalance favoring Class 1. The Index of Balanced Accuracy at 0.64 suggests a relatively fair model. At $\epsilon = 2$, the model's performance diminishes, resulting in lower precision, recall, and F1-scores. Substantial decreases in specificity for Class 0 highlight an imbalance in favor of Class 1, reflected in an Index of Balanced Accuracy of 0.50 a less balanced model. At $\epsilon = 1$, further reduction in ϵ leads to significant performance deterioration, with Class 0 specificity reaching 0, indicating severe unfairness. The Index of Balanced Accuracy drops to 0.32, portraying a highly unfair model.

Results								
DP- Private								
Dataset	(ϵ, δ) -dp		Pre	Rec	Spe	F1	Geo	Iba
PIMA	$(8, 10^{-4})$ -dp	0	0.81	0.77	0.51	0.79	0.63	0.40
		1	0.45	0.51	0.77	0.48	0.63	0.38
	$(2, 10^{-4})$ -dp	0	0.77	0.69	0.31	0.73	0.46	0.24
		1	0.23	0.31	0.69	0.27	0.46	0.21
	$(1, 10^{-4})$ -dp	0	0.75	0.63	0.00	0.68	0.03	0.01
		1	0.00	0.00	0.63	0.00	0.03	0.008
Breast Cancer Wisconsin	$(8, 10^{-4})$ -dp	B	1.00	0.79	1.00	0.88	0.89	0.80
		M	0.62	1.00	0.79	0.76	0.89	0.78
	$(2, 10^{-4})$ -dp	B	0.96	0.70	0.87	0.81	0.78	0.63
		M	0.43	0.87	0.70	0.57	0.78	0.60
	$(1, 10^{-4})$ -dp	B	0.91	0.63	0.67	0.74	0.65	0.43
		M	0.25	0.67	0.63	0.36	0.65	0.40

Table.4. DP-Private Binary Classification Model Evaluation Metrics for Pima and Breast Cancer Wisconsin Datasets.

For the Breast Cancer dataset at $\epsilon = 8$, both classes (B and M) demonstrate high precision, recall, and F1-scores, indicating a balanced performance. The Index of Balanced Accuracy at 0.89 suggests a well-fair model. At $\epsilon = 2$, decreasing ϵ results in lower precision, recall, and F1-scores. Reduced specificity for Class B implies an imbalance in favor of Class M, reflected in an Index of Balanced Accuracy of 0.79—a less fair model. At $\epsilon = 1$, further reduction in ϵ leads to significant performance deterioration, with Class B specificity decreasing to 0.67, indicating a substantial imbalance. The Index of Balanced Accuracy drops to 0.65, depicting a less fair model.

a) Impact of ϵ on Metrics and model quality:

Higher values of ϵ generally lead to better model performance, while lower values result in increased privacy and reduced utility. As ϵ decreases, the models become less accurate and more unfair.

b) Fairness - Distribution Trade-off:

Fairness is reflected in the balance of performance metrics across classes. In other words A well-balanced model should have similar metrics for both classes. Notably, adding

privacy has a more pronounced effect on the class with fewer samples.

c) Privacy-Utility/Fairness Trade-off:

There exists a trade-off between privacy (controlled by ϵ) and model utility/fairness. Smaller ϵ values provide stronger privacy guarantees but may sacrifice model quality.

Class distribution is a pivotal factor in model performance. The class imbalance in the PIMA dataset can result in accuracy biases, favoring the majority class. This is why the model for breast cancer performs better, benefiting from a more even class distribution. In practical scenarios, selecting an appropriate value of ϵ involves considering the desired level of privacy and the acceptable trade-off with model performance. Achieving a balance between privacy and utility is crucial, especially in applications where fairness is a key consideration.

V. CONCLUSION

In this study, we conducted a thorough empirical analysis to examine how imbalanced data influences the effectiveness and equity of differentially private deep learning models. Utilizing a real-world dataset encompassing imbalanced samples, we delved into the intricate trade-offs involving utility, privacy, and data distribution. Our findings indicate that differential privacy (DP) generally exerts a negative influence on both the utility and performance of deep models, particularly for underrepresented classes. Moreover, we observed that imbalanced data exacerbates the disparity in utility and fairness between minority and majority classes. Our choice of metrics allowed us to comprehensively investigate the diverse aspects of imbalanced data's impact on our experiment with privacy-preserving deep models. While our results align with our initial expectations regarding the effects of DP and imbalanced data, it's important to note that this study represents a constrained exploration of the intricate interplay between fairness and utility, considering their trade-offs with privacy and data distribution. Nonetheless, our consistent findings emphasize that the presence of imbalanced data has an adverse effect on the utility and fairness of privacy-preserving deep models.

REFERENCES

- [1] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership inference attacks against machine learning models. In 2017 IEEE Symposium on Security and Privacy (SP), pages 3–18, 2017.
- [2] Yugeng Liu, Rui Wen, Xinlei He, Ahmed Salem, Zhikun Zhang, Michael Backes, Emiliano De Cristofaro, Mario Fritz, and Yang Zhang. MI-doctor: Holistic risk assessment of inference attacks against machine learning models, 2021.
- [3] N. Carlini, S. Chien, M. Nasr, S. Song, A. Terzis, and F. Tramer. Membership inference attacks from first principles. arXiv, 2021a.
- [4] Matt Fredrikson, Somesh Jha, and Thomas Ristenpart. Model inversion attacks that exploit confidence information and basic countermeasures. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, pages 1322–1333, 2015.
- [5] Cynthia Dwork. Differential privacy: A survey of results. In International Conference on Theory and Applications of Models of Computation, Xi'an, China, April 2008.
- [6] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In Advances in Cryptology—EUROCRYPT, pages 486–503, 2006a.
- [7] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Proc. of the Third Conf. on Theory of Cryptography (TCC), pages 265–284, 2006b.
- [8] T. T. Cai, Y. Wang, and L. Zhang, “The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy,” arXiv preprint arXiv:1902.04495, 2019.
- [9] BLANCO-JUSTICIA, Alberto, SANCHEZ, David, DOMINGO-FERRER, Josep, et al. A critical review on the use (and misuse) of differential privacy in machine learning. ACM Computing Surveys, vol. 55, no 8, p. 1-16, 2022.
- [10] T. Farrand, F. Mireshghallah, S. Singh, and A. Trask, “Neither private nor fair: Impact of data imbalance on utility and fairness in differential privacy,” in Proceedings of the 2020 Workshop on Privacy-Preserving Machine Learning in Practice, pp. 15–19, 2020.
- [11] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang, “Deep learning with differential privacy”, October, In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, 2016.
- [12] I. Mironov, “Renyi differential privacy,” in 2017 IEEE 30th Computer Security Foundations Symposium (CSF), pp. 263–275, IEEE, 2017.
- [13] Facebook. Opacus: Train pytorch models with differential privacy, September 2021. [Online; posted 01-September-2021].
- [14] Google. Tensorflow privacy, September 2021. [Online; posted 01 - September-2021].
- [15] García, V., Mollineda, R.A., Sánchez, J.S. Index of Balanced Accuracy: A Performance Measure for Skewed Class Distributions. In: Araujo, H., Mendonça, A.M., Pinho, A.J., Torres, M.I. (eds) Pattern Recognition and Image Analysis. IbPRIA 2009. Lecture Notes in Computer Science, vol 5524. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02172-5_57. (2009).

Residual network (resnet-18) for laparoscopic image distortion classification

Belmokeddem Mohamed
Biomedical Engineering Department,
Faculty of Technology
University Abou-bekr Belkaid of
Tlemcen
Tlemcen, Algeria
mohamed.belmokeddem@univ-tlemcen.dz

khemis kamila
Biomedical Engineering
Department, Faculty of Technology
University Abou-bekr Belkaid of
Tlemcen
Tlemcen, Algeria
kamila.khemis@univ-tlemcen.dz.

Loudjedi Salim
Surgery B Tlemcen Hospital,
Department of medicine
University Abou-bekr Belkaid of Tlemcen
Tlemcen, Algeria
ismatsalim.loudjdimouedden@univ-tlemcen.dz

Abstract— Diminished laparoscopic video quality directly affects a surgeon's visibility and can compromise the outcomes of computational tasks in robot-assisted surgery. To address this challenge, numerous solutions have been proposed based on the detection and classification of laparoscopic video distortions.

In this work, we propose a method based on Residual networks (ResNet18) for the automatic detection and classification of noise 'NO', smoke 'SM', uneven illumination 'UI', defocus blur 'DB', and motion blur 'MB' in laparoscopic videos. We have obtained an accuracy of 98.75% for training and 97.97% for validation.

The high accuracy scores across the classes emphasize the model's capability to generalize well and make accurate predictions.

Keywords—Laparoscopic video, distortion classification, deep learning.

I. INTRODUCTION

Guaranteeing Adequate Video Quality is a Vital Prerequisite for Uninterrupted Minimally Invasive Surgery (MIS) Procedures. Laparoscopic Videos Are Prone to Various Types of distortions, encompassing Noise, Smoke, Uneven Illumination, Defocus Blur, and Motion Blur. These diverse distortions comprise the five categories of laparoscopic video distortion [1]. In response to the challenge of laparoscopic video distortion, researchers established the Laparoscopic Video Quality Database (LVQ). This database includes a collection of 200 videos, featuring five distinct distortion categories and four different intensity levels [1].

Numerous deep-learning techniques have been applied to detect and categorize these five distortion categories. These methods encompass the utilization of a single Deep Neural Network (DNN) [1], Convolutional Neural Networks like ResNet50, as well as the integration of ResNet and a Fully Connected Neural Network (FCNN) [2].

In this study, we employ Residual Networks (ResNet18) for the automated detection and classification of these distortions. Our computational framework was initially trained on an extended version of the LVQ database, referred to as the LVQ Challenge dataset.

II. MATERIALS AND METHODS

A. Dataset description

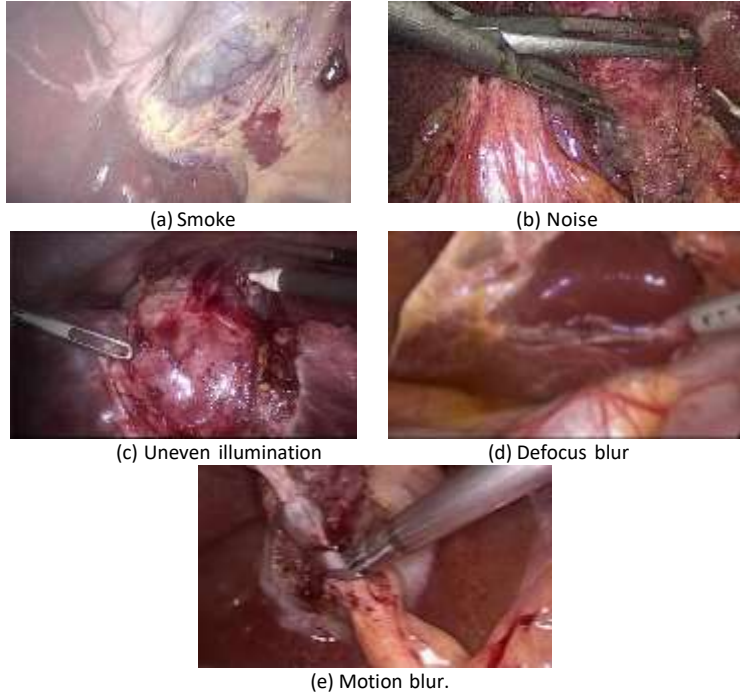
This paper utilizes the LVQ database Challenge, as developed by Khan, Beghdadi, Cheikh, et al. [1] [3]; This database consists of 20 reference videos, each spanning 10 seconds in duration. These reference videos have been extracted from the Cholec80 dataset, originally introduced by Twinanda et al. [4]. The Cholec80 dataset encompasses ten distinct categories of scene variations, including bleeding (BL), grasping and burning (GB), multiple instruments (MI), irrigation (IR), clipping (CL), stretching away (SA), cutting (CU), stretching forward (SF), organ extraction (OE), and burning (BU). Each of the reference videos has been intentionally subjected to five different categories of distortions, with four varying levels of severity as shown in Table 1.

This extensive dataset encompasses a total of 400 videos with single distortion and 400 videos with multiple distortions. The five primary categories of distortion within this dataset include smoke, noise, uneven illumination, defocus blur, and motion blur.

In our study, we specifically focused on utilizing the videos from the five single distortion classes to train our model. Each of these single distortion classes consists of 80 videos, resulting in a balanced and comprehensive dataset.

To evaluate the performance of the proposed methodology, we have used a total of 20660 frames (4132 frames for each class) extracted from the five laparoscopic video classes of LVQ dataset with a rate of 5 fps, as shown in Table 2.

We resize video frames to 224*224 pixels to apply them to ResNet-18. Eighty percent, (80%) of image, data from each class (3305 images) was used for training and 20% of images were reserved for testing.



containing five units. This alignment with the number of classes within our dataset ensures a tailored approach. For model training, we utilize a laparoscopic dataset and employ the cross-entropy loss function, a commonly used metric in classification tasks. The cross-entropy loss function is minimized during training. We accomplish this using the Adam optimizer with a learning rate of 0.01. Our training framework is implemented within the PyTorch framework and executed on an NVIDIA T1000 GPU with 12 GB of memory. During model training, we utilize mini-batches with a size of 10 and conduct training over 50 epochs.

III. RESULTS

For evaluating the performance of distortion classification, we have evaluated the classification performance of our proposed architecture. As shown in Fig.2, the matrix of confusion of the classification by class. Table 3 shows the accuracy of the classification of our model by class. The distortions present in the LVQ database are denoted as follows: SM for smoke, WN for noise, UI for uneven illumination, DB for blur due to defocus, and MB for blur due to motion.

Fig.1. Distorted images extracted from the LVQ database [1]

TABLE I. DATASET DESCRIPTION [1]

Resolution of video	Frame rate	Number of distortions	Number of levels
512 X 288	25 fps	5	4

TABLE II. TRAINING DATA SUMMARY

Dataset	Number of frames
Number of defocus blurred frames	4132
Number of motion blurred frames	4132
Number of smoked frames	4132
Number of uneven illumination frames	4132
Number of noised frames	4132
Number of total frames	20660

B. Methods

In this paper, we harness the ResNet architecture to tackle the classification of laparoscopic video distortions. ResNet-18, a convolutional neural network renowned for its depth, consists of 18 layers and boasts 2.37 million trainable parameters. This architecture is renowned for its remarkable generalization capabilities [5]. Given its prowess, we have opted to leverage the ResNet architecture, specifically replacing the output fully connected layer with a new layer

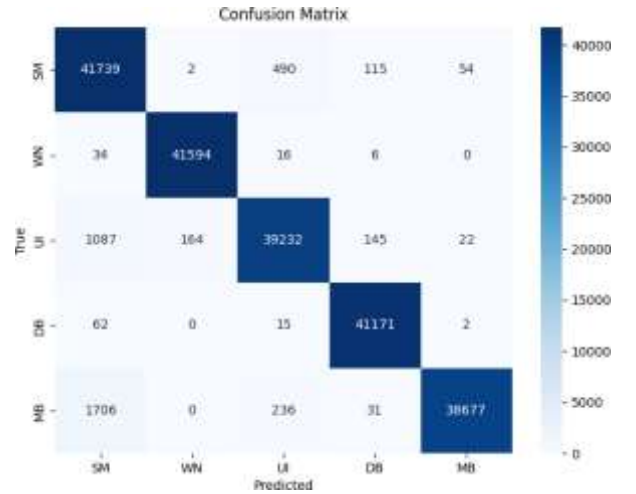


Fig.2. Confusion matrix of the proposed solution

TABLE III. PREDICTION RATE BY DISTORTION'S CLASS

Class	Smoke	Noise	Uneven illumination	Defocus blur	Motion blur
Prediction rate %	98.44	99.86	96.51	99.80	95.14

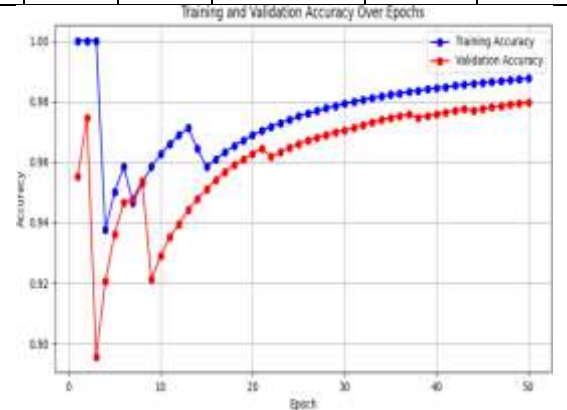


Fig. 3. Accuracy evolution with the number of epochs for our approach

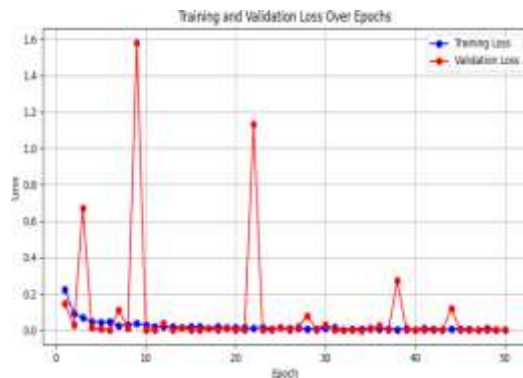


Fig. 4. Loss function evolution with the number of epochs for our approach

IV. DISCUSSION

The results of using this approach are highly encouraging, and exhibit impressive prediction rate scores:

Smoke ('SM'): Achieving a prediction rate of 98.44% for the 'SM' class is remarkable. This high value suggests that the model excels at correctly identifying frames affected by smoke, which is a critical factor in laparoscopic video quality assessment.

Noise ('WN'): With a prediction rate of 99.86% for the 'WN' class, the model shows exceptional performance in recognizing noisy frames. Minimal misclassifications in this category indicate the model's robustness in handling noise-related distortions.

Uneven Illumination ('UI'): A prediction rate of 96.51% for the 'UI' class reflects the model's ability to distinguish frames with uneven illumination. Though slightly lower than the previous classes, this score remains impressive.

Defocus Blur ('DB'): The 'DB' class attains an outstanding prediction rate of 99.80%. This highlights the model's proficiency in identifying frames affected by defocus blur, which is a significant factor in laparoscopic video quality assessment.

Motion Blur ('MB'): With a prediction rate of 95.14% for the 'MB' class, the model demonstrates its capability to classify frames exhibiting motion blur accurately. Although the prediction rate is slightly lower than some other classes, it remains a commendable performance.

These results collectively indicate that the ResNet-18 model is highly effective in recognizing and classifying frames affected by various laparoscopic video distortions. The high prediction rates across the classes emphasize the model's capability to generalize well and make accurate predictions, even when faced with different types of distortions commonly encountered in laparoscopic videos.

Moreover, the balanced distribution of frames across classes, with 4132 samples per class, ensures that the model is not biased toward any specific category. This balanced dataset contributes to the model's capacity to provide reliable predictions across all classes.

To show the interest of this approach, Fig.3 illustrates the accuracy evolution of both training and validation datasets over epochs. Thus, the convenience of curves can be observed. At the number of epochs=50, we have a Training Accuracy of 98.75% and a Validation Accuracy of 97.97%.

Fig. 4 illustrates the loss functions evaluated on the training

and validation datasets over epochs. Thus, it can be observed the convergence of the loss validation and loss training to zero.

V. CONCLUSION

The primary objective of this study was the classification of laparoscopic video distortions. To achieve this, we utilized the ResNet18 architecture.

The results obtained, particularly in terms of accuracy, are not only remarkable but also highly encouraging. They provide a strong foundation for expanding the process of detecting distortions in laparoscopic videos.

ACKNOWLEDGMENT

There are no conflicts to declare.

No funding was received to conduct this research and publication of this article.

All authors have seen and approved the manuscript.

VI. REFERENCES

- [1] Z.A.Khan et al., "Towards a video quality assessment based framework for enhancement of laparoscopic videos". J. Med. Imaging 11316, 2020.
- [2] ZA.Khan, A.Beghdadi, M.Kaaniche, FA.Cheikh, O.Gharbi, "A neural network based framework for effective laparoscopic video quality assessment", Comput. Med. Imaging Graph 101, 2022.
- [3] Khan, Z.A., Beghdadi, A., Kaaniche, M. and Cheikh, F.A., 2020, October. Residual networks based distortion classification and ranking for laparoscopic image quality assessment. In 2020 IEEE International Conference on Image Processing (ICIP) (pp. 176-180). IEEE.
- [4] Twinanda .P, Shehata S, Mutter D, Marescaux J, Mathelin M, Padoy N. EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Videos. IEEE Trans. on Medical Imaging 2016.
- [5] He K, Zhang X, Ren S, Sun J, Deep residual learning for image recognition, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778. Doi: 10.1109/CVPR.2016.90.

Artificial intelligence techniques in the smart grid: a review

Bezzar Nour El Houda
Electrical Engineering
Department, Mining
Laboratory University of
Larbi Tebessi, Tebessa,
12002, Algeria
nourelhouda.bezzar@univ-
tebessa.dz

Meraoumia Abdallah
Laboratory of
Mathematics, Informatics,
and Systems (LAMIS)
University of Larbi
Tebessi, Tebessa, 12002,
Algeria
ameraoumia@univ-
tebessa.dz

Houam Lotfi
Laboratory of
Mathematics, Informatics,
and Systems (LAMIS)
University of Larbi
Tebessi, Tebessa, 12002,
Algeria
lotfi.houam@univ-tebessa.dz

Abstract—The smart grid or the new generation of energy came as an alternative to the traditional energy systems that no longer meet the increasing demand for electricity consumption with the increasing population, using bi-directional flows of electricity and information compatible with the capabilities of communication, advanced computing, control, and self-healing. It appeared to enhance the reliability and efficiency of power distribution with minimal impacts and to establish an automated power delivery network. Artificial intelligence (AI) has become one of the fastest-growing areas of technology and can play an important role in many applications, especially in the energy field. This paper provides a structured review of current research into some artificial intelligence techniques that can be applied to load prediction, power grid stability assessment, fault detection, and safety issues in smart grid and power systems, and aims to introduce the applications of artificial intelligence in enhancing and improving the reliability and flexibility of smart grid systems.

Keywords—*smart grid, artificial intelligence, faults detection, load forecasting*

I. INTRODUCTION

As the electricity market has been reformed renewable energy scenarios have been implemented, and the energy system presents the characteristics of openness, uncertainty, and complexity. Building the smart power grid has become a distinct trend and opportunity. [1] The smart grid is a transformation of the energy pattern from an electromechanically controlled system to an electronic one. The network systems consist of information management, control techniques, communication techniques, and field devices that coordinate electrical operations. These technologies, in turn, changed the planning and operational problems found in the traditional network to the ability to monitor or measure operations and to communicate data to operating centers. Some of the relevant problem areas in smart grids include load prediction, stability assessment, network security identification, and fault detection. AI Techniques support the digital power grid, through power system optimization, robust user behavior analysis, fault diagnosis, and more. The application of artificial intelligence faces many

challenges and problems, including the lack of reliability, incomplete infrastructure, and the lack of algorithms for the energy industry. Artificial intelligence is a powerful tool to drive the smart grid more clearly. [2] Artificial intelligence techniques allow the use of staggering amounts of data to create intelligent machines that can handle tasks that require human intelligence. There are other ways to achieve artificial intelligence systems, such as machine learning, which is a branch of artificial intelligence, which are neural networks, robots, expert systems, and fuzzy logic, as we mentioned, despite the accuracy and reliability of artificial intelligence systems, it remains subject to many challenges in smart grid. Two types of artificial intelligence can be used, virtual and physical, and artificial intelligence systems in the smart grid are divided into artificial narrow intelligence (ANI) It is the only type of artificial intelligence that has been successfully achieved so far that is highly intelligent in completing the tasks directed at it, specifically designed to perform individual tasks such as facial or voice recognition, driving a car, searching the Internet. Artificial general intelligence (AGI) is the concept of a machine with intelligence that simulates human intelligence, whether in understanding or the ability to learn and solve problems. Artificial superintelligence (ASI) it is the superiority of machines over human intelligence. It is a source of inspiration for science fiction. The volume of artificial intelligence research and its applications in the smart grid has increased in recent years, and since it is not possible to provide a comprehensive review of techniques of artificial intelligence in the smart grid, whether in predicting the load or evaluating the stability of the power grid, detecting faults and ensuring network security, our paper presents some Techniques Artificial intelligence in some fields and sheds light on some potential future applications of artificial intelligence techniques. The remainder of this paper is organized as follows. Section 2 presents smart grid features. Section 3 surveys artificial intelligence techniques. Section 4 discusses some of the Artificial intelligence techniques in smart grids. Section 5 summarizes What is the future of artificial intelligence in smart networks. A short summary conclude the paper in Section 6.

II. SMART GRID AND ITS FEATURES

Countries and various institutions know the smart grid with several definitions, but it is one definition. The smart grid is the network that allows the transmission of information and energy in two directions from each node in the network. It has an optimal performance that provides users with optimal services.

The different features of the smart grid make it possible to distinguish it from the traditional electric grid; the smart grid contains the following features [3]:

- Two-Way Communication.
- Integration Of Renewable Energy Resources (RERs) into The Grid
- Advanced Metering Infrastructure (AMI) System.
- Advanced Energy Storage Technologies.
- Data Management and Processing.
- Real-Time Operation and Control.
- Cyber-Physical Security of SG.

III. ARTIFICIAL INTELLIGENCE TECHNIQUES

Given the revolutionary transformation of the modern energy system, the components of the network have become more distributed, along with the infrastructure for metering and communication, where energy resources are integrated by including the energy network with the basic communication system. [4] To improve the performance of the smart grid by supporting wide applications such as distributed power management, [5] system state prediction, [6] and cyber-attack security, large amounts of data are placed due to the inability of traditional computational techniques to process the data of smart grid systems known to be in huge quantities as intelligence techniques artificial and others. To meet the challenges, many researchers have been wasted studying artificial intelligence techniques to improve the performance of the smart grid.

Artificial intelligence techniques in the smart grid are classified into the following areas:

A. Machine Learning

Is a term that refers to continuous learning and predictions from the data available. It consists of different algorithms that analyze the data to produce decisions or predictions regarding the current context through a set of instructions.

B. Deep Learning

Is a subfield of machine learning that deals with algorithms inspired by the structure and function of the brain called artificial neural networks. Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.

C. Supervised learning

Create a mathematical model for the training data, which according to this mathematical model, where the supervised learning algorithms contain regression and classification algorithms.

D. Unsupervised learning

Recognize and describe the hidden data and rules in it so that unsupervised learning algorithms contain data of one class, grouping, and dimensionality reduction.

E. Reinforcement learning (RL)

Raising the system to the value of the output function, which is done by learning in some cases, and this learning has succeeded in areas we mention, including driving without a plane, and automated chess.

Traditional machine learning tries to put some laws in the samples and analyzes new data based on these laws. [1] Algorithms include traditional machines such as support vectors (SVM) and Bayesian methods.

IV. ARTIFICIAL INTELLIGENCE TECHNIQUES IN SMART GRIDS

The application of artificial intelligence techniques is divided into predicting the power load, [2] the application of artificial intelligence techniques is divided into forecasting the power load, Power grid stability assessments that include frequency stability, voltage stability, and small signal stability, then fault detection and ensuring the security of the smart grid.

A. Load forecasting

With the application of renewable energy scenarios such as solar energy, wind energy, and tidal energy, the uncertainty of the operation of the smart grid increases a lot. Forecasting the energy load makes the energy production and load identical in real-time and thus becomes important work for the electricity grid. Forecasting energy or energy load it is the discovery and determination of the strength of the demand in the network. [7,8] Forecasting is divided into three sections: Short-term (STLF), medium-term (MTLF), and long-term (LTLF), where the time ranges from a few minutes to a year or more.

TABLE I. TECHNIQUES FOR LOAD FORECASTING

Author (Ref.)	Techniques		
	Year	Objective	Technique
[12]	2017	STLF	RNN
[13]	2017	LTLF	LSTM
[14]	2018	LTLF	Fuzzy, ANN
[15]	2018	LTLF	LSTM, GRU
[16]	2019	MTLF	DBN
[27]	2019	MTLF	DNN
[18]	2020	STLF	CNN, Ensem
[19]	2020	LTLF	LSTM, GRU

Author (Ref.)	Techniques		
	Year	Objective	Technique
[20]	2021	MTLF	LSTM, ETS, Ensemble
[21]	2021	MTLF	SVR

B. Power Grid Stability Assessment

It includes frequency stability (FSA), Transient Stability Assessment (TSA), Voltage Stability Assessment (VSA), Small-Signal Stability Assessment, or oscillatory stable assessment (OSA); [9, 10] the purpose of these matters is to ensure the reliability and security of the power system. The stability of a power system is the ability to remain in equilibrium. Traditional stability assessment models are complex and require large computational resources. Therefore, network stability assessment methods have been applied.

TABLE II. AI TECHNIQUES FOR THE POWER SYSTEM STABILITY ASSESSMENT

Author (Ref.)	Techniques		
	Year	Objective	Technique
[22]	2017	TSA	ELM, TF
[23]	2017	VSA	SVR, FL
[24]	2018	TSA	RNN, LSTM
[25]	2018	TSA	SVM, ANN
[26]	2019	TSA	SVM
[27]	2019	OSA	PSO
[28]	2020	OSA	CNN
[29]	2020	VSA	Decision tree
[30]	2021	VSA	Random Forest

C. Faults Detection and protection of flexible equipment in power system

Equipment based on power electronics technology is considered flexible equipment used in several fields, including alternating and direct current transmission, generation and storage of renewable energy, power distribution systems, small grids, and others. Diagnosing faults and protecting flexible equipment in the power system is a line of defense that ensures the safety of equipment, which in turn contributes to isolating faults quickly and avoiding the expansion of damage and error. [1] Deep learning can gain profound advantages in introducing new knowledge to expand the sample space and to express clearly the characteristics of equipment error at various levels.

TABLE III. AI TECHNIQUES FOR POWER SYSTEM FAULTS DETECTION

Author (Ref.)	Techniques		
	Year	Objective	Technique
[31]	2017	PV FD	PNN
[32]	2017	Line trip	LSTM, SVM

Author (Ref.)	Techniques		
	Year	Objective	Technique
		FD	
[33]	2018	Line trip FD	AE, SVM
[34]	2018	WT FD	ANN
[35]	2019	HIFD	ELM
[36]	2019	PV FD	GPR
[37]	2020	Line FD	ELM
[38]	2020	PV FD	ANN
[39]	2021	FD	Ensemble

D. Power network security protection

A Smart grid is a highly automated power transmission network that allows information and energy to flow two-way from each node of the grid. [1] Smart grid has more perfect performance and can provide users with a series of value-added services. The deep interaction of the information flow aims to make the power system in the face of many threats, as the network attack has the characteristics of strong concealment throughout the incubation period. Deep learning can identify network attacks, detect malware, provide protection, and ensure network security.

TABLE IV. AI TECHNIQUES FOR POWER NETWORK SECURITY PROTECTION

Author (Ref.)	Techniques		
	Year	Objective	Technique
[40]	2018	Survey	DL, RL
[41]	2019	Attacks detection	RL
[42]	2020	Survey	ML
[43]	2020	Survey	AI

V. FUTURE OF ARTIFICIAL INTELLIGENCE IN SMART GRIDS

Smart Grids aim to achieve full self-learning, responsive, adaptive, self-healing, automotive, and cost-effective. [11] Future directions or opportunities to achieve advanced smart grid systems are discussed as follows.

- Integration with cloud computing: The integration of artificial intelligence with cloud computing plays an important role in achieving smart grid systems by enhancing security and strength and reducing interruptions.
- Fog computing: Fog computing processes raw data locally and has multiple advantages such as energy efficiency, scalability, and flexibility.
- Transfer learning: [1] Transfer learning from training data requirements where researchers are motivated to use them to solve the problem of insufficient data.
- Consumer behavior prediction: With the help of fog computing, demand-side management has become a vital

task of management participation of users in energy systems. Energy consumption contributes significantly to the response to demand for the energy consumer side.

VI. CONCLUSION

The power system is in a revolutionary transformation from the traditional electric grid system to the smart grid system. In this case, the energy system methods dictate the advancement of suitable solutions. Thus, artificial intelligence techniques are applied in smart grid systems with promising results. In this paper, we presented an overview of the modern applications of artificial intelligence techniques, which came in four areas (that is, load forecasting, power grid stability assessment, faults detection, and security problems). We also discussed the future trends of applying artificial intelligence techniques in the smart grid.

REFERENCES

- [1] Jiao, J. (2020, June). Application and prospect of artificial intelligence in smart grid. In IOP Conference Series: Earth and Environmental Science (Vol. 510, No. 2, p. 022012). IOP Publishing.
- [2] Omitaomu, O.A.; Niu, H. Artificial Intelligence Techniques in Smart Grid: A Survey. *Smart Cities* 2021, 4, 548–568. <https://doi.org/10.3390/smartcities4020029> Metke, A. R., & Ekl, R. L. Smart grid security technology. In 2010 Innovative Smart Grid Technologies (ISGT) (pp. 1–7). IEEE. (2010).
- [3] Kim, S. C., Ray, P., & Reddy, S. S. (2019). Features of smart grid technologies: an overview. *ECTI Transactions on Electrical Engineering, Electronics, and Communications*, 17(2), 169–180.
- [4] Foruzan, E.; Soh, L.K.; Asgarpour, S. Reinforcement learning approach for optimal distributed energy management in a microgrid. *IEEE Trans. Power Syst.* 2018, 33, 5749–5758.
- [5] Zhang, L.; Wang, G.; Giannakis, G.B. Real-time power system state estimation and forecasting via deep unrolled neural networks. *IEEE Trans. Signal Process.* 2019, 67, 4069–4077.
- [6] Karimipour, H.; Dehghantanha, A.; Parizi, R.M.; Choo, K.K.R.; Leung, H. A deep and scalable unsupervised machine learning system for cyber-attack detection in large-scale smart grids. *IEEE Access* 2019, 7, 80778–80788.
- [7] Tong, C.; Li, J.; Lang, C.; Kong, F.; Niu, J.; Rodrigues, J.J. An efficient deep model for day-ahead electricity load forecasting with stacked denoising auto-encoders. *J. Parallel Distrib. Comput.* 2018, 117, 267–273.
- [8] Zheng, J.; Xu, C.; Zhang, Z.; Li, X. Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network. In Proceedings of the 2017 51st Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 22–24 March 2017; pp. 1–6.
- [9] Xu, Y.; Dong, Z.Y.; Zhao, J.H.; Zhang, P.; Wong, K.P. A reliable intelligent system for real-time dynamic security assessment of power systems. *IEEE Trans. Power Syst.* 2012, 27, 1253–1263.
- [10] You, S.; Zhao, Y.; Mandich, M.; Cui, Y.; Li, H.; Xiao, H.; Fabus, S.; Su, Y.; Liu, Y.; Yuan, H.; et al. A Review on Artificial Intelligence for Grid Stability Assessment. In Proceedings of the 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), Tempe, AZ, USA, 11–13 November 2020; pp. 1–6.
- [11] Ali, S.S.; Choi, B.J. State-of-the-Art Artificial Intelligence Techniques for Distributed Smart Grids: A Review. *Electronics* 2020, 9, 1030.
- [12] Shi, H.; Xu, M.; Li, R. Deep learning for household load forecasting—A novel pooling deep RNN. *IEEE Trans. Smart Grid* 2017, 9, 5271–5280.
- [13] Zheng, J.; Xu, C.; Zhang, Z.; Li, X. Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network. In Proceedings of the 2017 51st Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 22–24 March 2017; pp. 1–6.
- [14] Ali, D.; Yohanna, M.; Ijasini, P.M.; Garkida, M.B. Application of fuzzy–Neuro to model weather parameter variability impacts onelectrical load based on long-term forecasting. *Alex. Eng. J.* 2018, 57, 223–233.
- [15] Kumar, S.; Hussain, L.; Banarjee, S.; Reza, M. Energy load forecasting using deep learning approach-LSTM and GRU in spark cluster. In Proceedings of the 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT), Kolkata, India, 12–13 January 2018; pp. 1–4.
- [16] Liu, Z.; Sun, X.; Wang, S.; Pan, M.; Zhang, Y.; Ji, Z. Midterm power load forecasting model based on kernel principal component analysis and back propagation neural network with particle swarm optimization. *Big Data* 2019, 7, 130–138.
- [17] Jiang, W.; Tang, H.; Wu, L.; Huang, H.; Qi, H. Parallel processing of probabilistic models-based power supply unit mid-term load forecasting with apache spark. *IEEE Access* 2019, 7, 7588–7598.
- [18] Moon, J.; Jung, S.; Rew, J.; Rho, S.; Hwang, E. Combination of short-term load forecasting models based on a stacking ensemble approach. *Energy Build.* 2020, 216, 109921.
- [19] Dong, M.; Grumbach, L. A hybrid distribution feeder long-term load forecasting method based on sequence prediction. *IEEE Trans. Smart Grid* 2019, 11, 470–482.
- [20] Dudek, G.; Pełka, P.; Smył, S. A Hybrid Residual Dilated LSTM and Exponential Smoothing Model for Midterm Electric Load Forecasting. *IEEE Trans. Neural Networks Learn. Syst.* 2021, doi:10.1109/TNNLS.2020.3046629.
- [21] Rai, S.; De, M. Analysis of classical and machine learning based short-term and mid-term load forecasting for smart grid. *Int. J. Sustain. Energy* 2021, 1–19. doi:10.1080/14786451.2021.1873339.
- [22] Tang, Y.; Li, F.; Wang, Q.; Xu, Y. Hybrid method for power system transient stability prediction based on two-stage computing resources. *IET Gener. Transm. Distrib.* 2017, 12, 1697–1703.
- [23] Amroune, M.; Musirin, I.; Bouktir, T.; Othman, M.M. The amalgamation of SVR and ANFIS models with synchronized phasor measurements for on-line voltage stability assessment. *Energies* 2017, 10, 1693.
- [24] James, J.; Hill, D.J.; Lam, A.Y.; Gu, J.; Li, V.O. Intelligent time- adaptive transient stability assessment system. *IEEE Trans. Power Syst.* 2017, 33, 1049–1058.
- [25] Baltas, N.G.; Mazidi, P.; Ma, J.; de Asis Fernandez, F.; Rodriguez, P. A comparative analysis of decision trees, support vector machines and artificial neural networks for on-line transient stability assessment. In Proceedings of the 2018 International Conference on Smart Energy Systems and Technologies (SEST), Seville, Spain, 10–12 September 2018; pp. 1–6.
- [26] Hu, W.; Lu, Z.; Wu, S.; Zhang, W.; Dong, Y.; Yu, R.; Liu, B. Real-time transient stability assessment in power system based on improved SVM. *J. Mod. Power Syst. Clean Energy* 2019, 7, 26–37.
- [27] Kamari, N.; Musirin, I.; Ibrahim, A.; Halim, S. Intelligent swarm-based optimization technique for oscillatory stability assessment in power system. *IAES Int. J. Artif. Intell.* 2019, 8, 342.
- [28] Shi, Z.; Yao, W.; Zeng, L.; Wen, J.; Fang, J.; Ai, X.; Wen, J. Convolutional neural network-based power system transient stability assessment and instability mode prediction. *Appl. Energy* 2020, 263, 114586.
- [29] Meng, X.; Zhang, P.; Xu, Y.; Xie, H. Construction of decision tree based on C4. 5 algorithm for online voltage stability assessment. *Int. J. Electr. Power Energy Syst.* 2020, 118, 105793.
- [30] Liu, S.; Shi, R.; Huang, Y.; Li, X.; Li, Z.; Wang, L.; Mao, D.; Liu, L.; Liao, S.; Zhang, M.; et al. A data-driven and data-based framework for online voltage stability assessment using partial mutual information and iterated random forest. *Energies* 2021, 14, 715.

- [31] Garoudja, E.; Chouder, A.; Kara, K.; Silvestre, S. An enhanced machine learning based approach for failures detection and diagnosis of PV systems. *Energy Convers. Manag.* 2017, 151, 496–513.
- [32] Zhang, S.; Wang, Y.; Liu, M.; Bao, Z. Data-based line trip fault prediction in power systems using LSTM networks and SVM. *IEEE Access* 2017, 6, 7675–7686.
- [33] Wang, Y.; Liu, M.; Bao, Z.; Zhang, S. Stacked sparse autoencoder with PCA and SVM for data-based line trip fault diagnosis in power systems. *Neural Comput. Appl.* 2019, 31, 6719–6731.
- [34] Helbing, G.; Ritter, M. Deep Learning for fault detection in wind turbines. *Renew. Sustain. Energy Rev.* 2018, 98, 189–198.
- [35] AsghariGovar, S.; Pourghasem, P.; Seyedi, H. High impedance fault protection scheme for smart grids based on WPT and ELM considering evolving and cross-country faults. *Int. J. Electr. Power Energy Syst.* 2019, 107, 412–421.
- [36] Fazai, R.; Abodayeh, K.; Mansouri, M.; Trabelsi, M.; Nounou, H.; Nounou, M.; Georghiou, G.E. Machine learning-based statistical testing hypothesis for fault detection in photovoltaic systems. *Sol. Energy* 2019, 190, 405–413.
- [37] Haq, E.U.; Jianjun, H.; Li, K.; Ahmad, F.; Banjerdpongchai, D.; Zhang, T. Improved performance of detection and classification of 3-phase transmission line faults based on discrete wavelet transform and double-channel extreme learning machine. *Electr. Eng.* 2020, 103, 953–963.
- [38] Hussain, M.; Dhimish, M.; Titarenko, S.; Mather, P. Artificial neural network based photovoltaic fault detection algorithm integrating two bi-directional input parameters. *Renew. Energy* 2020, 155, 1272–1292.
- [39] Niu, H.; Omitaomu, O.A.; Cao, Q.C. Machine Committee Framework for Power Grid Disturbances Analysis Using Synchrophasors Data. *Smart Cities* 2021, 4, 1–16.
- [40] Zhang, D.; Han, X.; Deng, C. Review on the research and practice of deep learning and reinforcement learning in smart grids. *CSEE J. Power Energy Syst.* 2018, 4, 362–370.
- [41] Ni, Z.; Paul, S. A multistage game in smart grid security: A reinforcement learning solution. *IEEE Trans. Neural Netw. Learn. Syst.* 2019, 30, 2684–2695.
- [42] Cui, L.; Qu, Y.; Gao, L.; Xie, G.; Yu, S. Detecting false data attacks using machine learning techniques in smart grid: A survey. *J. Netw. Comput. Appl.* 2020, 170, 102808.
- [43] Ali, S.S.; Choi, B.J. State-of-the-Art Artificial Intelligence Techniques for Distributed Smart Grids: A Review. *Electronics* 2020, 9, 1030.

Forecasting Algeria's General Industrial Index: A Comprehensive Data-Driven Analysis and Predictive Insights

Siham Bouguern
Jawsak Group Algeria
siham.bouguern@jawsak.com

Abstract— The manufacturing sector serves as the cornerstone of a nation's economic, productive, and social influence. Given its substantial share of global energy consumption, embracing data-driven strategies represents a pivotal approach to informed decision-making. Machine learning techniques, renowned for their capacity to address intricate challenges, are increasingly favored for energy forecasting. This research introduces and assesses four analytical models - namely, Linear Regression, Exponential Smoothing, ARIMA, and SARIMA - in the context of predicting Algeria's General Industrial Index. In this study, these four models are applied to assess and predict Algeria's general industrial production for the forthcoming decade (2021–2030) by utilizing a 50-years dataset provided by The National Office of Statistics (NOS). The findings of this study reveal that the four forecasting methods provides a unique perspective on the future trajectory of the General Industrial Index in Algeria. Linear Regression is optimistic, Exponential Smoothing suggests stable growth, ARIMA predicts a decline, and SARIMA reinforces the declining trend with seasonality considerations. This paper significantly contributes to the literature by providing a comprehensive analysis of Algeria's general industrial production using four distinct forecasting models. The findings offer valuable insights for policymakers and industry stakeholders, facilitating informed decisions to sustain and enhance manufacturing output in Algeria.

Keywords— *manufacturing industries, index of industrial production, data analytics, forecasting, linear regression, exponential smoothing, ARIMA, SARIMA*

I. INTRODUCTION

We currently reside in the age of data, where the significance of data lies not in its sheer existence, but rather in its potential to inform decision-making processes. To address this, the field of "Data Analytics" has emerged, drawing on techniques and methods from data mining and statistical analysis, with contributions from scholars and practitioners in artificial intelligence (AI), algorithms, and database domains. These methods are designed to extract actionable insights from vast and diverse datasets [1]. A report by the McKinsey Global Institute in 2011 emphasizes that the analysis of extensive datasets will become a fundamental driver of competitiveness, productivity growth, and innovation, showcasing the diverse applications of big data in value creation [2]. Consequently, businesses are increasingly embracing precision marketing strategies to remain competitive and safeguard their profit margins. As a result, forecasting models have gained widespread use in precision marketing to comprehend and meet customer demands [3].

In the realm of intelligent manufacturing, industrial big data serves a dual purpose. It not only empowers enterprises to accurately perceive changes in the internal and external

environment but also facilitates data-driven analysis and decision-making to optimize production processes, reduce costs, and enhance operational efficiency [4]. Predictive analytics, a domain of growing significance, involves the application of statistical techniques to analyze historical data for predicting future events and behaviors [5]. Accurate and sophisticated demand forecasts are pivotal in making informed decisions related to inventory management, purchasing, and assortment planning, especially in the era of Industry 4.0. This era is characterized by reduced fulfillment times and the generation of vast amounts of data from advanced technologies, offering opportunities for data-driven decision-making, particularly in the context of demand forecasting [6]. Classical time series forecasting models have a long-standing history in forecasting within the business environment. Notably, the AutoRegressive Integrated Moving Average (ARIMA) model, known for its accuracy, expresses the time series structure through a relational equation, making its results highly persuasive for decision-makers. The Seasonal AutoRegressive Integrated Moving Average (SARIMA) model extends the capabilities of ARIMA by incorporating seasonality [7].

Global energy consumption is anticipated to surge by more than 50% by the year 2030 [8]. Algeria, in particular, presents an economic landscape marked by its heavy reliance on oil exports, comprising more than 97% of its total exports, and standing as one of the world's foremost exporters of natural gas. Moreover, the Algerian economy exhibits limited diversification in its production base, with a predominant emphasis on gas extraction industries [9]. On one hand, the escalation in gas and oil production is driven by the imperative to satisfy the burgeoning energy demands of households and domestic industries, often characterized by their high energy consumption, a common trait in oil-rich nations. Conversely, on the other hand, the amplification of gas and oil production serves the purpose of bolstering export volumes, thereby securing additional financial resource [9].

In this paper we analyzed the national industrial production in Algeria over the period 1970 - 2020. Using and comparing between the following models: Linear Regression, Exponential Smoothing, ARIMA, and SARIMA. The remainder of the paper is organized as follows: Section 2 represents the state of the art. Section 3 describes the evolution of industrial production in Algeria. Section 4 presents the methodology. Section 5 discusses the critical analysis and findings and section 6 concludes the paper along with some perspectives for future work.

II. STATE OF THE ART

A. Data Analysis for Forecasting in Manufacturing Sector

The industrial and manufacturing sector uses the most energy worldwide. To reduce global energy consumption, it's crucial to adopt strategies and techniques that improve energy efficiency and management. Belhadi et al. in [1] have devised an original model for encapsulating key features of Big Data Analytics (BDA) within the domain of manufacturing processes. This was achieved through a synthesis of current research outcomes and a comprehensive case study within a prominent manufacturer of phosphates derivatives. The purpose was to highlight the capabilities of BDA in manufacturing processes and provide suggestions for advancing research in this area. The insights derived from their study are expected to assist companies in comprehending the capabilities of BDA, recognizing potential impacts on their manufacturing processes, and facilitating the development of more efficient BDA-enabling infrastructure, thereby minimizing the risk of plagiarism detection [1]. While Qin & Chiang [2] delineate the trajectory of progress in machine learning and AI, propelled by advancements in statistical learning theory spanning the past two decades and underscored by the commercial achievements of prominent big data enterprises. Subsequently, they delve into an examination of the distinctive attributes of process manufacturing systems, followed by a concise overview of research and development in data analytics over the preceding three decades [2]. The investigation conducted by Wang et al. [4] offers an all-encompassing examination of interconnected subjects, including the definition of big data, as well as methodologies grounded in both model-driven and data-driven approaches. The study delves into the framework, evolution, pivotal technologies, and applications of BDA within the context of intelligent manufacturing systems. Additionally, the exploration accentuates the challenges faced and opportunities presented for prospective research in this domain [4].

The increasing popularity of machine learning techniques for predicting energy consumption is attributed to their proficiency in addressing complex non-linear challenges. Mawson & Hughes [8] undertake a comparative analysis, evaluating the efficacy of two deep neural networks—feed-forward and recurrent—in predicting energy consumption and workshop conditions within manufacturing facilities. The predictive models rely on a combination of production schedules, climatic conditions, thermal properties of the facility building, and considerations of building behavior and use [8]. Amidst the COVID-19 pandemic, numerous studies have been undertaken, and in [5], Sheng et al. offer a review focusing on methodological innovations in the exploration of BDA. The study elucidates on how these innovations can be optimally employed to scrutinize current organizational challenges. Notably, the authors provide insights into methodologies encompassing descriptive/diagnostic, predictive, and prescriptive analytics, elucidating their potential application in the examination of unforeseen and impactful events, exemplified by the global crisis resulting from COVID-19. The study emphasizes the implications of such events for managers and policymakers [5].

Furthermore, the utilization of Predictive Analytics (PA) in the oversight of supply chains has gained considerable prominence in recent years, particularly with a focus on demand forecasting. As expounded in [2], Falatouri et al. offer

a comprehensive overview of methodologies in Retail Supply Chain Management (SCM) and undertake a comparative analysis of two selected methods. The analysis is grounded in the examination of more than 37 months of authentic retail sales data obtained from an Austrian retailer. The study involves the training and evaluation of SARIMA and LSTM (Long Short-Term Memory) models, both of which demonstrated satisfactory to commendable results based on the provided data [6]. Seyedan and Mafakheri [3] conduct an inquiry into the applications of predictive BDA in demand forecasting within supply chains. Their investigation aims to categorize these applications, discern gaps in existing research, and offer insights for prospective studies. The authors classify algorithms and their applications in SCM into distinct categories, encompassing time-series forecasting, clustering, KNN (K-nearest-neighbors), neural networks, regression analysis, support vector machines, and support vector regression [3]. The advancement of AI has facilitated remarkably precise demand forecasting. However, it is essential to recognize that achieving heightened forecast accuracy does not automatically result in reduced inventory costs or enhanced service levels within supply chain and inventory management. In response to this consideration, Shibayama et al. [7] have devised a framework for demand forecasting that prioritizes both accuracy and interpretability. The framework leverages time series decomposition and ARIMA methods. Specifically, Seasonal-trend decomposition using locally estimated scatterplot smoothing (STL) is employed to break down a time series into its components—trend, seasonality, and residual—providing decision makers with a comprehensible foundation for understanding changes in demand. This approach is designed to support decision-makers in making informed choices in the realm of demand forecasting [7].

B. CO₂ Emissions Forecasting

The article by Dragomir et al. [10] centers on the evaluation of CO₂ emissions and their correlation with the national industrial production of Romania. Through an examination of CO₂ emissions and the production of primary industrial goods, the study draws two noteworthy conclusions: Firstly, the trajectory of total production significantly influences emissions. Secondly, alterations in the industrial structure contribute distinctly to emissions during specific periods, such as 1993–1998 and 2002–2012, wherein increased production aligns with corresponding changes in the CO₂ emission trends [10]. In the study by Zhou et al. [11] conducted in China, three primary sources of carbon emissions spanning the period from 1990 to 2017 were identified: the energy industry, fuel combustion in other industries, and industrial processes. The paper outlines the development of a driving force model for each emission source using multiple linear regression. Utilizing these models, the study generates forecasts for both carbon intensity and total CO₂ emissions from the year 2018 to 2030. The findings indicate a continued reduction in both CO₂ emission intensity and total emissions; however, the study underscores the necessity for additional efforts to attain the objectives outlined in the Paris Agreement [11].

Within the realm of developing nations, Algeria emerges as a noteworthy contributor to CO₂ emissions, securing the third position among African countries in this context. Addressing this concern, Bouziane et al. [12] conducted similar studies in Algeria. In their work, they delineate a hybrid approach employing Artificial Neural Networks

(ANN) and an agent-based architecture for predicting carbon dioxide (CO₂) emissions originating from various energy sources in the city of Annaba, utilizing authentic data. The system is composed of multiple autonomous agents, classified into two types: forecasting agents, responsible for predicting the production of specific energy types using ANN models, and core agents, tasked with critical functions such as calculating equivalent CO₂ emissions and managing the simulation [12]. In their analytical review, Hicham & Moussa [13] investigate the interconnections among energy consumption, economic growth, and CO₂ emissions in Algeria spanning the period from 1970 to 2017. Employing a structural Vector Autoregression (VAR) approach, the study's outcomes indicate that a positive shock in CO₂ emissions corresponds to an increase in both economic growth and energy consumption. Additionally, the study reveals that a positive shock in energy consumption has a minimal positive effect on economic growth but a substantial negative impact on CO₂ emissions [13]. The paper by Bouznit and Pablo-Romero [9] seeks to examine the relationship between CO₂ emissions and economic growth in Algeria, considering factors such as energy use, electricity consumption, exports, and imports. The study assesses the validity of the Environmental Kuznets Curve (EKC) hypothesis for the period spanning 1970 to 2010, employing the Autoregressive Distributed Lag model extended to incorporate break points. The findings affirm the applicability of the EKC hypothesis for Algeria, suggesting a nonlinear relationship between CO₂ emissions and economic growth during the specified time frame [9]. The study of Chekouri et al. [14] conducted a stochastic impact by regression on population, affluence, and technology model to identify the determinant factors driving CO₂ emissions in Algeria during the period 1971–2016. The method of partial least squares regression is applied to eliminate multicollinearity problems. The results indicate that the population has a positive and significant effect on CO₂ emission. The energy use is found to be the second most contributing factor to CO₂ emissions followed by urbanisation and affluence (GDP per capita) [14]. In their work, Abounoori and Bagherpour [15] introduced a Hybrid Regression Neural Network approach with the aim of achieving improved fitness compared to traditional Regression Analysis and Neural Network methods. Through a comparative analysis of the estimated results obtained from Regression Analysis and Neural Networks against those derived from the Hybrid Neural-Regression method, the study highlights the superior performance of the latter approach [15].

