Advances in Science, Technology & Engineering Systems Journal

Special Issue

Computing, Engineering and Multidisciplinary Sciences

2025

www.astesj.com
ISSN: 2415-6698

Editorial

The Special Issue on *Computing, Engineering and Multidisciplinary Sciences 2025* in the *Advances in Science, Technology and Engineering Systems Journal (ASTES Journal)* reflects the evolving landscape of modern research, where the boundaries between disciplines are increasingly fluid and innovation thrives at their intersections. In an era defined by rapid technological progress and complex global challenges, the integration of computing methodologies with engineering principles and multidisciplinary scientific approaches has become essential. This issue brings together a diverse collection of contributions that demonstrate how collaborative, cross-domain thinking is driving advancements in both theoretical understanding and practical applications.

A central theme emerging from this issue is the transformative role of computing as an enabling force across scientific and engineering domains. From artificial intelligence and data analytics to high-performance computing and intelligent automation, computational tools are no longer auxiliary but foundational to innovation. The papers featured herein illustrate how these technologies are being harnessed to model complex systems, optimize engineering processes, and extract meaningful insights from vast datasets. Such approaches are not only enhancing efficiency and accuracy but are also opening new avenues for discovery across disciplines ranging from environmental science to biomedical engineering.

Equally significant is the emphasis on engineering solutions that are informed by multidisciplinary perspectives. The contributions highlight how integrating knowledge from physics, chemistry, biology, and social sciences leads to more robust and adaptable systems. Whether addressing sustainable energy challenges, designing resilient infrastructure, or developing smart technologies for urban environments, the research underscores the necessity of holistic thinking. This convergence enables researchers to move beyond narrow problem-solving frameworks and develop solutions that are better aligned with real-world complexities.

The issue also captures the growing importance of collaborative research ecosystems. Many of the studies presented are the result of partnerships between academia, industry, and research institutions, reflecting a broader shift toward cooperative innovation. Such collaborations facilitate the translation of theoretical research into practical implementations, ensuring that scientific advancements have tangible societal and economic impacts. Furthermore, the inclusion of multidisciplinary perspectives fosters creativity and encourages the exploration of unconventional ideas, which are often the catalyst for breakthrough innovations.

Another noteworthy aspect of this collection is its attention to emerging challenges and ethical considerations. As technologies become more integrated into daily life, questions related to data privacy, system security, sustainability, and societal impact become increasingly critical. The contributions in this issue not only present technical advancements but also engage with these broader concerns, highlighting the importance of responsible innovation. By addressing these dimensions, the research aligns with the global imperative to develop technologies that are not only effective but also equitable and sustainable.

From an editorial standpoint, this Special Issue represents a commitment to advancing knowledge at the convergence of computing, engineering, and multidisciplinary sciences. It provides a platform for researchers to share insights that transcend traditional disciplinary boundaries and encourages the development of integrated approaches to complex problems. The diversity of topics and methodologies showcased here reflects the dynamic and interconnected nature of contemporary research.

The works presented in this issue collectively illustrate that the future of innovation lies in collaboration, integration, and adaptability. By bridging disciplines and fostering a culture of interdisciplinary inquiry, this Special Issue contributes to a deeper understanding of how computing and engineering can be leveraged within a broader scientific context to address the pressing challenges of our time and to shape a more resilient and forward-looking global landscape.

Guest Editor

Prof. Wang Xiu Ying

ADVANCES IN SCIENCE, TECHNOLOGY AND ENGINEERING SYSTEMS JOURNAL

Special Issue **March 2024**

CONTENTS

Advances in Science, Technology and Engineering Systems Journal
Vol. 10, No. 2, 1-11 (2025)
www.astesj.com
Special Issue on Computing, Engineering and Multidisciplinary Sciences

ASTES Journal
ISSN: 2415-6698

Generative Artificial Intelligence and Prompt Engineering: A Comprehensive Guide to Models, Methods, and Best Practices

Maikel Leon*

Department of Business Technology, Miami Herbert Business School, University of Miami, Miami, Florida, USA

ARTICLE INFO

Article history:
Received: 06 January, 2025
Revised: 12 February, 2025
Accepted: 14 February, 2025
Online: 07 March, 2025

Keywords:
Generative AI
Large Language Models
Prompt Engineering

ABSTRACT

This article enhances discussions on Generative Artificial Intelligence (GenAI) and prompt engineering by exploring critical pitfalls and industry-specific advantages. It begins with a foundational overview of AI evolution, emphasizing how generative models such as GANs, VAEs, and Transformers have revolutionized language processing, image generation, and drug discovery. Prompt engineering is highlighted as a key methodology for directing model outputs with precision and ethical awareness, enabling applications in Natural Language Processing (NLP), content personalization, and decision support. The revised sections detail how prompt engineering can be misapplied, underscoring common errors like overly restrictive or ambiguous prompts that compromise GenAI's accuracy, ethicality, and creative capacity. Equally, the paper showcases high-impact use cases in finance, education, healthcare, and beyond, illustrating how carefully formulated prompts can strengthen risk detection, enhance student learning, improve clinical decision-making, and foster product innovation. The expanded discussion of industry alignment illustrates the tangible value these techniques offer across diverse sectors, ultimately reinforcing the notion that prompt engineering is central to maximizing GenAI's transformative potential. Future directions address emerging trends, from multimodal fusion and domain-specific fine-tuning to adaptive prompt designs that leverage real-time user feedback, further solidifying the role of responsible prompt engineering in shaping the next generation of intelligent and ethically aligned AI solutions.

1. Introduction

Artificial intelligence (AI) has progressed remarkably since its early stages, transitioning from systems governed by rigid rules to more adaptive, data-driven approaches capable of discerning intricate patterns [1]. Recently, the emergence of GenAI and robust prompt engineering techniques has redefined how organizations, researchers, and practitioners approach an expansive range of tasks—including, but not limited to, text analytics, visual design, product innovation, and strategic decision-making [2]. This document extensively evaluates these methodologies, presenting how GenAI fuels the creation of entirely new data. At the same time, prompt engineering directs and molds model outputs for improved precision, originality, and ethical oversight [3].

AI originated in the mid-20th century with influential figures such as Alan Turing, whose seminal inquiries into computational processes laid the groundwork for machine-mediated reasoning. During the 1950s and 1960s, there was pronounced enthusiasm for symbolic manipulation and expert systems [4], propelling significant investment and heightened aspirations for AI breakthroughs. Over time, these expectations were periodically unmet, leading to intervals known as AI winters, marked by waning research backing and tempered academic excitement. Nonetheless, incremental advancements in ML methodologies, algorithmic efficiency, and representational frameworks persisted. By the late 1980s and 1990s, rekindled interest in neural-network-based models, coupled with improving hardware performance, catalyzed the resurgence of AI. The advent of big data analytics and deep learning architectures in the early 2010s firmly established AI as a predominant technological force, furnishing the foundations for innovative developments in GenAI and modern ML paradigms [5].

GenAI can be seen as a type of advanced Machine Learning (ML) algorithm category that produces novel yet meaningful data points by internalizing learned distributions within extensive datasets. Such approaches extend beyond conventional classification and prediction paradigms, paving the way for generating text, images, music, and complex molecular structures [6]. Prompt engineering, by contrast, constitutes a systematic process of formulating concise, context-specific prompts to orient generative models to-

*Corresponding Author: Maikel Leon (mleon@miami.edu)

ward desired behaviors. As Large Language Models (LLMs) evolve in capacity and scope, carefully devised prompts serve as a vital interface, shaping clarity, domain-specific alignment, and overall efficacy [7].

In addition to providing historical context, this paper aims to:

- Highlight gaps in existing research where more systematic experimentation on GenAI and prompt engineering is needed.
- Compare popular GenAI models at a conceptual level and discuss how they suit different tasks.
- Provide practical guidance on best practices for prompt design, with examples illustrating successful and unsuccessful outcomes.
- Expand on ethical and societal implications, focusing on the bias, potential misuse, and long-term effects of generative AI in diverse industries.

2. Large Language Models

LLMs represent a transformative advancement in AI, designed to comprehend, generate, and engage in human-like language interactions. With remarkable proficiency, these models leverage vast amounts of data and sophisticated algorithms to perform many language-related tasks. The development of LLMs is driven by creating more intuitive and versatile AI systems that can seamlessly integrate into various facets of business and society. By enabling machines to understand and generate natural language, LLMs bridge the gap between human communication and machine processing, facilitating more effective and meaningful interactions [8].

The capabilities of LLMs extend beyond simple text generation; they encompass understanding context, maintaining coherence over extended discourse, and adapting to diverse linguistic styles and domains. This versatility positions LLMs as essential tools for enhancing productivity, automating complex tasks, and fostering innovation across multiple industries [9]. As organizations increasingly seek to harness the power of AI to gain competitive advantages, LLMs offer a robust foundation for developing intelligent applications that can drive strategic decision-making and operational excellence.

2.1. History

The evolution of LLMs can be traced back to the early developments in NLP and ML. Initial attempts focused on rule-based systems and statistical methods, which laid the groundwork for more advanced models. The introduction of neural networks marked a significant milestone, enabling machines to learn from data more flexibly and on a larger scale. However, transformer architectures in the mid-2010s revolutionized the field, providing the necessary framework for building models with unprecedented capacity and performance.

The release of models such as BERT and GPT series demonstrated the potential of LLMs to perform a wide range of tasks with minimal task-specific training. These models capitalized on large-scale pre-training on diverse datasets, followed by fine-tuning for specific applications, achieving state-of-the-art results in various benchmarks [10]. The continual scaling of model parameters and training data has further enhanced the capabilities of LLMs, enabling them to generate more coherent and contextually relevant outputs. This historical trajectory underscores the rapid advancements in computational power, data availability, and algorithmic innovations that have propelled LLMs to the forefront of AI research and application [11].

2.2. Architecture

The architecture of LLMs is primarily based on transformer networks, which utilize self-attention mechanisms to process and generate language. Unlike traditional recurrent neural networks, transformers can handle long-range dependencies and parallelize computations more efficiently, making them well-suited for training on extensive datasets. The core components of an LLM architecture include multiple layers of attention and feed-forward neural networks, which collectively enable the model to capture complex linguistic patterns and contextual relationships.

A typical LLM consists of an encoder and a decoder, although many modern architectures, such as the GPT series, employ only the decoder component for generative tasks. The self-attention mechanism allows the model to weigh the importance of different words in a sequence, facilitating a deeper understanding of context and meaning. Positional encoding is also incorporated to retain the order of words, which is crucial for maintaining coherence in generated text. Layer normalization, residual connections, and dropout techniques are employed to enhance training stability and prevent overfitting [3].

The scalability of LLM architectures is a key factor in their success. By increasing the number of layers, attention heads, and parameters, LLMs can achieve higher levels of performance and adaptability. This scalability is complemented by advancements in distributed computing and parallel processing, which enable the training of extremely large models on vast datasets [12]. The architectural innovations in LLMs have improved their ability to generate high-quality text and expanded their applicability to a broader range of tasks, including translation, summarization, and conversational agents.

Large-scale transformer training often incorporates gradient accumulation to handle very large batch sizes without exceeding GPU memory. Some frameworks use mixed-precision training (e.g., FP16) to reduce memory usage and speed computation. When scaling to billions of parameters, advanced optimizers like Layer-wise Adaptive Rate Scaling (LARS) can further improve convergence in deep networks. While these methods do not alter the fundamental self-attention architecture, they are critical for practical, large-scale LLM implementations.

2.2.1. Comparison with Other GenAI Models

Although transformers dominate many modern NLP tasks, other generative architectures retain niche advantages. RNNs and LSTMs can be more efficient for simpler tasks or smaller datasets, albeit with limitations in handling long-range context. VAEs offer interpretable latent spaces, supporting tasks like anomaly detection or data compression. Meanwhile, GANs excel in image and audio synthesis, though maintaining equilibrium between generator and discriminator can be challenging. The choice of architecture often hinges on domain constraints, data availability, and the desired trade-off between interpretability and performance.

In practical deployments, modern LLMs are typically evaluated on domain-specific tasks or industry benchmarks, such as human evaluation of text coherence, code generation accuracy, or specialized QA metrics. For instance, some organizations measure how well a large transformer-based model answers financial or legal queries compared to in-house experts or test chatbot performance on thousands of real customer interactions. These benchmarks provide pragmatic insights into how architectural differences (e.g., number of attention heads) and training optimizations translate into real-world improvements in quality and user satisfaction.

2.3. Applications

LLMs have found applications across various industries, leveraging their ability to understand and generate natural language to drive innovation and efficiency. In healthcare, LLMs analyze medical records, diagnose, and personalize patient care through tailored communication. By processing vast amounts of unstructured data, these models can identify patterns and insights that inform clinical decision-making and improve patient outcomes. They play a critical role in risk assessment, fraud detection, and customer service automation in the financial industry. They can analyze market trends, generate financial reports, and provide real-time support through intelligent chatbots, enhancing operational efficiency and enabling more informed investment strategies. Similarly, LLMs facilitate personalized learning experiences in the education sector by adapting educational content to individual student needs, automating grading, and providing instant feedback [13].

The realm of content creation and media has also been transformed by LLMs, which generate articles, scripts, and marketing materials with minimal human intervention. This capability accelerates content production and allows for greater customization and scalability. Additionally, LLMs enhance human-machine interactions through virtual assistants and conversational agents that engage users in meaningful and contextually relevant dialogues [14].

Beyond these sectors, LLMs are instrumental in research and development, aiding in literature reviews, hypothesis generation, and data synthesis. Their ability to process and generate language at scale makes them invaluable tools for accelerating innovation and fostering collaborative efforts across disciplines. The broad applicability of LLMs underscores their potential to drive significant advancements and create new opportunities in an increasingly data-driven and interconnected world.

2.4. Debrief

The widespread adoption of LLMs underscores their profound impact on technological and organizational landscapes. Their ability to process and generate human-like language has enhanced existing processes and paved the way for new applications and business models. Organizations leveraging LLMs benefit from increased efficiency, reduced operational costs, and the ability to deliver more personalized and engaging experiences to their stakeholders.

LLMs also play a pivotal role in enabling data-driven decision-making by providing deeper insights into vast and complex datasets. Their capacity to analyze unstructured data sources, such as text and speech, complements traditional data analysis methods, offering a more comprehensive understanding of market trends, customer behavior, and operational performance. This integration of LLMs into analytical frameworks empowers businesses to make more informed and strategic decisions, fostering a culture of innovation and continuous improvement [15].

Moreover, LLMs' scalability and adaptability remain relevant to evolving business needs and technological advancements. As models grow in size and complexity, their capabilities expand, allowing them to tackle more sophisticated tasks and integrate seamlessly with emerging technologies such as the Internet of Things (IoT) and augmented reality (AR). This adaptability enhances the longevity of LLM investments and ensures that organizations can stay ahead in a competitive and rapidly changing environment.

However, deploying LLMs also requires carefully considering ethical and operational challenges. Issues such as data privacy, bias in generated content, and the potential for misuse require robust governance frameworks and responsible AI practices. Addressing these challenges is essential for maximizing the benefits of LLMs while minimizing potential risks, ensuring that their integration into business processes aligns with organizational values and societal expectations.

LLMs have established themselves as indispensable tools in modern business, driving innovation and competitive advantage. Their ongoing development and integration into various sectors promise to unlock new possibilities and redefine the boundaries of what is achievable with AI. As organizations continue to navigate the complexities of digital transformation, LLMs will undoubtedly play a central role in shaping the future of work, communication, and strategic decision-making [16].

3. Understanding GenAI

Recognizing AI's historical trajectory and the underlying mechanisms that empower generative models can help data scientists and business innovators unlock fresh avenues for problem-solving and creative invention across multiple domains.

3.1. Evolution of AI: From Rule-Based to Generative Models

Early AI systems hinged on predefined logic rules and static processes. Although useful in constrained scenarios, these approaches could not adapt to subtle or evolving tasks. As the volume of digitized data swelled and computational resources advanced, ML and data-centric models gradually supplanted rigid rule-based tools [17]. Researchers came to appreciate that constructing models capable of generating new data instances and distinguishing and classifying existing data expanded the range of potential AI applications [18]. Out of this realization, GenAI emerged, providing a fertile ground for synthesizing novel text, images, or even decision-support insights. This paradigm shift spurred breakthroughs—from creating coherent language passages to rendering high-resolution images—and revolutionized industrial processes by streamlining innovation, elevating personalized consumer services, and minimizing labor-intensive manual workflows.

3.2. Key GenAI Models: RNNs, LSTMs, GPT, and More

Several well-established models form the backbone of GenAI:

- Restricted Boltzmann Machines (RBMs): Early probabilistic frameworks that represent data distributions by connecting observed units with latent variables, serving as precursors to deeper generative architectures.
- Variational Autoencoders (VAEs): Leverage probabilistic encoders and decoders to map data into a latent representation, facilitating structured data generation and meaningful interpolation in a continuous space.
- Generative Adversarial Networks (GANs): Employ a two-model framework (generator and discriminator) in competitive training. The generator aims to produce highly realistic outputs, while the discriminator evaluates authenticity, thereby driving continuous improvement in generation quality [19].
- Recurrent Neural Networks (RNNs): Capture sequence dependencies through hidden states that evolve, though they often grapple with gradient-related challenges for long sequences. Despite these drawbacks, they were a fundamental step in modeling linguistic structures [20].
- Long Short-Term Memory (LSTMs): Introduce gating mechanisms to mitigate vanishing gradients, enabling more reliable handling of extended text sequences or time-dependent phenomena [21].
- Transformers (for instance, GPT): Employ attention mechanisms that operate in parallel across sequences, allowing for significantly enhanced scalability and performance in language-oriented tasks, including in-context learning and context retention over considerable text spans.

Depending on project needs—such as data volumes, output format (text, image, audio), and available computational bandwidth—each model class delivers unique strengths and may be strategically adopted for best results [22].

Although Transformers often outperform alternative GenAI architectures in large-scale text tasks, VAEs or GANs may be more suitable for image synthesis or anomaly detection. At the same time, RNNs or LSTMs can be simpler to train on smaller datasets. The best choice hinges on domain constraints and resource availability.

3.3. Popular Use Cases for GenAI

GenAI, by synthesizing robust and contextually appropriate data, has permeated a multitude of sectors:

- NLP: Powers automated summarization, content creation, and question-answering systems, improving customer support, knowledge dissemination, and overall operational efficiency [23].
- Image Synthesis and Editing: GANs and VAEs underpin image-to-image translation, style transfers, and photorealistic rendering, reshaping digital design and online product visualization.
- Music and Audio Generation: Sequence-based and transformer-based audio frameworks facilitate the composition of musical scores and synthetic voices, transforming entertainment and interactive voice technologies.
- Drug Discovery and Material Science: Generating new molecular and structural formulations accelerates R&D cycles, reducing the time needed for validation and optimization.
- Anomaly Detection and Pattern Recognition: Generative models model standard patterns and detect deviations, thus supporting robust fraud prevention and quality assurance initiatives.
- Data Augmentation: Generative techniques enhance predictive accuracy for various supervised learning endeavors by producing new training instances that enrich underrepresented classes [24].
- Simulation and Scenario Planning: Synthetic data fuel the simulation of market fluctuations, supply-chain constraints, or policy trade-offs, helping businesses refine their strategic planning processes [25].

Informal surveys in some organizations reveal that managers find generative models boost efficiency in drafting or analytics tasks by up to 90% in pilot projects, yet about 60% express concerns about explainability or compliance. This underscores the balance between pragmatic gains and the need for rigorous oversight in domains like insurance or healthcare [26].

4. Prompt Engineering

The importance of prompt engineering has risen in parallel with the widespread integration of transformer-based architectures like GPT, which respond directly to prompt instructions when generating outputs. A nuanced approach to crafting these prompts can dramatically influence model performance [27].

4.1. Why it Matters

Prompt engineering entails developing precise instructions—referred to as prompts—that guide generative models, specifically LLMs, toward the intended results. A well-crafted prompt can evoke succinct, context-appropriate text, structured data, or specialized solutions while minimizing irrelevant or illogical responses. As AI models expand in complexity, the prompt design stands at the forefront of practical deployment, shaping outcomes for marketing communications, medical informatics, or content moderation [28]. Effectively engineered prompts can ensure an organization's brand remains consistent, guarantee technical accuracy in specialized domains, and uphold cultural sensitivities while aligning with operational goals.

4.2. Prompt Types

Varying degrees of specificity characterize prompt design:

- Explicit Prompts: Clearly define content type and structure. Example: "Summarize the subsequent article in four bullet points emphasizing its principal assertions."

- Implicit Prompts: Pose more open-ended queries, leaving the model to interpret context. Example: "Reflect on the foundational ideas presented in the text."
- Creative Prompts: Intentionally inspire novel perspectives or imaginative responses. Example: "Compose a futuristic narrative inspired by the primary discoveries in this research paper."

The appropriate style depends on the degree of creative freedom desired and the degree of detail required to meet organizational or scholarly benchmarks.

4.3. Best Practices

The effectiveness of prompt engineering is supported by key guidelines, such as:

- Be clear and concise: Articulate instructions unambiguously to avert confusion or extraneous responses.
- Provide context: Incorporate relevant domain insights, background data, or salient references in the prompt.
- Specify the desired format: Indicate structural expectations, such as enumerated lists or succinct paragraphs.
- Encourage multiple attempts: Solicit multiple outputs or iterative feedback to refine clarity and precision.
- Balance guidance and freedom: Overly restrictive prompts may hamper creativity, while excessively broad prompts risk losing focus.
- Evaluate and iterate: Continuously refine your prompt strategies in response to model performance metrics and expert feedback.

Below are brief illustrations of how small differences in prompt design can lead to significantly different results:
Successful Prompt: "Draft a 200-word press release introducing our new data analytics platform. Emphasize speed, security, and user-friendliness. Include a short quote from the CEO."
Analysis: This prompt's clarity on style, length, and key features (speed, security, user-friendliness) helps align the generated text with the organization's marketing goals.

Unsuccessful Prompt: "Write something about our new product."
Analysis: Overly vague instructions may produce meandering or irrelevant text, failing to highlight critical selling points or match the intended brand tone.

Another Failure Example: "Provide advice for diagnosing all diseases in humans using only three bullet points."
Analysis: This is both overly ambitious and excessively constrained. It encourages the model to produce incomplete or erroneous medical advice, which poses ethical and practical risks.

5. Practical Applications of Prompt Engineering

Custom prompts steer AI models toward producing more accurate results but also empower these models to address complex language-based and data-driven questions spanning countless fields.

5.1. Improving NLP Tasks with Custom Prompts

Prompt engineering has shown substantial benefits for language-centric processes:

- Text Summarization: Prompts can delineate target length, audience, or detail level, thereby generating succinct yet comprehensive overviews.
- Sentiment Analysis: Focus model attention on emotional cues within consumer feedback, supporting targeted marketing and brand strategy.
- Text Generation: Maintain thematic continuity and organizational voice across marketing, corporate communications, or public announcements.
- Question-Answering: Embed contextual hints and clarifications in prompts to bolster factual veracity and interpretive depth.
- Text Classification: Restrict the model to specific labels or categories, improving classification consistency in legal or customer-service contexts.
- Machine Translation: Strengthen stylistic adherence and domain-specific diction in translations by offering pertinent glossaries or examples.

5.2. Creativity and Diversity in AI-Generated Content

Content that thrives on innovation often harnesses prompt engineering to boost ideation and novelty:

- Idea Generation: Prompt the AI to merge unrelated concepts or shift narrative points of view, expanding creative frontiers in writing or media.
- Constraint-Based Challenges: Mandate the use of specific structures, lexical elements, or rhetorical forms, fostering more unconventional outputs.
- Iterative Refinement Loops: Feed a model's output as a new prompt, encouraging sophisticated evolution of narratives, character details, or design concepts.

5.3. AI Ethics and Bias

Thoughtful, prompt design can serve as a bulwark against harmful outputs and biases:

- Encouraging Fairness and Inclusivity: Instruct the model to include various perspectives, effectively broadening discourse.
- Avoiding Harmful Stereotypes: Explicitly discourage hateful or derogatory content concerning respectful outcomes.
- Promoting Fact-Checked Content: Require the model to cite verifiable sources, thus curtailing misinformation and preserving credibility.

5.4. Personalization

Prompt engineering is pivotal for delivering tailored experiences:

- Incorporating User Preferences: Integrate a user's reading or purchase history directly into prompts for heightened personalization.

- Adjusting Language and Tone: Align the AI's outputs with brand guidelines or adapt voice and register for professional, informal, or technical contexts.
- Adaptive Learning and Tutoring: Dynamically reshape prompts in educational platforms based on each learner's prior responses, fostering individualized instruction [29].

By combining concise prompts with domain-specific data, organizations can harness AI to craft resonant, personalized communications at scale.

6. Improper Approaches

Misaligned or poorly structured prompts can undermine the effectiveness of even the most sophisticated GenAI models. Practitioners risk producing misleading, irrelevant, or harmful outputs by overlooking crucial clarity, context, or ethical considerations. This section investigates some of the most common pitfalls in prompt engineering, providing concrete examples and analytical commentary to illuminate why these approaches fail to achieve reliable, high-quality results [30].

6.1. Common Pitfalls

Poor prompt design often results from a lack of domain understanding, ambiguous phrasing, or inadequate consideration of user or organizational needs. Additionally, improper prompts can propagate undesirable biases, inaccuracies, and unproductive responses. The subsections below highlight frequent missteps, illustrating how ill-structured prompts may compromise GenAI systems' ultimate performance and trustworthiness.

6.2. Ambiguous Directives

One of the most frequent errors involves delivering instructions that are too broad, vague, or contradictory for the model to parse effectively. Such prompts frequently produce meandering or nonsensical outputs, undermining the project's objectives.

- Example: "Tell me something interesting."
- Example: "Explain the world in one sentence."

6.3. Excessive Constraints

Another improper approach involves prompts that impose stringent parameters on the model, minimizing the system's creative or inferential latitude. While clear guidance is necessary, an overly constrained prompt can stifle potentially insightful outputs.

- Example: "Answer only with exactly five words about a complex topic."
- Example: "Provide a single solution to the problem without referencing any data."

Some prompts inadvertently demand conflicting outputs or request content that cannot accurately be produced, mainly when the demands surpass model capabilities or reference non-existent data.

- Example: "Describe exactly how to cure all diseases, using five references from the future."
- Example: "Generate a precise political forecast for the next 50 years with no uncertainty."

6.4. Neglect of Context and Ethical Boundaries

Prompts that omit critical social, cultural, or ethical contexts can inadvertently lead to insensitive or biased outputs.

- Example: "Rank various cultures from best to worst based on your data."
- Example: "Generate a statement that supports discrimination against a particular group."

6.5. Debrief and Corrective Insights

The above examples illustrate how ill-conceived prompts compromise the efficacy and reliability of GenAI models. Overly broad instructions yield unfocused or convoluted content, and excessively restrictive prompts stifle the system's ability to produce nuanced, value-added information. Prompts that present contradictory or unfeasible demands provoke confusion and false claims, while a lack of ethical or cultural awareness risks perpetuating harmful stereotypes or biases [31].

To mitigate these pitfalls, practitioners should adopt a structured and mindful approach to prompt design, balancing clear directives with sufficient latitude for creativity and interpretive reasoning. Whenever possible, prompt engineers should also engage in iterative testing, monitoring, and refinement to identify and correct problematic outputs before they propagate widely. Ultimately, fostering robust prompt engineering practices and ethical oversight is pivotal for harnessing the full potential of GenAI without compromising accuracy, inclusivity, or social responsibility [26].

7. Challenges and Limitations of Prompt Engineering

Recognizing and addressing hurdles in prompt engineering is indispensable for creating more reliable, transparent, and equitable AI systems.

7.1. Limitations and Biases

Despite major strides, generative models are still prone to:

- Training Data Biases: Historical datasets may omit or underrepresent specific demographics, amplifying systemic inequalities.
- Contextual Gaps: Extended or complex queries sometimes outstrip the model's capacity to maintain accurate references.
- Unpredictable Outputs: Even meticulously designed prompts may yield unexpected results, necessitating continued vigilance in reviewing outputs.

A large-scale study revealed that about 34% of AI-generated job postings contained gendered wording that subtly favored male candidates [26], underscoring the necessity of auditing datasets and prompts for unintended stereotypes. Additionally, adversarial "prompt injection" attempts have extracted sensitive data from generative systems in around 5% of test cases [30], highlighting the importance of monitoring and securing user-facing AI tools.

7.2. Balance between Guidance and Flexibility

Achieving optimal outcomes hinges on prompt specificity without hampering the model's innate innovation ability. Restrictive parameters can confine creative latitude, but an overly broad prompt can cause the system to diverge from strategic objectives. Iteration is frequently essential, with teams methodically refining prompt techniques and obtaining counsel from domain experts or stakeholders to pinpoint the intersection of control and originality.

7.3. Quality and Reliability in AI-Generated Content

High-caliber outputs span more than bare factual accuracy:

- Rigorous Testing and Evaluation: Employ formal metrics, human review boards, or user studies to assess the clarity, relevance, and truthfulness of outputs.
- Continuous Model Improvement: Gather real-world feedback, adapt models to new data, and fine-tune prompts to maintain high-performance standards.
- Monitoring and Maintenance: Track shifts in the model's outputs over time to detect unwanted biases or content drift.

8. Sectors and Industries Poised to Benefit from GenAI and Prompt Engineering

GenAI and prompt engineering offer a powerful toolkit for organizations seeking to enhance data analysis, drive product innovation, and strengthen communication strategies. These systems can substantially improve operational efficiency and unlock fresh growth opportunities by synthesizing new information and adjusting model outputs to align with specific objectives. In what follows, we examine several high-impact sectors poised to reap transformative advantages from these emerging technologies [32].

8.1. Financial Services and Banking

Financial institutions increasingly rely on advanced algorithms to decode intricate market signals and inform strategic decision-making. Integrating GenAI with thoughtfully constructed prompts elevates these capabilities by delivering precise, contextually relevant results.

- Automated Report Generation: Personalized investment summaries and forecasts, accelerating client services and enhancing transparency.
- Risk Assessment and Fraud Detection: Prompt-driven anomaly detection systems sift through extensive financial datasets, swiftly uncovering irregular patterns and thwarting fraudulent activities.
- Scenario Simulation: Synthetic data generation under diverse what-if prompts aids banks in evaluating market responses and regulatory changes.
- Customer Engagement: Chatbots and virtual assistants, guided by structured prompts, provide precise financial guidance at scale.

A practical prompt example in the finance sector might state: "Given two years of daily trading data for a major stock index, generate a set of five plausible volatility scenarios and highlight the potential risk factors associated with each scenario."

8.2. Education and Training

In educational contexts, GenAI has the potential to reshape pedagogy, course development, and student engagement. Coupled with refined prompt engineering, these solutions ensure that outputs are tailored to diverse learning levels and subject areas [33].

Below are the principal domains where GenAI and prompt engineering can advance teaching and learning:

- Customized Learning Materials: Context-sensitive prompts yield course outlines, practice problems, or revision guides, accommodating varied learner needs and abilities.
- Student Assessment: Prompts that reflect targeted cognitive skills help educators diagnose proficiency gaps and design more personalized intervention strategies.
- Intelligent Tutoring: Responsive learning assistants calibrated through precise prompts, model problem-solving processes, and offer step-by-step guidance.
- Course Content Updates: By processing recent scholarly findings, GenAI enables rapid revisions to lesson plans, ensuring instructional content is current [34].

A practical prompt example in the education sector could be: "Generate three progressively more challenging exercises on introductory algebra designed for high school students, and provide step-by-step solutions to each problem."

In a three-month pilot, a public high school leveraged an LLM-based tutor that auto-graded assignments and answered student questions. Teachers reported a 25% reduction in grading time and a 15% improvement in student engagement [26].

8.3. Healthcare and Medical Research

Healthcare organizations benefit substantially from GenAI and prompt engineering's capacity to improve diagnostic accuracy, accelerate drug development, and optimize various administrative functions [35].

Below are prime applications where GenAI can enrich healthcare delivery and research outcomes:

- Personalized Treatment Options: Patient-specific prompts empower generative models to simulate outcomes for alternative therapies, supporting medical practitioners in selecting the best treatment plans.
- Drug and Therapy Discovery: Automated mechanisms for molecular analysis help identify promising compounds at lower cost and in shorter timeframes than conventional methods.
- Real-Time Diagnostic Support: Prompt-driven systems evaluate patient data quickly, furnishing prioritized recommendations or alerts in high-pressure clinical settings.

- Patient Communication: Language models, directed by carefully structured prompts, generate coherent responses to frequently asked questions, reducing administrative overhead and clarifying medical guidance.

A practical prompt example in healthcare might be: "Using anonymized patient records, propose three possible treatment plans for a middle-aged patient with Type 2 diabetes and coexisting hypertension, and highlight relevant clinical guidelines."

From an overarching perspective, GenAI increases care precision and expedites key medical research tasks. By examining patient data holistically and issuing prompt-driven insights, these systems allow healthcare professionals to devote more time to complex clinical judgments and patient-centered decision-making.

8.4. Other High-Impact Domains

Beyond finance, education, and healthcare, various industries can harness GenAI's predictive, analytical, and creative functionalities to achieve superior efficiency and innovation [36].

Below are examples of sectors where GenAI and prompt engineering can add remarkable value:

- Manufacturing and Supply Chain: Predictive maintenance schedules and automated design prototypes can be generated via prompt-driven models, streamlining production and curbing operational bottlenecks.
- Legal and Compliance: Drafting briefs, preparing compliance documents, and reviewing regulations become more efficient when orchestrated through prompts that yield structured, domain-consistent outputs.
- E-Commerce and Retail: Highly personalized product recommendations and dynamic design ideas build more substantial customer experiences, providing real-time adaptability in a competitive marketplace.
- Marketing and Content Creation: Targeted campaigns and novel messaging formats can be generated through prompts that integrate brand narratives with pertinent consumer data.

A practical prompt example relevant to manufacturing might be: "Given a set of sensor readings from a production line over the past month, generate three potential equipment failure scenarios and propose a maintenance schedule to mitigate these risks."

The scope of GenAI's relevance spans an impressive spectrum of operations. From predictive analytics in supply chain management to data-driven marketing and retail innovation, these technologies aid decision-makers in tackling complexity, reducing uncertainties, and driving sustained organizational growth [37].

8.5. Unleashing Value Across Multiple Domains

In sectors ranging from financial services to healthcare and beyond, GenAI demonstrates its capacity to automate and enrich core operations, leveraging prompt engineering as a bridge between user-driven requirements and machine intelligence [38]. By carefully structuring these prompts, organizations capture the precise insights or solutions they seek, enabling wide-ranging efficiencies, faster innovation, and personalized offerings. As data-centric methodologies continue to converge, GenAI's ability to provide adaptive outputs at scale is a pivotal catalyst for economic and societal progress. Ultimately, thorough implementation of prompt engineering practices positions businesses and institutions to thrive in today's dynamic, information-driven landscape [39].

9. Emerging Trends in Prompt Engineering

These forward-looking areas will likely expand GenAI's potential, growing its acceptance and strategic importance across numerous enterprises and academic domains.

9.1. Utilizing Advanced AI Models and Techniques

Ongoing research continues to amplify generative modeling:

- Fine-Tuning and Transfer Learning: Adapt large foundational models to specific tasks, such as legal analysis or molecular design, by integrating domain-specific layers [40].
- Multimodal AI Models: Fuse text, images, audio, and video to enrich comprehension and facilitate context-aware generation.
- Contextual AI and Memory Mechanisms: Introduce extended memory or retrieval systems that broaden a model's ability to manage complex, multi-step interactions and references.

