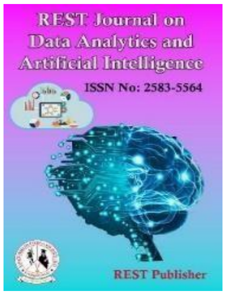
REST Journal on Data Analytics and Artificial Intelligence
Vol: 2(3), September 2023
REST Publisher; ISSN: 2583-5564
Website: http: //restpublisher.com/journals/jdaai/
DOI: https://doi.org/10.46632/jdaai/2/3/14



Assessing Explainable in Artificial Intelligence: A TOPSIS Approach to Decision-Making

***Srinivasa Rao Kolusu**
Sr. Technical Account Manager, Dallas, Texas, USA.
*Corresponding author: *srk8082@ieee.org*

Abstract: *Explainable in Artificial Intelligence (AI) is the ability to comprehend and explain how AI models generate judgments or predictions. The complexity of AI systems, especially machine learning models, is increasing. understanding their reasoning process becomes crucial for ensuring trust, fairness, and accountability. Explainable AI (XAI) helps demystify the "black box" character of sophisticated models, Deep neural networks, for example, which allows users to to grasp how inputs are transformed into outputs. In many AI system judgments can have a big impact on industries including healthcare, banking, and law making transparency a necessity. Explainable also aids in identifying and mitigating biases, improving model performance, and complying with regulatory requirements. As AI technologies evolve, there is an increasing emphasis on balancing model accuracy with interpretability, making some AI systems remain ethical, transparent, and in line with human values. In artificial intelligence (AI) research, Explainable is essential for fostering confidence, guaranteeing responsibility, and enhancing The openness of artificial intelligence systems. As Artificial intelligence models, especially intricate ones like deep learning, become more widely adopted, understanding their Processes for making decisions are crucial for validating their outcomes. The goal of explainable AI (XAI) research is to create models interpretable so that users can comprehend the decision-making process. This is particularly crucial in high-stakes industries like healthcare, banking, and law, where poor or prejudiced choices can have serious repercussions. Explainable also supports regulatory compliance, model improvement, and ethical AI deployment. An approach to decision-making known as TOPSIS (Technique for Order of Preference by Similarity to Ideal Answer) evaluates how far an alternative is from the worst-case situation and how close it is to the ideal solution. The worst-case solution shows the lowest values, while the ideal solution shows the best values given the desired criteria. Each alternative is given a similarity score by TOPSIS, which ranks them according to how near the ideal answer they are. This method is frequently used to enhance decision-making in a variety of domains, including business, engineering, environmental research, and healthcare. Alternative: LIME (Local Interpretable Model), SHAP (Shapley Additive Explanations), Deep LIFT (Deep Learning Important Features), Anchor Explanations, ICE (Individual Conditional Expectation), Counterfactual Explanations, Rule-based Explanation Systems, Saliency Maps (for CNNs), Integrated Gradients, XAI for Healthcare. Evaluation preference: Interpretability, Accuracy of Explanations, User Trust, Computational Complexity, Scalability, Flexibility. The results indicate that XAI for Healthcare ranks highest, while Saliency Maps (for CNNs) holds the lowest rank.*