III. ANALYSIS OF THE EVOLUTION OF INDUSTRIAL PRODUCTION IN ALGERIA (1970 - 2020)

In Algeria, the industrial sector witnessed significant growth in the 1970s and early 1980s. However, during the 1990s, this sector, and the overall economy, faced a period of weak performance. The Ministry of Industry and Restructuring reported a 25.8% decline in industrial production between 1989 and 1998, with production capacity utilization remaining low at around 30% to 60%. Additionally, the industrial sector's value-added saw insufficient representation, and exports of industrial products stagnated.

However, in 1998, there was a notable recovery, as indicated by a 10.50% growth in the industrial production index compared to previous years, according to ministry data. The National Office of Statistics (NOS) in 2010 attributed the

declining growth rate to a significant decrease in hydrocarbon production levels in the fourth quarter of 2008 (-7.5%) and the first three quarters of 2009 (-9.9%, -8.7%, and -3.6%), resulting in an overall growth rate of -5.4% for the year.

Despite the overall decline in industrial activity from 1998 to 2008, dropping from 69.0% to 52.8%, manufacturing industries managed to achieve a modest growth of 1.9% in 2008, even though there was a 2.4% decline in the third quarter of the same year following several years of negative growth rates since 1999. In more recent years, the industrial sector in Algeria has shown some resilience and growth, as indicated by the data for 2010-2020, though with fluctuations [16]. Figure 1 presents the evolution and the variations of the annual index of industrial production of the national public sector:

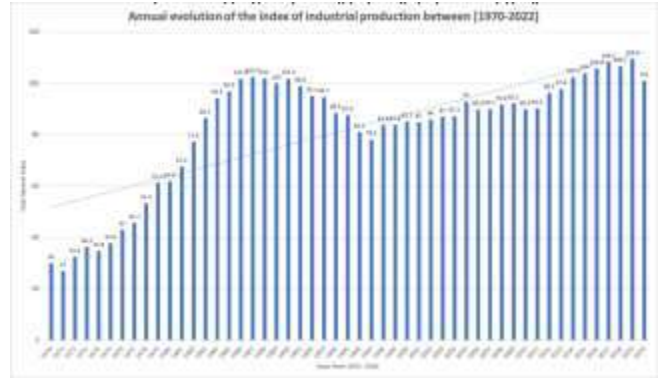


Fig. 1. Annual evolution of the index of industrial production between [1970-2020]

IV. METHODOLOGY

A. Data Collection and Preprocessing

The datasets of the General Index were collected by The National Office of Statistics (NOS) of Algeria [17] starting from 1970 until 2020. The 50 years of values recorded for all industries including: Water-Energy, Hydrocarbons Industries, Mines and Quarries, S.M.M.E.E.I (Steel, Metallurgical, Mechanical, Electrical and Electronic Industries), Materials of Construction, Ceramic, Glass, Chemistry, Rubber, Plastics, Agro-Food industry, Tobacco, Matches, Textile industries, Leather and Footwear Industries, Wood, Cork, Paper Industries, and other diverse Industries. In addition to the Total General Index, Total non-hydrocarbon industries index, and Total Manufacturing Industries Index.



Fig. 2. Python code snippet for checking missing values

Figure 2 shows the results of missing value search, as indicated there are no missing values.

B. Descriptive Analysis of the Used Data

In Figure 3 the main information of the variables used in this analysis are shown. While Figure 4 presents the main descriptive statistics of the dataset used in this analysis.

```

Information on Dataframe:
=====
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype  
---  ------  -
0   Year     51 non-null    int64  
1   GI       51 non-null    float64 
dtypes: float64(1), int64(1)
memory usage: 944.0 bytes
None

```

Fig. 3. Python code output snippet for dataset information

```

Descriptive Statistics:
=====

```

	count	mean	std	min	25%	50%	75%	max
Year	51.0	1995.000000	14.866089	1970.0	1982.50	1995.0	2007.50	2020.0
GI	51.0	82.076471	23.337134	27.0	77.75	89.9	98.35	109.6

Fig. 4. Python code output snippet for descriptive statistics

To display the general trend of the Algerian Total General Index from 1970 to 2020, Figure 5 is presented.



Fig. 5. Algeria total general index [1970-2020]

C. Selection of Forecasting Methodologies

In selecting a forecasting model, careful consideration of the data's characteristics, model interpretability, and underlying assumptions is crucial. With the available models, including Linear Regression, Exponential Smoothing, ARIMA, and SARIMA, the choice primarily depends on the specific context. Here are the main selection criteria:

- **Data Characteristics:** Considering the nature of time series data, including its stationarity, seasonality, and trends.
- **Interpretability:** Evaluating the model's ability to provide interpretable forecasts and insights.
- **Simplicity:** Choosing a model that balances complexity with the research's practicality.

- **Robustness:** Assessing the model's resistance to outliers and extreme values in the data.
- **Relevance:** Ensuring the selected model aligns with the research objectives and domain knowledge.
- **Performance Validation:** Conducting rigorous performance assessment, such as cross-validation, to gauge the model's accuracy.
- **Limitations:** Acknowledging and consider the limitations inherent to the chosen model.
- **Seasonality:** If present, select models capable of accounting for seasonality patterns.
- **Domain Expertise:** Leveraging domain-specific knowledge when making the final model selection.

V. CRITICAL ANALYSIS AND FINDINGS

A. Method 1: Linear Regression

The Linear Regression model suggests a relatively optimistic outlook for the General Industrial Index in Algeria. The forecast indicates a consistent upward trend in the Index over the next ten years.

```

Forecast results Using Linear Regression:
=====

```

	Year	Forecasted Index
0	2021	113.510118
1	2022	114.719104
2	2023	115.928090
3	2024	117.137077
4	2025	118.346063
5	2026	119.555050
6	2027	120.764036
7	2028	121.973023
8	2029	123.182009
9	2030	124.390995

```

=====

```

Fig. 6. Forecast results of the Linear Regression Model

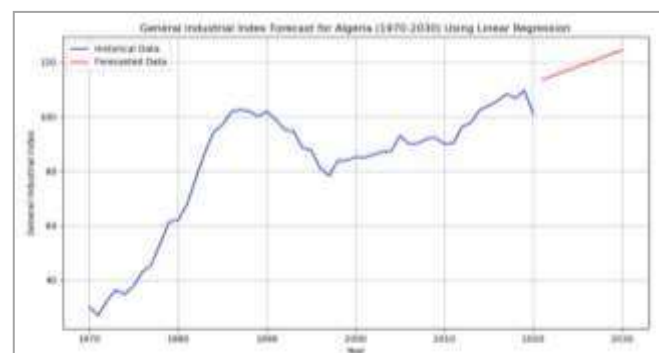


Fig. 7. General industrial index forecast for Algeria using Linear Regression Model

According to this model, the industrial sector is expected to experience a steady and positive growth trajectory, with the Index increasing at an annual rate of approximately 1.22 index points. This implies confidence in the Algerian industrial sector's potential for expansion and development (See Figure 7).

B. Method 2: Exponential Smoothing

The Exponential Smoothing model provides a forecast with a relatively stable trajectory for the General Industrial Index in Algeria. It anticipates a moderate and consistent growth pattern with minor fluctuations.

Forecast results Using Exponential Smoothing:		
	Year	Forecasted Index
0	2021	100.505293
1	2022	100.685924
2	2023	100.216032
3	2024	101.746320
4	2025	100.751280
5	2026	103.081338
6	2027	103.461049
7	2028	101.666272
8	2029	103.546164
9	2030	101.751232

Fig. 8. Forecast results of Exponential Smoothing Model

Although the growth rate is relatively low, this model suggests that the Index will maintain its stability over the coming decade, with occasional short-term variations. It portrays a scenario of steady, albeit slow, progress in the industrial sector (See Figure 9).

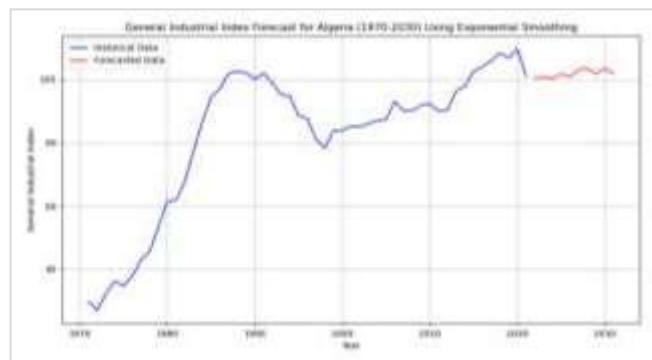


Fig. 9. General industrial index forecast for Algeria using Exponential Smoothing Model

C. Method 3: ARIMA

The ARIMA model, in contrast to the previous models, paints a less optimistic picture for the General Industrial Index in Algeria. It forecasts a declining trend in the Index over the next ten years, indicating potential challenges or structural changes within the industrial sector.

The forecasted decrease in the Index is relatively significant, with the Index declining each year. This model implies that the industrial sector might face headwinds and contraction in the near future (See Figure 11).

D. Method 4: SARIMA

The SARIMA model, which accounts for seasonality and trends in the data, presents a forecast that aligns with the declining trend seen in the ARIMA model. It suggests a decreasing trajectory for the General Industrial Index in Algeria over the next ten years.

Forecast results Using ARIMA:

	Year	Forecasted Index
0	2021	99.232382
1	2022	97.963523
2	2023	94.459775
3	2024	93.549295
4	2025	92.076708
5	2026	90.539457
6	2027	89.808227
7	2028	88.831117
8	2029	88.069560
9	2030	87.531452

Fig. 10. Forecast results of ARIMA Model



Fig. 11. General industrial index forecast for Algeria using AMIRA Model

While the decline is not as sharp as in the ARIMA model, it still indicates a challenging period ahead for the industrial sector. The SARIMA model highlights the importance of considering seasonality in the context of the forecast (See Figure 13).

Forecast results Using SARIMA:

	Year	Forecasted Index
0	2021	98.152887
1	2022	98.029634
2	2023	96.769332
3	2024	98.208912
4	2025	97.305021
5	2026	99.395185
6	2027	99.709836
7	2028	99.063797
8	2029	101.991889
9	2030	99.660582

Fig. 12. Forecast results of SARIMA Model

In summary, each of the four forecasting methods provides a unique perspective on the future trajectory of the General Industrial Index in Algeria. Linear Regression is optimistic, Exponential Smoothing suggests stable growth, ARIMA predicts a decline, and SARIMA reinforces the declining trend with seasonality considerations.



Fig. 13. General industrial index forecast for Algeria using SAMIRA Model

VI. CONCLUSION AND FUTURE WORK

In this paper we applied four different methods to test and forecast the general industrial production for Algeria over the next 10 years (2021–2030). Using the 30 years dataset collected by The National Office of Statistics (NOS), we observed the trend of Algerian production of main industrial. In general, there is a degree of disagreement among the models regarding the future trend of the General Industrial Index in Algeria. Linear Regression forecasts steady growth, Exponential Smoothing suggests stability with minor fluctuations, ARIMA predicts a declining trend, and SARIMA indicates a decreasing trend with some variations.

Our major contribution to the literature is the analysis of industrial production. By collecting and evaluating this data, technical measures can be taken to maintain a good level of manufacturing production in Algeria. In conclusion, forthcoming studies could expand these results with additional appraisal of the different common factors. This assessment of the uncertainties in a resolution is an essential element of reasonable, accurate information policy-making. Future work could study and select the most useful method given the domain knowledge.

REFERENCES

- [1] A. Belhadi, K. Zkik, A. Cherrafi, Y. M. Sha'ri and S. E. fezazi, "Understanding the capabilities of Big Data Analytics for manufacturing process: insights from literature review and multiple case study," *Computers & Industrial Engineering*, 2019.
- [2] S. J. Qin and L. H. Chiang, "Advances and opportunities in machine learning for process data analytics," *Computers and Chemical Engineering*, 2019.
- [3] M. Seyedan and F. Mafakheri, "Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities," *Journal of Big Data*, 2020.
- [4] J. Wang, C. Xu, J. Zhang and R. Zhong, "Big data analytics for intelligent manufacturing systems: A review," *Journal of Manufacturing Systems*, 2021.
- [5] J. Sheng, J. Amankwah-Amoah, Z. Khan and X. Wang, "COVID-19 Pandemic in the New Era of Big Data Analytics: Methodological Innovations and Future Research Directions," *British Journal of Management*, 2021.
- [6] T. Falatouri, F. Darbanian, P. Brandtner and C. Udokwu, "Predictive Analytics for Demand Forecasting – A Comparison of SARIMA and LSTM in Retail SCM," in *3rd International Conference on Industry 4.0 and Smart Manufacturing*, 2022.
- [7] H. Shibayama, A. Ishigaki, T. Kobe, T. Ueda, D. Arimizu and T. Nakamura, "Design of a Framework for Demand Forecasting Using Time Series Decomposition-Based Approach," in *4th Asia Pacific Conference of the Prognostics and Health Management*, 2023.

- [8] V. J. Mawson and B. R. Hughes, "Deep learning techniques for energy forecasting and condition monitoring in the manufacturing sector," *Energy & Buildings*, 2020.
- [9] M. Bouznit and M. d. P. Pablo-Romero, "CO2 emission and economic growth in Algeria," *Energy Policy*, p. 93–104, 2016.
- [10] C. M. B. Dragomir, G. Podaru and C. Muntenita, "A Bayesian Analysis of CO2 Emissions and National Industrial Production in Romania," in *Mechanisms and Machine Science*, 2022, pp. 683–690.
- [11] Y. Zhou, J. Zhang and S. Hu, "Regression analysis and driving force model building of CO2 emissions in China," *Scientific Reports*, p. 2021.
- [12] S. E. Bouziane, M. T. Khadir and J. Dugdale, "A collaborative predictive multi-agent system for forecasting carbon emissions related to energy consumption," *Multiagent and Grid Systems*, pp. 39–58, 2021.
- [13] A. Hicham and C. Moussa, "A Structural VAR analysis of Energy Consumption, Economic Growth and CO2 Emissions: Evidence from Algeria," *Journal of Managerial and Economic Research*, pp. 101–114, 2019.
- [14] S. M. Chekouri, A. Chibi and M. Benbouziane, "Examining the driving factors of CO2 emissions using the STIRPAT model: the case of Algeria," *International Journal of Sustainable Energy*, 2020.
- [15] E. Abounoori and M. Bagherpour, "Estimation of Industrial Production Costs, Using Regression Analysis, Neural Networks or Hybrid Neural - Regression Method?," *Iranian Economic Review*, vol. 11, no. 16, pp. 17–29, 2006.
- [16] N. Bouri, "The manufacturing sector in Algeria: analysis and interpretation of annual variations," *Revue du Lareid*, 2020.
- [17] "Industrial Production Index - National Public Sector: 1970 to 1976," The National Office of Statistics, 2020.

APPENDICES

Appendix A: Code snippet

```

1.1 Importing Libraries & Dataset
# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score

# Load the dataset
data = pd.read_csv('data/industrial_production.csv')

1.2 Exploratory Data Analysis (EDA)
# Display the first few rows of the dataset to inspect the data
data.head()

# Extract the year and index data
years = data['Year'].values.reshape(-1, 1)
index = data['Index'].values

# Check for missing values
data.isnull().sum()

Year    0
Index    0
dtype: object

2 Method 1: Linear Regression
# Create a linear regression model
model = LinearRegression()
model.fit(years, index)

# Predict the index for the next 10 years
future_years = np.arange(2021, 2031).reshape(-1, 1)
forecasted_index = model.predict(future_years)

# Print the historical data and the forecasted values
pd.figure(figsize=(10, 6))
plt.scatter(years, index, label='Historical Data', color='b')
plt.plot(future_years, forecasted_index, label='Forecasted Data', color='r')
plt.xlabel('Year')
plt.ylabel('General Industrial Index')
plt.title('General Industrial Index Forecast for Algeria (1970-2030) Using Linear Regression')
plt.grid(True)

# Save the figure
plt.savefig('General Industrial Index Forecast for Algeria (1970-2030) Using Linear Regression.png')

# Show the plot
plt.show()

```

3 Method 2: Exponential Smoothing

```
# Extract the year and index data
years = data['Year']
index = data['I']
# Create a DataFrame with the extracted data
df = pd.DataFrame({'Year': years, 'Index': index})

# Generate a date-based index
start_year = df['Year'].min()
end_year = df['Year'].max()
date_index = pd.date_range(start=f'{start_year}-01-01', end=f'{end_year}-12-31', freq='B')
# Set the date-based index for the DataFrame
df.set_index(date_index, inplace=True)

# Perform exponential smoothing forecasting
model = ExponentialSmoothing(df['Index'], seasonal='add', seasonal_periods=12)
result = model.fit()

# Forecast the index for the next 10 years
forecasted_index = result.predict(start=end_year, end=end_year + 10)

# Generate a date-based index for the forecasted values
future_index = [df.index[-1] + timedelta(days=365*i) for i in range(1, 11)]
forecasted_index.index = future_index

# Plot the historical data and the forecasted values
plt.figure(figsize=(12, 8))
plt.plot(df.index, df['Index'], label='Historical Data', color='b')
plt.plot(forecasted_index.index, forecasted_index.values, label='Forecasted Data', color='r')
plt.xlabel('Year')
plt.ylabel('General Industrial Index')
plt.title('General Industrial Index Forecast for Algeria (1970-2030) Using Exponential Smoothing')
plt.legend()
plt.grid(True)
# Save the figure
plt.savefig('General Industrial Index Forecast for Algeria (1970-2030) Using Exponential Smoothing.png')
# Show the plot
plt.show()
```

4 Method 3: ARIMA

```
# Fit an ARIMA model
model = ARIMA(df['Index'], order=(5, 1, 4)) # Set the initial the order parameter based on your data
result = model.fit()

# Forecast the index for the next 10 years
forecasted_index = result.predict(start=end_year, end=end_year + 10)

# Generate a date-based index for the forecasted values
future_index = [df.index[-1] + pd.Timedelta(days=365*i) for i in range(1, 11)]
forecasted_index.index = future_index

# Plot the historical data and the forecasted values
plt.figure(figsize=(12, 8))
plt.plot(df.index, df['Index'], label='Historical Data', color='b')
plt.plot(forecasted_index.index, forecasted_index.values, label='Forecasted Data', color='r')
plt.xlabel('Year')
plt.ylabel('General Industrial Index')
plt.title('General Industrial Index Forecast for Algeria (1970-2030) Using ARIMA')
plt.legend()
plt.grid(True)
# Save the figure
plt.savefig('General Industrial Index Forecast for Algeria (1970-2030) Using ARIMA.png')
# Show the plot
plt.show()
```

5 Method 4: SARIMA

```
# Fit a SARIMA model
# For now set the order and seasonal_order parameters based on your data and domain knowledge
order = (5, 1, 4)
seasonal_order = (1, 1, 1, 12)
model = SARIMA(df['Index'], order=order, seasonal_order=seasonal_order)
result = model.fit()

# Forecast the index for the next 10 years
forecasted_index = result.get_forecast(steps=10)

# Generate a date-based index for the forecasted values
future_index = [df.index[-1] + pd.Timedelta(days=365*i) for i in range(1, 11)]
forecasted_index.index = forecasted_index.predicted_mean.index
forecasted_index.index = future_index

# Plot the historical data and the forecasted values
plt.figure(figsize=(12, 8))
plt.plot(df.index, df['Index'], label='Historical Data', color='b')
plt.plot(forecasted_index.index, forecasted_index.values, label='Forecasted Data', color='r')
plt.xlabel('Year')
plt.ylabel('General Industrial Index')
plt.title('General Industrial Index Forecast for Algeria (1970-2030) Using SARIMA')
plt.legend()
plt.grid(True)
# Save the figure
plt.savefig('General Industrial Index Forecast for Algeria (1970-2030) Using SARIMA.png')
# Show the plot
plt.show()
```

Balanced exploration and exploitation properties of multi-objective sooty tern optimizer

Farid Boumaza
dept. Computer Science
(LAPECI)Laboratory of Parallel,
Embedded architectures and Intensive
Computing, University of Oran1 Oran,
Algeria farid.pgia@gmail.com

Djaafar Zouache
dept. Computer Science University of
Bordj Bou ArreridjBordj Bou Arreridj,
Algeria djaafarzouache@yahoo.fr

Abou El hassane Benyamina
dept. Computer Science
(LAPECI)Laboratory of Parallel,
Embedded architectures and Intensive
Computing, University of Oran1 Oran,
Algeria benyanabou@yahoo.fr

Abstract—The Sooty Tern Optimization Algorithm (STOA)[1], a newly designed metaheuristic drawing inspiration from the migratory and predatory behaviors of the sea bird Sooty Tern, presents several notable advantages, including minimal parameterization. However, it is confined to addressing single-objective problems exclusively. In this work, we have enhanced the properties of exploration and exploitation to effectively penetrate the area of search. Subsequently, we extended the STOA to a multi-objective version called MOSTOA (Multi-Objective Sooty Tern Optimization Algorithm). This enhanced algorithm is designed to address multiple objectives in diverse problem domains. The MOSTOA utilizes an archive repository to store and retrieve the optimal solutions generated throughout the optimization cycle. From this population archive, leaders are chosen to guide the solutions of the main population towards promising search locations. Furthermore, the utilization of the grid mechanism and dynamic archiving approach serves the purpose of achieving a harmonious equilibrium between convergence and variety inside the final Pareto set. These strategies ensure that the obtained solutions exhibit both high quality and spread across the objective space. The proposed MOSTOA is validated on various well-known benchmarks functions. In addition, its performance is assessed in comparison to well-established cutting-edge algorithms. Our method produces very competitive results and, in most circumstances, exhibits improved convergence behavior with a good variety of solutions, as demonstrated by the experimental findings.

Keywords— Multi-Objective optimization, Swarm Intelligence, Dominance relation, Metaheuristic, Pareto Front

I. INTRODUCTION

In recent years, there have been many multi-objective optimizers proposed with the purpose of addressing multi-objective optimization problems (MOPs), such as: MODA [1], MOMFO [2], MOALO [3], MOGWO [4], MOCSSO [5], MOHHO [6] ...etc. As a consequence of the contradictory characteristics of objectives, there is generally no one optimum solution in multi-objective optimization, but instead a set of proposal known as Pareto optimal (Ps) [7]. Since no other solutions exist in the area search that are superior for all of the objectives covered, these solutions are considered optimum. As expected, significant challenges were posed that need specific strategies to handle them, including the harmony between the exploration and the exploitation properties, since these are the essential criteria in multi-objective optimization. In this paper, we present our suggested algorithm based on an equilibrium of exploration and exploitation properties. The remainder of this study is arranged in the following manner: Section 2 offers the core concepts of MOPs and provides a brief background on STOA. Section 3 presents our innovative Multi-Objective Sooty Tern Optimization Algorithm

(MOSTOA). Section 4 contains the results of the experiment and discussions, while Section 5 summarizes our findings and offers recommendations for future research.

II. BACKGROUND

A. Multi-Objective Optimization

The typical structure of a MOP can be characterized in the following manner:

$$\begin{aligned} & \text{Maximize : } f_m(x), \quad m = 1, 2, \dots, M \\ & \text{Subject to : } g_j(x) = 0, \quad j = 1, 2, \dots, J \\ & \quad \quad \quad h_k(x) = 0, \quad k = 1, 2, \dots, K \\ & \quad \quad \quad L_i \leq x_i \leq U_i, \quad i = 1, 2, \dots, n \end{aligned} \quad (1)$$

In this context, x represents a solution comprising n choice individuals (x_1, x_2, \dots, x_n) that must adhere to J and K constraints. The function M quantifies the number of objectives, while L_i and U_i denote the minimum and maximum bounds for each decision individual.

B. Pareto dominance

Dominance relation. A solution $x^{(i)}$ is considered to dominate another solution $x^{(j)}$, represented as $x^{(i)} < x^{(j)}$ when two specific conditions hold true [8] :

- $\forall m \in \{1, \dots, M\} : f(x^{(i)}) \leq f(x^{(j)})$
- $\exists \mu \in \{1, \dots, M\} : \phi(x^{(i)}) < \phi(x^{(j)})$

Non-dominated collection. Within a given solution collection, the non-dominated solutions can be defined as those within the subset $A' \subseteq A$ that are not outperformed by any other member within the same collection A [9].

Pareto solution collection. The collection of non-dominated solutions within the entire area search S is referred to as the Pareto optimal set [10].

B. Sooty Tern Optimization Algorithm (STOA)

Sooty terns, as omnivorous seabirds, sustain themselves through a diverse diet comprising earthworms, insects, and fish. These avian species tend to congregate in colonies and employ swarm intelligence to efficiently locate and capture their prey.

Within the realm of the sooty tern's predation process, migration and attack emerge as two pivotal behaviors,

offering essential heuristic foundations for constructing the STOA [11]. The subsequent section delves into the mathematical representations of these migration and attack behaviors

Exploration Phase. Sooty terns embark on their migratory journeys primarily to mitigate the risk of mid-air collisions, a phenomenon that can be succinctly framed as:

$$L_{st} = K_A \times R_{st}(t) \quad (3)$$

$$K_A = L_f - (t \times (L_f / \text{MaxT})) \quad (4)$$

In this context, L_{st} represents the desired position of a search individual to avoid collisions with others. $R_{st} \square t \square$ denotes the current position of the search individual at iteration t , where

K_A signifies movement within the search space. The variable L_f serves as a control parameter responsible for progressively reducing the search amplitude (K_A) in a linear manner from its initial value L_f to zero. Following avoiding collisions, the search individuals efficiently go towards the vicinity of the optimal individual by employing the subsequent mathematical expressions:

$$D_{st} = L_B \times (R_{bst}(t) - R_{st}(t)) \quad (5)$$

$$L_B = 0,5 \times \text{Rd} \quad (6)$$

D_{st} quantifies the positional disparity between the search individual R_{st} and the best individual R_{bst} . Meanwhile, L_B plays a pivotal role in improving exploration and is determined by a uniformly distributed random variable Rd , which ranges between 0 and 1.

In the end, every individual has the opportunity to adjust its position based on the best search individual.

$$E_{st} = L_{st} + D_{st} \quad (7)$$

E_{st} signifies the disparity between the individual and the fittest individual.

Exploitation Phase. During their hunt for prey, sooty terns display a distinctive spiral flight pattern, which can be characterized as follows:

$$s_1' = \sin(j) \times \text{Rad} \quad (8)$$

$$s_2' = \cos(j) \times \text{Rad} \quad (9)$$

$$s_3' = j \times \text{Rad} \quad (10)$$

$$\text{Rad} = a \times e^{hb} \quad (11)$$

In this context, Rad denotes the distance between successive turns in a spiral, 'i' is a number within $[0 < h < 2\pi]$, and 'a' and 'b' are constants that shape the spiral, both set to 1 for our analysis. Subsequently, we use the following equation to update the search individual's position:

$$R_{st}(t+1) = (E_{st} \times (s_1' + s_2' + s_3')) \times R_{bst}(t) \quad (12)$$

In the context of this study, the term ' $R_{st}(t+1)$ ' denotes the process responsible for updating the positions of other search individuals while ensuring the preservation of the optimal solution.

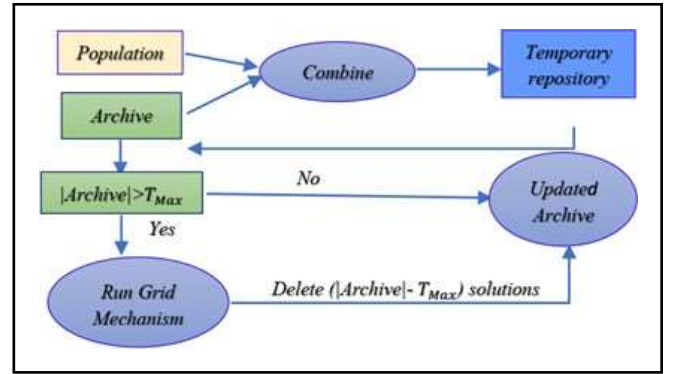
III. THE EXTENDED SOOTY TERN OPTIMIZATION ALGORITHM FOR ADDRESSING THE MOP

This part outlines the proposed Multi-Objective Sooty Tern

Optimization Algorithm (MOSTOA), designed for tackling MOPs. In MOP, the effectiveness of an algorithm hinges significantly on its elitism strategies, which involve acquired during optimization. This archive houses the Pareto front, necessitating careful updates throughout the optimization process to balance both convergence and diversity. Moreover, the selection of the global best solution, denoted as R_{bst} , plays a pivotal role in MOSTOA. It directly influences the movements of sooty tern within the search space, thereby impacting the exploitation of promising regions.

The archive update mechanism serves to distinguish potential solutions within the main population, separating accepted and rejected ones. To achieve this, MOSTOA combines the current population and archive into a temporary repository. From this repository, it identifies and stores the set

of non-dominated solutions into the archive. However, this may lead to a new Pareto set that surpasses the maximum archive size (T_{max}). In such cases, an adaptive grid cells strategy, similar to the one outlined in [12], is employed as a density estimator. Its purpose is to remove the most crowded solutions ($T_{max} - |\text{archive}|$) from the archive, with the aim of retaining well-distributed solutions in the search space. For a visual representation of this archiving strategy, refer to Fig. 1.



The grid strategy serves the purpose of partitioning the external archive into a specific number of hypercubes, with each hypercube containing a defined number of solutions. This mechanism functions as a density estimator, with the quantity of solutions within a particular hypercube representing the density of that hypercube. Additionally, the density of a particular solution corresponds to the density of the hypercube it belongs to. This density measurement plays a critical role in identifying both the sparsely populated and densely populated hypercubes. New non-dominated solutions are inserted into the less crowded hypercubes, while certain solutions are removed from the densely populated ones when the archive reaches its capacity. Consequently, this approach helps maintain a well-distributed set of solutions within the archive.

A. Global best solution (R_{bst}) selection approach

In our proposed multi-objective algorithm, a key strategy revolves around the selection of the global best solution, denoted as R_{bst} . This particular individual plays a pivotal role in guiding its fellow solutions towards promising areas within the search space, with an emphasis on achieving both convergence and diversity.

To identify R_{bst} in the MOSTOA, we adopt a method that involves selecting it from the less densely populated

hypercubes in the objective space. This selection process is based on a probability calculation using the following formula:

$$P_i = v / T_i \quad (13)$$

Here, 'v' represents a constant value greater than 1, while ' T_i ' signifies the number of solutions contained within the i^{th} hypercube. Essentially, this probability calculation favors hypercubes with fewer solutions, effectively prioritizing less crowded hypercubes for selection.

Furthermore, once an appropriate hypercube is chosen using this method, a roulette-wheel selection mechanism is employed to randomly pick an individual as the R_{bst} . This

C. The Pseudocode of MOSTOA algorithm

Algorithm: Multi-Objective-STOA

```

Initialize the population of Sooty Tern:  $X_i$ ;  $i=1...N$ ;
Evaluate each solution in  $X_i$ ;
Set  $T_{max}$  as the size of the archive population;
Initialize the archive (Arch) with the non-dominated
solution from  $X_i$ ;
While  $t < MaxT$  do
    For  $i=1$  to  $N$  do
        Select a global best solution  $R_{bst}$ ;
        Update the positions of individuals using Eq.12
    End
    Update the parameters  $K_A$  and  $L_B$ 
    Evaluate the new population
    Update the current archive (Arch)
    If Arch is full then
        Run the grid strategy;
        Remove the last ( $T_{max}-|archive|$ ) solutions;
    End
End
Return the final archive Arch

```

II. RESULTS AND DISCUSSION

In assessing the effectiveness of the novel Multi-Objective Sooty tern optimization algorithm (MOSTOA), a set of benchmark functions from the ZDT series [13] is employed.. To gauge the competitiveness of our approach, a comprehensive qualitative and quantitative evaluation is conducted in comparison to three established algorithms in the field, namely, MOEA/D [14], MOGWO [4], and MSSA [15]. For the benchmarking process, a uniform configuration is maintained across all algorithms, encompassing a population size of 100, an archive size of 100, and a fixed number of generations set at 1000.

To enable a rigorous quantitative comparison, two widely recognized performance metrics, namely Inverted Generational Distance (IGD) [16] and Hypervolume (HV) [17], are harnessed. IGD serves as a measure of convergence, offering insights into the algorithm's ability to approach the true Pareto front, while HV quantifies the diversity within the obtained Pareto front. These metrics collectively provide a comprehensive assessment of MOSTOA's performance against its contemporaries. *The finding of ZDT test functions*

A. The finding of ZDT test functions

The statistical outcomes, encompassing the best, worst, mean, median, and standard deviation (std) values of the IGD and HV metrics derived from the algorithms, are presented in Tables I and II, with the superior outcomes highlighted in bold. Furthermore, in Figures 2 and 3, we visually depict the most optimal Pareto fronts obtained by each algorithm.

TABLE I. COMPARATIVE RESULTS OF IGD METRIC

		MOSTOA	MOGWO	MSSA	MOEA/D
ZDT1	best	8,74E-05	6,10E-03	7,20E-03	3,14E-03
	worst	8,00E-05	5,89E-04	3,15E-03	5,41E-04
	mean	8,32E-05	1,50E-03	4,73E-03	1,15E-03
	median	8,29E-05	1,29E-03	4,46E-03	7,71E-04
	std	1,92E-06	9,63E-04	1,06E-03	7,65E-04
ZDT2	best	2,52E-04	2,31E-02	1,16E-02	6,24E-02
	worst	8,90E-05	7,93E-04	3,65E-03	3,99E-04
	mean	1,07E-04	6,45E-03	5,30E-03	2,20E-02
	median	9,79E-05	1,59E-03	5,10E-03	1,86E-02
	std	3,03E-05	9,16E-03	1,43E-03	1,54E-02
ZDT3	best	1,33E-04	3,26E-03	1,16E-02	2,39E-02
	worst	9,21E-05	3,06E-04	4,50E-03	6,03E-03
	mean	1,07E-04	9,66E-04	7,18E-03	1,04E-02
	median	1,05E-04	8,02E-04	6,93E-03	1,03E-02
	std	8,91E-06	5,93E-04	1,78E-03	3,07E-03
ZDT4	best	8,24E-05	7,19E-01	5,01E-01	2,56E+00
	worst	7,47E-05	2,89E-02	6,93E-02	2,03E-01
	mean	7,74E-05	2,88E-01	1,94E-01	1,16E+00
	median	7,76E-05	2,34E-01	1,74E-01	9,38E-01
	std	1,67E-06	2,09E-01	9,52E-02	6,88E-01
ZDT6	best	6,61E-05	2,99E-03	4,21E-03	1,69E-01
	worst	6,13E-05	2,88E-04	6,68E-04	3,52E-04
	mean	6,34E-05	1,45E-03	2,05E-03	3,35E-02
	median	6,34E-05	1,87E-03	1,74E-03	3,35E-03
	std	1,12E-06	8,96E-04	1,00E-03	4,86E-02

Best results are marked in bold

The table I presents compelling evidence of the superior performance of our proposed method across a range of test problems, namely ZDT1, ZDT2, ZDT3, ZDT4, and ZDT6, as assessed by the IGD metric. Notably, our MOSTOA demonstrates rapid convergence to the Pareto set in scenarios where ZDT1 exhibits a convex front, and ZDT2 exhibits a non-convex front (See Fig.2 and Fig .3).

TABLE II. COMPARATIVE RESULTS OF HV METRIC

		MOSTOA	MOGWO	MSSA	MOEA/D
ZDT1	best	7,22E-01	7,09E-01	5,96E-01	7,11E-01
	worst	7,22E-01	6,48E-01	4,83E-01	5,90E-01
	mean	7,22E-01	6,93E-01	5,48E-01	6,85E-01
	median	7,22E-01	6,96E-01	5,50E-01	7,05E-01
	std	9,12E-05	1,09E-02	2,58E-02	3,73E-02
ZDT2	best	4,46E-01	4,23E-01	3,13E-01	4,38E-01
	worst	4,45E-01	9,09E-02	1,91E-01	0,00E+00
	mean	4,46E-01	3,39E-01	2,61E-01	7,52E-02
	median	4,46E-01	4,10E-01	2,65E-01	2,06E-02
	std	2,29E-04	1,37E-01	2,59E-02	1,04E-01
ZDT3	best	5,85E-01	6,00E-01	6,46E-01	5,80E-01
	worst	5,82E-01	5,48E-01	4,70E-01	1,30E-01
	mean	5,84E-01	5,81E-01	5,55E-01	4,22E-01
	median	5,84E-01	5,81E-01	5,51E-01	4,18E-01
	std	5,16E-04	1,07E-02	4,57E-02	8,87E-02
ZDT4	best	7,22E-01	9,09E-02	0,00E+00	0,00E+00
	worst	7,22E-01	0,00E+00	0,00E+00	0,00E+00
	mean	7,22E-01	5,86E-03	0,00E+00	0,00E+00
	median	7,22E-01	0,00E+00	0,00E+00	0,00E+00
	std	2,71E-05	2,27E-02	0,00E+00	0,00E+00
ZDT6	best	3,90E-01	3,85E-01	3,76E-01	3,83E-01
	worst	3,90E-01	3,26E-01	2,52E-01	0,00E+00
	mean	3,90E-01	3,59E-01	3,38E-01	2,08E-01
	median	3,90E-01	3,50E-01	3,50E-01	2,81E-01
	std	2,67E-05	1,95E-02	3,19E-02	1,76E-01

Best results are marked in bold

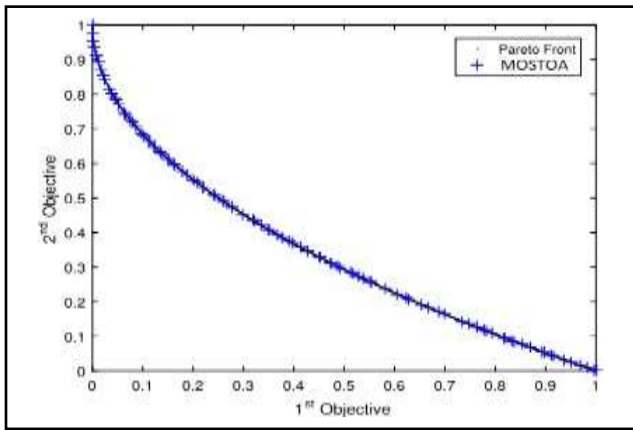


Fig.2. Pareto Front of the MOTOA algorithm for ZDT1

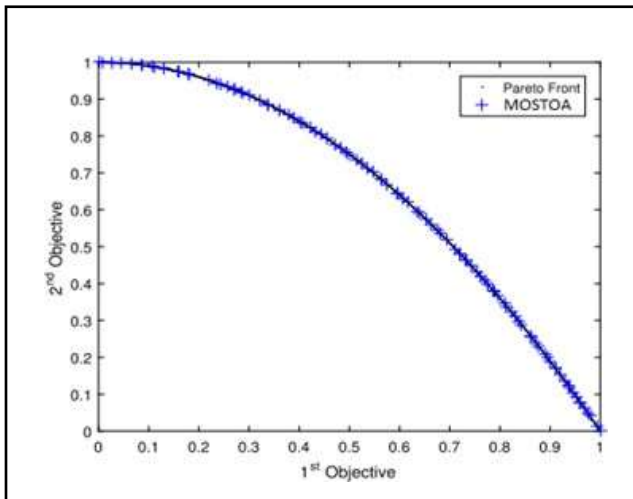


Fig.3. Pareto Front of the MOTOA algorithm for ZDT2

In accordance with the Hypervolume (HV) metric, the results presented in Table II provide strong statistical evidence of the superior performance of the MOSTOA algorithm when compared to other comparative methods across all test cases. Consequently, the proposed MOSTOA algorithm stands out as the top-performing approach among those examined. Visual representations in Figures 2-3 further illustrate that the non-dominated solutions generated by the MOSTOA algorithm exhibit a well-distributed alignment with the true Pareto front.

II. CONCLUSION

In our research, we have implemented three key strategies aimed at striking an optimal balance between exploration and exploitation attributes to extend the conventional Single-Objective Optimization Algorithm (STOA) into its Multi-Objective counterpart, known as MOSTOA. Firstly, we incorporate a population archive, serving as a dedicated repository for storing and retrieving non-dominated solutions acquired throughout the optimization process. Secondly, we employ a leader's solution to guide the primary population towards promising areas within the search space. Lastly, we utilize a density estimator, specifically based on the Grid Adaptive Strategy, to ensure comprehensive coverage of the search space. The comparative results, assessed using established metrics within the multi-objective optimization domain, have

conclusively demonstrated the effectiveness of our proposed MOSTOA. It exhibits superior convergence and diversity when compared to alternative algorithms.

REFERENCES

- [1] S. Mirjalili, « Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems », *Neural computing and applications*, vol. 27, p. 1053-1073, 2016.
- [2] D. Zouache, F. B. Abdelaziz, M. Lefkir, et N. E.-H. Chalabi, « Guided Moth–Flame optimiser for multi-objective optimization problems », *Ann Oper Res*, vol. 296, n° 1-2, p. 877-899, janv. 2021, doi: 10.1007/s10479-019-03407-8.
- [3] S. Mirjalili, P. Jangir, et S. Saremi, « Multi-objective ant lion optimizer: a multi-objective optimization algorithm for solving engineering problems », *Applied Intelligence*, vol. 46, p. 79-95, 2017.
- [4] S. Mirjalili, S. Saremi, S. M. Mirjalili, et L. dos S. Coelho, « Multi-objective grey wolf optimizer: a novel algorithm for multi-criterion optimization », *Expert systems with applications*, vol. 47, p. 106-119, 2016.
- [5] D. Zouache, Y. Ould Arby, F. Nouioua, et F. Ben Abdelaziz, « Multi-objective chicken swarm optimization: A novel algorithm for solving multi-objective optimization problems », *Computers & Industrial Engineering*, vol. 129, p. 377-391, mars 2019, doi: 10.1016/j.cie.2019.01.055.
- [6] F. Boumaza, Ab. Benyamina, D. Zouache, et L. Abualigah, « An improved Harris Hawks optimizer Based on Bi-Goal Evolution and Multi Leader Selection Strategy for Multi-Objective Optimization », *ingénierie des systèmes d'information*, vol. 28, n° 5, p. 1135-1150, oct. 2023, doi: 10.18280/isi.280503.
- [7] H. Wang *et al.*, « Multi-objective optimization of a hydrogen-fueled Wankel rotary engine based on machine learning and genetic algorithm », *Energy*, vol. 263, p. 125961, 2023.
- [8] J. He, N. Han, M. Xia, T. Sun, et H. Ghaebi, « Multi-objective optimization and exergoeconomic analysis of a multi-generation system based on biogas-steam reforming », *International Journal of Hydrogen Energy*, vol. 48, n° 55, p. 21161-21175, juin 2023, doi: 10.1016/j.ijhydene.2022.12.093.
- [9] P. Jangir, H. Buch, S. Mirjalili, et P. Manoharan, « MOMPA: Multi-objective marine predator algorithm for solving multi-objective optimization problems », *Evol. Intel.*, vol. 16, n° 1, p. 169-195, févr. 2023, doi: 10.1007/s12065-021-00649-z.
- [10] K. Tamashiro, E. Omine, N. Krishnan, A. Mikhaylov, A. M. Hemeida, et T. Senjyu, « Optimal components capacity based multi-objective optimization and optimal scheduling based MPC-optimization algorithm in smart apartment buildings », *Energy and Buildings*, vol. 278, p. 112616, janv. 2023, doi: 10.1016/j.enbuild.2022.112616.
- [11] G. Dhiman et A. Kaur, « STOA: A bio-inspired based optimization algorithm for industrial engineering problems », *Engineering Applications of Artificial Intelligence*, vol. 82, p. 148-174, juin 2019, doi: 10.1016/j.engappai.2019.03.021.
- [12] C. Cca, « Handling multiple objectives with particle swarm optimization », *IEEE Transactions on Evolutionary Computation*, vol. 8, n° 3, p. 256-279, 2004.
- [13] E. Zitzler, K. Deb, et L. Thiele, « Comparison of multiobjective evolutionary algorithms: Empirical results », *Evolutionary computation*, vol. 8, n° 2, p. 173-195, 2000.
- [14] Q. Zhang et H. Li, « MOEA/D: A multiobjective evolutionary algorithm based on decomposition », *IEEE Transactions on evolutionary computation*, vol. 11, n° 6, p. 712-731, 2007.
- [15] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, et S. M. Mirjalili, « Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems », *Advances in engineering software*, vol. 114, p. 163-191, 2017.
- [16] Y. Tian, X. Zhang, R. Cheng, et Y. Jin, « A multi-objective evolutionary algorithm based on an enhanced inverted generational distance metric », in *2016 IEEE Congress on Evolutionary Computation (CEC)*, juill. 2016, p. 5222-5229. doi: 10.1109/CEC.2016.7748352.
- [17] N. Riquelme, C. Von Lüken, et B. Baran, « Performance metrics in multi-objective optimization », in *2015 Latin American Computing Conference (CLEI)*, oct. 2015, p. 1-11. doi: 10.1109/CLEI.2015.7360024.

Orchestration of POIs ubiquitous contexts: a review of recommendation systems based on matrix factorization model

Mhamed HADJHENNI
EEDIS Laboratory
Djilali Liabes University
Sidi Bel Abbes, Algeria
md.hadjhenni@univ-chlef.dz

Nassim DENNOUNI
LIA laboratory
Hassiba Benbouali University
Chlef, Algeria
n.dennouni@univ-chlef.dz

Zohra SLAMA
EEDIS Laboratory
Djilali Liabes University
Sidi Bel Abbes, Algeria
zslama@yahoo.com

Abstract— Currently, to take full advantage of the capabilities of Artificial Intelligence (AI), Smart tourism must use Context-Aware Recommendation Systems (CARS) to orchestrate the evolving contexts of users with smartphones in order to improve their travel experiences. This type of orchestration allows points of interest (POIs) recommendations to be personalised according to the ubiquitous context of tourists during their visits. Recommending the next POIs to visit can be based on collaborative filtering techniques founded on memory or models such as matrix factorisation (MF). This paper explains the contribution of approaches that integrate contexts into models, such as MF, compared to collaborative filtering approaches without context. Consequently, this survey shows that collaborative filtering techniques using MF considerably alleviate the problems associated with the cold start of CARS and that the three types of orchestration of tourist contexts (pre-filtering, post-filtering or context modelling) improve their satisfaction.

Keywords— CARS, POI, orchestration, collaborative filtering, Matrix factorization, context pre-filtering, context post-filtering, context modelling.

I. INTRODUCTION

Artificial intelligence offers several potential applications in the tourism industry. AI is a promising way for tourists to discover critical data, enabling better mobility, improved decisions and an exceptional tourism experience [1]. The rise of AI is particularly relevant to almost every type of business. In the tourism sector, AI is employed for a wide range of purposes, including boosting levels of individualization, personalizing consumer suggestions and providing fast reaction times, even when staff are unavailable. The existence of artificial intelligence in the industrial environment has become so vital that it assists and communicates with customers, thereby increasing the quality of engagement. AI techniques include recommender systems (RSs), which enable relevant information to be proposed to users from a large mass of information.

The main types of RSs are: (1) content-based RSs, which use the characteristics of items to suggest articles, products, or content similar to those the user has already enjoyed or consulted. They analyze the intrinsic attributes of items, such as keywords, categories, or themes, to offer suggestions that are personalized and relevant to the user's preferences [2].

(2) Recommendation systems based on collaborative filtering identify a user's preferences based on the behavior or evaluations of other similar users using data collected on past

interactions. These systems recommend items by predicting the user's interests based on the tastes or actions of other individuals sharing similarities in their choices [3].

The combination of the two previous categories is known as (3) hybrid RSs, which makes it possible to benefit from the advantages of both categories [4].

RSs can be used in several domains, including tourism, for travel recommendations of the next POI to visit. RSs problems include the cold-start problem that occurs when a new user or item is added [5]; in this case, the system cannot generate a personalised recommendation due to a lack of information. RSs based on MF can solve this problem; in this article, we will study RS based on MF and see their contribution to the POI recommendation domain. Another problem with RSs is that users are not satisfied with the recommendations they receive, mainly because RSs need to consider the user's context. Integrating context into RSs can improve user satisfaction with the recommendations they receive.

Context is any information that can be used to characterize the situation of an entity. An entity can be a person, place or object considered relevant to the interaction between a user and an application, including the user and the application itself [6]. In specific domains, RSs use contextual information such as time, location, and companion to improve the recommendation process. This type of system is called a Context-Aware Recommendation System (CARS) [7] [8]. This contextual approach improves the accuracy of recommendations by adapting those to specific situations, offering more relevant and engaging user experiences. It also provides a better understanding of users' changing needs, improving satisfaction and loyalty by offering better-targeted recommendations tailored to the user's context.

In this article, we surveyed works related to CARSs based on collaborative filtering; particularly those based on model and matrix factorization techniques. We have discussed and compared work on FM and context integration to propose future directions for this type of technique in tourism.

This paper is structured as follows. Section II reviews the relationship between orchestration, recommender systems and context for the tourism domain. After that, in Section III, we present work related to the definition of context and approaches to incorporate context within collaborative filtering-based recommender systems using matrix factorization. Section IV will be devoted to comparing different collaborative filtering-based RSs (memory-based and matrix factorization-based). In section V, we discuss the use of these techniques for the POI recommendation domain. Finally, in section VI, we summarize the contributions of our paper in order to propose future perspectives for our work.

II. PRELIMINARY NOTIONS

In this section, we look at the principle of POI orchestration using RS/CARS and the personalization of POI recommendations based on the integration of user contexts.

A. RS and CARS

Most RS approaches focus on recommending the most relevant items to users and ignore contextual information, such as the time, place and companionship of other people (e.g. watching movies or dining out with friends or family). For this reason, traditional RSs deal with only two types of entities, users and items, and do not consider context in the recommendation process. However, it is essential to integrate contextual information into the recommendation process to suggest items to users in many applications, such as a travel recommendation application. For example, by using the context of the year's season, the travel RS can suggest skiing in winter and the beach in summer.

In the literature, three approaches can be used with CARS, depending on when the context was injected [9]:

Contextual pre-filtering consists of selecting a subset of data for the context in which it is found and limiting the recommendation process to this subset. The RS builds a model for each context [10]. For example, if a user wants to take a trip at the weekend, only POIs that are open at the weekend can be recommended, and only the scores of users who visited POIs over the weekend are used to predict the score. This pre-filtering can cause score prediction problems if the system does not have enough data (data sparsity problem).

In **contextual post-filtering**, the RS does not consider contextual data during the recommendation process. The results of the recommendation algorithms are reorganized according to the context to produce the list of items to be recommended [10]. For example, a system that recommends touristic places will use the user's geolocation (location context) and may decide to omit subsequent recommendations for places that are too far from the user's location.

Finally, the **contextual modelling** approach incorporates contextual information directly into the recommendation process to predict item scores. A recommendation is no

longer considered a function with parameters such as items and users but a function described with items, users and context variables [10].

B. Orchestration and RS

Orchestrating tourist visits involves planning and coordinating various aspects to offer tourists a harmonious experience. Using RS/CARS involves recommendation algorithms to personalize itineraries according to tourists' preferences.

RSs based on Collaborative Filtering (CF) can be divided into two main categories: memory-based and model-based collaborative filtering [3]. Memory-based collaborative filtering exploits similarity between users or items, while model-based collaborative filtering uses mathematical modelling or machine learning techniques (matrix factorization or neural networks) to predict user preferences.

C. Orchestration, context and CARS

Integrating the ubiquitous context in tourism relies on using technologies such as GPS to adjust real-time recommendations according to location, weather and other factors. Various approaches facilitate the orchestration of contexts for mobile tourists, using sensors, mobile applications, wearable devices and augmented reality. These systems collect location, time, weather conditions and tourist behavior data, offering personalized recommendations with CARS. Visit orchestration encompasses trip planning, post-arrival navigation, content personalization, and integrated customer service to resolve any problems encountered by tourists effectively. The aim is to deliver a seamless, personalized travel experience adapted to ubiquitous contexts, representing a key area of research and development to improve overall tourist satisfaction.

III. LITERATURE REVIEW

Matrix factorization techniques use different models, including Matrix factorization (MF) [11], Sparse Linear Method (SLIM) [12], Singular Value Decomposition (SVD) [13], Probabilistic Matrix Factorization (PMF) [14] and Non-negative Matrix Factorization (NMF) [15]. In This section, we will explain the operating principles of FM-based RSs AND SLIM RS.

A. RS FM-based functioning

The recommendation system based on matrix factorization decomposes a matrix of user-item ratings into two smaller matrices: a user matrix and an item matrix, which capture the underlying relationships between users and items.

In MF [11], users and items are represented by vectors \vec{p} and \vec{q} , respectively. The values of these vectors indicate the weights of K (e.g., $K = 5$) latent factors. Therefore, the ranking prediction is described by Equation 1.

$$\hat{r}_{it} = \vec{p}_u \cdot \vec{q}_t \quad (1)$$

The user and item biases rating (deviations) can be added to the prediction function, as shown in equation 2, where μ represents the overall average rating in the dataset, b_u and b_i represent the user and item rating biases, respectively [16].

$$\hat{r}_{it} = \mu + b_u + b_i + \vec{p}_u \cdot \vec{q}_t \quad (2)$$

Loss Function:

To train the model, we need to define a loss function that measures the difference between the predicted values and the actual values of the R matrix. The root mean square error (RMSE) is one of the most commonly used loss functions.

$$Loss = \sum (R_{ij} - \hat{R}_{ij})^2 \quad (3)$$

The matrix factorization can be done by using optimization techniques, such as stochastic gradient descent, to adjust the values of U and V to minimize the loss function. From that, we can learn the latent user and item characteristics. This training process is iterative, and the model converges to values of U and V that best minimize the difference between the predicted and actual values of the R matrix. Once the model has been trained, personalized recommendations can be generated by calculating predictions for items not rated by a user and ranking them according to these predictions.