By pursuing these developments, AI practitioners aim to engineer platforms that seamlessly navigate real-time complexities and provide actionable insights in increasingly sophisticated scenarios.

9.2. The Fusion of Human and AI Creativity

Collaboration between humans and advanced AI stands to shape entirely new creative spaces:

- Co-Creation and Brainstorming: Teams use AI-driven outputs as starting points or augmentations, applying professional acumen to polish and refine.
- Augmented Expertise: Specialists in medicine, law, and research can benefit from draft analyses or visual aids generated by AI, thereby improving both efficiency and quality.
- Ethical and Responsible Creativity: Ensuring content respects cultural norms and includes multiple perspectives remains critical as AI influences more creative fields.

In this paradigm, AI is a computational tool and a partner in imagining and shaping novel concepts.

9.3. Prompt Engineering in the AI-Driven Economy

Prompt engineering is quickly becoming a cornerstone of digital transformation and value creation:

- Business Communications: Automated drafting accelerates the production of proposals, memos, and reports, boosting organizational throughput.
- Accelerated Research and Development: Synthesizing data expedites prototyping and hypothesis testing, unlocking new opportunities across science and engineering.
- Personalized Customer Experiences: Companies cultivate stronger relationships and heightened customer loyalty by aligning outputs to individual preferences.

- Fostering Creativity and Innovation: Interactive, prompt-driven tools inspire fresh product concepts and services, bolstering long-term growth and strategic adaptability.

Astute, prompt engineering becomes a linchpin for maintaining a competitive edge as more enterprises integrate AI into their core processes.

10. Master Prompt Engineering with these critical tips and best practices

By adhering to practical guidelines and thoughtfully revising prompts, individuals and organizations can harness the extensive benefits of AI-generated content, mitigating common risks [41].

10.1. Embarking on a Journey into Prompt Engineering

- Understand your AI model's underpinnings, including architecture, training data, and known performance characteristics.
- Define concrete objectives and performance benchmarks, enabling systematic assessment of content quality and success metrics.
- Initiate with simple prompts, gradually incorporating nuanced details and complex constraints to refine outputs.
- Experiment with input style and parameter variations, collecting multiple model outputs for comparative assessment.
- Employ quantitative measures—like relevance or perplexity—and human review to evaluate clarity, factual accuracy, and creative fidelity.
- Share insights with the broader AI community, fostering a collaborative environment that accelerates best practices and innovation.
- Continuously prioritize ethics and accountability by detecting and addressing biases or misinformation that may surface.

10.2. Developing a Productive Workflow

1. Planning and Research: Delve into pertinent literature, existing architectures, and domain-specific challenges.
2. Objective Setting: Pinpoint the precise functions your generated content serves, from brief synopses to imaginative expansions.
3. Drafting Initial Prompts: Assemble multiple prototype prompts that reflect different degrees of specificity.
4. Testing and Evaluation: Assess the model's outputs systematically, noting clarity, functional relevance, and creativity.
5. Iteration and Refinement: Fine-tune prompts based on performance data and critique from stakeholders or subject experts.
6. Monitoring and Maintenance: Remain vigilant regarding shifts in model outputs, updating your prompts and data sources as required.
7. Collaboration and Continuous Learning: Facilitate team communication and stay current on emerging findings and tools in prompt engineering.

10.3. Addressing Typical Challenges

Although prompt engineering is powerful, it comes with notable obstacles:

- AI Model Constraints: Context window limits and domain mismatches may lead to oversights or inaccuracies.
- Managing Bias: Employ systematic checks, post-processing steps, or filtering to flag or mitigate biased outputs.
- Avoiding Overly Restrictive Prompts: Over-constraining the prompt can stifle the model's inventive dimension, while vague prompts can produce irrelevant or meandering text.
- Guaranteeing Content Quality: Clearly specify formatting criteria, ensure factual correctness, and regularly test model responses for consistency.
- Fostering Creativity: Utilize open-ended prompts or iterative feedback methods to overcome generative stagnation.
- Addressing Ethical Considerations: Incorporate disclaimers, implement reliability checks, and uphold societal standards and regulatory requirements.

10.4. Measuring the Success of Your Results

Several metrics help evaluate performance and impact:

- Content Quality (clarity, syntactical correctness, lexical richness)
- Content Novelty (degree of originality, minimal overlap with existing data)
- Content Accuracy (verified alignment with reliable sources, domain fidelity)
- User Satisfaction (positive engagement scores, survey feedback)
- Efficiency (reduced time and computational overhead)
- Adaptability (applicability to diverse tasks or evolving scenarios)

Organizations can enhance AI-driven initiatives by tracking these parameters and iterating on prompt designs to meet or surpass evolving objectives.

11. Conclusion

This document has examined the synergistic domains of GenAI and prompt engineering, revealing how they collectively anchor a new wave of intelligent data-processing and content-generating systems. GenAI empowers the automated production of novel outputs, spanning textual discourse, graphical elements, and molecular constructs, whereas prompt engineering channels these capabilities toward specific objectives with ethical and contextual finesse. Appreciating the historical evolution, prominent model structures, and most prevalent applications in these fields enable practitioners and scholars to optimize their AI deployments across various industrial and research contexts. As advancements continue toward broader multimodal applications, deeper contextual understanding, and more prosperous human-AI collaboration, prompt engineering remains indispensable for shaping productive, trustworthy, and forward-thinking AI solutions that address the emerging priorities of a data-driven global economy.

12. Future Works

Future efforts might refine formal frameworks for prompt creation, systematically linking prompt attributes to desired outcome characteristics. Another exciting direction involves the development of adaptive prompts that draw on iterative user interactions and feedback, fine-tuning a model's output dynamically in real-time. Additionally, exploring next-generation multimodal or continuous data streams could enhance generative models' versatility, enabling them to flourish in use cases involving real-time sensor data or live video feeds. Lastly, ensuring transparency, equity, and regulatory compliance in GenAI remains a priority for public confidence and ethical standards. By delving further into these research domains, the community can guide future AI systems toward excellent reliability, interpretability, and responsible innovation, facilitating breakthroughs in how societies and industries harness the power of intelligent automation.

While this paper focuses on general best practices and conceptual frameworks, future work could include direct experimental comparisons of distinct prompting strategies or GenAI architectures. Evaluating prompt styles, output quality, or error rates across multiple domains would provide more substantial evidence of each method's effectiveness and improve real-world applicability.

Furthermore, a systematic literature review could shed more light on multimodal prompt design or domain-specific fine-tuning. A deeper analysis of long-term societal implications—covering potential misuse, emergent behaviors, and ethical oversight—would strengthen preparedness for rapid AI adoption. Such expansions could broaden the reference pool to reflect contemporary sources and top-tier AI conference findings, complementing the foundational material presented here.

Acknowledgment

Maikel Leon would like to acknowledge the support provided by the ICSRI – Intelligent Computer Systems Research Institute. We thank the anonymous reviewers for their constructive criticism, valuable comments, and suggestions.

References

[1] M. León, R. Bello, K. Vanhoof, "Cognitive Maps in Transport Behavior," in 2009 Eighth Mexican International Conference on Artificial Intelligence, 179–184, IEEE, 2009, doi:10.1109/MICAI.2009.31.

[2] M. Leon, L. Mkrtchyan, B. Depaire, D. Ruan, K. Vanhoof, "Learning and clustering of fuzzy cognitive maps for travel behaviour analysis," Knowledge and Information Systems, **39**(2), 435–462, 2013, doi:10.1007/s10115-013-0616-z.

[3] M. León, "Fuzzy Cognitive Maps as a Bridge Between Symbolic and Sub-Symbolic Artificial Intelligence," International Journal on Cybernetics & Informatics, **13**(4), 57–75, 2024, doi:10.5121/ijci.2024.13405.

[4] M. Leon, "Aggregating Procedure for Fuzzy Cognitive Maps," The International FLAIRS Conference Proceedings, **36**(1), 2023, doi:10.32473/flairs.36.133082.

[5] A. Ghimire, J. Prather, J. Edwards, "Generative AI in Education: A Study of Educators' Awareness, Sentiments, and Influencing Factors," 2024, doi:10.48550/ARXIV.2403.15586.

[6] M. León, N. M. Sánchez, Z. Z. García, R. B. Pérez, "Concept Maps Combined with Case-Based Reasoning in Order to Elaborate Intelligent Teaching/Learning Systems," in Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007), 205–210, IEEE, 2007, doi:10.1109/ISDA.2007.11.

[7] M. León, G. Nápoles, R. Bello, L. Mkrtchyan, B. Depaire, K. Vanhoof, "Tackling Travel Behaviour: An Approach Based on Fuzzy Cognitive Maps," International Journal of Computational Intelligence Systems, **6**(6), 1012–1039, 2013, doi:10.1080/18756891.2013.816025.

[8] M. León, "Comparing LLMs Using a Unified Performance Ranking System," 2024, doi:10.5121/ijcsit.2023.15103.

[9] M. León, H. DeSimone, "Advancements in Explainable Artificial Intelligence for Enhanced Transparency and Interpretability across Business Applications," Advances in Science, Technology and Engineering Systems Journal, **9**(5), 9–20, 2024, doi:10.25046/aj090502.

[10] M. León, "Toward the Application of the Problem-Based Learning Paradigm into the Instruction of Business Technology and Innovation," International Journal of Learning and Teaching, **10**(5), 571–575, 2024, doi:10.18178/ijlt.10.5.571-575.

[11] H. DeSimone, M. León, "Leveraging Explainable AI in Business and Further," in 3rd IEEE Opportunity Research Scholars Symposium, 2024, doi:10.1109/ORSS.2024.1234567.

[12] M. León, "Harnessing Fuzzy Cognitive Maps for Advancing AI with Hybrid Interpretability and Learning Solutions," Advanced Computing: An International Journal, **15**(5), 1–23, 2024, doi:10.5121/acij.2024.150501.

[13] M. León, "Generative AI as a New Paradigm for Personalized Tutoring in Modern Education," International Journal on Integrating Technology in Education, **13**(3), 49–63, 2024, doi:10.5121/ijite.2024.13304.

[14] M. León, "Benchmarking Large Language Models with a Unified Performance Ranking Metric," International Journal on Foundations of Computer Science & Technology, **14**(4), 15–27, 2024, doi:10.5121/ijfcst.2024.14402.

[15] M. León, "The Needed Bridge Connecting Symbolic and Sub-Symbolic AI," International Journal of Computer Science, Engineering and Information Technology, **14**(1), 1–19, 2024, doi:10.5121/ijcseit.2024.14101.

[16] M. León, "Leveraging Generative AI for On-Demand Tutoring as a New Paradigm in Education," International Journal on Cybernetics & Informatics, **13**(5), 17–29, 2024, doi:10.5121/ijci.2024.13502.

[17] M. León, G. Nápoles, C. Rodríguez, M. M. García, R. Bello, K. Vanhoof, "A Fuzzy Cognitive Maps Modeling, Learning and Simulation Framework for Studying Complex System," in New Challenges on Bioinspired Applications: 4th International Work-conference on the Interplay Between Natural and Artificial Computation (IWINAC 2011), 243–256, Springer Berlin Heidelberg, 2011, doi:10.1007/978-3-642-21326-7_27.

[18] G. Nopoles, M. L. Espinosa, I. Grau, K. Vanhoof, R. Bello, Fuzzy cognitive maps based models for pattern classification: Advances and challenges, volume 360, 83–98, Springer Verlag, 2018.

[19] M. León, L. Mkrtchyan, B. Depaire, D. Ruan, R. Bello, K. Vanhoof, "Learning Method Inspired on Swarm Intelligence for Fuzzy Cognitive Maps: Travel Behaviour Modelling," in Artificial Neural Networks and Machine Learning–ICANN 2012: 22nd International Conference on Artificial Neural Networks, Lausanne, Switzerland, September 11-14, 2012, Proceedings, Part I, 718–725, Springer Berlin Heidelberg, 2012, doi:10.1007/978-3-642-33269-2_90.

[20] G. Nápoles, Y. Salgueiro, I. Grau, M. Leon, "Recurrence-Aware Long-Term Cognitive Network for Explainable Pattern Classification," IEEE Transactions on Cybernetics, **53**(10), 6083–6094, 2023, doi:10.1109/TCYB.2022.3142284.

[21] G. Nápoles, M. Leon, I. Grau, K. Vanhoof, "FCM Expert: Software Tool for Scenario Analysis and Pattern Classification Based on Fuzzy Cognitive Maps," International Journal on Artificial Intelligence Tools, **27**(07), 1860010, 2018, doi:10.1142/S0218213018600102.

[22] M. León, B. Depaire, K. Vanhoof, "Fuzzy Cognitive Maps with Rough Concepts," in Artificial Intelligence Applications and Innovations: 9th IFIP WG 12.5 International Conference, AIAI 2013, Paphos, Cyprus, September 30–October 2, 2013, Proceedings, 527–536, Springer Berlin Heidelberg, 2013, doi:10.1007/978-3-642-41142-7_53.

[23] F. Hoitsma, A. Knoben, M. León, G. Nápoles, "Symbolic Explanation Module for Fuzzy Cognitive Map-Based Reasoning Models," in Artificial Intelligence XXXVII: 40th SGAI International Conference on Artificial Intelligence, AI 2020, Cambridge, UK, December 15–17, 2020, Proceedings, 21–34, Springer International Publishing, 2020, doi:10.1007/978-3-030-63799-6_3.

[24] M. León, G. Nápoles, M. M. García, R. Bello, K. Vanhoof, "A Revision and Experience Using Cognitive Mapping and Knowledge Engineering in Travel Behavior Sciences," Polibits, (42), 43–50, 2010, doi:10.17562/PB-42-6.

[25] H. DeSimone, M. Leon, "Explainable AI: The Quest for Transparency in Business and Beyond," in 2024 7th International Conference on Information and Computer Technologies (ICICT), IEEE, 2024, doi:10.1109/icict62343.2024.00093.

[26] M. Alier, F.-J. García-Peñalvo, J. D. Camba, "Generative Artificial Intelligence in Education: From Deceptive to Disruptive," International Journal of Interactive Multimedia and Artificial Intelligence, **8**(5), 5, 2024, doi:10.9781/ijimai.2024.02.011.

[27] J. Su, W. Yang, "Unlocking the Power of ChatGPT: A Framework for Applying Generative AI in Education," ECNU Review of Education, **6**(3), 355–366, 2023, doi:10.1177/20965311231168423.

[28] M. Leon, "Business Technology and Innovation Through Problem-Based Learning," in Canada International Conference on Education (CICE-2023) and World Congress on Education (WCE-2023), CICE-2023, Infonomics Society, 2023, doi:10.20533/cice.2023.0034.

[29] H. Wang, A. Tlili, R. Huang, Z. Cai, M. Li, Z. Cheng, D. Yang, M. Li, X. Zhu, C. Fei, "Examining the applications of intelligent tutoring systems in real educational contexts: A systematic literature review from the social experiment perspective," Education and Information Technologies, **28**(7), 9113–9148, 2023, doi:10.1007/s10639-022-11555-x.

[30] E. A. Alasadi, C. R. Baiz, "Generative AI in Education and Research: Opportunities, Concerns, and Solutions," Journal of Chemical Education, **100**(8), 2965–2971, 2023, doi:10.1021/acs.jchemed.3c00323.

[31] D. BAİDOO-ANU, L. OWUSU ANSAH, "Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning," Journal of AI, **7**(1), 52–62, 2023, doi:10.61969/jai.1337500.

[32] E. Struble, M. Leon, E. Skordilis, "Intelligent Prevention of DDoS Attacks using Reinforcement Learning and Smart Contracts," The International FLAIRS Conference Proceedings, **37**(1), 2024, doi:10.32473/flairs.37.1.135349.

[33] X. Zhai, X. Chu, C. S. Chai, M. S. Y. Jong, A. Istenic, M. Spector, J.-B. Liu, J. Yuan, Y. Li, "A Review of Artificial Intelligence (AI) in Education from 2010 to 2020," Complexity, **2021**, 1–18, 2021, doi:10.1155/2021/8812542.

[34] C.-C. Lin, A. Y. Q. Huang, O. H. T. Lu, "Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review," Smart Learning Environments, **10**(1), 2023, doi:10.1186/s40561-023-00260-y.

[35] W. Holmes, K. Porayska-Pomsta, K. Holstein, E. Sutherland, T. Baker, S. B. Shum, O. C. Santos, M. T. Rodrigo, M. Cukurova, I. I. Bittencourt, K. R. Koedinger, "Ethics of AI in Education: Towards a Community-Wide Framework," International Journal of Artificial Intelligence in Education, **32**(3), 504–526, 2021, doi:10.1007/s40593-021-00239-1.

[36] K. Zhang, A. B. Aslan, "AI technologies for education: Recent research & future directions," Computers and Education: Artificial Intelligence, **2**, 100025, 2021, doi:10.1016/j.caeai.2021.100025.

[37] L. Chen, P. Chen, Z. Lin, "Artificial Intelligence in Education: A Review," IEEE Access, **8**, 75264–75278, 2020, doi:10.1109/ACCESS.2020.2988510.

[38] G. Nápoles, I. Grau, R. Bello, M. León, K. Vanhoof, E. Papageorgiou, "A Computational Tool for Simulation and Learning of Fuzzy Cognitive Maps," in 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 1–8, IEEE, 2015, doi:10.1109/FUZZ-IEEE.2015.7338005.

[39] G. Nápoles, J. L. Salmeron, W. Froelich, R. Falcon, M. Leon, F. Vanhoenshoven, R. Bello, K. Vanhoof, Fuzzy Cognitive Modeling: Theoretical and Practical Considerations, 77–87, Springer Singapore, 2019, doi:10.1007/978-981-13-8311-3_7.

[40] M. León, "The Escalating AI's Energy Demands and the Imperative Need for Sustainable Solutions," WSEAS Transactions on Systems, **23**, 444–457, 2024, doi:10.37394/23202.2024.23.46.

[41] G. Nápoles, F. Hoitsma, A. Knoben, A. Jastrzebska, M. Leon, "Prolog-based agnostic explanation module for structured pattern classification," Information Sciences, **622**, 1196–1227, 2023, doi:10.1016/j.ins.2022.12.012.

Advances in Science, Technology and Engineering Systems Journal
Vol. 10, No. 4, 1-13 (2025)
www.astesj.com
Special Issue on Computing, Engineering and Multidisciplinary Sciences

ASTES Journal
ISSN: 2415-6698

Machine Learning Methods for University Student Performance Prediction in Basic Skills based on Psychometric Profile

Glender Brás[*], Samara Leal, Breno Sousa, Gabriel Paes, Cleberson Junior, João Souza, Rafael Assis, Tamires Marques, Thiago Teles Calazans Silva

UNA University Center, Belo Horizonte, 30640-070, Brazil

ARTICLE INFO

Article history:
Received: 15 March, 2025
Revised: 10 July, 2025
Accepted: 13 July, 2025
Online: 31 July, 2025

Keywords:
Psychometric Profile Assessment
Machine Learning
Student Performance

ABSTRACT

Ensuring the quality of higher education in Brazil presents a complex challenge, intensified by factors that directly affect students' academic performance. The pervasive influence of social media and the overconsumption of superficial digital content undermine students' ability to engage in deep comprehension, critical thinking, and the practical application of knowledge. Furthermore, inadequate preparation during the preceding educational years hinders students' ability to adapt to the academic demands of higher education, leading to difficulties in academic progression and increased dropout rates. In view of the above, this paper explores the potential of Machine Learning models (ML) in predicting the academic performance of higher education students within the Ânima Educação ecosystem, Brazil. The contribution of this work is the development of an artificial intelligence-based assessment tool called AILA that recommends personalized study content for fundamental skills such as Portuguese and Mathematics, based on the psychometric profile of each student. This approach aims to optimize the learning process by addressing individual needs, enhancing academic performance, and overcoming the challenges faced by students in the contemporary educational landscape. Psychometric profile data were collected from approximately 41,296 incoming students of the Ânima Educação universities on the following dimensions: learning, social intelligence, emotional management, socio-emotional skills, teaching method, and knowledge area of the students. The AILA ML models presented good results in predicting students' basic skills performance in the binary and regression approaches. Specifically, the CatBoost model showed an accuracy of 0.74 in predicting scores on the Portuguese and Mathematics and Logical Reasoning proficiency tests.

1. Introduction

This paper is an extension of the paper originally presented at the *2024 IEEE 12th International Conference on Intelligent Systems* [1]. The 2024 Map of Higher Education in Brazil, published by the Semesp Institute, reveals that more than a half of university students (57.2%) drop out before completing their courses [2]. A potential contributor to the high student dropout rate in Brazil's higher education system is the deficient quality of public basic education, largely due to insufficient government investment in the sector. Furthermore, the growing influence of social media and the excessive consumption of superficial digital content have contributed to a reduced capacity for deep comprehension, critical analysis, and the practical application of knowledge. In [3], the authors argued that excessive daily internet use can lead to family conflicts, impaired communication, superficial relationships, learning difficulties, anxiety disorders, and attention deficits. Consequently, students who enter university lacking essential competencies may struggle to maintain academic progress, ultimately leading some to withdraw.

A practical illustration of the substandard quality of education in Brazil is provided by the nation's performance in the Program for International Student Assessment (PISA), a study administered by the OECD (Organization for Economic Cooperation and Development). PISA evaluates the knowledge of 15-year-old students in the domains of reading, mathematics, and science. In the PISA 2022 study, which assessed 81 countries, Brazil ranked among the 20 lowest-performing nations in two of the three evaluated subjects [4]. Moreover, the country scored below the OECD average across all domains. The results indicate that approximately 50% of Brazilian students failed to reach the minimum proficiency level in the assessed areas.

Insufficiency in basic skills represents one of the main challenges faced by university students, directly impacting their academic per-

[*]Corresponding Author: Glender Brás, UNA University Center, Belo Horizonte, 30640-070, Brazil, glender.medeiros@ulife.com.br

formance. According to [5], difficulties in academic writing and text interpretation compromise the ability of students to organize ideas, write coherent texts, and adequately express their knowledge, thus hindering their performance in assessments and academic assignments. Similarly, the study of [6] reveals that gaps in fundamental mathematical skills, such as fractions and problem-solving, can hinder academic progress, especially in courses that require quantitative reasoning. These limitations often necessitate remedial coursework, which extends the duration of studies and contributes to higher dropout rates.

Furthermore, the basic skills play a critical role in the development of higher-order cognitive abilities, such as analytical thinking and problem-solving [7]. The absence of these fundamental skills can significantly impair a student's ability to engage with more complex academic content and tasks. Without a solid foundation in essential areas like literacy and numeracy, students often struggle to make connections between different concepts, analyze problems critically, and apply learned knowledge to real-world situations. As a result, their ability to develop academic autonomy is severely hindered. This lack of autonomy in turn affects their capacity to manage learning independently, which is vital for success in higher education. Students who are unable to independently navigate through academic challenges often rely heavily on external support, such as remedial courses or additional tutoring, which further extends their time in the education system.

In this way, the lack of basic skills not only directly impacts academic performance but also prevents students from developing the necessary self-regulation and problem-solving strategies that are essential for lifelong learning and success in the professional world. Consequently, addressing these deficiencies early in a student's academic career is essential for fostering both academic independence and long-term educational success [8].

Given this, a number of studies have explored the application of machine learning (ML) as a strategy to address the aforementioned issues. Some of the studies involve the use of data analysis and prediction techniques to assess students' academic performance and psychometric profiles. Psychometric profiling is a way to understand an individual's psychological and behavioral characteristics. This is done through psychometric tests that measure the cognitive abilities, personality traits, motivation, interests, and attitudes and can be used to understand how a person thinks, learns, and behaves in different situations [8].

Contemporary psychometrics has shown a significant potential in enhancing psychological assessments, particularly through the integration of ML algorithms [9]. These algorithms possess the ability to improve their performance over time, learning from new data and experiences. This capability allows the inclusion of probabilistic relationships within computer programs, enabling more nuanced and accurate predictions of individual behavior and characteristics. Unlike traditional educational technology tools that rely on predefined, rigidly programmed rules to process inputs and generate outputs, ML models offer a dynamic and adaptive approach. They can continuously evolve based on incoming data, providing greater flexibility and responsiveness in the analysis. This adaptability is a key advantage, as it enables psychometric tools to stay relevant and effective as new patterns and insights emerge, making them far more powerful in real-world applications where data and conditions are constantly changing.

In [10], for instance, a hybrid regression model was proposed to enhance the accuracy of predicting student grades in various subjects. Additionally, an optimized multi-label classifier was developed to qualitatively predict the factors that influence student performance. The model employs three dynamic weighting techniques: collaborative filtering, fuzzy set rules, and Lasso linear regression. This integration of techniques enables a more flexible and adaptable analysis of the variables that impact learning.

Despite significant advances in recent literature on student performance prediction, several limitations persist, including challenges related to imbalanced datasets, unreliable data sources, and concerns regarding the transparency and quality of artificial intelligence (AI) models [11, 12]. These limitations pose substantial challenges in the practical application of predictive tools within real-world educational environments. Specifically, they hinder the ability to generate accurate and reliable insights into student performance, which are essential for formulating evidence-based learning strategies. In the absence of high-quality data and robust model transparency, the reliability of predictions is compromised, making it difficult for educators to make informed decisions. Consequently, this undermines the effectiveness of personalized learning interventions and hampers the creation of adaptive educational strategies that can cater to the diverse needs of students. Overcoming these challenges is crucial for ensuring that predictive analytics can be used to meaningfully enhance educational outcomes and improve the overall learning experience.

In view of the above, this paper presents the AILA (Artificial Intelligence for Learning Assistance) project, that aims to study and implement an AI-based algorithm that utilizes multiple ML models to evaluate students' performances and suggest appropriate academic content to assist the students of Ânima Educação. This tool enhances the accuracy of student performance predictions by integrating diverse data sources, enabling more effective management of unbalanced data sets and improving the transparency of its recommendations.

AILA was developed to enable a more accurate diagnosis of learning gaps, thereby supporting personalized interventions aimed at improving student performance. Leveraging machine learning models, AILA generates individualized learning plans, ensuring targeted support aligned with each student's specific needs. This innovative tool aligns with the institutional and academic objectives of higher education institutions by enhancing retention rates, minimizing the need for remedial instruction, and promoting a more efficient academic trajectory. Ultimately, the algorithm offers a data-driven approach to learning, fostering continuous improvement in student outcomes and advancing educational quality.

The case study of AILA was implemented among incoming students at Ânima Educação Group, with the objective of providing personalized recommendations based on each student's psychometric profile. The Ânima Educação Group is a prominent private educational organization in Brazil, operating 25 educational brands and managing over 500 educational centers nationwide, with a student population of approximately $400,000$.

The data collection process for the mapping phase of this study occurred in two distinct methods. First, a series of questionnaires were answered by university students who have recently entered their courses at one of Ânima's higher education institutions. The Likert scale [13], a prevalent instrument in questionnaire design, was used to employ a five-point scale ranging from "never" to "always." This scale is commonly utilized in the examination of attitudes, beliefs, and behaviors. The questionnaires are organized into three dimensions of knowledge: socio-demographic, socio-cognitive, and socio-emotional. In addition, the students completed Portuguese and Logical Reasoning tests.

These data were pre-processed and, based on the quantitative average of the scores obtained in the questionnaires/tests, the students are mapped as Naive, Beginner, Apprentice or Advanced and receive a recommendation based on this taxonomy. After this mapping, the ML models are then used to predict the students' scores in the Portuguese Language and Logical Reasoning diagnostic tests. This prediction is done considering three approaches:

- **Binary Classification**: The prediction is whether the students achieved high performance (1) or low performance (0) in these tests, based on the average scores of the population and;
- **Multi-class Classification**: The target variable is the classes of the taxonomy.
- **Regression**: The models predict the students' scores;

The case study demonstrates the effectiveness of AILA's machine learning models in predicting academic performance by integrating students' psychometric variables, thereby enhancing predictive accuracy. This approach underscores the importance of incorporating AI into teaching methodologies to support both students and educators. By identifying students' strengths, weaknesses, and potential academic difficulties in advance, AILA enables the provision of targeted resources to mitigate challenges and foster academic success

A total of 41,296 students completed the aforementioned questionnaires between September 2023 and October 2024 via a custom-developed web application. In addition to data collection, this application features a user-friendly interface that presents the learning content recommended by the models in a clear and accessible manner. The following ML models were employed: CatBoost, Decision Tree (DT), Random Forest (RF), XGBoost, Support Vector Classifier (SVC), and Support Vector Regressor (SVR), all of which demonstrated strong performance across regression, binary, and multi-class classification tasks. For example, the CatBoost model achieved an accuracy of 0.74 in predicting proficiency scores in Portuguese and logical-mathematical assessments using a binary classification approach. In contrast, under a multi-class configuration, the XGBoost and Decision Tree Classifier yielded better results. A comprehensive analysis and discussion of these findings are presented in Section 5.

The rest of this paper is organized as follows: Section 2 presents a summary of the literature, containing works that explore the use of ML models to assess student performance. In Section 3 concepts and methods relevant to this research are discussed. Section 4 explains the processes carried out to test the models and provides a comparative analysis of their performance. Finally, Section 6 presents our conclusions about the study.

2. Related Works

According to [14], predicting students' academic performance has become an increasingly complex task due to the growing volume and variety of data within educational systems. The authors argue that prevailing predictive methods remain insufficient for accurately identifying the most appropriate techniques to assess student performance in higher education institutions. Additionally, the identification of factors influencing student performance remains an underexplored area requiring further investigation, highlighting the need to determine which variables exert the most significant impact on academic outcomes.

Given this, a number of studies have been conducted in the literature to explore the use of ML models as a strategy to assess student performance in various domains. Among these studies, some are particularly noteworthy due to their relevance to the current research and the significant contributions they have made. In [14], the authors conducted a comprehensive review of 162 studies that utilized ML techniques to predict student performance between 2010 and 2022. The study of [15] proposes an intelligent system based on ML to predict students' academic performance, taking into account factors such as attendance, grades, and participation in activities. Algorithms such as Random Forest and Support Vector Machines (SVM), which have been shown to be effective in analyzing academic data, were used. The model developed showed an accuracy of 85%, standing out for its ability to accurately predict performance, with great potential for personalizing pedagogical interventions and optimizing educational outcomes. A key point observed in that research is that the appropriate choice of variables (features) can significantly influence the quality of the predictions.

The use of deep neural networks (DNN) to assess the quality of English language teaching is explored in [16], offering a more effective alternative to traditional methods. With an accuracy rate of 97%, the model is able to process large amounts of data and capture the semantic nuances present in texts, facilitating evaluation in a scalable and less subjective way. The research demonstrates how automating the feature extraction process can reduce cost and time, while improving the accuracy and consistency of scores, bringing an innovative solution to the field of language teaching.

The study of [17] utilizes ML algorithms to identify low-engagement students in a social science course at the Open University (OU) and assess how engagement affects performance. The analysis included variables such as education level, assessment scores, and interactions with virtual learning environment (VLE) activities. Several ML models, including decision trees and gradient-boosted classifiers, were tested, with the best performance in accuracy and recall. A dashboard was developed to help instructors monitor student engagement and provide timely interventions, further exploring the relationship between engagement and course assessment scores.

In [18], the use of ML models is proposed to predict the development of university students' skills over the course of their studies. By analyzing performance data in assessments and extracurricular activities, the authors were able to identify patterns that allow them to predict the evolution of students' cognitive and socio-emotional skills. The results show that deep learning models are effective, achieving 90% accuracy, and provide an agile way to tailor pedagogical approaches to students' individual needs.

The application of ML in the development of flexible learning environments is examined in [19], highlighting its potential to enable new forms of personalized instruction. The study shows how ML models can be used to adapt the content and pace of teaching to the needs of each student, resulting in greater engagement and better academic outcomes. In addition, automating the assessment process helps eliminate human bias, promoting a fairer and more accurate way of assessing students. Based on data collected in real time, the study suggests how curricula and pedagogical strategies can be continuously adjusted to promote more inclusive learning. [20] presents a comprehensive analysis of the literature on how ML has been applied to identify characteristics that affect students' academic performance. The review of 84 publications found that academic and demographic variables, such as grade history and attendance, are the most commonly studied. The study indicates that, although existing models yield satisfactory performance, incorporating additional factors—such as family dynamics and students' psychological characteristics—could enhance predictive accuracy. Moreover, it emphasizes that expanding educational databases is essential to optimizing personalized interventions.

A detailed review of the use of ML in online education is presented in [21], with a particular focus on enhancing student skill acquisition. The study shows that techniques such as content personalization, automatic correction, and progress prediction have been effective in optimizing learning. However, the authors also highlight important challenges, such as privacy issues and model accuracy, and suggest that more research should be done to overcome these limitations. Collaboration between educators, researchers and platform developers is also seen as essential to maximize the positive impact of ML in education. In [22], the authors reviews the main applications of AI and ML in digital education, covering topics such as intelligent tutors, dropout prediction, adaptive learning and process automation. The work shows that artificial neural networks and SVM are the most widely used algorithms, with an emphasis on predictive models aimed at preventing dropout and improving student performance.

The study [23] explores ML application to predict the development of university students' basic skills throughout their course. Using algorithms such as RF and SVM, the research analyzed academic performance data and practical activities to identify the students most likely to succeed or struggle. The ML model proved effective in identifying patterns, allowing for faster and more personalized interventions, which could be crucial in optimizing students' learning and academic development.

In [24], the author focus is on predicting which students are at risk of dropping out of courses on online learning platforms such as MOOCs and Learning Management Systems (LMS). Using machine learning and deep learning algorithms, the study analyzed variables such as performance in assessments, engagement, and online behavior to identify students at risk at an early stage. The model, based on the RF algorithm, achieved excellent results in terms of precision and recall, demonstrating how engagement data can significantly improve the effectiveness of predictions. This allows for timely intervention to prevent students from dropping out and improve their academic performance.

The work of [25] investigates the use of deep neural networks (DNN) to predict the academic performance of students in a data structures course. The model achieved 89% accuracy when using the SMOTE oversampling technique based on students' grades from previous courses. In addition, DNN outperformed other ML algorithms, such as SVM and RFs, on several performance metrics. The research suggests that the model could be a valuable tool for identifying at-risk students early in the semester, enabling early intervention to improve academic outcomes.

Thus, several studies have been proposed to use ML models to improve the design of educational systems and create a more personalized and effective learning experience. However, challenges persist regarding feature selection, dataset size and balance, and the explanatory power and reliability of these models. Furthermore, there is a lack of studies that have applied the results of these models in real-world settings and presented the models' predictions and recommendations through a user-friendly interface.

3. Main Concepts

This section provides an overview of the main concepts and methods used in this study, aiming to make the reading smoother and the understanding more accessible. The goal is to clarify how ML techniques can be applied to assess academic performance, contributing to promoting student success. Section 3.1 outlines the fundamental concepts underlying student performance analysis and their application within the scope of this study. Section 3.2, in turn, presents and defines the ML models employed to predict deficiencies in essential academic skills.

3.1. Elements for Student Performance Analysis

As defined by [26], performance can be understood as the way someone or something acts or behaves, measured by its output. In the educational context, student performance refers to the assessment of students based on criteria that consider essential competencies for the current scenario [27]. In this study, performance analysis is conducted through three main approaches: the Likert Scale [13], the Item Response Theory (IRT) [28], and the principles of Psychometrics [8].