Keywords: *LIME, SHAP, ICE, TOPSIS.*

1. INTRODUCTION

Because consumers need to feel secure and trusted about the processes and reasoning underlying automated decision-making across all domains, Explainable in AI has become a renewed focus of current research various sectors, including autonomous vehicles, healthcare diagnostics, and banking as well as finance. Even while Explainable in Although artificial intelligence (AI) has gained a lot of attention recently, the field's roots can be traced back to earlier decades, when knowledge-driven expert systems were the main focus of AI development. Throughout the history of artificial intelligence, Explainable has been defined, comprehended, and used in a variety of research fields, such as expert systems, machine learning, recommendation algorithms, and neural-symbolic learning and reasoning approaches. We present a historical perspective on AI that can be explained. We look at how Explainable was previously conceptualised, how it is currently understood, and what could be understood in the future. We wrap up by putting up standards providing justifications that we believe will be essential to developing systems with explanations that are understandable to humans. [1] In tasks including picture identification, audio processing, and language translation, Businesses, public agencies, and society as a whole have seen the transformative power of artificial intelligence (AI) and machine learning (ML), which can perform on par with or better than humans. Although deep learning (DL) models are quite accurate, they are frequently criticized for being opaque and operating as "black boxes." These models rely on an immense number of weight parameters sometimes numbering in the millions or billions which encode knowledge gleaned from training data. The sheer scale of these parameters and their complex relationships make it challenging to interpret how these models function in connection to real-world phenomena. This opacity presents significant challenges, particularly in key areas including healthcare, law, economics, and privacy, where openness and explain ability are essential. As AI and ML applications, including DL, continue to expand in areas like digital health, legal systems, transportation, finance, and security, the demand for interpretable and transparent solutions is becoming increasingly critical. [2] Over the past ten years, the field of explainable artificial intelligence, or XAI, has shown remarkable growth. This surge has resulted in the emergence of a vast array of domain-specific and context-oriented techniques aimed at the elucidation Building machine learning (ML) models and developing intelligible justifications for human users. Concurrently, research has focused on exploring methods to assess the efficacy of The topic of explainable artificial intelligence, or XAI, has grown significantly during the last decade can largely be attributed to the rapid ascent in the popularity of ML, particularly deep learning, which has found numerous applications across various commercial sectors, from e-commerce to gaming, as well as in fields such as computer vision, criminal justice, healthcare, and military simulations, to name a few. Regrettably, the majority of models developed using ML and deep learning have earned the designation of 'black-box' by researchers, due to their intricate, non-linear architectures that are exceedingly challenging to interpret and clarify to stakeholders. [3] Our daily lives now incorporate artificial intelligence (AI). the statistics website Statista projects that the AI sector will generate 2.59 trillion dollars in revenue globally, up from 480 billion dollars in 2017. by 2021 in US dollars According to Gartner, AI is an unavoidable technology and one of the Top 10 Strategic Technology Trends for 2018. These elements, together with immersive experiences, digital twins, event-driven thinking, and continuous adaptive security, are shaping the next generation of digital business models and ecosystems. As a result, society is being significantly impacted by the development of AI. [4] In fact, artificial intelligence has already permeated our daily lives, and we have become used to it making choices for us. Examples include Facebook friend recommendations, Google search results with tailored advertisements, and Netflix and Amazon recommendations for movies and products. However, it's critical to comprehend the reasoning behind significant judgments, including medical diagnosis. This emphasises how important it is to clarify AI results. Unfortunately, despite its seeming stability in terms of outcomes and forecasts, AI algorithms particularly Machine Learning (ML) algorithms—face opacity issues that make it difficult to understand how they operate inside. This makes the problem much more complicated since there are serious hazards associated with depending on a system that is unable to offer answers. [5] Among the many industries that have embraced cutting-edge information technology, artificial intelligence (AI) is at the forefront. Even though artificial intelligence has been around for a few decades it is widely acknowledged that intelligent machines now possess critical capabilities in learning, reasoning, and adapting. Thanks to these strengths, AI methodologies are reaching unmatched performance levels as they learn to tackle ever-more intricate computational challenges, positioning them as essential for the advancement of Human civilization. The complexity of AI-enhanced systems has escalated to such a degree that minimal human involvement is necessary in their creation and application. As decisions made by these systems increasingly impact human lives particularly idomains like medicine, law, and defense there is a growing necessity to grasp the foundation upon which these