Matrix factorization can also be applied to the tourism sector to recommend Points of Interest (POIs) and discover interesting places. Matrix Factorization for user-POI (Point of Interest) recommendations are based on the decomposition of a user-POI interaction matrix into two smaller matrices: a user matrix (U) and a POI matrix (V).

Here is a simplified example of matrix factorization with some detailed formulae involved. Suppose we have a user-POI interaction matrix (R) of dimensions M x N, where M represents the number of users and N is the number of POIs. Each element of the matrix R (R_{ij}) represents the interaction of user i with POI j , for example, a rating given by the user to a POI.

The aim is to decompose this matrix R into two matrices, U and V, so their product approximates the matrix R as closely as possible. This process models user preferences and POI characteristics as vectors.

User matrix (U)

The matrix U has dimensions M x K, where K is the number of latent features (factors). Each row of the matrix U represents a user, and each column represents a latent characteristic. The value U_{ik} represents the extent to which user i is influenced by latent feature k .

POI matrix (V)

The matrix V has dimensions K x N. Each row of the matrix V represents a latent feature, and each column represents a POI. The value V_{kj} represents the extent to which latent feature k is present in POI j .

Predicting interactions

The prediction of the interaction (score) of a user i with a POI j is obtained by the scalar product between the user vector U_i and the POI vector V_j and the addition of a global bias (or user and POI biases):

$$R_{ij} = U_i * V_j + G_bais + U_bais_i + P_bais_j \quad (4)$$

The biases **G_bais**, **U_bais_i** and **P_bais_j** (global bias, user bias and POI bias) are terms added to account for global effects and the particularities of each user and POI.

Using the following example, we will explain how a model-based recommendation system works using matrix factorization. Consider the following matrix of ratings given by users U1, ..., U5 to POI1,...POI5.

P(5*2)

	K1	K2
U1	0.16	0.25
U2	0.67	0.26
U3	0.72	0.07
U4	0.81	0.05
U5	0.96	0.83

Q (2*5)

	POI 1	POI 2	POI 3	POI 4	POI 5
K1	0.49	0.86	0.52	0.03	0.69
K2	0.58	0.15	0.82	0.50	0.46

\hat{R}

	POI 1	POI 2	POI 3	POI 4	POI 5
U1	0.22	0.17	0.28	0.13	0.22
U2	0.48	0.61	0.56	0.15	0.58
U3	0.39	0.63	0.43	0.06	0.53
U4	0.43	0.71	0.47	0.05	0.58
U5	0.95	0.95	1.18	0.44	1.04

N iterations After N=100000, we will have:

P (5*2)

	K1	K2
U1	1.93	0.75
U2	-1.30	1.75
U3	1.74	0.37
U4	0.79	1.27
U5	0.14	2.51

Q (2*5)

	POI 1	POI 2	POI 3	POI 4	POI 5
K1	0.46	2.33	2.05	-0.05	-0.11
K2	1.07	0.28	0.28	2.75	1.60

\hat{R}

U/P	POI 1	POI 2	POI 3	POI 4	POI 5
U1	1.69	4.72	4.17	1.98	0.99
U2	1.27	-2.55	-2.18	4.87	2.95
U3	1.19	4.16	3.67	0.93	0.40
U4	1.72	2.20	1.97	3.46	1.95
U5	2.75	1.05	0.99	6.91	4.01

B. RS SLIM-based functioning

SLIM is a model-based collaborative filtering method that aims to solve the Top-N recommendation problem. SLIM is based on a sparse matrix factorization approach, emphasizing

U/P	POI 1	POI2	POI3	POI4	POI5
U1	?	5	4	2	1
U2	1	?	?	5	3
U3	1	4	4	1	?
U4	?	?	2	?	2
U5	3	1	1	?	?

For this matrix R, we will define the number of latent factors with the value $K=2$ and carry out several iterations to obtain the matrix factorization of R.

First iteration:

We randomly initialize the matrix P (U, K) and Q(K, I). In our case, the matrices P(U,K) and Q(K,I) are defined as follows:

smoothing to obtain a sparse matrix that models the relationships between items. This approach produces high- quality recommendations while being efficient in terms of computational complexity. It is derived from the FM method, but instead of looking for $\hat{R} \approx P \cdot Q$, we look for $\hat{R} \approx RQ$ in the SLIM method. Q can represent the deviation matrix between items (POIs), the SLIM I variant, with K= number of items (POIs). Alternatively, Q can represent the deviation matrix between users, and this is the SLIM U variant, with K= number of users.

FM / SLIM comparison

- FM focuses on modelling the latent characteristics of users and items, while SLIM focuses on modelling the similarity between items or users.
- FM is best suited for personalized recommendations based on latent characteristics, whereas SLIM can recommend items similar to those the user has previously enjoyed.
- FM may be more appropriate for available explicit data, such as ratings, as it can predict accurate ratings.
- SLIM may be more appropriate when diversifying recommendations based on item similarities.
- It is also possible to integrate these two approaches to obtain a combination of the advantages of both methods.

C. CARS FM/SLIM-based functioning

Based on the variants for integrating context into a recommendation system, we can have the following CARS alternatives with FM and SLIM.

FM/SLIM with Pre-filtering

This type of system combines the advantages of model-based collaborative filtering with context-dependent customization. Adding context in pre-filtering means contextual data, such as location or time of day, are considered before the recommendation process. This approach adapts suggestions according to the specific context in which the user interacts with the system.

FM/SLIM with post-filtering

FM /SLIM RS incorporates context after the initial generation of recommendations. This approach adjusts suggestions according to the user's specific context, and the system becomes more flexible in adapting to the evolving preferences of users in different situations.

FM/SLIM with Context Modelling

Using context modelling with matrix factorization allows the system to integrate dynamically contextual data

into modelling relationships between users and items. This approach enables real-time adaptation to context variations, improving recommendations' relevance.

IV. COMPARISON OF FM-BASED APPROACHES

In the literature, in addition to the FM and SLIM methods, there are other variants of the FM method for RSs, such as:

(1) The **SVD** method decomposes a matrix R into three matrices, capturing the most significant components and truncating them to reduce dimensionality while preserving important information. SVD factors a matrix into its main components, representing users and items in a space of reduced dimensions. This method, used in recommender systems, results in the original matrix R decomposed as follows: $R = U \cdot \Sigma \cdot V^T$.

(2) The **PFA** method is often used when the data is not simply continuous but follows probabilistic distributions. It is used to model user-item interactions using probabilistic concepts. PMF aims to approximate a user-item matrix R by two feature matrices, P for users and Q for items, by maximizing the probability of observing existing evaluations.

(3) The **NMF** method is similar to matrix factorization but restricts the resulting matrices to positive values, allowing a more straightforward interpretation of the components. It can be seen as a variant of matrix factorization but with additional constraints on component values.

In order to be able to compare the above approaches with each other and with collaborative filtering techniques, we have set the following criteria:

a) Cold start

The cold-start problem in recommender systems occurs when new users or items without a significant history are introduced; it compromises the accuracy of similarity-based collaborative filtering.

b) Data sparsity

Data sparsity concerns the scores available in user-item matrices, where most users have only rated a fraction of the available items. It creates a crucial challenge: predicting preferences for unrated items. It complicates the task of recommendation algorithms, as they have to infer relationships between users and items based on incomplete data. Approaches to solving this problem include the use of matrix completion techniques and matrix factorization models to improve the quality of recommendations despite sparse data.

c) Context integration (temporal, geographical, seasonal)

Context integration, whether temporal, geographical, or seasonal, is crucial when evaluating recommender systems. Exploring and using different contexts contributes to an in- depth understanding of user preferences, ensuring personalized and relevant recommendations.

d) Prediction significance

The significance of predictions in a recommendation system evaluates the system's ability to provide understandable explanations behind the suggestions it generates. This criterion encompasses the model's explicability, the decision-making process's transparency, consideration of

the specific characteristics of an item or a user's preferences, and the possibility of the user having control over the recommendation process.

A straightforward interpretation of predictions increases user confidence and facilitates a more informed and satisfying interaction. By focusing on the comprehensibility of the system, the meaning of predictions promotes a transparent and adaptable user experience.

e) Recommendation personalization

Personalizing recommendations involves adapting suggestions according to each user's preferences. The aim is to create a unique user experience, improving the relevance of suggestions and user satisfaction.

f) The diversity of recommendations

Diversity refers to a RS ability to offer various suggestions rather than focusing on similar choices. The aim is to avoid redundancy by presenting users with diverse items or content, reflecting different aspects of their preferences, to ensure a richer user experience.

g) Scalability

A scalable recommender system must handle massive datasets efficiently without compromising the speed and quality of recommendations. It often involves parallel computing techniques, task distribution, or optimized data structures. Scalability is essential to guarantee a fluid and responsive user experience, even in environments with diverse users and items.

a) Implementation complexity

Higher implementation complexity can lead to higher costs and longer development times. However, moderate complexity can be justified if it enables higher performance or advanced functionality.

V. DISCUSSION

Techniques involving matrix decomposition are based on matrix factorization. From Table I, we can see the advantage of FM-based approaches over memory-based recommender systems. The integration of context (in pre-filtering or post-filtering) significantly contributes to

improving the recommendation quality and the integration of context in the modelling context, which can dynamically adapt the recommendation results to the user's current context.

Our comparison of memory-based and FM-based approaches to recommender systems (see Table I) shows that memory-based approaches do not solve the cold-start and sparsity problems. On the contrary, the FM-based variants of RS can solve these problems. The significance of the predictions or the explanation of the recommendation results obtained is guaranteed by memory-based approaches. On the other hand, this criterion is not guaranteed by approaches based on matrix factorization since they use latent factors whose meaning cannot be determined. Memory-based approaches do not guarantee scalability; on the contrary, this criterion is guaranteed by matrix factorization-based approaches but with the disadvantage of implementation complexity. All the approaches to recommender systems with collaborative filtering (memory-based and FM-based) allow for the integration of context (in pre-filtering, post-filtering, and context modelling). Results from [17][11][18] show that SVD-based prediction algorithms can effectively address the challenge of sparse data by exploiting hidden correlations. SVD performance can be improved by incorporating user and feature biases. This technique is best suited to predicting tourist check-ins. On the other hand, the FM and SLIM methods are well suited to feedback in rating or check-in format.

In the literature, there are limited dedicated FM-based RSs for POI recommendation. We have: (1) **GeoMF**: Joint Geographical Modelling and Matrix Factorization for Point-of-Interest Recommendation [19], (2) **GEOMFTD**: GEOMF with Time Dependencies POI Recommendation [20], (3) **ReGS**: Review geographical Social [21], (4) **FSS-FM**: Feature-Space Separated Factorization Model [22] and (5) **LGLMF**: Local Geographical based Logistic Matrix Factorisation Model for POI Recommendation [23].

Table I. Comparison of Collaborative Filtering Approaches (**MBCF**: Memory-Based Collaborative Filtering, **MF**: Matrix Factorization, **SLIM**: Sparse Linear Methods, **SVD**: Singular Value Decomposition, **PFA**: Probabilistic Factor Analysis and **NMF** Non-negative Matrix Factorization).

Criteria	MBCF ¹	FM ¹	SLIM ¹	SVD ⁴	PFA ⁶	NMF ⁶
Solving the cold start problem	No	Yes	Yes	Yes	Yes	Yes
Solving the data sparsity problem	No	Yes	Yes	Yes	Yes	Yes
Context integration (temporal, geographical, etc.)	Yes	Yes	Yes	Yes	Yes	Yes
Adaptability to real-time changes in context	No	Yes	Yes	Yes	Yes	Yes
Prediction significance	Yes	No	No	No	No	No
Personalised recommendations	Yes	Yes	Yes	Yes	Yes	Yes
Diversity of recommendations taken into account	Yes	Yes	Yes	Yes	Yes	Yes
Scalability	No	Yes	Yes	Yes	Yes	Yes
Implementation complexity	No	Yes	Yes	Yes	Yes	Yes
Context integration : Pre-filtering, Post-filtering	Yes	Yes	Yes	Yes	Yes	Yes
Context Modelling	Yes	Yes	Yes	Yes	Yes	Yes

These systems use datasets such as foursquare, Gowalla, or JiePang (a Chinese location-based social network) to make recommendations based on feedback such as tourist check-ins. In the literature, few approaches adopt the principle of matrix factorization, which considers several context variables such as the tourist's location and the time of the visit. For these context variables, it is necessary to consider the tourist's companion and the weather on the day of the visit. For these context variables, we propose to perform CARS based on matrix factorization with implicit and explicit data (ratings) from other datasets such as (yelp or JiePang). We also propose considering the ubiquitous context of tourists or POIs and orchestrating the integration of these context variables to select the variables that increase tourist satisfaction. Datasets with such context data are non-existent, so building such datasets for different regions of the world would be interesting. We also propose to develop CARS based on pre-filtering and post-filtering. The evaluations and comparisons of these CARS help us to determine the best method of context integration, and to evaluate which type of data (explicit or implicit) gives better results. This evaluation can take into account the tourists' satisfaction and the remedy for the cold start problem and the sparsity problem.

VI. CONCLUSION

In this article, we present the state of the art of collaborative filtering-based RSs using matrix factorisation as a model. We then explain how FM-based approaches can alleviate the cold-start problem associated with data sparsity. We also explored how context can be integrated into the CARS FM approach to improve tourist satisfaction. This type of integration can be associated with orchestrating the different context variables using pre-filtering, post-filtering, and context modelling to produce CARS for POI. As perspectives related to our work, we wish to make online evaluations of the orchestration of context variables in FM-based RS to estimate tourist satisfaction with metrics such as Click Through (CTR) or others. This evaluation requires the implementation of several variants of the FM approach integrating context and using several datasets such as YELP or JiePang.

ACKNOWLEDGMENT

This work has been supported by the PRFU project N° COOLO7UN020120220001 entitled "Formalisation of point-of-interest recommendation systems and performance evaluation for mobile users," Department of Computer Science, Hassiba Ben Bouali University, Chlef, Algeria.

REFERENCES

- [1] Tussyadiah, I., & Miller, G. (2019). Perceived impacts of artificial intelligence and responses to positive behaviour change intervention. In *Information and Communication Technologies in Tourism 2019: Proceedings of the International Conference in Nicosia, Cyprus, 2019* (pp. 359-370). Springer International Publishing.
- [2] Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web: methods and strategies of web personalization* (pp. 325-341). Berlin, Heidelberg: Springer Berlin Heidelberg.
- [3] Candillier, L., Jack, K., Fessant, F., & Meyer, F. (2009). State-of-the-art recommender systems. In *Collaborative and Social Information Retrieval and Access: Techniques for Improved User Modelling* (pp. 1-22). IGI Global.
- [4] Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6), 734-749.
- [5] Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002, August). Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 253-260).
- [6] Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., & Steggles, P. (1999). Towards a better understanding of context and context-awareness. In *Handheld and Ubiquitous Computing: First International Symposium, HUC'99 Karlsruhe, Germany, September 27-29, 1999 Proceedings 1* (pp. 304-307). Springer Berlin Heidelberg.
- [7] Baltrunas, L. (2008, October). Exploiting contextual information in recommender systems. In *Proceedings of the 2008 ACM conference on Recommender systems* (pp. 295-298).
- [8] Karatzoglou, A., Amatriain, X., Baltrunas, L., & Oliver, N. (2010, September). Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 79-86).
- [9] Panniello, U., Tuzhilin, A., Gorgoglione, M., Palmisano, C., & Pedone, A. (2009, October). Experimental comparison of pre-vs. post-filtering approaches in context-aware recommender systems. In *Proceedings of the third ACM conference on Recommender systems* (pp. 265-268).
- [10] Adomavicius, G., Sankaranarayanan, R., Sen, S., & Tuzhilin, A. (2005). Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information systems (TOIS)*, 23(1), 103-145.
- [11] Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30-37.
- [12] Ning, X., & Karypis, G. (2011, December). Slim: Sparse linear methods for top-n recommender systems. In *2011 IEEE 11th international conference on data mining* (pp. 497-506). IEEE.
- [13] Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. T. (2000). Application of dimensionality reduction in recommender system-a case study.
- [14] Salakhutdinov, R., & Mnih, A. (2008, July). Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In *Proceedings of the 25th international conference on Machine learning* (pp. 880-887).
- [15] Cai, D., He, X., Han, J., & Huang, T. S. (2010). Graph regularized nonnegative matrix factorization for data representation. *IEEE transactions on pattern analysis and machine intelligence*, 33(8), 1548-1560.
- [16] Zheng, Y., Mobasher, B., & Burke, R. D. (2015, July). Incorporating Context Correlation into Context-aware Matrix Factorization. In *CPCR+ ITWP@ IJCAI* (p. 7).
- [17] Takács, G., Pilászy, I., Németh, B., & Tikk, D. (2008, August). Investigation of various matrix factorization methods for large recommender systems. In *Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition* (pp. 1-8).
- [18] Rendle, S., Gantner, Z., Freudenthaler, C., & Schmidt-Thieme, L. (2011, July). Fast context-aware recommendations with factorization machines. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (pp. 635-644).
- [19] Lian, D., Zhao, C., Xie, X., Sun, G., Chen, E., & Rui, Y. (2014, August). GeoMF: joint geographical modelling and matrix factorization for point-of-interest recommendation. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 831-840).
- [20] Griesner, J. B., Abdesslem, T., & Naacke, H. (2015, September). POI recommendation: Towards fused matrix factorization with geographical and temporal influences. In *Proceedings of the 9th ACM Conference on Recommender Systems* (pp. 301-304).
- [21] Xing, S., Liu, F., Zhao, X., & Li, T. (2018). Points-of-interest recommendation based on convolution matrix factorization. *Applied intelligence*, 48, 2458-2469.
- [22] Cai, L., Xu, J., Liu, J., & Pei, T. (2018). Integrating spatial and temporal contexts into a factorization model for POI recommendation. *International Journal of Geographical Information Science*, 32(3), 524-546.
- [23] Rahmani, H. A., Aliannejadi, M., Ahmadian, S., Baratchi, M., Afsharchi, M., & Crestani, F. (2020). LGLMF: local geographical based logistic matrix factorization model for POI recommendation. In *Information Retrieval Technology: 15th Asia Information Retrieval Societies Conference, Hong Kong, China, Proceedings 15* (pp. 66-78).

Automatic Opinion Extraction from Football-Related Social Media: A Gazetteer and Rule-Based Approach

Atmane Hadji

Department of Computer Science, University Center A.
Boussouf Mila, 43000 Mila, Algeria.
a.hadji@centre-univ-mila.dz

Mohmed-Khireddine Kholadi

Department of Mathematics and Computer Science,
HAMMA lakhdar, El oued University, El-Oued, Algeria.
MISC Laboratory of Abdelhamid Mehri university of
Constantine 2, Algeria.
kholadi@univ-eloued.dz

Abstract—Sentiment analysis on social networks has become a highly active area of research in recent years. With the explosion of social media and the massive amount of user-generated data, it has become crucial to understand the opinions and sentiments expressed online. Sentiment analysis is used to categorize expressed feelings in various ways, such as negative, positive, or neutral. The aim of this work is to enhance techniques for researching and extracting opinions. The main idea is to identify opinions within a set of documents or texts available online for exploitation by other systems. In this study, we present an approach based on an opinion detection system on social networks (Facebook) regarding the UEFA Champions League. The implementation of this solution was carried out using the GATE platform (General Architecture for Text Engineering). This work thus contributes to the field of sentiment and opinion analysis in social networks by employing Gazetteers and leveraging the JAPE rules (Java Annotation Patterns Engine).

Keywords— *Sentiment Analysis ; Social Networks; Gazetteer ;JAPE, UEFA Champions League.*

I. INTRODUCTION

Thanks to the Internet, there is the possibility of discovering the opinions and feelings of a large number of people, which can be very useful in making informed decisions and forming an opinion on a given subject. In this work, we focus on analyzing opinions expressed regarding the UEFA Champions League. We study the information and comments posted on social networks (Facebook) in order to extract information about fan preferences, team performances, influential players, and emerging trends in the competition. Our goal is to understand the opinions of supporters and analyze the sentiments and preferences expressed towards the UEFA Champions League. Analysts and sports media use these opinions to assess the strengths and weaknesses of teams, predict match results, and provide comments and analyses to fans. People's opinions provide insight into the perception of the game, team strategies, and the impact of individual performances. In this context, we proposed an opinion analysis approach based on a Gazetteer. The aim of this work is to present some techniques to enhance the automatic detection of opinions and sentiments from comments on social networks.

The goal of using a Gazetteer in opinion extraction is to strengthen the performance of the given results and improve their quality.

II. BACKGROUND AND RELATED WORKS

The extraction of opinions based on rules involves using predefined patterns or guidelines to identify and extract subjective information, sentiments, or attitudes expressed in text data. This approach is often used in natural language processing (NLP) and sentiment analysis tasks. Here's a background overview of opinion extraction based on rules:

- **Rule-Based NLP:** Rule-based NLP relies on a set of predefined linguistic patterns, grammatical rules, or heuristics to process and analyze text data. These rules are designed by linguists or NLP experts and are used to capture specific linguistic structures, sentiments, or entities within the text.
- **Subjectivity and Sentiment Analysis:** Opinion extraction is a subtask of sentiment analysis, which aims to determine the sentiment or emotion expressed in a piece of text. Subjectivity refers to the extent to which a statement is influenced by personal feelings, opinions, or beliefs.
- **Key Components: Linguistic Patterns:** Rules are typically defined based on linguistic patterns, syntactic structures, or semantic cues. These patterns may include specific keywords, parts of speech, or syntactic relationships that are indicative of opinions or sentiments.
- **Gazetteers:** A gazetteer is a list of words or phrases associated with specific categories or entities. It can be used in conjunction with rules to identify named entities or specific terms related to opinions.
- **Regular Expressions:** Regular expressions are powerful tools for defining complex patterns in text. They can be employed to capture various linguistic features that indicate opinions.

Our proposed approach is a hybrid approach merging between the two previous approaches cited. In the following, we will mention some related works :

The work proposes by [1] a rule-based algorithm for sentiment analysis, incorporating POS tagging during data preprocessing using Python. The user interface utilizes HTML with highlighted phrases, emojis, and GIFs to convey emotions. The analysis indicates that the proposed algorithm excels in predicting sentence sentiments and does so in a remarkably short time frame.

The authors [2] suggest a novel rule-based approach for extracting aspects from product reviews. This method relies on intuitive knowledge and the analysis of grammatical structures in sentences to identify both explicit and implicit aspects. The evaluation of the system on two popular datasets demonstrates that it achieves a higher detection accuracy compared to state-of-the-art aspect extraction techniques for both datasets.

The study realized by [3] investigates the application of aspect-based sentiment analysis in the legal domain to extract valuable information from legal opinion text. The authors introduce a rule-based approach for conducting aspect-based sentiment analysis, aiming to determine the sentiment expressed in a sentence regarding each legal party involved in a court case, considering these parties as the aspects of analysis.

III. APPROACH PROPOSED

The following architecture (Figure 1) depicts the detailed design of our opinion analysis system. The proposed system consists of several stages:

- **Data Collection** : We get information from social network (Facebook) online. We processed comments related to fan opinions semi-automatically.
- **Pretreatment** : In this step, we identified the comments related to the Champions League, then processed them in the next step.
- **Processing and Text Analysis**

We carried out the first phase of this work which is the research and collection of opinions on a sample of 70 comments from Facebook pages posted on the ChampionsLeague (chosen at random). after this phase we analyzed the text within the framework of automatic processing of a natural language by GATE [4] according to the order of steps: (Tokenization, Sentence Splitter, Part Of Speech Tagger, ...)

.However, the gazetteers were created for the indexing and extraction of opinions. The steps of text processing is as follows:

A. Tokenization

The Tokenizer divides text into simple words such as numbers, punctuation marks and many different types. For example, we have different words in Majestic and Minuscule, and among certain types of punctuation, etc. There is a

"Token" annotation in the box, it should not be changed for different applications or text types.

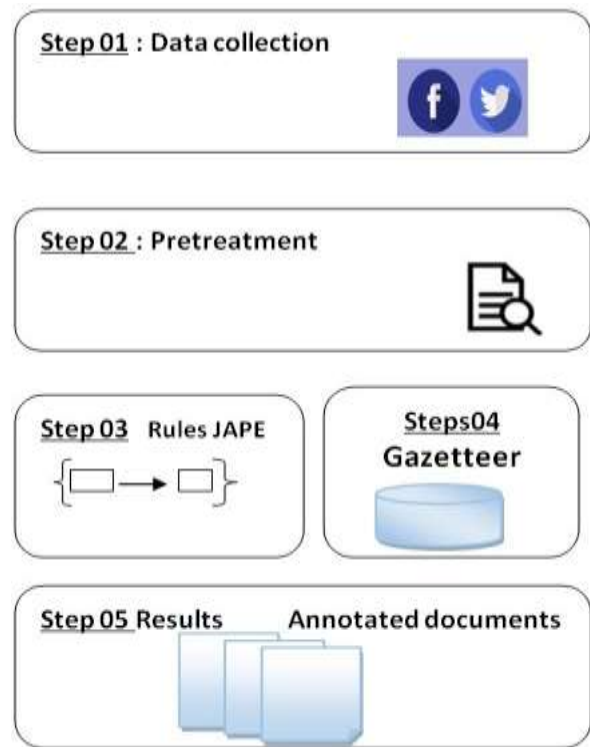


Fig. 1. General architecture of the proposed system

B. Sentence Splitter

The sentence splitter is a cascade of finite-state transducers that segments text into sentences. This module is required for the tagger. The separator uses a list of gazetteer abbreviations to help distinguish phrase marking points from other types.

C. Part Of Speech Tagger

The tagger used is a modified version of the Brill tag, which assigns a part-of-speech tag to each word or symbol in the text. It is based on a lexicon and a set of default rules, which were learned from a large corpus from the Wall Street Journal. These elements can be adjusted manually if necessary.

Two additional lexicons are available: one for texts entirely in uppercase and the other for texts entirely in lowercase. To use them, simply load the appropriate lexicon, replacing the default one. In any case, the default rule set should always be used.

D. Semantic Tagger

The semantic marker of ANNIE (A Nearly-New Information Extraction System) is based on the JAPE (Java Annotation Patterns Engine) language. It uses rules that act on annotations, features and values assigned in previous sentences. It also includes rules that act on annotations, features and values that need to be assigned manually.

E. Gazetteer Creation

The Gazetteer plays a vital role in identifying named entities in the text based on predefined lists. It consists of different lists, such as city names, organization names, days of the week, etc. These lists are plain text files, where each entry is on a separate line. The Gazetteer contains not only specific entities, but also relevant keywords such as company abbreviations (e.g. "Ltd."), titles (e.g. "Dr."), etc. The lists are then compiled into finite state machine structures, which map text tokens to identified entities .

TABLE I. GAZETEEER LIST EXAMPLE

	Positive Opinion	Negative Opinion	Neutral Opinion
Match Opinion	Exhilarating, highly entertaining, memorable match, Historic remnants Remnants, Historic match	Frustrating Underwhelming Subpar, Dismal, Mediocre Sloppy	neutral fan, evenly matched Balanced match, Thrilling game well fought , neutral observers
Team Opinion	Effective, Dominant Skilled,Cohesive, Effective Competitive	execution fell, vulnerable,underwhelming Weak,Inept,Inefficient	Evident,steady, Competent Capable, Decent, Adequate

In this study, we take two opinions, match opinion and team opinion, where supporters' opinions which are divided into a negative opinion, a positive opinion, and a neutral opinion, and the words which express which are among the lists from the gazetteers.



Fig. 2. Gazetteer list in GATE of Opinion Match Positive and Negative

F. JAPE Rules :

A JAPE grammar [5] consists of a set of phases, each of which consists of a set of pattern/action rules. The phases execute sequentially and constitute a cascade of finite-state transducers on the annotations. The left side (LHS) of the rules consists of a description of the annotation pattern. The right

side (RHS) consists of annotation manipulation instructions. Corresponding annotations on a ruler's LHS can be referenced on the RHS by means of labels attached to the pattern elements. Below is an example of a JAPE rule.



Fig. 3. Example of JAPE Rules

IV. RESULTS AND EVALUATION

After running the corpus with the use of JAPE and Gazetteer rules (figure 4), the system is now able to detect the entities named "Opinion Match Positive", "Opinion Match Negative" and "Opinion Match Neutral" corresponding to opinions on a Champions League match.

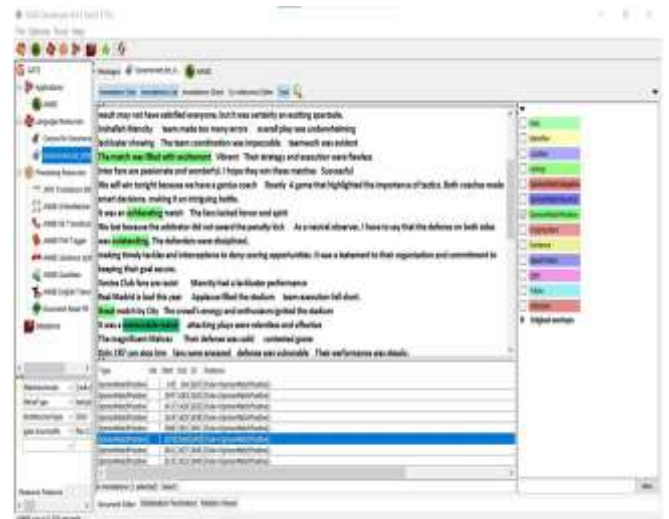


Fig. 4. Gazetteer list of Opinion Match Positive and Negative

We processed a corpus containing 70 comments of 1000 words from social media (Facebook) using JAPE rules and using the Gazetteer we calculated the percentage: Recall,

precision and F-measure are widely used measures in NLP assessments;

- Precision is the percentage of correct results among the results obtained .

$$\text{Precision} = \frac{\text{Number of NE correctly recognized}}{\text{Number of NE recognized}}$$

- Recall is the percentage of correct results among the results that must be found .

We present formulas for evaluation such as precision, recall which are widely used measures in NLP evaluations

$$\text{Recall} = \frac{\text{Number of correctly recognized NE}}{\text{Number of corpus NE}}$$

- The F-measure is the combination of precision and recall and their weighting. The formula for the F-measure is as follows:

$$F - \text{measure} = \frac{2(\text{precision} \times \text{recall})}{\text{precision} + \text{recall}}$$

V. ANALYSIS AND DISCUSSION

A. Information Analysis

According to the (Table II) it shows that opinions are extracted from two main categories: “Match” and “Team”. Each of these categories is then subdivided into positive, negative and neutral opinions.

When it comes to opinions on the matches, there is a relatively balanced distribution between positive (20), negative (16) and neutral (17) opinions. This suggests that match comments are quite varied in terms of the sentiment expressed.

Regarding opinions on teams, we notice that positive opinions (29) are more frequent than negative (16) and neutral (9). This could indicate a more positive trend towards teams.

TABLE II. OPINION EXTRACTION RESULTS

	Named Entity		Precision	Recal l	F- measure
Positive Opinion Match	20	38	0.71	0.80	0.75
Negative Opinion Match	16	30	0.65	0.84	0.73
Neutral Opinion Match	17	32	0.67	0.84	0.77
Psitive Opinion Team	29	54	0.80	0.85	0.82
Negative Opinion Team	16	29	0.62	0.69	0.65
Neutral Opinion Team	9	17	0.70	0.76	0.73

The results obtained in this study are highly satisfactory, as demonstrated by the good Precision and excellent Recall rates (see to Table II and Figure 5).

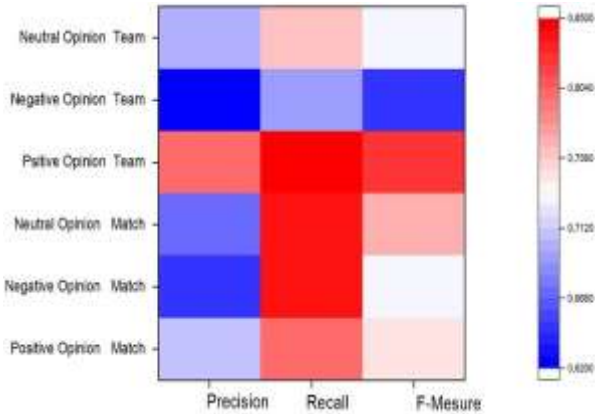


Fig. 5. Our Approach Evaluation

B. Overall Performance:

Overall, the performance of named entity opinion extraction appears to be quite robust, with acceptable precision,recall, and F-measure values for most categories. These results indicate that the system is capable of correctly identifying opinions in the text.

C. Category-wise Analysis:

The results are divided into three main categories: "Positive Opinion", "Negative Opinion", and "Neutral Opinion", each further subdivided into "Match" and "Team". Let's examine them in more detail:

- **Positive Opinion Match:** This category exhibits high precision (0.71), indicating that the majority of mentions extracted as positive opinions are indeed accurate. The recall is also strong (0.8), demonstrating that most positive opinions have been captured. The balanced F-measure (0.75) suggests a good trade-off between precision and recall.
- **Negative Opinion Match:** Performance is also good for this category, with a precision of 0.65 and a recall of 0.84. The F-measure of 0.73 indicates a good balance between precision and recall.
- **Neutral Opinion Match:** This category shows similar performance to the previous categories, with a precision of 0.67, a recall of 0.84, and an F-measure of 0.77.
- **Positive Opinion Team:** This category displays high precision (0.8), solid recall (0.85), and a balanced F-measure (0.82), indicating very good performance in extracting positive opinions about teams.

- **Negative Opinion Team:** While performance is lower than that of the "Positive Opinion Team" category, it remains acceptable, with a precision of 0.62, a recall of 0.69, and an F-measure of 0.65.
- **Neutral Opinion Team:** This category exhibits relatively low precision (0.05), indicating that the majority of mentions extracted as neutral opinions are incorrect. However, the recall is high (0.76), showing that most neutral opinions have been identified. The F-measure of 0.73 reflects a balance between precision and recall.

D. Comparison of Categories:

Overall, the "Positive Opinion" and "Negative Opinion" categories have similar performances, with balanced precision and recall. The "Neutral Opinion" categories have lower precision but higher recalls. The results demonstrate that extracting positive and negative opinions is more successful than extracting neutral opinions, especially for teams. It might be useful to explore specific strategies to improve the precision of neutral opinions.

In summary, the results show that the system is capable of extracting opinions with generally good performance. However, there are variations across categories, suggesting potential for continuous improvement, particularly in extracting neutral opinions. These results provide a solid foundation for opinion analysis in the context of matches and teams. They can be valuable for sports analysts, media, and sports organizations to evaluate public reactions and trends.conclusion

After In conclusion, our study is part of the context of the analysis of sentiments and opinions in social networks, with a particular emphasis on the Football Champions League. Through the use of advanced techniques such as Gazetteers and JAPE Rules, we have developed an innovative approach to extracting and understanding opinions expressed online. Our work demonstrated the effectiveness of this approach, allowing more precise detection and better quality of the extracted opinions.

Leveraging the GATE platform, we have implemented an opinion detection system on Facebook, providing a valuable tool for analysts, sports media and football fans. This methodology can be extended to other domains and social networks, thus opening new perspectives for sentiment analysis research .

VI. References

- [1] U. Vudatha ,K. Prema “ Aspect Based Sentiment Analysis Using Rule Based Approach ” First International Conference on Advances in

- Computing and Future Communication Technologies (ICACFCT) , India, 2021.
- [2] P. Soujanya , C. Erik , K. Lun-Wei , G. Chen , G. Alexander “A Rule-Based Approach to Aspect Extraction from Product Reviews” Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP), pages 28–37, Dublin, Ireland, 2014.
- [3] R. Isanka , R.M. Chanika , K. Dilini ,S. Nisansa S ,R. Gathika , A. Shehan Pererak , “ Rule-Based Approach for Party-Based Sentiment Analysis in Legal Opinion Texts ”<https://arxiv.org/pdf/2011.05675.pdf>
- [4] GATE: <http://gate.ac.uk/>, last accessed 2022/12/20.
- [5] D. Thakker, P. Photosand ,R. Tanaka, Taha Osman, P. Lakin, (2009.) “GATE JAPE Grammar Tutorial Version 1.0. UK: PA Photos”,
- [6] Gate.ac.uk , " Chapter 8 JAPE : Regular Expressions over Annotations,"[Online],Available: <https://gate.ac.uk/sale/tao/splitch8.html>.
- [7] P. Malo., A. Sinha , P. Takala , P. Korhonen ,J. Wallenius , “Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts”, JASIST, 2014.
- [8] LI F., “ The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach”. Journal of Accounting Research, vol. 48, no 5, p. 1049-1102, 2010.
- [9] T. Loughran , B. McDonald , “The Use of Word Lists in Textual Analysis”, Journal of Behavioral Finance, vol. 16, no 1, p. 1-11,2015.
- [10] X. Mao , S. Chang , J Shi ,F. Li , R. Shi , “Sentiment-Aware Word Embedding for Emotion Classification”, Applied Sciences, vol. 9, p. 14,2019.
- [11] P. Kumar ,J. Prabhu , “Role of sentiment classification in sentiment analysis: a survey”, AISC,vol. 1168, p. 196-209, 2018.

Optimizing Arabic spam filtering through unsupervised and ensemble learning approaches

Marouane KIHAL

Laboratory of Medical Informatics (LIMED), Faculty of
Exact Sciences, University of Bejaia, 06000
Bejaia, Algeria
marouane.kihal@univ-bejaia.dz

Lamia HAMZA

Laboratory of Medical Informatics (LIMED), Faculty of
Exact Sciences, University of Bejaia, 06000
Bejaia, Algeria
lamia.hamza@univ-bejaia.dz

Abstract—

This paper presents a new Ensemble Learning approach for filtering Arabic spam. The proposed approach utilizes four unsupervised Machine Learning algorithms, including One Class Support Vector Machine (OCSVM), the Histogram-Based Outlier Score (HBOS), Local Outlier Factor (LOF) and Isolation Forest (IF), to construct a robust spam filter. The performance of our proposed approach is evaluated on a textual Arabic dataset. The experimental results show that our model achieves more than 84% of accuracy outperforming other Machine Learning algorithms. The use of Ensemble Learning and multiple unsupervised algorithms in our approach proves to be a promising solution for effective Arabic spam filtering.

Keywords—*Spam filtering; Arabic Spam; Unsupervised Learning; Ensemble Learning; Machine Learning*

I. INTRODUCTION

Spam is unsolicited messaging or content that is frequently distributed to a large number of recipients. It may be employed for maliciousness, advertising or commercial purposes. To avoid this problem, researchers and companies are proposing filtering software and spam detectors based on Artificial Intelligence (AI). Until today, computer users and especially social network users still suffer from spam. Consumer scams, threatened privacy, security breaches and lost productivity for businesses are the most relevant dangers of spams. To protect against these dangers, researchers have proposed several filtering techniques.

Nowadays with the evolution of Machine Learning algorithms, several filters have been proposed, such as K-Nearest Neighbors (KNN) [1], Support Vector Machine (SVM) [2], ETC. Most of the existing work focuses on supervised learning, but recently there are some new works based on unsupervised learning, which attempts to teach a Machine Learning algorithm information that is neither classified nor labeled, and to allow this algorithm to react to this information independently.

However, the problem of spam remains, especially spam written in languages other than English, such as Arabic, which really causes a problem because of the complexity of this language and the very few resources available to train models

in Arabic. That is why we have chosen to address Arabic spam.

In this paper, we propose a new spam filter based on unsupervised learning and ensemble learning. Our architecture is based on a late fusion of four Machine Learning algorithms namely One Class Support Vector Machine (OCSVM), the Histogram-Based Outlier Score (HBOS), Local Outlier Factor (LOF) and Isolation Forest (IF).

To summarize, the primary contributions of this study are:

- Proposal of an effective Arabic anti-spam filter based on four unsupervised learning.
- Utilizing ensemble learning fusion for result optimization.
- We compared our model with supervised and unsupervised learning algorithms.
- Our model surpassed the Machine Learning models, achieving 84,78% accuracy, 99,99% precision, 84,78% recall, and 91,76% F1-score.

The remainder of this paper is structured as follows:

Section 2 entails an overview of related works. Section 3 presents our proposed approach. Section 4 discusses the experimentation results. Section 5 gives the conclusion and perspectives for future research.

II. RELATED WORKS

In this section we present related works in two parts: unsupervised learning spam filter and arabic spam filter.

A. Unsupervised Learning Spam filter

In [3], the authors presented a text-mining framework for extracting textual signatures from unlabeled documents. The system contains preprocessing stages like term reduction and matrix formation in addition to Latent Semantic Analysis, which boosts semantics. As a supplement to textual signatures in spam detection, it also covers the extraction of HTML and URL signatures. The article assesses the detection accuracy of the various types of signatures as well as the accuracy of the created campaign clusters.

In [4], the authors used M-DBSCAN's unsupervised learning capabilities to identify spam and legitimate emails. Modified Density-Based Spatial Clustering of Applications with Noise (M-DBSCAN) was used in the method's implementation. The online test uses the N-representative points that were retrieved from each cluster. These points are created during the training phase of the distance-based spam email detection process.

In [5], the authors used the digest procedure to cluster emails, and then they used the DBSCAN clustering method. Comparing the results to the regular DBSCAN, the accuracy was found to have improved. This study demonstrates the efficacy of DBSCAN and clustering algorithms for email spam detection.

B. Arabic Spam filter

Najadat et al.[6] proposed a keyword-based technique for identifying fake reviews in Arabic. To extract keywords from Arabic text, the weight of a term, the Term Frequency-Inverse Document Frequency (TF-IDF) matrix, and filter approaches have been utilized. Then, Supervised Machine Learning algorithms including Decision Tree, kNN, SVM, and Nave Bayes have been used for classification.

The Gradient Boosting algorithm and the most efficient hyperparameter selection were used by [7] to provide a novel method for improving Arabic spam filtering effectiveness.

In [8], the authors provided a method that does not rely on hand-crafted features that are frequently time-consuming to get and are designed for a specific type of low-quality information, but instead automatically extracts textual features using deep learning techniques. They additionally set up a rapid system that makes use of a selection of arabic textual features to spot spammy Twitter accounts in real time.

III. OUR METHOD

Our approach is based on two concepts: unsupervised learning to get the primary decision on whether the Arabic content is spam or not, by using four unsupervised algorithms. Moreover, the second notion of our approach is ensemble learning to make the final decision by fusing the results of the four previous algorithms.

A. Unsupervised Learning

In the context of spam filtering, unsupervised learning is a technique for data analysis where a model finds key patterns in a dataset without the use of labels. Instead of depending just on labels, the algorithms used in this method seek to cluster related data or reveal hidden patterns in order to determine if content is Spam or Ham.

a) One class SVM:

One Class Support Vector Machines (OCSVM) [9] is a kind of outlier detection method. This unsupervised learning method used for developing the capacity to distinguish test samples from one class from test samples from other classes.

B. The Histogram-based Outlier Score:

The Histogram Based Outlier Score (HBOS) [10] determines the outlier score for that variable using the histogram of each variable. The multivariate outlier score for an observation can be calculated by summing the outlier scores of all the variables. The HBOS is an effective

unsupervised method to identify anomalies since histograms are simple to create.

a) Local Outlier Factor:

The Local Outlier Factor (LOF) [11] algorithm calculates the local density deviation of a particular data point with respect to its neighbors. It is an unsupervised anomaly identification technique. The samples that have a significantly lower density than their neighbors are considered as outliers.

b) Isolation Forest:

The binary tree-based Isolation Forest (IF) algorithm[12] works well with massive quantities of data since it has a linear time complexity and needs little memory. Fundamentally, the system does a quick density estimation approximation, recognizing data points with noticeably low density estimates as anomalies.

C. Ensemble Learning

Ensemble learning improve results by combining various models. This technique permits the production of higher prediction performance as compared to employing a single model.

A number of fusion techniques exist, including decision-level fusion, which combines feature classification decisions, model-level fusion, which is based on the relationships between features in various algorithms, and rule-level fusion, which employs techniques like majority voting or weighted fusion [13]. "Fig.1" shows the general architecture of our proposed system.

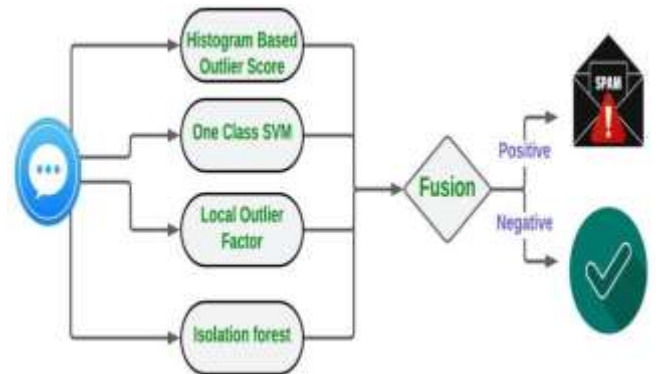


Fig. 1. Our system architecture.

We propose to make the classification with the four algorithms already mentioned (IF, HBOS, LOF and One class SVM). Therefore, we will arrive at four decisions on the content: 1 for spam, 0 for not spam. The final decision is determined by averaging the decisions made by unsupervised Machine Learning classifiers. As shown in equation 1:

$$D(f) = \text{Round}((D(\text{OCSVM}) + D(\text{HBOS}) + D(\text{LOF}) + D(\text{IF})) / 4). \quad (1)$$

Where $D(f)$ is the final decision, and $D(\text{OCSVM})$, $D(\text{HBOS})$, $D(\text{LOF})$, and $D(\text{IF})$ are the decision of the four unsupervised algorithms.

IV. EXPERIMENTATIONS

A. Dataset

Our model was trained and tested using health-related spam campaigns[14]. Between May 2018 and November 2020, this unbalanced dataset was collected from prominent Arabic hashtags using Twitter's standard search application programming interface (API). The dataset composed of 3000 tweets; 2500 training and 500 testing dataset.

B. Results

For the results we used four evaluation metrics most commonly used in spam detection [15], namely Accuracy, Precision, Recall and F1-score.

Table 1 shows the experimental results of our model in comparison with seven known Machine Learning algorithms, four unsupervised above-mentioned and three supervised ML such as: Naïve Bayes, Logistic Regression, and Support Vector Machine.

TABLE I. HEALTH-RELATED SPAM CAMPAIGNS RESULTS

Classifier	Justesse	Precesion	Recall	F1-Score
Isolation Forest	79,35%	93,59%	83,91%	88,48%
HBOS	82,61%	97,44%	84,44%	90,48%
Local Outlier Factor	83,70%	98,72%	84,62%	91,12%
One class SVM	80,43%	94,87%	84,09%	89,16%
Naïve Bayes	68,85%	33,71%	54,12%	41,54%
Logisitic Regression	79,92%	57,14%	7,33	13,00%
Support Vector Machine	80,86%	81,81%	8,25%	15,00%
Our method	84,78%	99,99%	84,78%	91,76%

From the results displayed in the table 1, we conclude that our model has overcome the Machine Learning algorithms under four evaluation metrics with accuracy equal to 84.78%, a precision 99.99%, Recall 84.78% and F1-score equally to 91.76%.

V. CONCLUSION

In this paper, we have proposed a new approach based on Unsupervised and Ensemble Learning to effectively filter Arabic textual spam, using four powerful unsupervised Machine Learning algorithms, namely One Class Support Vector Machine (OCSVM), The Histogram-Based Outlier Score (HBOS), Local Outlier Factor (LOF) and Isolation Forest (IF). The experiment's findings show that our suggested model surpassed the individual supervised and unsupervised Machine Learning algorithms in terms of four evaluation metrics, achieving an accuracy of 84.78%, a precision of 99.99%, Recall of 84.78% and a F1-score equally to 91.76%.

These findings demonstrate the potency and reliability of our suggested approach, which may be applied as a successful countermeasure to the spreading Arabic textual spam across a variety of internet platforms. Future research can examine how well this method can be applied to different domains and datasets.

Acknowledgment

This work has been sponsored by the General Directorate for Scientific Research and Technological Development, Ministry of Higher Education and Scientific Research (DGRSDT), Algeria.

- [1] L. FIRTE, C. LEMNARU, and R. POTOLEA. "Spam detection filter using KNN algorithm and resampling". In : Proceedings of the 2010 IEEE 6th international conference on intelligent computer communication and processing. IEEE, 2010. p. 27-33.
- [2] RC. PATIL and DR. PATIL, "Web spam detection using SVM classifier". In : 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO). IEEE, 2015. p. 1-4.
- [3] F. QIAN, A. PATHAK, YC. HU, ZM. Mao and Y. Xie, "A case for unsupervised-learning-based spam filtering". ACM SIGMETRICS performance evaluation review, 2010, vol. 38, no 1, p. 367-368.
- [4] M. MANAA, A. OBAID, M. DOSH. "Unsupervised approach for email spam filtering using data mining". EAI Endorsed Transactions on Energy Web, 2021, vol. 8, no 36.
- [5] A. HARISINGHANEY, A. DIXIT, S. GUPTA, A. Arora. "Text and image based spam email classification using KNN, Naïve Bayes and Reverse DBSCAN algorithm". In : 2014 International Conference on Reliability Optimization and Information Technology (ICROIT). IEEE, 2014. p. 153-155.
- [6] H. NAJADAT, MA. ALZUBAIDI and I. QARQAZ. "Detecting Arabic spam reviews in social networks based on classification algorithms". Transactions on Asian and Low-Resource Language Information Processing, 2021, vol. 21, no 1, p. 1-13.
- [7] M. KIHAL and L. HAMZA. "Enhancing Efficiency of Arabic Spam Filtering Based on Gradient Boosting Algorithm and Manual Hyperparameters Tuning". In : International Conference on Applied CyberSecurity. Cham : Springer Nature Switzerland, 2023. p. 49-56.
- [8] R. ALHARTHI, A. ALHOTHALI and K. MORIA. "A real-time deep-learning approach for filtering Arabic low-quality content and accounts on Twitter". Information Systems, 2021, vol. 99, p. 101740.
- [9] M. MANEVITZ, and M. YOUSEF, "One-class SVMs for document classification". Journal of machine Learning research, 2001, vol. 2, no Dec, p. 139-154. M. GOLDSTEIN, and A. DENGEL, "Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm". KI-2012: poster and demo track, 2012, vol. 1, p. 59-63.
- [10] O. ALGHUSHAIRY, R. ALSINI, T. SOULE, "A review of local outlier factor algorithms for outlier detection in big data streams". Big Data and Cognitive Computing, 2020, vol. 5, no 1, p. 1.
- [11] FT. LIU, KM. TING, Z. ZHOU, "Isolation forest". In : 2008 eighth IEEE international conference on data mining. IEEE, 2008. p. 413-422.
- [12] X. DONG, Z. YU, W. CAO. "A survey on ensemble learning". Frontiers of Computer Science, 2020, vol. 14, p. 241-258.
- [13] M. KIHAL and L. HAMZA. "Robust multimedia spam filtering based on visual, textual, and audio deep features and random forest". Multimedia Tools and Applications, 2023, p. 1-19.
- [14] Imam N (2020), Health-related Spam Campaigns, Mendeley Data, V1, <https://doi.org/10.17632/rgrvt5x4tk.1>

Investigating explainable AI for enhanced atherosclerosis detection and decision transparency

Amel Laidi
Faculty of Engineering, LIMOSE Laboratory
M'Hamed Bougara University
Boumerdes, Algeria
a.laidi@univ-boumerdes.dz

Mohammed Ammar
Engineering Systems and Telecommunication Laboratory
University M'Hamed Bougara
Boumerdes, Algeria
m.ammar@univ-boumerdes.dz

Abstract— Atherosclerosis, a cardiovascular disease, is commonly diagnosed through non-invasive imaging methods like Coronary CT Angiography (CCTA). Deep learning algorithms have demonstrated remarkable potential in assisting with the classification of CCTA images. However, the inherent opacity in the decision-making processes of black-box AI models presents a significant challenge in the medical field. Healthcare professionals require clear insights into the rationale behind AI system recommendations. In this research, we investigate the advantages of Explainable AI (XAI) algorithms in the context of a previously established deep learning model for atherosclerosis classification from CCTA images. Our study not only highlights the capability of XAI in elucidating the model's decision-making process but also showcases its potential in identifying misclassified cases, thereby contributing to the refinement and enhancement of the model's performance.

Keywords— deep learning, atherosclerosis, coronary CT angiography, Explainable Artificial Intelligence, GradCam.

I. INTRODUCTION

The extensive use of AI technology in the medical field, particularly in atherosclerosis screening, has drastically transformed disease diagnostics. Through the use of machine learning and deep learning techniques, AI has enabled early, precise, and comprehensive diagnosis of coronary atherosclerotic heart disease (CAD), a common cardiovascular disease with high morbidity, disability, and societal burden [1]. However, the prevalent reliance on Deep Neural Networks (DNNs) in AI models poses a substantial challenge regarding their explainability and transparency. This opacity in DNN models restricts the comprehension of how they reach diagnostic conclusions, which proves to be a significant obstacle in safety-critical medical domains. To address this gap, Explainable AI (XAI) methodologies have emerged to elucidate the decision-making processes of these models; they can aid in adhering to regulatory prerequisites and ethical standards, thus elevating the overall efficiency and impartiality of the system. [2]

Moreover, prior research has highlighted the importance of explainable AI in diverse fields, exemplified by works like [3]. The research conducted by Suryani et al. [4] emphasizes the critical role of AI-based tools in identifying lung tumors from chest X-ray images. The proposed methodologies in

their work, such as Seg-Grad-CAM (Semantic Segmentation via Gradient-Weighted Class Activation Mapping), align with the objectives of our investigation in providing precise and accurate localization of lesions or abnormalities within medical images.