The Likert Scale is a widely used instrument for assessing perceptions and preferences, and it is classified as a summative assessment method [13]. It offers response options arranged in a progressive order, typically ranging from strong disagreement to strong agreement. In the context of this study, the Likert Scale is employed to measure an individual's self-perceived proficiency in a given skill, using a five-point scale with the following categories: "never," "rarely," "sometimes," "often," and "always".

The concept of psychometrics can be approached in various ways, one of which is its application in assessing an individual's psychological traits [29]. In this sense, psychometrics involves the development of measurement tools, such as tests, scales, and questionnaires, to perform a precise and valid analysis of different aspects of human behavior [8]. Moreover, psychometrics extends beyond simple measurement by emphasizing the quality and accuracy of assessment instruments. Its primary aim is to ensure that these instruments yield consistent results while effectively capturing the constructs they are intended to measure [30]. Guided by these principles, this study applied psychometric concepts to construct the profile of the analyzed students, as detailed in Section 4.

3.2. Machine Learning (ML)

The field of machine learning focuses on developing algorithms capable of learning directly from data rather than following predefined commands [31]. The primary goal is to build computational systems that, when fed with a dataset, can generate models to make predictions, classifications, or identifications based on the acquired knowledge. Within this context, we applied various ML approaches to address the proposed problem, including Decision Tree, Random Forest, Neural Network, and Support Vector Machine. The ML techniques used in this study were:

- **Decision Tree (DT):** According to [32], a Decision Tree [33] is a hierarchical, branched structure composed of nodes and branches. At each internal node, a decision is made based on a test applied to input variables, guiding the flow along specific branches. The terminal nodes, or leaf nodes, provide the predicted values of the target variable or the associated probability distributions. According to [34], the Decision Tree, known for its quick understanding and ease of implementation, is often adopted in decision support systems in the healthcare field. Its versatility makes it applicable in various domains, offering benefits in terms of efficiency and simplified operation. With the ability to generate clear and easy-to-understand analyses, DT establishes itself as a valuable tool to optimize decision-making processes in different contexts, standing out for its accessibility and effectiveness in various areas of knowledge.

- **Random Forest (RF):** The Random Forest method, as described by [33], consists of a set of classifications based on decision trees, where each tree is influenced by a random vector. This vector is generated independently and follows a uniform distribution among the trees in the forest. Various approaches can be applied to construct these vectors, including bagging, estimated selection of splits, output randomization, and estimated attribute selection. The fundamental principle of this model lies in the independence of the generated vectors for each tree individually. By gathering a large number of trees and combining their decisions, the technique aims to increase accuracy in data classification for ML problems. This is achieved because, after building the trees, the final prediction is determined based on the most voted class by the ensemble [35].

- **Support Vector Machine (SVM):** Introduced by [36], SVM is a ML technique for binary classification problems. Its approach involves transforming input vectors through a nonlinear mapping into a high-dimensional space, where a linear decision boundary is constructed with specific characteristics that ensure good generalization capability. Originally, this model was developed to handle perfectly separable datasets but was later improved to accommodate scenarios where data exhibits overlap and is not completely separable.

- **XGBoost:** As detailed by [37], XGBoost is a ML model based on decision trees, designed to maximize computational efficiency and predictive accuracy. This method is distinguished by using ”boosting,” an approach where each subsequent tree seeks to correct the errors of the previous ones. XGBoost differs from conventional methods by integrating regularization mechanisms that minimize the risk of overfitting, applying, among other strategies, data partitioning and parallelism in tree construction. Its ability to scale processing of large data volumes makes it a popular choice for a wide range of predictive problems. Additionally, the model enhances classification accuracy by coordinating the integration of multiple trees, focusing on reducing bias and variance. According to [37], XGBoost was developed to be an effective, versatile, and easy-to-implement tool, making it a predominant choice in competitions and practical applications.

- **CatBoost:** It is an ML algorithm that stands out for its ability to optimize performance in decision tree-based models, especially when dealing with categorical data. The algorithm's main innovation is the use of Ordered Target Statistics, a method that improves the encoding of categorical variables and thus reduces the risk of overfitting by preventing information from leaking improperly during training [38]. Additionally, CatBoost implements a symmetric boosting approach, which ensures greater training efficiency and enhances the model's ability to generalize to new data. These advancements make CatBoost an effective solution for classification and regression problems, particularly useful in contexts with large data volumes and challenging tasks, such as those in finance and marketing sectors [39].

4. Proposed Modeling

In this section, we present the models implemented for understanding the relationship between the psychometric profiles of the students and their performance in basic skills. The primary objective of this study is to develop the ML models, which aim to assess both the knowledge of the Portuguese Language and Mathematical Reasoning. The implementation of these methods will facilitate the determination of the relationship between different psychometric domains and the teaching and learning of these fundamental subjects, which can support the students with the recommendation of relevant content to improve their abilities. The information was modeled based on three output configurations:

- **Binary Classification**: categorizes students into two groups: “high performance” (1) or “low performance” (0).

- **Multi-class classification**: classifies students based on the four classes or taxonomies: Naive, Beginner, Apprentice, and Advanced, with Naive having the lowest scores and Advanced the highest.

- **Regression**: the regression analysis, in turn, is used to predict the scores in the assessments, showing the improvement of the students. In addition, logistic regression, based on the idea of 'propensity scores', is used to identify the correlation between student grades and other factors, helping to understand the elements that influence their grades [40].

The development of the predictive models was guided by the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology [41], ensuring a structured, systematic, and replicable approach. The use of CRISP-DM enabled a logical progression from data understanding to model implementation, ensuring that each step effectively contributed to the quality and accuracy of the predictions. Figure 1 shows the process of this methodology.

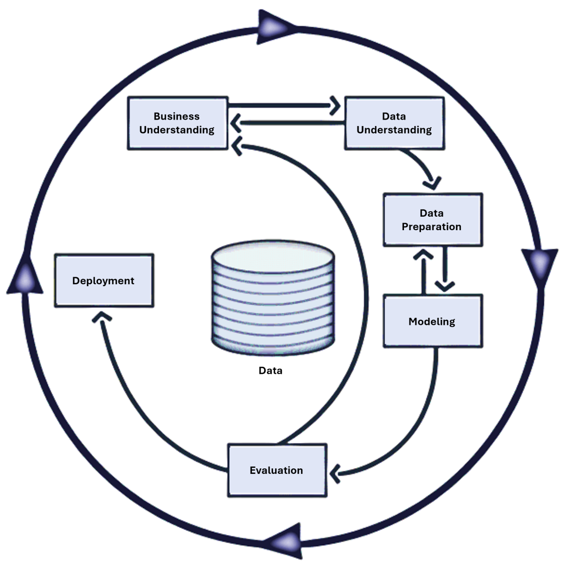

Figure 1: Phases of CRISP-DM applied to this study.

The application of CRISP-DM began with **business and data understanding**, during which the project objectives were defined and the characteristics of the data collected from student questionnaires were analyzed. These questionnaires — comprising emotional states, cognitive behaviors, and academic routines — underwent a rigorous preparation process conducted by educators from Ânima, which was essential to ensuring data quality and reliability.

During the **data preparation** phase, standardization and encoding techniques were applied to ensure data consistency and compatibility, optimizing it for subsequent modeling. This careful preparation was essential for generating a robust dataset that served as the foundation for building effective predictive models.

Following the data preparation stage, the process advanced to the **modeling** and **evaluation** phases, during the ML models were trained, fine-tuned, and rigorously assessed for performance. Finally, in the **deployment** phase, AILA's models and their artifacts were structured to ensure reusability and seamless integration into production environments. A web-based application was developed to make the models' personalized learning recommendations accessible to students in a user-friendly and practical manner, thereby completing the CRISP-DM cycle and establishing this study as a practical contribution to real-world educational contexts.

4.1. Data Understanding and Preparation

As previously mentioned, the data used to train and test the ML models were collected through psychometric questionnaires designed to assess students' emotional states, cognitive behaviors, and academic routines, along with diagnostic tests measuring proficiency in fundamental skills (Portuguese and Logical Reasoning). These instruments were administered to incoming students at Ânima Educação institutions, as illustrated in Figure 2.

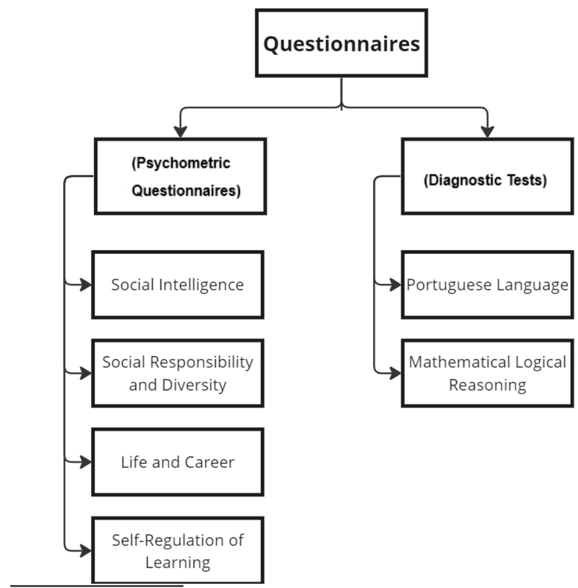

Figure 2: Questionnaire Flowchart

It is important to note that the aforementioned data were processed in accordance with Brazil's General Data Protection Law (Lei Geral de Proteção de Dados Pessoais – LGPD), which safeguards privacy and ensures information security. Participation in the study was limited to students who provided full consent, reinforcing ethical research practices and legal compliance established by Ânima Educação institutions. Although participation was voluntary, several measures were taken to mitigate self-report bias and preserve the diversity and representativeness of the sample. These included the use of validated psychometric instruments with clear, neutrally worded items; the assurance of anonymity and confidentiality to reduce social desirability effects; the incorporation of consistency checks across similar items; and the inclusion of control questions to identify inattentive responses. Moreover, behavioral frequency-based questions were prioritized over abstract self-assessments. Prior to responding, students were clearly informed about the purpose of the study and were encouraged to answer honestly

All data collected in this study were securely stored on AWS infrastructure, through an account managed by Ânima. Access to both the source code and the MongoDB database was strictly limited to the project team, in compliance with privacy and data protection requirements. The application and database are hosted on an EC2 instance, ensuring controlled and secure access aligned with industry-standard data security practices.

The psychometric questionnaires were constructed using the Likert Scale, discussed in Section 3.1, and subjected to a statistical

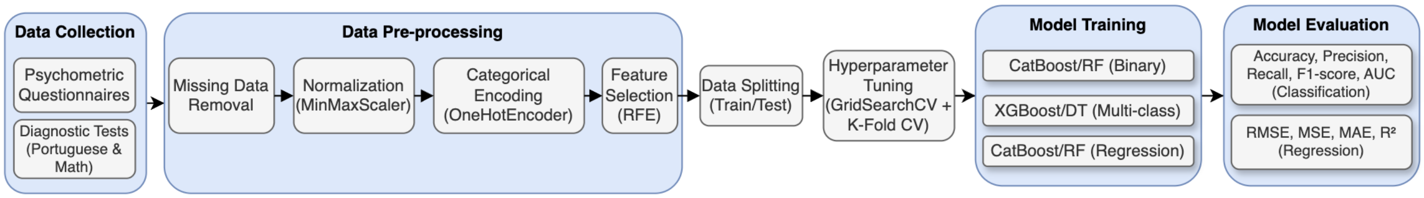

Figure 3: ML Workflow

analysis to aim to validate their effectiveness in mapping the correct profiles based on population analysis. The diagnostics were developed by the specialized professionals in pedagogical intervention.

4.2. Modeling

Figure 3 provides an overview of the ML workflow designed for this study. This modeling framework encapsulates the main stages of the process, from the collection of psychometric and diagnostic data to the training and evaluation of predictive models. This representation provides a clear and systematic overview of the process, enhancing methodological transparency and supporting the reproducibility of the study.

4.2.1. Data Pre-processing

The preliminary processing of data is of the utmost importance for the construction of predictive models, as it ensures the quality of the information. To this end, we employed data derived from questionnaires described in Section 4.1. To ensure data standardization and compatibility, we implemented variable normalization and coding techniques.

We applied the numerical data to undergo *MinMaxScaler* normalization, which scales the values between 0 and 1. This normalization enhances model stability by mitigating distortions caused by features with varying scales and units, such as the number of clicks on VLE activities compared to assessment scores [42]. By standardizing the numerical data, the model can more effectively learn relevant patterns without being disproportionately influenced by any single feature.

For categorical variables, the *OneHotEncoder* method was used, converting each category into binary representations. This approach ensures that the model does not assign any hierarchical or ordinal relationships between categories, treating each one independently and without bias [42]. By encoding variables such as different VLE activities (e.g., forums, resources) separately, the model can better capture the impact of each activity on student engagement. These preprocessing techniques enhance the quality of the data and improve the ability of the ML models to accurately predict student engagement.

4.2.2. Used Models

A comparative analysis was conducted to evaluate the efficacy of several models to predict student's performance in this study. The analysis revealed that XGBoost exhibited superior performance in terms of efficiency in decision trees and the capacity to process substantial volumes of data. CatBoost also demonstrated satisfactory results, particularly in scenarios involving disorganized data. Additionally, simple decision trees were utilized for comparative purposes, along with Random Forest, which was noteworthy for its stability and predictive capacity. For complex data, SVM was used, as it performs well on many variables [43].

4.2.3. Parameter Adjustment and Validation

The models were adjusted using the method *GridSearchCV*, an exhaustive search of various combinations of adjustments, with the aim of identifying the best option for prediction. To prevent the model from learning too much from the training data alone, k-fold cross-validation was used, ensuring a good forecast on new data. Studies demonstrate that this practice improves the prediction of academic performance by reducing statistical errors [44].

4.2.4. Evaluation Metrics

Accuracy is a widely used metric for evaluating the performance of a specific model, reflecting the ratio of correct predictions to total observations [45]. This metric can be used to further evaluate a model by measuring the ratio of correctly predicted positive cases to total predicted positive cases. This metric is advantageous in situations with high costs of false positive results [46].

The classification and regression methods were assessed using various methodologies, enabling a more comprehensive analysis [47]. Accuracy, which represents "hits," performs optimally with balanced data, while the area under the ROC curve (AUC) is more suitable for unbalanced data [45]. Additionally, the accuracy of positive predictions and the hit rate on positives were evaluated, which are fundamental in academic settings [46]. In the context of regression models, the mean squared error (MSE) was employed as a simplification, prioritizing significant errors while making predictions in a linear fashion [48]. The root mean squared error (RMSE) is the average error of the model expressed in the variable.

5. Case Study

This section details the case study developed to test the validity of the proposed methodologies, contextualizing the study and its results. The study involved 41,296 incoming students from various Ânima Educação institutions. The primary objective was to assist students in overcoming their challenges from the very beginning of their higher education journey, thereby optimizing their academic trajectory right from the outset. To achieve this, the selection of the student sample for the algorithms was based on the availability of complete individual data. Specifically, only students with comprehensive information from all the applied tests were included in the sample.

The distribution of binary classes reveals a significant imbalance in both Portuguese and Math scores. In both subjects, the vast majority of students scored below the classification threshold, while a considerably smaller fraction achieved or exceeded this mark. This class imbalance demands attention in the development of predictive models, as it can negatively impact model performance and generalization capabilities (Figure 4).

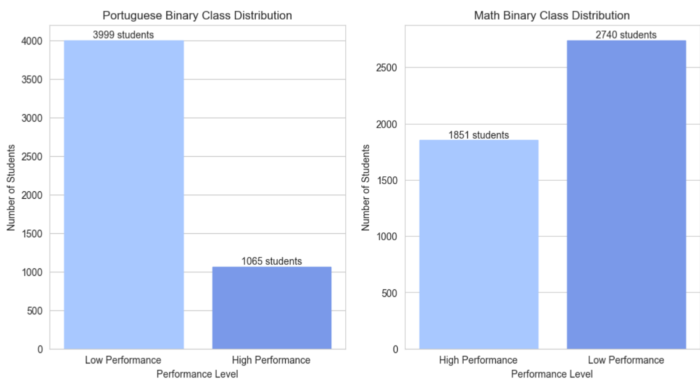

Figure 4: Class distribution Binary Classification

Furthermore, the distribution of learning taxonomies presents interesting patterns. In Portuguese, the "Beginner" category shows the highest concentration of students, while the "Advanced" category is the least represented. In Math, a similar pattern is observed, with the "Beginner" category predominating and "Advanced" being the least common. This uneven distribution among learning categories suggests a need for differentiated pedagogical approaches to address the specific needs of each group (Figure 5).

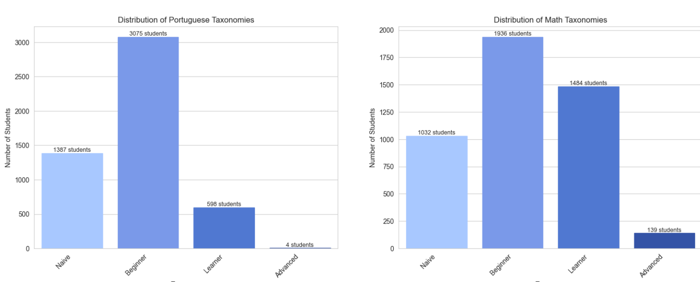

Figure 5: Class distribution Multi-class Classification

This approach was essential, as the supervised learning algorithms used in the study require a full set of labeled data to effectively learn and make accurate predictions. Incomplete data could lead to biased models or reduced prediction accuracy, making it crucial to ensure that only students with complete data were considered. By focusing on students with full datasets, the study aimed to maximize the reliability and validity of the predictions generated by the ML algorithms.

5.1. Experiments

To carry out the experiments, a process was structured that involved analyzing the students' answers to a series of questions, with the aim of measuring different aspects of learning. From these answers, a consolidated data set was generated in a CSV file, containing the features shown in Table 1. The implementation of the AILA algorithm was carried out in the Visual Studio Code environment, using the Jupyter extension to facilitate interactive code execution. The language chosen was Python, due to its wide range of specialized ML libraries.

Table 1: Features and Outputs

Feature	**Output**
Modality	Logical/Math Reasoning Tax.
Knowledge Area	Portuguese Lang. Tax.
Tax. Learning	Score Logical/Math Reasoning
Tax. Soc. Intell.	Score Portuguese Lang.
Tax. Life/Career	-
Tax. Emot. Mgmt.	-
Tax. Soc. Resp.	-
Score Learning	-
Score Soc. Intell.	-
Score Life/Career	-
Score Emot. Mgmt.	-
Score Soc. Resp.	-

Initially, the data was subjected to a preparation pipeline, which included removing missing values and transforming the variables. To avoid bias in the models, incomplete entries were eliminated, resulting in 4,499 lines. Subsequently, the numerical attributes were normalized using *MinMaxScaler*, ensuring that all the variables were on the same scale, in the 0 to 1 range. *OneHotEncoder* was used to transform categorical variables into numerical representations suitable for machine learning algorithms. The least relevant variables, based on their predictive importance, were gradually removed by Recursive Feature Elimination (RFE), reducing the dimensionality of the data set and improving the performance of the models.

This study used the approach of splitting the data into training and test sets with the help of scikit-learn's train-test-split function. This technique makes it possible to split the data randomly, so that a fraction of it is used to train the model, while the other fraction is used to evaluate its performance.

After preparing the data, the model's hyperparameters were optimized using *GridSearchCV*, a method that systematically goes through a grid of predefined values to find the best combination of parameters. In the experiment, cross-validation was applied with three divisions (cv=3), and the accuracy metric was used as the evaluation criterion. The optimized set of hyperparameters was then selected, and the final model was adjusted based on this configuration, ensuring better predictive performance.

5.2. Results

The experiments were carried out with the aim of predicting the academic performance of the university's students and showed that the models based on boosted decision trees (Boosting) and Random Forest obtained the best results, especially in binary classification. The selection of attributes included variables related to the dimensions of learning, social intelligence, emotional management, and other socio-emotional skills, as well as the teaching method and area of knowledge of the students.

The study also highlighted the influence of teaching methods and students' subject knowledge on academic performance. Certain pedagogical strategies appeared to promote better learning outcomes, while discipline-specific factors also played a role in shaping students' outcomes. These findings suggest that personalized interventions, tailored to students' academic and socio-emotional profiles, could be instrumental in improving educational outcomes. Future research could explore the impact of these interventions and further refine predictive models to support data-driven decision-making in educational settings.

5.2.1. Binary Classification

The *CatBoostClassifier* and *RandomForestClassifier* models performed better in predicting students' proficiency in Portuguese and logical and mathematical reasoning. In predicting Portuguese, the CatBoost model obtained the best results, while for logical and mathematical reasoning, the *RandomForest* model performed better, as shown in Table 2.

In addition, it was observed that the models showed high metrics for class 0, which encompasses the students with the greatest difficulty in the subjects assessed. This result is particularly relevant, as it indicates that the models are able to more accurately identify the students who need the most attention and pedagogical support. The high precision and recall for class 0 reinforce the ability of these approaches to correctly discriminate between students with difficulties, making them useful tools for targeted educational interventions.

The joint analysis of the confusion matrices from the *CatBoostClassifier*, shown in Figure 6, complements these findings. For Portuguese Language, there is a good performance in identifying students in class 0, but lower accuracy in identifying those in class 1. In the case of Logical and Mathematical Reasoning, the model produced more balanced results between the classes. This combined visualization highlights the models' focus on correctly identifying students with low performance, which is the central objective of this study.

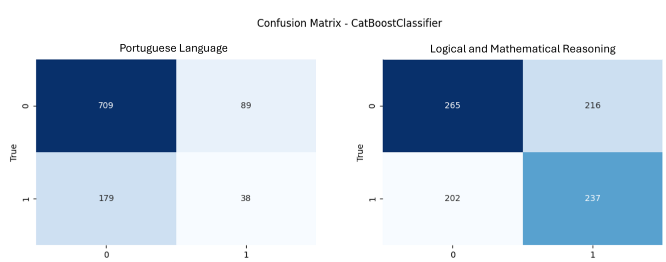

Figure 6: Confusion Matrix - CatBoostClassifier

The superior performance of the Boosting and Random Forest models can be justified by the fact that these techniques employ ensemble learning, reducing bias and variance [49]. CatBoost, in particular, uses effective processing of categorical variables and reduces the impact of overfitting, while Random Forest benefits from the aggregation of multiple decision trees, promoting robustness to the model [38].

Table 2: Binary Classification Results. PT = Portuguese language; LR = Logical and Mathematical Reasoning. Acc = Accuracy; Prec 0 = Precision for class 0; Prec 1 = Precision for class 1; F1 = F1-Score (weighted).

Model	Type	Acc	Prec 0	Prec 1	F1
CatBoost	PT	0,74	0,80	0,30	0,22
XGB	PT	0,71	0,80	0,26	0,22
DecisionTree	PT	0,53	0,82	0,25	0,35
RandomForest	PT	0,71	0,80	0,26	0,28
SVC	PT	0,53	0,82	0,25	0,35
CatBoost	LR	0,55	0,57	0,52	0,53
XGB	LR	0,52	0,55	0,50	0,51
DecisionTree	LR	0,56	0,57	0,54	0,53
RandomForest	LR	0,56	0,57	0,54	0,53
SVC	LR	0,55	0,56	0,53	0,51

5.2.2. Multi-class Classification

With regard to multi-class classification, the XGBoost and *DecisionTreeClassifier* models showed the best results for Portuguese language and logical and mathematical reasoning, respectively (Table 3). These results suggest that although Boosting models remain effective, the complexity of predicting multiple classes may have affected overall accuracy.

Table 3: Multi-class Classification Results. PT = Portuguese language; LR = Logical and Mathematical Reasoning. Acc = Accuracy; Prec = Macro-averaged Precision; F1 = Macro-averaged F1-Score.

Model	Type	Acc	Prec	F1
CatBoost	PT	0,47	0,48	0,47
XGB	PT	0,56	0,48	0,49
DecisionTree	PT	0,56	0,49	0,51
RandomForest	PT	0,43	0,48	0,45
SVC	PT	0,35	0,47	0,38
CatBoost	LR	0,37	0,36	0,36
XGB	LR	0,40	0,38	0,38
DecisionTree	LR	0,42	0,40	0,39
RandomForest	LR	0,37	0,37	0,37
SVC	LR	0,35	0,37	0,35

As shown in Figure 5, the used dataset has a much larger number of students with low performance than students with high performance. This can bias the model's learning, especially when it involves multi-class classification, and make it difficult to identify different levels among the students. For this reason the multi-class models found a low accuracy, as shown in Table 3.

Despite these limitations, the multi-class classification models still provide complementary insights regarding the distribution of student performance levels. The scarcity of examples in intermediate and high-performance classes hinders the models' ability to learn discriminative patterns for these categories, contributing to the overall lower accuracy. For the model to better classify student performance levels through the multi-class approach, more data would be needed in the other classes. However, since the critical task of identifying students with unsatisfactory performance is handled by the binary classification model—more robust and less affected by class imbalance—the limitations of the multi-class model do not compromise the practical objectives of this study. Thus, multi-class results should be viewed as a secondary analytical tool, useful for exploratory interpretation, while pedagogical decision-making is grounded on the binary model's outputs.

5.2.3. Regression

Although the *CatBoostRegressor* and *XGBRegressor* achieved the best results for Portuguese, and the *RandomForestRegressor* for logical and mathematical reasoning (Table 4), the regression models generally exhibited low predictive power. This limitation may be related to the fact that the domain scores used as predictors are not necessarily strong indicators of performance in another domain, as each one assesses distinct cognitive skills. Therefore, attempting to predict performance in a specific area based on performance in others may fail to adequately capture the unique characteristics of each evaluated competency.

Table 4: Regression Model Performance. PT = Portuguese language; LR = Logical and Mathematical Reasoning. RMSE = Root Mean Squared Error; MSE = Mean Squared Error; MAE = Mean Absolute Error; R^2 = Coefficient of Determination.

Model	Type	RMSE	MSE	MAE	R^2
CatBoost	PT	0.16	0.03	0.13	0.05
XGB	PT	0.16	0.03	0.14	0.04
DecisionTree	PT	0.17	0.03	0.14	-0.02
RandomForest	PT	0.16	0.03	0.14	0.04
SVR	PT	0.17	0.03	0.13	0.03
CatBoost	LR	0.25	0.06	0.22	0.01
XGB	LR	0.25	0.06	0.22	0.00
DecisionTree	LR	0.25	0.06	0.22	-0.04
RandomForest	LR	0.25	0.06	0.22	0.01
SVR	LR	0.25	0.06	0.22	-0.02

5.3. Discussion of the Results

The results obtained confirm that the Boosting and Random Forest models are highly effective for binary classification problems, corroborating previous studies that highlight their ability to capture complex patterns and reduce overfitting through regularization [50] [37].

The analysis of variable importance in the CatBoost model, used to predict students' performance in the Portuguese language, reveals which features had the greatest influence on the predictions. Figure 7 presents a bar chart ranking the most relevant features based on their importance values.

It is observed that scores related to life and career and learning had the greatest impact, followed by social intelligence and emotional management. These results suggest that socioemotional skills play a crucial role in students' performance, not only in content mastery but also in their ability to apply such knowledge in the exam.

Additionally, categorical variables such as the field of study (Humanities, Biological and Health Sciences, among others) and the mode of instruction (in-person or not) also showed some influence, although to a lesser extent, in predicting the results. This analysis reinforces the need for educational policies that consider not only technical knowledge but also interpersonal and emotional skills, which directly impact students' academic success in the Portuguese language exam.

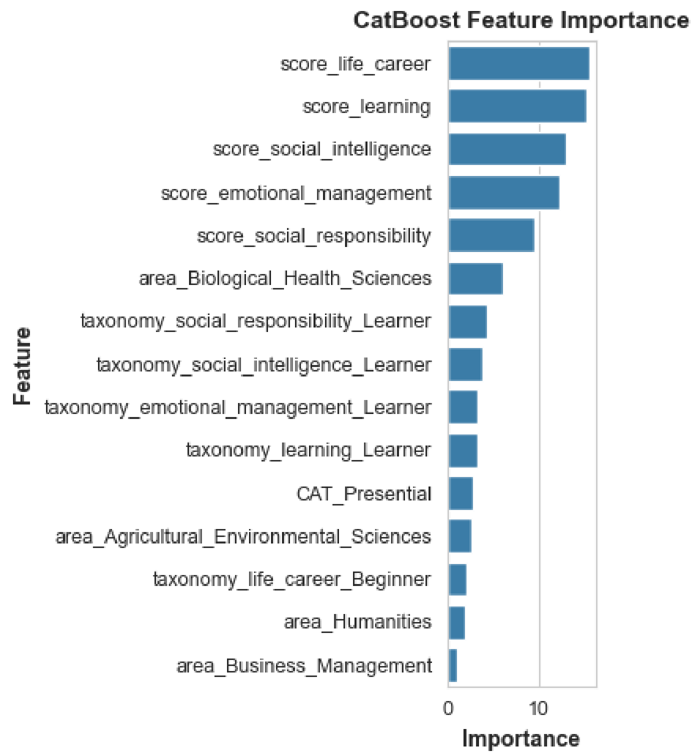

Figure 7: Feature Importances – CatBoost Model

However, multiclass prediction proved to be more challenging, possibly due to the imbalance in the data, which can guide the models learning to a good accuracy in one class but not in the others. Regarding regression, the relatively low R^2 values suggest that other factors, such as individual aspects of the students and unmeasured external factors, may have influenced academic performance.

5.4. Practical application of the recommendation

AILA employs the responses of students to psychometric questionnaires and diagnostic tests to categorize them according to a taxonomy comprising four levels: The designation of Naive, Beginner, Apprentice or Advanced is utilized to categorize individuals based on their level of expertise or proficiency in a particular domain. In accordance with the classification that has been presented, the platform provides recommendations that are customized to align with each student's unique profile.

Figure 8 presents a simulated example of the AILA's recommendation interface, which displays content suggestions categorized by dimension (e.g., logical reasoning, social intelligence, life and career). The recommendations presented are based on the student's level in each domain, with the objective of strengthening specific skills.

The practical implementation in question establishes a connection between psychometric assessment and pedagogical guidance, promoting a more student-centered learning experience.

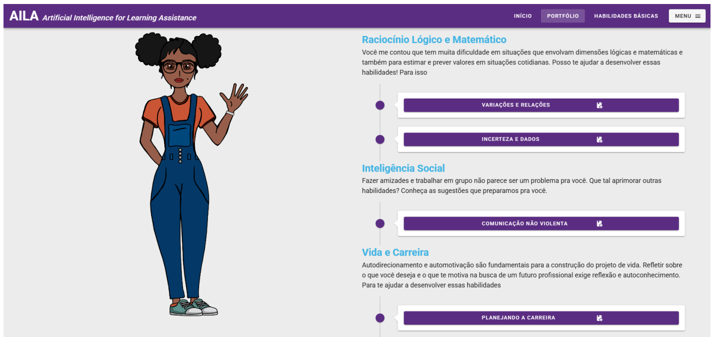

Figure 8: Simulated interface displaying personalized recommendations generated by AILA based on psychometric classification

6. Conclusion

This work explored ML models to university students performance in basic skills such as Portuguese language and Logical Reasoning. The study has been applied in a real case involving more than 40, 000 students at several institutions of Ânima Educação, an educational group in Brazil.

The results reinforce the effectiveness of using ML to predict the academic performance of higher education students and show that the models implemented are effective for predicting student performance in basic skills, especially for predicting students who will have difficulty, which is the main objective of this study. The high accuracy in class zero, in the case of binary classification, reinforces this point and can provide support to intervention actions on the part of the university. Similarly, the regression models demonstrated effectiveness in predicting the student's score and can also serve as an important method for this purpose.

As for the multi-class classification task, the models presented a low accuracy and showed that the proposed models are still not sufficient to find several levels among the students satisfactory. This is justified by the imbalance of data in the used dataset and the low amount of data in classes that represent high student performance, that is, the low number of students with high performance in basic skills in the research carried out. This confirms that data balance is crucial for this type of task.

In addition, the models proved to be effective in finding a correlation between psychometric profile and performance in basic subjects, which suggests that emotional and psychosocial factors can influence the learning process of students. For future work, it will be important to conduct new tests with more advanced students in order to improve the performance of the multi-class model, thus generating advances in the learning of the model for these cases.

However, a central limitation of this study lies in the pronounced imbalance in class distribution, particularly in the context of multi-class classification. The scarcity of students with high performance in both Portuguese Language and Logical-Mathematical Reasoning restricted the models' ability to learn representative patterns across all performance levels. This imbalance contributed to the lower accuracy observed in the multi-class models and compromised the regression results by reducing the diversity and richness of the training data. Moreover, the reliance on correlated data, such as psychometric profiles based on self-reported responses, may introduce potential bias, as subjective answers can reflect inconsistent or distorted perceptions. Future work should address these limitations through the collection of more balanced datasets, the application of techniques to mitigate self-report bias, and the use of data augmentation and resampling methods, aiming to improve the models' generalization capacity and robustness.

References

[1] G. Brás, S. S. Leal, B. Sousa, C. Junior, J. Souza, Z. Mendes, G. Paes, "Machine Learning Models for Basic Skills Identification in University Students," in 2024 IEEE 12th International Conference on Intelligent Systems (IS), 1–9, IEEE, 2024, doi:10.1109/IS61756.2024.10705207.

[2] I. Semesp, "Mapa do Ensino Superior no Brasil," in 14ª Edição, 39, 2024.

[3] T. d. O. Silva, Os impactos sociais, cognitivos e afetivos sobre a geração de adolescentes conectados às tecnologias digitais, Undergraduate thesis (tcc), Universidade Federal da Paraíba, João Pessoa, Brazil, 2016, submitted June 14, 2016; bibliographic study; advisor: Lebiam Tamar Gomes Silva.

[4] D. Salinas, F. Avvisati, R. Castaneda Velle, "PISA 2022 Results: The State of Learning and Equity in Education – Volume I," Technical report, Organisation for Economic Co-operation and Development (OECD), Paris, France, 2023, doi:10.1787/53f23881-en.

[5] R. Bailey, "Student writing and academic literacy development at university," Journal of Learning and Student Experience, **1**, 7–7, 2018.

[6] F. Ngo, "Fractions in college: How basic math remediation impacts community college students," Research in Higher Education, **60**, 485–520, 2019, doi:10.1007/s11162-018-9519-x.

[7] L. Aranda, E. Mena-Rodríguez, L. Rubio, "Basic skills in higher education: An analysis of attributed importance," Frontiers in Psychology, **13**, 752248, 2022, doi:10.3389/fpsyg.2022.752248.

[8] G. Kuan, Y. C. Kueh, N. Abdullah, E. L. M. Tai, "Psychometric properties of the health-promoting lifestyle profile II: cross-cultural validation of the Malay language version," BMC public health, **19**(1), 1–10, 2019, doi:10.1186/s12889-019-7109-2.

[9] V. R. Franco, "Aprendizado de Máquina e Psicometria: Inovações Analíticas na Avaliação Psicológica," PePisic Periodicos de Psicologia, **20**, a.c, 2021, doi:10.15689/ap.2021.2003.ed.

[10] A. Alshanqiti, A. Namoun, "Predicting Student Performance and Its Influential Factors Using Hybrid Regression and Multi-Label Classification," IEEE Access, **8**, 203827–203844, 2020, doi:10.1109/ACCESS.2020.3036572.

[11] E. P. F. Kelvin dos Santos, "A Previsão de Evasão em Cursos de Graduação Utilizando Machine Learning," Caderno Virtual, **1**(58), 13–15, 2024.