AI-driven decisions are made. While early AI systems were relatively straightforward to interpret, Deep Neural Networks and other opaque decision-making frameworks have become more popular in recent years (DNNs). [6] Deep Learning (DL) models, like DNNs, are successful in practice due to a combination of effective learning algorithms and their broad parametric range. DNNs are complicated black-box models because of this range, which has innumerable layers and parameters. Transparency, which is the antithesis of black-box features, is the endeavor to gain a clear understanding of how a model operates. As machine learning in a black box (ML) models gain traction for significant predictions in essential domains, the call for transparency grows stronger from various stakeholders in AI. The risk lies in generating and utilizing decisions that lack justification, legitimacy, or do not permit detailed explanations regarding their behavior. Providing explanations that underpin a model's output is vital, especially in precision medicine, where professionals demand more insight than a mere binary prediction to assist in their diagnoses. Additional instances involve security and driverless cars applications, and money, to name a few. Generally speaking, people are hesitant to employ methods that are not instantly interpretable., controllable, and trustworthy, especially in view of the growing demand for ethical AI. It's well accepted that focusing primarily on performance makes systems more ambiguous. The clarity of a model is traded off against performance, thus this idea has some validity. However, a deeper understanding of a system can enable the rectification of its flaws. [7] It is anticipated that advancements in machine learning (ML) techniques will lead to the creation of AI systems with the ability to observe, learn, make judgments, and act independently. However, they won't be able to tell human users about their decisions and actions. This shortcoming is especially important for the Department of Defense, which faces challenges that need the creation of more intelligent, autonomous, and symbiotic systems. Explainable AI will be essential if humans are to successfully understand, trust, and manage these artificially intelligent partners. This issue prompted DARPA to launch its explainable artificial intelligence (XAI) project in May 2017. For instance, referring to this initiative as explainable AI rather than interpretable, intelligible, or transparent AI emphasizes DARPA's objective of employing persuasive justifications to increase human comprehension of AI systems. The XAI team's interest in the human psychology of explanation, which stems from the extensive study and knowledge in the social sciences, is also reflected in it. [8] The development of AI systems with perception, learning, decision-making, and autonomous behaviour is possible because to advancements in machine learning (ML) techniques. But these technologies will have a hard time explaining their choices and behaviours to human operators. The Department of Defense is especially affected by this shortcoming because of the difficulties it encounters, which call for the creation of more intelligent, independent, and collaborative systems. In order for consumers to understand, trust, and successfully manage these AI partners, explainable AI will be essential. DARPA's explainable artificial intelligence (XAI) initiative was launched in May 2017 to address this demand. Explainable AI, according to DARPA, is defined as AI systems that can communicate their logic to a human operator, recognise their advantages and disadvantages, and offer predictions about how they will behave in the future. This program's designation as explainable AI (rather than interpretable, intelligible, or transparent AI, for example) reflects DARPA's goal of creating AI systems that can be effectively explained to humans in order to make them easier to comprehend. It also demonstrates the XAI team's emphasis on the human psychology underlying explanation, which is guided by a wealth of social scientific research and experience. [9] One area of computer science that aims to understand is artificial intelligence, or AI. the operations of intelligent beings to develop software that mimics their behaviors. Initially, these models were focused on supporting clinical decision-making. For example, early AI systems paid particular attention to tasks such as interpreting electrocardiogram (ECG) signals, diagnosing medical conditions, selecting appropriate treatments, and analyzing clinical reasoning. Consequently, a substantial amount of AI research has aimed to alleviate the burdens faced by healthcare practitioners. [10] Clarity in predictive modeling systems is crucial for their effectiveness in healthcare. In the absence of openness, fostering trust among healthcare professionals and integrating predictive models into their routines becomes challenging. Explainable AI has gained a lot of attention in recent years (XAI). In order to elucidate factors related to the input, output, performance, and methodology (how) of the predictive models, the XAI system should address information-driven explanation queries. In order to investigate why, why not, what if, how to do that, and what keeps it that way when users provide examples to clarify the predictions, user interfaces also need to include instance-based explanation inquiries. Users must create examples and understand how the predictive model makes its decisions. Revolutionizing the AI framework requires including end users (such as physicians, patients, and operational leaders) from the very beginning of data analysis and maintaining a continuous conversation between change-responsive AI and AI-informed transformations. AI's success in the healthcare industry depends on everyone's