In this context, our research aims to leverage the advancements in explainable AI methodologies for the task of atherosclerosis screening through CCTA images, ensuring not only the accuracy of disease diagnosis but also fostering trust and understanding in AI-assisted medical decision-making. The model under study is one that has been previously developed [5], fine-tuned, and specialized for the specific task of atherosclerosis screening.

II. BACKGROUND

A. What is AI ? and why deep learning?

Artificial intelligence is a wide term that refers to all types of computer systems trained to perform tasks that are normally associated with human intelligence or abilities, such as perception, reasoning, learning, and decision-making. AI systems are designed to analyze large amounts of data, recognize patterns, and make predictions or decisions based on that data.

There are different types of AI systems, including:

- **Rule-based systems:** AI systems that use pre-defined rules to make decisions or perform tasks. They are limited to the rules that have been programmed into them and cannot adapt to new situations.
- **Machine learning systems :** AI systems that can learn from data and improve their performance over time. There are different types of machine learning, including supervised learning, unsupervised learning, and reinforcement learning.
- **Deep learning systems :** a subset of machine learning systems that use artificial neural networks to analyze large amounts of data and learn from it. Deep learning systems are particularly effective in tasks that involve image and speech recognition, natural language processing, and other complex tasks.

Deep learning involves the use of artificial neural networks that consist of multiple layers of interconnected nodes, each of which performs a specific function in the processing of data. The input data is fed into the network, and the network gradually learns to recognize patterns in the data by adjusting the weights of the connections between nodes.

Deep learning has shown tremendous promise in computer vision and healthcare applications. It can analyze and interpret complex images, such as X-rays or MRI scans, with greater accuracy than traditional image analysis techniques. This ability has enabled medical professionals to diagnose diseases and conditions at an early stage, leading to better treatment outcomes. Additionally, deep learning has been used in healthcare to improve patient outcomes through personalized treatment plans. By analyzing vast amounts of patient data, deep learning algorithms can identify patterns and correlations that would be impossible for human doctors to detect. This has the potential to lead to more accurate diagnoses, better treatment plans, and ultimately, improved patient outcomes. Overall, the combination of deep learning, computer vision, and healthcare has the potential to revolutionize the way we diagnose and treat diseases, ultimately leading to better healthcare for everyone. [6]

B. Deep learning for atherosclerosis detection from Coronary CT Angiography

Atherosclerosis is a common cardiovascular disease that is characterized by the accumulation of fatty deposits in the walls of arteries. Coronary CT Angiography (CCTA) is the key technique for atherosclerosis screening, it is a non-invasive imaging technique that can be used to visualize the coronary arteries and detect the presence of atherosclerotic plaques [7]. However, interpreting CCTA images can be challenging and time-consuming, particularly when there are multiple plaques or complex plaque morphology.

Deep learning algorithms have been developed to assist in the classification of CCTA images by automatically identifying and characterizing atherosclerotic plaques based on their morphology, composition, and location within the coronary arteries. These algorithms can analyze large amounts of data and learn to recognize patterns that are indicative of different types of plaques, such as calcified or non-calcified plaques.

In a recent study [5], we created a deep learning model and trained it on a publicly available dataset.

Dataset :

The dataset we used is an open-source collection of Coronary CT Angiography images for screening atherosclerosis. It consists of Mosaic Projection View (MPV) images of 18 views of straightened coronary arteries from 500 patients, created by combining unique ray-traced projections. The dataset was partitioned into 300 training images, 100 testing images, and 100 validation images. To balance the dataset, the training images were augmented 6-fold. The validation dataset contains one randomly selected artery per normal case and one diseased case. [8]

Model:

The model we used was a pretrained Residual network (ResNet) [9] with 101 layers.

ResNet was first presented at the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2015. It innovated by incorporating additional branches within its architecture. Specifically, one branch, known as the identity block, forwards information unchanged, while the other, the convolutional block, processes data akin to a standard layer. The unaltered data, referred to as the "residual," is combined with the original signal traversing the network without modification. This architectural division ensures that one branch solely transmits gradients without modifying them. Deep residual networks are constructed by stacking these blocks, enabling robust learning through powerful gradient propagation.

The concept of residual blocks is visually demonstrated in the reference Fig. 1. These blocks' core lies in the "jump connection," which defines the essence of the residual blocks.

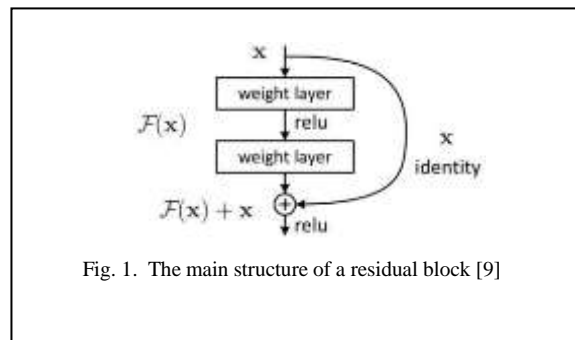


Fig. 1. The main structure of a residual block [9]

Residual Networks emerged as a solution to a prevalent issue encountered in deeper and more intricate networks—vanishing and exploding gradients. This phenomenon arises during the training of very deep networks when derivatives or slopes become exceedingly large or small, sometimes exponentially so, thereby complicating the training process.

There has been evidence that residual networks are easier to optimize and the gain in their accuracy is directly proportional with a considerable increase in their depth [10]. We used the resnet101 which has 101 layers.

Hyperparameters:

Mini-batch size : a segmented subset extracted from the training dataset via shuffling and partitioning. Its size can range from a single image to encompassing all examples within the training dataset.

Given the model's size, we trained on a mini-batch size of 64, which is neither too big nor too small. We used a learning rate of 10^{-4} which allowed the model to learn fast enough without missing local minima.

Dropout: a regularization technique employed during the training phase, where it randomly deactivates multiple

neurons within the network based on a pre-defined probability. This exclusion prevents their participation in both forward and backward propagation. Consequently, this alters the model's architecture during each iteration, fostering a more resilient training process. The implementation of Dropout notably enhances accuracy and substantially diminishes the required training time.

We used 50% dropout, which made the model more universal.

Number of training epochs: the number of cycles the model would go through the dataset. Training for too long epochs can cause the model to overfit, while not training enough can lead to underfitting.

The model trained for 20 epochs which made it perform well without overfitting.

The model took about 50 hours to train. It achieved 95.21% of accuracy, 90.48% positive predictive value, and 95.6% negative predictive value. The detailed results are illustrated in Fig. 2.

Output class	NEGATIVE	1058 88.8%	49 4.1 %	95.6 % 4.4 %
	POSITIVE	8 0.7 %	76 6.4 %	90.5 % 9.5 %
		99.2 % 0.8 %	60.8 % 39.2 %	95.2 % 4.8 %
		NEGATIVE	POSITIVE	

Target Class

Fig. 2. Confusion matrix for the Resnet model

C. Challenges and limitations of black-box AI models

Black-box AI models, which refer to AI systems that are difficult or impossible to interpret, particularly Deep learning models, have become increasingly popular in the medical field for tasks such as atherosclerosis screening. However, there are several challenges and limitations associated with these models that need to be addressed to ensure their safe and effective use in clinical settings.

Not understanding the full process happening inside the model leads to a lack of transparency in the decision-making process. This can be particularly problematic in the medical field, if healthcare professionals cannot understand how an AI system arrived at a particular diagnosis or recommendation, they may be hesitant to act on it.

To address these challenges and limitations, recent research worked on developing AI models that are more transparent and interpretable, and to ensure that they are trained on unbiased and representative data. Additionally, there needs to be ongoing monitoring and evaluation of AI

systems to ensure that they remain accurate and up to date with changes in the data and patient population. By addressing these challenges and limitations, the potential of black-box AI models to improve atherosclerosis screening and other medical tasks can be fully realized.

D. Explainable AI Algorithms

Explainable AI (XAI) algorithms are a type of AI system that is designed to be transparent and interpretable, allowing humans to understand how the AI system arrived at its decision or recommendation. XAI algorithms are intended to address the limitations of black-box AI models, which are often difficult or impossible to interpret.

There are several popular explainable AI algorithms that are used to provide transparency and interpretability to AI models. In this paper we took interest in three algorithms:

LIME (Local Interpretable Model-agnostic Explanations): LIME is a model-agnostic method for explaining the predictions of any black-box model. LIME works by creating a simpler, interpretable model that is locally faithful to the black-box model around the instance being explained. The simpler model can then be used to provide an explanation for the black-box model's prediction. [11]

LIME works by generating perturbations around the instance being explained and measuring the impact of these perturbations on the model's output. The algorithm then uses these perturbations to create a simpler, interpretable model that is locally faithful to the black-box model around the instance being explained.

LIME can provide a way to explain the prediction of any black-box model, without requiring knowledge of its internal workings. LIME is also a model-agnostic algorithm, meaning that it can be used with any type of model, including neural networks, decision trees, and support vector machines.

Grad-CAM (Gradient-weighted Class Activation Mapping): Grad-CAM is an algorithm for generating heatmaps that visualize the importance of different regions of an image for a given classification decision. Grad-CAM works by computing the gradients of the output class score with respect to the feature maps in the final convolutional layer of a neural network. The resulting gradient-weighted maps are then used to generate a final heatmap that visualizes the importance of different regions of an image for a given classification decision. [12]

The advantage of Grad-CAM over black-box models is that it provides a way to visualize the decision-making process of the neural network, which can be difficult to interpret using traditional methods. Grad-CAM also provides a way to identify which regions of an image are most important for the model's decision, which can be particularly useful in medical imaging.

Occlusion Sensitivity: Occlusion sensitivity is a technique for visualizing the importance of different regions of an image for a given classification decision by systematically occluding different parts of the image and

measuring the resulting change in the model's output. By comparing the model's output for the original image and the occluded images, it is possible to identify the most important regions of the image for the model's decision. [13]

Occlusion sensitivity is also a simple and easy-to-implement algorithm, making it a useful tool for interpreting the decisions of machine learning models.

E. XAI algorithms in atherosclerosis classification.

The use of explainable AI algorithms in atherosclerosis classification offers the typical advantages such as increased transparency and improved decision making, but it also adds another layer to the task of the classification. The main advantage of classifying the images directly, without first segmenting them, is the rapidity of the task. When using XAI algorithms, we can add a layer on the explanation, where the highlighted areas are most likely areas of high activity, in our case it could be the location of buildup. This not only allows for a quick detection of the disease, but it also guides the doctor to which region to investigate for higher risk of plaque.

III. RESULTS AND DISCUSSION

For a full analysis of the model, we ran the three algorithms on a number of images from the validation dataset. We tested true positives, true negatives, and false negatives.

False negatives were not part of the scope of this study, because they do not pose an urgent risk. A healthy patient that has been falsely classified as sick would be kept under medical care and eventually identified as healthy with further tests. While sick patients that were classified as healthy can be discharged without medical care, leading to further complications of their health.

A. True positive

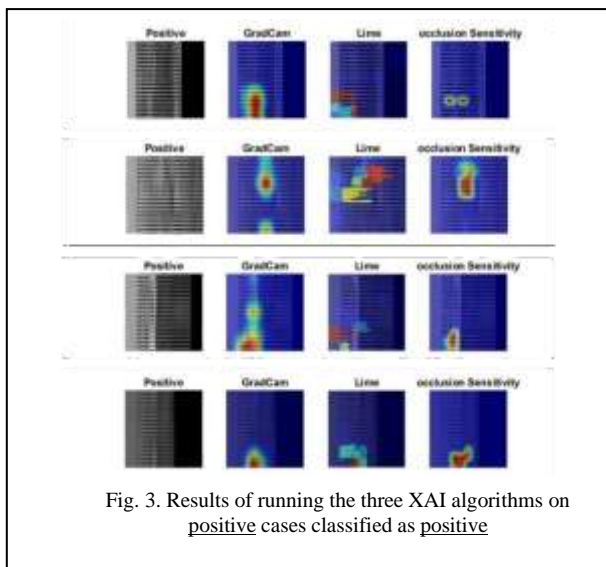


Fig. 3. Results of running the three XAI algorithms on positive cases classified as positive

True positive (TP) is a term used to describe a correctly predicted positive case. In other words, it is the case where the model correctly identifies a positive example as positive. In our case, the model has 90.48% positive predictive value, so it should be good at classifying positive cases.

At first sight, all three algorithms show similar activity at the same regions of the image (Fig. 3). This means that the deep learning model we used relied heavily on this region for its final decision.

The focus is on one small region, being highlighted in red, with no significant activity anywhere else.

The explainable models offer an extra layer of performance, it shows the region with higher probability of plaque and residue, which is useful for the rest of the medical care process.

B. True Negative (TN)

True negative (TN) is a term used to describe a correctly predicted negative case. In other words, true negative occurs when the actual class of an image is negative, and the model correctly predicts it as negative.

Our model has a 95.6% negative predictive value; and although it is considered as an impressive performance, it would help more to understand the reason behind the classification.

Both GradCam and Lime show a more spread-out activity (Fig. 4), which is quite different from positive cases. However, occlusion sensitivity for TN does not look too different from the one of TP, so it cannot be used for comparison.

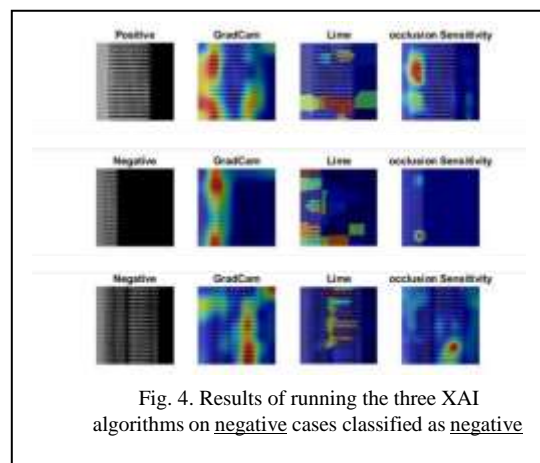


Fig. 4. Results of running the three XAI algorithms on negative cases classified as negative

C. False Negative

A false negative (FN) is an error in binary classification where a negative outcome is predicted when the actual outcome is positive. In other words, a false negative occurs when a model fails to recognize a positive image or when it incorrectly classifies a positive example as negative. False negatives are particularly important in medical diagnoses as

they can lead to the dismissal of sick people when their case can be critical or even fatal.

All three algorithms show an activity that is less focused than true positives (Fig. 5), which explains the error in classification. It still is different from true negatives. The images have small areas of high activity (red regions), but also significantly large areas with lesser activity (orange and yellow regions).

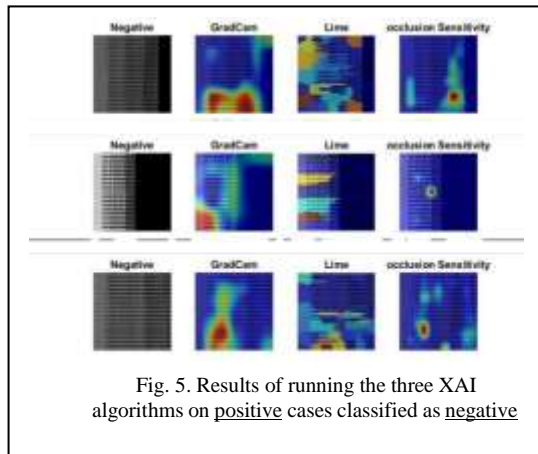


Fig. 5. Results of running the three XAI algorithms on positive cases classified as negative

D. Comparison between FN and TN

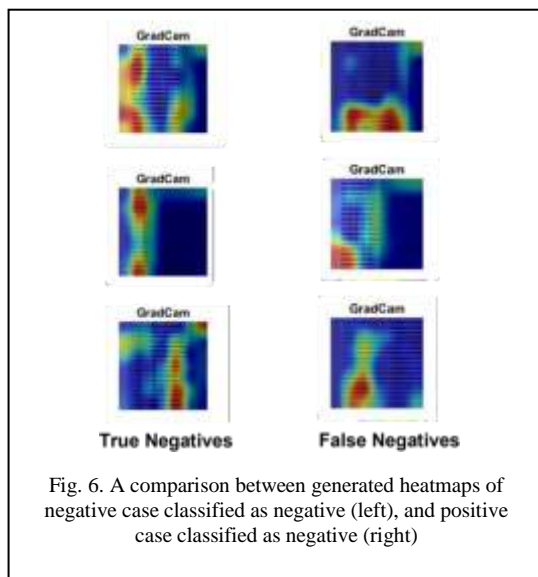


Fig. 6. A comparison between generated heatmaps of negative case classified as negative (left), and positive case classified as negative (right)

As mentioned earlier, false negatives are of critical importance in medical applications, because they refer to patients who are positive but were classified by the model as negative. Their dismissal can lead to fatal complications.

In this experiment, we compared the GradCam analysis of true negative cases and false negative cases. The choice of GradCam was due to it being the only algorithm that showed a significant variation in behavior.

Fig. 6 shows that true negative cases generate heatmaps with a larger surface for high activity (red area) that is well spread vertically. This suggests that a larger region of the

image was considered for the final decision of the model. Unlike true positive cases, where only a small region was considered, probably the location of plaque.

As for false negatives, the heatmaps have a smaller localized surface for high activity (red area), with other region with lesser activity (orange and yellow areas). Having the lesser activity can explain WHY the images were falsely classified as negative, a larger area was considered for the final decision.

In conclusion, GradCam shows a tangible difference between false negatives and true negatives, showing significant activity in small areas for false negatives, indicating not only that they were wrongfully classified, but the region for the physician to focus on before making their final decision. Eventually improving the performance of the model.

IV. CONCLUSION

Artificial intelligence has significantly transformed healthcare applications, demonstrating human-level performance on classification tasks in a fraction of the time. However, the opacity in the decision-making process has hindered the trust of many healthcare professionals.

Explainability algorithms serve as a vital bridge to unlock the full potential of AI, augmenting trust alongside superior performance. In this study, we revisited a deep learning model designed for atherosclerosis screening via Coronary CT Angiography. To gain insights into the model's reasoning, we applied three prominent explainable AI algorithms: Gradient-weighted Class Activation Mapping (GradCam), Local Interpretable Model-agnostic Explanations (LIME), and Occlusion sensitivity. The results not only provided additional information regarding classifications but also exhibited a degree of performance enhancement.

Positive cases correctly classified as positive (True Positives, TP) exhibited high focused activity in a specific area, suggesting the presence of plaque. Conversely, positive images erroneously classified as negative (False Negatives, FN) displayed similar concentrated activity, prompting healthcare professionals to reevaluate these regions for a final determination regarding the presence of plaque and atherosclerosis.

Negative cases correctly classified as negative (True Negatives, TN) showcased a more diffuse heatmap, indicating that the model considered a broader image area for its decision-making. GradCam proved to be the most effective algorithm in distinguishing between FN and TN cases.

The primary contribution of this study lies in revealing that the deep learning model primarily relies on the detection of localized high activity, likely representing the position of plaque, thereby rendering the model more trustworthy. Furthermore, the study provides valuable insights by identifying the probable plaque locations in positive cases and, most importantly, offers a means to differentiate true negative cases from false negatives. This augments the

overall performance of the model and reduces the risk of prematurely dismissing patients with underlying conditions.

While the study provides valuable insights, it is limited by the number of samples and should be further generalized in future research. Additionally, it is advisable to test other models with varying performance for comparative analysis.

REFERENCES

- [1] H. Lu, Y. Yao, L. Wang, J. Yan, S. Tu, Y. Xie, W. He and others, "Research progress of machine learning and deep learning in intelligent diagnosis of the coronary atherosclerotic heart disease," *Computational and Mathematical Methods in Medicine*, vol. 2022.
- [2] C. Molnar, *Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models*, Lulu. com, 2020.
- [3] S. G. Kim, S. Ryu, H. Kim, K. Jin and J. Cho, "Enhancing the Explainability of AI Models in Nuclear Power Plants with Layer-wise Relevance Propagation," in *Proceedings of the Transactions of the Korean Nuclear Society Virtual Autumn Meeting.*, Jeju, Korea, 2021.
- [4] A. I. Suryani, C.-W. Chang, Y.-F. Feng, T.-K. Lin, C.-W. Lin, J.-C. Cheng and C.-Y. Chang, "Lung Tumor Localization and Visualization in Chest X-Ray Images Using Deep Fusion Network and Class Activation Mapping," *IEEE Access*, vol. 10, pp. 124448--124463, 2022.
- [5] A. Laidi, M. Ammar, M. E. H. Daho and S. Mahmoudi, "Deep Learning Models for Coronary Atherosclerosis Detection in Coronary CT Angiography," *Current Medical Imaging*, vol. 19, 2023.
- [6] A. Rajkomar, J. Dean and I. Kohane, "Machine learning in medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347--1358, 2019.
- [7] M. Portegies, P. Koudstaal and M. Ikram, "Cerebrovascular disease," in *Handbook of Clinical Neurology*, F. B. D. F. S. Michael J. Aminoff, Ed., Elsevier, 2016, pp. 239-261.
- [8] V. Gupta, M. Bigelow, B. Erdal, L. Prevedello and R. White, "Image dataset for a CNN algorithm development to detect coronary atherosclerosis in coronary CT angiography," *Mendeley Data*, vol. 1, 2019.
- [9] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [10] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [11] M. T. Ribeiro, S. Singh and C. Guestrin, "" Why should i trust you?" Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016.
- [12] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization.," *arXiv preprint arXiv:1610.02391*.
- [13] "visualizing and understanding convolutional networks.," in *European conference on computer vision*, 2014.

Toward Mitigating Tourists' Indifference in POI recommendation systems using MAS

Zoulikha MAGHNI SANDID, Zohra SLAMA
EEDIS Laboratory, Djilali Liabes University of Sidi Bel
Abbes, Algeria
maghnisandid.z@yahoo.com, zslama@yahoo.com

Nassim DENNOUNI
HIGHER SCHOOL OF MANAGEMENT-TLEMCEM
dennouninas@gmail.com

Abstract— POI recommendation is one of the artificial intelligence techniques used to personalize a user's experience in the field of smart tourism. However, this technique suffers from the problem of sparse data due to the indifference of the ratings of the places visited by the user. To mitigate this problem, we propose in this work a Multi-agent System for Reconciling POI Recommendation Algorithms (MSRPRA) using three types of POI recommendation algorithm that exploit user ratings, check-ins during visits and explicitly declared trust relationships between users. Additionally, a voting system is employed to merge the results of these three algorithms.

Keywords—*Point of Interest recommendation; multi-agent system; collaboratif filtering;*

I. INTRODUCTION

Nowadays, artificial intelligence (AI) is integrated in several fields such as e-tourism [1], e-commerce [2], etc. The recommendation is one of the techniques of AI that can be used by these fields to personalize and improve the user experience. This technique uses AI algorithms (deep learning [3], decision tree [4], etc.) to exploit the potential of the vast amounts of data collected during the use of systems to make them more intelligent.

These systems include smart tourism, which uses ubiquitous technologies and AI algorithms to enhance the tourist experience by assisting them on their trip to discover new tourism destinations [5]. Artificial Intelligence is currently supporting this field to make the tourist experience increasingly personalized according to the needs of each tourist [6]. Indeed, to recommend the most relevant places for tourists, AI recommends Points of Interest (POIs) using the tourist's profile, which can either be (1) declared by the tourist through their demographic data, preferred categories, etc., or (2) inferred from their evaluations, the check-ins he carried out during the tourist visit, etc [7]. In this work, we will focus on the implicit profile of the tourist, which can be formed using information gathered from the visit's history such as browsing duration, clicks, comments, etc. [8]. Our interest in this type of profile is justified by the fact that tourists do not tend to express their preferences or personal data when using smart tourism tools [7].

In the literature, Recommendation Systems (RS) [8] are divided into two main classes: (1) Collaborative Filtering Recommendation Systems (CFRS) and (2) Content-Based Recommendation Systems (CBRS). The CBRS [7] match the

content description of the POIs with the tourist's profile, whereas the CFRS [9] [10] calculate the similarity between tourists to predict the POIs to visit. These systems rely on the explicit ratings of the POIs given by tourists to compute similarities used to initiate their recommendation processes. However, the majority of users tend to overlook rating the places they've visited, which can lead to cold-start problems due to data scarcity. To resolve this cold-start issue in RS due to the indifference of POI ratings, several studies combine the results of multiple recommendation algorithms [11]. This type of solution helps mitigate data scarcity problems in ratings by incorporating other types of data such as check-ins or the number of comments per POI. Nevertheless, reconciling the lists of POIs obtained by recommendation algorithms using these different types of data (rating, check-in, and comments) presents a real challenge. To address this challenge, several recent studies in the literature [13-17] have utilized multi-agent systems (MAS) as a solution to user indifference issues because they allow the merging of POI lists obtained through the simultaneous use of multiple recommendation algorithms.

For these reasons, in this article we propose a Multi-Agent System for Reconciling POI Recommendation Algorithms, which we will call MSRPRA. This system uses three types of agent: (1) context agents that collect the context of each user, (2) method agents that return ordered lists of POIs according to the algorithm used and the context adopted, and (3) a coordinator agent that selects the most relevant POIs by merging these lists of POIs.

The rest of the document is structured as follows. In section 2, we present the state of the art on SR development approaches using SMAs in order to explain our research motivations. In section 3, we present our proposed method for reconciling different recommendation algorithms using an SMA-based approach. Before concluding, in Section 4 we detail our analysis of the problem and discuss the advantages and limitations of our approach. Finally, in Section 5, we summarize the contributions of our article in order to propose some perspectives for our work.

II. RELATED WORK

In recent years, several research projects have focused more on intelligent Recommendation Systems (RS) that aim to satisfy users' preferences using Multi-Agent Systems (MAS) [12]. Multi-Agent Recommendation Systems (MARS) are SRs that integrate autonomous and cooperative agents to improve the quality of recommendations.

In [13], the authors propose a multi-agent recommendation application that can recommend tourist locations in the city of Valencia (Spain) and propose a plan of activities. However, this application requires the intervention of the tourist at each stage to specify their needs and note the places they have visited.

In the same context, in [14], the authors proposed a MARS for e-tourism using reputation-based collaborative filtering, enabling the recommendation of services such as hotels, places to visit, and restaurants. The system assigns an agent for each service, and each agent returns the list of that particular service. However, although this approach leverages the strengths of MAS in distributing and cooperating on recommendation tasks, this solution presents the issue of a sparse matrix.

In another context, to personalize websites, the authors in [15] integrated two recommendation techniques: association rules and collaborative filtering. They incorporated agents to reduce response time and to separate processing from model updates to enhance performance. Their approach is incremental; the system integrates additional information after each session. However, their system requires a lot of updates at the end of each session.

A centralized MAS architecture was proposed in [16], to recommend locations to users. The server agent collects the data from the client agents, and then calculates the predictions of the evaluations. The list of the most relevant places is sent to the client agent, which in turn selects those within 1 km of the user's current location and displays them to the tourist.

A video recommendation system based on MAS was proposed in [17]. The authors propose 7 agents (location, age, financial, identity, personality, needs, social) and an information centre agent (ICA) responsible for collecting and processing information. Each agent specifies whether the POI is interesting or not, then the ICA ranks the POIs. However, in this system, the history is not exploited, and each agent does not classify the films, he only expresses his interest or not towards the film.

Table 1 represents a synthesis of recommendation works based on MAS.

Works	Application domain	Used Techniques	Recommended objects
[13]	Tourism	Demographic data filtering, Content-based filtering	Location and plan of activities
[14]	Tourism	Collaborative filtering based on reputation	hotels, places to visit, restaurants
[15]	Web pages	Association rules and collaborative filtering	Next web page
[16]	Points of interest	Sentiment analysis	Next location
[17]	Movies	Content-based filtering	List of movies
Our approach	Tourism	Collaborative filtering	List of POIs

Previous works have used agents for their autonomies and their abilities to cooperate with each other in order to accomplish global tasks. However, these systems do not address the problem of indifference in the rating of places visited by users. This problem arises when users do not explicitly express their preferences (providing ratings for POIs or leaving comments). To mitigate this indifference problem, we will explore the possibility of using SMAs to exploit the temporal and geographical behavior of tourists and existing trust relationships with other users.

III. PROPOSED METHOD

The indifference of item/POI scoring is a real problem in the field of recommendation in general and tourism more specifically. Therefore, in this section, we propose a Multi-Agent System for Reconciling POI Recommendation Algorithms (MSRPRA) to alleviate this problem.

This system consists of five agents, which are intelligent, autonomous and cooperative. We propose three agent roles: (1) the context agent, which collects the context of each user, (2) the method agent, which returns ordered lists of POIs according to the algorithm used and the context adopted, and (3) the coordinator agent, which selects the most relevant POIs by merging these lists of POIs. The following figure illustrates the architecture of our system:

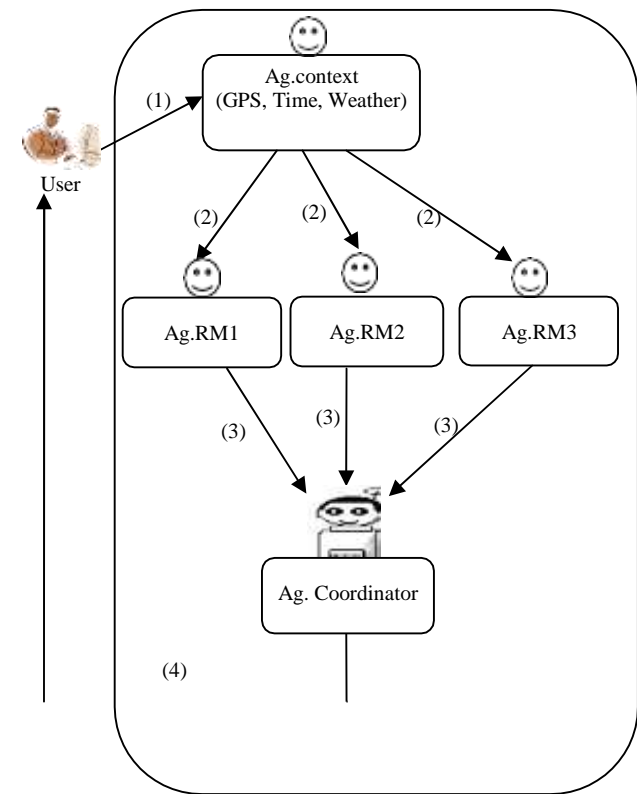


Fig.1. General Architecture of our system

As shown in fig1, in step (1) the context agent collect context data (location, weather, time) from user phone, and selects the unvisited POIs that are closest to the user's current location, are currently available and are appropriate to the

current weather situation. In step (2) the context agent sends the preliminary list to the recommendation agents (Ag.RM1, Ag.RM2 and Ag.RM3). On receipt of the preliminary list of POIs, each recommendation agent applies its own recommendation method to select only the most relevant top-k locations, and each agent, in step (3), sends a sorted list (list1, list2 and list3) to the coordinating agent. In the final step (4), the coordinating agent applies a negotiation technique to aggregate the three lists into a final list, which is returned to the user.

In the following, we explain the role of each agent in our system.

A. Context agent

This agent collects the user's contextual data, such as the location deduced from the Smartphone's GPS, the date and time of day and the weather, which can be retrieved using a weather API. This data is used to select preliminary list of POIs that are closest to the user's current location, are currently available and are appropriate to the current weather situation.

B. Recommendation Agents

These agents exploit the history of places visited (the number of visits, the ratings given to POIs already visited, etc.) as well as the trust relationships explicitly declared by the user. In our system, we propose three agents: Ag.RMMethod1, Ag.RMMethod2 and Ag.RMMethod3 where each agent performs a recommendation method.

These three POI recommendation methods use three types of similarity: Pearson similarity [18], Jaccard similarity [19] and similarity based on trust [20].

Each agent uses a type of similarity to calculate predictions for unvisited POIs based on tourists' previous visit history. Next, each agent returns the top-k most relevant places ranked in descending order of prediction values. Finally, we obtain three lists of POIs, with each list associated with a single agent.

C. Coordinator Agent

In MAS, negotiation is a form of coordination among agents that allows them to communicate and manage conflicts [21]. There are several negotiation approaches, such as game theory, heuristics, and argumentation and voting. Game theory is based on mathematics, where agents are in a game and their goal is to maximize their gains [22]. However, an agent must anticipate the behavior of all the others and find the optimal solution, which requires high a computational cost. Heuristics were developed to solve the problem of game theory. In these negotiation models, agents rely on their reasoning and strategies for decision-making without pursuing the optimal solution [23]. However, this means they do not explore the entire space of solutions. In game theory and heuristics, agents only focus on proposals. These methods do not allow agents to explain the motivations behind their choices. Argumentation is an approach that allows agents, in addition to presenting proposals, to provide explanations for these proposals and

their decisions, whether to accept or reject them [24]. Each agent has the possibility of accompanying its proposals with arguments to explain why the others should consider them favorably. However, this technique requires additional communication costs. The vote is a technique in the theory of collective decision-making or social choice that allows the selection of one alternative from various possible options. It provides participants with the opportunity to express their preferences among a set of solutions [21].

For our system, we have chosen the voting technique, because this technique seems to be adapted to our design, where the agents do not coordinate.

In our design, the recommendation agents do not communicate. Each agent is autonomous and its objective is to return a list of places which is calculated based on one of the three proposed recommendation methods. These agents communicate with the coordinator to send it the lists.

The coordinating agent's objective is to return the final sorted list of places to recommend to our user. After receiving the three lists: list1, list2 and list3 from the agents: Ag.RMMethod1, Ag.RMMethod2 and Ag.RMMethod3 respectively, the agent applies the voting technique to aggregate the three lists into a final list.

Borda Method: The Borda voting system is frequently used in situations involving the selection of multiple winners; it relies on a weighted vote. This method involves each agent ranking alternatives in order of preference and assigning a positive number of points, 'n', to the top-ranked alternative, 'n-1' points to the second-ranked alternative, and 1 (or 0) point to the last alternative on the list. The number 'n' must be less than or equal to the number of alternatives [25].

Example: We assume we have a set L of 10 unvisited locations: $L = \{l_1, l_2, l_3, l_4, l_5, l_6, l_7, l_8, l_9, l_{10}\}$, and each agent must return the top-5 locations. $N = 4$. We have chosen to present three scenarios in the following:

Scenario 1: the three agents return 3 lists with the same locations.

TABLE II. LISTS OF LOCATIONS TRIED BY EACH AGENT FOR SCENARIO 1

List_AgRM1	List_AgRM2	List_AgRM3	N
l_3	l_2	l_3	4
l_6	l_6	l_8	3
l_2	l_1	l_1	2
l_1	l_3	l_2	1
l_8	l_8	l_6	0

For each location, we calculate the score as follows:

$$l_3: 4+1+4= 9$$

$$l_6: 3+3+0= 6$$

$$l_2: 2+4+1= 7 \rightarrow \text{Final list is: List}_F= \{l_3, l_2, l_6, l_1, l_8\}$$

$$l_1: 1+2+2= 5$$

$$l_8: 0+0+3= 3$$

Scenario 2: one of the agents returns a location that doesnot exist in the other lists.

TABLE III. LISTS OF LOCATIONS TRIED BY EACH AGENT FOR SCENARIO 2

List_AgRM1	List_AgRM2	List_AgRM3	N
l_1	l_2	l_3	4
l_5	l_4	l_2	3
l_2	l_6	l_1	2
l_4	l_1	l_4	1
l_6	l_5	l_5	0

For each location, we calculate the score as follows:

$$l_1: 4+1+2= 7$$

$$l_2: 2+4+3= 9$$

$$l_3: 0+0+4= 4 \rightarrow \text{Final list is: List}_F = \{l_2, l_1, l_4, l_3, l_5\}$$

$$l_4: 1+3+1= 5$$

$$l_5: 3+0+0= 3$$

$$l_6: 0+2+0= 2$$

Scenario 3: the three lists are different.

In this scenario, we need to expand the lists. The coordinator agent must request recommendations from the agents to send sorted lists of all unvisited places. Subsequently, we follow a similar procedure as in the previous scenarios and return the final list of top places.

IV. DISCUSSION

To mitigate the problem of user indifference towards place ratings, we used three recommendation algorithms based on three distinct data sources: user ratings, check-ins during visits, and explicitly declared trust relationships among users. To reconcile these algorithms, we employed the Borda voting method. As illustrated in the example in the previous section, we observed that regardless of the scenario, our system can merge the three lists, and we noticed that the final list contain the most relevant places from each algorithm.

The use of these methods allowed us to mitigate the cold start problem because each agent returns a list of the topPoints of Interest (POIs) to recommend by utilizing multiple data sources.

The Multi-Agent Systems enabled us to efficiently reconcile and merge the lists obtained from the threerecommendation algorithms described earlier.

Unlike the work [13] [14], our approach is solely based on user behavior (history, trust relationships, and ratings). However, this is still inadequate as integrating demographic data and Point of Interest (POI) characteristics could bring more accuracy to our system.

V. CONCLUSION

In this research, we designed a Multi-Agent System for Reconciling POI Recommendation Algorithms (MSRPRA). This system uses three types of POI recommendation algorithm: the first algorithm is based on Pearson similarity, the second algorithm uses Jaccard similarity and the last algorithm relies on trust relationships between users to calculate their similarities. These algorithms exploit three types of data: (1) user ratings, (2) check-ins during visits and explicitly declared trust relationships between users. Initial simulation results indicate that our MSRPRA can be a valuable tool for efficiently gathering crucial data to alleviate the problem of indifferent ratings of places visited by users. This work contributes to the field of POI recommendation, particularly in the context of Smart tourism, where the problem of data scarcity prevents the cold start of SRs thus causing tourist dissatisfaction.

Future research could focus on enhancing this system by integrating other algorithms such as Deep Learning or other types of data, like tourists' comments. Overall, our research establishes a solid foundation for implementing an intelligent smart tourism application for the Oran province.

VI. REFERENCES

- [1] S. DOLNICAR, Market segmentation for e-tourism. In : Handbook of e-Tourism. Cham : Springer International Publishing, 2022. pp. 849-863.
- [2] J. VIPIN, M. BINDOO, A. SATYENDRA. "An overview of electronic commerce (e-Commerce)". *Journal of Contemporary Issues in Business and Government*, 2021, vol. 27, no 3, pp. 665-670.
- [3] J. D. Kelleher, *Deep learning*. MIT press, 2019.
- [4] B. DE VILLE, "Decision trees". *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 5, no 6, pp. 448-455, 2013.
- [5] C. HUDA, A. RAMADHAN, A. TRISETYARSO, et al. "Smart Tourism Recommendation Model: A Systematic Literature Review", *International Journal of Advanced Computer Science and Applications*, 2021, vol. 12, no 12.
- [6] K. Kabassi. "Personalizing recommendations for tourists", *Telematics and Informatics* 27.1, 2010, pp. 51-66.
- [7] Z. Abbasi-moud, N.H. Vahda, J. Sadri, "Tourism recommendation system based on semantic clustering and sentiment analysis", *Expert Systems with Applications*, 2021. 167. 114324. 10.1016/j.eswa.2020.114324.
- [8] F. Ricci, L. Rokach, B. Shapira, *Recommender Systems Handbook*. Springer US, Boston, MA, 2015. <https://doi.org/10.1007/978-1-4899-7637-6>
- [9] J.B. Griesner, T. Abdessalem, H.Naacke, « Intégration des Influences Géographique et Temporelle pour la Recommandation de Points d'Intérêt », conférence Extraction et Gestion des Connaissances, 2016, pp.153-158
- [10] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, et Y. Rui. "Geomf: Joint geographical modeling and matrix factorization for point-of-interest recommendation". *KDD '14*. 2014.
- [11] L. Cai, J. Xu, J. Liu, T. Pei, "Integrating spatial and temporal contexts into a factorization model for POI recommendation", *international journal of geographical information science*, 2018, VOL. 32, NO. 3, 524-546.
- [12] Selmi, Afef, Zaki Brahmi, and M. Gammoudi. "Multi-agent recommender system: State of the art", *Proceedings of the 16th international conference on information and communications security*, 2014.
- [13] L. Sebastia, A. Giret, et I. Garcia, "A Multi Agent Architecture for Tourism Recommendation", in *Trends in Practical Applications of Agents and Multiagent Systems*, Y. Demazeau, F. Dignum, J. M.

- Corchado, J. Bajo, R. Corchuelo, E. Corchado, F. Fernández-Riverola, V. J. Julián, P. Pawlewski, et A. Campbell, Éd., in *Advances in Intelligent and Soft Computing*, vol. 71. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, p. 547-554. doi: [10.1007/978-3-642-12433-4_64](https://doi.org/10.1007/978-3-642-12433-4_64).
- [14] P. Bedi, S. K. Agarwal, V. Jindal, Richa, “ MARST: Multi-Agent Recommender System for e-Tourism Using Reputation Based Collaborative Filtering “, in *Databases in Networked Information Systems*, vol. 8381, 2014, pp. 189-201. doi: [10.1007/978-3-319-05693-7_12](https://doi.org/10.1007/978-3-319-05693-7_12).
- [15] A. J. Morais, E. Oliveira, et A. M. Jorge, “A Multi-Agent Recommender System”, in *Distributed Computing and Artificial Intelligence*, vol. 151, in *Advances in Intelligent and Soft Computing*, vol. 151, 2012, pp. 281-288. doi: [10.1007/978-3-642-28765-7_33](https://doi.org/10.1007/978-3-642-28765-7_33).
- [16] C. Cavallaro, G. Verga, E. Tramontana, et O. Muscato, “ Multi-Agent Architecture for Point of Interest Detection and Recommendation”, Workshop "From Objects to Agents", 2019.
- [17] A. M. Mansour, M. A. Obeidat, et B. Hawashin, “A novel multi agent recommender system for user interests extraction”, *Cluster Comput*, vol. 26, n° 2, pp. 1353-1362, avr. 2023, doi: [10.1007/s10586-022-03655-7](https://doi.org/10.1007/s10586-022-03655-7).
- [18] J. Tang, X. Hu, H. Liu, “Social recommendation: a review.”, *Social Network Analysis and Mining* 3, 2013, pp. 1113-1133.
- [19] S. Bag , S.K. Kumar, and M.k. Tiwari. “An efficient recommendation generation using relevant Jaccard similarity”, *Information Sciences* 483, 2019, pp. 53-64.
- [20] M. Ye, P. Yin, W.C. Lee, D.L. Lee, “Exploiting geographical influence for collaborative point-of-interest recommendation”, *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, 2011.
- [21] J. Ferber, *Les systèmes multi-agents: vers une intelligence collective*, InterEditions, 1995.
- [22] D.G. Pruitt. *Negotiation behavior*. Academic Press, 2013.
- [23] S.Kraus. “Negotiation and cooperation in multi-agent environments”. *Artificial intelligence*, 94(1-2) ,1997, pp. 79–97.
- [24] S. Kraus, N. NIRKE, K. Sycara. “Reaching agreements through argumentation”. In *Proceedings of the Twelfth International Workshop on Distributed Artificial Intelligence*, 1993, pp. 233–247.
- [25] D. Bouyssou, P. Perny, « Aide multicritère à la décision et théorie du choix social », 2016.

MLP-powered smart application to enhance efficiency and productivity in Algerian agriculture

M'hamed Mancer

LINFI Laboratory

Mohamed Khider University

Biskra, Algeria

mohamed.mancer@univ-biskra.dz

Labib Sadek Terrissa

LINFI Laboratory

Mohamed Khider University

Biskra, Algeria

terrissa@univ-biskra.dz

Soheyb Ayad

LINFI Laboratory

Mohamed Khider University

Biskra, Algeria

s.ayad@univ-biskra.dz

Hamed Laouz

LINFI Laboratory

Mohamed Khider University

Biskra, Algeria

laouz.hamed@univ-biskra.dz

Abstract—Artificial Intelligence (AI) has emerged as a pivotal driver of global economic growth, expanding its applications across various sectors. This paper introduces a smart application for crop selection tailored to the Algerian environment, aimed at enhancing smart agriculture in Algeria. By harnessing AI's capabilities to enhance efficiency and productivity, our system promises to stimulate economic growth and contribute to the well-being of the agricultural sector. It assists in making informed crop choices, maximizing yields, and resulting in significant time and cost savings for farmers. Our study presents a comprehensive analysis of the Multi-Layer Perceptron Classifier (MLP) model, standing out with an accuracy rate of 91.81%. This underscores its potential as a reliable tool for crop selection in Algerian agriculture.

Index Terms—Smart application, Machine learning, Crop selection, MLP classifier, Smart agriculture, Algerian environment.

I. INTRODUCTION

Evolved beyond its technological origins, AI stands as a driving force propelling global economic growth. Its widespread applications across sectors like healthcare and finance are reshaping societies and economies on a global scale. One arena where AI, Blockchain technology, and smart agriculture are set to make an indelible mark is agriculture in Algeria—an environment characterized by unique challenges and opportunities. This integration offers unprecedented potential to reshape the farming landscape. Several related works have underscored the efficacy of AI-driven and Blockchain-based solutions in agriculture, inspiring our endeavour [1] [9].

In related works, significant strides have been made in AI-driven agricultural solutions. For instance, in [2], a crop recommendation system based on various machine learning models, including Support Vector Machine (SVM), Random Forest (RF), Light Gradient Boosting Machine (LGBM), Decision Tree (DT), and K-Nearest Neighbors (KNN), was developed. These models trained using a dataset containing 22 crop types, demonstrated high accuracy, with RF performing best at 99.24%. This system is critical for informed decision-making, reducing land resource wastage for farmers and the government.

In [3], authors introduced a crop recommendation system leveraging machine learning algorithms and IoT sensors to collect soil moisture, temperature, macro-nutrient content,

and other environmental factors. Various algorithms, such as Artificial Neural Networks (ANN), Random Forest, Logistic Regression, and K-Nearest Neighbor (KNN), were assessed to optimize the crop recommendation system, aiding precise crop yield predictions for sustainable practices.

The aim of [4] is to develop a smart crop recommendation system using machine learning to enhance agricultural productivity. Gathering input from farmers, the system utilizes Random Forest, XGBoost, and Decision Tree algorithms, with XGBoost achieving the highest accuracy at 99.31%. The system's recommendations are expected to improve crop yields and reduce farmer losses.

Researchers in [5] introduced a crop recommendation system leveraging soil and climate information, using machine learning models. The XG-Boost classification model excelled in accuracy, recall, precision, and F1 score, resulting in a website aiding informed crop selection for farmers.

This study [6] focuses on a crop recommendation system using machine learning algorithms like Support Vector Machines, Random Forest, and Naive Bayes. The Naive Bayes classifier outperformed Random Forest and SVM with a 99.09% accuracy rate.

The primary focus of [7] is to use machine learning methods to create a crop recommendation system that meets farmers' needs. Achieving a remarkable 99.5% accuracy rate, this system considers variables like climate suitability, soil conditions, and market demands.

This study proposes a smart application tailored for the Algerian agricultural setting, utilizing MLP models that consider intricate soil characteristics and climate conditions. This advancement significantly enhances agricultural productivity and sustainability within the Algerian environment.

II. METHODOLOGY

This section outlines the system components and our methodology, as demonstrated in Fig. 1. Further details will be elaborated upon in the following subsections.

A. Dataset

The dataset utilized in this proposed system, sourced from Kaggle, comprises 22 different crop types and a total of

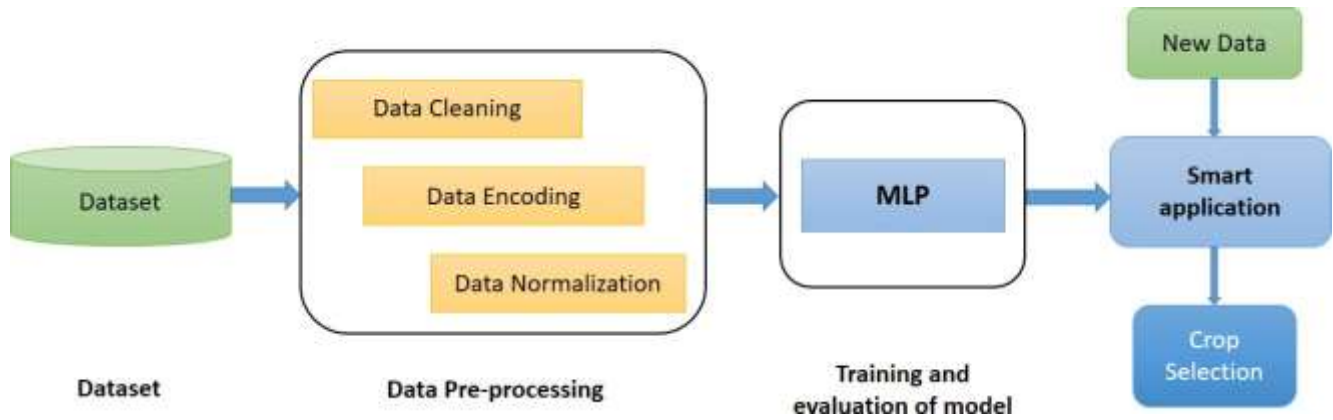


Fig. 1. System Architecture.

2200 data entries. This dataset is rich in agricultural parameters, including Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, pH, and rainfall. Each of these parameters plays a vital role in understanding and predicting crop behaviour.

N (Nitrogen): Nitrogen is an essential nutrient for plant growth and is often required in large quantities. It is a primary component of chlorophyll, which is crucial for photosynthesis. Nitrogen deficiency can lead to stunted growth and reduced crop yields [8].

P (Phosphorus): Phosphorus is vital for energy transfer within plants and is involved in the formation of DNA, RNA, and ATP. It plays a key role in root development and flowering. Insufficient phosphorus can result in poor root growth and delayed maturity [8].

K (Potassium): Potassium is critical for various plant processes, including photosynthesis, water uptake, and enzyme activation. It helps plants resist diseases and improves overall stress tolerance. A lack of potassium can lead to reduced fruit quality and increased susceptibility to pests [8].

Temperature: Temperature affects the growth and development of crops. Each crop has an optimal temperature range for growth. Variations outside this range can affect germination, flowering, and fruiting, potentially leading to reduced yields [10].

Humidity: Humidity levels influence the rate of transpiration, which affects a plant's water and nutrient uptake. High humidity can encourage fungal diseases, while low humidity may lead to water stress in plants [10].

pH: Soil pH is a measure of its acidity or alkalinity. Different crops have specific pH preferences for optimal growth. Deviations from the ideal pH range can affect nutrient availability in the soil and, consequently, plant health [11].

Rainfall: Adequate rainfall is essential for crop irrigation, as it provides the necessary moisture for plant growth. Droughts or excessive rainfall can have adverse effects on crop production, potentially leading to crop failure [12].

However, the development of our intelligent application is tailored to the specific agricultural conditions in Algeria.

To ensure relevance and accuracy, we carefully curated the dataset by selecting only those crops that are well-suited to the environmental conditions prevalent in Algeria. This selection process resulted in a refined dataset featuring eleven crops: maize, chickpea, mungbean, lentil, pomegranate, banana, grapes, watermelon, muskmelon, apple, and orange.

B. Data preprocessing

In this section, we meticulously cleaned the dataset by eliminating duplicate entries, addressing missing data using appropriate imputation techniques, and identifying and treating outliers to enhance data reliability. Categorical data underwent transformation through one-hot encoding to render it compatible with machine learning algorithms, minimizing potential bias. To facilitate model development, we judiciously split the dataset into distinct training and testing sets, allocating 70% for training and 30% for testing. This proportion was carefully chosen to ensure effective model training, robust performance assessment, and thorough evaluation, resulting in a well-structured dataset perfectly primed for analysis and the development of our intelligent application.

C. Model and Performance Evaluation

1) MLP Classifier (Multi-Layer Perceptron Classifier): The MLP Classifier represents a type of artificial neural network employed for classification tasks in machine learning and falls within the broader category of feedforward neural networks. Comprising multiple layers of interconnected nodes, including an input layer, one or more hidden layers, and an output layer [13], the primary function of the MLP Classifier involves learning intricate patterns and relationships within input data to make predictions or classifications based on acquired knowledge. Each node within the neural network operates as a perceptron, conducting weighted summations of its inputs, applying an activation function, and conveying the outcome to the subsequent layer. The hidden layers act as intermediate stages to extract and transform features from the input data [14]. Renowned for their capability to manage non-linear relationships and intricate data structures, MLP Classifiers

prove suitable for a diverse spectrum of classification tasks [15].

2) *Performance Evaluation*: Performance evaluation in machine learning is a critical phase that examines the effectiveness and accuracy of predictive models. Metrics such as accuracy, precision, recall, and F1 Score offer a comprehensive understanding of their predictive capabilities.

Accuracy: Accuracy represents the ratio of correctly predicted outcomes to the total number of predictions generated by the model. It evaluates the consistency of the model's predictions with actual outcomes. Although accuracy is a straightforward metric, it might not be the most suitable choice for datasets with imbalanced distributions [16].

Precision: Precision, also known as positive predictive value, measures the model's ability to accurately identify positive instances among all instances it labels as positive. Precision is particularly valuable in scenarios prioritizing the reduction of false positives, such as in medical diagnosis [16], [17].

Recall: Recall, commonly referred to as sensitivity, quantifies the model's effectiveness in recognizing all relevant instances within the dataset. It represents the ratio of true positives to the total actual positives. Recall is especially relevant when minimizing false negatives is a critical objective, such as in tasks involving the identification of spam emails [17].

F1 Score: The F1 Score is the harmonic mean of precision and recall. This metric provides a balanced evaluation, particularly useful in situations where the dataset exhibits class imbalance. The F1 Score is essential when mitigating both false positives and false negatives concurrently [17].