[12] A. da Silva Franqueira, B. F. C. ZanettiCarlos, A. L. Bitencourt, D. Z. Franco, E. H. B. de Oliveira, Érica Rafaela dos Santos Campos, H. G. M. Júnior, J. da Cruz Chagas, "Análise impulsionada por IA para previsão de desempenho estudantil," Cuadernos de Educación y Desarrollo, **16**(5), 14–17, 2024.

[13] A. M. Feijó, E. F. R. Vicente, S. M. Petri, "O uso das escalas Likert nas pesquisas de contabilidade," Revista Gestão Organizacional, **13**(1), 27–41, 2020, doi:10.22277/rgo.v13i1.5112.

[14] A. Abu Saa, M. Al-Emran, K. Shaalan, "Factors affecting students' performance in higher education: a systematic review of predictive data mining techniques," Technology, Knowledge and Learning, **24**(4), 567–598, 2019, doi:10.1007/s10758-019-09408-7.

[15] D. Petkovic, K. Okada, M. Sosnick, A. Iyer, S. Zhu, R. Todtenhoefer, S. Huang, "Work in progress: a machine learning approach for assessment and prediction of teamwork effectiveness in software engineering education," in 2012 frontiers in education conference proceedings, 1–3, IEEE, 2012, doi:10.1109/FIE.2012.6462205.

[16] J. Zhu, C. Zhu, Z. Wang, "Application of Machine Learning in English Language Teaching Quality Assessment," in 2024 International Conference on Interactive Intelligent Systems and Techniques (IIST), 300–304, IEEE, 2024, doi:10.1109/IIST62526.2024.00041.

[17] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, "Student engagement predictions in an e-learning System and their impact on student course assessment scores," Computational intelligence and neuroscience, **2018**(1), 6347186, 2018, doi:10.1155/2018/6347186.

[18] J. G. Valen-Dacanay, T. D. Palaoag, "Exploring The Learning Analytics Of Skill-Based Course Using Machine Learning Classification Models," in 2023 11th International Conference on Information and Education Technology (ICIET), 411–415, IEEE, 2023, doi:10.1109/ICIET56899.2023.10111210.

[19] A. Ravuri, M. Lourens, S. Aswini, G. Nijhawan, R. S. Zabibah, R. Chandrashekar, "Improving Personalized Education: A Machine Learning Method for Flexible Learning Environments," in 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 1715–1720, IEEE, 2023, doi:10.1109/UPCON59197.2023.10434888.

[20] I. Issah, O. Appiah, P. Appiahene, F. Inusah, "A systematic review of the literature on machine learning application of determining the attributes influencing academic performance," Decision analytics journal, **7**, 100204, 2023, doi:10.1016/j.dajour.2023.100204.

[21] Y. Christian, Y. Choo, N. Yusof, "Systematic literature review on the use of machine learning in online learning in the context of skill achievement," Journal of Theoretical and Applied Information Technology, **102**(6), 2466–2479, 2024.

[22] H. Munir, B. Vogel, A. Jacobsson, "Artificial intelligence and machine learning approaches in digital education: A systematic revision," Information, **13**(4), 203, 2022, doi:10.3390/info13040203.

[23] D. Petkovic, M. Sosnick-Pérez, S. Huang, R. Todtenhoefer, K. Okada, S. Arora, R. Sreenivasen, L. Flores, S. Dubey, "Setap: Software engineering teamwork assessment and prediction using machine learning," in 2014 IEEE frontiers in education conference (FIE) proceedings, 1–8, IEEE, 2014, doi:10.1109/FIE.2014.7044199.

[24] M. Adnan, A. Habib, J. Ashraf, S. Mussadiq, A. A. Raza, M. Abid, M. Bashir, S. U. Khan, "Predicting at-risk students at different percentages of course length for early intervention using machine learning models," Ieee Access, **9**, 7519–7539, 2021, doi:10.1109/ACCESS.2021.3049446.

[25] A. Nabil, M. Seyam, A. Abou-Elfetouh, "Prediction of students' academic performance based on courses' grades using deep neural networks," IEEE Access, **9**, 140731–140746, 2021, doi:10.1109/ACCESS.2021.3119596.

[26] A. Namoun, A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: A systematic literature review," Applied Sciences, **11**(1), 237, 2020, doi:10.3390/app11010237.

[27] A. M. J. d. Andrade, Desempenho acadêmico, permanência e desenvolvimento psicossocial de universitários: relação com indicadores da assistência estudantil, Master's thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, 2014, doi:10.1590/S1414-40772017000200014.

[28] E. A. C. d. Araujo, D. F. d. Andrade, S. L. V. Bortolotti, "Teoria da resposta ao item," Revista da Escola de Enfermagem da USP, **43**, 1000–1008, 2009, doi:10.1590/S0080-62342009000500003.

[29] R. Primi, "Psicometria: fundamentos matemáticos da Teoria Clássica dos Testes," Avaliação Psicológica, **11**(2), 297–307, 2012.

[30] C. F. Collares, W. L. P. Grec, J. L. M. Machado, "Psicometria na garantia de qualidade da educação médica: conceitos e aplicações," Science in Health, **3**(1), 33–49, 2012.

[31] G. M. d. M. Paixão, B. C. Santos, R. M. d. Araujo, M. H. Ribeiro, J. L. d. Moraes, A. L. Ribeiro, "Machine learning na medicina: revisão e aplicabilidade," Arquivos brasileiros de cardiologia, **118**, 95–102, 2022, doi:10.36660/abc.20200596.

[32] C. A. Meira, L. H. Rodrigues, S. A. Moraes, "Análise da epidemia da ferrugem do cafeeiro com árvore de decisão," Tropical Plant Pathology, **33**, 114–124, 2008, doi:10.1590/S1982-56762008000200005.

[33] L. Breiman, "Random forests," Machine learning, **45**, 5–32, 2001.

[34] J. P. da Silva Funchal, D. F. Adanatti, "Um Estudo Sobre a Classificação de Risco na Área da Saúde Utilizando Árvores de Decisão," iSys-Brazilian Journal of Information Systems, **9**(3), 89–111, 2016.

[35] T. M. Oshiro, Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica, Ph.D. thesis, Universidade de São Paulo, 2013.

[36] C. Cortes, V. Vapnik, "Support-vector networks," Machine learning, **20**, 273–297, 1995.

[37] T. Chen, C. Guestrin, "Xgboost: A scalable tree boosting system," in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 785–794, 2016, doi:10.1145/2939672.2939785.

[38] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, A. Gulin, "CatBoost: unbiased boosting with categorical features," Advances in neural information processing systems, **31**, 2018.

[39] J. T. Hancock, T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," Journal of big data, **7**(1), 94, 2020, doi:10.1186/s40537-020-00369-8.

[40] P. J. L. Adeodato, "Data Mining Solution for Assessing Brazilian Secondary School Quality Based on ENEM and Census Data," 13th CONTECSI International Conference on Information Systems and Technology Management, online, 2016.

[41] J. L. C. Ramos, R. L. Rodrigues, J. C. S. Silva, P. L. S. de Oliveira, "CRISP-EDM: uma proposta de adaptação do Modelo CRISP-DM para mineração de dados educacionais," in Simpósio Brasileiro de Informática na Educação (SBIE), 1092–1101, SBC, 2020, doi:10.5753/cbie.sbie.2020.1092.

[42] K. Anguraj, B. Thiyaneswaran, G. Megashree, J. Preetha Shri, S. Navya, J. Jayanthi, "Crop recommendation on analyzing soil using machine learning," Turkish Journal of Computer and Mathematics Education (TURCOMAT), **12**(6), 1784–1791, 2021.

[43] J. An, "Using CatBoost and Other Supervised Machine Learning Algorithms to Predict Alzheimer's Disease," in 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA), 1732–1739, 2022, doi:10.1109/ICMLA55696.2022.00265.

[44] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, **12**, 2825–2830, 2011, doi:10.5555/1953048.2078195.

[45] C. Liu, M. White, G. Newell, "Measuring the Accuracy of Species Distribution Models: A Review," in Proceedings of the 18th World IMACS / MODSIM Congress, 4241–4246, Victoria, Australia, 2009, reviewed accuracy metrics like discrimination capacity and reliability in SDMs; introduced both threshold-dependent and threshold-independent indices.

[46] M. Sokolova, G. Lapalme, "Learning Opinions in User-Generated Web Content," Natural Language Engineering, **17**(4), 541–567, 2011, doi:10.1017/S135132491100012X, first published online 11 March 2011; Cambridge University Press.

[47] M. Bekkar, H. K. Djemaa, T. A. Alitouche, "Evaluation Measures for Models Assessment over Imbalanced Data Sets," Journal of Information Engineering and Applications, **3**(10), 27–38, 2013, open access; ISSN 2224-5782 (print), 2225-0506 (online).

[48] R. L. Chambers, H. Chandra, N. Tzavidis, "On Bias-Robust Mean Squared Error Estimation for Pseudo-Linear Small Area Estimators," Survey Methodology, **37**(2), 153–170, 2011, develops a simpler, bias-robust MSE estimator for pseudo-linear small area estimators, including EBLUP, MBDE, and M-quantile predictors.

[49] T. Hastie, R. Tibshirani, J. Friedman, "The elements of statistical learning. Springer series in statistics," New York, NY, USA, 2001.

[50] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," Advances in neural information processing systems, **30**, 2017.

Advances in Science, Technology and Engineering Systems Journal Vol.10, No. 3, 15-20 (2025)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology 2024-25

ASTESJ

ISSN: 2415-6698

The Impact of Digitalization on Shipbuilding as Measured by Artificial Intelligence (AI) Maturity Models: A Systematic Review

Dharmender Salian[1*], Geeta Sandeep Nadella [1], Gasan Elkhodari [2], Rabih Neouchi [2], Steven Brown [1], Eduard Babulak [3], Raed Sbeit [1]

[1]*University of the Cumberlands, Williamsburg, KY 40769, USA*

[2]*Naveen Jindal School of Management Business School, The University of Texas at Dallas, Richardson, TX 75080, USA*

[3]*National Science Foundation, Alexandria, VA 22314, USA*

ARTICLE INFO

Article history:
Received: 02 February, 2025
Revised: 17 April, 2025
Accepted: 18 April, 2025
Online: 22 May, 2025

Keywords:
Digitalization
Shipping
Transformation
Technologies
Maturity

ABSTRACT

Artificial Intelligence (AI) is reshaping the global shipbuilding sector, yet existing maturity models fail to capture the domain-specific complexities of this capital-intensive industry. This study reviews over 50 AI maturity models and introduces a specialized framework tailored for shipbuilding. The proposed model outlines four progressive stages—Beginner, Innovation, Integration, and Expert—across eight key dimensions: culture, resilience, sustainability, strategy, customer focus, organizational integration, connectivity, and production efficiency. A hybrid benchmarking approach involving comparative analysis of major shipbuilders such as China State Shipbuilding Corporation(CSSC), General Dynamics National Steel and Shipbuilding Company(NASSCO), and Hyundai Heavy Industries(HHI), as well as synthesis from literature, was used to validate the relevance and coverage of each dimension. The framework provides a roadmap for operational modernization and links digital maturity to measurable outcomes such as delivery timelines, production scalability, and environmental performance. Policy recommendations highlight the need for targeted investments, workforce reskilling, and public-private collaboration to enable sustainable and AI-enabled growth in the U.S. shipbuilding sector.

1. Introduction

Artificial intelligence (AI) is redefining industrial processes globally, with the shipbuilding sector increasingly adopting AI to improve efficiency, safety, and competitiveness [1]. Nations such as China, South Korea, and Japan dominate the global shipbuilding landscape due to superior infrastructure, automation, and AI-driven capabilities [2].

Training shipyard workers in modern shipbuilding techniques and using AI will be imperative for global shipyards [3]. Organizations must understand and adapt artificial intelligence to specific uses and requirements[4]. Real-time decision-making relies on statistics, econometrics, math, simulations, and optimization to collect and analyze high-speed data from multiple sources [5]. Using current and new web data can assist organizations in identifying their competitors [6]. The Chinese shipyards enjoy state-of-the-art infrastructure, automation, and government subsidies, allowing them to reach economies of scale and construct multiple ships simultaneously [7]. China State Shipbuilding Corporation, the world's largest shipbuilding conglomerate, now owns numerous research institutes and various shipyards and builds a third of all ships worldwide. Government and shipyards work closely together to achieve their national strategic goal of being a world leader in the maritime industry [7].

The shipbuilding industry remains a vital contributor to national economies, particularly in the United States, where the sector supports over 100,000 jobs, generates $9.9 billion in labor income, and contributes $12.2 billion to GDP annually [4].

*Corresponding Author: Dharmender Salian, University of the Cumberlands, Williamsburg, KY 40769, USA, dsalian0302@ucumberlands.edu

https://dx.doi.org/10.25046/aj100303

However, despite its strategic significance, the U.S. shipbuilding sector has significant obstacles that limit its capacity to compete internationally, such as aging infrastructure, high labor costs, and a shortage of people with digital skills [8], [9]. Conversely, shipyards in China and South Korea enjoy substantial government backing, cutting-edge automation, and state-sponsored training programs, which allow them to increase production capacity and swiftly incorporate AI technologies [8]. As noted by the World Bank, the availability of a skilled labor force and policies promoting industrial transformation are key enablers of digital readiness and economic resilience in heavy industries [8]. These contrasts underscore the urgency for targeted digitalization strategies in U.S. shipyards, where AI maturity assessments can help guide sustainable modernization efforts [10].

The adoption of automation and artificial intelligence in shipbuilding can result in cost reductions, enhanced safety, and faster production cycles [11]. However, this transformation varies across regions. Chinese shipyards have rapidly embraced AI-powered robotics and analytics, whereas their U.S. counterparts continue to rely on conventional systems that prioritize operational resilience [7].

This paper aims to summarize and evaluate existing AI maturity models, assess the digitalization levels of leading shipbuilding nations, identify gaps in current AI maturity assessment frameworks as they apply to the shipbuilding industry, and propose a specialized AI maturity model tailored to address the sector's unique challenges and characteristics.

This paper evaluates how digitalization influences shipbuilding outcomes, using AI maturity models as the assessment tool. Our research objectives are:

1. To review global AI maturity models applicable to manufacturing and maritime sectors.
2. To identify limitations in existing frameworks when applied to shipbuilding.
3. To propose and validate a tailored AI maturity model for shipbuilding.
4. To assess how AI maturity affects efficiency, delivery timelines, and sustainability.

By answering research questions related to AI maturity's role in digital transformation, model effectiveness, and comparative insights, we provide a strategic lens for shipbuilding modernization.

2. Literature Review and Theoretical Foundation

We reviewed 50 scholarly sources across IEEE, Springer, ACM, and ScienceDirect databases using Boolean keywords ("AI maturity model," "shipbuilding," "digital transformation"). The review found that existing models, such as the Digital Maturity Model (DMM) [12], [13], Global Big Data Maturity Model (GBDMM) [14], do not adequately reflect the shipbuilding sector's complexity.

Key dimensions such as resilience, sustainability, and connectivity are inconsistently applied. Moreover, few models account for shipbuilding's unique regulatory, infrastructural, and labor requirements. These dimensions can be used to quantitatively evaluate an organization's AI adoption maturity practices, providing a comprehensive framework for qualitatively evaluating and improving AI adoption maturity practices. Table 1 shows the dimensions of the available maturity models, and Figure 1 shows AI dimensions.

Table 1: Dimensions in Existing AI Maturity Models

Dimension	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[14]
Culture		X		X	X		X	X	X	
Resilience	X	X	X		X					
Sustainability		X	X							
Strategy	X	X	X	X	X	X	X	X	X	X
Customer	X		X	X	X		X	X	X	
Organization	X	X	X	X	X	X	X	X	X	X
Connectivity	X	X	X	X	X		X	X	X	X
Expansive Growth	X	X		X	X		X	X	X	
Production	X		X	X		X		X		

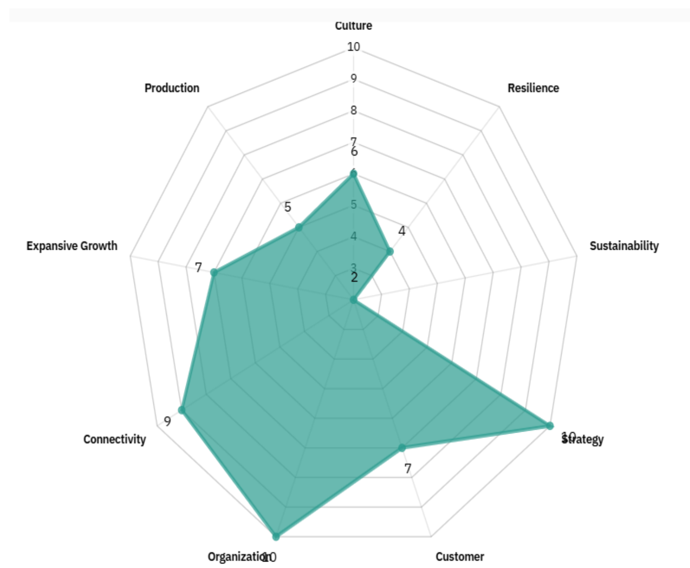

Figure 1: AI Dimensions

Note: The AI dimensions above are derived from the table and referenced models. This comparison highlights inconsistencies in how different models address key elements relevant to digital transformation in complex industries such as shipbuilding.

3. Comparative Case Study: U.S. vs. China

U.S. Shipbuilding:

- General Dynamics NASSCO focuses on defense contracts, producing 2–3 vessels per year [24].
- 2023 revenue (Marine Systems): $12.5B [24].
- Constraints: high labor costs, limited automation, reliance on military demand [25].

China Shipbuilding:

- China State Shipbuilding Corporation (CSSC) constructs ~33% of global ships [26].
- Backed by government subsidies, high R&D, and integrated AI systems [26].

Table 2 presents the annual turnover figures (in USD billions) for four major shipbuilding organizations from 2019 to 2023. China's CSSC shows a steady rise in turnover, from approximately $7.19 billion in 2019 to around $10.77 billion in 2023, reflecting its expanding global presence and state-backed initiatives. South Korea's HD Hyundai Heavy Industries (HD HHI) maintains moderate growth, with revenue climbing from $7.12 billion in 2019 to $9.2 billion in 2023. In contrast, Japan's Mitsubishi Heavy Industries (MHI) exhibits relative stability, with turnover fluctuating slightly between $1.42 and $1.67 billion, indicating a more consistent but less expansive market footprint than its peers.

Table 2: Annual Turnover (USD Billion)

Year	General Dynamics*	CSSC**	HD HHI***	Mitsubishi Heavy Industries Ltd. (MHI)****
2023	42.3	~10.77	9.2	1.42
2022	39.4	~8.14	7.05	1.42
2021	38.5	~8.16	7.25	1.50
2020	37.9	~7.55	7.03	1.67
2019	39.4	~7.19	7.12	1.66

*General Dynamics' total revenue includes revenue from other segments, including NASSCO. NASSCO does not disclose figures annually, but they are included in the segment's operations [24], [25]
**In 2019, China State Shipbuilding Industry Corporation and China Shipbuilding Industry Corporation merged, significantly boosting revenue for CSSC [26]
***HD Hyundai Heavy Industries Co., Ltd. (HHI) is a leading shipbuilding company based in Ulsan, South Korea [27]
****Historically, shipbuilding & ocean development have contributed approximately 3%–6% to MHI's total revenue [28]

4. Gaps in Existing AI Maturity Models

Most maturity models:

- Use generic stages (e.g., planning, integration, optimization)
- Lack of empirical application in the shipbuilding context
- Do not align AI dimensions with shipbuilding KPIs like delivery speed, modular construction, or regulatory compliance.

5. Proposed Shipbuilding AI Maturity Framework

Our framework includes four stages (Beginner, Innovation, Integration, Expert) and evaluates eight dimensions: Culture, Resilience, Sustainability, Strategy, Customer Focus, Organizational Integration, Connectivity, and Production Efficiency. Each stage is validated using benchmarking data.

Table 3 summarizes the four-stage progression across the eight proposed dimensions, offering a roadmap for shipbuilding organizations to assess and advance their AI maturity in alignment with industry goals.

Table 3: AI Maturity Framework

Dimension	Beginner	Innovation	Integration	Expert
Culture	Low AI literacy, resistance to change	Early experimentation with AI; supportive mindset emerging	AI embraced across teams; moderate adoption	The AI-centric culture embedded across the enterprise
Resilience	Reactive responses to disruptions	Basic forecasting using AI tools	Adaptive systems supported by AI for risk management	AI-driven autonomous resilience planning
Sustainability	Minimal awareness of green AI applications	Pilot initiatives for energy optimization	AI is used to optimize emissions and waste	Sustainability embedded as a strategic goal powered by AI

Strategy	No formal AI strategy	Isolated AI pilot programs aligned with select goals	AI aligned with business KPIs and strategic planning	AI drives strategic transformation across the organization
Customer Focus	Limited digital interaction	AI used in selected touchpoints (e.g., support bots)	Personalized services using AI analytics	Customer AI insights drive anticipatory service models
Organizational Integration	Siloed departments, ad-hoc AI efforts	Cross-functional AI collaboration begins	AI integrated into core business workflows	AI fully embedded in enterprise-wide processes
Connectivity	Low data integration, outdated systems	Partial IoT/IT-OT convergence	Real-time data pipelines and secure communications	Fully connected, interoperable, and secure digital ecosystem
Production Efficiency	Manual-heavy operations, low visibility	Initial automation in select operations	AI-optimized scheduling and predictive maintenance implemented	AI enhances throughput, uptime, and intelligent resource use

Benchmarking Approach:

A hybrid benchmarking approach validated the dimensions of the proposed AI maturity model (culture, strategy, connectivity, and sustainability):

- A literature-based review compared key dimensions across 10 AI maturity models from various domains (manufacturing, government, digital transformation). Commonly recurring dimensions were retained for inclusion.
- We conducted a benchmarking analysis using data from leading shipbuilders such as China State Shipbuilding Corporation (CSSC), Hyundai Heavy Industries (HHI) [27], and General Dynamics NASSCO [24], [25]. Publicly available performance data (e.g., delivery cycles, digital investment, vessel throughput, and automation level) were aligned with model dimensions to confirm relevance.
- The benchmarking highlighted that connectivity and sustainability dominate Asian shipyards (especially CSSC), while cultural alignment and strategic integration are critical for U.S. shipyards to improve. resilience. Table 4 shows the benchmarking analysis along different dimensions.

Table 4: Benchmarking Analysis

Dimension	Found in Literature?	Evident in CSSC?	Evident in NASSCO?	Included in Model?
Culture	Yes	Partial	Partial	Yes
Strategy	Yes	Yes	Yes	Yes
Connectivity	Yes	High	Limited	Yes
Sustainability	Moderate	High	Developing	Yes

6. Future Outlook and Validation Roadmap

As shipbuilding evolves under the pressure of environmental regulations and global competition, integrating AI, robotics, digital twins, and IoT-driven systems will be central to enhancing shipyard performance and sustainability. Nations such as China and South Korea are already advancing smart shipyards with high levels of automation, supported by public R&D funding and specialized technical education. For the United States to remain globally competitive, it must invest in modernizing commercial shipyards, foster public-private innovation ecosystems, and develop AI-skilled talent pipelines.

While the benchmarking approach in this study aligns AI maturity dimensions with operational benchmarks from leading shipbuilders, future validation efforts are essential to strengthen the model's practical application. These may include:

- Field trials in selected shipyards to assess maturity progression.
- Surveys of digital adoption across U.S. and international shipyards.
- Expert panels involving maritime engineers, defense contractors, and AI strategists will refine and validate model dimensions.
- Case-based longitudinal studies to track the impact of AI adoption on delivery efficiency and sustainability metrics.

Policymakers, business executives, and shipyard operators looking to speed up digital transformation in the maritime industry will find the model's acceptance as a strategic tool easier with the help of a well-organized validation roadmap offering empirical support.

7. Conclusion

This study proposes a sector-specific AI maturity model to guide digital transformation in shipbuilding. It bridges the gap between generic models and shipyards' unique operational challenges. Future validation through real-world pilots and international benchmarking is recommended.

Amid fierce global competition, particularly from shipbuilding powerhouses such as China, Japan, and South Korea, these nations have secured leadership through advancements in automation, AI integration, and proactive industrial policy [7] In contrast, American shipyards face structural challenges such as higher labor and material costs, aging infrastructure, and limited automation in commercial operations [24], [25]. America must reduce reliance on foreign shipbuilders and re-establish the U.S. as a key player in the global maritime landscape. It bridges the gap between generic models and the unique operational challenges of shipyards.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Y.-G. Lee, C.-H. Lee, Y.-H. Jeon, and J.-H. Bae, "Transformative Impact of the EU AI Act on Maritime Autonomous Surface Ships," *Laws*, **13**(5): 61, 2024, DOI: 10.3390/laws13050061.

[2] P. C. Hong, Y. S. Park, D. W. Hwang, and M. J. Sepehr, "A growth theory perspective on the competitive landscape of shipbuilding: a comparative study of Japan, Korea, and China," *Maritime Economics & Logistics*, **26**(3): 462–489, 2024, DOI: 10.1057/s41278-023-00279-5.

[3] B. F. Socoliuc, A. A. Suciu, M. E. Popescu, D. A. Plesea, and F. Nicolae, "Shipyard Manpower Digital Recruitment: A Data-Driven Approach for Norwegian Stakeholders," *Economies*, **13**(1): 16, 2025, DOI: 10.3390/economies13010016.

[4] "The Big Data bandwagon," *Strategic Direction*, **36**(10): 13–14, 2020, DOI: 10.1108/SD-08-2020-0144.

[5] A. Cakir, Ö. Akın, H. F. Deniz, and A. Yılmaz, "Enabling real time big data solutions for manufacturing at scale," *Journal of Big Data*, **9**(1): 118, 2022, DOI: 10.1186/s40537-022-00672-6.

[6] G. Cappellesso and K. M. Thomé, "Technological innovation in food supply chains: systematic literature review," *British Food Journal*, ahead-of-print(ahead-of-print), 2019, DOI: 10.1108/BFJ-03-2019-0160.

[7] L. Jiang and S. P. Strandenes, "Assessing the cost competitiveness of China's shipbuilding industry," *Maritime Economics & Logistics*, **14**(4): 480–497, 2012, DOI: 10.1057/mel.2012.17.

[8] World Bank, "Skilled Labor Force and Industrial Transformation," *Global Competitiveness Report*, [vol. and issue missing], [pages missing], [year missing].

[9] D. Nadolny and M. Block, "Labor cost structures and competitiveness in U.S. shipbuilding," *International Journal of Maritime Engineering*, **165**(A2): 123–134, 2021.

[10] F. K, "Legacy Systems and Interoperability in Large-Scale Engineering Projects: A Case Study of U.S. Shipyards," *System Engineering*, **24**(3): 210–225, 2022.

[11] A. Martins, "Dynamic capabilities and SME performance in the COVID-19 era: the moderating effect of digitalization," *Asia-Pacific Journal of Business Administration*, **15**(2): 188–202, 2023, DOI: 10.1108/APJBA-08-2021-0370.

[12] E. Omol, P. Abuonji, and L. Mburu, "SMEs' digital maturity: analyzing influencing factors and the mediating role of environmental factors," *Journal of Innovative Digital Transformation*, **2**(1): 19–36, 2025, DOI: 10.1108/JIDT-01-2024-0002.

[13] M. Khraiwesh, "Measures of Organizational Training in the Capability Maturity Model Integration (CMMI)," *International Journal of Advanced Computer Science and Applications*, **11**(2): [pages missing], 2020, DOI: 10.14569/IJACSA.2020.0110274.

[14] S. Mouhib, H. Anoun, M. Ridouani, and L. Hassouni, "Towards a Global Big Data Maturity Model," in *2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS)*, IEEE, pp. 1–5, 2020, DOI: 10.1109/ICDS50568.2020.9268720.

[15] J. Hu and S. Gao, "Research and Application of Capability Maturity Model for Chinese Intelligent Manufacturing," *Procedia CIRP*, **83**: 794–799, 2019, DOI: 10.1016/j.procir.2019.05.013.

[16] N. B. Yams, V. Richardson, G. E. Shubina, S. Albrecht, and D. Gillblad, "Integrated AI and Innovation Management: The Beginning of a Beautiful Friendship," *Technology Innovation Management Review*, **10**(11): 5–18, 2020, DOI: 10.22215/timreview/1399.

[17] W. Chen, C. Liu, F. Xing, G. Peng, and X. Yang, "Establishment of a maturity model to assess the development of industrial AI in smart manufacturing," *Journal of Enterprise Information Management*, **35**(3): 701–728, 2022, DOI: 10.1108/JEIM-10-2020-0397.

[18] T. Paschou, M. Rapaccini, C. Peters, F. Adrodegari, and N. Saccani, "Developing a Maturity Model for Digital Servitization in Manufacturing Firms," *[Conference Book Title Missing]*, pp. 413–425, 2020, DOI: 10.1007/978-3-030-43616-2_44.

[19] "Digital Maturity Model: Achieving digital maturity to drive growth," *Deloitte*, [year missing], [volume/issue/page missing].

[20] L. Canetta, A. Barni, and E. Montini, "Development of a Digitalization Maturity Model for the Manufacturing Sector," in *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, IEEE, pp. 1–7, 2018, DOI: 10.1109/ICE.2018.8436292.

[21] G. Valdés, M. Solar, H. Astudillo, M. Iribarren, G. Concha, and M. Visconti, "Conception, development and implementation of an e-Government maturity model in public agencies," *Government Information Quarterly*, **28**(2): 176–187, 2011, DOI: 10.1016/j.giq.2010.04.007.

[22] F. Blatz, R. Bulander, and M. Dietel, "Maturity Model of Digitization for SMEs," in *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, IEEE, pp. 1–9, 2018, DOI: 10.1109/ICE.2018.8436251.

[23] M. Kırmızı and B. Kocaoglu, "Digital transformation maturity model development framework based on design science: case studies in manufacturing industry," *Journal of Manufacturing Technology Management*, **33**(7): 1319–1346, 2022, DOI: 10.1108/JMTM-11-2021-0476.

[24] P. Novakovic, "General Dynamics: Annual Report 2023," *General Dynamics*, Mar. 2024.

[25] J. Saballa, "General Dynamics to Build More US Navy Replenishment Ships in $6.7B Deal," *The Defense Post*, [vol./issue/page/year missing].

[26] S. Wang, "China State Shipbuilding Corporation's Role in Smart Shipbuilding," *Marine Technology Reports*, **58**(2): 21–26, 2023.

[27] HD Hyundai, "Annual Report 2023," *Hyundai Heavy Industries Group*, 2024.

[28] Mitsubishi Heavy Industries Ltd., "Mitsubishi Heavy Industries Shipbuilding News," *Report*

Advances in Science, Technology and Engineering Systems Journal Vol. 10, No. 3, 21-28 (2025)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology 2024-25

ASTESJ

ISSN: 2415-6698

Cooperative Game Theory for Grid Service Pricing: A Utility-Centric Approach

Faraz Farhidi[1*], Yahia Baghzouz[2], Maxim Rusakov[3]

[1]*Department of Economics, Georgia State University, Atlanta, 30303, USA*

[2]*Department of Electrical & Computer Engineering, University of Nevada Las Vegas, Las Vegas, 89154, USA*

[3]*Evolution Networks, Ramat Gan, Israel*

A R T I C L E I N F O

Article history:
Received: 18 April, 2025
Revised: 22 May, 2025
Accepted: 24 May,2025
Online: 28 May, 2025

Keywords:
Cooperative game
Energy arbitrage
Peak load management
Price schemes

A B S T R A C T

This study presents a novel alternative to traditional Net Energy Metering (NEM) by proposing a set of innovative pricing schemes for solar customers participating in utility-led grid service programs through the aggregation of Distributed Energy Resources (DERs). Grounded in cooperative game theory, the proposed framework facilitates equitable and efficient value allocation among key stakeholders, namely customers, utilities, and aggregators—based on their respective marginal contributions to grid performance and system cost reductions. In contrast to legacy NEM structures, which typically remunerate customers at retail rates and inadequately incentivize storage adoption, load flexibility, or temporal optimization, this approach enables new revenue opportunities by embedding DERs within coordinated grid service portfolios. The pricing mechanisms developed herein are centered on two critical grid services: energy arbitrage and peak load management. These services are provisioned by the excess capacity of customer-owned DERs, particularly rooftop photovoltaic systems and behind-the-meter battery storage. Through the implementation of a Grid Services Set (GSS) and a complementary Grid Services Rider (GSR) tariff structure, participating customers voluntarily permit automated utility coordination of their devices in return for performance-based compensation. An integrated optimization algorithm co-optimizes DER dispatch across both distribution-level operational requirements and real-time wholesale market opportunities, such as those found in the Energy Imbalance Market. This enables strategic charging during periods of surplus or negative pricing and discharging during price peaks. The proposed model contributes to the advancement of Non-Wires Alternatives (NWAs) by reducing reliance on conventional infrastructure upgrades and enhancing grid flexibility and resilience. It also offers a regulatory-aligned pathway for harmonizing DER integration with utility planning objectives, renewable energy targets, and climate adaptation strategies. By fostering a cooperative paradigm between utilities and customers, the framework promotes prosocial grid behavior, scalable DER participation, and innovation in the evolving landscape of decentralized energy systems.

List of Abbreviations:

- DER: Distributed Energy Resource
- GSS: Grid Services Set
- GSR: Grid Services Rider
- ISB: Integrated Service Bundle
- NEM: Net Energy Metering
- MSP: Marginal Supply Price
- AC: Avoided Costs
- CR: Customer Revenue (Compensation Rate)
- EIM: Energy Imbalance Market
- VoLL: Value of Lost Load
- LMP: Locational Marginal Price
- PBR: Performance-Based Regulation

*Corresponding Author: Faraz Farhidi, Georgia State University E-mail: faraz.farhidi@gmail.com

https://dx.doi.org/10.25046/aj100304

1. Introduction

The Grid Services Set (GSS) is designed to effectively leverage customer-owned Distributed Energy Resources (DERs) to enhance grid operational efficiency and reduce system-wide costs of service delivery [1]. This framework reimagines the role of customer DERs—such as rooftop solar, battery storage, and smart appliances—not as passive elements, but as dynamic assets that can contribute to real-time grid reliability and resilience [2].

In the proposed approach, residential customers voluntarily opt into an Integrated Service Bundle (ISB) authorizing automated control of their DER assets within a utility-managed framework that ensures consumer protection, incentivizes energy storage adoption, and facilitates scalable energy savings through home energy management systems. By providing a more integrated and dynamic mechanism for DER participation, the ISB approach seeks to address the shortcomings of existing policies, such as net energy metering (NEM). Traditional NEM programs, implemented through net metering tariff riders (NMR), compensate customers at the retail rate for the electricity exported to the grid. This has raised equity concerns, reduced dispatch efficiency, and provided limited incentives for adopting flexible loads or storage technologies [3].

In jurisdictions such as California and Nevada, rapid adoption of Distributed Energy Resources (DERs) has outpaced the evolution of compensation models, creating policy and operational challenges. For example, California's Net Billing Tariff (NBT), also known as NEM 3.0, has replaced traditional Net Energy Metering (NEM) with a value-based export rate that better reflects grid impacts. Similarly, Nevada's revised NEM program, established under Assembly Bill 405, implements a tiered rate structure and time-of-use pricing to incentivize consumption-shifting and storage adoption. Yet both approaches often fail to fully capture the grid value of flexible DER dispatch and offer limited support for coordinated grid services.