III. EXPERIMENTAL RESULTS

In this study, we employed a dataset encompassing 22 diverse crop varieties and meticulously narrowed down our focus to 11 crops suitable for cultivation in Algeria's multifaceted agricultural terrain. The dataset comprises seven crucial features, incorporating nutrient composition, temperature, humidity, precipitation, and soil pH levels. To optimize our crop selection model utilizing Multilayer Perceptron (MLP), we identified the most influential features for crop prediction, which specifically include temperature, humidity, rainfall, and pH. Deliberately excluding nutrient-related features from the predictive set is a strategic decision, as it allows farmers greater control over nutrient management. Accordingly, we prioritize environmental factors such as temperature, humidity, rainfall, and pH in our MLP model, recognizing their more direct impact on crop selection and yield optimization.

The MLP model has demonstrated remarkable effectiveness in crop classification within the Algerian agricultural context, as depicted in Fig. 2. Boasting an outstanding accuracy rate of 91.81%, this model furnishes reliable recommendations for crop selection.

Moreover, achieving a precision of 92.25% assures that when the model predicts a crop, it is highly likely to be accurate. The model's recall rate of 91.81% indicates its ability

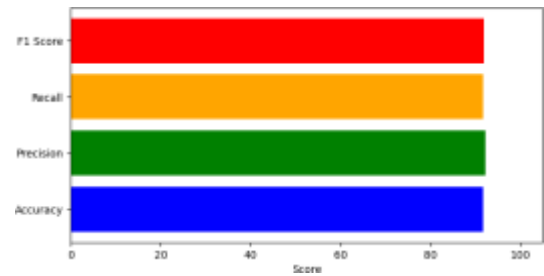


Fig. 2. Crop Classification Metrics Overview.

to identify a significant portion of actual crop varieties, thereby minimizing the risk of missed planting opportunities.

The well-balanced F1 Score, standing at 91.87%, signifies a practical equilibrium between minimizing false predictions and accurately capturing the majority of actual crop classifications.

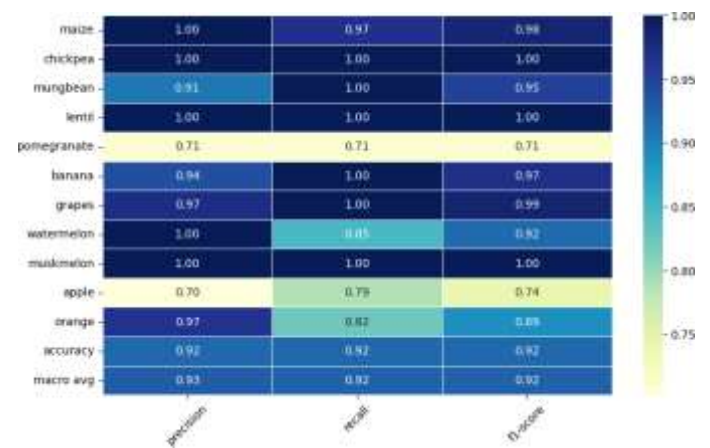


Fig. 3. Model Performance Report.

The performance metrics of the MLP model in classifying different crop types indicate its proficiency in agricultural applications, as depicted in Fig. 3. Particularly noteworthy are the remarkable precision and recall values, achieving an F1-score of 1.00 for crops like Chickpea and Lentil, indicating perfect accuracy. This demonstrates the model's exceptional performance in correctly identifying these specific crops. However, for crops like Grapes, Watermelon, and Oranges, lower precision and recall result in reduced F1-scores. These discrepancies suggest that while the model excels in certain crop classifications, there is room for improvement in accurately distinguishing others.

IV. CONCLUSION AND FUTURE WORKS

In conclusion, the MLP model demonstrates significant potential for crop classification in Algerian agriculture, delivering exceptional accuracy, precision, and recall for various crop types, ensuring reliable crop selection recommendations. This indicates its promising role in optimizing agricultural practices and enhancing crop yield. Looking ahead, there are avenues for improvement. Firstly, focusing on enhancing the

model's capability to distinguish crops with lower precision and recall, like Grapes, Watermelon, and Orange, can be beneficial. Additionally, incorporating more advanced deep learning architectures and expanding the dataset's size and diversity could further augment the model's capabilities, offering substantial benefits to the Algerian farming community.

REFERENCES

- [1] Mancer, M., Terrissa, L., Ayad, S. & Laouz, H. A Blockchain-based approach to securing data in smart agriculture. *2022 International Symposium On INnovative Informatics Of Biskra (ISNIB)*. pp. 1-5 (2022)
- [2] Kathiria, P., Patel, U., Madhwani, S. & Mansuri, C. Smart Crop Recommendation System: A Machine Learning Approach for Precision Agriculture. *Machine Intelligence Techniques For Data Analysis And Signal Processing: Proceedings Of The 4th International Conference MISP 2022, Volume 1*. pp. 841-850 (2023)
- [3] Chauhan, A., Tsunduru, A., Parveen, K., Tokala, S., Hajarathaiah, K. & Enduri, M. A Crop Recommendation System Based on Nutrients and Environmental Factors Using Machine Learning Models and IoT. *2023 International Conference On Information Technology (ICIT)*. pp. 453-458 (2023)
- [4] Gawade, S., Rout, G., Kochar, P., Ahire, V. & Namboodiri, T. AGRO-FERDURE: Intelligent Crop Recommendation System For Agriculture Crop Productivity Using Machine Learning Algorithm. *2023 International Conference On Computer, Electronics & Electrical Engineering & Their Applications (IC2E3)*. pp. 1-9 (2023)
- [5] Desai, M. & Ansari, N. An Innovative Method to Increase Agricultural Productivity using Machine Learning-based Crop Recommendation Systems. *2023 Second International Conference On Augmented Intelligence And Sustainable Systems (ICAISS)*. pp. 645-651 (2023)
- [6] Ramachandra, A., Ankitha, G., Divya, I., Vandana, P. & Jagadeesh, H. Crop Recommendation Using Machine Learning. *2023 International Conference On Data Science And Network Security (ICDSNS)*. pp. 1-5 (2023)
- [7] Dahiphale, D., Shinde, P., Patil, K. & Dahiphale, V. Smart Farming: Crop Recommendation using Machine Learning with Challenges and Future Ideas. (TechRxiv,2023)
- [8] Gerloff, G. & Others Plant efficiencies in the use of nitrogen, phosphorus, and potassium.. *Plant Adaptation To Mineral Stress In Problem Soils. Proceedings Of A Workshop Held At The National Agricultural Library, Beltsville, Maryland, November 22-23, 1976*. pp. 161-173 (1976)
- [9] Mancer, M., Akram, K., Barka, E., Okba, K., Sihem, S., Harous, S., Athamena, B. & Houhamdi, Z. Blockchain Technology for Secure Shared Medical Data. *2022 International Arab Conference On Information Technology (ACIT)*. pp. 1-6 (2022)
- [10] Ferrante, A. & Mariani, L. Agronomic management for enhancing plant tolerance to abiotic stresses: High and low values of temperature, light intensity, and relative humidity. *Horticulturae*. **4**, 21 (2018)
- [11] Neina, D. The role of soil pH in plant nutrition and soil remediation. *Applied And Environmental Soil Science*. **2019** pp. 1-9 (2019)
- [12] Lawson, D. & Rands, S. The effects of rainfall on plant–pollinator interactions. *Arthropod-Plant Interactions*. **13**, 561-569 (2019)
- [13] Hampshire II, J. & Pearlmutter, B. Equivalence proofs for multi-layer perceptron classifiers and the Bayesian discriminant function. *Connectionist Models*. pp. 159-172 (1991)
- [14] Taud, H. & Mas, J. Multilayer perceptron (MLP). *Geomatic Approaches For Modeling Land Change Scenarios*. pp. 451-455 (2018)
- [15] Bisong, E. & Bisong, E. The multilayer perceptron (MLP). *Building Machine Learning And Deep Learning Models On Google Cloud Platform: A Comprehensive Guide For Beginners*. pp. 401-405 (2019)
- [16] Stallings, W. & Gillmore, G. A note on “accuracy” and “precision”. *Journal Of Educational Measurement*. **8**, 127-129 (1971)
- [17] Goutte, C. & Gaussier, E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. *European Conference On Information Retrieval*. pp. 345-359 (2005)

Recommendation System: a review of trust techniques

Sarah MEDJROUD
Department of Computer Science
Hassiba Ben Bouali University
Chlef, Algeria
s.medjrour@univ-chlef.dz

Nassim DENNOUNI
LIA Laboratory, UHBC
Higher School of Management
Tlemcen, Algeria
dennouninas@gmail.com

Mourad LOUKAM
Department of Computer Science
Hassiba Ben Bouali University
Chlef, Algeria
m.loukam@univ-chlef.dz

Abstract— This article highlights the growing application of artificial intelligence (AI) with a particular focus on the use of Implicit Trust-based Recommender Systems (ITRS). These systems leverage trust relationships between users inferred from their past actions such as reviews, check-ins, and clicks. In this article, two types of approaches are discussed. Firstly, explicit trust approaches which aim to improve the accuracy of recommendations by taking into account users' explicit declarations of trust. And secondly, implicit trust approaches are examined, which utilize implicit trust relationships among users inferred from their past behaviors. The overall analysis of these two approaches underscores the benefits of implicit trust in reducing the need for active user participation and alleviating the cold start problem of data sparsity. In conclusion, the article opens the way to new perspectives for item recommendation in the field of smart tourism using AI.

Keywords— Recommendation Systems, Explicit Trust, Implicit Trust, Cold Start, Data Sparsity.

I. INTRODUCTION

Today, the utilization of artificial intelligence is on the rise for analyzing data obtained from the historical records of tourists on smart tourism websites and mobile apps. This analysis aims to extract information about their past locations, preferences, travel habits, and more.

AI-based recommendation systems analyze this data to identify items that are likely to interest a tourist at a given time. This may include restaurants, shops, tourist attractions, etc. Machine learning algorithms are often used to personalize recommendations based on the individual tastes and needs of each tourist.

Implicit Trust-based Recommender Systems (ITRS) focus on using implicit trust relationships among tourists to improve the accuracy of item recommendations. These trust relationships are inferred from tourists' past actions, such as item ratings, check-ins of visited places, clicks, etc. these ITRSs are often based on collaborative filtering approaches to compute preference similarities between tourists.

Implicit data, such as clicks, view times, downloads, are commonly used to infer implicit trust between users. These data are more abundant than explicit ratings (evaluations) and do not require active solicitation from the user. On the other hand, approaches based on probabilistic models, such as the matrix factorization model based on implicit trust, can be used to reduce tourist participation. Recent developments in this area include the use of hybrid systems combining approaches based on implicit trust with collaborative filtering methods of

explicit ratings taking into account the contextual information of each tourist.

In this article, we discussed in the second section, works that use explicit trust in the recommendation of items for users without a demographic profile. Then, we focused on implicit trust in the third section to show its advantage over the explicit declaration of trust relationships. These two trust inference methods can use the indirect relationships between users to improve the effectiveness of recommendations and alleviate the cold start problem due to data sparsity. For this reason, before concluding, we discussed the contribution of these two approaches to the field of tourism, highlighting their potential to use trust propagation to overcome limitations related to data sparsity and tourist reliability. Finally, this state-of-the-art review on the contribution of trust to item recommendation in general, and specific items in particular, allowed us to understand the advantages and limitations of existing approaches in order to propose new perspectives for item recommendation using AI.

II. RECOMMENDER SYSTEMS BASED ON EXPLICIT TRUST

In this section, we will present a state-of-the-art review of works that rely on explicit trust declarations among users in order to classify them according to criteria such as the method used, the data set used, the performance metrics implemented evaluations and the adoption of the principle of trust propagation or not.

Golbeck and Hendler [1] established a website called FilmTrust, which integrates semantic web-based social networks with User Trust (UT) to generate movie recommendations. The ratings associated with this type of trust within this social network are used for calculating similarities within their own FilmTrust platform. Then, a comparison is made between the user's actual rating, the movie's average rating, and the rating suggested by the automatic collaborative filtering (ACF) algorithm. Finally, these predicted ratings are also compared with the nearest neighbor prediction algorithm, which is based on Pearson correlation.

In order to favor trust-based approaches over collaborative filtering (CF) techniques, Massa and Avesani [2] propose an algorithm that diffuses trust across a network to identify users that an active user can trust. This algorithm leverages explicit trust information provided by users of the Epinions.com site by allowing them to indicate the level of trust they attribute to each other. Then, this level of trust is related to how users perceive the relevance of reviews provided by specific users.

Finally, the evaluation of the predictions provided by the proposed algorithm is carried out using metrics such as mean absolute error (MAE), mean absolute error per user (MAUE), and user/rating coverage.

In this same context, Jamali and Ester [3] found that ratings provided by highly trusted friends for articles analogous to a targeted article are more reliable than all other ratings. For this reason, the authors implemented a random walk technique called TrustWalker to merge the recommendation strategies based on trust and on articles to avoid the impact of noisy data by considering a sufficient number of evaluations. Using this technique, the system calculates the trust in its predictions based on the Epinions dataset. To evaluate recommendation errors, the authors combined root mean square error (RMSE) and a coverage metric to calculate FMeasure to compare their approaches with the methods: TidalTrust, MoleTrust, CF Pearson and item-based models.

On the other hand, Jamali and Ester [4] implemented a model-based method named SocialMF designed to recommend content within social networks using matrix factorization techniques. These models incorporate trust propagation mechanisms and use matrix factorization to predict future user ratings. This method uses the Flixster and Epinions datasets and RMSE as the evaluation metric. To measure the effectiveness of this method, Jamali and Ester compare it to three models: the basic matrix factorization approach proposed in [5], user-based collaborative filtering (a memory-based approach) and the STE model from [6]. This comparison shows that the SocialMF method allows a significant reduction in recommendation errors (RMSE), especially in the case of new users.

Guo et al. [7] proposed a method called “Merge” aimed at explicitly integrating trusted neighbors identified by users into recommender systems to alleviate the cold start problem. In their experiments, these authors used three real-world datasets from FilmTrust, Flixster and Epinions to compare their proposed method with different approaches, including TrustAll, a collaborative filtering technique using the Pearson Correlation Coefficient (PCC) measure to calculate the similarity between users, the MoleTrust method [8]. The evaluation of the obtained results was carried out using performance measures of accuracy and coverage of notations such as mean absolute error (MAE). These results demonstrated that the Merge method does not require trust propagation.

TABLE I. RSS BASED ON EXPLICIT TRUST

Author	Method	Dataset	Evaluation metrics	Trust Propagation
FilmTrust Golbeck 2006 [1]	CF memory based	FilmTrust	MAE	Yes
MoleTrust Massa 2009 [2]	TARS ¹	Epinions	MAE MAUE Ratings coverage Users coverage	Yes
TrustWalker Jamali	CF memory	Epinions	RMSE Coverage	No

¹ Trust-Aware Recommender Systems

2009 [3]			FMeasure	
SocialMF Jamali 2010 [4]	CF Model based (MF)	Epinions Flixster	RMSE	Yes
Merge Guo 2012 [7]	TARS	FilmTrust Flixster Epinions	MAE Rating coverage	Yes

III. RECOMMENDER SYSTEMS BASED ON IMPLICIT TRUST

In this section, we will present a state-of-the-art review of works that use implicit trust relationships between users in order to classify them according to criteria such as the method used, the data set used, the evaluation metrics implemented and the adoption of the principle of trust propagation or not.

Pitsilis and Marshall [9] use similarity measures between users according to the Pearson coefficient while integrating their trust relationships. Then, they calculate the predictions using the Resnick formula. This trust approach is asymmetric and requires the calculation of an initial prediction from a MovieLens dataset coming from a Collaborative Filtering system. Finally, to evaluate this approach, an average error percentage is calculated.

In the same context, O’Donovan and Smyth [10] propose a collaborative filtering approach aimed at improving the accuracy of recommendations. They suggest integrating trust mechanisms to refine the selection and weighting of recommendation partners in the recommendation process. They propose several trust calculation models based on past rating behaviors of individual profiles, encompassing both profile-level and item-level. In their experiment, they use the well-known MovieLens dataset and implement Resnick’s prediction methodology. To evaluate the accuracy of their trust-based prediction techniques after computing an initial prediction, they compare the average recommendation error of their approach with Resnick’s standard prediction method.

Papagelis et al. [11] present an approach to mitigate data sparsity using trust inferences. This approach uses the trust established between two users through Pearson Correlation based on ratings. Then, the system propagates this trust using transitive relationships between users within a social network where the experimental data comes from a movie SR called MRS. Finally, this work is evaluated using MAE and Receiver Operating Characteristic (ROC) sensitivity to compare it to collaborative filtering techniques that do not consider transitive associations.

In [12], the authors present an approach that argues that the accuracy of a user’s past predictions plays a crucial role in determining its reliability. This approach improves traditional recommendation techniques by integrating trust into the CF recommendation process. In their work, Hwang and Chen calculate an initial prediction using Resnick’s prediction formula and a trust score from user rating data. Then, they exploit the spread of trust in the trust network to infer indirect relationships and define these two key measures of trust: global trust and local trust. Then, they calculate the PCC and use the Resnick prediction formula for the final rating

predictions. Finally, they calculate the MAE and coverage to compare the recommendation accuracies of different trust configurations with those generated by the standard CF technique.

In [13], the authors proposed an algorithm called "trusted k-nearest recommenders" (kNR) which defines trust as the distance between the user's rating and the recommender's rating, increasing as the gap decreases. The highest level of trust is given to users who rated an item exactly as the user did. In their work, Lathia et al. used metrics such as MAE and root mean square error (RMSE) to compare the performance of their method against two similarity measures: the weighted-Pearson correlation coefficient and the proportion of co-rated items.

Lifen uses conventional similarity measures derived from Collaborative Filtering (CF) to establish potential trust between correlated entities. Trust is expressed as opinions, modeled using subjective logic. The formation of opinions relies on the availability of evidence. Therefore, the degree of trust between entities depends on the perceived similarity of their choices to enable transitivity of trust [14]. The author evaluates the effectiveness of his approach by using the "MovieLens" data set to calculate the divergence between the predicted rating generated by his method and the actual rating provided by the entity. This evaluation is performed by using the percentage of successful estimates obtained to compare the performance of his proposal against the average of all recommendations.

Yuan et al. [15] propose an implicit trust-aware recommendation system, iTARS, with the maximum trust propagation distance of iTARS, improving the conventional eTARS (explicit TARS) model by predicting ratings without explicit trust statements. They use similarities between users to generate implicit trust between users. Recommendations are weighted based on active users' implicit trusts in the recommenders to generate the predicted ratings. They define the implicit trust network as dynamic: a user can join at any time if it has a high similarity with other users and shares a certain number of rated items in common with an existing user of this implicit trust network. They use the Epinions dataset. They examined their proposed iTARS by comparing it to eTARS, based on rating prediction accuracy, rating prediction coverage, and computational complexity.

This paper [16] focuses on how e-governments can support businesses in the problem of selecting a reliable business partner to carry out reliable business transactions. In the process of selecting a business partner, trust or reputation information is crucial and has a significant influence on a business user's decision whether or not to do business with other business entities. Shambour and Lu propose a hybrid trust-enhanced CF recommendation approach (TeCF), which integrates the implicit trust filtering and user-based enhanced CF approaches. This study measures the trustworthiness of a given user based on their history of trusted recommendations. The MovieLens dataset is used to evaluate the performance of the proposed recommendation approach. They use different metrics to evaluate the quality of the recommendations produced, including Resnick's MAE and the coverage metric. They compare the recommendation performance of the proposed implicit trust filtering, enhanced user-based CF, and hybrid TeCF approaches with Resnick's user-based CF approach.

Bedi and Sharma [17] proposed a system that integrates a notion of dynamic trust between users by selecting a restricted and high-quality neighborhood based on the Ant Colony Optimization (ACO) algorithm to generate recommendations. They used data from both the Jester dataset and the MovieLens dataset. To evaluate the effectiveness of their proposed system, they implemented the traditional CF method by calculating the similarity between users using Pearson's correlation coefficient and predicting the ratings using of Resnick's prediction formula. Evaluation metrics such as precision, recall, and F1 measures were used to evaluate their TARS approach.

Shambour and Lu [18] proposed a recommendation system based on Collaborative Filtering that combines trust information and semantic data to address the problems of sparsity and cold start. They evaluate the trustworthiness of a user by measuring the accuracy of the predictions they made as a recommender for the active user by calculating initial predictions, using the mean square difference (MSD) method. The Jaccard measure is used as a weighting scheme to assess the relationship between common ratings and the set of all rated items, thereby calculating the derived implicit trust. Incorporating trust propagation makes it possible to infer trust and establish new ties among users who do not have direct trust ties. Using MovieLens and Yahoo! Webscope R4E like datasets, Shambour and Lu used metrics such as MAE and Coverage to compare their approach with benchmark algorithms like Resnick's user-based CF [19], Sarwar item-based CF [20], O'Donovan-Trust [10] and semantic filtering based on Ruiz-Semantic [21].

In this article, Zhang et al. [22] present a method to solve the sparsity problem by constructing a small-world implicit trust network. The construction of this small-world implicit trust network is based on the grouping of users (user clustering) and the implicit trust relationships between them. They choose the MovieLens 100K dataset as the experimental data. The performance of their proposed ARA algorithm is evaluated using MAE and F-measure metrics. To evaluate the effectiveness of the ARA algorithm, the authors perform a comparative analysis with the following algorithms: (1) CF, a classic user-based collaborative filtering algorithm; (2) O'Donovan, a collaborative filtering algorithm based on the item-level trust model proposed by O'Donovan; and (3) basic MF, a collaborative filtering algorithm based on basic matrix factorization.

In this paper, Shambour and Lu [23] present an effective recommendation approach called Hybrid User-Item Trust-based (HUIT), which integrates implicit trust information of users and items. A key aspect of their proposed approach is the inclusion of user and item reputations. The main goal of the HUIT approach is to improve the quality of recommendations by expanding the neighborhoods of active users and target items. This extension is achieved by integrating alternative information from historical assessments, seamlessly integrating implicit trust details as part of Collaborative Filtering.

To deal with data sparsity and Cold Start user problems, the HUIT approach leverages the intuitive features of implicit trust and trust propagation among users in the user-based implicit trust model, taking into account the reputation of users. The performance evaluation of the HUIT recommendation approach involves experiments conducted on three datasets: MovieLens, Yahoo! Webscope R4 and

FilmTrust. Evaluation metrics, including standard MAE and coverage metrics, are used for comprehensive analysis. The results of the HUIT recommendation approach are systematically compared with those of benchmark recommendation algorithms, including Resnick user-based CF [19], Sarwar item-based CF [20], O'Donovan user-based trust [10], and Kim item-based trust [24].

Roy et al. present a new similarity approach based on implicit trust between users [25]. This approach aims to improve the precision and reliability of predictions by considering similarity as asymmetric in the face of evolving user interests. These authors calculate user trust scores using the psychology forgetting curve, confidence, mean square difference (MSD), and trust between users to create a trust matrix. Finally, to evaluate their method, the authors used the MovieLens dataset and the MAE to measure the accuracy of the recommendations using their proposed method and they compared the quality of the recommendations obtained with those calculated by the different traditional trust-based approaches, such as TFS [18], JMSD [26], O'Donovan-Trust [10] and Resnick-UCF [19].

Zahir et al. [27] proposed an innovative trust-based method called AgreeRelTrust. This method does not require the initial prediction calculation because it merges the positive and negative agreements between users as well as their relative activities to obtain the trust relationship. To evaluate their method using datasets such as GroupLens and MovieLens, Zahir et al. used MAE and RMSE to measure the accuracy of their model's predictions. These measurements made it possible to compare the results of their method with the fundamental kNN approach and the O'Donovan trust-based method.

TABLE II. RSS BASED ON IMPLICIT TRUST

Author	Method	Dataset	Evaluation Metrics	Trust Propagation
Pitsilis 2004 [9]	CF	MovieLens	Mean Error Rate	No
O'donovan 2005 [10]	CF	MovieLens	Mean Error Rate	No
Papagelis 2005 [11]	CF	Movie RS "MRS"	MAE ROC	Yes
Hwang 2007 [12]	CF	MovieLens	MAE Coverage	Yes
Lathia 2008 [13]	CF	MovieLens	MAE RMSE Coverage	No
Lifen 2008 [14]	CF	MovieLens	divergence predicted /real rate	Yes
Yuan 2010 [15]	TARS	Epinions	MAE Coverage	Yes
Shambour 2011 [16]	User-based (CF)	MovieLens	MAE Coverage	Yes
Bedi 2012 [17]	TARS	Jester MovieLens	Precision Recall F-measure	No

Shambour 2012 [18]	CF User / item- based	MovieLens Yahoo! Webscope R4	MAE Coverage	Yes
Zhang 2014 [22]	CF	MovieLens	MAE F-measure	Yes
Shambour 2015 [23]	CF User and item- based	MovieLens Yahoo! Webscope R4 FilmTrust	MAE Coverage	Yes
Roy 2015 [25]	CF	MovieLens	MAE	Yes
Zahir 2019 [27]	Memory based CF	GroupLens MovieLens Jester	RMSE MAE	No

IV. ANALYSIS AND DISCUSSION

The two families of recommendation approaches cited above are based on the use of two types of trust (explicit and implicit) between users and use evaluation metrics such as mean absolute error (MAE), root mean square error (RMSE), and other metrics to estimate the performance of approaches.

The first family of trust approaches focuses on explicit declarations of trust between users and the diffusion of trust across a network while the second family of trust approaches emphasizes the use of implicit trust relationships between users and the propagation of these trust relationships.

In this article, several works on explicit trust between users are cited such as those of Golbeck and Hendler, Massa and Avesani, Jamali and Ester, Guo et al., etc. On the other hand, several works on implicit trust are mentioned such as the works of Pitsilis and Marshall, O'Donovan and Smyth, Papagelis et al., Hwang and Chen, Lathia et al., Yuan et al., Shambour and Lu, Bedi and Sharma, etc.

In the following, Table III compares explicit trust and implicit trust.

TABLE III. COMPARISON BETWEEN EXPLICIT AND IMPLICIT TRUST

Comparison criteria	Explicit trust	Implicit trust
Declaration of trust relationships	Yes	No
Using Direct Trust Propagation	Yes	Yes
Actual trust values (reliability)	Yes	No
Inferred (calculated) trust values	Yes	Yes
Evolution with user behavior	No	Yes
Addressing the problem of data sparsity relating to trust relationships	No	Yes
A considerable computational cost	No	Yes
Mitigating the cold start problem	No	Yes

In summary, this article places more emphasis on implicit trust relationships and the contribution of trust propagation compared to explicit trust declarations and their diffusions.

V. CONCLUSION

In conclusion, this article explores the growing application of artificial intelligence (AI) in the field of smart tourism, highlighting the use of Implicit Trust-based Recommender Systems (ITRS) to improve the relevance of item recommendations for tourists. In this article, the focus is on ITRSs, which leverage implicit trust relationships among tourists, inferred from their past actions such as reviews, check-ins, and clicks. For this reason, the article contains two distinct sections. The first looks at work using explicit trust, while the second focuses on implicit trust. The analysis highlights the benefits of implicit trust as it reduces the need for active user participation and alleviates the cold start problem of data sparsity. In summary, this state-of-the-art provides a comprehensive understanding of the advantages and limitations of existing approaches, opening the way to new perspectives for item recommendation in the context of smart tourism using AI.

ACKNOWLEDGMENT

This work has been supported by project PRFU N° COOLO7UN020120220001 entitled “Formalization of points of interest recommendation systems and evaluation of their performance for mobile users”, Department of Computer Science, Hassiba Ben Bouali University, Chlef, Algeria.

REFERENCES

- [1] Golbeck, J., & Hendler, J. A. (2006, January). FilmTrust: movie recommendations using trust in web-based social networks. In CCNC (Vol. 2006, pp. 282-286).
- [2] Massa, P., & Avesani, P. (2009). Trust metrics in recommender systems. In *Computing with social trust* (pp. 259-285). London: Springer London.
- [3] Jamali, M., & Ester, M. (2009, June). Trustwalker: a random walk model for combining trust-based and item-based recommendation. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 397-406).
- [4] Jamali, M., & Ester, M. (2010, September). A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 135-142).
- [5] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In *NIPS 2008*, volume 20.
- [6] H. Ma, I. King, and M. R. Lyu. Learning to recommend with social trust ensemble. In *SIGIR 2009*, pages 203–210.
- [7] Guo, G., Zhang, J., & Thalmann, D. (2012). A simple but effective method to incorporate trusted neighbors in recommender systems. In *User Modeling, Adaptation, and Personalization: 20th International Conference, UMAP 2012, Montreal, Canada, July 16-20, 2012. Proceedings 20* (pp. 114-125). Springer Berlin Heidelberg.

- [8] Massa, P., Avesani, P.: Trust-aware recommender systems. In: *Proceedings of the ACM Conference on Recommender Systems*, pp. 17–24 (2007).
- [9] G. Pitsilis and L. Marshall. A model of trust derivation from evidence for use in recommendation systems. Technical Report. University of Newcastle upon Tyne, 2004.
- [10] J. O'Donovan and B. Smyth. Trust in recommender systems. In *Proceedings of the 10th International Conference on Intelligent User Interfaces (IUI)*, 2005.
- [11] Papagelis, M.; Plexousakis, D.; Kutsuras, T. Alleviating the sparsity problem of collaborative filtering using trust inferences. *International conference on trust management*. Springer, 2005, pp. 224–239.
- [12] C.-S. Hwang and Y.-P. Chen. Using trust in collaborative filtering recommendation. In *New Trends in Applied Artificial Intelligence*. 2007.
- [13] N. Lathia, S. Hailes, and L. Capra. Trust-based collaborative filtering. In *Trust Management II*, 2008.
- [14] Lifan, L. (2008, December). Trust derivation and transitivity in a recommendation trust model. In *2008 International Conference on Computer Science and Software Engineering (Vol. 3, pp. 770-773)*. IEEE.
- [15] Yuan, W.; Shu, L.; Chao, H.C.; Guan, D.; Lee, Y.K.; Lee, S. ITARS: trust-aware recommender system using implicit trust networks. *IET communications* 2010, 4, 1709–1721.
- [16] Shambour, Q., & Lu, J. (2011). A hybrid trust-enhanced collaborative filtering recommendation approach for personalized government-to-business e-services. *International Journal of Intelligent Systems*, 26(9), 814-843.
- [17] Bedi, P., & Sharma, R. (2012). Trust based recommender system using ant colony for trust computation. *Expert Systems with Applications*, 39(1), 1183-1190.
- [18] Q. Shambour and J. Lu. A trust-semantic fusion-based recommendation approach for e-business applications. *Decision Support Systems*, 2012, 54:768–780.
- [19] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl, GroupLens: an open architecture for collaborative filtering of netnews, in: *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, 1994, pp. 175–186.
- [20] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, Item-based collaborative filtering recommendation algorithms, in: *Proceedings of the 10th International Conference on World Wide Web, WWW, Hong Kong, China, ACM Press*, 2001, pp.285–295.
- [21] M. Ruiz-Montiel, J. Aldana-Montes, Semantically enhanced recommender systems, in: R. Meersman, P. Herrero, T. Dillon (Eds.), *On the Move to Meaningful Internet Systems: OTM 2009 Workshops*, Springer, Heidelberg, 2009, pp. 604–609.
- [22] Zhang, F., Wang, H., & Yi, H. (2014). An Adaptive Recommendation Method Based on Small-World Implicit Trust Network. *J. Comput.*, 9(3), 618-625.
- [23] Shambour, Q., & Lu, J. (2015). An effective recommender system by unifying user and item trust information for B2B applications. *Journal of Computer and System Sciences*, 81(7), 1110-1126.
- [24] H.-N. Kim, A.-T. Ji, G.-S. Jo, Enhanced prediction algorithm for item-based collaborative filtering recommendation, in: K. Bauknecht, B. Pröll, H. Werthner (Eds.), *E-Commerce and Web Technologies*, in: *Lecture Notes in Computer Science*, vol.4082, Springer, Heidelberg, 2006, pp.41–50.
- [25] Roy, F., Sarwar, S. M., & Hasan, M. (2015). User similarity computation for collaborative filtering using dynamic implicit trust. In *Analysis of Images, Social Networks and Texts: 4th International Conference, AIST 2015, Yekaterinburg, Russia, April 9–11, 2015, Revised Selected Papers 4* (pp. 224-235). Springer International Publishing.
- [26] Bobadilla, J., Serradilla, F., Bernal, J.: A new collaborative filtering metric that improves the behavior of recommender systems. *Knowl.-Based Syst.* 23(6), 520–528 (2010).
- [27] Zahir, A., Yuan, Y., & Moniz, K. (2019). AgreeRelTrust—a simple implicit trust inference model for memory-based collaborative filtering recommendation systems. *Electronics*, 8(4), 427.

Biometric authentication using invariant Gray Level Co-occurrence Matrices (GLCM)

Mokhtari Aicha
Biomedical Engineering Department
Faculty of Technology
University of Abou Bekr Belkaid
Tlemcen 13-000, Algeria
aicha.mokhtari@univ-tlemcen.dz

Hadj Slimane Zine Eddine
Biomedical Engineering Department
Faculty of Technology
University of Abou Bekr Belkaid
Tlemcen 13-000, Algeria
zineeddine.hadjslimane@univ-tlemcen.dz

Abstract— Biometric authentication plays a crucial role in ensuring security, access control and identity verification, with applications in various sectors and contexts. Biometric authentication in the context of access control is a resilient security technique employed to oversee and manage entry to physical locations, digital platforms, or valuable assets. It verifies the identities of individuals by assessing their distinct biological or behavioral attributes. This approach ensures that it enhances security and accountability by granting access only to authorized individuals. This article introduces an authentication method centered on retinal images, enhancing security levels through the utilization of the Invariant Gray Level Co-occurrence Matrices (GLCM) technique for feature extraction and employing a Random Forest classifier for classification. Our evaluation, conducted on the RDB database, yielded highly favorable results, achieving a flawless accuracy of 100%.

Keywords—Authentication, retina, GLCM, Random Forest.

I. INTRODUCTION:

Advances in information and communication technologies (ICT) open wider doors to the accessibility of information and knowledge, resulting in a significant transformation of the institutional framework of society. Digital platforms, services and technologies have transformed societies into interconnected ecosystems, facilitating monitoring, data collection and analysis. Globalization and the rapid pace of change accelerate this development, affecting various aspects of public life and disrupting traditional norms.

Innovation is exponentially growing and driving an increase in the speed of technological advancements. These advancements include new technologies like the Internet of Things, artificial intelligence, biometric authentication, and other digital technologies.

Biometric authentication is an access control and management system.

This system ensures security and controls the entry and exit of premises or territory.

Biometric authentication involves the verification or identification of individuals through their distinct physical or behavioral attributes. This method ensures secure access control and identity verification by leveraging unique biological traits that are challenging to duplicate or counterfeit.

These systems analyze and compare specific features such as fingerprints, iris or retinal patterns, facial characteristics, voiceprints, or behavioral traits like typing patterns or gait. The collected biometric data is securely stored in a database or on a device for future reference.

When someone seeks access to a system, facility, or device, the system compares their biometric information to the stored data to verify their identity. If the biometric traits match the stored template, the authentication is successful, granting access. Conversely, if there is a significant mismatch, the system denies access. Figure 1 illustrates biometric system authentication

These systems find applications in various industries and contexts, including physical access control to buildings or restricted areas, unlocking devices like smartphones or laptops, authorizing financial transactions, and ensuring secure identification for government or law enforcement purposes.

Among the biometric modalities, retinal recognition is a very accurate and reliable method of biometric identification. In the current state of the art, the field of biometric retina has witnessed significant advancements. Many studies have focused on the use of retinal vascular network as biometric keys. For instance, Sadeghpour et al. [1] introduce and assess

a cohort-based dissimilarity vector representation for safeguarding retinal vascular templates in biometric systems, emphasizing irreversibility and unlinkability of the protected reference templates. The proposed approach achieved a mean AUC of over 98% in classifying cohort-based dissimilarity vectors, with a maximum accuracy drop of 1-2% compared to unprotected templates across three datasets: ESRID, VARIA, and Messidor-2. To improve existing methods for identifying retinal images, Noori. A [2] proposes an innovative approach for evaluating such images. The suggested method relies on the histogram of Hue matrix values

in the HSC color format, along with the LBP and HOG algorithms. According to the DRIVE database, the mean precision of the proposed method is 98.5%. Betaouaf et al. [3] have developed an automated identity verification algorithm based on the complex vascular network structure in the human retina. Their approach involves using watershed transformation for retinal vasculature segmentation and a hit or miss transform for detecting bifurcation and intersection points. Additionally, they utilize automatic image registration and compare biometric data models through the Iterative Closest Point (ICP) algorithm.

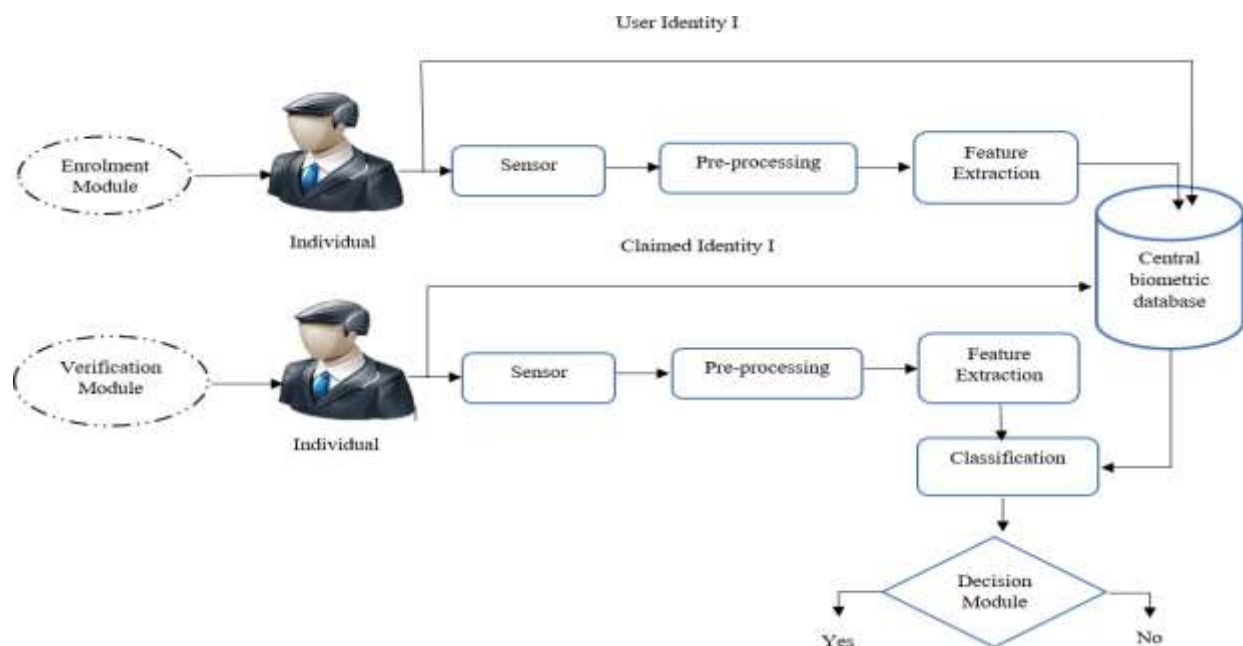


Fig. 1. Biometric system authentication.

Vignan's Nirula and co-authors [4] have developed a biometric verification system for personal identification on mobile devices. The system uses adaptive histogram equalization, Gaussian filtering, top-hat transformation, binary morphological reconstruction, and the KNNRF classifier. The system achieves a 60% false acceptance rate, 90% false rejection rate, and 94% accuracy on the Drive database.

Saba A. Tuama and Loay E. George present a distinctive method centered on retinal vascular patterns and personal recognition [5]. This innovative technique employs the Euclidean distance metric during the matching phase and leverages the spatial distribution of local vascular density as a feature vector. The system evaluates its efficacy using two databases, DRIVE and STARE, and it reveals a remarkable 100% accuracy rate for both datasets.

Jarina B. and her team [6] present a three-step procedure for individual identification utilizing color retinal images: initial vessel extraction through parabolic model fitting, subsequent feature extraction, and matching

based on Euclidean distance. The performance of their proposed method is comprehensively assessed using a variety of pathological images, a local hospital database, two authentication databases (RIDB and VARIA), and three standard databases (DRIVE, HRF, and Messidor).

Biometric retina technology extends beyond vascular patterns to encompass other unique features of the retina for identification and verification purposes. This approach in retinal biometrics focuses on aspects other than blood vessels to establish identity. By analyzing aspects of the retina beyond blood vessels, including the overall retinal texture. Our research methodology comprises three distinct phases:

1. In the first stage, we employ CLAHE to preprocess the data.
2. Following the data preprocessing, the next step involves the extraction of unique features using the Gray-Level Co-occurrence Matrix (GLCM) analysis with invariance properties.

3. Subsequently, the features extracted in the previous phase undergo a classification process, utilizing a Random Forest algorithm.

The outcomes of our investigation demonstrate an exceptional accuracy rate of 100% during rigorous testing conducted on the RIDB (Retinal Identification Data Base) database. This paper comprises three principal sections: the methodology, where we elaborate on our approach; the experimental results and subsequent discussion; and, finally, a concluding section that encapsulates our findings and insights.

II. METHODOLOGY:

Within this section, we outline our methodology, commencing with image preprocessing, followed by the extraction of characteristics using GLCM, and concluding with the classification process.

A. Preprocessing :

In this step, we convert the input image to gray level and apply adaptive histogram equalization to eliminate contrast and brightness problems Fig.2 present results of this step.

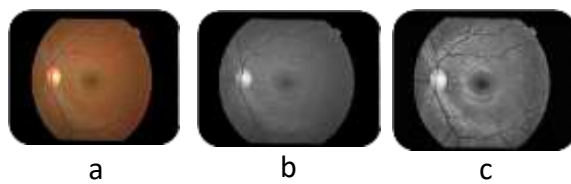


Fig.2.Preprocessing step: a: input image, b: gray image, c: image preprocessed.

B. Feature extraction:

The Gray-Level Co-occurrence Matrix (GLCM) is a widely used method for texture feature extraction in image processing. It characterizes the spatial relationship of pixel intensity values in an image by computing the occurrence of pairs of pixel values at various pixel displacements or offsets.

Gray-level invariant Haralick features constitute a collection of statistical metrics employed to assess the textural characteristics of an image. These features hold substantial significance in the domains of image analysis and computer vision, finding extensive utility in diverse applications such as image categorization, segmentation, and the comprehensive analysis of textures.

When computing the GLCM, one can incorporate various measures or statistics derived from the matrix to extract features. These features are commonly used to describe the texture properties of an image. However, the GLCM itself is invariant to certain transformations such as rotation, translation, or scaling [7]. To address the lack of

invariance, we employ in combination with the GLCM based features. A combination of these approaches can significantly enhance the invariance of GLCM based features and improve the robustness of texture analysis in image processing tasks.

The texture features computed from GLCMs are:

1. Autocorrelation: Measures the similarity between an image and a shifted version of itself.
2. Contrast: Gauges the variation in intensity between adjacent pixels in the image.
3. Correlation: Measures the linear dependency between the intensity values of neighboring pixels.
4. Difference Average: Calculates the mean absolute difference among pixel intensities in the image.
5. Difference Entropy: Measures the randomness or disorder of the differences between pixel intensities.
6. Difference Variance: Measures the variance of the differences between pixel intensities.
7. Dissimilarity: Determines the mean absolute difference between pixel intensities.
8. Energy: Also referred to as Angular Second Moment, quantifies the uniformity or smoothness of the image.
9. Entropy: Measures the randomness or disorder of pixel intensities in the image.
10. Homogeneity (Inverse Difference Moment): Assesses the proximity of the element distribution in the GLCM to the diagonal.
11. Maximum Correlation Coefficient: Measures the maximum correlation coefficient between intensity values at different pixel distances and orientations.
12. Maximum Probability: Measures the highest probability of occurrence of pixel pairs in the GLCM.
13. Sum Average: Measures the average sum of the pixel pairs at different distances and orientations.
14. Sum Entropy: Measures the randomness or disorder of the sum of pixel pairs at different distances and orientations.
15. Sum of Squares: Measures the sum of squared elements in the GLCM.
16. Sum Variance: Measures the variance of the sum of pixel pairs at different distances and orientations.

C. Classification :

This step involves feature matching. Initially, we collected various variables to train the random forest model for classification. Furthermore, we assigned a unique identifier (ID) to each individual. In a binary classification approach, label chosen ID with the value 1, while assigning zero to other IDs. This method makes it

possible to predict and distinguish the chosen ID from other identifiers.

A random forest serves as a meta-estimator that trains multiple decision tree classifiers on different subsets of the dataset and employs averaging to enhance predictive accuracy while mitigating the risk of over fitting. It is a popular choice in many machine-learning applications due to its effectiveness and ease of use. Through experimentation and thorough evaluation on our specific dataset, we choose a number of trees to 100.

III. RESULTS AND DISCUSSION:

To verify the effectiveness of our algorithm, we utilize a publicly dataset called RIDB [8]. The Retina Identification Database (RIDB) consists of Retinal Fundus Images taken using the TOPCON-TRC camera. It comprises 100 images, each with a resolution of 1504 x 1000 pixels, stored in JPEG format. These images were captured from 20 individuals, with five samples per person, and none of the individuals had any retinal diseases. We employed RIDB for training and testing purposes in developing a retinal recognition system. retinal recognition is one of the most precise biometric methods used to authenticate and identify individuals.. We carried out all tests using MATLAB 2016 on a personal computer with an Intel Celeron CPU N3050.

I based the evaluation of the system's performance on the following performance criteria:

1) False Acceptance Rate (FAR):

This probability refers to the likelihood of wrongly identifying an unknown user as a known user. This rate is essential to evaluate the security of the biometric system. Its calculation is as follows:

$$FAR = \frac{\text{the number of false acceptances}}{\text{he number of identification attempts}} \times 100$$

2) False rejection rate (FFR):

The probability that a known user will be rejected by the biometric system. This rate is the probability that a known user will be rejected by the biometric system. This rate helps to evaluate the level of comfort of use of the biometric system. Its formula is as follows:

$$FRR = \frac{\text{the number of false recognitions}}{\text{the number of identification attempts}} \times 100$$

3) Equal Error Rate (EER) :

overall accuracy and usability of a biometric authentication system.

4) The ROC curve:

Serves as a valuable tool to visualize and evaluate the effectiveness of binary classification models. It enables informed decision making regarding the balance between correctly identifying true positives and the occurrence of false positives, achieved through the adjustment of the classification threshold.

In this part, we present the results for two identifiers 2 and 10 (we choose the IDs numbers) as an example of classification. This method makes it possible to increase the precision. Fig.3 and Fig. 4 shows the ROC curve, false acceptance rate and false rejection rate of the proposed system using the RIDB database. It shows a good separation distance between the genuine and imposter classes.

Additionally, we calculate the average accuracy across various IDs and proceed to evaluate the performance of our proposed method against established state-of-the-art techniques. The table presented below presents both the outcomes of our suggested approach and those of existing methods. This table provides a summary of the release year for several approaches and their respective performances on the RIDB dataset. Our findings indicate that our method has the highest accuracy, along with a substantial confidence interval.

Table: Result of our method and other methods.

Method	application	Dataset	Accuracy (%)
Our method	authentication	RIDB	100
[3]	authentication	DRIVE	100
[9]	identification	DRIVE	97.5
[10]	identification	DRIVE, RIDB	100
[11]	authentication	DRIVE	0.999
[12]	authentication	VARIA	97.16

IV. CONCLUSION

In our pursuit of developing trustworthy digital retinal templates for identity verification, we concentrated on identifying unique and dependable retinal characteristics.

We focused on textural features extracting by invariant GLCM. To validate the efficacy of our proposed method, we subjected it to rigorous testing using the RIDB

Represents the point at which the False Acceptance Rate (FAR) and False Rejection Rate (FRR) are equal. In other words, the Equal Error Rate is the threshold or operating point where the system's performance reaches a balance between incorrectly accepting impostors and incorrectly rejecting legitimate users. It is a critical metric for assessing the dataset.

The results were exceptionally impressive, achieving a remarkable accuracy rate of 100%. This accomplishment underscores the resilience and

reliability of our approach in accurately confirming individuals' identities based on their retinal biometrics.

By harnessing these distinctive retinal features, we have paved the way for the creation of highly reliable digital retinal templates. This breakthrough carries significant potential for improving identity verification systems across various domains, including security applications.

Leveraging such unique retinal characteristics empowers us to construct precise and secure identity authentication systems, instilling confidence in the realm of biometric authentication.

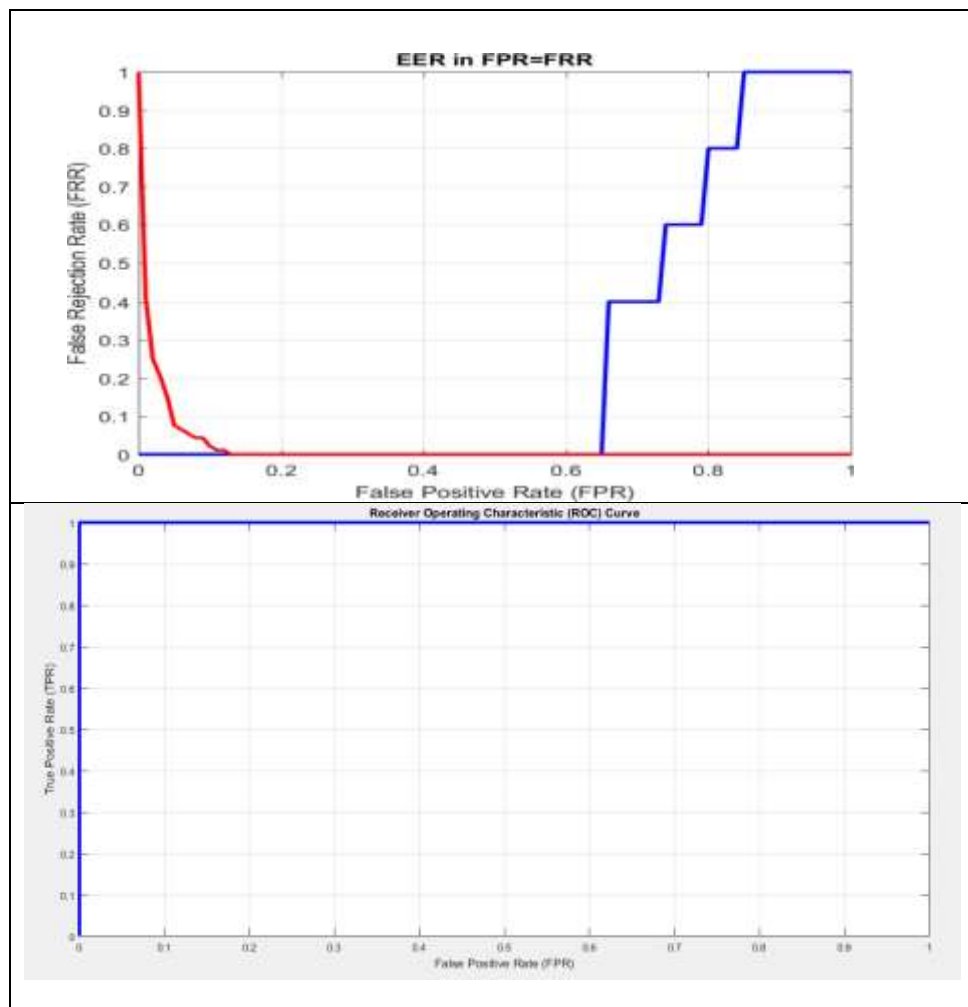


Fig.3. the ROC curve, false acceptance rate, and false rejection rate of the proposed system when employing the RIDB database for ID 2.

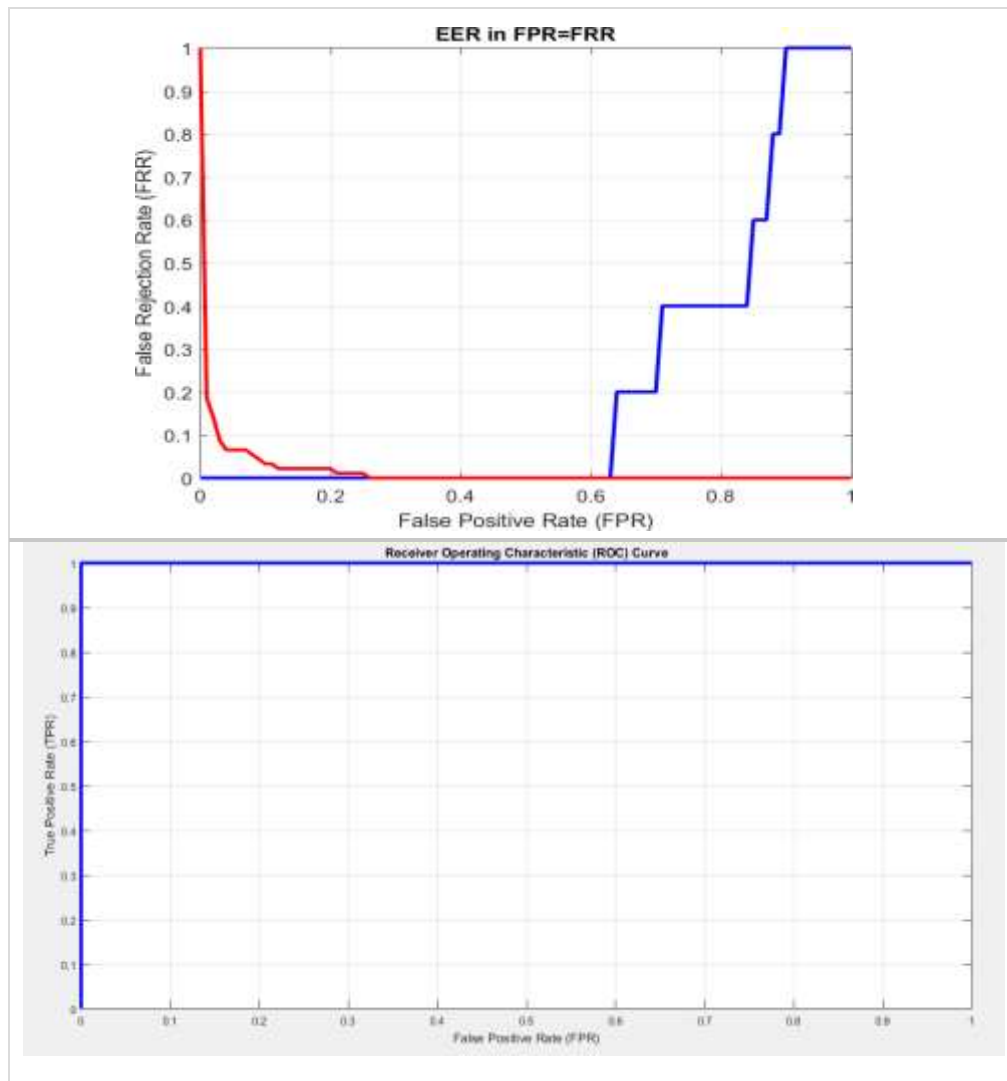


Fig.4. the ROC curve, false acceptance rate, and false rejection rate of the proposed system when employing the RIDB database for ID 10.

REFERENCES

- [1] M. Sadeghpour, A. Arakala, S. A. Davis and K. J. Horadam, "Protection of Sparse Retinal Templates Using Cohort-Based Dissimilarity Vectors," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 5, no. 2, pp. 233-243, April 2023, doi: 10.1109/TBIOM.2023.3239866.
- [2] A Noori, "Retinal Vessels Combining LBP and HOG". arXiv preprint arXiv:2206.01658, Computer Science - Computer Vision and Pattern Recognition, 2022, 10.48550/arXiv.2206.01658
- [3] TH Betaouaf, E Decenci re, A Bessaid. "Automatic biometric verification algorithm based on the bifurcation points and crossovers of the retinal vasculature branches". International Journal of Biomedical Engineering and Technology (2020), 32(1), 66-82, doi: 10.1504/IJBET.2020.104677.
- [4] BMS Rani, AJ Rani. « Biometric retinal security system for user identification and authentication in smartphones". International Journal of Pure and Applied Mathematics (2018), 119(14), 187-202, doi: 10.21506/j.ponte.2018.2.3.
- [5] SA Tuama, LE George. "Automatic Human Recognition Based on the Geometry of Retinal Blood Vessels Network", DOI: 10.9734/bpi/crst/v1.
- [6] J Mazumdar, SR Nirmala. "Person identification using parabolic model-based algorithm in color retinal images", International Journal of Electrical and Electronic Engineering & Telecommunications (2019), doi: 10.18178/ijeetc.8.6.358-366.
- [7] T L fstedt, P Brynolfsson, T Asklund, T Nyholm, A Garpebring, "Gray-level invariant Haralick texture features", PloS one, 2019, vol. 14, no 2, p. e0212110, <https://doi.org/10.1371/journal.pone.0212110>
- [8] "RIDB: a dataset of fundus images for retina based person identification", Data in Brief, vol.33, p.106433, 2020, Elsevier, <https://doi.org/10.1016/j.dib.2020.106433>.
- [9] F Sadikoglu, S Uzelaltinbulat, "Biometric Retina Identification Based on Neural Network" Procedia Computer Science Volume 102, 2016, Pages 26-33 Elsevier, <https://doi.org/10.1016/j.procs.2016.09.365>.
- [10] S Sultan, M Faris Ghanim, "Human Retina Based Identification System Using Gabor Filters and GDA Technique", JOURNAL OF COMMUNICATIONS SOFTWARE AND SYSTEMS, VOL. 16, NO. 3, SEPTEMBER 2020, (DOI: 10.24138/jcomss.v16i3.1031
- [11] MSA Malik, Q Zahra, IU Khan, M Awais, G Qiao, "An Efficient Retinal Vessel. Journal of Medical Imaging and Health Informatics, 10(10), 2481-2489, DOI: <https://doi.org/10.1166/jmihi.2020.3180>.
- [12] R. Manjula Devi, P. Keerthika, P. Suresh, Partha Pratim Sarangi, M. Sangeetha, C. Sagana, K. Devendran, Chapter 5 - Retina biometrics for personal authentication, Machine Learning for Biometrics, Academic Press, 2022, p87-104, ISBN 9780323852098, <https://doi.org/10.1016/B978-0-323-85209-8.00005-5>.

AI-based pathogenicity classification of flea and tick-borne diseases and seasonal distribution of some species in Algeria

Nouredine Rabah sidhoum¹, Mehdi Boucheikhchoukh¹, Nouredine Mechouk^{2,3}

1. Biodiversity and Ecosystems Pollution Laboratory, Faculty of Life and Nature Sciences, Chadli Bendjedid University, El Tarf 36000, Algeria.

2. Ecology of Terrestrial and Aquatics Systems Laboratory (EcoSTaQ), Department of Biology, Faculty of Science,

Badji Mokhtar University, Annaba 23200, Algeria.

3. Department of Parasitology and Parasitic Diseases, Faculty of Veterinary Medicine, University of Agricultural Sciences and Veterinary Medicine of Cluj-Napoca, Calea Mănăştur 3-5, Cluj-Napoca 400372, Romania.

Abstract—Vector-borne diseases and their classification are paramount epidemiological importance since they are an important cause of death or debilitating situations, especially in African countries. In this study, a database of vector-borne diseases was generated out of Algerian data to train two different classifiers and compare the fitness of each one. In addition, 1037 fleas and ticks were sampled in different regions of the central north of the country to understand the seasonal activity of these vectors and their implication in transmitting diseases to humans and animals.

Keywords—ANN; epidemiology; Arthropods; north-Africa; zoonoses

I. INTRODUCTION

In North Africa, the coexistence of various arthropod-borne pathogens poses a substantial public health challenge [1], [2]. Ticks and fleas transmit these pathogens, encompassing a spectrum of diseases with diverse clinical manifestations. Among these are TG group *Rickettsia felis* and *R. typhi*, bartonellosis, *Dipylidium caninum*, and plague, predominantly transmitted by fleas [3], [4]. In contrast, ticks are vectors cable of transmitting SFG group rickettsias, certain bartonellosis, and filarial helminths [5], [6]. Understanding the dynamics of these pathogens is critical for developing effective preventive measures and interventions.

The prevalence and distribution of these arthropod-borne pathogens vary across the region, and several factors influence their transmission [7]. These factors encompass a complex interplay between humans, animals, and arthropod vectors. Thus, it is imperative to employ advanced tools and methodologies to categorize these pathogens, shedding light on their spatial distribution, host specificity, and the competence of arthropod vectors.

This study seeks to employ artificial neural network models to classify the pathogens mentioned above, focusing on four fundamental aspects: reproduction rate, prevalence, vector/host presence, and vectorial competence. Prevalence is a critical indicator of the risk posed by a particular pathogen in a given region. By discerning the prevalence detected in humans, animals, and arthropods, we can gain insights into the potential burden of these diseases on public health and animal populations. Vector presence, which pertains to the geographic

distribution of arthropod vectors and their vectorial competence, plays a pivotal role in transmitting these pathogens. In summary, this research contributes to our knowledge of tick and flea-borne pathogens in North Africa, enhancing our capacity to mitigate the impact of these diseases on human and animal populations. Through the innovative application of artificial neural network models, we aspire to offer a more nuanced classification that integrates factors critical to disease prevalence, vector presence, and vectorial competence, ultimately advancing our understanding of the transmission of these pathogens in the region.

II. MATERIALS AND METHODS

A. Study area and ticks and fleas sampling

Between June 2022 and July 2023, a sampling was conducted in four provinces of the central north of Algeria: Algiers, Boumerdes, Blida, and Tizi Ouzou.

The sampling targeted domestic and peridomestic, farms and stray and wild animals. The localization of visited counties during the fieldwork is illustrated in Figure 1.

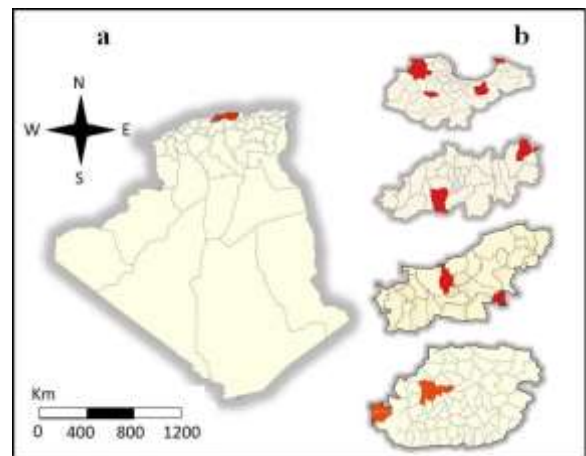


Figure 1: (a) Study area (Algiers, Blida, Boumerdes and Tizi-Ouzou) (b) Visited counties (Aïn Benian, Cheraga, Bab Ezzouar, Bab Hassen, Oued Smar, El Marsa, Bouarfa, Meftah, Zemouri, Timezirt, M'kira, Tizi Ghennif, and Tizi Ouzou).

The arthropods were conserved in 70° ethanol micro-tubes at T=-12C°. The identification of ectoparasites was established according

to taxonomic keys [8], [9] the fieldwork was completed by data analysis using artificial neural networks based on published data from the country.

B. Artificial neural network classifiers

In order to generate the dataset for artificial neural network analysis a systematic review was carried out following the guidelines of Preferred Reporting Elements for Systematic Reviews and Meta-analysis (PRISMA). The focus was on references related to flea and tick-borne pathogens in Algeria. The search, conducted up to November 2023, targeted materials published online in English and French, including peer-reviewed publications, theses, DVM and M'sdissertations, local authorities reports, and published communications.

In addition to data regarding the reported prevalence in humans, animals, and arthropods, the presence or absence of the arthropod vector (fleas and ticks), the geographical location (urban, rural, forest, or desert), the competence of the arthropod in transmitting the disease and the type of animal species (domestic, peridomestic, feral, synanthropic, or wild) were noted.

Seven covariates were considered, with the pathogenicity as a dependent variable.

Each covariate was attributed a score based on an appropriate scale.

The pathogenicity score was determined using classifications from NIH, APSA, ePATHogen, and German scales, ranging from (1) for non-pathogenic microorganisms to (4) for highly pathogenic and communicable ones. Scores for pathogenic and highly pathogenic organisms were (2) and (3), respectively. A binary score (0/1) denoted the absence/presence of host generalist (or host-specific) fleas. Host behavior was assessed based on lifestyle and contact with humans, with scores ranging from (1) for wild animals to (5) for pets. Locations were individually examined using coordinates from consulted studies and reports, employing Google Earth (Version 7.3.6) and the United Nations DESA database (2018 revision) to estimate scores related to urbanization and population density. The lowest scores (1) and (2) were assigned to desertic and forest lands, while suburban and rural areas received a score of (3), and urban areas were assigned the highest score (4). Prevalences of detected pathogens in humans, animals, and fleas were extracted or calculated based on available data and incorporated into the database (Supplementary Material).

Statistical analyses, including multilayer perceptron and radial basis function neural networks, were conducted using IBM-SPSS version 26.0.0 (IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp).

Information regarding data analysis can be found in the **supplementary material 1**.

III. RESULTS

A. Sampling outcomes

During the sampling, 162 animals were examined. Among them, 126 (77.78%) were infested by at least one ectoparasite species. A total number of 1037 ticks and fleas was recovered from infested animals. Details about ectoparasite species with their relative count are depicted in Figure 2.

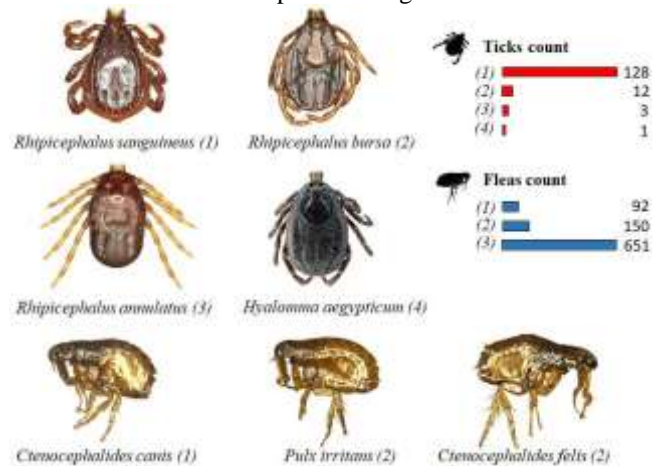


Figure 2: Identified flea and tick species with the number of collected specimens.

Figure 3 shows the bimestrial chart of the ectoparasites in terms of their numbers.

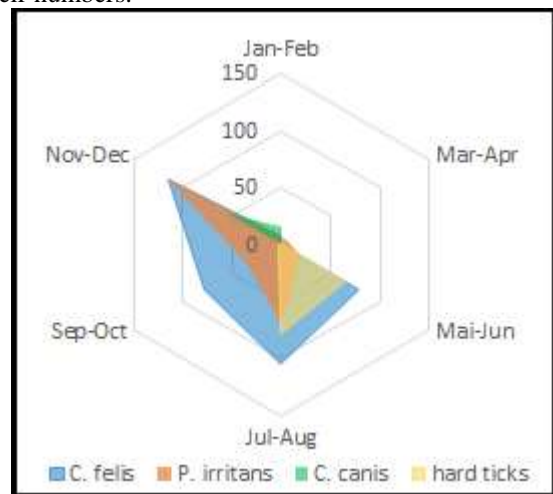


Figure 3: Radar chart of recovered fleas and ticks seasonal activity

B. Pathogens classifying models

Tick-borne pathogens that were previously detected in Algeria and considered in this study are the following: *Anaplasma bovis*, *A. centrale*, *A. marginale*, *A. ovis*, *A. phagocytophilum*, *A. platys*, *Babesia bigemina*, *B. bovis*, *B. ovis*, *B. vogeli*, *Borrelia burgdorferi*, *B. hispanica*, *B. theileri*, *B. turicatae*, *Bartonella bovis*, *B. dromedarii*, *B. tamiae*, *Coxiella brunetti*, *Ehrlichia canis*, *Hepatozoon canis*, *Mycoplasma wenyonii*, *Rickettsia aeschlimannii*, *R. africae*, *R. conorii*, *R. helvetica*, *R. massilae*, *R. monacensis*, *R. raoulti*, *R.*

sibirica, *R. slovaca*, *Theileria annulata*, *T. equi*, *T. orientalis*, and *T. ovis*.

Several flea-borne pathogens have also been reported in Algeria, including *Bartonella claridgeiae*, *B. elizabethae*, *B. henslae*, *B. tribocorum*, *B.v. berkhoffi*, *Dipylidium caninum*, *Rickettsia felis*, *R. typhi*, and *Yersinia pestis*.

The training and testing results of each models are as follows:

Table 1 represents the cross entropy error of training and testing as well as percentage of incorrect predictions.

TABLE I. MULTILAYER PERCEPTRON SUMMARY

Training	Cross Entropy Error	83,274
	Percent Incorrect Predictions	28,4%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error
	Training Time	0:00:00,03
Testing	Cross Entropy Error	28,098
	Percent Incorrect Predictions	21,3%

Figure 4 represents the ROC (Receiving Operating Characteristic) curve of the multilayer perceptron model.

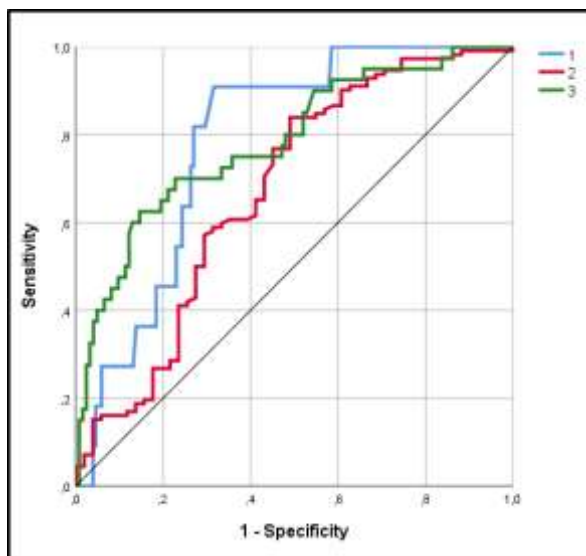


Figure 4: performance of the multilayer perceptron model in classifying the pathogenicity of tick and flea-borne pathogens in Algeria according to published data.

Table 2 displays the sum of squares error and the percentage of incorrect predictions in the training/testing phase for the radial basis function model.

TABLE II. RADIAL BASIS FUNCTION SUMMARY

Training	Sum of Squares Error	19,151
	Percent Incorrect Predictions	22,5%
	Training Time	0:00:00,18
Testing	Sum of Squares Error	11,014 ^a
	Percent Incorrect Predictions	30,8%

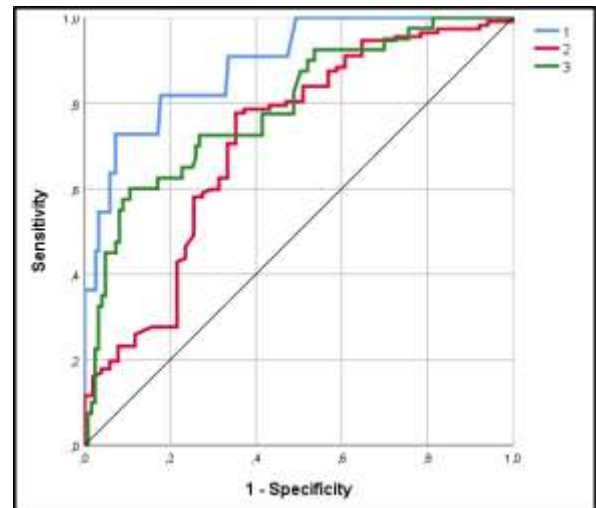


Figure 5: performance of the radial basis function model in classifying the pathogenicity of tick and flea-borne pathogens in Algeria according to published data.

Figure 5 represents the radial basis function model's ROC curve (sensitivity/specificity).

IV. DISCUSSION

Algeria has a warm Mediterranean climate [10] that generally promotes arthropods' burden. Therefore, the results of the small survey that we conducted led to seven different species of ectoparasites. Similar data have been previously reported from different parts of the country [11], [12].

For fleas, *C. felis* and *P. irritans* were the most abundant fleas. Both species had a seasonal activity ranging from May to December, while for ticks *R. sanguineus (s.l.)* was the most abundant tick in the studied hosts with a restricted activity from May to August. These results align with previous studies [13], [14].

Regarding pathogens classifying models, when comparing the summary of both models, it appears that the MLP (multilayer perceptron) model had a better performance than the RBF (radial-based function) model, with reduced training time and error in its prediction, even though it's known that MLPs are prone to overfitting for small-sized datasets, as the one exploited in this study [15]. However, this case does not prove that MLP is the ideal classifier for this task.

The dataset, consisting of only 163 samples, cannot adequately represent a sufficient number of pathogenicity classes. This necessitates the use of a larger dataset [15]. Both classifiers yielded similar results in classifying pathogens from the second risk group (the second level of pathogenicity), constituting most of the dataset. Furthermore, both models exhibited a similar trend in ROC curve analysis for classifying non- pathogenic microorganisms. However, this cannot be a deciding criterion in choosing one over the other, as only eleven samples were categorized as non-pathogenic. This smallsample size makes it insufficient for both models to be thoroughly trained and tested, as described by [16].

Nevertheless, reconsidering the covariates may be appropriate to efficiently determine the pathogenicity of pathogens since some of the covariates, such as vectorial competence and the presence of ticks or fleas, are more epidemiologic than clinical criteria.

Finally, this study brought some insights into some ectoparasites distribution and seasonal activity in the central north of Algeria with a trial on classifying the circulating pathogens in this region, which needs to be completed by further studies to have full insight into flea and tick-borne diseases and their reservoirs in the country.

ACKNOWLEDGMENT

Dr. Farid Tritri, a veterinary practitioner at M'kira Veterinary Clinic, realized a part of the sampling (Tizi Ouzou).

REFERENCES

[1]. El Hamzaoui, B., Zurita, A., Cutillas, C., & Parola, P. (2020). Fleas and flea-borne diseases of North Africa. *Acta tropica*, 211, 105627.

[2]. Perveen, N., Muzaffar, S. B., & Al-Deeb, M. A. (2021). Ticks and tick-borne diseases of livestock in the Middle East and North Africa: A review. *Insects*, 12(1), 83.

[3]. Bitam, I., Parola, P., De La Cruz, K. D., Matsumoto, K., Baziz, B., Rolain, J. M., ... & Raoult, D. (2006). First molecular detection of *Rickettsia felis* in fleas from Algeria. *The American journal of tropical medicine and hygiene*, 74(4), 532-535.

[4]. Bitam, I., Baziz, B., Kernif, T., Harrat, Z., Parola, P., & Raoult, D. (2009). Molecular detection of *Rickettsia typhi* and *Rickettsia felis* in fleas from Algeria. *Clinical microbiology and infection*, 15, 255-256.

[5]. Kernif, T., Djerbouh, A., Mediannikov, O., Ayach, B., Rolain, J. M., Raoult, D., & Bitam, I. (2012). *Rickettsia africae* in *Hyalomma dromedarii* ticks from sub-Saharan Algeria. *Ticks and tick-borne diseases*, 3(5-6), 377-379.

[6]. Benredjem, W., Leulmi, H., Bitam, I., Raoult, D., & Parola, P. (2014). *Borrelia garinii* and *Rickettsia monacensis* in *Ixodes ricinus* ticks, Algeria. *Emerging infectious diseases*, 20(10), 1776.

[7]. Pherez, F. M. (2007). Factors affecting the emergence and prevalence of vector-borne infections (VBI) and the role of vertical transmission (VT). *Journal of vector-borne diseases*, 44(3), 157-164.

[8]. Beaucournu, J. C., & Launay, H. (1990). Les puces (siphonaptera) de France et du Bassin méditerranéen occidental: par Jean-Claude Beaucournu et Henri Launay. Lechevalier.

[9]. Estrada-Peña, A., Mihalca, A. D., & Petney, T. N. (Eds.). (2018). *Ticks of Europe and North Africa: a guide to species identification*. Springer.

[10]. Ammam, I., Rahal, M., Rahal, K., & Bitam, I. (2022). Epidemiological study and identification of some flea species infesting wild rabbits (*Oryctolagus cuniculus*) and cape hares (*Lepus capensis*) in northern Algeria. *Iraqi Journal of Veterinary Sciences*, 36(1), 159-165.

[11]. Amrani, S., Seaward, M. R., Sipman, H. J., & Feuerer, T. (2018). Lichenological exploration of Algeria II: checklist of lichenized, lichenicolous, and allied fungi. *Herzogia*, 31(2), 817-892.

[12]. Sidhoum, N. R., Boucheikhchoukh, M., Mechouk, N., & Deak, G. (2023). An overview of fleas (Siphonaptera) in wild and domestic mammals from Algeria with new data from the central north and south of the country. *Acta Tropica*, 107004.

[13]. Hassissen, L., Kebbi, R., Ayad, A., & Nait-Mouloud, M. (2019). Seasonal activity of ticks infesting domestic dogs in Bejaia province, Northern Algeria. *Onderstepoort Journal of Veterinary Research*, 86(1), 1-6.

[14]. Li, Y., Pont, M. J., Jones, N. B., & Twiddle, J. A. (2001). Applying MLP and RBF classifiers in embedded condition monitoring and fault diagnosis systems. *Transactions of the Institute of Measurement and Control*, 23(5), 315-343.

[15]. Li, Y., Pont, M. J., Jones, N. B., & Twiddle, J. A. (2001). Applying MLP and RBF classifiers in embedded condition monitoring and fault diagnosis systems. *Transactions of the Institute of Measurement and Control*, 23(5), 315-343.

[16]. Sug, H. (2010). The effect of training set size on the performance of neural networks for classification. *WSEAS Trans Comput*, 9, 1297-1306.

Enhancing education decision-making with deep learning for Arabic spoken digit recognition

Zineb TOUATI-HAMAD
*Laboratory of mathematics, informatics
and systems (LAMIS)*
Echahid Cheikh Larbi Tebessi University
Tebessa, 12002, Algeria
zineb.touatihamad@univ-tebessa.dz

Mohamed Ridda LAOUAR
*Laboratory of mathematics, informatics
and systems (LAMIS)*
Echahid Cheikh Larbi Tebessi University
Tebessa, 12002, Algeria
ridda.laouar@univ-tebessa.dz

Abstract—In the realm of education, it becomes imperative to gauge students' learning progress, enabling well-informed decisions and effective support. Recent years have witnessed the ascent of deep learning as a potent instrument for speech recognition. It provides a more exact and efficient examination of people receiving speech treatment. This research offers a deep learning model centered on convolutional neural networks (CNN) suited for the categorization of Arabic spoken digits spanning from 0 to 9. Our model is rigorously trained on a broad dataset including recordings of spoken Arabic numbers, covering authentic Arabic speakers of various ages and ability levels. The findings speak much about the capabilities of our CNN-based algorithm. It attains an outstanding accuracy rate in identifying and classifying Arabic spoken digits, claiming an overall accuracy of 96.10%. Furthermore, we dive into the larger ramifications of our results within the educational environment. This emphasizes the potential of our strategy to better the evaluation of adult learners' speech therapy and to create more effective support measures. This adaptable methodology finds relevance across many educational environments, making voice recognition technology in speech therapy for adult learners more accessible and productive.

Keywords—Deep Learning, Convolutional Neural Networks, Speech Recognition, Adult Learning, Speech Therapy, Arabic Spoken Digit, Decision Making.

I. INTRODUCTION

In recent years, deep learning has emerged as a promising technology for speech recognition and has shown significant success in various languages [1]. However, the Arabic language presents unique challenges due to its complex phonological system, where subtle differences in pronunciation can alter the meaning of words [2]. Therefore, the application of deep learning in Arabic speech recognition is of particular interest in the field of education, where accurate assessment of language proficiency is essential for effective decision making and support for students.

The use of ASR technology in adult education has gained significant attention in recent years due to its potential to assess and support learners' progress in speech therapy.

However, the current ASR models based on neural networks, particularly MLPs, have limitations related to their learning style [3]. MLPs consider each neuron as independent and assign a different weight to each incoming signal, which can result in suboptimal recognition rates. CNNs, on the other hand, are designed to handle pattern recognition tasks and have demonstrated promising results in speech recognition.

Moreover, the proposed CNN-based model aims to recognize Arabic speech numbers from 0 to 9, which is a significant challenge due to the complex nature of the Arabic language. To address this challenge, we use the UCI

Machine Learning Repository's Spoken Arabic Digit Dataset to train and evaluate our model. This dataset includes recordings of Arabic speech numbers pronounced by different speakers, which provides a diverse set of data for training and testing the proposed model.

The use of CNNs in ASR has shown remarkable results in recent studies, where it has achieved state-of-the-art performance in various speech recognition tasks [4]. The proposed model can contribute to improving the performance of ASR technology in adult learning by providing accurate assessments of learners' progress in Arabic spoken digit recognition as a part of their speech therapy. Ultimately, this can lead to more effective decision making and support for adult learners, enabling them to achieve their speech therapy goals.

The structure of this paper comprises several sections. The second section presents the background of the study. The third section outlines the research motivation. The fourth section presents the related works. The fifth section describes the methodology employed in the study. The sixth section details the experiments. Finally, the paper concludes with a conclusion and suggests avenues for future research.

II. BACKGROUND

A. Speech recognition

Speech recognition is a rapidly growing field in computer science, with a wide range of applications in various domains, including education. Automatic speech recognition (ASR) technology has the potential to provide valuable insights into student learning and support their progress in language learning [5]. In the context of education, ASR technology can be used to assess students' speaking and listening skills, providing feedback on pronunciation and identifying areas for improvement. One of the main challenges in ASR is the variability of speech signals, which can be affected by factors such as speaker accent, background noise, and speaking rate. To address these challenges, deep learning techniques have shown promising results in speech recognition tasks. Convolutional neural networks (CNNs) have gained popularity in recent years, as they have been shown to be effective in handling complex patterns and achieving high accuracy in speech recognition tasks.

B. Decision support systems

Decision support systems (DSS) are computer-based tools that help people make better decisions [6]. They use data, models, and algorithms to analyze complex problems and provide recommendations to decision-makers. DSS can be used in a variety of fields, including business, healthcare, and government. They are especially useful for tackling problems that are too complex for humans to solve alone.

DSS have been around since the 1960s, but they have become much more powerful and sophisticated in recent years, thanks to advances in machine learning and big data analytics.

C. Deep learning

Deep learning (DL) is a subfield of machine learning that involves the use of neural networks with many layers [7]. It is inspired by the structure and function of the human brain and has been used to achieve breakthroughs in a wide range of applications, from image recognition to natural language processing. Convolutional neural networks (CNNs) are a type of deep learning architecture that has been particularly successful in image recognition tasks [8]. They work by using multiple layers of filters to detect features in images, and they have been used to achieve state-of-the-art performance on many benchmark datasets.

CNNs have become a popular choice for various pattern recognition tasks. CNNs can extract features from speech signals, such as mel-frequency cepstral coefficients (MFCCs), which can be used to train a model for speech recognition [9]. Moreover, CNNs have shown impressive results in speech recognition tasks, demonstrating their potential to improve the accuracy of ASR systems.

In the context of the Arabic language, there has been a growing interest in ASR technology, particularly in the field of Arabic speech recognition. However, the complex nature of the Arabic language, which includes various dialects and phonetic variations, poses a challenge for ASR systems. Therefore, the use of deep learning techniques, particularly CNNs, can contribute to improving the accuracy of ASR systems for Arabic speech recognition.

III. RESEARCH MOTIVATION

The Arabic language is one of the most widely spoken languages in the world. It is a language of great cultural significance. However, Arabic is a morphologically rich and highly ambiguous language, with many colloquial dialects that differ from Modern Standard Arabic (MSA), used in formal communication [10]. This linguistic complexity presents significant challenges for Arabic Automatic Speech Recognition (AASR) systems, particularly for adult learners undergoing speech therapy. Recognizing spoken words accurately is critical for effective communication and treatment outcomes. However, the complexity of the language can pose significant challenges that impede the progress in therapy and hinder the development of speech and language skills in adult learners.

To address this challenge, we propose using Convolutional Neural Networks (CNNs) for Arabic speech number recognition, which has shown impressive performance in ASR applications. The use of CNNs is particularly effective at handling individual variations in the speech signal and improving the speaker invariance of the acoustic model [10].

Therefore, the main motivation for this research is to develop an accurate and reliable ASR system for Arabic spoken digit recognition that can be integrated into a decision support system to aid adult learners' speech therapy processes in the Arabic language. This will help to bridge the gap in the current Arabic education system, which lacks

effective and accessible tools for adults to improve their speech and language skills.

IV. RELATED WORKS

Several research studies have been conducted using deep learning models for speech recognition in various languages, including Arabic. These studies have shown that deep learning models, such as Convolutional Neural Networks (CNNs), can significantly improve speech recognition accuracy compared to traditional methods. For example, in [10] the authors proposed a robust method for Arabic speech recognition using deep CNNs. Similarly, in [11] the authors provided an overview of deep learning-based Arabic speech recognition and highlighted the effectiveness of using CNNs in this field.

In the context of education, decision support systems have been developed to support student learning in various areas, such as academic performance prediction, student engagement, and personalized learning. The authors of [12] developed an intelligent decision support system to predict students' academic performance in Jordanian public universities.

Our contribution to this field is the development of a decision support system that focuses on adult learners' speech therapy progress in Arabic spoken digit recognition. The system employs deep learning models, particularly CNNs, for speech recognition and classification. The system's goal is to aid adult learners in improving their speech therapy progress by providing real-time feedback on their pronunciation of Arabic spoken digits from 0 to 9. The system is designed to be interactive and engaging, with visual and auditory feedback that motivates adults to practice and enhance their skills.

V. PROPOSED MODEL

In this study, a convolution neural network (CNN) approach was implemented and the architecture of the network was modified and improved by experimenting with various combinations of parameters (see Fig. 1).

The use of CNNs for processing spectrograms represented as 2D images was explored. Discriminatory features were extracted from the MFCC frequency in the selected dataset by applying a medium-sized CNN with varying numbers of convolutional layers and filters of different sizes. The number of convolutional layers was selected based on the complexity of the data, and after two or three layers, the gain in accuracy became stable while learning took a long time. The CNN operation with several filters was used to obtain a feature map, which was then passed through the activation function ReLU. The Max-pooling operation was used to obtain more significant features by selecting the maximum value of the other values of the map. This process was repeated with another convolutional layer and Max-pooling operation to obtain new features, which were then passed to two fully connected layers. A dropout layer was included to prevent overfitting, and the last layer contained the Softmax function, which provided the probability distribution on each class. The Adagrad stochastic gradient algorithm was used to optimize the network, and the categorical cross-entropy loss function was used since the study involved a multi-class classification problem (see Fig. 1).

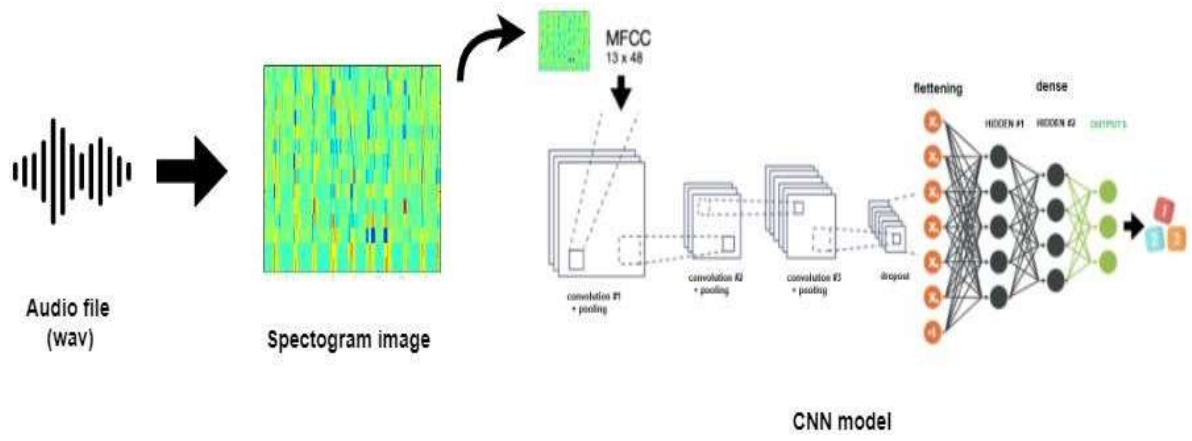


Fig. 1. Proposed architecture

Layer (type)	Output Shape	Param #
conv2d_12 (Conv2D)	(None, 256, 256, 16)	448
max_pooling2d_12 (MaxPoolin g2D)	(None, 128, 128, 16)	0
conv2d_13 (Conv2D)	(None, 128, 128, 32)	4640
max_pooling2d_13 (MaxPoolin g2D)	(None, 64, 64, 32)	0
conv2d_14 (Conv2D)	(None, 64, 64, 64)	18496
max_pooling2d_14 (MaxPoolin g2D)	(None, 32, 32, 64)	0
dropout_13 (Dropout)	(None, 32, 32, 64)	0
flatten_6 (Flatten)	(None, 65536)	0
dense_13 (Dense)	(None, 256)	16777472
dropout_14 (Dropout)	(None, 256)	0
dense_14 (Dense)	(None, 10)	2570

Total params: 16,803,626
Trainable params: 16,803,626
Non-trainable params: 0

Fig. 2. The CNN model

VI. EXPERIMENTS

A. Dataset

The Spoken Arabic Digit dataset is a collection of mel-frequency cepstrum coefficients (MFCCs) corresponding to spoken Arabic digits from 0 to 9.



Fig. 3. SpokenArabicDigits Dataset Logo [13]

It includes recordings from 88 native Arabic speakers, consisting of 44 males and 44 females, between the ages of 18 and 40. Each speaker was asked to pronounce each digit ten times, resulting in a total of 880 recordings for each digit.

The dataset provides a valuable set of data, with 75% allocated for training and 25% for testing, for speech recognition models, particularly for the recognition of Arabic numbers. The dataset is publicly available in the UCI Machine Learning Repository [13].

TABLE I. DATASET DESCRIPTION [13]

Data Set Characteristics:	Multivariate, Time-Series
Attribute Characteristics:	Real
Associated Tasks:	Classification
Number of Instances:	8800
Number of Attributes:	13
Missing Values?	No

B. Results and discussion

In this study, the proposed model was trained and evaluated using a 75% training dataset and a 25% test dataset, as previously mentioned. A series of tests were conducted to identify the optimal hyper-parameters for the model. These parameters, which include the network structure (such as the number of neurons and layers, activation functions), batch size, and number of iterations, play a significant role in the performance of the model during training. Once a suitable model was identified with minimum error rate and maximum accuracy, it was then tested on the independent test subset.

The experiment involved 10-class classification on a dataset of 6600 training spectrograms and 2200 test spectrograms. The model was trained for a number of epochs ranging from 10 to 100. The results of the experiment are presented in the Fig.3.

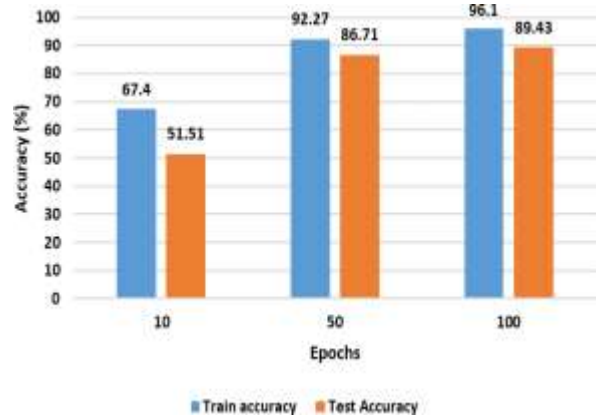


Fig. 4. Accuracy and loss results for different epochs

The results of the experiment indicate that the model begins to learn the classes from the first 10 epochs, and its accuracy improves as the number of epochs increases until it reaches an acceptable value of 96.10% for the learning dataset and 89.43% for the testing dataset. It is observed that the model begins to stabilize after reaching a certain threshold of epochs, and the increase in the number of epochs is not as significant as at the beginning. However, it is worth noting that while increasing the number of epochs improves the classification of Arabic digit, the execution time also increases with each iteration.

VII. CONCLUSION AND FUTURE WORKS

The ability of a computer to recognize human speech or voices poses a new challenge for modern scientific research, particularly as communication between humans and electronic devices has increased significantly. Consequently, researchers are striving to develop intelligent software capable of speech recognition without any human intervention.

In this study, we designed a system for automatic recognition of spoken Arabic digit using a CNN model to aid decision-making processes regarding adult learners' speech therapy progress.

We provided two subsets of data, one for learning and the other for testing, containing Arabic digit from 0 to 9 given by MFCC coefficients. The MFCC coefficient parameterization is widely used in this field to extract features and has also produced good results. Every recognition model undergoes a classification step, and we chose convolutional neural networks (CNN) in this study, which yielded satisfactory results after recognizing completely unknown speech samples. Therefore, CNNs proved to be a better technique for learning and identifying new data than other neural networks. Our research work serves as a starting point for launching future projects, such as continuous speech recognition, and also proposes ideas for improving our model, such as testing it on larger databases, adding additional training data to a database, and incorporating different languages.

ACKNOWLEDGMENT

The authors wish to acknowledge the support received from the Algerian General Directorate of Research (DGRSTD) and the Laboratory of Mathematics, Informatics and Systems (LAMIS) at the University of Larbi Tebessi in conducting this research

REFERENCES

- [1] L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: An overview," in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 8599-8603.
- [2] Z. Touati-Hamad, M. R. Laouar, I. Bendib, and S. Hakak, "Arabic Quran Verses Authentication Using Deep Learning and Word Embeddings," in International Arab Journal of Information Technology, vol. 19, no. 4, 2022, pp. 681-688.
- [3] D. Palaz, M. M. Doss, and R. Collobert, "Convolutional neural networks-based continuous speech recognition using raw speech signal," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015, pp. 4295-4299.
- [4] O. Abdel-Hamid, A. R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional neural networks for speech recognition," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 10, 2014, pp. 1533-1545.
- [5] J. Van Doremalen, L. Boves, J. Colpaert, C. Cucchiari, and H. Strik, "Evaluating automatic speech recognition-based language learning systems: A case study," in Computer Assisted Language Learning, vol. 29, no. 4, 2016, pp. 833-851.
- [6] S. Eom and E. Kim, "A survey of decision support system applications (1995-2001)," in Journal of the Operational Research Society, 2006, vol. 57, pp. 1264-1278.
- [7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," in Nature, vol. 521, no. 7553, 2015, pp. 436-444.
- [8] J. Naranjo-Torres et al., "A review of convolutional neural network applied to fruit image processing," in Applied Sciences, vol. 10, no. 10, 2020, p. 3443.
- [9] A. Chowdhury and A. Ross, "Fusing MFCC and LPC features using 1D triplet CNN for speaker recognition in severely degraded audio signals," in IEEE Transactions on Information Forensics and Security, 2019, vol. 15, pp. 1616-1629.
- [10] E. R. Abdelmaksoud, A. Hassen, N. Hassan, and M. Hesham, "Convolutional Neural Network for Arabic Speech Recognition," in The Egyptian Journal of Language Engineering, vol. 8, no. 1, 2021, pp. 27-38.
- [11] R. Amari et al., "Deep Convolutional Neural Network for Arabic Speech Recognition," in International Conference on Computational Collective Intelligence, September 2022, pp. 120-134.
- [12] Y. S. Alsalmán, N. K. Abu Halemah, E. S. AlNagi, and W. Salameh, "Using Decision Tree and Artificial Neural Network to Predict Students Academic Performance," in 2019 10th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 2019, pp. 104-109.
- [13] M. Bedda and N. Hammami, "Spoken Arabic Digit," UCI Machine Learning Repository, 2010. [Online]. Available: <https://doi.org/10.24432/C52C9Q>.

Deep Learning Models for Left Atrial Segmentation in MRI

Meriem Triki

Engineering Systems and Telecommunication Laboratory
M'Hamed Bougara University
Boumerdes, Algeria
m.triki@univ-boumerdes.dz

Mohammed Ammar

Engineering Systems and Telecommunication Laboratory
University M'Hamed Bougara
Boumerdes, Algeria
m.ammar@univ-boumerdes.dz

Abstract— The left atrium (LA) segmentation is a crucial and essential procedure in the field of cardiac imaging due to its significance in cardiovascular health and its role in diagnosing heart conditions. Magnetic resonance imaging (MRI) is a non-invasive medical imaging technique that has become indispensable in the cardiovascular field, especially for visualizing and evaluating atrial myopathy. In recent years, deep learning has emerged as the approach with the highest adoption rate for segmenting cardiac images. For the purpose of segmenting the left atrium from MRI images, we applied three Unet variation architectures to compare the optimal one, and we experimented with dice and cross-entropy loss.

Keywords— *left atrial, segmentation, cardiac MRI, deep learning, U-net.*

I. INTRODUCTION

The left atrium is a vital component of the heart's structure and holds significant importance in the fields of cardiology and cardiac health. Due to the intricate variations in its structure among patients, accurately understanding it becomes challenging, often resulting in misdiagnosis and ineffective treatment [1].

Left atrial segmentation identifies and delineates the boundaries of the left atrium in medical imaging, typically using techniques from image analysis and computer vision. This segmentation task holds significant importance in cardiac imaging, as it allows precise quantification of the left atrium's size and shape, which can be critical in diagnosing and monitoring various heart conditions including atrial fibrillation, mitral valve regurgitation, and other cardiac diseases.

Noninvasive medical imaging became essential in cardiovascular medicine. The exceptional measurement precision offered by MRI positions it as a perfect tool for monitoring the advancement and treatment outcomes of cardiovascular diseases [1].

Existing models of left atrial segmentation are based on encoder-decoder architectures like FCN [2] and U-Net [3]. For instance, [4] segmented the left atrial from magnetic resonance (MR) images with two consecutive 3D U-Net architectures. The initial 3D U-Net aimed to identify and

coarsely extract the region of interest, while the second network focused on performing detailed segmentation from the cropped images. [5] devised an automated left atrial segmentation technique utilizing a FCN to enhance the accuracy of left atrial structural delineation. This method integrates a dual-path, multi-scaled architecture capturing both local atrial tissue geometry and the broader positional information of the left atrium. [6] proposed a deep 2D U-Net, a derivation of the original U-Net, was introduced for automatic left atrium segmentation from GE-MRIs. This method employed multi-task learning by augmenting the network depth and incorporating an additional classification branch.[7] modified the 3D U-Net [4] by incorporating a hierarchical aggregation unit (HAU) serving as a trunk branch and an attention unit (AU) in the encoder path for 3D left atrial segmentation. To address the segmentation issues in GE-MRI for the left atrium, [8] created two image segmentation networks using FCN and U-Net architectures. They improved the segmentation accuracy by modifying the dice loss function, aiming to reduce the impact of imbalances between positive /negative samples. [9] constructed a modified 3D U-Net by integrating dilated convolutions at the deepest layer of the encoder and incorporating residual connections, aiming to merge and capture local and global information. [10] proposed a two-stage method for LA segmentation from LGE MRI. The method involves an Otsu-based localization phase followed by fine segmentation, utilizing both 2-D and 3-D pipelines which is based on the original U-Net architecture. [1] developed a method for automatic left atrial region segmentation in MRI scans of patients with LAE. This method comprises two parts: the first one, a U-Net architecture with Gaussian blur and channel weight neural network (GCW-UNet) segments the left atrial region. The subsequent part reconstructs the 2D segmented left atria into a 3D model.

The achievements of these studies demonstrate that designed U-Net-based architectures can perform satisfactorily on tasks like left atrial segmentation. There are, however, several U-Net versions, including Residual U-Net [11], Dense U-Net [12], V-Net [13], Attention U-Net [14], U-Net++ [15], SegResNetVAE [16], to mention a handful. Due

to the various U-Net architectures available, selecting the optimal one is challenging.

In this study, we compare the U-net variation for left atrial segmentation. We tested 3DU-Net [17], Attention U-Net [14], and SegResNet [16] variants and experimented with the Dice and Cross Entropy loss function.

II. METHOD

A. Dataset

The dataset employed in this work from Left Atrial Segmentation Challenge Dataset (LASC2013) [18], sourced from STACOM'13 in MICCAI'13. This dataset comprises 30 MRI images in NIFTI format, with 20 images designated as the training set, accompanied by manually labeled segmentation images. The remaining images constitute the test set, lacking manual segmentation labels. Each sample contained 100–130 slices. Therefore, we confined the dataset to the 20 MRI images with labels, partitioned into 15 images for training purposes and 5 images for validation.

B. Data Preprocessing

As an initial step, all files were stacked in such a way that each example has the shape $(320 \times 320 \times 110)$ and the same-sized segmentation mask. Then, on the borders of each volume, redundant background voxels are cropped. This has no effect on the dataset because it provides no significant information, which the network could ignore, but it does reduce its size and hence its computation burden.

C. Model Architectures

To determine the optimal U-Net variation architecture, we studied some models, including 3D-UNet [17], Attention-UNet [14], and SegResNet [16]. Here, we present brief descriptions of each of these models.

3D U-Net [17] (shown in Fig. 1) is composed of two essential stages: a contracting encoder, whose primary purpose is to examine the entire image, and a full-resolution segmentation produced by a successively expanding decoder. A rectified linear unit (ReLU) and two $3 \times 3 \times 3$ convolutions are present in every encoder layer. Following each of them, a max pooling layer of $2 \times 2 \times 2$ is added. In a similar manner, two $2 \times 2 \times 2$ up-convolutions, two $3 \times 3 \times 3$ convolutions, and ReLU comprise the decoder stage. To suit the amount of labels, the final convolution layer $1 \times 1 \times 1$ decreases the output channels [17].

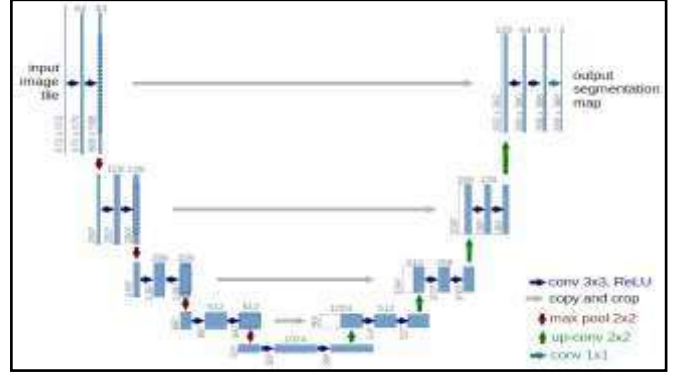


Fig. 1. The 3D U-Net architecture.

Attention U-Net [14] expands the basic U-Net by incorporating an attention gate (seen in Fig. 2) in the decoder section to emphasize important features that move through the skip connections. Prior to the concatenation process, the attention gate modifies the encoder's feature map to merge only the relevant activations. It figures out which parts of the encoder's feature map are most significant, taking cues from the contextual information provided by the previous decoder block's feature map. This is accomplished by multiplying the weight values determined by the attention gate by the encoder feature map. These weight values, which are limited to the interval $(0, 1)$, indicate the degree of focus that the neural network is applying to a certain pixel [14].

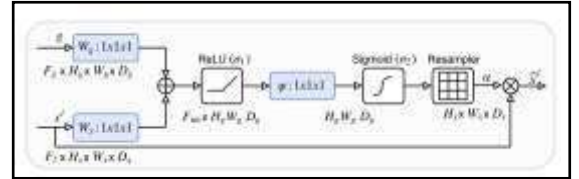


Fig. 2. The attention gate's architecture. Attention weights (α) get multiplied by input features (x_1). To calculate α , the feature map from the associated encoder level and the input features (x_1) are first processed by $1 \times 1 \times 1$ convolution and then added together. After that, a further $1 \times 1 \times 1$ convolution and ReLU activation are used. Finally, trilinear interpolation is used to upsample attention weights [14].

SegResNet is based on [16] but the variational autoencoder (VAE) mentioned in the article is not included in the module. The method relies on a CNN architecture featuring an encoder-decoder design. It employs an asymmetrical larger encoder to capture image features and a smaller decoder for reconstructing the segmentation mask (shown in Fig. 3). In the encoder, ResNet is implemented as blocks. Each block comprises two $3 \times 3 \times 3$ convolutions accompanied by normalization and ReLU activation, followed by an additive identity skip connection. The decoder follows a similar structure to the encoder, but it contains a single block for each spatial level. Each decoder level commences by decreasing the feature count by a factor of two using $1 \times 1 \times 1$ convolutions while simultaneously increasing the spatial dimension

utilizing 3D bilinear operations, then adding encoder output from the same spatial level. The resulting decoder output matches the original input feature size both in terms of spatial dimensions and feature size after 1x1x1 convolution and a sigmoid function [16]

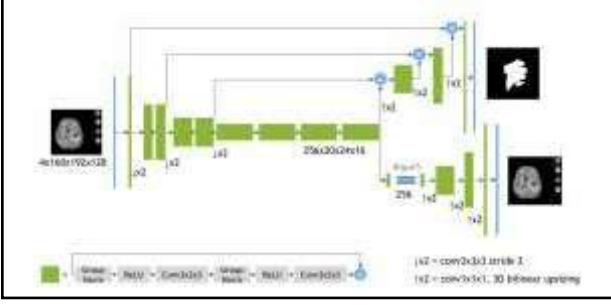


Fig.3. The architecture of SegResNet. Every green block represents a ResNet-like block incorporating group normalization [16].

D. Loss function

When handling medical images, it is usual for the anatomy segment to occupy minor areas of the image, which could result in a significant bias in the background; thus, the foreground regions are typically underrepresented or missing in the subsequent predicted segmentations. To solve this class unbalance, we applied DiceCross-Entropy loss, which is a hybrid of dice loss [13] and cross-entropy loss.

III. EXPERIMENTS AND RESULTS

A. Augmentation

The quantity of training data has a significant impact on their excellent segmentation task performance. The data augmentation technique can have a big impact when these networks are applied to tiny datasets. It is used to reduce the issue of overfitting by increasing the dataset during training, thereby improving the performance of network architecture.

During the training, we apply to each image/mask pair a composition of image transformations such as RandomFlip,

CropForeground, and Random Scale and Shift Intensity, which provides an expanded dataset that offers a wider range of image variations. As a result, during training, random transformations prevent the network from fixating on specific features within its perceptive field by constantly shifting these features across the field.

B. Evaluation Metric

We have evaluated our results using the Dice Coefficient score, referred to as the overlapping index, which measures the extent of overlap between the ground truth and the predicted output.

For a given class, G_i and P_i indicate the ground truth and predicted values for voxel i . The Dice score metric is defined by the following formula:

$$Dice(G, P) = \frac{2 \sum_{i=1}^I G_i P_i}{\sum_{i=1}^I G_i + \sum_{i=1}^I P_i} \quad (1)$$

$$\sum_{i=1}^I G_i + \sum_{i=1}^I P_i$$

C. Experimental Results

To compare the segmentation models of various U-Net variations as described earlier, we implemented our work in PyTorch¹ and MONAI². We set the learning rate at 1×10^{-5} and employed the Adam optimizer.

According to our experiments (the detailed results are shown in Table 1), the 3D U-Net achieves the highest average Dice score. SegResNet has a similar score as of Attention U-Net, but their training time is two times greater than that of a 3D U-Net.

Table1. Averaged Dice scores of the data comparing the 3D U-Net, SegResNet, Attention-Unet models.

Model	3D U-Net	SegResNet	AttentionU-net
DSC (Train)	0.9294	0.8240	0.8027
DSC (Test)	0.9005	0.8098	0.7774

Fig.4 below presents the visual comparison of segmentation effects achieved by the previous approaches using predicted segmentation images from one dataset example, contrasted with manual segmentation images. The first row displays the MRI images earmarked for segmentation, while the second row shows the manual labeling of the left atrial structures by an expert. Commencing from the third row are the segmentation outcomes for various models. Regarding the results obtained, the 3D-UNet showed the best performance, successfully segmenting the left atrium along with its appendages (LAA). On the other hand, the remaining models achieved decent segmentation but failed to encompass the left atrium's appendages. An inherent limitation across all three models, notably visible in the Attention U-Net, was the inability to emphasize key features in cardiac MRI images, primarily due to sequence connections and low contrast between the left atrium and its surrounding tissues.