The GSS model enables the parallel development of a Grid Services Rider (GSR)—a new tariff mechanism that outlines how participating customers are compensated for providing grid services. These services include but are not limited to voltage support, frequency regulation, peak shaving, and load shifting. The design of the GSR involves establishing metering protocols, defining billing determinants, quantifying the locational and temporal value of grid services, and implementing equitable and transparent settlement procedures.

Historically, customer-owned DERs have participated in utility-administered programs (e.g., demand response, interruptible tariffs) or in regional transmission organization/independent system operator (RTO/ISO) markets through aggregators [4]. However, participation has been limited due to the complexity of compliance, technical barriers, and a lack of coordination across devices and programs. Large commercial and industrial (C&I) customers with advanced energy management capabilities often dominate such programs, while smaller residential customers remain underrepresented [5].

In response, utilities and third-party aggregators are exploring new paradigms that simplify participation for residential customers and enable DER coordination through bundled offerings such as the ISB. Unlike conventional price-based coordination (e.g., "price-to-devices" strategies where utilities broadcast dynamic price signals to IoT-connected devices for self-scheduling [6]), the ISB emphasizes direct automated control and pre-negotiated compensation structures, simplifying participation and ensuring performance fidelity. This approach also supports distribution-level grid optimization—an increasingly important goal as electrification and DER penetration accelerate.

Moreover, competitive procurement mechanisms, where utilities solicit grid services from third-party aggregators, represent another emerging strategy, albeit with distinct implementation complexities and scalability challenges [7]. In contrast, the GSR/ISB framework offers a scalable, utility-centric pathway for integrating DERs into grid operations while maintaining regulatory oversight and aligning with public policy goals.

A foundational element of this work is the concept of excess DER capacity, which refers to the portion of a customer's DER resource that is not consumed onsite and is thus available to provide grid services. Properly tracking and monetizing this capacity requires accurate measurement of behind-the-meter energy flows and clear attribution of services performed. The proposed GSR tariff defines the mechanisms through which this excess capacity is converted into Grid Services Revenue (GS Revenue), offering solar customers an alternative to traditional NEM compensation schemes. Two specific grid services—(1) capacity reservation during critical system peaks and (2) responsive discharge during load ramps—are identified as illustrative use cases for this compensation model [8, 9].

To ensure fair and efficient distribution of the benefits arising from the aggregation and deployment of DER assets, this paper applies a cooperative game theory framework. In doing so, it proposes a utility-centric mechanism to allocate value among stakeholders—including utilities, aggregators, and individual customers—based on their marginal contributions to system reliability and cost reduction. The cooperative game theory lens has been previously applied to energy markets to explore fair revenue distribution, coalition formation, and incentive compatibility [10, 11]. In this context, the framework ensures that all parties benefit proportionately from participation, which is critical to sustained engagement and trust in utility programs.

This paper presents a game-theoretic pricing framework for DER-enabled grid services, drawing on cooperative game theory to ensure fair value allocation among stakeholders. It develops and simulates new pricing models for energy arbitrage and peak load management, incorporates real-world tariff examples, and evaluates the potential of DER coordination to support Non-Wires Alternatives (NWAs). The remainder of the paper details the design of the GSS/GSR mechanism, the cooperative value-sharing structure, simulation results, and policy implications.

2. Methodology: Cooperative Game Theory and Tariff Modeling

This service targets energy arbitrage opportunities within the Western Energy Imbalance Market (EIM), a real-time wholesale electricity market operated by the California Independent System

Operator (CAISO) that allows participants to buy and sell electricity in five-minute and fifteen-minute intervals across balancing authority areas. By leveraging co-optimization strategies, the proposed model enables Distributed Energy Resources (DERs), when aggregated under utility or aggregator management, to actively participate in this market and generate incremental Grid Services Revenue (GS Revenue) beyond local distribution-level benefits.

The underlying optimization algorithm is designed to maximize the net economic value derived from arbitrage by dynamically scheduling DER charging and discharging cycles. Specifically, the algorithm identifies periods of surplus generation—such as midday hours when solar production exceeds load demand—characterized by low or negative locational marginal prices (LMPs). During these periods, energy is stored in DER systems (e.g., home batteries, electric vehicles) under utility or aggregator control. Later, during periods of high system stress or elevated market prices, the stored energy is discharged and sold back into the grid, creating a price spread from which revenue is derived.

This model aligns with prior work demonstrating the potential of DERs to participate in energy arbitrage and ancillary services markets [12, 13]. By operating across both temporal price differentials and locational constraints, the model contributes to overall market efficiency while providing system-level benefits such as load balancing, renewable integration support, and peak demand reduction. Moreover, it highlights the dual-use potential of DERs, which can simultaneously serve local reliability needs and generate value in wholesale markets supported by recent developments in FERC Orders 2222 and 841, which expand access for aggregated DERs to wholesale markets [14].

Importantly, the cooperative game theory approach proposed in this paper ensures that the value generated through arbitrage is equitably distributed among participating customers, the utility, and other stakeholders based on their contributions to system performance. This contrasts with more centralized optimization paradigms, offering a fair and incentive-compatible structure for residential DER participation. The model also incorporates risk-adjusted dispatch constraints, including availability, degradation cost of storage devices, customer-defined operational limits, and forecast uncertainty, ensuring both robustness and customer satisfaction.

In sum, this arbitrage service extends the Grid Services Set (GSS) from a purely distribution-grid operational model to one that is interoperable with real-time market signals, supporting the vision of a transactive, prosumer-enabled grid.

To establish a transparent and equitable pricing mechanism for event-based grid services, this paper draws upon foundational principles from cooperative game theory [15, 16], particularly in scenarios where bargaining power is assumed to be equally distributed among stakeholders. The core intuition is to determine a "fair market rate" for DER-enabled grid services that simultaneously improves the net payoff for both the utility and participating customers. This framework departs from competitive or adversarial pricing schemes and instead focuses on joint value creation and benefit sharing, which is central to achieving a sustainable "win-win" equilibrium.

The cooperative model is conceptualized as a two-state system, distinguishing between the baseline case of non-cooperation and the potential for enhanced collaboration through contractual participation in grid service programs.

2.1 Non-Cooperative Baseline

In the non-cooperative scenario, the utility continues its operations under business-as-usual conditions without engaging customers in DER-driven event-based services. Customers consume energy and are billed according to their existing rate structures—typically flat rates or tiered pricing—without receiving compensation for any grid-supporting actions their DERs might be capable of. Under this scenario, no formal coordination exists between the utility and its customers regarding resource dispatch or grid service contributions.

The financial outcomes for each party in this state are modeled as follows:

- Utility Payoff per kWh:

$$U_{baseline} = FR - MSP - AC$$

Where:

- FR = Flat Rate charged to the customer per kWh
- MSP = Marginal Supply Price (i.e., cost to procure electricity from the wholesale market or EIM)
- AC = Avoided Costs, including capacity deferral, ancillary service costs, or reduced grid congestion, attributable to potential DER participation

This formulation defines the utility's net revenue per kWh without DER compensation or coordination, excluding fixed charges for simplicity. This condition is particularly relevant when $FR < MSP + AC$, as it suggests the utility may be incurring a loss for each kilowatt-hour delivered, making cooperative alternatives more attractive.

- Customer Payoff:

$$CR_{baseline} = 0$$

Since customers are not compensated for their flexibility or DER participation, they accrue no financial benefit from supporting grid services and only incur standard retail charges. This scenario sets the baseline for evaluating the marginal improvement offered by cooperation.

2.2 Cooperative Agreement with Grid Service Compensation

In the cooperative scenario, customers enter into a formalized grid service arrangement with the utility, wherein they agree to allow their DERs (e.g., batteries, smart inverters, thermostats) to be dispatched or managed in alignment with grid needs. In exchange, customers receive credit or payment (CR) for their participation, while the utility benefits from the avoided costs and potentially enhanced operational efficiency.

For such cooperation to be rationally attractive to both parties, their respective payoffs under cooperation must exceed those under non-cooperation. The utility revenue in this case adjusts to reflect the cost of compensating the customer:

- **Utility Revenue under Cooperation:**

  $U_{coop} = FR - MSP - AC - CR$

The condition for utility participation is:

$U_{coop} \geq 0 \Rightarrow FR - MSP - AC - CR \geq 0$

This inequality implies that the utility will only agree to share a portion of the avoided cost (via CR) if its net revenue remains non-negative, or ideally, improves. If the utility is experiencing a negative margin in the baseline case (i.e., FR < MSP + AC), the cooperative arrangement becomes not only viable but economically advantageous, as the avoided losses can be partially reallocated to customer compensation without creating a net loss.

- **Customer Revenue:**

  CR > 0

Under this scheme, customers receive a tangible benefit for their grid contributions, creating a clear economic incentive to participate. The cooperative framework, particularly when modeled through Shapley values or Nash bargaining solutions, can further refine the exact division of surplus based on marginal contributions, ensuring allocative efficiency and fairness.

2.3 Simulation of Tariff Schemes

The cooperative model offers several compelling advantages. It transforms passive energy consumers into active grid participants, incentivizes demand flexibility, and internalizes DER benefits into utility planning processes. Furthermore, because the model is grounded in mutual surplus generation, it creates self-enforcing agreements that do not rely on heavy-handed regulatory mandates or subsidies.

In practice, this framework can be expanded to accommodate a variety of rate designs, including time-of-use pricing, critical peak pricing, or even real-time locational prices, depending on market maturity and metering infrastructure. Moreover, the model is extensible to scenarios where customer bargaining power is not equal—e.g., in low-income or underserved communities—by incorporating weighted utility functions or social welfare constraints into the cooperative solution.

Customers' payoffs would be CR (+ FR), not only do they avoid paying the rate, but they also receive compensation for helping to improve grid reliability. To derive the fair rate for the grid service, we solve the following Nash equation incorporating the bargaining powers for both sides:

$$Max\ U\ \{(|-MSP-AC-CR+FR|)^{P}(CR)^{1-p}\}$$

Where $FR < MSP + AC$, and p is the bargaining power between the utility and the customers.

By solving the first-order condition, we derive the customer compensation rate (CR) as:

CR = (1-p)(MSP + AC – FR)

For simplicity, we can assume 50-50 benefit sharing (an equal bargaining power between utility and customers, where p = 0.5); thus, CR would be 0.5 * (MSP + AC – FR).

We can use any real time EIM nodal price as the MSP in the above formula.

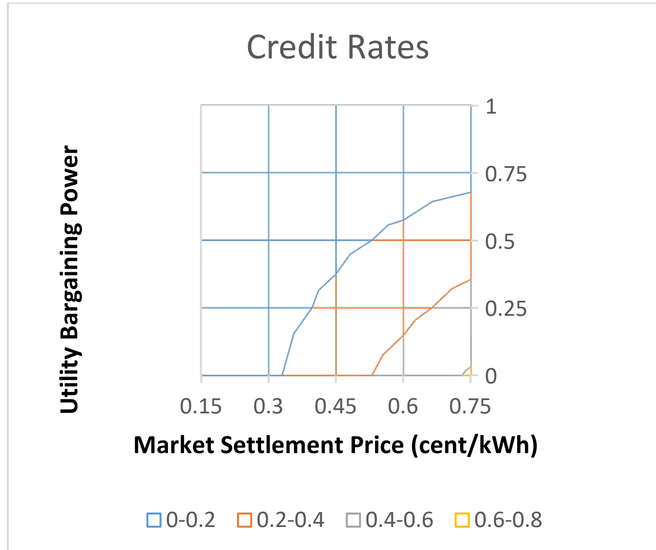

Figure 1: Simulated credit rates under different scenarios

Figure above shows a basic simulation when utility's bargaining power (P: vertical axis) changes from 0 to 1, and market price (MSP: horizontal axis) changes from 15 to 75 cents per kWh.

2.4 Energy Arbitrage Tariff Scheme Structure

The following formula can be used to compensate the customers when such service is being called:

CR = 0.5 * [MSP + ACGC + ACTC]

The avoided costs are subject to change based on the annual confirmation of the GRC filing for each utility. Here, I assume arbitrary rounded values for the purpose of this practice.

AC could be one or a combination of the costs:

Avoided cost of generation capacity (ACGC) = $ 0.03 kWh

Avoided cost of transmission capacity (ACTC) = $ 0.01 kWh

Avoided cost of distribution capacity (ACDC) = $ 0.015 kWh.

Substituting the above values, and setting a random market settlement price (MSP) to $50 MWh, we get the customers' compensation as follow:

CR = 0.5 * [0.05 + 0.03 + 0.01] = $0.045.

3. Peak Load Management

The peak load management service seeks to minimize both the system peak load and the distribution peak load at managed aggregation points through load shaping and load shed [17]. The proposed grid service offers a dual-purpose economic and emergency dispatch framework for Distributed Energy Resource (DER) assets, targeting both peak load management and distribution-level reliability enhancement. It plays a foundational role in the Grid Services Set (GSS) by enabling utilities to optimize DER dispatch across time and space in alignment with system-wide operational objectives. At its core, the service executes a real-time or near real-time optimization process whose primary objective function is to minimize total system energy procurement costs, specifically during peak demand periods, while simultaneously mitigating local distribution grid stress through location-specific dispatch incentives.

From a systems integration perspective, the formulation harmonizes transmission-level and distribution-level objectives by embedding dual-pricing signals within a single optimization framework. On the transmission side, the model ingests real-time wholesale market prices, particularly those related to system peak events or high locational marginal prices (LMPs), typically observed in the Energy Imbalance Market (EIM) or day-ahead markets. On the distribution side, it incorporates locational "shadow prices", derived from distribution-level constraints such as transformer loading, feeder congestion, or equipment thermal limits. These shadow prices act as proxies for the marginal reliability value of DER dispatch at specific nodes, enabling the system operator to prioritize areas of the grid that are more vulnerable to overload or failure during high demand.

Operationally, when the probability of distribution equipment overload—such as transformer overheating or feeder voltage violations—exceeds a defined threshold, the dispatch algorithm adjusts DER instructions to prioritize local reliability over broader system economic objectives. In such scenarios, devices located within constrained zones are directed to export energy or reduce consumption in a way that alleviates stress on the most at-risk aggregation points, thus preventing equipment damage or service interruptions. Conversely, during periods of normal or low distribution system risk, the same optimization algorithm reverts to a market-cost minimization objective, leveraging DER flexibility to reduce utility exposure to wholesale market price volatility, particularly during regional peaks or scarcity pricing events.

Importantly, this dynamic optimization process respects the operational constraints and preferences of DER-owning customers. It factors in variables such as state-of-charge limitations for batteries, comfort bands for smart thermostats, and usage patterns for behind-the-meter systems to ensure that customer experience and participation willingness are preserved. This constraint-sensitive design is critical for maintaining trust and ensuring consistent engagement in voluntary or incentive-based programs.

In cases of unexpected emergency conditions, such as system faults, weather-related disruptions, or load-forecasting errors that result in unforeseen peaks, the service includes a pre-configured rapid dispatch protocol. This protocol allows eligible DERs—particularly battery storage systems and fast-responding inverter-based technologies—to act as 10-minute spinning reserves. Devices enrolled under this protocol receive advanced configuration settings that dictate their behavior in emergency events, allowing them to respond without requiring real-time optimization or operator intervention. This capability not only strengthens distribution system resilience, but also aligns with broader grid modernization goals, such as increasing non-wire reliability options and reducing reliance on traditional spinning reserve sources.

By merging economic dispatch with reliability-based dispatch logic and enabling rapid fallback mechanisms, this service represents a multifunctional tool for modern grid operations. It enhances distribution system reliability, reduces peak demand charges, facilitates renewable integration by improving grid flexibility, and enables DERs to participate meaningfully in both energy and ancillary services markets. Moreover, architecture establishes a platform for future market-based dispatch mechanisms, potentially allowing DERs to participate in locational capacity markets or transactive energy systems where grid constraints and energy prices are jointly optimized.

3.1. Peak Load Management Tariff Schemes

To ground the proposed methodology in a realistic context, we construct a stylized example of a utility service area with moderate DER penetration. The scenario includes customer-owned rooftop solar, battery systems, and smart inverters, operating under typical Western U.S. pricing dynamics. For simulation purposes, we assume a market settlement price of \$50/MWh, avoided generation costs of \$0.03/kWh, and a residential VoLL of \$7/kWh. These inputs are used to demonstrate the energy arbitrage and peak load management compensation formulas developed in this study.

CR = 0.5 * [MSP*(1+LL) + ACGC + ACTC + ACDC + E(ICE| Utility Residents)*CDF.Norm] (load forecast, 1.1*transformer rate, load STD)]

Value of lost load (VoLL) = Expected value of interruption cost estimation (\$7 kWh for residents, that can be adjusted for inflation based on the CPI in 2016 [when the ICE calculation was estimated] and current year) * cumulative normal distribution, where X is the forecasted load, mean is the transformer/feeder capacity, and the standard deviation of the historical load on that transformer/feeder; the probability function looks as follow using excel formula:

CDF.Normal (Forecasted load, transformer rate, standard deviation between actual and backtest/backcast, True)

Line loss (LL) = 8% of the load at the peak

As an illustrative example, consider a standard substation transformer in the western region of Las Vegas with a rated capacity of 37 MVA. If the forecasted load is 36 MVA and the historical load standard deviation is 6.74 MVA, and assuming a market settlement price (MSP) of \$50/MWh, the resulting customer credit would be:

CR = 0.5 * [0.05*(1.08) + 0.03 + 0.01 + 0.015 + 7*(0.24)] = \$0.8945 kWh.

4. Discussion

The proposed Grid Services Set (GSS) and its associated cooperative pricing schemes represent a paradigmatic shift in the integration of Distributed Energy Resources (DERs) into regulated utility frameworks. Traditionally, customer-sited solar and storage assets have been compensated through static models such as net energy metering (NEM), which, despite their simplicity, have increasingly been critiqued for their misalignment with the actual value streams that DERs provide to the grid [18]. By moving beyond NEM toward a dynamic, service-based compensation framework, the GSS introduces a game-theoretic, value-reflective approach that fosters symbiotic cooperation between utilities and DER-owning customers.

From a cooperative game theory perspective, the proposed tariff design formalizes a benefit-sharing coalition between utilities and customers. Customers, in return for providing real-time grid services—such as energy arbitrage, peak shaving, and voltage support—are compensated not just for their exported kWh, but for the marginal grid value their actions create. This aligns with the Shapley value framework for cooperative games [19], where each participant is remunerated in proportion to their contribution to the coalition's total value. Such structuring addresses the free-rider problems inherent in flat or volumetric compensation schemes.

4.1. Energy Arbitrage and Market Synergies

A central component of the GSS is the energy arbitrage pricing model, which leverages hourly price signals from the Energy Imbalance Market (EIM) and enables DER participants to buy and store electricity during low-price periods and discharge or export during high-price windows. This approach mirrors utility-scale arbitrage strategies already employed by grid operators and independent power producers but adapts them to the residential and commercial customer scale through automated control systems and smart contracts.

This democratized arbitrage model benefits utilities by:

- Shaving peaks and reducing marginal procurement costs
- Improving load shape and net demand predictability
- Minimizing dependence on peaker plants, which are often carbon-intensive and expensive to operate

Simultaneously, customers gain access to non-linear revenue streams beyond flat-rate bill reductions, making participation more economically attractive and sustainable long-term. The application of formula-based compensation models, adjusted dynamically to market prices and system needs, ensures transparency and predictability in customer payments while remaining value-aligned with system conditions.

4.2 Peak Load Management and Reliability Contributions

Another key innovation in the proposed framework is the integration of DERs into distribution-level peak load management. By deploying localized DER dispatch in a coordinated fashion, either through virtual power plant (VPP) aggregations or utility-orchestrated demand response, the grid can mitigate distribution and system-level constraints more efficiently [20]. This is especially critical in high-DER penetration environments where feeder-level constraints, reverse power flow, and voltage excursions become more prevalent.

Importantly, the use of Value of Lost Load (VoLL) as part of the compensation metric recognizes the reliability value that customer DERs contribute during high-stress grid events. This valuation approach is consistent with reliability-centered planning in utilities and reflects current best practices in performance-based ratemaking and resource adequacy compensation [21]. Incorporating VoLL reinforces customer engagement while addressing equity concerns by compensating for both energy and capacity value provided.

4.3 Implementation Challenges and Regulatory Considerations

Despite the theoretical and practical benefits of the GSS model, several challenges require careful consideration for successful implementation:

- Automated DER Participation and Customer Trust

Effective participation in the GSS framework depends heavily on real-time automated control of DERs, either via customer-side energy management systems or utility aggregation platforms. This raise concerns around customer autonomy, data privacy, and cybersecurity—areas that are increasingly scrutinized under evolving federal and state guidelines. Transparent governance structures, opt-in/opt-out flexibility, and clear data ownership policies will be essential for fostering long-term customer trust.

- Advanced Metering and Billing Infrastructure

The proposed pricing schemes require granular metering (e.g., 5-minute intervals) and advanced billing platforms capable of real-time settlements and post-hoc performance validation. While many utilities are investing in AMI (Advanced Metering Infrastructure), not all service territories are equally prepared. Therefore, regulatory support and cost recovery mechanisms must be aligned to facilitate these capital expenditures, particularly in vertically integrated utility structures.

- Policy Alignment and Market Integration

Full deployment of the GSS model will also require harmonization with state-level policy directives, including renewable portfolio standards, decarbonization mandates, and equity goals. Pilot programs, sandbox testing environments, and performance-based regulation (PBR) models may serve as intermediaries to test the framework's effectiveness before wider rollout. Moreover, coordination with wholesale markets (e.g., ISO/RTOs) is necessary to avoid value duplication and ensure accurate settlement of grid services at both distribution and transmission levels.

4.4 Statistical Inference on Value Distribution

To evaluate the robustness of the proposed pricing scheme, we simulated a range of market settlement prices (MSP) from \$30/MWh to \$75/MWh and applied corresponding avoided cost values with ±20% variability, reflecting annual utility cost filings. The resulting customer compensation rates (CR) varied between \$0.035/kWh and \$0.10/kWh. A Monte Carlo simulation with 10,000 trials, drawing MSP and avoided cost parameters from triangular distributions, yielded an expected CR of \$0.062/kWh with a standard deviation of \$0.011. This inference supports the conclusion that even under cost volatility, the cooperative scheme consistently generates nontrivial value for participating customers. Moreover, 95% of the simulated outcomes exceeded a baseline zero-compensation NEM scenario, indicating statistical dominance of the cooperative framework.

5. Conclusion and future directions

This paper proposes a utility-centric, cooperative game-theoretic framework for pricing distributed energy resources (DERs) that participate in grid service programs. The study introduces the Grid Services Set (GSS) and the associated Grid Services Rider (GSR) tariff as scalable mechanisms to integrate customer-owned DERs—such as rooftop solar and battery storage—into both distribution-level operations and real-time wholesale markets. The proposed compensation structure departs from traditional Net Energy Metering (NEM) models by reflecting the marginal grid value of DER contributions, rather than static volumetric offsets. Through cooperative value-sharing principles, particularly those derived from Shapley value and Nash bargaining concepts, the framework ensures equitable distribution of system benefits among utilities, aggregators, and customers.

Key findings include:

- The demonstration of cooperative pricing schemes that internalize avoided capacity, reliability, and market arbitrage benefits into customer compensation.
- A dual optimization approach that co-optimizes DER dispatch for both grid resilience (e.g., peak load management) and market revenue (e.g., energy arbitrage in the Energy Imbalance Market).
- The use of risk-adjusted and customer-sensitive constraints to balance economic efficiency with customer participation willingness and equity.

These findings collectively support a shift toward dynamic, service-based DER valuation that can align utility financial interests with policy goals around decarbonization, affordability, and grid modernization.

Future research directions include:

- Empirical testing and validation through pilot programs in diverse regulatory and market environments to assess the real-world feasibility and customer responsiveness to cooperative DER pricing.
- Integration of advanced forecasting and optimization tools, including machine learning algorithms, to enhance the precision of dispatch schedules and pricing signals under uncertainty.
- Exploration of differentiated pricing strategies to account for socioeconomic factors, ensuring equitable participation across income levels and geographies.
- Institutional design and governance research to determine optimal structures for utility-aggregator-customer coordination, particularly in vertically integrated versus deregulated markets.
- Regulatory analysis to identify pathways for harmonizing GSR-type tariffs with performance-based regulation and wholesale market participation frameworks, such as those enabled by FERC Orders 841 and 2222.

By advancing both the theoretical and practical foundations for cooperative DER integration, this study contributes to a more adaptive and equitable energy system in the face of increasing decentralization and climate imperatives.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was supported in Part by the Solar Energy Technology Office of the U.S. Department of Energy under Grant No. DE-EE0009022.

References

[1] P. Denholm, R. M. Margolis, J.M. Milford, "Production cost modeling for high levels of photovoltaics penetration," National Renewable Energy Laboratory 2010, doi: 10.2172/924642

[2] T. Navidi, A. El Gamal, R. Rajagopal, "Coordinating distributed energy resources for reliability can significantly reduce future distribution grid upgrades and peak load," Joule, **7**(8), 1769-1792, 2023, DOI: 10.1016/j.joule.2023.06.015.

[3] S. Borenstein, "The Private Net Benefits of Residential Solar PV: The Role of Electricity Tariffs, Tax Incentives, and Rebates," Journal of the Association of Environmental and Resource Economists, **4**(S1), S85–S122, 2017, DOI: 10.1086/691978.

[4] R. Matsuda-Dunn, L. Leddy, E. Hotchkiss, M. Gautam, M. Abdelmalak, "What Role Do Aggregators Play in Power System Security and Resilience?" Preprint. Golden, CO: National Renewable Energy Laboratory. NREL/CP6A40-85649, 2023, DOI: 10.1109/RWS58133.2023.10284615.

[5] S.P. Burger, M. Luke, "Business models for distributed energy resources: A review and empirical analysis," Energy Policy, 109, 230–248, 2017, DOI: 10.1016/j.enpol.2017.07.007.

[6] J.L. Mathieu, D.S. Callaway, S. Kiliccote, "Examining uncertainty in demand response baseline models and variability in automated responses to dynamic pricing," In IEEE Conference on Decision and Control, 2011, DOI: 10.1109/CDC.2011.6160795.

[7] R. Hledik, T. Lee, "Load flexibility: Market potential and opportunities in the United States," In Variable Generation, Flexible Demand (pp. 195-210). Academic Press, 2021, DOI: 10.1016/B978-0-12-823810-3.00001-7.

[8] E. Hittinger, J. Siddiqui, "The challenging economics of US residential grid defection." Utilities Policy 45: 27-35, 2017, DOI: 10.1016/j.jup.2016.11.003.

[9] F. Farhidi, M. Rusakov, M. "An Application of Excess Solar and Storage Capacity Optimization for Grid Services," Journal of Energy and Power Technology, **6**(3), 1-25, 2024, DOI: 10.21926/jept.2403017.

[10] W. Saad, Z. Han, H.V. Poor, T. Başar, "A noncooperative game for double auction-based energy trading in microgrids," In 2011 IEEE International Conference on Smart Grid Communications, 2011, DOI: 10.1109/SmartGridComm.2011.6102323.

[11] K. Khezeli, H. Firoozi, "A Cooperative Game Theory Approach for Fair Cost Allocation in Peer-to-Peer Energy Trading," Sustainable Energy, Grids and Networks, 21, 100299, 2020, DOI: 10.1016/j.segan.2020.100299.

[12] R. Sioshansi, P. Denholm, T. Jenkin, J. Weiss, J. "Estimating the value of electricity storage in PJM: Arbitrage and some welfare effects," Markets Operated by Regional Transmission Organizations and Independent System Operators, 2020, Docket No. RM18-9-000.

[13] F. Farhidi, K. Madani, "A game theoretic analysis of the conflict over

Iran's nuclear program," IEEE International Conference on Systems, Man, and Cybernetics, pp. 617-622, 2015, DOI: 10.1109/SMC.2015.118.

[14] K. Madani, F. Farhidi, S. Gholizadeh. "Bargaining Power in Cooperative Resource Allocations Games," Algorithms 15, no. 12: 445, 2022, DOI: 10.3390/a15120445.

[15] S. Iqbal, M. Sarfraz, M., Ayyub, M., Tariq, R.K. Chakrabortty, M.J. Ryan, B. Alamri, "A comprehensive review on residential demand side management strategies in smart grid environment," Sustainability, **13**(13), 7170, 2021, DOI: 10.3390/su13137170.

[16] S. Borenstein, "The Private Net Benefits of Residential Solar PV: The Role of Electricity Tariffs, Tax Incentives and Rebates," Journal of the Association of Environmental and Resource Economists, 4(S1), S85-S122. 2017, DOI: 10.1086/691978.

[17] L.S. Shapley, "A Value for n-person Games," Ann. Math. Stud. 28, 307–318, 1953, DOI: 10.1515/9781400881970-018.

[18] H. Gao, T. Jin, C. Feng, C. Li, Q. Chen, C. Kang, "Review of virtual power plant operations: Resource coordination and multidimensional interaction," Applied energy, 357, 122284, 2024, DOI: 10.1016/j.apenergy.2023.122284.

[19] M. Najafi, A. Akhavein, A. Akbari, M. Dashtdar, "Value of the lost load with consideration of the failure probability," Ain Shams Engineering Journal, *12*(1), 659-663, 2021, DOI: 10.1016/j.asej.2020.05.012.

[20] Energy Economics, **31**(2), 269–277, 2009, DOI: 10.1016/j.eneco.2008.10.005.

[21] E. Hittinger, J.F. Whitacre, J. Apt, J. "What properties of grid energy storage are most valuable?" Journal of Power Sources, 206(1), 436–449, 2010, DOI: 10.1016/j.jpowsour.2011.12.023.

Advances in Science, Technology and Engineering Systems Journal Vol. 10, No. 2, 28-34 (2025)

www.astesj.com

Special Issue on Computing, Engineering and Multidisciplinary Sciences

ASTESJ

ISSN: 2415-6698

AI-Based Photography Assessment System using Convolutional Neural Networks

Surapol Vorapatratorn[*], Nontawat Thongsibsong

Center of Excellence in AI and Emerging Technologies, School of Applied Digital Technology, Mae Fah Luang University, Chiang Rai, 57100, Thailand

A R T I C L E I N F O

Article history:
Received: 12 January, 2025
Revised: 01 March, 2025
Accepted: 02 March, 2025
Online: 18 March, 2025

Keywords:
Automated assessment
Deep learning
Convolutional neural networks
Education Technology
Image classification
AI in education

A B S T R A C T

Providing timely and meaningful feedback in photography education is challenging, particularly in large classes where manual assessment can delay skill development. This paper presents M-Stock, an AI-based automated photo evaluation system that uses Convolutional Neural Networks (CNNs) to assess student photography assignments on web browser. M-Stock evaluates both technical aspects (such as lighting, composition, and exposure) and creative elements, providing students with real-time, formative feedback. The system was trained on a diverse dataset, including student submissions and commercial standards, achieving an overall accuracy of 97.18% with an average prediction speed of 46.1 milliseconds per image. Experiments assessed the system's performance across varying resolutions and batch sizes, confirming its scalability and suitability for real-time classroom use. Additionally, a pilot study with students indicated that M-Stock's feedback positively impacted their technical skills and encouraged self-directed learning. The results demonstrate M-Stock's potential as a transformative tool for photography education, combining high accuracy, immediate feedback, and pedagogical value to support continuous learning. Future improvements will focus on refining creative assessments and expanding the system's applicability to other visual arts disciplines.

1. Introduction

In recent years, digital technology has revolutionized the way photography is taught, offering students unprecedented access to resources and tools for developing their skills. University-level courses on photography increasingly emphasize both theoretical knowledge and practical expertise, aiming to produce competent professionals equipped for the rapidly evolving media and creative industries [1]. However, as photography courses expand in scope and enrolment, especially in digital classrooms, educators face significant challenges in efficiently assessing student work [2]. The task of providing timely, meaningful feedback is often hindered by the volume of student submissions, which can delay the developmental process of photography skills [3].

Traditional assessment methods for photography assignments are often manual and time-consuming, leading to delays that can impede learning and limit student engagement. Studies have highlighted that real-time feedback plays a significant role in accelerating skill acquisition in domains requiring both technical precision and creative expression [4]. Given this, automated assessment systems powered by artificial intelligence (AI) have emerged as promising tools for enhancing the learning experience. AI technologies, especially deep learning, have shown considerable potential in automating visual assessments, enabling more personalized, consistent, and timely feedback for students [5].

Despite these advancements, current AI-based assessment systems in photography education primarily focus on evaluating technical attributes, such as lighting, composition, and exposure, often overlooking the creative and subjective aspects critical to artistic development. Furthermore, many existing tools provide only summative feedback, offering a one-time evaluation rather than iterative feedback that supports continuous learning and improvement. Addressing these gaps requires an assessment platform that can balance both technical and creative evaluations while also offering formative, actionable feedback.

[*]Corresponding Author: Surapol Vorapatratorn, Mae Fah Luang University, Surapol.vor@mfu.ac.th

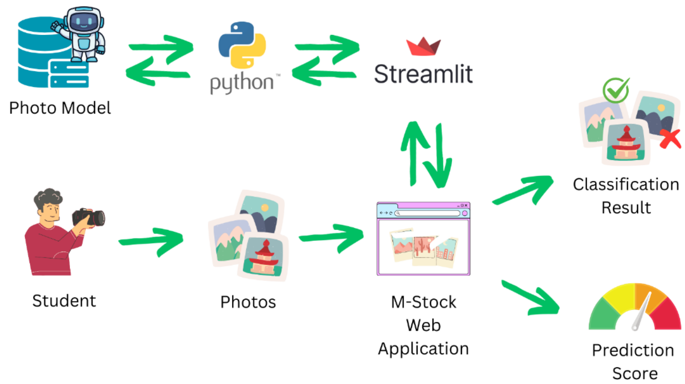

Figure 1: Overall structure of our proposed system

This paper introduces M-Stock (Mae Fah Luang University Photo Stock), an AI-driven automated photo evaluation platform designed to support student learning in photography by providing real-time feedback on both technical and artistic elements of their work. Using Convolutional Neural Networks (CNNs), M-Stock evaluates photographs based on predefined criteria developed in consultation with industry standards and educational experts, thus ensuring both relevance to the professional field and pedagogical value. In addition, M-Stock is built with scalability and ease of use in mind, allowing seamless integration into classroom environments where students can receive immediate feedback on their submissions. Overall structure of our proposed system as depicted in Figure 1.

2. Related Work

The integration of artificial intelligence (AI) in education has shown promising potential to enhance learning outcomes by providing personalized and adaptive feedback across various fields. In creative education, AI-driven assessment tools have been increasingly applied, yet challenges remain, particularly in domains like photography, where both technical and creative competencies are essential. This section reviews recent advancements in AI-supported educational systems, focusing on automated assessment in creative disciplines and identifying key gaps that the M-Stock system aims to address.

2.1. AI in Education for Automated Assessment

AI technologies, particularly deep learning, have transformed educational assessment by enabling automated grading and personalized feedback systems. These systems have proven effective in evaluating diverse student outputs, including essays, problem-solving exercises, and visual projects, providing more timely feedback than traditional methods. Adaptive learning environments and intelligent tutoring systems use AI to tailor educational content and assessment to individual learners' needs, which has been shown to improve learning efficiency and engagement [6]. Furthermore, as 21st-century learning frameworks emphasize critical thinking, creativity, and lifelong learning [7], AI-based assessments must evolve to support these skills, especially in creative subjects like photography. Furthermore, AI technologies have been applied in the context of stock photography. Platforms such as Shutterstock and Adobe Stock have incorporated AI algorithms to evaluate the quality of images submitted by photographers, offering real-time feedback and ensuring that only images meeting commercial standards are accepted [8]. This use of AI for large-scale image evaluation highlights its potential for integration into photography education, where it can be used to assess student submissions and provide immediate feedback on technical aspects such as focus, lighting, and composition [9]. However, most existing systems in this category are optimized for structured and quantifiable tasks, such as quizzes and assignments that focus on objective metrics. This approach is limited in addressing subjective assessments, such as those required in photography education, which involve creative expression and aesthetic judgment.