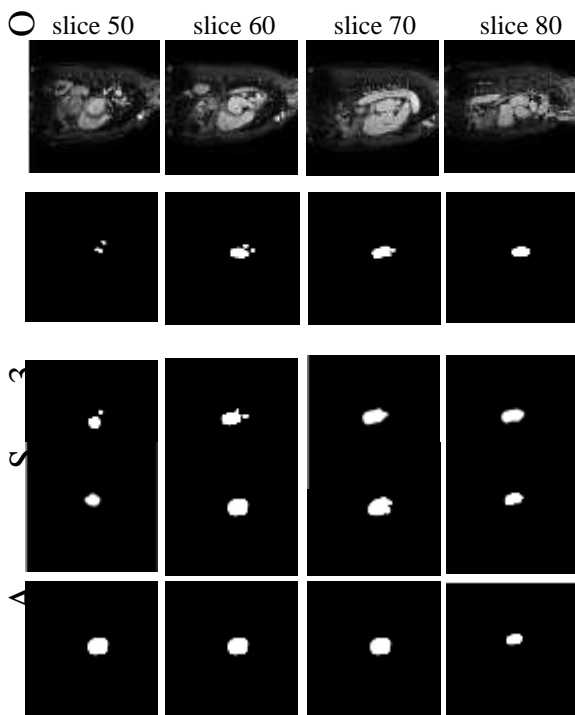


Fig.4. Segmentation results for different networks.

IV. CONCLUSION

The study is limited to a small number of samples, the changes in scale, and the high imbalance between the left atrial structure and its surrounding areas. These factors made challenges in accurately segmenting the left atrial, making it difficult to achieve precise delineation. These limitations might affect future research general findings. To improve this, future studies could explore and compare more U-Net variant models, helping to better understand how to segment the left atrial effectively.

REFERENCES

- [14] WONG, Kelvin KL, ZHANG, An, YANG, Ke, et al. GCW-UNet segmentation of cardiac magnetic resonance images for evaluation of left atrial enlargement. *Computer Methods and Programs in Biomedicine*, 2022, vol. 221, p. 106915.
- [15] LONG, Jonathan, SHELHAMER, Evan, et DARRELL, Trevor. Fully convolutional networks for semantic segmentation. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015. p. 3431-3440.
- [16] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation.
- [17] JIA, Shuman, DESPINASSE, Antoine, WANG, Zihao, et al. Automatically segmenting the left atrium from cardiac images using successive 3D U-nets and a contour loss. *Springer International Publishing*, 2019. p. 221-229.
- [18] Xiong, Z.; Fedorov, V.V.; Fu, X.; Cheng, E.; Macleod, R.; Zhao, J. Fully Automatic Left Atrium Segmentation From Late Gadolinium Enhanced Magnetic Resonance Imaging Using a Dual Fully Convolutional Neural Network. *IEEE Trans. Med. Imaging* 2019, 38, 515– 524.
- [19] Chen, Chen, BAI, Wenjia, et RUECKERT, Daniel. Multi-task learning for left atrial segmentation on GE- MRI. *Springer International Publishing*, 2019. p. 292- 301.
- [20] LI, Caizi, TONG, Qianqian, LIAO, Xiangyun, et al. Attention based hierarchical aggregation network for 3D left atrial segmentation. *Springer International Publishing*, 2019. p. 255-264.
- [21] LIU, Yashu, DAI, Yangyang, YAN, Cong, et al. Deep learning based method for left atrial segmentation in GE-MRI. *Springer International Publishing*, 2019. p. 311-318.
- [22] Vesal, S.; Ravikumar, N.; Maier, A. Dilated Convolutions in Neural Networks for Left Atrial Segmentation in 3D Gadolinium Enhanced-MRI. In *International Workshop on Statistical Atlases and Computational Models of the Heart*; Springer: Berlin/Heidelberg, Germany, 2019.
- [23] BORRA, Davide, ANDALÒ, Alice, PACI, Michelangelo, et al. A fully automated left atrium segmentation approach from late gadolinium enhanced magnetic resonance imaging based on a convolutional neural network. *Quantitative Imaging in Medicine and Surgery*, 2020, vol. 10, no 10, p. 1894.
- [24] HE, Kaiming, ZHANG, Xiangyu, REN, Shaoqing, et al. Deep residual learning for image recognition. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. p. 770-778.
- [25] HUANG, Gao, LIU, Zhuang, VAN DER MAATEN, Laurens, et al. Densely connected convolutional networks. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. p. 4700- 4708.
- [26] MILLETARI, Fausto, NAVAB, Nassir, et AHMADI, Seyed-Ahmad. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In : *2016 fourth international conference on 3D vision (3DV)*. Ieee, 2016. p. 565-571.
- [27] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.:Attention u-net: Learning where to look for the pancreas (2018)
- [28] ZHOU, Zongwei, RAHMAN SIDDIQUEE, Md Mahfuzur, TAJBAKHSH, Nima, et al. Unet++: A nested u-net architecture for medical image segmentation. *Springer International Publishing*, 2018. p. 3-11.
- [29] MYRONENKO, Andriy. 3D MRI brain tumor segmentation using autoencoder regularization. *Springer International Publishing*, 2019. p. 311-320.
- [30] ÇIÇEK, Özgün, ABDULKADIR, Ahmed, LIENKAMP, Soeren S., et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation. *Springer International Publishing*, 2016. p. 424-432.
- [31] Tobon-Gomez, C.; Geers, A.J.; Peters, J.; Weese, J.; Pinto, K.; Karim, R.; Ammar, M.; Daoudi, A.; Margeta, J.; Sandoval, Z.; et al. Benchmark for Algorithms Segmenting the Left Atrium From 3D CT and MRI Datasets. *IEEE Trans. Med. Imaging* 2015, 34, 1460– 1473.
- [32] ZHANG, Zhilu et SABUNCU, Mert. Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*, 2018, vol. 31.

Recommendation Systems: techniques, challenges and applications

Djelloul BETTACHE
Computer Science department
Hassiba Benbouali University Of Chlef
d.bettache@univ-chlef.dz

Nassim DENNOUNI
Higher school of management
Tlemcen, Algeria
dennouninas@gmail.com

Abstract— The rapid expansion of the Internet, mobile technology, and e-commerce has created an overwhelming surplus of information. In response, recommendation systems have been developed to sort and prioritize relevant information for users. These systems empower individuals to discover customized knowledge, products, and services. Since their inception, researchers have diligently worked on improving recommendation systems by utilizing a variety of filtering techniques to enhance both the user experience and system performance. This paper offers an initial examination of recommendation systems that rely on filtering methods, discussing the challenges they encounter and the fields in which they are applied.

Keywords— Collaborative filtering, Content-based filtering, Recommendation system, recommendation technique, recommendation algorithm.

I. INTRODUCTION

When people have many options to choose from, it can be confusing to pick the right one. That's why recommendation systems were created. They help us decide by giving suggestions based on our preferences [1]. People have always sought advice from others for various reasons. These systems help by narrowing down our choices and suggesting what might be best for us[2]. The huge amount of information and user preferences nowadays make these recommendation systems more important. The first computer-based recommendation system was made in 1992, called Tapestry[3]. It helped manage lots of documents like emails and news articles, using human input to be more effective. The main reason for creating recommendation systems is to reduce the load of information and make it easier for users to find what they want. This benefits both users and the companies providing the service[4]. Nowadays, many big companies like Google, Twitter, LinkedIn, Netflix, and Amazon use recommendation systems to help them make decisions that increase their profits and reduce risks[5]. Some well-known recommendation systems today include Group Lens, Amazon.com, Netflix, Google News, Youtube, Instagram and Facebook.

II. TECHNIQUES

Numerous approaches have been proposed for creating an effective recommendation system. Among them, two serve as basic principles on which other methods are built: content filtering and collaborative filtering. These methods are then extended and, in the current RS techniques, they are categorized as follows:

A. Content Filtering Recommendation System)

Content-based (CB) techniques use the characteristics of an item to find recommendations for a user based on their past preferences. This method works in two main steps: it creates a user profile using the characteristics generally preferred by the user, and then compares the characteristics of each item to this user profile, recommending items that have a high degree of similarity [1,6]. To use content-based filtering, an article profile is built, which includes the essential characteristics of the article. For example, in the case of a film, the profile might include details such as cast, director, year of release and genre.

CB filtering is a simple approach that does not rely on user feedback. Sometimes a single preference can lead to recommendations for several items. This method works well when information about items is well structured and readily available, as is the case for films, songs, products or books. However, there are limitations, as not all items have a detailed description, making it difficult to measure similarity between items[7]. These recommendation systems also tend to provide consistent but somewhat static results over time.

B. Collaborative Filtering (CF)

Collaborative filtering (CF) is the most widely used recommendation technique. Its basic principle is that individuals with similar interests tend to share preferences for new and upcoming items[8]. This approach is based on two key principles. Firstly, it identifies a group of users with similar interests, known as 'nearest neighbours', whose opinions are aggregated to form the basis of recommendations. Secondly, it extends this idea by creating a larger group to exert a more significant influence on the recommendation process.

Collaborative filtering techniques are applied in a variety of fields, taking advantage of large datasets, and have found applications in finance, weather forecasting, environmental sensing and e-commerce. These techniques use ratings or preferences provided by users for various items to predict other items likely to appeal to an active user[1,8].

Collaborative methods are categorized into three groups: memory-based approaches, model-based approaches, and hybrid approaches.

1) Memory-Based Filtering

Memory-based methods are known for their simplicity and ease of implementation. The most common approach in this category is memory-based neighbourhood filtering[7], which predicts preferences by referring to users with similar tastes to the user being queried or to items that resemble the item being queried. The effectiveness and efficiency of the

neighbourhood technique is highly dependent on how the similarity between users or items is calculated. This method can be improved by various techniques such as default voting, inverse user frequency and case amplification. Memory-based techniques fall into two categories: user-based filtering and item-based filtering.

a) *User-based*: filtering determines the predicted preferences for an active user's items by examining their similarity to other users who have rated the same items[1,10].

b) *Item-based filtering*: on the other hand, calculates predictions by measuring the similarity between items. This method retrieves all the items rated by an active user, evaluates the similarities of these items to the target item and selects the N most similar items to predict the active user's preference for the target item. Model-Based Filtering[1,10].

2) Model-Based Filtering

Model-based techniques leverage data mining and machine learning methodologies to forecast a user's item preferences. They encompass a variety of methods, including association rules, clustering, decision trees, artificial neural networks, Bayesian classifiers, regression, link analysis, and latent factor models[1,6]. Among these, latent factor models are the most extensively studied and commonly used model-based approaches. These models streamline the user-item preference matrix by reducing dimensionality and learning latent variables that aid in predicting user preferences for items during the recommendation process. Techniques in this category encompass matrix factorization, singular value decomposition, probabilistic matrix factorization, Bayesian probabilistic matrix factorization, and nonnegative matrix factorization.

3) Hybrid Filtering

Hybrid filtering techniques is created by merging the strengths of two or more filtering methods and overcomes their individual limitations[9]. This approach leads to more effective and improved recommendation results. An example of this is the combination of memory-based and model-based approaches to create a hybrid filtering system, resulting in enhanced prediction accuracy and operational efficiency.

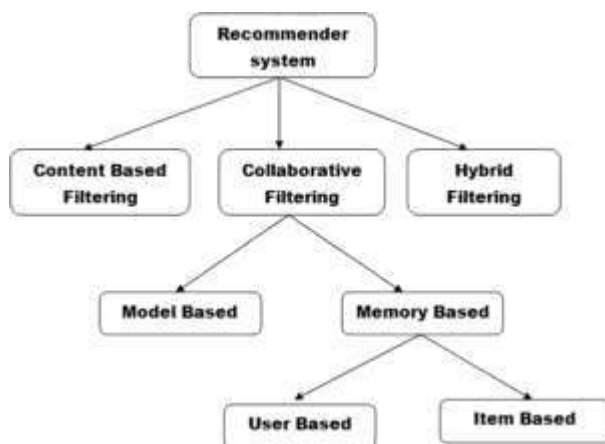


Fig. 1. Recommender System techniques

III. CHALLENGES

Recommender systems (RS) face several challenges, including:

A. Data Sparsity

Numerous e-commerce and online shopping platforms employ recommender systems to evaluate extensive item catalogs. However, as the item sets grow larger, the user-item interaction data becomes increasingly sparse, presenting a challenge for many recommender systems[1,8]. Limited user ratings or preferences can result in less accurate predictions. Additionally, new items are challenging to recommend until they have received sufficient user ratings, and new users may struggle to receive relevant recommendations due to their limited preference history.

B. Scalability

As the number of users and items grows, the computational resources required for recommendation tasks can become a significant challenge[8].

C. Cold Start

New users and items pose a challenge because the system lacks historical data to make accurate suggestions until more interactions occur[8].

D. Evaluation Metrics

Determining the effectiveness of RS poses challenges as traditional evaluation metrics may not reflect the quality of recommendations accurately[8,11].

E. Gray Sheep

Users whose preferences don't align with any particular group are referred to as "gray sheep." These users pose a challenge for collaborative filtering systems as they don't fit neatly into the typical user profiles. Additionally, there's a special class of users known as "black sheep" whose highly idiosyncratic behavior makes generating recommendations nearly impossible. To address these issues, an optimal combination of content-based and collaborative filtering, often referred to as hybrid techniques, can help mitigate the problems associated with "gray sheep" users [8].

F. Synonymy

Synonymy is the phenomenon where several identical or highly similar items are listed under different names or entries. Many recommender systems struggle to identify this underlying connection, resulting in the separate treatment of these products[1] [8].

Addressing these challenges is an ongoing effort in the field of recommender systems to enhance the quality and usability of recommendation services[1,8].

IV. APPLICATIONS

Recommender systems have seen significant growth and use in various service domains. Their applications now include personal, social and business services, all of which are of practical importance in our daily lives and have a significant impact. In general, the applications of recommender systems can be categorised as follows:

A. Social Network

Online social networks like Facebook, Instagram, Twitter, and LinkedIn are significant digital platforms where users not only share details about their daily lives, hobbies, and interests but also engage and interact with other users[6] [9]. The widespread adoption of these social networking services has led to a substantial growth in user-generated data.

B. E-Commerce

This system was created to offer guidance to online shoppers. It's widely used, particularly in the field of e-commerce, and relies on ratings and preferences to generate recommendations. Tagging and user reviews are also utilized to establish connections between users and items[9]. Well-known e-commerce platforms like iTunes, Amazon, and eBay make use of these recommendation systems to enhance the shopping experience for their users.

C. Entertainment

As the volume of movies, videos, and music continues to expand, users often find it challenging to discover content that aligns with their preferences. This has prompted the development of more efficient and personalized recommendation systems. Collaborative filtering is a widely used technique in these systems[1]. Specifically, for video content like TV shows on platforms such as Netflix and YouTube, a combination of social and context-aware techniques complements the traditional content-based and collaborative approaches effectively[6].

D. Contents

In recent times, recommender systems have emerged as a vital component of the e-content system, enabling users to discover information and knowledge within digital libraries. They are applied in various domains, including personalized web pages, recommending new articles, and filtering emails[8].

E. Service Oriented

The Internet and mobile devices have created significant opportunities for accessing diverse types of information. This, in turn, has spurred the development of service-based recommendation systems across various domains, including tourist recommendations, travel services, matchmaking services, and consultation services[1,6].

F. Tourism

With the growing interest in travel, the tourism industry has embraced recommendation systems to suggest tourist destinations, travel routes, and transportation options. These recommendation systems rely on situational data, including reviews, location information, user details, time, and weather, which are collected through social networking sites (SNS). This has led to a surge in research focusing on SNS-based recommendation systems in the tourism service sector[12].

G. Education

Traditional classroom education is evolving into a new form known as "Smart Learning" or e-learning, where education takes place online. This shift is driven by the widespread use of smart devices and advancements in wireless networks. Smart education can tap into extensive digital

resources and offer personalized learning experiences tailored to individual learners' needs, goals, talents, and interests, all without constraints of time and space[1,8]. This new educational approach aligns with the digital age's learning trends, and education services employing recommendation systems play a crucial role in delivering learning resources that cater to each learner's style and knowledge level. This leads to an effective and efficient learning experience, providing personalized learning content to learners.

V. CONCLUSION

Recommendation systems have become part of our daily lives. With the development of the internet and the amount of information available, it is important to have better recommendation systems that work well. These systems help users find things they like and might not otherwise have discovered. This article describes the different recommendation systems, how they work, the challenges they face and where they are used.

REFERENCES

- [1] Malik, S., Rana, A., & Bansal, M. (2020). A survey of recommendation systems. *Information Resources Management Journal (IRMJ)*, 33(4), 53-73.
- [2] Nagarnaik, P., & Thomas, A. (2015, February). Survey on recommendation system methods. In *2015 2nd international conference on electronics and communication systems (ICECS)* (pp. 1603-1608). IEEE.
- [3] Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using collaborative filtering to weave an information tapestry. *Comm. ACM* 35(12), 61–70 (1992)
- [4] Patel, K., & Patel, H. B. (2020). A state-of-the-art survey on recommendation system and prospective extensions. *Computers and Electronics in Agriculture*, 178, 105779.
- [5] Bouneffouf, D., Bouzeghoub, A., Ganarski, A.L.: Risk-aware recommender systems. In: *Neural Information Processing*, pp. 57–65. Springer, Berlin, Heidelberg (2013).
- [6] Zuva, T., Ojo, S. O., Ngwira, S., & Zuva, K. (2012). A survey of recommender systems techniques, challenges and evaluation metrics. *International Journal of Emerging Technology and Advanced Engineering*, 2(11), 382-386.
- [7] Afsar, M. M., Crump, T., & Far, B. (2022). Reinforcement learning based recommender systems: A survey. *ACM Computing Surveys*, 55(7), 1-38.
- [8] Shapira, B., Rokach, L., & Ricci, F. (2022). *Recommender systems handbook*.
- [9] Raza, S., & Ding, C. (2022). News recommender system: a review of recent progress, challenges, and opportunities. *Artificial Intelligence Review*, 1-52.
- [10] Papadakis, H., Papagrigoriou, A., Panagiotakis, C., Kosmas, E., & Fragopoulou, P. (2022). Collaborative filtering recommender systems taxonomy. *Knowledge and Information Systems*, 64(1), 35-74.
- [11] Zangerle, E., & Bauer, C. (2022). Evaluating recommender systems: survey and framework. *ACM Computing Surveys*, 55(8), 1-38.
- [12] Ricci, F. (2022). Recommender systems in tourism. In *Handbook of e-Tourism* (pp. 457-474). Cham: Springer International Publishing.

The impact of AI on credit risk assessment: A SWOT analysis

AMRANI Meriem

PHD Student,
Superior School of Management
Tlemcen, Algeria
meriem.amrani94@gmail.com

SAIDI Tarik

Doctor
Superior School of Management
Tlemcen, Algeria
saidi_t5@yahoo.fr

Abstract : The use of artificial intelligence (AI) in credit evaluation has become a central topic in the financial sector. This study focuses on the use of AI in credit risk assessment using a SWOT analysis. The goal is to highlight how AI is revolutionizing the credit evaluation process by offering improved accuracy, operational efficiency, and credit offer customization. However, there are also challenges to be addressed, such as complexity, bias, and transparency. To take advantage of AI in a responsible and effective manner, lenders must invest in the resources and skills necessary to implement and manage AI. They must also ensure that AI algorithms are designed and tested to minimize bias, and provide increased transparency on how AI algorithms make decisions.

Keywords: AI , risk , credit, evaluation, challenges

I. INTRODUCTION

At the core of the financial industry, credit risk assessment is a critical activity that influences billions of dollars in transactions every day. It is the cornerstone on which trust in loans, investments, and business transactions is built. When a financial institution or a private lender considers granting a loan, issuing a credit card, or investing in bonds, credit risk analysis is the linchpin that determines whether the transaction will be approved, at what interest rate, or whether it will simply be declined. Traditional credit risk assessors have long relied on analysis methods that, while valid, had their limitations. Models based on historical creditworthiness criteria and manual evaluations were subject to human error and could sometimes overlook early risk indicators. In a world where financial markets are in constant flux, and economic uncertainty is the norm, there is an increasing imperative to have more robust and sophisticated tools to evaluate credit risk.

This is where artificial intelligence (AI) comes into play, a technological revolution that has rapidly stormed the financial sector. AI, with its ability to analyze vast amounts of data, detect complex patterns, and make decisions based on real-time information, is revolutionizing how we approach credit risk assessment. AI's applications in this field are vast and continuously expanding, providing a fresh perspective on how we evaluate borrowers' creditworthiness. From sophisticated machine learning models to natural language processing algorithms capable of

dissecting unstructured information, AI is fundamentally transforming how financial institutions, credit rating agencies, and private lenders assess credit risk. In this research, we will explore the following issue:

What are the challenges and risks associated with the use of AI in credit risk management?

II. LITERATURE REVIEW

A. DEFINITION OF CREDIT RISK :

"Credit risk is the risk that a borrower will not repay a loan. It is a financial risk that is inherent to lending and is a major concern for banks and other financial institutions". (BCBS, 2010)

"Credit risk is the risk of loss due to a borrower's failure to repay a loan or meet contractual obligations". (Mishkin, 2012)

B. Overview of different credit risk assessment methods:

Credit risk is the probability that a borrower will default on a loan. It is an important risk for banks and other financial institutions, which use a variety of methods to assess and manage it.

The main credit risk assessment methods are as follows:

- Qualitative approaches: These approaches focus on qualitative factors, such as the borrower's solvency, ability to repay, and financial situation.

- Quantitative approaches: These approaches focus on quantitative factors, such as the borrower's financial data, credit history, and industry.

A. Qualitative approaches

Qualitative approaches are often used for low-risk loans, such as consumer loans or mortgages. They focus on the following factors:

➤ Borrower solvency: Solvency is the ability of a borrower to repay a loan. It is typically assessed based on factors such as the borrower's income, assets, and liabilities.

➤ Borrower ability to repay: Ability to repay is the amount of money that the borrower can afford to put towards repaying a loan. It is typically assessed based on factors such as the borrower's disposable income and the amount of other debts the borrower has.

➤ Borrower financial situation: The borrower's financial situation is an overall assessment of their financial

health. It is typically assessed based on factors such as income, assets, liabilities, credit history, and industry.

B. Quantitative approaches

Quantitative approaches are often used for higher-risk loans, such as business loans or loans to individuals with a poor credit history. They focus on the following factors:

- **Borrower financial data:** The borrower's financial data includes the borrower's balance sheets, income statements, and cash flow statements. This data can be used to assess the borrower's solvency, ability to repay, and financial situation.

Machine learning

- **Borrower credit history:** The borrower's credit history is a record of their previous payments on loans. It can be used to assess the likelihood that the borrower will repay a new loan.

- **Borrower industry:** The borrower's industry can affect their likelihood of default. For example, businesses operating in cyclical industries are more likely to go bankrupt in a recession. (Hasan, 2023)

C. Introduction to AI and its impact on the banking sector :

Artificial intelligence (AI) is a rapidly evolving field with the potential to transform many industries, including banking. AI-powered technologies are already being used in a variety of ways in the banking sector, from improving customer service to detecting fraud.

a) Impact of AI on the banking sector

- The impact of AI on the banking sector is likely to be significant in the coming years. AI-powered technologies have the potential to:

- **Improve customer service:** AI can be used to automate tasks such as customer service chatbots and virtual assistants. This can free up human employees to focus on more complex tasks, such as providing personalized advice to customers.

- **Detect fraud:** AI can be used to analyze large amounts of data to identify patterns that may indicate fraud. This can help banks to protect themselves from financial losses.

- **Make lending decisions:** AI can be used to assess the creditworthiness of borrowers more accurately than traditional methods. This can help banks to reduce the risk of lending to borrowers who are likely to default.

- **Personalize products and services:** AI can be used to collect data about customers' preferences and behaviors. This data can be used to personalize products and services to meet the needs of individual customers.

b) Challenges of AI in the banking sector

- While AI has the potential to revolutionize the banking sector, there are also some challenges that need to be addressed. These challenges include:

- **Data privacy and security:** AI-powered technologies often require access to large amounts of data. This data needs to be protected from unauthorized access.

- **Bias:** AI models can be biased, which can lead to discrimination against certain groups of people.

- **Explainability:** It can be difficult to explain how AI models make decisions. This can make it difficult to trust AI-powered technologies (Cai, 2023)

III. AI applications in credit risk assessment

A. Machine learning

- a) **Definition:** Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate in predicting outcomes without being explicitly programmed to do so. ML algorithms use historical data as input to predict new output values.

b) Different types of machine learning:

- **Supervised learning:** Supervised learning algorithms are trained on a set of labeled data, where each input is associated with a known output. Once trained, the algorithm can be used to predict the output for new inputs.

- **Unsupervised learning:** Unsupervised learning algorithms are trained on a set of unlabeled data, where the output is not known. The algorithm learns to identify patterns in the data without being explicitly told what to look for.

- **Reinforcement learning:** Reinforcement learning algorithms learn to perform a task by trial and error. The algorithm is rewarded for taking actions that lead to desired outcomes and penalized for taking actions that lead to undesired outcomes.

c) Applications of machine learning in credit risk assessment:

- **Credit scoring:** ML algorithms can be used to develop credit scoring models that predict the likelihood of a borrower defaulting on a loan.

- **Fraud detection:** ML algorithms can be used to develop fraud detection models that identify fraudulent loan applications.

- **Early warning systems:** ML algorithms can be used to develop early warning systems that identify borrowers who are at risk of defaulting on a loan. (China, 2023)

D. Predictive analytics

- a) **Definition:** Predictive analytics is the use of data, statistical algorithms, and machine learning techniques to identify patterns and make predictions about future events.

b) Different types of predictive analytics:

- **Regression:** Regression is a statistical technique that is used to predict continuous variables, such as the amount of a loan that a borrower will default on.

➤ **Classification:** Classification is a statistical technique that is used to predict discrete variables, such as whether or not a borrower will default on a loan.

➤ **Clustering:** Clustering is an unsupervised learning technique that is used to identify groups of similar data points.

d) **Applications of predictive analytics in credit risk assessment:**

➤ **Credit scoring:** Predictive analytics can be used to develop credit scoring models that predict the likelihood of a borrower defaulting on a loan.

➤ **Fraud detection:** Predictive analytics can be used to develop fraud detection models that identify fraudulent loan applications.

➤ **Early warning systems:** Predictive analytics can be used to develop early warning systems that identify borrowers who are at risk of defaulting on a loan. (OECD, 2023)

E. Robotization

a) **Definition:** Robotization is the use of robots to automate tasks that were previously performed by humans.

b) **Different applications of robotization in credit risk assessment:**

➤ **Data entry and processing:** Robots can be used to automate the entry and processing of loan applications, credit reports, and other data.

➤ **Credit decisioning:** Robots can be used to automate the decision-making process for loan applications.

➤ **Collections:** Robots can be used to automate the collections process for delinquent loans. (OECD, 2023)

IV. *Advantages and limitations of using AI in credit risk assessment*

A. Advantages

- **Improved accuracy:** AI algorithms can be trained on large datasets of historical loan performance to identify patterns that are difficult for human analysts to detect. This can lead to more accurate credit risk assessments. (Barr, 2020)

- **Reduced bias:** AI algorithms can be designed to be more objective and less prone to human biases than traditional credit scoring methods. This can help to ensure that all borrowers are treated fairly. (Barr, 2020, p. 1389)

- **Increased efficiency:** AI algorithms can automate many of the tasks involved in credit risk assessment, freeing up human analysts to focus on more complex tasks. This can lead to significant efficiency gains for lenders. (Barr, 2020, p. 1391)

B. Limitations

- **Complexity:** AI algorithms can be complex and difficult to understand, even for experts. This can make it difficult to audit and monitor AI-powered credit risk assessment systems (Barr, 2020, p. 1393)

- **Data requirements:** AI algorithms require access to large datasets of high-quality data to be effective. This can be a barrier for smaller lenders or lenders that specialize in niche markets. (Barr, 2020, p. 1395)

- **Regulatory uncertainty:** The regulatory landscape for AI in credit risk assessment is still evolving. This can create uncertainty for lenders and make it difficult to invest in AI solutions. (Barr, 2020, p. 1397)

V. *Future prospects for using AI in credit risk assessment*

"Credit risk assessment is a critical process for financial institutions, as it allows them to determine the likelihood that a borrower will default on a loan." (Institute, 2018)

A. Current trends

"The use of AI in credit risk assessment is on the rise." (Accenture, 2020)

According to a study by McKinsey Global Institute, AI could save financial institutions up to 20% of the costs associated with credit risk assessment. (Institute, 2018, p. 1)

B. Future outlook

- **AI has the potential to transform credit risk assessment in several ways, including:** (Institute, 2018, p. 1)
- **Improved model accuracy:** AI allows financial institutions to use larger and more complex datasets to develop more accurate models. (Institute, 2018)
- **Reduced costs:** AI allows financial institutions to automate manual tasks, which reduces costs. (Accenture, 2020, p. 3)
- **Improved speed:** AI allows financial institutions to assess credit risk more quickly. (Gartner, 2022, p. 1)
- **Improved inclusion:** AI allows financial institutions to access more diverse data, which can lead to improved financial inclusion. (Gartner, 2022)

VI. *SWOT analysis for the use of AI in credit risk assessment*

i. Strengths

- **Improved accuracy:** AI models can learn from historical data to identify patterns and associations that are not easily identifiable by traditional credit risk assessment methods. This can lead to more accurate credit scores that can help lenders make better lending decisions.
- **Increased efficiency:** AI models can automate many of the tasks involved in credit risk assessment, such as data collection and analysis. This can free up lenders to focus on other tasks, such as building relationships with customers.

- Reduced fraud: AI models can be used to detect and prevent fraudulent loan applications. For example, AI models can be trained to identify patterns of suspicious behavior, such as multiple loan applications from the same IP address.
 - Increased access to credit: AI models can be used to develop new credit scoring models that can assess the creditworthiness of borrowers with thin or no credit files. This can help expand access to credit to underserved populations, such as young people and immigrants.
 - Ability to identify patterns and trends: AI can identify patterns and trends in data that are not visible to the naked eye. This can help financial institutions make more informed credit decisions.
 - Ability to automate tasks: AI can be used to automate tasks such as credit history checks and customer scoring. This can free up analysts' time to focus on more complex tasks.
- ii. Weaknesses
- Bias: AI can be biased by the data on which it is trained. It is important to take steps to mitigate this bias, such as collecting and using diverse data.
 - Errors: AI is still an emerging technology, and there is a risk of errors. It is important to implement quality controls to identify and correct errors.
 - Cost: AI can be expensive to implement and maintain. Financial institutions need to be prepared to invest in AI to get the most out of it.
- iii. Opportunities
- Use of AI for predictive analytics: AI can be used to develop models that can identify borrowers who are at risk of default before they actually do. This could help lenders take steps to prevent defaults, such as offering payment assistance or restructuring loans.
 - Use of AI for real-time analytics: AI can be used to monitor the financial health of borrowers in real time. This could help lenders identify and address potential problems before they become too serious.
 - Improved credit decision accuracy: AI can be used to improve the accuracy of credit decisions. This can reduce losses for financial institutions and improve customer satisfaction.
 - Personalization of credit offers: AI can be used to personalize credit offers to meet the individual needs of customers. This can improve customer satisfaction and increase revenue for financial institutions.
 - Development of new financial products and services: AI can be used to develop new financial products and services that meet the needs of customers. This can create new growth opportunities for financial institutions.
- iv. Threats
- Regulation: Regulation may limit the use of AI in the financial sector. Financial institutions need to be aware of applicable regulations in order to comply.
 - Competition: Competition may increase, making it difficult for financial institutions to differentiate

themselves. Financial institutions need to invest in innovation to stay competitive.

- Cyberattacks: Cyberattacks can compromise AI systems, which can lead to financial losses. Financial institutions need to implement security measures to protect their AI systems.
- Development of new types of fraud: Fraudsters could develop new techniques to circumvent AI systems.

VII. CONCLUSION

Artificial intelligence (AI) has the potential to significantly improve credit risk assessment. It can enable financial institutions to better understand the risks associated with a loan or investment, and make more informed decisions about whether to grant them.

However, the use of AI in credit risk assessment also raises some challenges, including model complexity, bias, and security. Financial institutions must be aware of these challenges and put measures in place to mitigate them.

A. Perspectives

AI is a rapidly evolving technology, and its applications in credit risk assessment are still under development. Financial institutions must keep abreast of the latest AI innovations to stay competitive and improve their performance.

B. Recommendations

- Invest in research and development: Financial institutions should invest in research and development to improve AI technologies. This will allow them to get the most out of AI and stay ahead of the competition.
- Implement risk management measures: Financial institutions should implement risk management measures to mitigate the risks associated with the use of AI. This includes implementing quality controls to identify and correct errors, and implementing security measures to protect AI systems from cyberattacks.
- Train staff: Financial institutions should train their staff on AI technologies. This will allow staff to understand the benefits and risks of AI and use it effectively.

By following these recommendations, financial institutions can make the most of AI in credit risk. AI has the potential to transform this field, helping financial institutions make more informed credit decisions, personalize credit offerings, and develop new financial products and services.

VIII. REFERENCES

1. Accenture. (2020). *The future of credit risk assessment: The rise of AI and machine learning*. Londres, Royaume-Uni : Accenture.
2. Barr, M. S. (2020). The impact of artificial intelligence on credit risk assessment. *Harvard Law Review*, 133(6), 1387.
3. BCBS. (2010). *The Basel III framework for credit risk*. Basel, Switzerland: Bank for International Settlements.
4. Cai, Y. L. (2023). Introduction to AI and its impact on the banking sector. *Financial Innovation*, 9(1), 1 à 15.
5. China, D. (2023, June 28). Deloitte Insights. 2(2), 1-5. doi:10.1016/j.deliint.2023.06.002
6. Gartner. (2022). Gartner predicts 60% of organizations will implement AI-powered credit risk decisioning by 2024,. Stamford, CT, USA.
7. Hasan, I. e. (2023). Overview of different credit risk assessment methods. *International Journal of Financial Studies*, Volume 11(1), 1 à 15.
8. Institute, M. G. (2018). services, Artificial intelligence in financial.
9. Mishkin, F. S. (2012). *The economics of money, banking, and financial markets* (éd. 11ème). (Addison-Wesley, Éd.) Boston, MA.
10. 10. OECD. (2023, July 20). OECD Insights. 2(3), 1-5. doi:10.1016/j.deliint.2023.07.001

Financial forecasting using an ARIMA model

KARA TERKI Nesrine
Higher School of Management-
Tlemcen
Random Statistics and Modeling
Laboratory, Tlemcen Algeria
kara.nesrine@yahoo.fr

Abstract— Sales forecasts are often the most important element of the business plan. They reflect the company's ability to innovate, sell and produce. They are the reflection, in money, of all the company's objectives and strategies. Artificial intelligence plays a key role in making these forecasts since they are based on iterative processes that continually consider new data and improve over time. For this purpose, a modeling test will be proposed using the Autoregressive Integrated Moving Average model and we will use it and the software R to make sales predictions.

Keywords— Artificial intelligence, Sales prediction, Autoregressive Integrated Moving Average model.)

I. INTRODUCTION

Financial forecasting is the process by which a company determines its expectations for future results. When a company carries out its financial forecasts, it seeks to provide itself with the means to express its objectives and priorities in order to ensure their internal consistency.

Forecasting can also help a business identify the assets or budget needed to achieve its goals and priorities.

Artificial Intelligence -powered financial forecasting is an iterative process that continually considers new data and improves over time. Machine learning algorithms can adapt and refine their predictions based on feedback and validation of actual results.

A common example of financial forecasting is forecasting a company's sales. Since most financial statement accounts are related to sales, sales forecasting can help a company make other financial decisions that help it achieve its goals. However, if sales were to increase, the resulting expenses to produce the additional sales would also increase. Each forecast has an impact on the overall financial situation of the company.

A sales forecast is a prediction of the quantity of products a company will sell in the future (see Maricourt, Rd).

A forecast is different from sales goals, which are the sales a business hopes to achieve. Forecasts help the business determine its direction. It is in calculating the financial impact of these forecasts that financial modeling comes into play.

Financial modeling takes financial forecasts and builds a predictive model that helps a company make sound business decisions. It influences many decisions: recruitment, resource management, definition of objectives, budget, etc. It allows anticipation of commercial unforeseen events and the implementation of appropriate solutions.

Financial modeling is the process by which a company constructs its financial representation. The created model is used to make business decisions. Financial models are the mathematical models developed by a company in which the variables are linked together.

There are two categories of models to account for a time

series. The first considers that the data is a function of time ($y = f(t)$). This category of model can be fitted by the least squares method, or other iterative methods.

A second category of models seeks to determine each value of the series according to the values which precede it ($y_t = f(y_{t-1}, y_{t-2}, \dots)$). This is the case of ARIMA ("AutoRegressive – Integrated – Moving Average") models. This category of models was popularized and formalized by Box and Jenkins (1976). Financial forecasting and modeling can be used in budgeting, investment research, project financing and capital raising.

Although forecasting is a simple process, it can become a source of problems. Indeed, poor sales forecasts can cause cash flow difficulties, increases in storage costs or even the risk of obsolescence of unsold items...

In this work we have the monthly sales of an automobile brand over a period from January 1976 to December 2018, the goal is to predict those of the year 2019 and compare them with actual sales. These monthly sales during year i , with $i = 1 \dots n$, $n = 43$, are denoted $y_i = (y_{i1}, y_{i2}, \dots, y_{i12})$, it is quite clear that y_i is a vector data.

II. DATA

The data chosen for this work are as follows: monthly car sales in a city in the United States over a period from January 1976 to December 2019. These data come in part from the R software.

In the extremely competitive case of the automobile industry, monthly sales are closely monitored by Top Management because this information is directly correlated with turnover and market share. The data table and graph are given in Figure 1.

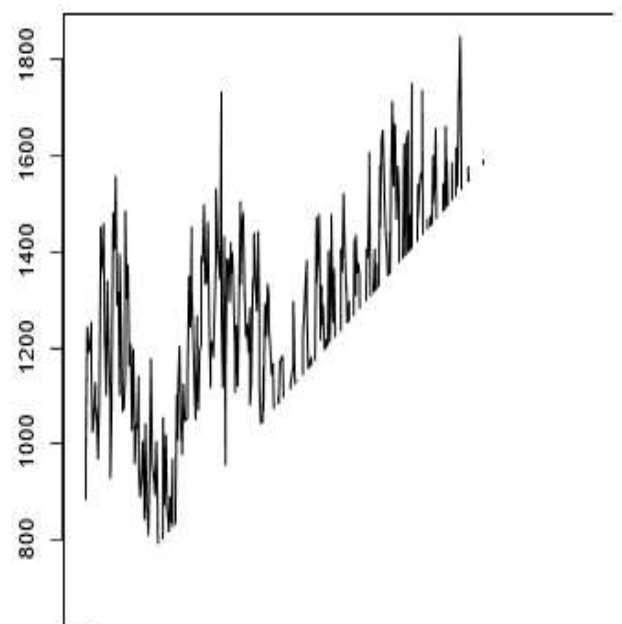


Figure 1: USV sales

A. Data validation

a. Reserved data:

The data reserved for the forecast are those of the last 12 months of the series. These therefore extend from January 2019 to December 2019.

b. Validation criteria:

To validate our predictors, we will use the criteria MSE (also denoted eq) mean square error and RMAE relative mean absolute error. These are given by the general formulas:

$$MSE = 1/m \sum_{i=1}^m |\hat{y}_{n+1}(t_i) - y_{n+1}(t_i)|^2$$

$$RMAE = \frac{1}{m} \sum_{i=1}^m |\hat{y}_{n+1}(t_i) - y_{n+1}(t_i)| / |y_{n+1}(t_i)|$$

- n: number of data =12

- m number of discretizations=12

- $\hat{y}_{n+1} = \hat{y}_3$: predicted data

- $y_{n+1} = y_{44}$: reserved data

B. Relevance of the choice of criteria:

The MSE criterion can be used for positive and negative errors. It has the advantage of keeping the same order of magnitude as the data. This index provides an indication of the dispersion or variability of the quality of the prediction. The MSE can be related to the variance of the model.

Often, MSE values are difficult to interpret because one is not able to tell whether a variance value is low or high. To overcome this effect, it is more interesting to normalize the MSE so that this indicator is expressed as a percentage of the average value of the observations which we will note RMSE. This can be used to give more meaning to the indicator.

The RMAE criterion has the advantage that it is based on a percentage, which is practical to handle. We will use it to show the influence of the choice of indicator on the validation of the prediction model.

C. ACF representations:

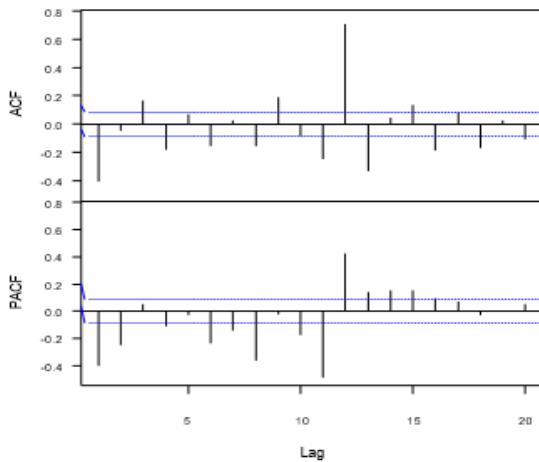


Figure 2: Time series: y_dif1

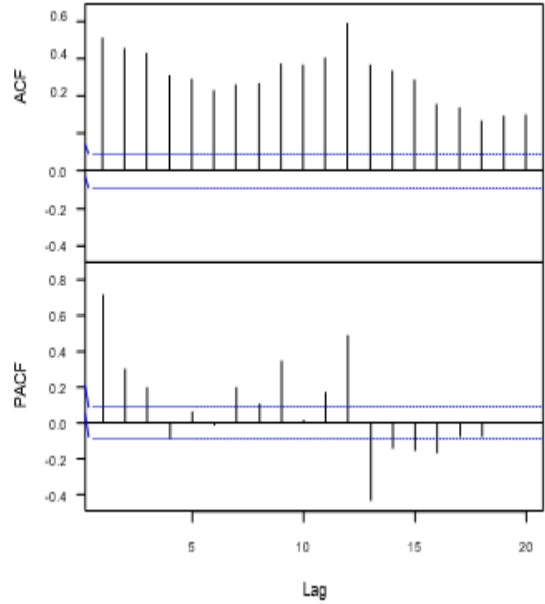


Figure 3: Time series: USVSales

The ACF output in the Figure 2 shows a slow decay towards 0, which reflects a non-stationarity problem. We therefore want to differentiate.

The ACF output in the Figure 3 of the series thus differentiated still presents a slow decrease towards 0 for the multiples of 12 but the lag 1 autocorrelation is negative, so the series does not need to be differentiated.

We cancel the first differentiation and this time we carry out the differentiation I - B12 .we obtain the following ACF:

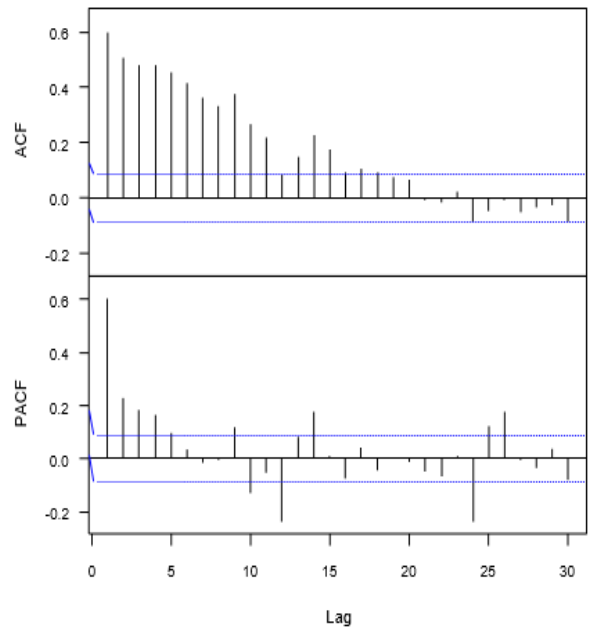


Figure 4: Time series: y_dif2

III. MODEL:

Based on the previous study then considering some ARIMA(p,q,d) models (P,Q,D) we choose using the principle of minimization of the AIC criterion, the following ARIMA model (1,0,3) (0,1,2) defined by

$$(I - a_1L)(I - L^{12})y_t = (I + m_1L + m_2L^2 + m_3L^3)(I + sm_1L^{12} + sm_2L^{24})e_t$$

Coefficients:

$$\begin{matrix} a_1 & m_1 & m_2 & m_3 & sm_1 & sm_2 \\ 0.9728 & -0.5871 & -0.1034 & 0.0649 & -0.5773 & -0.2420 \end{matrix}$$

$$\text{s.e. } 0.0123 \quad 0.0461 \quad 0.0501 \quad 0.0452 \quad 0.0430 \quad .0414$$

IV. PREDICTION :

Using the R software and considering the model proposed above, we can forecast sales for the year 2019, the results of this forecast are given in the following table and the graphs below:

Table I: difference between sales / sales prediction for the year 2019.

\hat{y}_4 :sales prediction	y_{44} : sales	Difference
1228.144	1171.503	56.641
1372.168	1288.278	83.890
1676.122	1642.750	33.372
1479.488	1372.659	106.829
1644.233	1628.074	16.159
1585.640	1554.748	30.892
1493.603	1443.947	49.656
1578.758	1685.342	-106.584
1453.505	1315.632	137.873
1412.431	1380.180	32.251
1392.667	1446.483	-53.816
1625.936	1565.023	60.913

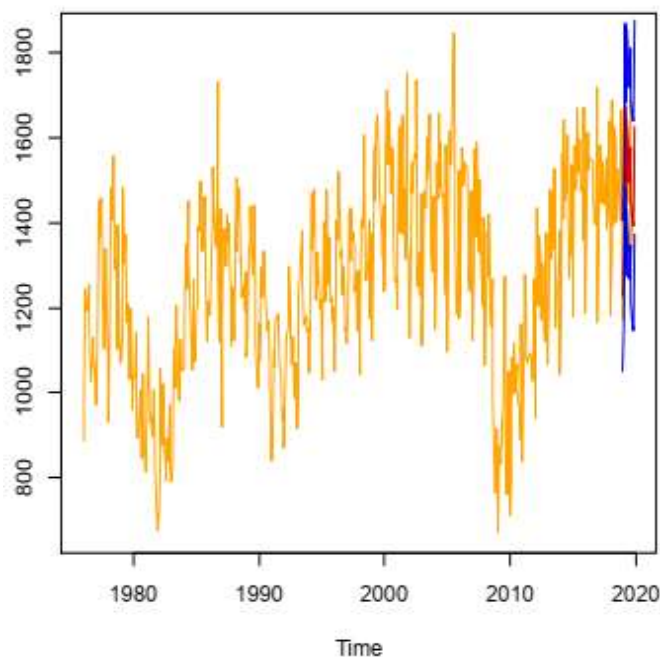


Figure 5: Time series: USV Sales

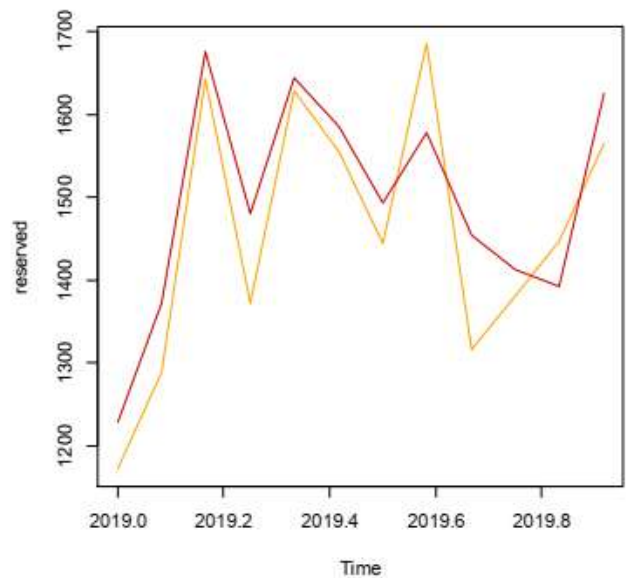


Figure 6 : comparison between actual sales made during the year 2019 and the predicted values using the proposed model

Calculation of validation criteria:

$$\text{root}(MSE) = 73.29666$$

$$RMAE = 0.04527691$$

The interpretation of the error criteria depends on the series and the forecast quality required. In the present case, a RMAE of 4.5% seems satisfactory a priori.

V. CONCLUSION

In the previous graph, the red curve represents the actual sales made during the year 2019 and the orange one represents the predicted values using the proposed model. We notice that the two curves are quite close and similar. What is even more satisfactory are the very low values of the RMAE and root criteria of MSE (eq) equal to 4.53% and 73.297 respectively. Remember that the eq criterion has the advantage of keeping the same order of magnitude as the data, by expressing it as a percentage of the average value of the observations we obtain RMSE=5% which is a fairly low value.

VI. REFERENCES

- [1] Box, G.E.P. & Jenkins, G.M. (1976). Time series analysis : Forecasting and control. Oakland, CA : Holden-Day.
- [2] Maricourt, Rd. La prévision des ventes, Edition PUF, Paris,1985, p.15.
- [3] Sylvain Rubenthaler(Université Nice Sophia Antipolis). Séries chronologiques (avec R),(Cours et exercices) Master1 IM, 2016-2017
- [4] Regis Bourbonnais, Michel Terraza. Analyse des Séries temporelles : application à l'économie et à la gestion. Ed Dunod, Paris 2004..

Usage of artificial intelligence in online shopping: Algerian customer attitude

Hamza Fatima Zohra
PhD. dept. of commercial sciences
Higher school of management
Tlemcen, Algeria
hamzafatimazohra96@hotmail.com

Khedim Amel
PhD. dept. of commercial sciences
Higher school of management
Tlemcen, Algeria
amelkhedim@yahoo.fr

Abstract: *This study investigates the use of Artificial Intelligence (AI) in Algerian online shopping applications. Data was collected from 220 respondents aged 18-35 on Jumia, Ouedkniss and Facebook Marketplace. The Technology Acceptance Model (TAM) was applied, and independent variables such as perceived usefulness, ease of use, trust, and performance were analyzed through SPSS. Results showed a significant relationship between perceived usefulness, perceived ease of use, trust, performance and attitude. Keywords—Artificial Intelligence (AI), Attitude, Online Application.*

I. INTRODUCTION

In today's digital age, where online shopping is the standard, every brand must work to make their product or service stand out in the sea of choices available to consumers. This involves the use of Artificial Intelligence (AI) to customize the buying process, making it smooth and efficient, from gathering customer information to processing payments. This results in the creation of a distinct brand identity and added value for consumers.

Artificial Intelligence (AI) is a multidisciplinary field dedicated to creating intelligent machines that mimic human actions and responses. It has applications in various areas, such as speech recognition, learning, planning, and problem-solving. The incorporation of AI enhances user experiences by tailoring them to individual preferences, potentially enabling search results based on user thoughts or sensory cues.

In the current digital landscape, where online purchases are the norm, every brand must strive to ensure their product or service stands out among the numerous options available to consumers. This involves utilizing Artificial Intelligence (AI) to customize the purchasing process, ensuring it is seamless and efficient, from gathering customer information to completing payments. This approach results in the development of a unique brand identity and increased value for consumers.

AI plays a crucial role in brand development, particularly in search engines. When users input a product or service into a search engine, AI presents a range of options and predictions, sometimes deviating from the expected outcomes. This phenomenon is especially relevant in the e-commerce sector, where brands aim to improve visibility,

accuracy in presenting product information, and customer engagement. Ultimately, these efforts lead to enhanced brand value and customer satisfaction.

The main purpose of this research is to investigate the utilization of Artificial Intelligence in online shopping applications and its impact on Algerian consumer attitudes. Additionally, the objectives of the study can be further delineated as follows:

1. Identifying the factors that influence Algerian consumers' attitudes toward AI.
2. Developing a research model based on the Technology Acceptance Model (TAM) to define the factors affecting consumer attitudes toward AI.
3. Investigating the factors incorporated in the proposed research model.

Research Questions:

RQ1: What are the factors affecting Algerian consumers' attitude towards AI?

RQ2: Which research model could define the factors that affect the consumers' attitude towards AI?

II. THE CONCEPT OF ARTIFICIAL INTELLIGENCE MARKETING

I. Definition of Artificial Intelligence Marketing

Gijs Overgoor and associates (2019, p. 157) define marketing AI as the "creation of simulated entities that, leveraging their understanding of consumers, competitors, and the core company, suggest and/or execute marketing plans to attain the most advantageous marketing outcomes." [1] The integration of AI into marketing is anticipated to gain significant traction [2], offering a diverse array of potential applications for AI within the marketing domain [3]. Van Esch (2018) delineates AI as a multifaceted concept, viewed through the prism of human-computer interaction. AI is already exerting a substantial influence across nearly all sectors of Marketing, ranging from enhancing customer experiences to streamlining marketing operations and even affecting pivotal business decisions, albeit to differing degrees [4].

II. Online Shopping Application

Traditionally, people have used shopping lists written on paper, usually composed at home, to refresh their memory when shopping. Only five percent of paper-based shopping lists are edited in the supermarket.[5] Furthermore, the transition from paper-based to digital shopping lists has been slow.[6]

The rapid growth of online shopping has led to intense competition in the industry[7]. Factors influencing customer attitudes include trust in sellers, the ability to see products before purchase, a wide range of products, and convenient and efficient shopping anytime from anywhere. Studies have shown that both online and mobile experiences, including smartphone usage, are positively related to the ownership of an online shopping application[8]. Research by Al Dmour et al. (2014)[9] also indicates that consumers are more likely to purchase when mobile app quality is highly perceived.