2.2. Automated Assessment in Photography Education

In photography education, AI-based assessment tools have typically focused on evaluating technical attributes, such as exposure, sharpness, and composition. As professional photography requires both technical proficiency and artistic expression, it is critical that educational tools reflect industry standards and expectations [10]. Recent research by [11], has explored AI-supported assessment in photography, demonstrating the potential for Convolutional Neural Networks (CNNs) to classify images based on technical quality. Such systems provide valuable feedback for improving technical proficiency but often lack the capability to assess the creative and subjective qualities of an image. Moreover, many existing tools in photography education offer only summative feedback, which does not facilitate iterative improvement and skill refinement, both of which are critical for creative learning. Unlike these existing systems, M-Stock aims to bridge this gap by integrating both technical and creative evaluations, providing formative feedback that encourages continuous learning. The system's feedback is designed not only to assess basic technical aspects but also to guide students in enhancing their artistic interpretation and aesthetic sensibilities, offering a more comprehensive educational experience.

2.3. Convolutional Neural Networks for Image Classification

Convolutional Neural Networks (CNNs) have emerged as a robust tool for image classification, widely applied in various fields, including medical imaging, autonomous driving, and creative media [12]. CNNs excel at identifying spatial hierarchies and features in visual data, making them well-suited for assessing technical quality in photography. While CNNs have demonstrated high accuracy in image classification, most studies in this area have focused on technical metrics without exploring how these models might be adapted to assess creative and subjective qualities in educational contexts. Other deep learning models, such as transformers and attention-based networks, have also shown success in visual tasks, providing an alternative to CNNs. However, CNNs remain the primary choice for this study due to their well-established efficiency and proven effectiveness in photography-related tasks. Future iterations of M-Stock could explore alternative models or ensemble approaches to further enhance its evaluative capabilities, particularly for assessing creativity.

2.4. Existing Gaps in Automated Photography Assessment

Despite the advances in AI-based assessment tools, significant gaps remain in the automated evaluation of creative student outputs. Most current systems excel at objective assessments, but they struggle to capture subjective elements, such as artistic style and emotional impact, which are essential in photography education [13]. Additionally, the lack of iterative, formative feedback in current photography assessment tools limits their effectiveness in supporting continuous skill development. The need for systems that can provide nuanced, ongoing feedback on both technical and creative elements of student work remains largely unmet. In response to these challenges, M-Stock was designed to provide a balanced approach to automated photography assessment, incorporating both technical and creative evaluations. By integrating AI-based formative feedback, M-Stock addresses the limitations of existing systems, offering students timely, constructive feedback that promotes self-directed learning and skill enhancement.

3. Proposed Method

The M-Stock system was developed to automate the assessment of student photography, providing a balanced evaluation that addresses both technical and creative aspects of students' work. This section outlines the methodology used to design, implement, and evaluate the M-Stock system, focusing on data collection, model training, feedback mechanisms, and system architecture.

3.1. Data Gathering

The M-Stock system's training dataset combines images from two primary sources to cover diverse photography skills and quality levels: *Student Assignments from Photography Courses:* Images were collected from photography courses at Mae Fah Luang University. These assignments covered various topics such as fast shutter speed, long shutter speed, night light photography, composition and subject, aperture and depth of field, light and shadow, portrait photography, moving subjects, and product photography. The assignments were submitted via Google Classroom [14], and each image was categorized into three performance levels: Excellence, Good, and Bad, based on criteria established by instructors and photography experts as shown in Figure 2.

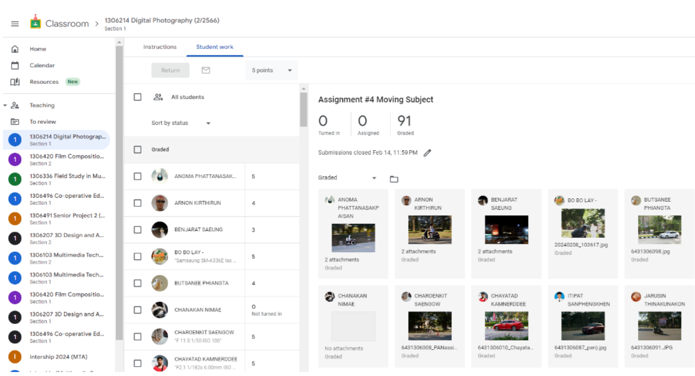

Figure 2: Digital Photography Assignment in Google Classroom

Commercial Standards from Shutterstock Submissions: To integrate professional criteria, the dataset includes student submissions to Shutterstock [15], labelled as either Accepted (commercially viable) or Rejected (commercially inadequate). This source introduces real-world standards into the model, making it robust for assessing quality in a manner that aligns with industry requirements as depicted in Figure 3. Images were stored in a server database, organized by assignment type and quality category. This collection strategy ensures that the M-Stock model can generalize well across different photography styles, skill levels, and educational contexts.

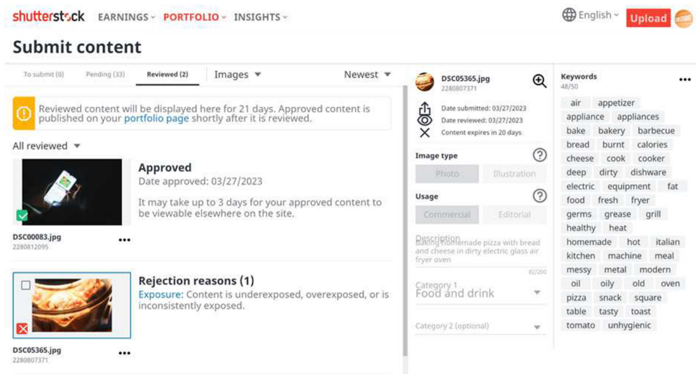

Figure 3: Assessment of uploaded photographs from Shutterstock

3.2. Model Training and Selection

The M-Stock system utilizes Convolutional Neural Networks (CNNs) [16] due to their strong performance in visual data analysis and spatial feature extraction. CNNs were chosen over alternative models, such as transformers, because of their efficiency in handling complex image data with lower computational requirements, making them suitable for real-time feedback in educational environments. The CNN architecture includes multiple convolutional layers, ReLU activations, batch normalization, max-pooling layers, and fully connected layers. A final softmax classifier predicts image categories (e.g., Excellence, Good, Bad, Accepted, Rejected), as shown in Figure 4.

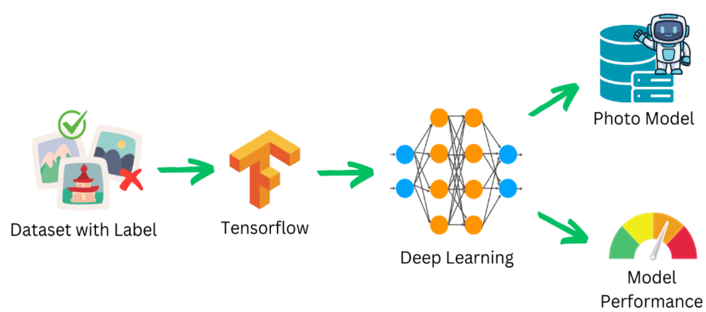

Figure 4: Model training and selection diagram

The training process involved the following steps:

Image Preprocessing: All images were resized to 800 x 800 pixels to maintain consistency, ensuring that the model could effectively extract meaningful features across various image types. *Model Optimization:* The Adam optimizer [17] was used to minimize the loss function (sparse categorical cross-entropy), ensuring that the model converged efficiently [18]. During training, performance metrics such as accuracy, prediction speed, and training time were monitored to evaluate the model's effectiveness. During training, performance metrics such as accuracy, prediction speed, and training time were monitored to evaluate the model's effectiveness. The training process was executed using Python 3.11.0 [19], Keras [20], and TensorFlow 2.13 [21]. *Evaluation Metrics:* In addition to accuracy, other metrics such as precision, recall, and F1 score were used to comprehensively assess the model's effectiveness. These metrics are essential in ensuring that the system's predictions are reliable across different types of assignments and quality levels.

3.3. Web Implementation and User Interface

The third component of the M-Stock system is the development of a user-friendly web application that allows students and instructors to interact with the model in real time. The web application was developed using Streamlit 1.31.0 [22], a Python-based framework that simplifies the deployment of machine learning models in web environments. Users initiate the M-Stock system by accessing the website via the URL http://datascience.mfu.ac.th/mstock/. The application's user interface is designed to be intuitive, enabling students to submit their photographs for assessment quickly and easily, Figure 5 illustrates the user interface of the homepage.

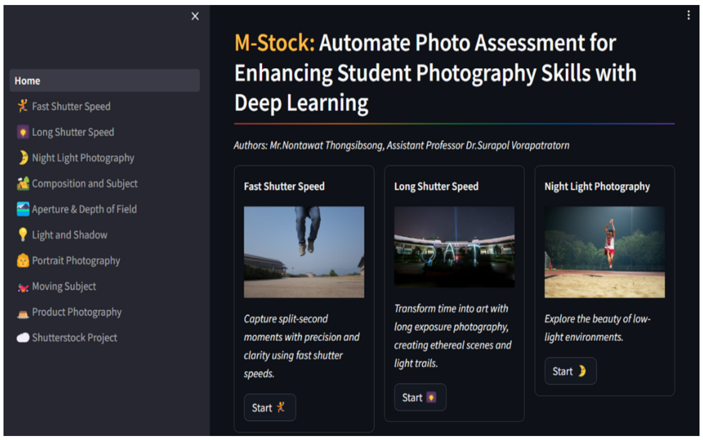

Figure 5: The user interface of the homepage

The submission process involves the following steps: Image Upload: Students select the assignment type and upload their photographs via the web interface. The system supports image formats such as JPG and PNG, with a maximum file size of 200 MB per image. Image Preprocessing and Classification: Once an image is uploaded, the system preprocesses it by resizing and standardizing the input. The pre-trained CNN model then classifies the image, providing a prediction and confidence score for each category (e.g., Excellence, Good, Bad). Feedback Delivery: The classification results are displayed immediately, allowing students to receive prompt feedback on their work. This feedback can help students identify areas for improvement and refine their photography skills iteratively. The M-Stock web application includes 11 pages: one homepage and ten assignment pages. Each assignment page corresponds to a specific photography lesson, where students can view sample photographs and submit their own work for evaluation. The web application's architecture ensures that it can scale to accommodate larger datasets and more complex assignments as the photography curriculum evolves. The Assignments page's user interface is depicted in Figure 6.

Overall, the M-Stock system combines the power of CNN-based image classification with a tailored feedback mechanism to support student learning in photography. Through its combination of technical rigor, creative assessment, and real-time feedback, M-Stock offers a novel solution for enhancing photography education in university settings. This method ensures that students receive immediate, meaningful feedback on their work, fostering continuous improvement and skill development in both technical and artistic aspects of photography.

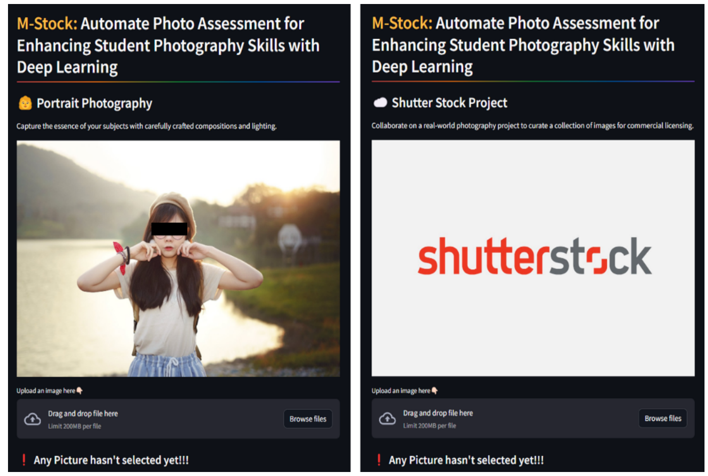

Figure 6: The user interface of the assignment page

4. Evaluation And Results

The M-Stock system was evaluated based on its classification accuracy, prediction speed, and training time, along with additional metrics such as precision, recall, and F1 score to provide a comprehensive assessment. Furthermore, a pilot study was conducted with students to gather qualitative feedback on their learning experience with M-Stock. This section presents the experimental setup, results, and analysis, demonstrating M-Stock's efficacy in supporting photography education.

4.1. Experimental Setup

The M-Stock system was tested in a virtualized server environment using VMware ESXi [23], running Windows Server 2016 [24] with an 8-core CPU (2.10 GHz) and 16 GB of RAM. The dataset for training and testing included 4,616 student images from various photography assignments and 244 Shutterstock images. and the necessary software tools, including Python 3.11.0, Keras with TensorFlow 2.13.0, and Streamlit 1.31.0 for web deployment. The dataset was divided into an 80% training set and a 20% test set, ensuring a robust model capable of handling diverse image categories. The system's scalability and performance were also evaluated under different image resolutions and batch sizes. Additionally, a small-scale pilot study with 30 students was conducted to assess the impact of M-Stock's feedback on learning outcomes.

4.2. Model Performance

The M-Stock system was evaluated for its ability to accurately classify student photography submissions across various assignment types, including technical and creative tasks. To assess the model's effectiveness, we measured several key metrics: accuracy, precision, recall, and F1 score for each assignment type. These metrics provide a comprehensive view of the model's classification performance, highlighting its strengths in technical precision and adaptability to different photography genres. The results are presented in Table 1

Table 1: Performance of each photo model for the M-Stock System

Photo Model	Accuracy (%)	Precision	Recall	F1-Score	Training Time (min.)	Prediction Speed (ms)
Fast Shutter Speed	96.72	0.96	0.95	0.96	28.7	47.1
Long Shutter Speed	96.93	0.97	0.96	0.97	39.1	46.7
Night Light Photography	98.36	0.98	0.97	0.98	45.5	50.6
Composition and Subject	98.77	0.99	0.98	0.99	70.4	43.7
Aperture, Depth of Field	95.33	0.95	0.94	0.95	14.7	47.9
Light and Shadow	95.47	0.94	0.95	0.94	35.4	45.3
Portrait Photography	99.53	0.99	0.99	0.99	119.2	43.4
Moving Subject	97.54	0.97	0.96	0.96	16.6	46.5
Product Photography	96.73	0.96	0.95	0.96	14.3	44.7
Shutterstock Project	96.39	0.95	0.94	0.94	20.3	45.2
Total (Average)	**97.18**	**0.97**	**0.96**	**0.96**	**40.4**	**46.1**

Accuracy: The system achieved an overall accuracy of 97.18%, with individual assignment accuracies ranging from 95.33% for Aperture, Depth of Field to 99.53% for Portrait Photography. The high accuracy demonstrates M-Stock's ability to consistently classify images across diverse photography tasks, from technical skills (e.g., Long Shutter Speed) to more composition-focused assignments (e.g., Composition and Subject). High accuracy in these varied tasks indicates that M-Stock can generalize well across different photographic techniques and styles, making it adaptable to a comprehensive photography curriculum. In addition to accuracy, we calculated precision, recall, and F1 scores for each assignment type to gain insights into M-Stock's classification reliability: *Precision:* High precision values (average of 0.97) indicate that M-Stock has a low rate of false positives, meaning it rarely misclassifies lower-quality images as higher quality. This is crucial in an educational context where students need accurate feedback to understand areas requiring improvement. *Recall:* The average recall of 0.96 shows M-Stock's effectiveness in identifying all images that meet specific quality criteria. High recall is especially important for technical assignments, as it ensures that the system accurately identifies images with correct exposure, composition, and other technical parameters. *F1 Score:* With an average F1 score of 0.97, M-Stock demonstrates a balanced performance in both identifying correct classifications and avoiding misclassifications. This score, the harmonic mean of precision and recall, confirms that the system provides reliable feedback, balancing sensitivity and specificity. The average prediction speed of 46.1 milliseconds per image shows that M-Stock provides rapid feedback, which is essential in real-time educational environments where students submit images and expect prompt responses. This quick feedback loop enables students to immediately identify mistakes and make improvements, reinforcing the learning process. The system's training time varies based on assignment type, with more complex tasks such as Portrait Photography taking longer (119.2 minutes) due to the intricate analysis required.

M-Stock's classification performance metrics demonstrate its effectiveness in providing real-time feedback across a wide range of photographic techniques. By maintaining high accuracy, precision, and recall across both technical and creative assignments, the system supports educators in delivering consistent, objective feedback to students. This capability is particularly beneficial in large classes, where individualized feedback is challenging to provide manually. With M-Stock, students can receive accurate, actionable feedback that promotes self-directed learning and skill refinement. Overall, M-Stock's classification performance confirms its suitability as a comprehensive educational tool, capable of assessing diverse photography tasks with high accuracy and efficiency. Future enhancements may involve refining these classification models further to increase precision and recall in more subjective creative categories, aligning with the evolving needs of photography education.

4.3. Scalability and Runtime Performance

The scalability of the M-Stock system was tested across various image resolutions and batch sizes to evaluate its capacity for handling large volumes of submissions in real-time classroom settings. Scalability is essential in educational applications where a high number of images may be submitted simultaneously, especially in large classes. The system was assessed under four different image resolutions—640x480 (low), 1280x720 (HD), 1920x1080 (Full HD), and 3840x2160 (4K)—to analyse the effect of image size on prediction speed and accuracy. For each image resolution, we measured the average prediction speed, batch processing time, and accuracy to determine the system's efficiency and robustness under increasing data sizes. Table 2 below illustrates these findings:

Table 2: Different Image Sizes Experiment Results

Image Size	Prediction Speed (ms)	Processing Time (Sec)	Acc. (%)
640 x 480	39.5	2.1	96.3
1280 x 720	46.1	2.5	97.2
1920 x 1080	52.4	3.2	97.8
3840 x 2160	74.6	5.4	98.1

These results show that the system maintains high accuracy across all resolutions, with a minimal decrease in prediction speed as image size increases. For low and HD resolutions, prediction times are under 50 milliseconds, allowing near-instantaneous feedback in real-time applications. Full HD and 4K images take slightly longer to process, but the prediction speeds are still well within acceptable limits for classroom use, ensuring efficient operation even for high-quality images. The accuracy remains high across resolutions, demonstrating that M-Stock's performance does not degrade with larger image sizes. The system's batch processing ability was evaluated to simulate high-demand situations where multiple students submit images simultaneously. We processed batches of 50 images at different resolutions, recording the total processing time required. M-Stock handled batch submissions with only a slight increase in processing time for higher-resolution images, completing a 50-image batch in approximately 2.1 seconds at low resolution and 5.4 seconds at 4K. This capability indicates that M-Stock is well-suited to handle real-time feedback needs in large classes, where simultaneous submissions are common.

4.4. User Satisfaction

To assess user satisfaction with the M-Stock system, a survey was conducted among 30 students and 5 instructors during the pilot study. The survey evaluated four key dimensions: ease of use, feedback clarity, perceived usefulness, and overall experience. Participants rated each dimension on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). The results are summarized in Table 3.

Table 3: User Satisfaction Survey Results

Dimension	Average	SD	Agreement (Score ≥ 4)
Ease of Use	4.7	0.3	93%
Feedback Clarity	4.5	0.4	87%
Perceived Usefulness	4.6	0.5	90%
Overall Experience	4.6	0.3	92%

The survey results indicate high levels of satisfaction across all dimensions. Students found the system's interface intuitive and straightforward, with an average score of 4.7 for ease of use. Feedback clarity received an average score of 4.5, reflecting the comprehensibility of the AI-generated evaluations. The system's ability to enhance photography skills was rated 4.6 on average, indicating its perceived effectiveness in promoting self-directed learning. Overall, users rated their experience with the system highly, with an average score of 4.6 and 92% agreement. Qualitative responses also highlighted specific benefits, such as the speed of feedback delivery and the ability to focus on iterative improvement. Some suggestions for enhancement included adding more nuanced assessments of creative aspects, such as artistic style and emotional impact. The results demonstrate that M-Stock effectively supports both teaching and learning objectives, providing a user-friendly, impactful solution for photography education.

Students also reported appreciating the quick turnaround time of feedback, which allowed them to adjust in near real-time. These findings suggest that M-Stock's formative feedback supports continuous learning, enhancing students' technical and creative skills. The results show that the M-Stock system performed exceptionally well in both educational and commercial contexts. The high accuracy rates across all categories demonstrate that the CNN model is capable of handling diverse photographic styles and quality levels. The relatively low prediction speed of 46.1 milliseconds per image allows the system to provide immediate feedback, which is crucial for enhancing the learning experience in photography courses. The results of the photo quality assessments for each assignment are displayed in Figure 7.

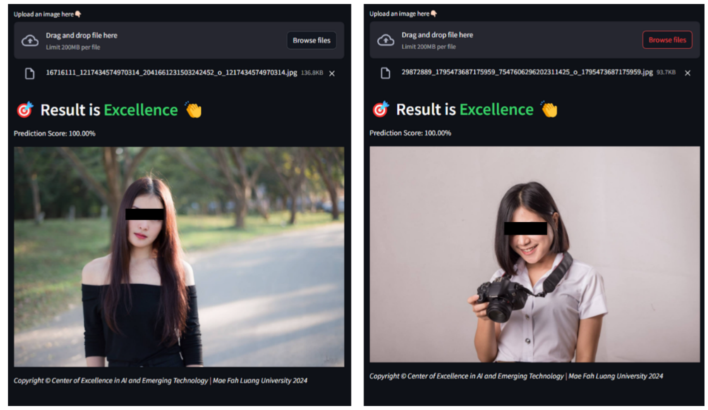

Figure 7: The photo quality assessment in 'Excellence' result

The portrait photography model, which achieved the highest accuracy (99.53%), required the longest training time (119.2 minutes). This indicates that more complex assignments, which involve intricate features such as lighting and composition in portrait photography, require more computational resources to train effectively. However, once trained, the model can classify images quickly and accurately. In contrast, simpler assignments, such as Product Photography and Moving Subject, required significantly less training time but still achieved high accuracy, indicating that the model can generalize well across different photography styles. The Shutterstock project data also yielded strong results, with an accuracy of 96.39%. This indicates that the system can meet industry standards for evaluating commercial photography, providing feedback that aligns with professional evaluation criteria. The results of the photo quality assessments for the Shutterstock project are presented in Figure 8.

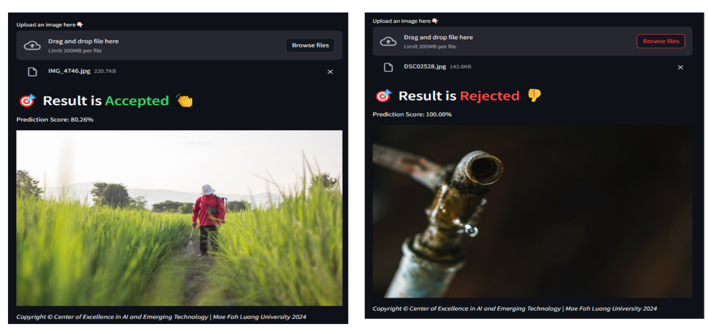

Figure 8: The photo quality assessment in 'Excellence' result

The evaluation results indicate that M-Stock performs reliably across technical metrics, while the pilot study confirms its positive impact on student learning. The high accuracy, coupled with quick feedback delivery, underscores M-Stock's suitability for real-time educational applications. The scalability tests further demonstrate that the system is robust enough to handle diverse classroom environments with high submission volumes. Overall, M-Stock provides a comprehensive assessment experience for photography students, offering both technical precision and creative guidance. Future work may explore expanding the system's feedback capabilities to include more nuanced assessments of creative elements, potentially incorporating reinforcement learning techniques to adapt feedback based on individual student progress.

5. Conclusion

The M-Stock system represents a significant advancement in photography education by leveraging the power of artificial intelligence to provide automated, real-time feedback on both technical and creative aspects of student submissions. By utilizing Convolutional Neural Networks (CNNs), the system achieved high accuracy (97.18%) and rapid prediction speeds (46.1 milliseconds per image), making it a reliable and scalable solution for dynamic classroom environments. Through a combination of quantitative evaluations and qualitative user feedback, the study demonstrated that M-Stock effectively enhances student learning experiences. Students reported improvements in their technical skills, self-directed learning, and overall engagement, while instructors appreciated the system's ability to maintain consistent evaluation standards across large class sizes. The system's ease of use and comprehensive feedback mechanisms make it a valuable tool for fostering continuous learning and skill development in photography courses.

Despite these achievements, challenges remain in assessing highly subjective creative elements, such as artistic style and emotional impact. Future iterations of M-Stock should incorporate advanced techniques, such as reinforcement learning or generative models, to provide deeper insights into these aspects. Additionally, expanding the platform to support other creative disciplines, such as graphic design and visual arts, could broaden its applicability and impact. In conclusion, M-Stock exemplifies how AI can transform education by addressing key limitations of traditional assessment methods. By combining technical rigor with creative evaluation, the system not only meets the evolving needs of photography education but also sets the stage for broader applications of AI in creative and technical learning environments.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The web server used in this study M-Stock, AI-Based Photography Assessment System using Convolutional Neural Networks was supported by the Center of Excellence in Artificial Intelligence and Emerging Technologies, School of Applied Digital Technology, Mae Fah Luang University.

References

[1] T. Crawford, C. DeLaney, Starting your career as a freelance photographer, Simon and Schuster, 2017.

[2] H. Alkan, O. Topuz, B. İnce, Ş. Kapıkıran, "The effects of basic photography education on quality of life, self-esteem, life satisfaction and moods in children with diplegic cerebral palsy: A randomized controlled study," Physical & Occupational Therapy in Pediatrics, **42**(1), 1–11, 2021.

[3] C.-M. Chen, "Personalized E-learning system with self-regulated learning assisted mechanisms for promoting learning performance," Expert Systems with Applications, **36**(5), 8816–8829, 2009.

[4] T. Zoltie, T. Shemwood, "Instructional design of a clinical photography course for undergraduate dental students," Journal of Visual Communication in Medicine, **42**(2), 47–51, 2019.

[5] M. Munakata, A. Vaidya, "Encouraging creativity in mathematics and science through photography," Teaching Mathematics and Its Applications: International Journal of the IMA, **31**(3), 121–132, 2012.

[6] Z. Yanhua, L. Jiaogang, "Research on the expression of photography language under digital technology," in 2020 5th International Conference on Electromechanical Control Technology and Transportation (ICECTT), IEEE: 228–230, 2020.

[7] M.C. Sahin, "Instructional design principles for 21st century learning skills," Procedia-Social and Behavioral Sciences, **1**(1), 1464–1468, 2009.

[8] D. Freer, Microstock photography: how to make money from your digital images, Routledge, 2008.

[9] K. Kemavuthanon, "Integrated E-project collaborative management system: Empirical study for problem-based learning project," in 2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE), IEEE: 1–5, 2017.

[10] L. Jacobs, Professional commercial photography: techniques and images from master digital photographers, Amherst Media, 2010.

[11] N. Thongsibsong, K. Kemavuthanon, S. Vorapatratorn, "Enhancing Student Photography Skills with Web-based Photo Stock and Learning System," in 2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE), IEEE: 454–457, 2023.

[12] A. Krizhevsky, I. Sutskever, G.E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, **60**(6), 84–90, 2017.

[13] N. Thongsibsong, "Photography Supporting System for Distance Learning in University," in 2021 13th International Conference on Information Technology and Electrical Engineering (ICITEE), IEEE: 15–18, 2021.

[14] B. Sahu, "Digital Tools for Educational Enhancement," Digital Narratives in Education, 78, 2024.

[15] Q. Jadwan, "How to get Money From Shutterstock Contributor in Smartphone," Hikamatzu| Journal of Multidisciplinary, **1**(1), 7–12, 2024.

[16] T. Kattenborn, J. Leitloff, F. Schiefer, S. Hinz, "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing," ISPRS Journal of Photogrammetry and Remote Sensing, 173, 24–49, 2021.

[17] A. Keras, "Keras 3 API documentation," Metrics, Regression Metrics Https://Keras. Io/Api/Metrics/Regression_metrics, 2024.

[18] M. Yeung, E. Sala, C.-B. Schönlieb, L. Rundo, "Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation," Computerized Medical Imaging and Graphics, 95, 102026, 2022.

[19] G. VanRossum, F.L. Drake, The python language reference, Python Software Foundation Amsterdam, The Netherlands, 2010.

[20] E. Dumić, "Learning neural network design with TensorFlow and Keras," in ICERI2024 Proceedings, IATED: 10689–10696, 2024.

[21] T. Developers, "TensorFlow," Zenodo, 2022.

[22] A. Streamlit, "Streamlit: A faster way to build and share data apps," Faster Way to Build and Share Data Apps, 2024.

[23] L. Patrão, VMware and vSphere Overview, Springer: 9–18, 2024.

[24] J. Krause, Mastering Windows Server 2016, Packt Publishing Ltd, 2016.

Advances in Science, Technology and Engineering Systems Journal Vol. 10, No.3, 29-44 (2025)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ
ISSN: 2415-6698

Explainable AI and Active Learning for Photovoltaic System Fault Detection: A Bibliometric Study and Future Directions

François Dieudonné Mengue[1,2], Verlaine Rostand Nwokam[3], Alain Soup Tewa Kammogne[4*], René Yamapi[5], Moskolai Ngossaha Justin[3], Bowong Tsakou Samuel[3], Bernard Kamsu Fogue[6]

[1]Fundamental Physics Laboratory, Department of Physics, Faculty of Sciences, University of Douala, Box 24 157, Douala, Cameroon

[2]Energy and Environment Technologies Department, National Committee for Development of Technologies (CNDT), Ministry of Scientific Research and Innovation (MINRESI), P.O Box 1457 Yaoundé, Cameroon. Email: mengue.dieudonn@gmail.com

[3]Department of Mathematics and Computer Science. Faculty of Science, University of Douala, P.O. Box 24157 Douala, Cameroon; rostand.nwokam@myiuc.com, justin.moskolai-ngossaha@enit.fr, samuelbowong@yahoo.fr

[4]Laboratory of Condensed Matter, Electronics and Signal Processing (LAMACETS). Department of Physic, Faculty of Sciences, University of Dschang, P.O. Box 67, Dschang, Cameroon Email: kouaneteoua@yahoo.fr, alain.kammogne@univ-dschang.org

[5]Fundamental Physics Laboratory, Department of Physics, Faculty of Sciences, University of Douala, Box 24 157, Douala, Cameroon Email: ryamapi@yahoo.fr

[6]Laboratoire Génie Production (LGP), ENIT, 47 Avenue d' Azereix, 65000 Tarbes E-Mail: bernard.kamsu-foguem@enit.fr

A R T I C L E I N F O

Article history:
Received: 22 April, 2025
Revised: 26 May, 2025
Accepted: 27 May, 2025
Online: 15 June, 2025

Keywords:
photovoltaic
faults prognostic
deep learning
Explainable IA
Active learning
bibliometric

A B S T R A C T

Persistent anomalies in modern photovoltaic (PV) systems present a formidable challenge, impeding optimal power output and system resilience. Artificial Intelligence (AI) has surfaced as a game-changing solution, yet existing research has merely scratched the surface of solar panel prognosis, leaving a critical void in leveraging AI's explainable nature and active learning capabilities. This pioneering study investigates AI methods for detecting and classifying critical faults in PV systems, pushing the boundaries of innovative methodologies for fault identification. We acknowledge that the opacity of AI methods can hinder their adoption, particularly among practitioners, thus emphasizing Explainable AI (XAI) in an exhaustive bibliometric analysis. This study showcases authors who thoroughly detail their development processes and underscores the indispensable role of human/expert interaction in active learning for labeling the most informative data. Our findings unveil a glaring underutilization of XAI in the solar panel domain, with China at the forefront of this field. This leadership is likely attributed to the robust research focus in Chinese universities and China's position as the world's leading solar panel producer. We delve into the potential role of human/expert involvement in designing and deploying deep learning predictive applications, highlighting methods that harmoniously integrate practical knowledge from human end-users through active learning. Our methodology encompasses extensive data collection, bibliometric analysis of collaborations between entities, researchers, and nations, and an examination of the most prevalent persistent faults. We conclude by strongly advocating for future studies to address the underutilization of XAI and active learning in AI-based defect prediction. Bridging this gap is crucial for pinpointing the root causes of solar panel defects and enhancing prognosis, positioning this research as indispensable for both scientists and industry professionals at the forefront of PV technology.

*Corresponding Author: Alain Soup Tewa Kammogne, University of Dschang, Cameroon, kouaneteoua@yahoo.fr

1. Introduction

1.1. Photovoltaic System

Photovoltaic energy is derived from the direct conversion of sunlight into electricity through solar cells, also known as photovoltaic cells, which form the basic component of a photovoltaic production chain. The assembly of multiple photovoltaic cells forms a photovoltaic module, commonly referred to as a solar panel [1]. A photovoltaic power plant is a solar installation comprised of numerous interconnected modules to generate electricity on a large scale, often intended to supply an electrical grid [2].

Over the past few decades, photovoltaic electricity production has experienced extremely significant growth worldwide. The International Energy Agency's forecasts for the year 2025 indicate that solar energy will provide approximately 60% of the total demand for renewable energy, placing it at the forefront of the most promising sources [3]. In terms of production, China positions itself as the world's leading producer of solar panels and has strongly committed to increasing its share of non-fossil fuel energy usage to 25% by 2030, aiming to satisfy 27.5% of the world's energy demand with solar energy by 2050 [4]. It is worth noting, however, that African nations are not lagging behind in this crucial fight against global warming and the adoption of green energies. For instance, Ghana has embarked on a major transition towards renewable energy. The Ghanaian government has adopted a master plan aimed at increasing electricity production capacity from 42.5 MW in 2015 to 1363.63 MW by 2030, with photovoltaic solar sources representing more than 50% of the total capacity [5]. Solar highways also offer enormous opportunities for Bangladesh [6]. However, the use of photovoltaic (PV) modules for energy production is not a simple task, due to potential degradation that can lead to a decrease in the performance and efficiency of PV solar installations [7]. According to recent studies, the degradation rate varies between 0.6% and 0.7% per year [8]. Therefore, it is imperative to examine the defects that may compromise the proper functioning of photovoltaic solar installations.

1.2. Fault Detection

Solar panel defects refer to any abnormalities or problems affecting the structure, performance, or durability of photovoltaic solar panels. These imperfections may encompass manufacturing, transportation, installation, environmental, or electrical defects that can cause a decrease in energy efficiency, a reduction in useful lifespan, or safety risks. Failures in PV systems result from various factors such as shading, module contamination, inverter failure, and variations in manufacturing or aging of photovoltaic modules, among others [9-10]. These elements can lead to a performance decrease of over 2.5%, progressively contributing to the deterioration of the affected component's longevity [10].