-Jumia: Jumia is a global corporation Algerian e-commerce site that was founded in 2014 with the goal and vision of being the only point of sale for retail in Africa via the adoption of best practices both online and offline. Jumia is the site with the greatest selection in Algeria, and it guarantees delivery in 43 wilayas. It has established Algeria's first e-commerce academy to help new entrepreneurs enhance their talents. Jumia is the top internet platform in Africa, and the first in Algeria, specialized in online shopping and sales, providing an easy online and mobile shopping platform, and simplifying life with simple internet apps and intelligent solutions.[10]

-Ouedkniss: The Oued Ness site is regarded one of Algeria's most successful initiatives, since it is the fourth most visited site in Algeria, according to Alexa rankings beginning in February 2022 (Datareportal, 2022). It is a website that specializes in facilitating business transactions between... Individuals and certain organizations, so that supply matches demand; in its most basic form, it simulates the global eBay experience. Lougoud was founded in 2006 by five students Friends, and has since grown to 38 million visitors in August 2010, at an average of 133,000 visits each day.[11]

-Facebook Marketplace: As a popular social media platform in the community, Facebook's marketplace is frequently used by its users to conduct online buying and selling operations.[12] Facebook Marketplace is the social network's classified-ad area that specializes on assisting individuals and companies in selling products locally. Marketplace is Facebook's entry into new markets in order to compete with businesses such as eBay and Craigslist. [13]

III. HYPOTHESES DEVELOPMENT

I. Relationship between consumer attitude and usage of artificial intelligence in online shopping application

To understand the acceptance towards AI in online shopping application in the context of Algerian consumers, we developed a research model based on the TAM model (Technology Acceptance Model) developed by F. Davis in 1989[14] which is able to identify the factors that affect the consumers' attitude towards AI, naming: Perceived Usefulness, Perceived Ease of Use, Perceived Trust, Perceived Performance/ Quality that will be taken for this study as independent variables, and Attitude as the dependent variable.

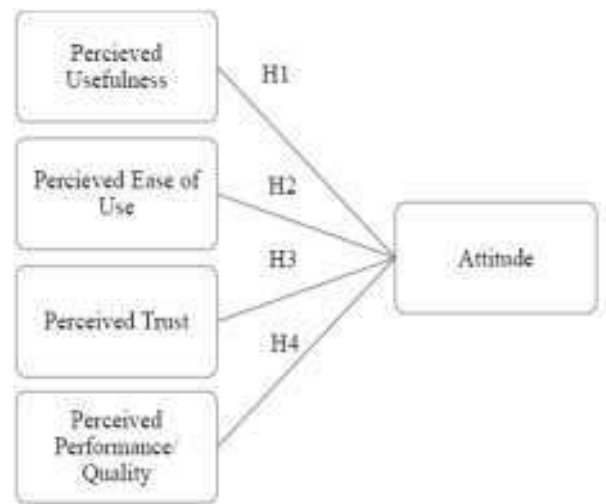


Figure 1. Proposed Conceptual Framework (Source: Develop for the research)

II. Hypothesis

- H1: There is significant positive relationship between perceived usefulness and consumer attitude.
- H2: There is significant positive relationship between perceived ease of use and consumer attitude.
- H3: There is significant positive relationship between perceived trust and consumer attitude.
- H4: There is significant positive relationship between perceived performance/ quality and consumer attitude.

IV. METHODOLOGY

This study will implement the quantitative research, by conducting a questionnaire. 220 copies of survey questionnaire will be distributed to target respondents through Google Form online. The data collected is analysed through SPSS software.

V. RESULTS AND DISCUSSION

The survey included 115 female respondents out of 220, with 47.65% being male. The target age group was 18-35, with 70.25% of respondents aged 18-29, followed by 29.75% aged 30-35. The majority of respondents were female, with a smaller proportion of male respondents. For the educational background, 47.8% of the respondents hold Master's degree. The majority of the respondents (70.15%) selected Facebook Marketplace as their most common used online shopping application and only 17.65% and 12.2% selected Jumia and Ouedkniss respectively. This shows that Facebook Marketplace has the most active users in Algeria currently. 38.05% of the respondents spend between 2000dza to 5000 on online shopping a month and 27.2% spend less than 2000dza a month, while 14.75% spend more than 10000dza a month.

The results show that 141 individuals selected 'Yes' confirming their enjoyment or preference towards AI assistance while using online shopping application while 69 individuals selected 'No'.

Table 1. Reliability Statistics for Actual Research

No.	Construct	Cronbach's Alpha	No. of items	Strength of Association
1	Attitude	0.871	3	Very Good Reliability
2	Perceived Usefulness	0.838	3	Very Good Reliability
3	Perceived Ease of Use	0.890	3	Very Good Reliability
4	Perceived Trust	0.703	3	Good Reliability
5	Perceived Performance	0.898	2	Very Good Reliability

Source: Develop for the research

Table 1. indicates results for test of reliability which is equally to the value of Cronbach's alpha. As an illustration, 4 constructs have very good reliability, as Perceived Performance has the highest alpha coefficient of 0.898, Perceived Ease of Use with 0.890. Besides that, constructs that have very good reliability including perceived usefulness as the Cronbach's alpha value of 0.838; while Attitude have 0.871. Lastly, there is one construct with good reliability, which is Perceived Trust with 0.703.

Table 2. Correlations

		Attitude	PT	PEU	PP	PU
Attitude (A)	Pearson Correlation	1	.362**	.649**	.531*	.486*
	Sig.		.000	.000	.000	.000
	N	220	220	220	220	220
Perceived Trust (PT)	Pearson Correlation	.362**	1	.471**	.327*	.335*
	Sig.	.000		.000	.000	.000
	N	220	220	220	220	220
Perceived Ease of Use (PEU)	Pearson Correlation	.649**	.471**	1	.527*	.484*
	Sig.	.000	.000		.000	.000
	N	220	220	220	220	220
Perceived Performance (PP)	Pearson Correlation	.531**	.327**	.527**	1	.327*
	Sig.	.000	.000	.000		.000
	N	220	220	220	220	220
Perceived Usefulness (PU)	Pearson Correlation	.486**	.335**	.484**	.327*	1
	Sig.	.000	.000	.000	.000	
	N	220	220	220	220	220

** Correlation is significant at the 0.01 level

Table 2. indicates the result of Pearson Correlation analysis among both independent variables (perceived usefulness, perceived ease of use, perceived trust, and perceived performance), and the dependent variable (attitude).

In accordance with table 2, the Pearson Correlation falls between 0.362 to 0.649, together with the significant for the entire variables is 0.000. On that account, this proves that

each independent variables and dependent variable possesses statically linear relationship seeing that it was positively correlated.

-Relationship between Perceived Usefulness and Attitude

Pearson correlation of Perceived Usefulness and Attitude was found to be low positive and statistically significant ($r = 0.486$, $p < 0.001$)

Table 2. interprets the p-value of perceived usefulness ($p = .000$) is lower than the significant level of 0.01. Therefore, the hypothesis is supported as the p-value is less than 0.01

-Relationship between Perceived Ease of Use and Attitude

Pearson correlation of Perceived Ease of Use and Attitude was found to be moderate positive and statistically significant ($r = 0.649$, $p < 0.001$)

Table 2. interprets the p-value of perceived ease of use ($p = .000$) is lower than the significant level of 0.01. Therefore, the hypothesis is supported as the p-value is less than 0.01

-Relationship between Perceived Trust and Attitude

Pearson correlation of Perceived trust and Attitude was found to be low positive and statistically significant ($r = 0.362$, $p < 0.001$)

Table 2. interprets the p-value of perceived trust ($p = .000$) is lower than the significant level of 0.01. Therefore, the hypothesis is supported as the p-value is less than 0.01.

-Relationship between Perceived Performance and Attitude

Pearson correlation of Perceived Usefulness and Attitude was found to be moderate positive and statistically significant ($r = 0.531$, $p < 0.001$)

Table 2. interprets the p-value of perceived usefulness ($p = .000$) is lower than the significant level of 0.01. Therefore, the hypothesis is supported as the p-value is less than 0.01

Conclusion

Because of the rapid advancement of digital technology, online shopping has transformed. Artificial Intelligence (AI) has been increasingly used in online commerce in recent years, since it is an effective tool for meeting fast changing consumer demand and increasing sales efficiency

In a nutshell, this study is completed through the guidance of TAM, as it provides a new insight on perceived usefulness, perceived ease of use, perceived trust, and perceived performance towards attitude. Moreover, the results shows that there is a positive significant relationship between perceived usefulness, perceived ease of use, perceived trust, and perceived performance towards attitude

REFERENCES

- [1] Chintalapati, S. and S.K.J.I.J.o.M.R. Pandey, Artificial intelligence in marketing: A systematic literature review. 2022. 64(1): p. 38-68.
- [2] Vishnoi, S.K., et al., Marketing intelligence and firm performance: Reviewing the mediating impact of customer relationships, customer satisfaction and customer loyalty. 2019. 18(2): p. 555-570.
- [3] GABELAIA, I.J.J.o.M.R. and C. Studies, The applicability of artificial intelligence marketing for creating data-driven marketing strategies. 2022. 2022(466404).
- [4] Hildebrand, C.J.N.M.I.R., The machine age of marketing: How artificial intelligence changes the way people think, act, and decide. 2019. 11(2): p. 10-17.

- [5] Block, L.G. and V.G.J.J.o.C.P. Morwitz, Shopping lists as an external memory aid for grocery shopping: Influences on list writing and list fulfillment. 1999. 8(4): p. 343-375.
 - [6] Heinrichs, F., D. Schreiber, and J. Schöning. The hybrid shopping list: Bridging the gap between physical and digital shopping lists. in Proceedings of the 13th international conference on Human computer interaction with mobile devices and services. 2011.
 - [7] Pratminingsih, S.A., et al., Factors influencing customer loyalty toward online shopping. 2013. 4(3): p. 104-110.
 - [8] Kim, M., et al., Mobile shopping through applications: Understanding application possession and mobile purchase. 2017. 39: p. 55-68.
 - [9] Al Dmour, H., M. Alshurideh, and F.J.L.S.J. Shishan, The influence of mobile application quality and attributes on the continuance intention of mobile shopping. 2014. 11(10): p. 172-181.
 - [10] Belkalem, L. and D. Adjaoud, Influence du marketing mobile sur le comportement d'achat du consommateur: cas de la boutique en ligne Jumia Algérie. 2018, Université Mouloud Mammeri.
- ع. الكريم. حماية المعلومات الشخصية للمستهلك الإلكتروني:
دراسة مقارنة بين زهيو [11]. الجزائري. 2022
Ouedkniss العالمي وموقع Ebay موقع
- [12] Piranda, D.R., et al., Online Marketing Strategy In Facebook Marketplace As A Digital Marketing Tool. 2022. 1(3): p. 1-8.
 - [13] TechTarget Contributor. Facebook Marketplace. 2018 July 18; Available from: <https://www.techtarget.com/whatis/definition/Facebook-Marketplace#:~:text=Facebook%20Marketplace%20is%20classified%20ad,services%20like%20eBay%20and%20Craigslist.>
 - [14] Davis, F.D.J.A.-S., MN, Al-Aufi, AS: Information Seeking Behavior and T. Adoption, Technology acceptance model: TAM. 1989: p. 205- 219.

How AI is helping marketers and advertisers?

Rahali Amina Souad
Higher school of management
Tlemcen Algeria
Souad_rahali@hotmail.fr

Tabet Derraz Imane
Higher school of management
Tlemcen Algeria
tabetderrazimane@yahoo.fr

Abstract—marketing uses artificial intelligence technologies to make automated decisions based on data collection, data analysis, and additional observations of audience or economic trends that may impact marketing efforts. AI is often used in digital marketing efforts where speed is essential. AI marketing tools use data and customer profiles to learn how to best communicate with customers, then serve them tailored messages at the right time without intervention from marketing team members, ensuring maximum efficiency. This article proposes a research agenda that addresses how marketers use artificial intelligence to gather insights into their customers and produce better content.

Keywords—marketing, artificial, intelligence, decision, customers

I. INTRODUCTION

It's evident that Artificial Intelligence (AI) is rapidly evolving and is poised to become a fundamental component of commercial entities worldwide. This is due to significant trends in AI-driven automation that are reshaping the landscape of AI adoption by enterprises.

Indeed, the impact of artificial intelligence (AI) on various industries, including marketing strategies, business models, sales processes, and customer service options, is a significant and evolving trend.

In an analysis of more than 400 AI use cases, across 19 industries and 9 business functions, McKinsey & Co. indicates that the greatest potential value of AI pertains to domains related to marketing and sales [1], through impacts on marketing activities such as next-best offers to customers [2], programmatic buying of digital ads [3], and predictive lead scoring [4]. This article tries to discuss the current state and future potential of using Artificial Intelligence (AI) in marketing. It highlights the need for a framework to understand AI's role in marketing and suggests areas where AI is expected to have a significant impact.

II. A FRAMEWORK FOR UNDERSTANDING UTILITIES OF AI IN MARKETING

AI is a computer science technology that teaches computers to comprehend and emulate human communication and behaviour. Based on the data provided, AI has created a new intelligent machine that thinks, responds, and performs jobs the same way people do. AI can do highly technical and specialised activities such as robotics, speech and picture recognition, natural language processing, problem-solving, etc. AI is a collection of several technologies capable of executing tasks that need human intelligence. When applied to standard commercial

processes, these technologies can learn, act, and perform with human-like intelligence. It simulates human intelligence in machines, saving us time and money in business transactions.

Marketing literature related to AI is relatively sparse, prompting this effort to propose a framework that describes both where AI stands today and how it is likely to evolve. Marketers plan to use AI in areas like segmentation and analytics (related to marketing strategy) and messaging, personalization and predictive behaviors (linked to customer behaviors) [5]. Thus, we also propose an agenda for future research, in which we delineate how AI may affect marketing. In so doing, we respond to mounting calls that AI be studied not only by those in computer science, but also studied by those who can integrate and incorporate insights from psychology, economics and other social sciences [6].

[7] defines marketing AI as “the development of artificial agents that, given the information they have about consumers, competitors, and the focal company, suggest and/or take marketing actions to achieve the best marketing outcome.” AI-marketing fusion will certainly grow in stature [8], and there will be more possibilities of the application of AI to marketing [9], [10], [11]. [12] defines AI as a multifaceted concept through the lens of a human-computer interaction. From customer experience to marketing operations and up to business decision-making, AI is already affecting almost all the functional themes of Marketing at varying degrees of severity.

AI has been a game-changer in almost every corner of the digital marketing world, from content marketing to social media marketing.

Although AI technologies have been used in marketing strategies for more than a decade, the new generation of natural language processing tools like ChatGPT has brought AI tech into the limelight.

An AI marketing tool will generally use machine learning to understand repetitive tasks like content generation for blogs or social media posts by analyzing large amounts of data. Automating tasks like this can help increase your ROI by helping you work more efficiently and reach more potential customers.

From content generation and data analysis to smart automation and sales forecasting, AI provides numerous opportunities that help businesses enhance their marketing activities, improve customer experience, and boost sales. (Fig 1).

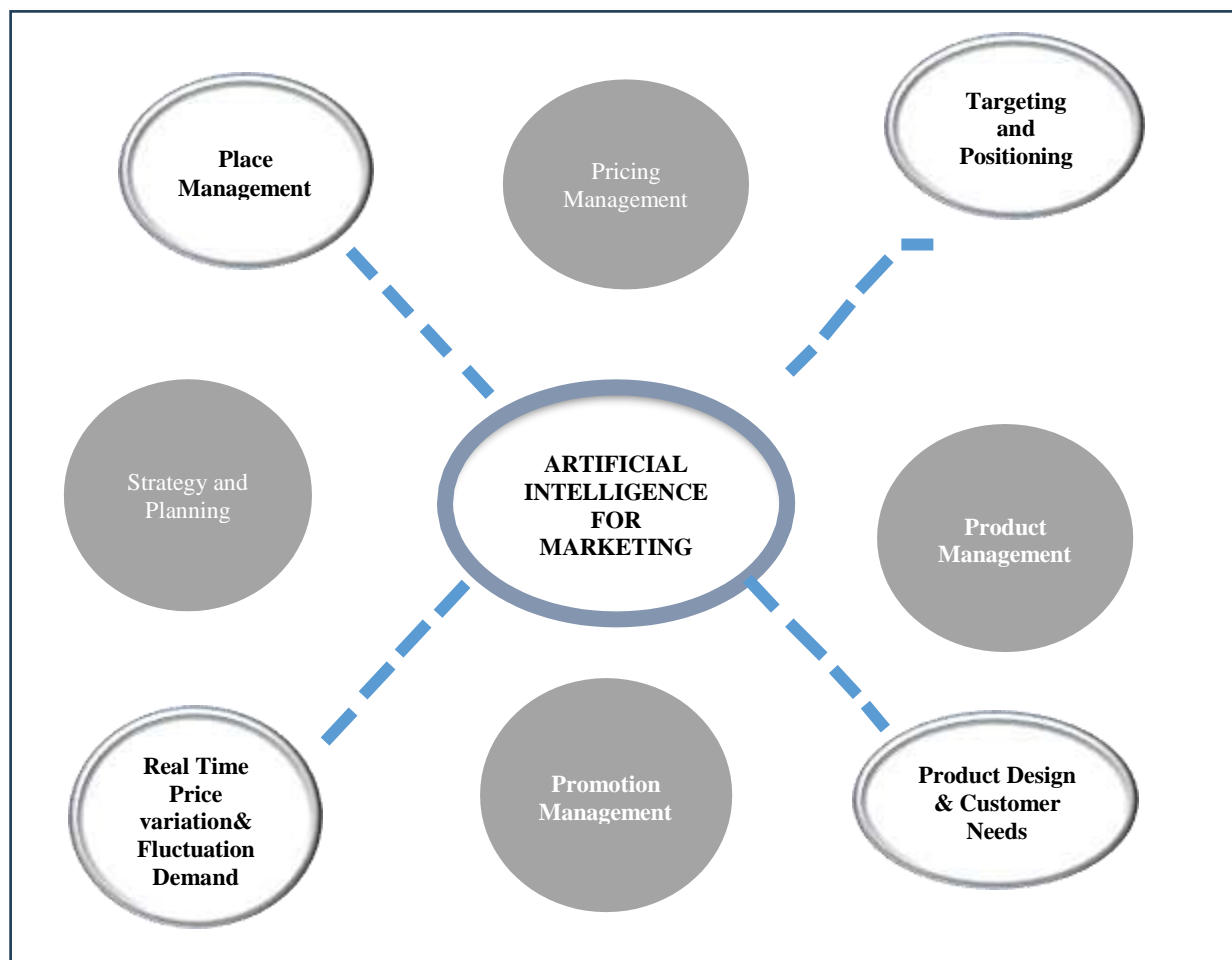


Fig. 1. Several Segments for AI applications in Marketing Domain

A. Artificial intelligence technologies in marketing field

Powerful social marketing platforms, like Sprout, weave together sophisticated AI technologies under the hood to provide the insights you need to succeed. Capabilities such as semantic classification, named entity recognition and aspect based sentiment analysis help you get targeted insights specific to your industry, while natural language processing helps you optimize social content and improve customer engagement—all leading to greater competitive advantage and share of voice.

AI technologies improve marketing efforts in several fields, here are the main known :

- Machine learning

Machine learning (ML) uses statistical methods to analyze social data for high-precision insights around customer experience, audience sentiment and other marketing drivers. Once trained, ML models automatically complete text mining, topic extraction, aspect classification, semantic clustering and other tasks to provide results in seconds.

AI-ML models get smarter as they process more data over time and so upgrade automatically, which is perfect for scaling your business operations while minimizing future investment in your tech stack.

- Natural language processing (NLP)

Natural language processing powers your AI marketing tool so it can semantically and contextually understand social listening data. It combines rules-based lexical and statistical methods, enabling you to scan a wide range of posts, messages, reviews or comments and extract critical information from it.

When NLP algorithms are coded for social listening, they can interpret the data even if it's splattered with colloquialisms, code switches, emojis, abbreviations, hashtags or spelling

mistakes. Natural language generation (NLG) further enhances the tool's capabilities to help you create high-performing copy for posts, customer responses and more.

This gives you access to a wider audience for outreach campaigns, stronger communication with existing customers and better returns on our investment in social.

- Semantic search

Semantic search algorithms are critical in NLP because they help understand the intent of a phrase or lexical string without depending on keywords. These algorithms extract relevant keywords and categorize them into semantic clusters. This eliminates chances of duplicates in text mining, especially where sentiment analysis is concerned, for an accurate measure of customer experience or brand performance.

Knowing exactly how strong your brand is in relation to your competitors and monitoring it against your benchmarks can help you alter marketing and sales strategies to achieve long-term business goals.

- Named entity recognition (NER) and neural networks

NER helps an AI platform identify named entities in big data. These entities could be important people, places or things such as CEOs, celebrities, locations, currencies, businesses and others. It can identify these entities even if they are misspelled. NER also is a key function in generating knowledge graphs because they establish a relationship between entities in order to derive context and insights from data.

Neural network (NN) algorithms, built to mimic how a human brain handles information, remember these interconnected data points and keep adding them to their knowledge database. This is what enables ML models to provide more precise results with time through deep learning.

Thus, you get to know why certain brands keep appearing in your social listening data, what new market trends are brewing, which influencers would be a great fit and many other insights that can help you strengthen your social marketing strategy.

- Sentiment analysis

Sentiment analysis is the process of measuring customer sentiment from feedback data and can be instrumental in helping with reputation management. Sentiment analysis algorithms analyze social listening data including survey responses, reviews and incoming messages, both in real-time and historically. They measure sentiment in every aspect that is extracted from the data and assign polarity scores in the range of -1 to +1. Neutral statements are counted as zero.

When analyzing social data where customers are talking about aspects of a business, sentiment analysis models consider the polarity score of each aspect. The scores are aggregated to provide an overall sentiment of the brand in terms of customer experience. This eventually gives you an idea of how well your business is performing.

With such insights available, you can grow your brand by evaluating and improvising social media content, shaping sales and marketing, improving brand management, better interpreting customer intent and so much more.

B. The spread of AI in marketing

AI is primarily concerned with user retention and lead conversion in digital marketing. It can guide a user in the direction that aligns with the business's goals by using intuitive AI chatbots, intelligent email marketing, interactive web design, and other digital marketing services. Several factors determine the impact of AI on digital marketing. ML, a subset of AI, is concerned with computer programmes that access data and use it to learn independently. It compiles data from various places, including social media accounts, menus, online reviews, and websites. AI then uses the information to produce and deliver content relevant to the audience. AI

software enables in-depth online analysis of restaurants and their customers.

Several key ways in which marketers can leverage artificial intelligence (AI) to enhance their marketing efforts and achieve more effective results. Here's a breakdown of the key points made in the statements : (Fig 2)

-Deeper Consumer Insights: AI can be used to analyze vast amounts of consumer data, allowing marketers to gain deeper insights into consumer behaviors, preferences, and trends. This understanding can help marketers create more targeted and relevant marketing campaigns.

-Categorizing and Guiding Customers: AI can assist in categorizing customers based on their preferences and behavior, and then guide them to the next step in their customer journey. This personalized approach can improve the overall customer experience.

-Increasing ROI: By using AI to analyze consumer data and target marketing efforts more effectively, marketers can increase their return on investment (ROI). This means they can spend their resources on strategies that are more likely to yield positive results.

-Avoiding Ineffective Advertising: AI can help marketers avoid wasting resources on ineffective advertising strategies. By understanding what consumers truly want, marketers can tailor their campaigns to be more engaging and less irritating to clients.

-Personalization: AI plays a significant role in personalizing marketing efforts. This can include personalizing websites, emails, social media posts, videos, and other marketing materials to better respond to individual customer demands. Personalization often leads to higher customer engagement and satisfaction.

-Automation: AI can automate tasks that previously required human intervention, reducing the need for human labor resources and saving time. This increased efficiency can lead to cost savings and improved productivity.



Fig.2: The impact of AI on marketing

C. An example about the transportation industry

The impact of artificial intelligence (AI) on various industries, including marketing strategies, business models, sales processes, and customer service options, is indeed a significant and evolving trend. The transportation industry is one of the sectors where AI is expected to bring transformative changes. Here are some key points regarding the influence of AI in the transportation industry:

Driverless, AI-Enabled Cars: The development of autonomous vehicles is a prominent example of AI's impact on the transportation industry. These vehicles use AI algorithms and sensors to navigate and make decisions on the road. This technology has the potential to revolutionize transportation by reducing the need for human drivers, improving safety, and increasing efficiency.

Altered Business Models: The introduction of AI-enabled autonomous vehicles can lead to new business models and revenue streams. For example, companies may shift from selling individual cars to providing transportation as a service (TaaS) where customers can use self-driving cars on-demand. This could have a profound impact on car manufacturers, rental companies, and ride-sharing services.

Changes in Customer Behavior: The widespread adoption of autonomous vehicles can lead to changes in customer behavior. People might be more willing to use on-demand transportation services, reducing the need for car ownership. Commuters may choose to work or relax during their journeys, which can impact other industries like entertainment and productivity software.

Data-Driven Insights: AI-equipped vehicles collect vast amounts of data on driving patterns, user preferences, and road conditions. This data can be valuable for marketing and customer insights, enabling companies to tailor their services more effectively to individual preferences.

Safety and Efficiency: AI in transportation can significantly enhance safety by reducing accidents caused by human error. It can also improve traffic flow, reduce congestion, and minimize fuel consumption, which has implications for sustainability and cost savings.

Regulatory Challenges: The adoption of AI in transportation faces regulatory and legal challenges, including liability in the event of accidents and ensuring the safety of autonomous vehicles. These issues will need to be addressed for AI-driven transportation to become widespread.

Infrastructure and Connectivity: To support AI-enabled transportation, there will be a need for advanced infrastructure and connectivity. Roads and transportation systems may need to be upgraded to accommodate self-driving vehicles and ensure a seamless communication network.

CONCLUSION

While AI has been proven to enhance our daily consumer lives—with the adoption of autonomous vehicle technology, fitness devices, voice assistants, and smartphones—it has

also served to improve the digital marketing efforts of businesses across many industries.

Since its launch in November 2022, ChatGPT, for example, has already made a significant impact in the digital marketing arena, with its innate ability to produce targeted AI-powered content on an ever-expanding range of subjects. A clear indication of how widespread AI is in our current landscape

If utilized creatively and with the consumer's wants, needs and desires at heart, the use of AI in marketing offers incredible rewards in terms of growth, evolution, development, and ongoing success in an increasingly competitive digital landscape.

AI will continue to play an increasingly important role in digital marketing by augmenting human capabilities, automating routine tasks, and providing valuable insights. But, the human element in creativity, strategy, and relationship building is likely to remain indispensable for the foreseeable future.

REFERENCES

- [1] Chen, T. F., Tan, T., & Ko, C. (2016). Application of artificial intelligence to cross-screen marketing: A case study of AI technology company. *Advances in Intelligent Systems Research*, 133, 517–519
- [2] Davenport, T., Guha, A., Grewal, D., Bressgott, T., & Davenport, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48, 24–42
- [3] Devang, V., Chintan, S., Gunjan, T., & Krupa, R. (2019). Applications of artificial intelligence in marketing. *Annals of Dunarea de Jos University of Galati. Fascicle I. Economics and Applied Informatics*, 25(1), 28–36. <https://doi.org/10.35219/eai158404094>
- [4] Huang, M. H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, 21(2), 155–172
- [5] Lashinsky, A. (2019). Artificial intelligence: Separating the hype from the reality. Retrieved February 11, 2019 from <http://fortune.com/2019/01/22/artificial-intelligence-ai-reality/>
- [6] Faggella, D. (2019a). Artificial intelligence industry — An overview by segment. <https://emerj.com/ai-sectoroverviews/artificial-intelligence-industry-an-overview-by-segment/>
- [7] Gkikas, D. C., & Theodoridis, P. K. (2019). Artificial intelligence (AI) impact on digital marketing research. In A. Kavoura, E. Kefallonitis & A. Giovanis (Eds.), *Strategic innovative marketing and tourism* (pp. 1251–1259). Springer
- [8] Jarek, K., & Mazurek, G. (2019). Marketing and artificial intelligence. *Central European Business Review*, 8, 46–55. <https://doi.org/10.18267/j.cebr.213>
- [9] Karimova, G. Z., & Shirkhanbek, A. (2019). Marketing artificial intelligence-creating the AI archetype for evoking the personality trust. *Academy of Marketing Studies Journal*, 23, 1–13.
- [10] Marinchak, C. L. M., Forrest, E., & Hoanca, B. (2018a). Artificial intelligence: Redefining marketing management and the customer experience. *International Journal of E-Entrepreneurship and Innovation*, 8(2), 14–24. <https://doi.org/10.4018/IJEEI.2018070102>
- [11] Rekha, A. G., Abdulla, M. S., & Asharaf, S. (2016). Artificial intelligence marketing: An application of a novel lightly trained artificial intelligence marketing: An application of a novel lightly trained support vector data description. *Journal of Information and Optimization Sciences*, 37, 681–691. <https://doi.org/10.1080/02522667.2016.1191186>
- [12] Van Esch, P. (2018). Marketing AI recruitment: The next phase in job application and selection. *Computers in Human Behavior*, 90, 215–222

Demystifying AI in consumer product logistics:

A case study of« Maersk »

Optimizing transport and supply chain management operations with artificial intelligence at Maersk

Rahou Mohamed Abdl Wahab
Phd student, Higher School of Management
Tlemcen, Algérie
LEREMA
rahou7766@gmail.com

Chikhi Kamel
Professor, Higher School of Management
Tlemcen, Algérie
LEREMA
kamel_chikhi@hotmail.com

Abstract

This study highlights the various applications of artificial intelligence (AI) in supply chain management, with a particular focus on the renowned company Maersk operating in the maritime logistics sector. It examines how Maersk utilizes AI to optimize maritime routes, container management, shipment tracking, demand forecasting, warehouse operations, enhance customer service, and strengthen maritime security. This contribution aims to assess the benefits and challenges associated with the adoption of AI in managing the logistics and supply chain of consumer goods while considering prospects for continuous improvement.

Keywords : *Artificial Intelligence (AI); Logistics; Supply Chain Management; Maersk.*

I. INTRODUCTION

Artificial intelligence (AI) has evolved to become a driving force in various sectors, transforming the way businesses operate and interact with their environment. In this context, the logistics and supply chain industry has also been profoundly influenced by AI advancements. This field of computer science aims to develop systems and software capable of performing tasks that require intelligent thinking, similar to human abilities.

Supply chains, the backbone of global trade, have become increasingly complex, requiring perfect coordination of goods movement across the globe. The use of AI in this context holds the promise of a major revolution. Operational efficiency, resource management, cost reduction, and customer satisfaction are among the areas that benefit from the advantages of AI.

In order to dispel the ambiguity surrounding the use of artificial intelligence in supply chain management and to enhance their performance, we have posed the following question: "How can artificial intelligence techniques be used in logistics operations and supply chain management?"

To answer this question, we conducted a case study that showcases the use of artificial intelligence techniques and its positive impact on the logistical performance of the company "Maersk."

The Objectives of the Study :

- Examine in detail the various applications of AI in the management of consumer goods supply chains, focusing on the example of "Maersk," a renowned company in the maritime logistics sector.
- Analyze how Maersk utilizes AI to optimize maritime routes, manage containers and shipment tracking, forecast demand, manage warehouses, improve customer service, and enhance maritime security.
- Evaluate the benefits and challenges associated with the adoption of AI in logistics and supply chain with a perspective on continuous improvement.

II. LITERATURE REVIEW

A. Previous Studies:

– The study by Alain Kiyindou, Gilbert Capo-Chichi, and Kossi Amessinou published online in the journal "Tic et société" in 2023 titled: "Artificial Intelligence and Mobility in Sub-Saharan Africa: Contribution to the Analysis of Applications in the Transportation and Logistics Domain." This study explored the crucial role of mobility for development in Africa, despite limited transport infrastructure. It delved into the potential of artificial intelligence to improve mobility. The study revealed persistent challenges, including the absence of supportive policies, lack of skills and funding, as well as variations in the adoption of smart mobility among African countries.

– "The application of artificial intelligence in logistics and express delivery," a study by Yun Zhang published in the "Journal of Physics Conference Series" (2019). In this study, the analysis of the impact of artificial intelligence on logistics revealed certain opportunities and challenges. The current logistics sector has a significant gap between its capabilities and the need for intelligent data collection. Specialized talent

and standardization are essential. The introduction of AI in logistics is changing traditional relationships that require a social credit system. To realize the potential of smart logistics, concerted efforts are needed to bridge these gaps.

– Hasan Balfaqih's study titled "Artificial intelligence in logistics and supply chain management: A perspective on research trends and challenges," published in the journal "Springer: Explore Business, Technology Opportunities and Challenges after the Covid-19 Pandemic." In this study, the researcher examined the impact of Industry 4.0 technology on the digital transformation of logistics and supply chain processes. This vision aims to increase automation, improve communication, and create a data-driven, connected supply chain ecosystem. The emphasis on integrating artificial intelligence into supply chain management has generated growing interest. This study provides a conceptual overview of these areas, encouraging further research and offering valuable insights for logistics and supply chain managers.

B. Artificial intelligence: Conceptual introduction

Artificial intelligence is a field of computer science and engineering that deals with the development of systems and software capable of performing tasks that require intelligent thinking, similar to human abilities (Briganti, 2023). This includes the ability to learn from data, make decisions, solve problems, and interact with the environment (Schuett, 2019). Artificial intelligence relies on a variety of technologies and concepts (Figure 1), including:

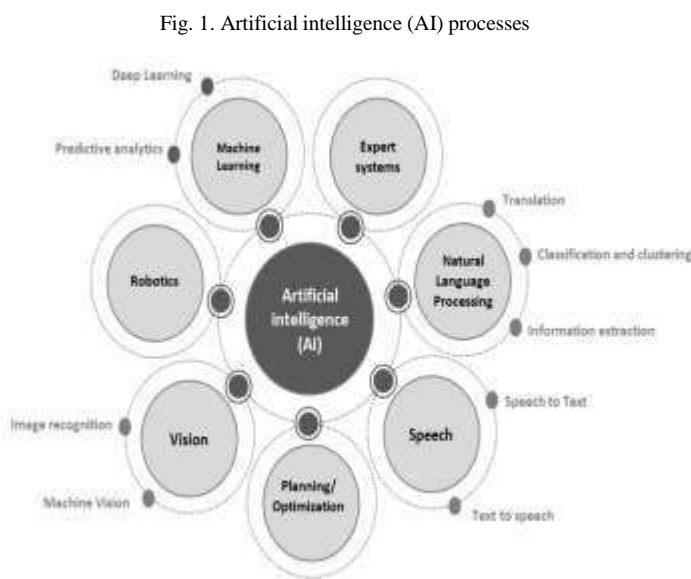


Fig. 1. Artificial intelligence (AI) processes

Source : Boris Barraud, (2019).

- Machine learning : A technology that allows industrial systems to understand and extract patterns from data and improve their performance over time (Zhou, 2021).
- Artificial neural networks : Models inspired by the neural networks in the human brain, used to solve complex problems and process data (Amellal, 2023).

- Natural language processing: Technology that enables systems to understand and produce human language and interact with humans through language (Briganti, 2023).
- Computer vision: Technology that allows systems to understand and analyze images and videos (Jiang, 2023).
- AI planning: Technologies that enable the development of models capable of planning and making decisions (Zhang & others, 2023).
- Intelligent robots: Design and development of robots using artificial intelligence technologies to perform various tasks (Briganti, 2023).

Artificial intelligence is used in a wide range of applications, including intelligent robots, medical diagnostic systems, smart surveillance, machine learning, video games, and many more. This field is constantly evolving and expanding to make advancements in various domains and industries.

C. The fascinating evolution of Artificial Intelligence: Past, Present, and Future

The development of artificial intelligence (AI) has a captivating history dating back to the period before 1956, often regarded as an incubation period for AI. Scientists and engineers sought to automate certain mental tasks using machines. In 1936, mathematician Alan Turing proposed a mathematical model of an ideal computer, laying the theoretical groundwork for future electronic computers (Zhang & Zhang, 1999). In 1943, neurophysiologists W. McCulloch and W. Pitts constructed the first neural network model, the M-P model, which imitated the structure and function of biological neurons. In 1949, Hebb proposed a learning mechanism based on neuropsychology, called "Hebb's learning rule," one of the early ideas of machine learning (Kuriscak & others, 2015). In 1952, IBM scientist Arthur Samuel developed a checkers-playing program capable of learning from the current position, marking the beginnings of evolutionary computing (Samuel, 2000). In 1956, John McCarthy coined the term "Artificial Intelligence" during the Dartmouth AI research project, becoming the "father" of AI. AI has achieved many successes since, particularly in machine learning, theorem proving, pattern recognition, problem-solving, expert systems, and natural language processing (McCarthy & others, 1995). In 1957, American psychologist Frank Rosenblatt introduced the "Perceptron" model, a turning point in neural network design. In the 1960s, the perceptron algorithm was implemented in physical hardware, the "Mark 1 Perceptron," illustrating machines' ability to perform recognition tasks (Freund & Schapire, 2000). AI also made progress in solving complex problems such as determining the molecular structure of chemical compounds. However, there have been periods of AI "winters," marked by a significant reduction in funding often related to unrealistic expectations. Despite these challenges, AI has experienced a resurgence thanks to major advancements in machine learning, the availability of massive data, increased computational power,

and impressive autonomous systems like self-driving vehicles and drones (Alexandre, 2023). Narrow AI, or ANI, has found many practical applications in our daily lives, from GPS navigation to search engines. However, ANI is limited to specific tasks. The future of AI lies in the development of autonomous systems capable of making complex decisions based on real-time data, such as autonomous vehicles (Mohri, 2018).

Currently, AI is at an exciting stage of its development with significant advancements in deep learning and neural networks. The use of high-quality data, increasing computational power, and cutting-edge algorithms have allowed AI to perform increasingly complex tasks. Areas such as computer vision, natural language processing, and robotics are experiencing exponential growth, paving the way for exciting new applications (Alexandre, 2023).

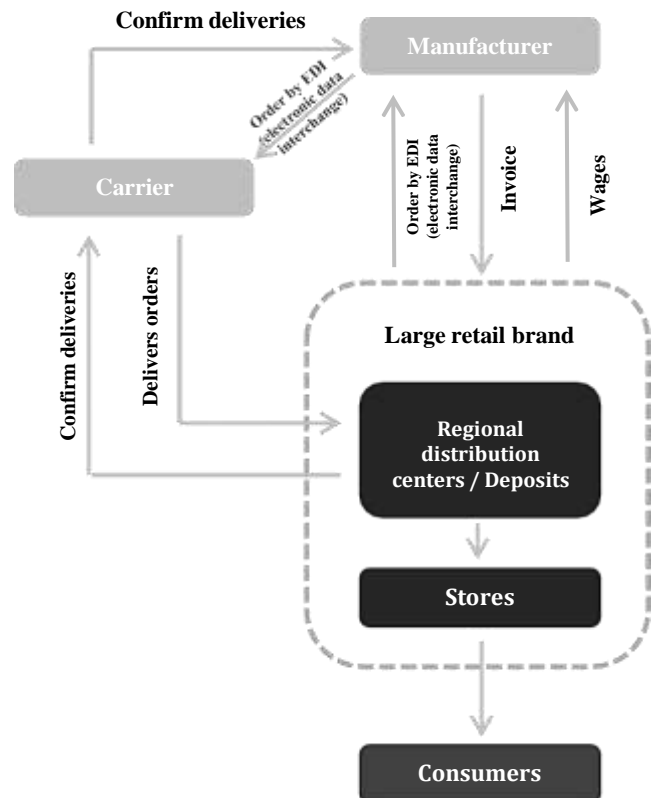
The future of AI holds even more promise. We can expect major breakthroughs in fields such as medicine, where AI can be used for medical diagnosis and drug discovery. Autonomous vehicles will become more widespread, improving road safety and mobility. AI will also be essential in addressing global challenges such as combating climate change and managing natural resources (Arvind & Vipin, 2023). There are also ethical and regulatory considerations, including privacy protection, transparency in AI system decisions, and accountability for errors. It is essential to develop strong ethical frameworks to guide the development and use of AI.

D. Logistics and supply chains: From efficiency to coordination

Logistics can be described as the process of planning, executing, and supervising operations to ensure efficient transportation and storage of goods, including services and relevant information, from their point of origin to the point of consumption to meet customer needs. This encompasses the internal and external movement of products (Mangan, Lalwani, & Calatayud, 2020). The French Association for Logistics (ASLOG) defines logistics as the art and method of making a specific product available at the right time, in the right place, at the lowest cost, and with the best possible quality to the consumer (ASLOG, 2022).

Another perspective on logistics is to ensure the right acquisition of the right product, in the appropriate quantity and quality, delivered to the right place, at the right time, for the right customer, at the right price, which can sometimes require clarity in complex environments (Takdjerd, 2022). For example, in many industrial sites, there may be multiple co-located companies, and even within a production chain, several subcontractors may collaborate with the manufacturer, making it essential to clearly define product ownership and location. Therefore, delivering to the right place may only represent part of the challenge; the goal is to deliver the product to the right customer at the right place.

Fig. 2. Description of the supply chain



Source : <https://www.mi-gb.com/conseil/logistique-la-chaine-dapprovisionnement/?lang=fr> Consulted on 20/10/2023.

Supply chain management goes beyond logistics and is a broader concept (Christopher, 2016). It is essentially an orientation and planning framework aimed at creating an overall strategy for the flow of products and information within a company. Supply chain management relies on this framework and seeks to establish coordination between the processes of different entities in the chain, namely suppliers, customers, and the company itself (Ferrahi, 2016).

Therefore, supply chain management focuses on managing relationships to achieve a more profitable outcome for all parties involved. This can present challenges as there may be situations where one party's individual interests must be set aside for the benefit of the entire chain.

III. METHODOLOGY

To gather information on Maersk's use of artificial intelligence (AI) in supply chain management, we have adopted a mixed approach combining primary and secondary data sources. Primary data collection methods involve direct communication with information and media representatives of the company, while secondary methods include reviewing reports and information available on Maersk's official website.

A. Primary data collection

Communication with Media and Information Representatives: We established communication with designated media and information representatives of Maersk through email and phone calls. A set of structured questions was presented to these representatives to obtain firsthand information on Maersk's AI applications in supply chain management. These questions covered:

- Supply chain management practices and strategies adopted by Maersk.
- The use of AI: Exploring the various ways Maersk utilizes AI technologies, including application areas.
- AI algorithms and techniques: Requesting information on the AI algorithms and techniques used by Maersk.
- Interviews and correspondence: Follow-up interviews or correspondence were conducted to gain better clarity and deeper understanding of certain questions.

B. Secondary data collection

Maersk's official website was examined to find reports, technical documents, and any public information related to the company's use of AI in logistics and supply chain management. Information from official publications and press releases was gathered and analyzed. The combination of primary and secondary data collection methods allows for a comprehensive and multifaceted understanding of Maersk's strategy around the use of AI for logistics and its associated applications, algorithms, and techniques. Primary data collection from company representatives provides insights from within the organization, while secondary data collection ensures that the study is based on publicly available and verifiable information. This research methodology is designed to facilitate a robust analysis of Maersk's use of AI in supply chain management and to ensure that the gathered information is reliable and representative of the company's practices and strategies in this field.

C. MAERSK

Maersk is a renowned Danish company in the field of transportation and logistics of consumer products, founded in 1904. It operates on a global scale and is primarily known for its activities in maritime transportation, freight, and port management. Its influence extends to various sectors such as container transportation, bulk cargo transportation, and integrated logistics (A.P. Møller-Mærsk Gruppen, 2023). Maersk's commitment to sustainability is demonstrated through various initiatives aimed at reducing its environmental footprint, including investments in more environmentally friendly technologies, as illustrated by the following key figures (Maersk, 2023):

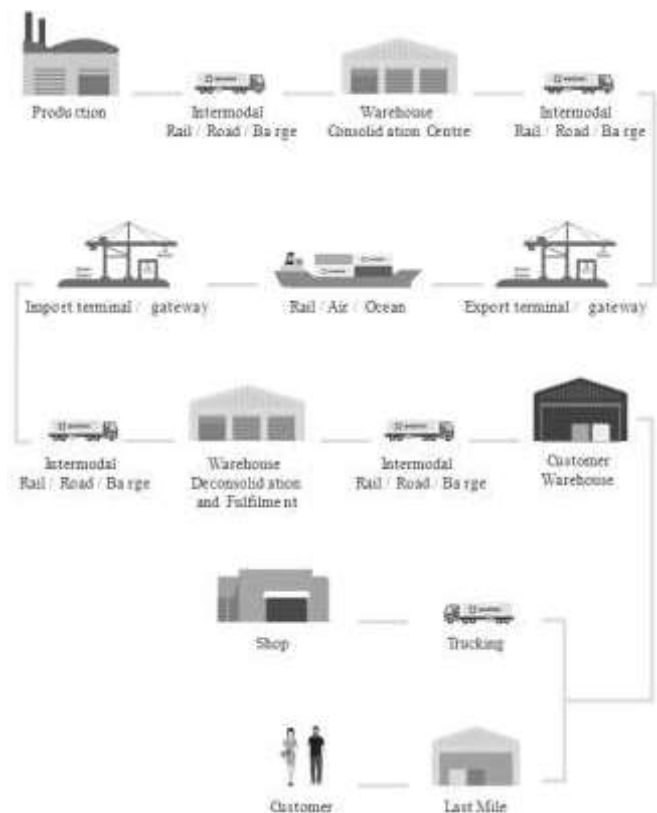
- Its fleet includes more than 730 container vessels operating internationally.
- It employs over 100,000 employees.

- The company manages 67 terminals in 42 countries and operates in a total of 130 countries.
 - Maersk serves more than 100,000 customers.
 - Its website, maersk.com, hosts more than 3 billion professional events.
 - With over 3 million square meters of warehouses located around the world, Maersk offers advanced logistics.
- Furthermore, Maersk has set an ambitious goal to achieve carbon neutrality in all its operations by 2040. These data highlight Maersk's global reach and its commitment to sustainability, underscoring its central role in the international logistics and transportation sector.

IV. RESULTS AND DISCUSSION

Maersk continuously strives to use the latest technologies to reduce time and costs while satisfying its customers. Based on the information we have gathered in our study, we will explain how the company employs artificial intelligence in its logistics operations:

Fig .3. Logistics processes at Maersk



Source : <https://www.maersk.com/fr-fr/supply-chain-logistics/warehousing-and-distribution/warehousing-services>, Consulted on 20/10/2013.

A. Optimization of maritime routes

- Maersk uses AI to optimize the routes of its vessels based on various factors, including weather conditions, ports, vessel capacity, and fuel costs.

- The process involves data collection, data modeling, setting clear objectives, utilizing optimization algorithms, and considering factors such as weather conditions, port calls, vessel capacity, and fuel costs.
- Real-time adjustments are made if necessary, and the optimal route is implemented in operational planning.

B. Container and shipment tracking

- Maersk uses AI systems to track the real-time location of containers and predict shipment arrival times.
- This allows Maersk's customers to gain real-time visibility into their goods and optimize their operations.
- The process involves data collection from IoT sensors, data preprocessing, data modeling for location and arrival predictions, integration with customers, operations optimization, and analysis of historical data.

C. Demand forecasting

- Maersk uses AI to predict the demand for maritime transport, aiding in capacity and operations planning.
- The process includes data collection, data preprocessing, using AI models for demand prediction, feature extraction, model training, forecast integration, capacity and operations optimization, and continuous improvement.

D. Warehouse management

- Maersk employs AI to automate inventory management, order picking, and other warehouse operations, reducing processing times.
- The process involves data collection, data preprocessing, automated inventory management, automated order picking, delivery route optimization, real-time monitoring, analysis of historical data, and continuous improvement.

E. Customer service and chatbots

- Maersk uses AI-based chatbots to provide 24/7 customer support, answer customer questions, and provide instant information.
- The process includes identifying use cases, data and content collection, selecting the chatbot platform, training the chatbot, integration into communication channels, customization, human backup support, performance analysis, and content updates.

F. Sensor data analysis

- Maersk analyzes data collected from IoT sensors on its vessels to identify trends, anomalies, and optimization opportunities.
- The process includes IoT data collection, data transfer, data storage, data preprocessing, using AI models for data analysis, anomaly detection, maintenance prediction, route and fuel

consumption optimization, result visualization, alerts, automated actions, and continuous improvement.

G. Maritime security

- Maersk uses AI to enhance the safety of its vessels by detecting potential threats, monitoring dangerous weather conditions, and helping to avoid collisions at sea.
- The process involves data collection, data transmission, data preprocessing, using AI models for threat detection, weather prediction, navigation assistance systems, crew training, data analysis, alerts, automated actions, and continuous improvement of maritime safety.

V. CONCLUSION

The integration of artificial intelligence into supply chain management has paved the way for a profound transformation in the logistics of consumer products. The example of Maersk demonstrates how AI has become an indispensable tool for improving efficiency, enhancing coordination, and meeting customer expectations. The benefits are numerous, ranging from optimizing maritime routes to improving warehouse management and maritime security. However, it is essential to acknowledge that challenges remain, such as managing sensitive data and the need to train staff in these new technologies. The future of logistics and supply chain lies in the ethical and responsible use of AI and the constant quest for improvements for better performance and sustainable customer satisfaction. AI is now the best collaborator of supply chains, promising to continue to reinvent this vital sector.

REFERENCES

- [1] A.P. Møller-Mærsk Gruppen. (2023, 8 9). Wikipédia : <https://fr.wikipedia.org/wiki/Maersk>. Consulted on 23/10/2013
- [2] Alexandre, G. (2023). IA: pour une histoire culturelle. *Revue d'histoire culturelle* (4).
- [3] Amellal, R. (2023). Application des réseaux de neurones artificielles dans l'optimisation des performances d'antennes reconfigurables. *Revue Internationale du Chercheur*, 4(3).
- [4] Arvind, A., & Vipin, M. (2023). Les risques liés à l'innovation : le cas de l'intelligence artificielle. *Open science*.
- [5] Briganti, G. (2023). Intelligence artificielle : une introduction pour les cliniciens. *Revue des Maladies Respiratoires*, 40(4), 208-313.
- [6] Christopher, M. (2016). *Logistics & Supply Chain Management* (éd. 5). Pearson UK.
- [7] Ferrahi, A. (2016). La gestion des risques de la chaîne logistique au cœur de la problématique d'efficacité du processus logistique : Etude exploratoire des entreprises en Algérie. *Les Cahiers MECAS*, 12(2), 299-315.
- [8] Freund, Y., & Schapire, R. (2000). Large margin classification using the perceptron algorithm. *Mach Learn*.
- [9] Jiang, J. (2023). Contribution à l'analyse biomécanique et ergonomique de poste de travail par capture du mouvement sans marqueurs et vision par ordinateur. *École Nationale Supérieure d'Arts et Métiers*. Paris: These doctorat.
- [10] Kurisak, E., & Marsalek, P. & Stroffek, J. & Toth, PG. (2015). Biological context of Hebb learning in artificial neural networks, a review. *Neurocomputing*, 27-35.

- [11] L'ASLOG. (2022). 12 3, 2022, sur [faq.logistique : https://www.faq-logistique.com/Logistique.htm#:~:text=L'ASLOG%20d%C3%A9finit%20la%20logistique,et%20avec%20la%20meilleur%20qualit%C3%A9%22](https://www.faq-logistique.com/Logistique.htm#:~:text=L'ASLOG%20d%C3%A9finit%20la%20logistique,et%20avec%20la%20meilleur%20qualit%C3%A9%22). Consulted on 20/10/2013.
- [12] Maersk. (2023, 1 1). About Maersk. MAERSK: <https://www.maersk.com/about>. Consulted on 21/10/2013.
- [13] Mangan, J., Lalwani, C., & Calatayud, A. (2020). *Global Logistics and Supply Chain Management* (éd. 4, illustrée). John Wiley & Sons.
- [14] McCarthy, J., & Minsky, M L. & Rochester, N. & Shannon, C E. (1995). *A proposal for the Dartmouth summer research project on artificial intelligence*. Stanford: AI Magazine.
- [15] Mohri, M. (2018). *Foundations of machine learning*. Cambridge: MIT Press.
- [16] Prajogo, D., & Sohal, A. (2013). Supply chain professionals: A study of competencies, use of technologies, and future challenges. *Int. J. Oper. Prod. Manag.*, 1532–1554.
- [17] Samuel, A. (2000). Some studies in machine learning using the game of checkers. *IBM J Res Dev*, 44(1.2), 20-26.
- [18] Schuett, J. (2019). *A Legal Definition of AI*. Frankfurt: Goethe University.
- [19] Takdjerad, N. (2022). La gestion des risques de la chaîne logistique à l'international: cas d'une entreprise: SARL SERINA HCAN BEN ALI. *Marketing and Business Research Review*, 2(1), 34-44.
- [20] Wamba, S., & Queiroz, M. (2020). Blockchain in the operations and supply chain management: Benefits, challenges and future research opportunities. *Int. J. Inf. Manag.* (52), 102064.
- [21] Zhou, Z.-H. (2021). *Machine Learning*. Springer Nature.
- [22] Zhang, L., & Zhang, B. (1999). A geometrical representation of McCulloch-Pitts neural model and its applications. *IEEE Transact Neural Networks*, 10(4).
- [23] Zhang, Y., & Liu, J. & Chen, z. Weixun, Li. (2023). Path Planning of Mobile Robots Based on Genetic Algorithm. *IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*. IEEE.
- [24] Zhi-Peng, L., & Sang-Joon, L. & Hyi-Thaek, C. (2021). The Effect of Blockchain Operation Capabilities on Competitive Performance in Supply Chain Management. *Sustainability*, 13(21).