Generally, these defects can be classified into three main categories: abrupt, intermittent, and incipient faults. Ground-fault or line-to-line short circuits, open-circuit defects, connector disconnections, and junction box anomalies are examples of abrupt faults that occur when part or all of the PV network is damaged. Partial shading or environmental fluctuations represent examples of temporary or intermittent fault sources that may fade or evolve over time (such as dust or contamination). The third type, called incipient faults, manifests as minor but potentially dangerous anomalies that evolve slowly over time, making their initial detection difficult. If not identified timely, their consequences can be hard to control. These faults can occur on both the DC side, i.e., PV module and DC-DC converter parts (such as yellowing and browning of solar cells, delamination, cracks, bubbles, and anti-reflective coating defects), as well as on the AC side, i.e., the inverter side (examples include bipolar transistor faults, overheating, aging, and degradation of connection cabling) [11-12].

In the literature, several approaches to detecting and diagnosing faults in solar systems are distinguished. These methods are primarily characterized by their ability to quickly detect malfunctions, to instantly analyze the necessary input data (whether climatic or electrical), and to be selective, i.e., able to distinguish the type of fault in question. Thus, they can be classified into two main families: Visual and thermal methods specific to detecting malfunctions such as discoloration, browning, soiling, hotspots, breakage, delamination, etc., and Electrical methods that focus on defective PV modules, strings, and matrices, including arc, ground, and diode faults, etc. [13-15]. However, it often happens that both approaches are combined to solve specific malfunctions, as is the case with Artificial Intelligence methods.

1.3. Contributions and Research Objectives of the Paper

Bibliometric research is a quantitative analysis method used to study scientific publications. It aims to measure and evaluate different aspects of scientific production, such as the number of publications, citations, journal impact, and collaboration networks between researchers and institutions. Bibliometrics also quantifies researchers' contributions, identifies research trends, tracks field evolution, and assesses research impact on the scientific community. It is widely employed in scientific policy decision-making, resource allocation, and researcher and institution performance evaluation. In recent years, numerous researchers and fields have shown interest in bibliometrics. References [16-18] respectively present the use of bibliometric techniques to examine basic research on blockchain technology in the energy sector, the state of research on using blockchain technology for environmental sustainability in the building sector, and recent progress in electrolyzer control technologies for hydrogen production. Other works like reference [19] analyze research trends in big data analysis of the Internet of Things (IOT) and fog computing in the health sector, and reference [20] performs a comprehensive analysis of research in Energy System Analysis (ESA) using statistical techniques. Reference [21] provides a detailed analysis of research on predictive maintenance 4.0 by artificial intelligence, while references [22-23] focus on emerging techniques and trends in equipment maintenance systems, and reference [24] presents the evolution of artificial intelligence. Reference [25] offers a literature review on machine learning in industrial applications.

In the field of solar energy, bibliometric analysis has been integrated to identify emerging research areas and provide an impartial overview of current research status. A key work to date is that of [26], where bibliometrics was used to analyze a large number of documents on artificial intelligence methods for detecting and diagnosing faults in photovoltaic systems published since 2023.

To our knowledge, no bibliometric study on the prognosis by artificial intelligence of critical anomalies for predictive maintenance of photovoltaic systems has been identified. This gap in the literature pertains to the targeted faults, which can lead to total or partial shutdown of electricity production in photovoltaic systems, remaining unexplored.

In this research, we undertake a bibliometric analysis of publications indexed in the Scopus database, focusing specifically on the prognosis of critical faults in photovoltaic systems through artificial intelligence methods. We pay particular attention not only to the explanatory and descriptive aspect of the design and application procedures of AI methods but also to highlighting studies on human involvement in the AI integration process. In photovoltaic engineering, fault diagnosis identifies and locates current anomalies through real-time or historical data analysis and inspection techniques, aiming for quick detection for immediate corrective measures. Prognosis predicts the evolution and severity of future problems using degradation models and historical data, enabling planning of preventive maintenance actions to extend solar system life and minimize production interruptions.

Several questions arise in this work, including: What are the current trends in AI involvement in diagnosing and prognosing faults in PV systems? What is the role of active learning (human/expert involvement) in machine learning implementation for this problem? What are the main challenges related to AI methods application in PV system lifespan prognosis?

Figure 1 presents the flowchart of actions for this work: (1) and (2): Data collection from the Scopus database using specific keywords related to the topic. (3): Selection of important documents based on quality criteria, aiming to create a database containing necessary and relevant information to determine hidden relationships automatically. (4) and (5): Processing, analysis, and visualization of selected documents using tools such as R Bibliometrix (Biblioshiny) [27] or VOSviewer [28]. Highlighted information includes main trends, collaboration networks, most cited documents, challenges in AI prognosis of defects, XAI trends, human involvement in AI implementation, publication trends over time, most productive authors, leading countries, and frequent keywords.

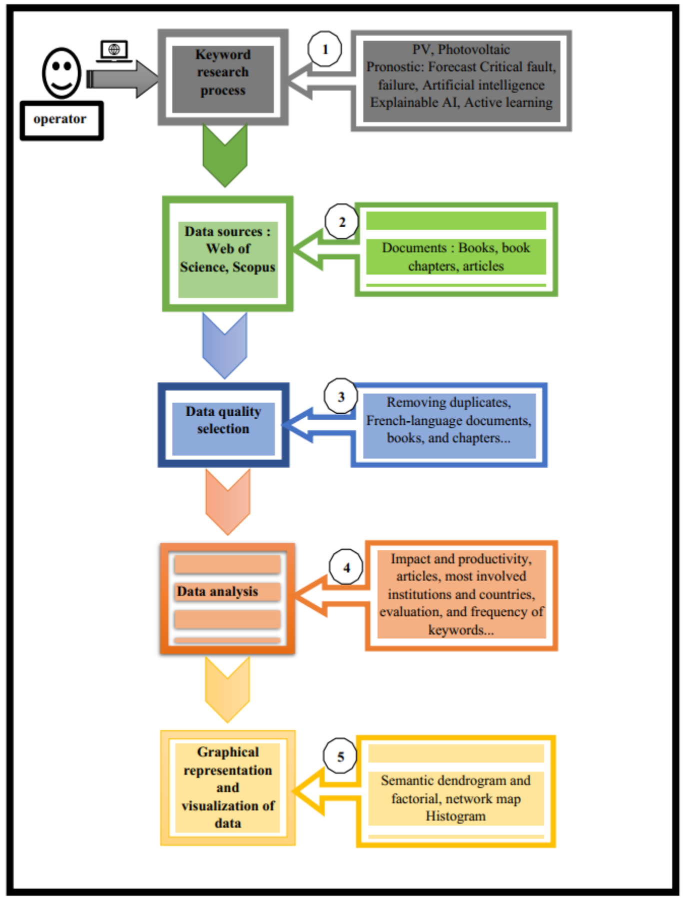

Figure 1: Methodological Framework for Bibliometric Analysis. Each step is color-coded and corresponds to a specific number

1.4. Explainable IA Learning and Active Learning

1.4.1. Explainable IA

Machine learning methods, particularly with the advent of neural networks (NN), are now widely used in engineering applications. This success has led to considerable adoption of machine learning (ML) in many scientific fields, including photovoltaic energy. However, most modern AI techniques suffer from the concept of the black box nature, which hinders their adoption by practitioners in many application areas. Explainability is a prerequisite to ensure the scientific value of the result. In this context, research directions such as explainable artificial intelligence (AI) [29], informed machine learning [30], or intelligible intelligence [31] have emerged.

The term XAI, for "eXplainable Artificial Intelligence," refers to a set of processes and methods aimed at making every result calculated by artificial intelligence understandable. It is a field of Machine Learning that seeks to precisely justify a given result by a model. Explainable AI (or XAI) is thus a research area in artificial intelligence that aims to create AI systems capable of explaining their decisions, actions, or results in a way that is understandable to humans. This approach distinguishes itself from many current AI systems, particularly those based on deep learning, which are often considered "black boxes" due to the difficulty in understanding their internal functioning [32-33].

Figure 2 above highlights the contrast between precision and explainability of the most commonly used models. Special attention is given to Artificial Neural Networks (ANNs), which constitute the majority of the most used AI methods in PV fault detection and prognosis. From this graph, it is clear that ANNs are very precise but have low explainability. The explainability of AI in general, or machine learning in particular, helps to answer questions such as: why did you make this decision? Why didn't you act differently? When will you succeed or fail? How do you correct the error? These questions should be addressed in the development and implementation of different AI methods.

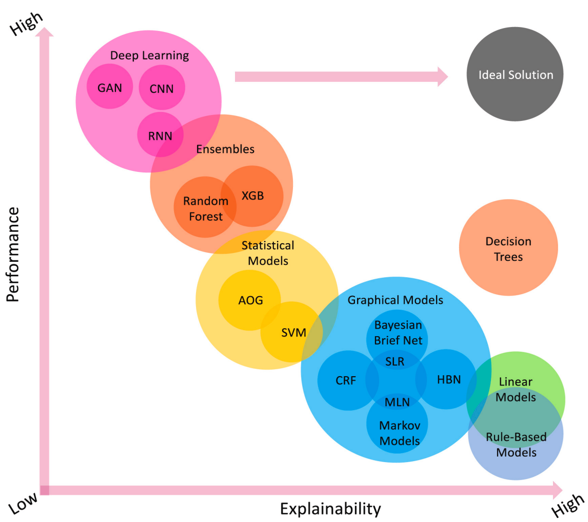

Figure 2: Comparison between Explainability and Accuracy for Different Machine Learning Approaches. **[34]**

To be more precise, terms and indicators that highlight this concept may include: AI interpretability, model transparency, explainability of algorithms, post-hoc analysis of decisions, methods for visualizing decision-making processes, detailed documentation of AI models, interpretation of AI results, identification of explanatory factors in AI predictions, etc.

1.4.2. Active Learning

Nowadays, discussing deep learning and artificial intelligence without addressing image analysis and interpretation, as well as their use for extracting essential information for purposes such as computer-aided detection, prognosis, treatment planning, intervention, and preventive maintenance of systems, is essential. However, the unique challenges posed by data analysis, particularly in the context of detecting and predicting defects in photovoltaic systems, suggest that it would be beneficial to maintain the presence of a human end-user in any system using deep learning, specifically, or artificial intelligence, in general [35].

Active learning is an approach to artificial intelligence that involves humans in the learning process to improve the efficiency and accuracy of the AI algorithm. In this approach, the algorithm selects the most informative data for which it needs labels, rather than relying on randomly labeled data [36]. Human involvement in the active learning process can take different forms, such as providing labels for data selected by the algorithm, correcting errors made by the algorithm, or providing feedback on the results of the algorithm. Active learning has many advantages over traditional machine learning, such as reducing the amount of labeled data required to train the AI algorithm and improving the accuracy of the AI algorithm [37-38].

Specifically, some terms and indicators that may highlight this method are: Interactive learning, Active data selection, Human feedback in the learning process, User query systems, Data labeling optimization, Data acquisition methods, Iterative improvement of the model through human interaction, Dynamic adaptation of the learning process to user needs.

2. Research Methodology

2.1. Bibliometric Analysis

Conducting a rigorous bibliometric analysis requires a structured approach. It begins with the clear formulation of research questions and the selection of suitable bibliometric methods. Data collection is then performed using reputable sources such as Web of Science, Scopus (as in this study), Google Scholar, or ScienceDirect, ensuring the reliability of the information. Once collected, the data is analyzed using advanced statistical techniques to uncover trends and correlations. The results are then carefully interpreted and visualized through specialized tools, enabling a comprehensive understanding of the research dynamics.

2.2. Data Collection from Scopus

For this study, we used Scopus as the primary source of bibliographic data [39-40]. This platform offers a wide range of scientific publications in various fields. Data were collected using relevant keywords for our research topic. We then applied data exploration techniques to identify the most relevant publications. The collected data includes information such as article titles, authors, affiliations, keywords, abstracts, and citation counts. Table 1 provides a recapitulation of the information gathered after the search.

2.3. Scanning and Keywords Search

To ensure robust and relevant conclusions, we developed a rigorous data collection methodology. On April 23, 2024, we began by exploring Scopus, a leading bibliometric database. Our search strategy was built around carefully selected keywords relevant to the field: "PV," "Photovoltaic," "Fault," "Failure," "Anomaly," "Detection," "Diagnosis," "Prognosis," "Classification," "Artificial Intelligence," "Machine Learning," "AI explainable," "XAI," "Active learning," and "Feedback experiment." These terms guided the identification of documents likely to inform our understanding of current issues and technological advances. Next, we applied strict selection criteria: only English-language documents published in internationally recognized journals and conferences were retained. This ensured the inclusion of high-quality, relevant materials.

The process yielded a curated corpus of 225 documents. Despite their volume, we noted a relative scarcity of studies addressing explainable AI and active learning in photovoltaic fault prediction. Nonetheless, the selected literature provides a solid foundation for analyzing trends and innovations in this area.

Overall, this rigorous and selective approach enhances the credibility of our findings and contributes meaningfully to academic discourse on contemporary energy challenges [41]

3. Analysis and Results

In bibliometric analysis, it is important to discuss the two main approaches for creating bibliometric maps: distance-based methods, such as multidimensional scaling [42], visualization of similarities (VOS) [43], force-directed placement [44-45], among others, and graph-based methods [46].

In the photovoltaic research field, a distance-based bibliometric mapping was performed using VOSviewer [47] and R-bibliometrix. This approach enables the visualization of relationships between different entities such as authors, laboratories, and countries, and measures their strength [48-49]. Using association strength as a similarity metric, co-occurrence maps of keywords were created to identify research subdomains in the photovoltaic field.

3.1. Main Information About the Collection

Table 1 provides an overview of the 225 publications collected from the Scopus search engine based on the selection criteria. It contains a total of 776 keywords and over 900 authors, with a detailed distribution between unique and multiple authors.

Table 1: Main Information about the Data Comprehensive Analysis Over 20 Years: 173 Sources, 225 Documents, highlighting 12.37% Annual Growth, Emphasizing Explainable AI and Active Learning for PV Fault Detection.

MAIN INFORMATION ABOUT DATA	
Timespan	2002:2024
Sources (Journals, Books, etc)	173
Documents	225
Annual Growth Rate %	12,37
Document Average Age	4,37
Average citations per doc	58,97
References	0
DOCUMENT CONTENTS	
Keywords Plus (ID)	2002
Author's Keywords (DE)	776
AUTHORS	
Authors	927
Authors of single-authored docs	36
AUTHORS COLLABORATION	
Single-authored docs	41
Co-Authors per Doc	4,58
International co-authorships %	0
DOCUMENT TYPES	
article	89
Book	43
book book	1
book chapter	10
conference paper	17
editorial	1
review	63
review book	1

Table 1 provides an overview of the main information from the research conducted over a period of more than 20 years, from 2002 to 2024. The research was based on 173 sources and analyzed 225 documents, indicating its comprehensive nature. The annual growth rate of 12.37% suggests evolving research in this field, particularly in the areas of explainable AI and active learning for PV fault detection. With an average of 58.97 citations per document, the work is increasingly recognized and utilized within the scientific community, bolstering its credibility. Involving 927 authors and encompassing various document types, including articles, books, book chapters, conference proceedings, editorials, and reviews, the research was conducted exhaustively to provide a comprehensive analysis of the subject. The significant number of authors highlights the importance of collaborative ideas in this field.

3.1.1. Annual Scientific Publication Trend

This section examines the annual trend of scientific publications in the field of interest. Analysis was conducted on the number of documents published each year between 2002 and 2024 using the Scopus database. Table 2 presents bibliometric data for articles published during this period, including measures such as the average number of citations per article (MeanTCperArt), the total number of articles (N), the average number of citations per year (MeanTCperYear), and the number of citable years (CitableYears).

Table 2: Annual Scientific Publication Trend: Average citations per article in solar panel defect research vary widely annually, ranging from 0.08 in 2024 to 395 in 2010, indicating evolving trends in explainable AI.

YEAR	***MEANTCPERA***	***N***	***MEANTCPERYEAR***	***CITABLEYE***

	RT			ARS
2002	22	1,00	0,96	23
2003	50	1,00	2,27	22
2004	38	1,00	1,81	21
2006	21	1,00	1,11	19
2007	135,5	2,00	7,53	18
2009	88,5	2,00	5,53	16
2010	395	4,00	26,33	15
2011	8,5	2,00	0,61	14
2012	270,17	6,00	20,78	13
2013	69	2,00	5,75	12
2014	90,33	3,00	8,21	11
2015	413,56	9,00	41,36	10
2016	63,75	4,00	7,08	9
2017	89,14	14,00	11,14	8
2018	56,44	9,00	8,06	7
2019	57,23	13,00	9,54	6
2020	42,38	26,00	8,48	5
2021	24,58	24,00	6,14	4
2022	16	39,00	5,33	3
2023	5,47	49,00	2,73	2
2024	0,08	13,00	0,08	1

It can be observed that the average number of citations per article (MeanTCperArt) varies significantly from year to year, with values ranging from 0.08 in 2024 to 395 in 2010. The total number of articles (N) has also increased over the years, from 1 in 2002 to 49 in 2023; this low value of MeanTCperArt in 2024 may be due to the fact that explainable AI is still under development in the field of prognosis and diagnosis of defects in solar panels. As indicated in the article by [50], there are still significant challenges to be addressed to make AI models more transparent and understandable to end-users.

The average number of citations per year (MeanTCperYear) has also varied over the years, with values ranging from 0.08 in 2024 to 41.36 in 2015. Regarding this downward trend in recent years, this may be partly due to the time it takes for articles to be read, cited, and integrated into the literature. As noted by Pan 2019 [49], there can be a lag of several years between the publication of an article and its actual impact on the field.

Finally, the decrease in the number of citable years (CitableYears) may be related to the rapid evolution of technology and methods used in the field of prognosis and diagnosis of defects in solar panels. Utama and colleagues highlighted this in their work in [50], where they explain that as more new AI methods are constantly being developed and implemented, this could make previous work obsolete more quickly.

3.1.2. Most Productive, Impact and Source Growth Dynamics

The data presented in Table 3 below show the most productive, impactful, and source journals in the field of solar energy, based on different bibliometric indicators such as the h-index, g-index, m-index, total number of citations (TC), number of publications (NP), and year of first publication (PY_start).

Table 3: *Most Productive, Impact, and Source Growth Dynamics*

Element	*h_ index*	*g_ index*	*m_ index*	*TC*	*NP*	*PY_ start*
Energies	7	13	1	187	16	2018
Renewable and Sustainable Energy Reviews	6	6	0,667	635	6	2016
IEEE Access	4	7	0,8	91	7	2020
Solar Energy	4	5	0,5	419	5	2017
Chemical Reviews	3	3	0,3	3493	3	2015
IEEE Transactions on Power Electronics	3	3	0,375	201	3	2017
Sustainability (Switzerland)	3	3	0,75	34	3	2021
Applied Energy	2	2	0,667	36	2	2022
Electronics (Switzerland)	2	3	0,333	24	3	2019
IEEE Power And Energy Society General Meeting	2	2	0,182	6	2	2014
International Journal of Heat and Mass Transfer	2	2	0,087	72	2	2002
International Journal of Hydrogen Energy	2	2	0,182	322	2	2014
International Transactions on Electrical Energy Systems	2	2	0,5	53	2	2021

It's apparent that the journal "ENERGIES" boasts the highest h-index (7), indicating it has published at least 7 articles each cited 7 times or more. However, "RENEWABLE AND SUSTAINABLE ENERGY REVIEWS" leads in total citations, with its articles garnering 635 out of 1759 total citations (36.1%).

Moreover, this journal has published the highest number of articles (6) since 2016, showcasing its influence. "IEEE ACCESS," although relatively recent, starting in 2020, has an h-index of 4 and 91 citations, representing 5.2% of the total, placing it third in impact and productivity. "SOLAR ENERGY" shares an h-index of 4 but boasts a higher total citation percentage of 23.8% (419 out of 1759), despite having fewer publications (5) since 2017. This indicates its articles have a significant impact in the solar energy field. Notably, some journals exhibit a high m-index, indicating uniform citations across their articles, like "CHEMICAL REVIEWS," with an m-index of 0.3. Additionally, newer journals such as "SUSTAINABILITY (SWITZERLAND)" and "APPLIED ENERGY" have also made an impact, starting in 2021 and 2022, respectively. "INTERNATIONAL JOURNAL OF HYDROGEN ENERGY" received 322 citations, representing 18.3% of total citations, making it a vital source in solar energy research, particularly regarding hydrogen energy.

This information aids in focusing on influential journals in solar energy and AI, guiding research toward high-quality, consistently cited articles. It underscores the growing interest in explainable AI and active learning in PV fault prognosis, despite their current underrepresentation. The involvement of prestigious journals like "ENERGY," "RENEWABLE AND SUSTAINABLE ENERGY REVIEWS," and "IEEE ACCESS" signals the necessity for further research in these areas. This suggests a potential pathway for enhancing solar panel defect detection using explainable AI and active learning, ultimately improving the reliability and performance of solar energy systems.

3.1.3. *Most Globally Cited Papers and References*

Table 4 presents an overview of the 20 most cited documents published in the Scopus database during the study period, with accompanying graphical representation in Figure 3. The columns include Total Citations, TC per Year, and Normalized TC, accounting for publication year and field.

Table 4: *Most Globally Cited Papers*

Paper	*Total Citations*	*TC per Year*	*Normalized TC*
[51]	2285	228,50	5,53
[52]	1256	96,62	4,65
[53]	1173	78,20	2,97
[54]	1096	109,60	2,65
[55]	421	52,63	4,72
[56]	291	48,50	5,08
[57]	255	17,00	0,65
[58]	232	29,00	2,60
[59]	231	46,20	5,45
[60]	224	17,23	0,83
[61]	206	18,73	2,28
[62]	190	23,75	2,13
[63]	189	27,00	2,98
[64]	187	23,38	2,10
[65]	157	8,72	1,16
[66]	152	21,71	2,39
[67]	152	9,50	1,72
[68]	150	2,35	2,35
[69]	147	29,40	3,47
[70]	138	69,00	25,23

From the table data, it can be seen that the most cited article is "LU L, 2015, Chem. Rev." with a total of 2285 citations, which also represents the highest number of citations per year (228.50) and normalized citations (5.53). This article has therefore had a significantly important impact in the studied research field. Other articles also have high citation numbers, such as "JØRGENSEN M, 2012, Adv. Mater." with 1256 citations and "TODESCHINI R, 2010, Mol. DESCRIPTORS CHEMOINFORMATICS" with 1173 citations. However, some articles have lower citation numbers, such as "SABNIS RW, 2010, Handb. Biol. Dye. Stain. Synth. Ind. Appl." with 255 citations and "MANNAN S, 2012, LEES' LOSS Prev. Process. Ind. Hazard Identif. Assess. Control. Fourth Ed." with 224 citations.

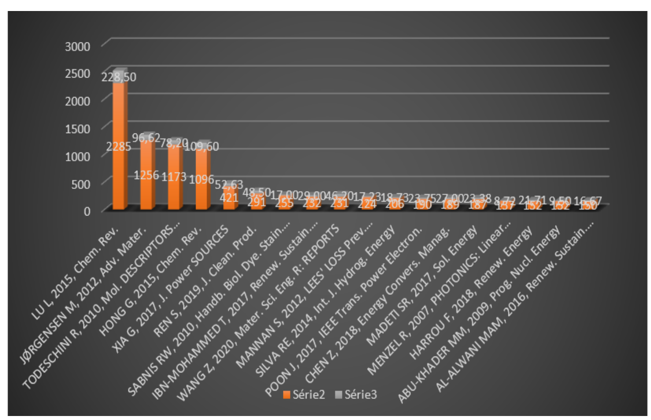

Figure 3: Most Global Cited Papers: Highlighting Significant Impact and Emerging Trends in Solar Panel Technology Through AI Applications, Especially in Explainable AI and Active Learning

By examining the normalized citation data, it can be seen that some articles have high numbers, such as "REN S, 2019, J. Clean. Prod." with 5.08 and "WANG Z, 2020, Mater. Sci. Eng. R: REPORTS" with 5.45, indicating that these articles have had a significant impact in their respective research fields, despite their lower total number of citations.

Figure 3 highlights the articles that have attracted the most attention in the domains of explainable AI and active learning. While some do not explicitly mention these fields, they contribute significantly by applying AI and machine learning to related areas such as solar technology, chemistry, materials science, and renewable energy. As noted in [71], explainable AI and active learning are rapidly growing research areas, marked by a steady rise in publications and citations across disciplines.

However, it is important to note that most of these articles focus on the use of AI and machine learning to improve the performance and efficiency of PV solar systems, rather than on the explainability and transparency of the models themselves. This highlights the need for more in-depth research in the field of explainable AI and active learning to develop more interpretable models and build trust in AI systems.

Furthermore, it should be noted that this analysis is based solely on data from the Scopus database, which may limit the generalizability of the results. Future studies could consider including other databases for a more comprehensive analysis of the literature on explainable AI and active learning applied to PV fault prognosis.

3.1.4. Most Productive and Highly Cited Authors

Table 5 shows the number of articles published by the most productive authors in the studied field, as well as their fractionalized number of articles and their number of citations.

Table 5: Most Productive and Highly Cited Authors: The table highlights the most productive and highly cited authors in the field, emphasizing their significant contributions to the research on solar panel fault diagnosis and prognosis, particularly through the use of AI methods.

Authors	*Articles*	*Articles Fractionalized*
SUN Y	5	1,15
ZHANG Y	5	1,03
HARROU F	4	0,90
LIU Z	4	0,46
TAGHEZOUIT B	4	0,90
WANG C	4	0,10
WANG H	4	0,41
WANG Z	4	0,69
YU H	4	0,10
BANSAL RC	3	1,37
BLAABJERG F	3	0,73
LIU C	3	0,29
LIU J	3	0,45

The table shows that SUN Y is the most productive author in this research field, with 5 published articles. However, when considering fractionalized articles—which account for each author's specific contribution to a publication—SUN Y attains a higher score (1.15) than other authors with the same number of publications. This indicates that SUN Y's contributions are relatively more substantial. Authors HARROU F, LIU Z, TAGHEZOUIT B, and WANG Z each published 4 articles, yet their fractionalized scores vary between 0.46 and 0.90, reflecting differences in their levels of involvement. Notably, BANSAL RC, despite having only 3 publications, holds the highest fractionalized article score of 1.37 among all authors, emphasizing the significant impact of his contributions to the field.

The relatively low number of articles and authors may be attributed to a lack of explicit focus on the explainability of artificial intelligence in the development process of AI methods for PV fault diagnosis and prognosis.

Regarding co-authorship, Table 6 presents data concerning the collaboration network between authors, where each line represents a node in this network. The "Cluster" field indicates that all nodes belong to the same cluster (Cluster 1).

Table 6: *Most Cited Authors*

Node	*Cluster*	*Betweenness*	*Closeness*	*PageRank*
zhang y	1	0	0,017857143	0,007070492
liu z	1	33	0,027777778	0,028781946
wang c	1	6,116666667	0,03125	0,036368497
wang h	1	6,116666667	0,03125	0,036368497
wang z	1	0	0,019230769	0,006478859
yu h	1	6,116666667	0,03125	0,036368497
blaabjerg f	1	0	0,017241379	0,005235847
liu c	1	42	0,026315789	0,026182215
liu p	1	21,78333333	0,032258065	0,039046276
liu y	1	30,5	0,027027027	0,02590086

The table presents the centrality measures of 10 nodes in a network, grouped into a single cluster. The centrality measures include betweenness centrality, closeness centrality, and PageRank. It can be observed that the node "liu z" has the highest betweenness centrality value (33), indicating that it is most often on the shortest paths between other nodes in the network. The nodes "wang c", "wang h", "yu h", and "blaabjerg f" all have low betweenness centrality values (less than 7), suggesting that they are less frequently on the shortest paths between other nodes. In terms of closeness centrality, the nodes "wang c", "wang h", and "yu h" have the highest value (0.03125), which means that they are closest to all other nodes in the network. Finally, the node "liu p" has the highest PageRank value (0.039046276), indicating that it is the most important in terms of incoming and outgoing connections with other important nodes in the network.

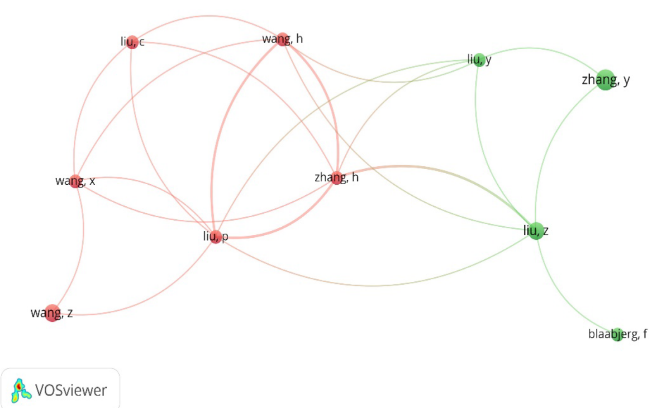

Figure 4: Co-author Network: Illustrates limited collaboration among authors in the specialized field, influenced by technical complexity and interdisciplinary communication challenges

Indeed, one can observe a limited number of active authors and less collaboration in the studied field, as depicted in the figure, which could be explained by several factors. Firstly, the domain's specificity and the specialized technical skills required to integrate approaches such as explainable AI and active learning may restrict the number of active authors. Secondly, the complexity of these methods and the communication challenges between researchers from different fields may diminish collaboration opportunities. Finally, in a developing field, competition to publish the first significant results may also hinder collaboration among researchers.

3.2. Citation Analysis

3.2.1. Co-Occurrence

The aim of this co-occurrence analysis is to illuminate potential relationships between two elements present in bibliographic works that appear together [72-73]. Utilizing certain standardized and automated methodologies [74-75], data can be extracted and visually represented using tools such as VOSviewer, R Bibliometrix, as demonstrated in this study for widespread application in conducting co-occurrence analyses of keywords across various areas of expertise.

a. Co-occurrence of all keywords.

Table 7: Most Cited Keywords

Words	*Occurrences*
fault detection	32
solar power generation	30
solar energy	22
photovoltaic cells	14
photovoltaics	13
electric inverters	12
photovoltaic systems	12
failure analysis	11
solar panels	11
deep learning	10
power quality	10
wind power	10
fault tolerance	9
learning systems	9

Table 7 presents the main keywords along with their frequency of occurrence. It is evident that specific learning techniques are currently utilized as scientific methodologies, producing favorable results in defect detection within photovoltaic systems. However, the explainability of AI is notably underrepresented, as reflected by its absence among the most frequently cited keywords. To deepen understanding of key terms, their clusters, associations, and temporal relevance, two bibliometric maps were generated. For clarity and focus, only keywords appearing at least five times were included in these maps.

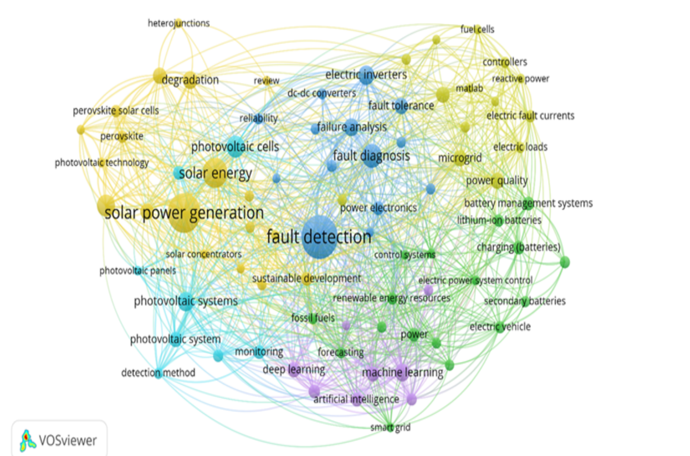

Figure 5: Visualization Map of Co-occurring Keywords Identifying 6 Main Clusters: Displays relationships between co-occurring keywords, highlighting thematic groupings and their associations within the research domain

The first map presents six keyword clusters, reflecting associations among the articles. Keyword co-occurrence—indicating simultaneous appearance—is shaped by proximity and similarity. Figure 5 visualizes these clusters, highlighting groups of keywords that frequently appear together and are organized by associated methods. Table 8 provides a summary of these co-occurring keyword clusters.

Table 8: Clusters of Co-occurring Keywords showcase thematic groupings

N cluster	Node
Cluster1	Carbon, degradation, heterojunction, life cycle, perovskite, perovskite solar cells, photovoltaic modules, photovoltaic technology, photovoltaics, reneweble energies, renewable energy, renew, solar celles, solar concentrators, solar energy, solar panels, solar power generation, sustainable developmer
Cluster2	Batteremy management system, charging (batteries), control systems, diital storage, electric power system control, electric power transmission, electric vehicle, electric vehicles, energy storage, forecasting, fossils fuels, lithium-ion batteries, optimization, power, reneweble energy resource, secondary batteries, smart grid.
Cluster3	dc-dc converters, electric drives, electric inverters, failure analysis, fault detection, fault diagnosis, fault tolerance, faults diagnosis, neural networks, power converters, power electronics, reability, timing circuits, topology.
Cluster4	Controllers, electric fault currents, electric loads, energy conversion, fault, fuel cells, MATLAB, microgrid, microgrids, performance, power quality, reactive power, sliding mode control, wind power.
Cluster5	Artificial intelligence, condition monitoring, deep learning, electric power system protection, fault detection, learning system, machine learning.
Cluster6	Detection method, efficiency, monitoring, photovoltaic cells, photovoltaic panels, photovoltaic system, photovoltaic systems, solar photovoltaic system.

Cluster 1 emphasizes carbon degradation, perovskite solar cells, photovoltaic technology, and renewable energy, highlighting sustainable development and the use of solar energy for electricity generation. Cluster 2 centers on battery management, electrical control systems, energy transmission, electric vehicles, and energy storage, addressing issues related to fossil fuels, lithium-ion batteries, and smart electrical grids. Cluster 3 focuses on electric power converters, electric drives, neural networks, and circuit reliability, dealing with fault detection, fault tolerance, and power electronics system topologies. Cluster 4 covers controllers, electrical fault currents, energy conversion, fuel cells, microgrids, as well as power quality, reactive power, and sliding mode control. Cluster 5 revolves around artificial intelligence, deep learning, state monitoring, and machine learning, targeting power supply protection and fault and failure detection. Finally, Cluster 6 concentrates on detection methods, efficiency, monitoring, photovoltaic cells, panels, and systems, specifically addressing fault detection and diagnosis in photovoltaic solar energy. The last two clusters primarily focus on PV fault diagnosis and related methodologies.

Moreover, human involvement in AI and the explainability of AI are less evident in several clusters outlined in the table. However, Cluster 5 focuses on machine learning, deep learning, and artificial intelligence, while also considering state monitoring and the protection of electrical power systems, implying human intervention to ensure system safety and reliability. Similarly, Cluster 6 addresses the detection and diagnosis of faults in photovoltaic systems, necessitating human interpretation of results and informed decision-making.

b. Co-occurrence of keywords by authors

In this map depicted in Figure 6, associations of keywords by author are linked to each other through three clusters. Similar to the previous case, the degree of co-occurrence of keywords, indicating their simultaneous appearance, is influenced by their proximity and similarity. Table 9 provides an overview of the clusters associated with the map.

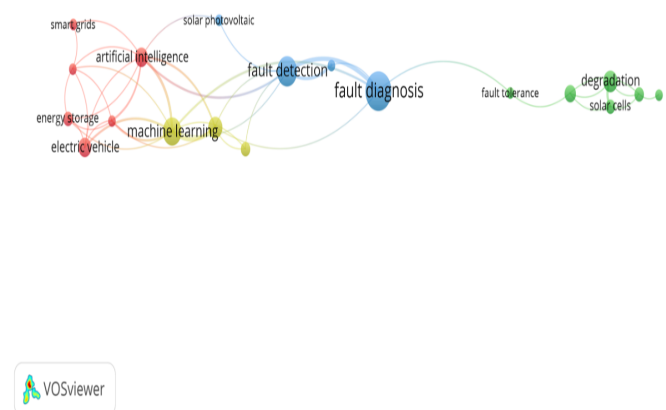

Figure 6: Visualization Map of Co-occurring Authors Keywords

Table 9: Clusters of Authors' Co-occurring Keywords illustrate diverse research focuses in photovoltaic systems, encompassing AI applications, system reliability, fault detection, and optimization for sustainable energy solutions.

N cluster	Node
Cluster1	Artificial intelligence, electric vehicle, energy storage, power electronics, prediction, smart grids
Cluster2	Degradation, fault tolerance, photovoltaics, reliability, solar cells, solar energy,
Cluster3	Fault detection, fault diagnosis, fault location, solar photovoltaic
Cluster4	Deep learning, machine learning, photovoltaic systems

Cluster 1 focuses on the applications of artificial intelligence, power electronics, and energy storage systems in smart electrical grids and electric vehicles. The authors of this cluster may be experts in areas such as machine learning, signal processing, and electrical system design. Research trends in this cluster may include the development of new machine learning algorithms for energy demand prediction, electric vehicle charging optimization, and energy management in smart grids or microgrids.

Cluster 2 focuses on the reliability and sustainability of photovoltaic systems, with a particular emphasis on degradation, fault tolerance, and reliability assessment. The authors of this cluster have expertise in materials, photovoltaic system design, and degradation modeling. Research trends in this cluster may include the development of new materials and system designs to improve the lifespan of photovoltaic systems, as well as the use of machine learning techniques to predict degradation and assess reliability.

Cluster 3 is centered on the detection, diagnosis, and localization of faults in photovoltaic systems. The authors of this cluster may be experts in signal processing, photovoltaic system design, and fault modeling. Research trends in this cluster include the development of new intelligent algorithms for fault detection and localization, as well as the use of machine learning techniques for fault diagnosis and classification.

Cluster 4 focuses on the applications of deep learning and machine learning in photovoltaic systems, with a particular emphasis on energy production prediction and system optimization. The authors of this cluster have expertise in machine learning, signal processing, and photovoltaic system design. Research trends in this cluster are focused on the development of new deep learning algorithms for energy production prediction and system optimization, as well as the use of machine learning techniques for fault detection and diagnosis.

Overall, the results of the cluster analysis suggest that research in the field of photovoltaic systems is diverse and rapidly evolving. Researchers are focusing on a wide range of topics, from the reliability and sustainability of systems to the optimization and prediction of energy production. The increasing use of machine learning and deep learning in photovoltaic research is also evident, with many authors exploring the applications of these techniques in fault detection. Furthermore, these results highlight the importance of collaboration and interdisciplinarity in photovoltaic research. Researchers need to work together to combine their expertise in materials, system design, signal processing, and machine learning to address the complex challenges facing the photovoltaic industry. Additionally, the research trends in the clusters suggest that the use of machine learning and deep learning in fault detection and diagnosis remains a promising avenue for improving the reliability and sustainability of photovoltaic systems.

3.2.2. Thematical Map and Thematic Evolution

To identify key research themes, keyword co-occurrence analysis is used to trace thematic evolution [76]. Each cluster reflects a conceptual theme, with the research period considered. Callon centrality measures inter-cluster connections, while Callon density assesses their internal strength and influence over time [77]. The size of each sphere in the diagram corresponds to the frequency of publications on that theme. The resulting strategic diagram places these sub-clusters in a two-dimensional space, offering a global view of the field's thematic evolution [78].

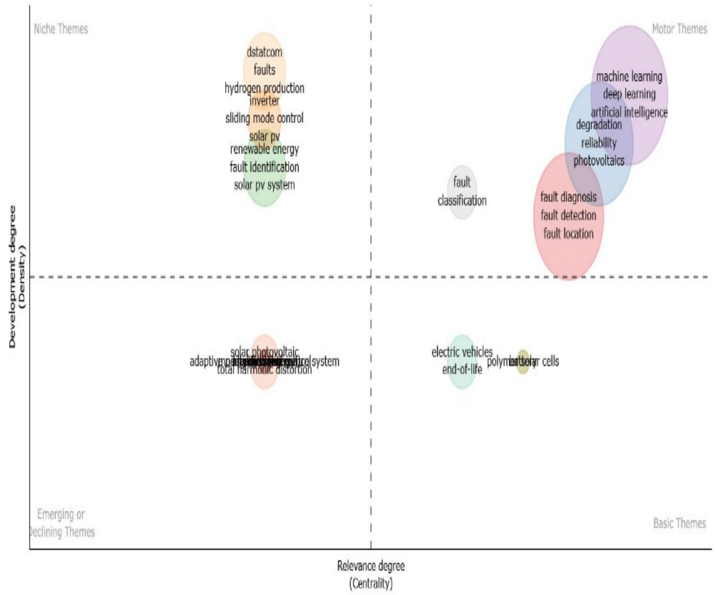

Figure 7: Strategic Map of Author Keywords illustrates thematic trends in solar panel defect diagnosis and prognosis using artificial intelligence, emphasizing improvements in system reliability and efficiency

Using keyword co-occurrence analysis to identify thematic trends in the research field, we generated a strategic map highlighting the main themes over the studied period (see Figure 7). The most frequent keywords, derived from author keywords, were analyzed. This analysis revealed ten themes, with four grouped into primary clusters. The first cluster comprising machine learning, deep learning, and artificial intelligence centers on AI techniques aimed at enhancing solar energy production systems. The second cluster, including degradation, reliability, and photovoltaics, addresses the durability and reliability of photovoltaic systems. The third cluster focuses on fault detection and location, featuring keywords such as fault degradation, fault detection, and fault location. Lastly, the fourth cluster pertains to fault classification within photovoltaic systems.

Examining these keyword clusters, it is clear that they all converge towards a central theme: the diagnosis/prognosis of solar panel defects using artificial intelligence. This trend reflects the growing importance of using artificial intelligence to improve the reliability and performance of photovoltaic systems, as well as to reduce maintenance and repair costs.

4. Issues Identified, Key Challenges and Future Directions

To interpret the results and the color legends of keywords in this figure, readers are referred to the corresponding colors.

Utilizing keyword co-occurrence analysis to identify thematic trends in a research field, we constructed a strategic map delineating the main themes for the studied period (refer to Figure 7). We considered the most frequent keywords in relation to author keywords. In this depiction, ten themes were identified, four of which were amalgamated into primary clusters.

The first cluster (machine learning, deep learning, artificial intelligence) concentrates on leveraging machine learning and artificial intelligence techniques to enhance the efficiency of solar energy production systems. The second cluster (degradation, reliability, photovoltaics) delves into the reliability and durability aspects of photovoltaic systems. The third cluster (fault degradation, fault detection, fault location) revolves around the detection and localization of faults within photovoltaic systems. Lastly, the fourth cluster (fault classification) pertains to the classification of faults within photovoltaic systems.

Upon scrutinizing these keyword clusters, it becomes apparent that they all converge towards a central theme: the diagnosis/prognosis of solar panel defects using artificial intelligence. This trend underscores the increasing significance of employing artificial intelligence to enhance the reliability and performance of photovoltaic systems while mitigating maintenance and repair costs.

Identified Problems, Key Challenges, and Future Directions in the Implementation of Explainable Artificial Intelligence and Active Learning for Solar Panel Defect Prognosis

Addressing the question of current trends in explainable AI in PV fault diagnosis/prognosis, it's essential to acknowledge that challenges associated with deploying AI systems on photovoltaic solar systems can stem from operational, organizational, technical, and data-related factors, alongside interpretability, transparency, and trust. However, our findings indicate that explainable AI (XAI) remains underutilized in the deployment of AI algorithms for PV fault prognosis. This is evidenced by the scant presence of XAI among the most cited keywords, the limited number of documents and authors retrieved from the Scopus database, and the minimal collaboration between institutions in this specific field.

Concerning human involvement in the machine learning process to emphasize active learning, a similar observation arises. Active learning, an AI method necessitating human interaction to refine algorithmic efficiency and accuracy, encounters challenges primarily related to data collection [36]. While data is pivotal for machine learning, its collection remains a formidable challenge. Human involvement, in various forms such as labeling data selected by algorithms or providing feedback on algorithmic results, is essential to optimize the learning process and enhance algorithmic performance. However, our findings indicate that active learning remains underutilized in the detection, diagnosis, and prognosis of solar panel defects.

Regarding the main challenges related to the application of AI methods in the prognosis of the lifetime of PV systems, several key points emerge:

- The availability and quality of data are paramount for training AI algorithms. Data must be massive, secure, accessible, and of high quality to ensure accurate predictions. Advanced data collection techniques such as drone-based or wireless sensor data collection can alleviate this challenge. Moreover, human involvement in labeling significant data can facilitate user understanding.
- PV systems are complex, influenced by numerous factors like solar irradiation, temperature, and humidity. Advanced AI models such as deep neural networks are increasingly being used to account for these complexities, yet there remains a gap between their precision and user-friendliness.
- The interpretability of AI models is crucial. While accurate, AI models can be challenging to interpret, hindering informed decision-making. Explainable AI techniques such as decision trees can enhance model interpretability and facilitate informed decision-making.

In summary, addressing challenges related to AI methods in PV system prognosis necessitates considering advanced data collection techniques, complex AI models, explainable AI techniques, and continuous model learning and adaptation. Recent studies have shown that these approaches can enhance the accuracy and interpretability of AI models, thus facilitating informed decision-making regarding PV system maintenance and replacement [79-80].

5. Discussion

In this section, we'll delve into the future trends and benefits of our bibliometric study on the use of artificial intelligence (AI) for detecting and diagnosing anomalies and defects in modern photovoltaic systems (PV), alongside presenting the limitations of this work.

Our bibliometric study has shed light on the promising emergence of Explainable Artificial Intelligence (XAI) and active learning as means to enhance the transparency and comprehension of AI methods applied to photovoltaic systems. Particularly, integrating human expertise in the data labeling process has proven to be a decisive simplifying approach for effective AI utilization in this field. The benefits of XAI and active learning have been substantiated by several recent studies, showcasing their potential to augment the accuracy and reliability of AI methods deployed in the photovoltaic domain. Moreover, our analysis has unveiled the predominant influence of China in the realm of explainable AI applied to photovoltaic systems. This dominance can be attributed to the pronounced focus of Chinese universities on research in this sector, coupled with China's status as the world's leading producer of solar panels. However, our study has also underscored a notable underinvestment in the domain of AI explainability (XAI) concerning solar panels. Additional initiatives aimed at exploring the utilization of XAI in this field are advocated to enhance the transparency and understanding of the employed AI methods.

Regarding future trends, we anticipate a continuous proliferation in the utilization of AI for detecting and diagnosing anomalies and failures in modern photovoltaic systems. Specifically, the ascendancy of deep learning and machine learning is projected to persist in this domain, owing to their capacity to process vast datasets and identify intricate patterns that would be challenging to discern using conventional methods.

Lastly, we believe that our study confers several benefits. Firstly, it furnishes a comprehensive overview of AI utilization in the photovoltaic systems domain, with particular emphasis on XAI and active learning. Furthermore, our work accentuates the advantages of incorporating human expertise in the data labeling process, thereby significantly facilitating AI utilization in this context. Lastly, our study identifies promising trends and research directions in this field, which can guide future endeavors for both researchers and practitioners.

5.1. Limitations

While this bibliometric study has provided valuable insights into the utilization of AI in detecting and diagnosing defects in photovoltaic systems, it bears certain limitations.

Primarily, the study focused solely on articles published in journals and conferences, thereby overlooking research works published as technical reports, theses, or patents. This might have led to an underrepresentation of certain trends or developments in the field. Additionally, employing specific keywords to identify relevant articles may have excluded pertinent works that did not utilize these keywords, potentially introducing a bias towards certain AI methods or approaches. Furthermore, the utilization of a single database, namely Scopus, may be perceived as a limitation, as it might not encompass all relevant publications in the domain of AI applied to photovoltaic systems. Lastly, the study exclusively focused on articles published in English, potentially excluding relevant articles published in other languages.

In summary, these limitations warrant consideration when interpreting the results of this bibliometric study. Future research could broaden the scope of the study by incorporating additional data sources such as Web of Science, IEEE Xplore, or Google Scholar, adopting different keywords, and considering languages other than English.

5.2. Future Word Orientation

To enhance the results and steer advancements and initiatives in the studied field, we propose the following recommendations:

- ➢ Encourage research on Explainable AI and Active Learning in the domain of photovoltaic systems to bolster the transparency and understanding of AI methods utilized.
- ➢ Develop standards for Explainable AI in photovoltaic systems to ensure the transparency, comprehensibility, reliability, and accessibility of the employed AI methods.
- ➢ Foster collaboration between AI researchers and solar energy experts to develop effective AI methods tailored to photovoltaic systems, emphasizing interdisciplinary collaboration to grasp the challenges and opportunities in this domain.

6. Conclusion

In conclusion, this bibliometric study has underscored the significance of explainable AI and active learning in detecting dysfunctions in photovoltaic systems. Despite the rapid increase in publications in this field, the majority of proposed methods often overlook the explainability aspect of AI and the necessity for human interaction in the learning process. This oversight has resulted in a proliferation of AI methods in theory but limited real-world applications. Additionally, we've identified key players in this domain, including prolific authors, active countries, and fruitful collaborations. These findings serve as a valuable resource for researchers to discern current trends and gaps in the literature and to forge productive collaborations to propel the field forward.

Based on our results, we have formulated several recommendations to improve future research. Firstly, it is crucial to consider the explainability aspect of AI in methods for detecting dysfunctions in photovoltaic systems. Secondly, integrating human interaction in the learning process is essential to enhance the accuracy and reliability of the proposed methods. Finally, encouraging collaborations between researchers, domain engineers, and social scientists is vital to develop more effective and practically applicable methods.

In summary, this bibliometric study has emphasized the significance of explainable AI and active learning in detecting dysfunctions in photovoltaic systems, while also highlighting current gaps in the literature. The proposed recommendations aim to facilitate the development of more effective and applicable methods, thereby contributing to the global energy transition towards more reliable and sustainable renewable energy sources.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

- ➢ The authors are grateful to Patrice ELE ABIAMA, Head Manager of the National Committee for Development of Technologies (CNDT)/ Ministry of Scientific Research and Innovation (Cameroon) for his valuable help and his supervision.

- The authors thank the Production Engineering Laboratory (LGP) of the National School of Engineers of Tarbes (ENIT) and also *Institut Universitaire de la Cote* (IUC) for the technical and material support during the evaluation and the redaction of this work.

References

[1] S. Arya, P. Mahajan, "Introduction to Solar Cells," *In: Solar Cells*, Springer, Singapore, 2023, 10.1007/978-981-99-7333-0_1

[2] P.G.V. Sampaio, M.O.A. González, "Photovoltaic solar energy: Conceptual framework," *Renewable and Sustainable Energy Reviews*, 74, 590–601, 2017, 10.1016/j.rser.2017.02.081.

[3] L. Melnyk, H. Sommer, O. Kubatko, M. Rabe, S. Fedyna, "The economic and social drivers of renewable energy development in OECD countries," *Problems and Perspectives in Management*, **18**(4), 37–48, 2020, 10.21511/ppm.18(4).2020.04

[4] F. Chien, C.C. Hsu, Z. Andlib, M.I. Shah, T. Ajaz, M.G. Genie, "The role of solar energy and eco-innovation in reducing environmental degradation in China: Evidence from QARDL approach," *Integrated Environmental Assessment and Management*, **18**(2), 555–571, 2022, 10.1002/ieam.4500

[5] A. Afful-Dadzie, S.K. Mensah, E. Afful-Dadzie, "Ghana renewable energy master plan: The benefits of private sector participation," *Scientific African*, 17, e01353, 2022, 10.1016/j.sciaf.2022.e01353.

[6] J. Jung, S. Han, B. Kim, "Digital numerical map-oriented estimation of solar energy potential for site selection of photovoltaic solar panels on national highway slopes," *Applied Energy*, 242, 57–68, 2019, 10.1016/j.apenergy.2019.03.101.

[7] S. Ansari, A. Ayob, M.S.H. Lipu, M.H.M. Saad, A. Hussain, "A review of monitoring technologies for solar PV systems using data processing modules and transmission protocols: Progress, challenges and prospects," *Sustainability*, 13(15), 8120, 2021, 10.3390/su13158120

[8] J.E.F. da Fonseca, F.S. de Oliveira, C.W.M. Prieb, A. Krenzinger, "Degradation analysis of a photovoltaic generator after operating for 15 years in southern Brazil," *Solar Energy*, 196, 196–206, 2020, 10.1016/j.solener.2019.11.086

[9] A. Mellit, G.M. Tina, S.A. Kalogirou, "Fault detection and diagnosis methods for photovoltaic systems: A review," *Renewable and Sustainable Energy Reviews*, 91, 1–17, 2018, 10.1016/j.rser.2018. 03.062

[10] A.A. Al-Katheri, E.A. Al-Ammar, M.A. Alotaibi, W. Ko, S. Park, H.J. Choi, "Application of artificial intelligence in PV fault detection," *Sustainability*, **14**(21), 13815, 2022, 10.3390/su142113815.

[11] R.G. Vieira, M. Dhimish, F.M. de Araújo, M.I. Guerra, "PV module fault detection using combined artificial neural network and sugeno fuzzy logic," *Electronics*, **9**(12), 2150, 2020, 10.3390/electronics 9122150.

[12] T. Rahman, A.A. Mansur, M.S. Hossain Lipu, M.S. Rahman, R.H. Ashique, M.A. Houran, E. Hossain, "Investigation of degradation of solar photovoltaics: A review of aging factors, impacts, and future directions toward sustainable energy management," *Energies*, **16**(9), 3706, 2023, 10.3390/en16093706.

[13] F. Aziz, A.U. Haq, S. Ahmad, Y. Mahmoud, M. Jalal, U. Ali, "A novel convolutional neural network-based approach for fault classification in photovoltaic arrays," *IEEE Access*, 8, 41889–41904, 2020, 10.1109/ACCESS.2020.2977116.

[14] G. Tina, F. Cosentino, C. Ventura, "Monitoring and Diagnostics of Photovoltaic Power Plants," in *Renewable Energy in the Service of Mankind Vol II*, A. Sayigh (ed.), Springer, Cham, 2016, 10.1007/978-3-319-18215-5_45.

[15] J.A. Tsanakas, L.D. Ha, F. Al Shakarchi, "Advanced inspection of photovoltaic installations by aerial triangulation and terrestrial georeferencing of thermal/visual imagery," *Renewable Energy*, 102, 224–233, 2017, 10.1016/j.renene.2016. 10.046.

[16] J.A. Tsanakas, L. Ha, C. Buerhop, "Faults and infrared thermographic diagnosis in operating c-Si photovoltaic modules: A review of research and future challenges," *Renewable and Sustainable Energy Reviews*, 62, 695–709, 2016, 10.1016/j.rser.2016.04.079.

[17] Q. Wang, R. Li, L. Zhan, "Blockchain technology in the energy sector: From basic research to real world applications," *Computer Science Review*, 39, 100362, 2021, 10.1016/j.cosrev.2021.100 362.

[18] Z. Yang, C. Zhu, Y. Zhu, X. Li, "Blockchain technology in building environmental sustainability: A systematic literature review and future perspectives," *Building and Environment*, 245, 110970, 2023, 10.1016/j.buildenv.2023.110970

[19] S.M. Abu, M.A. Hannan, P.J. Ker, M. Mansor, S.K. Tiong, T.I. Mahlia, "Recent progress in electrolyser control technologies for hydrogen energy production: A patent landscape analysis and technology updates," *Journal of Energy Storage*, 72, 108773, 2023, doi:10.1016/j.est.2023.108773.

[20] T. Saheb, L. Izadi, "Paradigm of IoT big data analytics in the healthcare industry: A review of scientific literature and mapping of research trends," *Telematics and Informatics*, 41, 70–85, 2019, 10.1016/j.tele.2019.03.005.

[21] D.F. Dominković, J.M. Weinand, F. Scheller, M. D'Andrea, R. McKenna, "Reviewing two decades of energy system analysis with bibliometrics," *Renewable and Sustainable Energy Reviews*, 153, 111749, 2022, 10.1016/j.rser.2021.111749

[22] A.T. Keleko, B. Kamsu-Foguem, R.H. Ngouna, et al., "Artificial intelligence and real-time predictive maintenance in industry 4.0: a bibliometric analysis," *AI Ethics*, 2, 553–577, 2022, 10.1007/s43681-021-00132-6.

[23] V.V.F. Grubisic, J.P.F. Aguiar, Z. Simeu-Abazi, "A Review on Intelligent Predictive Maintenance: Bibliometric analysis and new research directions," in *2020 International Conference on Control, Automation and Diagnosis (ICCAD)*, IEEE, Oct. 2020, pp. 1–6, 10.1109/ICCAD49821.2020.9260504.

[24] M.A. Noman, E.S.A. Nasr, A. Al-Shayea, H. Kaid, "Overview of predictive condition-based maintenance research using bibliometric indicators," *Journal of King Saud University - Engineering Sciences*, 31(4), 355–367, 2019, 10.1016/j.jksues.2018.02.003.

[25] M. Bertolini, D. Mezzogori, M. Neroni, F. Zammori, "Machine Learning for industrial applications: A comprehensive literature review," *Expert Systems with Applications*, 175, 114820, 2021, 10.1016/j.eswa.2021.114820.

[26] E.H. Sepúlveda-Oviedo, L. Travé-Massuyès, A. Subias, M. Pavlov, C. Alonso, "Fault diagnosis of photovoltaic systems using artificial intelligence: A bibliometric approach," *Heliyon*, 2023, 10.1016/j.heliyon.2023.e21491.

[27] M. Aria, C. Cuccurullo, "bibliometrix: An R-tool for comprehensive science mapping analysis," *Journal of Informetrics*, 11(4), 959–975, 2017, 10.1016/j.joi.2017.08.007

[28] N.J. van Eck, L. Waltman, "Visualizing Bibliometric Networks," in *Measuring Scholarly Impact*, Y. Ding, R. Rousseau, D. Wolfram (eds), Springer, Cham, 2014, pp. 285–320, 10.1007/978-3-319-10377-8_13.

[29] W. Samek, T. Wiegand, K.R. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," *arXiv preprint arXiv:1708.08296*, 2017, 10.48550/arXiv.1708.08296

[30] L. Von Rueden, S. Mayer, K. Beckh, B. Georgiev, S. Giesselbach, R. Heese, … J. Schuecker, "Informed machine learning – a taxonomy and survey of integrating prior knowledge into learning systems," *IEEE Transactions on Knowledge and Data Engineering*, **35**(1), 614–633, 2021, **DOI :**10.1109/TKDE. 2021.3079836.

[31] X. Liang, A. Mead, M.S.R. Siddiqui, L. Van Waerbeke, A. Zhitnitsky, "Axion quark nugget dark matter: Time modulations and amplifications," *Physical Review D*, **101**(4), 043512, 2020, 10.1103/PhysRevD.101.043512

[32] A. Adadi, M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)," *IEEE Access*, 6, 52138–52160, 2018, 10.1109/ACCESS.2018.2870052

[33] G. Yang, Q. Ye, J. Xia, "Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond," *Information Fusion*, 77, 29–52, 2022, 10.1016/j.inffus.2021.07.016

[34] S. Budd, E. C. Robinson, B. Kainz, "A survey on active learning and human-in-the-loop deep learning for medical image analysis," Medical Image Analysis, 71, 102062, 2021, 10.1016/j.media.2021.1020 62

[35] D. Gissin, S. Shalev-Shwartz, "Discriminative active learning," *arXiv preprint arXiv:1907.06347*, 2019, 10.48550/arXiv.1907.06347.

[36] P. Ren, Y. Xiao, X. Chang, P. Y. Huang, Z. Li, B. B. Gupta, … & X. Wang, "A survey of deep active learning," *ACM Computing Surveys (CSUR)*, vol. 54, no. 9, pp. 1–40, 2021. 10.1145/3472291.

[37] Z. Tu et Z. Xu, « Deep active learning for object detection: A survey », *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no 11, pp. 4417–4430, 2020. 10.32604/cmes.2023. 028018

[38] D. S. Weld et G. Bansal, "The challenge of crafting intelligible intelligence", *Communications of the ACM*, vol. 62, no 6, pp. 70-79, 2019. 10.1145/3282486

[39] M. R. Hosseini, M. Maghrebi, A. Akbarnezhad, I. Martek, M. Arashpour, "Analysis of citation networks in building information modeling research," Journal of Construction Engineering and Management, **144**(8), 04018064, 2018, 10.1061/(ASCE) CO.1943-7862.0001492.

[40] Y. Hong, D. W. Chan, A. P. Chan, J. F. Yeung, "Critical analysis of partnering research trend in construction journals," *Journal of Management in Engineering*, **28**(2), 82–95, 2012, doi:10.1061/(ASCE)ME.1943-5479.0000084.

[41] M. K. Lim, Y. Li, C. Wang, M. L. Tseng, "A literature review of blockchain technology applications in supply chains: A comprehensive analysis of themes, methodologies and industries," *Computers & Industrial Engineering*, 154, 107133, 2021, doi:10.1016/j.cie.2021.107133.

[42] S. Chung, L. F. Abbott, "Neural population geometry: An approach for understanding biological and artificial neural networks," *Current Opinion in Neurobiology*, 70, 137–144, 2021, doi:10.1016/j.conb.2021.10.010.

[43] N.J. van Eck, L. Waltman, J. van den Berg, U. Kaymak, "Visualizing the computational intelligence field [Application Notes]," *IEEE Computational Intelligence Magazine*, **1**(4), 6–10, 2006, doi:10.1109/MCI .2006.329702.

[44] R. Klavans, K. W. Boyack, "Quantitative evaluation of large maps of science," *Scientometrics*, **68**(3), 475–499, 2006, doi:10.1007/s11192-006-0125-x.

[45] N. Van Eck, L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, **84**(2), 523–538, 2010, doi:10.1007/s11192-009-0146-3.

[46] N. J. Van Eck, L. Waltman, R. Dekker, J. Van Den Berg, "A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS," *Journal of the American Society for Information Science and Technology*, **61**(12), 2405–2416, 2010, doi:10.1002/asi.21421.

[47] J. Wang, X. Zhao, X. Guo, B. Li, "Analyzing the research subjects and hot topics of power system reliability through the Web of Science from 1991 to 2015," *Renewable and Sustainable Energy Reviews*, 82, 700–713, 2018, doi:10.1016/j.rser.2017.09.064..

[48] J. Wang, M. K. Lim, C. Wang, M. L. Tseng, "The evolution of the Internet of Things (IoT) over the past 20 years," *Computers & Industrial Engineering*, 155, 107174, 2021, doi:10.1016/j.cie.2021.107174..

[49] W. Pan, L. Jian, T. Liu, "Grey system theory trends from 1991 to 2018: a bibliometric analysis and visualization," *Scientometrics*, 121, 1407–1434, 2019, doi: 10.1007/s11192-019-03256-z.

[50] C. Utama, C. Meske, J. Schneider, R. Schlatmann, C. Ulbrich, "Explainable artificial intelligence for photovoltaic fault detection: A comparison of instruments," *Solar Energy*, 249, 139–151, 2023, doi: 10.1016/j.solener.2022.11.018.

[51] R. Todeschini, V. Consonni, *Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices, references*, John Wiley & Sons, 2009..

[52] G. Hong, S. Diao, A. L. Antaris, H. Dai, "Carbon nanomaterials for biological imaging and nanomedicinal therapy," *Chemical Reviews*, **115**(19), 10816–10906, 2015, doi:10.1021/acs.chemrev.5b00008.

[53] G. Xia, L. Cao, G. Bi, "A review on battery thermal management in electric vehicle application," *Journal of Power Sources*, 367, 90–105, 2017, doi:10.1016/j.jpowsour.2017.09.046..

[54] S. Ren, Y. Zhang, Y. Liu, T. Sakao, D. Huisingh, C. M. Almeida, "A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: A framework, challenges and future research directions," *Journal of Cleaner Production*, 210, 1343–1365, 2019, doi:10.1016/j.jclepro.2018.11.025.

[55] R. W. Sabnis, *Handbook of biological dyes and stains: synthesis and industrial applications*, John Wiley & Sons, 2010.

[56] K. Amasyali, N. M. El-Gohary, "A review of data-driven building energy consumption prediction studies," *Renewable and Sustainable Energy Reviews*, 81, 1192–1205, 2018, doi: 10.1016/j.rser.2017. 04.095..

[57] Z. Wang, X. Wang, S. Cong, F. Geng, Z. Zhao, "Fusing electrochromic technology with other advanced technologies: A new roadmap for future development," *Materials Science and Engineering: R: Reports*, 140, 100524, 2020, doi: 10.1016/j.mser. 2019.100524.

[58] V. Venkatasubramanian, R. Rengaswamy, K. Yin, S. N. Kavuri, "A review of process fault detection and diagnosis: Part I: Quantitative model-based methods," *Computers & Chemical Engineering*, **27**(3), 293–311, 2003, doi: 10.1016/S0098-1354(02)00160-6..

[59] R. E. Silva, R. Gouriveau, S. Jemei, D. Hissel, L. Boulon, K. Agbossou, N. Y. Steiner, "Proton exchange membrane fuel cell degradation prediction based on adaptive neuro-fuzzy inference systems," *International Journal of Hydrogen Energy*, **39**(21), 11128–11144, 2014, doi: 10.1016/j.ijhydene.2014.05.005.

[60] J. Poon, P. Jain, I. C. Konstantakopoulos, C. Spanos, S. K. Panda, S. R. Sanders, "Model-based fault detection and identification for switching power converters," *IEEE Transactions on Power Electronics*, **32**(2), 1419–1430, 2016, doi: 10.1109/TPEL.2016.2541342.

[61] W. Wu, S. Wang, W. Wu, K. Chen, S. Hong, Y. Lai, "A critical review of battery thermal performance and liquid based battery thermal management," *Energy Conversion and Management*, 182, 262–281, 2019, doi: 10.1016/j.enconman.2018.12.051.

[62] S. R. Madeti, S. N. Singh, "A comprehensive study on different types of faults and detection techniques for solar photovoltaic system," *Solar Energy*, 158, 161–185, 2017, doi: 10.1016/j.solener.2017.08.069..

[63] R. Menzel, *Photonics: linear and nonlinear interactions of laser light and matter*, Springer Science & Business Media, 2007.

[64] F. Harrou, Y. Sun, B. Taghezouit, A. Saidi, M. E. Hamlati, "Reliable fault detection and diagnosis of photovoltaic systems based on statistical monitoring approaches," *Renewable Energy*, 116, 22–37, 2018, doi: 10.1016/j.renene.2017.09.048.

[65] M. M. Abu-Khader, "Recent advances in nuclear power: A review," *Progress in Nuclear Energy*, **51**(2), 225–235, 2009, doi: 10.1016/j.pnucene.2008.05.001.

[66] M. A. Al-Alwani, A. B. Mohamad, N. A. Ludin, A. A. H. Kadhum, & K. Sopian, "Dye-sensitised solar cells: Development, structure, operation principles, electron kinetics, characterisation, synthesis materials and natural photosensitisers," *Renewable and Sustainable Energy Reviews*, 65, 183–213, 2016, doi: 10.1016/j.rser.2016.06.045..

[67] C. B. Pandey, U. Kumar, M. Kaviraj, K. J. Minick, A. K. Mishra, & J. S. Singh, "DNRA: a short-circuit in biological N-cycling to conserve nitrogen in terrestrial ecosystems," *Science of the Total Environment*, 738, 139710, 2020, doi: 10.1016/j.scitotenv.2020.139710.

[68] M. Ghiasi, T. Niknam, Z. Wang, M. Mehrandezh, M. Dehghani, & N. Ghadimi, "A comprehensive review of cyber-attacks and defense mechanisms for improving security in smart grid energy systems: Past, present and future," *Electric Power Systems Research*, 215, 108975, 2023, doi: 10.1016/j.epsr.2022.108975.

[69] A. Das, P. Rad, "Opportunities and challenges in explainable artificial intelligence (XAI): A survey," *arXiv preprint* arXiv:2006.11371, 2020, doi: 10.48550/arXiv.2006.11371

[70] Y. Hong, D. W. Chan, A. P. Chan, J. F. Yeung, "Critical analysis of partnering research trend in construction journals," *Journal of Management in Engineering*, **28**(2), 82–95, 2012, doi: 10.1061/(ASCE)ME.1943-5479.0000084..

[71] R. Machlev, L. Heistrene, M. Perl, K. Y. Levy, J. Belikov, S. Mannor, Y. Levron, "Explainable Artificial Intelligence (XAI) techniques for energy

and power systems: Review, challenges and opportunities," *Energy and AI*, 9, 100169, 2022, doi: 10.1016/j.egyai. 2022.100169.

[72] B. Dong, G. Xu, X. Luo, Y. Cai, W. Gao, "A bibliometric analysis of solar power research from 1991 to 2010," *Scientometrics*, **93**(3), 1101–1117, 2012, doi: 10.1007/s11192-012-0675-0.

[73] M. Hajji, M. F. Harkat, A. Kouadri, K. Abodayeh, M. Mansouri, H. Nounou, M. Nounou, "Multivariate feature extraction based supervised machine learning for fault detection and diagnosis in photovoltaic systems," *European Journal of Control*, 59, 313–321, 2021, doi: 10.1016/j.ejcon.2020.03.004.

[74] M. Callon, J. P. Courtial, W. A. Turner, S. Bauin, "From translations to problematic networks: An introduction to co-word analysis," *Social Science Information*, **22**(2), 191–235, 1983, doi: 10.1177/053901883022002003.

[75] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, F. Herrera, "Science mapping software tools: Review, analysis, and cooperative study among tools," *Journal of the American Society for Information Science and Technology*, **62**(7), 1382–1407, 2011, doi: 10.1002/asi.21525..

[76] C. Cheng, F. Wang, Y. Tian, X. Wu, J. Zheng, J. Zhang, ... & J. Zhao, "Review and prospects of hydrate cold storage technology," *Renewable and Sustainable Energy Reviews*, 117, 109492, 2020, doi: 10.1016/j.rser.2019.109492.

[77] C. Koch, L. Hirth, "Short-term electricity trading for system balancing: An empirical analysis of the role of intraday trading in balancing Germany's electricity system," *Renewable and Sustainable Energy Reviews*, 113, 109275, 2019, doi: 10.1016/j.rser.2019.109275.

[78] X. Liang, A. Mead, M. S. R. Siddiqui, L. Van Waerbeke, A. Zhitnitsky, "Axion quark nugget dark matter: Time modulations and amplifications," *Physical Review D*, **101**(4), 043512, 2020, doi: 10.1103/PhysRevD. 101.043512..

[79] D. Rodríguez-Gracia, M. de las Mercedes Capobianco-Uriarte, E. Terán-Yépez, J. A. Piedra-Fernández, L. Iribarne, R. Ayala, "Review of artificial intelligence techniques in green/smart buildings," *Sustainable Computing: Informatics and Systems*, 38, 100861, 2023, doi: 10.1016/j.suscom.2023.100861..

[80] J. Hmad, A. Houari, A. E. M. Bouzid, A. Saim, H. Trabelsi, « A review on mode transition strategies between grid-connected and standalone operation of voltage source inverters-based microgrids », *Energies*, vol. 16, no. 13, p. 5062, 2023, doi: 10.3390/en16135062.