participation in the process as well as on improvements in AI algorithms. [11] Explainable artificial intelligence (XAI) has grown so rapidly that keeping up with developments or gaining an initial understanding has become increasingly challenging. To address this, articles and reviews offering taxonomies of XAI methods have emerged as valuable tools for providing a structured overview of the domain. However, this trend has resulted in multiple, often competing methods for creating taxonomies. This article looks at recent attempts to create taxonomies for XAI exploring the overarching challenges they face along with their respective strengths and weaknesses. Our review aims to raise awareness among scholars about the limitations and issues inherent in current taxonomic approaches. We argue that relying solely on one taxonomy may not adequately take a picture of the complexity of the XAI landscape. In order to remedy this, we suggest three potential solutions: the development of a new comprehensive taxonomy that integrates existing approaches, the creation of a centralized collection of XAI techniques, as well as the creation of a decision tree to assist in selecting appropriate methods.[12] Artificial Intelligence (AI), an academic discipline established in the 1950s, has rapidly evolved to capture significant interest across scholarly and practical fields. Global funding for AI technologies is expected to soar, rising from $38 billion in 2019 to an anticipated $98 billion by 2023. At its core, AI particularly through machine learning describes a system's capacity to examine outside data, extract meaningful insights, and adaptively utilize this knowledge to achieve specific goals. Enhanced computational power has driven remarkable advancements in methods for machine learning, such as artificial neural networks, enabling unprecedented levels of performance. Modern solutions increasingly employ bio-inspired paradigms. However, most AI systems today are categorized as "narrow" or "weak" AI, excelling at specialized tasks rather than exhibiting the general adaptability of "strong AI." Despite this limitation, contemporary algorithms have outperformed humans in several domains, including mastering complex games like Go and Poker and achieving superior accuracy in medical diagnostics, such as detecting breast cancer. These advancements are set to significantly shape the trajectory of work within socio-technical systems. [13] With the advancement of technology, the generation of big data across nearly all scientific domains has become commonplace. But evaluating such large data sets is difficult, especially artificial intelligence (AI) techniques are. Many of these methods, especially deep learning, operate as opaque "black box" models, making their internal mechanisms difficult to understand. This lack of transparency has raised serious concerns in crucial domains including criminal justice and healthcare, leading to a growing emphasis on developing explainable AI (XAI). Instead of adopting the conventional approach of outlining what XAI should ideally be, this paper takes a different perspective. It explores what XAI is by examining its practical, scientifically grounded characteristics rather than engaging in speculative aspirations, with a focus on framing these insights within a broader theoretical context beyond the realm of physics. [14] Argumentation and Explainable Artificial Intelligence (XAI) have become increasingly intertwined, with Argumentation playing a crucial role in enhancing AI systems' transparency. By detailing the reasoning process underlying an AI's decisions, Argumentation provides clarity in ambiguous situations and helps reconcile conflicting information. This review explores the synergy between Argumentation and XAI, focusing on key methodologies, research developments, and applications that utilize Argumentation to enhance AI Explainable. It examines how Argumentation contributes to addressing decision-making challenges, supporting justification, and enabling effective dialogue. Additionally, the survey highlights Argumentation's impact on creating interpretable systems across various domains, including General-purpose, robotics, security, the Semantic Web, medical informatics, and AI. Lastly, it discusses cutting-edge techniques that integrate Machine Learning with Argumentation Theory to advance the creation of more understandable predictive models. [15]

2. MATERIAL AND METHODS

Alternative:

1. LIME (Local Interpretable Model-agnostic Explanations): LIME is a method that locally approximates the predictions of black-box models (such ensemble techniques or deep neural networks) using interpretable models. LIME works by modifying the input data and tracking the predictions made by the black-box model. The altered data is then fitted to a straightforward, interpretable model (such a linear regression), which sheds light on the variables affecting the model's choice in that particular case. LIME is extremely flexible since it is "model-agnostic," meaning it may be used with any machine learning model.

2. SHAP (Shapley Additive Explanations): Cooperative game theory—more especially, Shapley values—is the foundation of SHAP which originated in the field of economics to fairly distribute the payoff of a game among its players. For a given prediction in machine learning, SHAP gives each feature a value that explains how it contributes to the model's output. SHAP has the advantage of providing a unified explanation framework, offering both global and local explanations. It is considered one of the most accurate and theoretically grounded approaches to model interpretability.

3. DeepLIFT (Deep Learning Important Features): A technique called DeepLIFT was created to provide an explanation for the predictions produced by deep neural networks. In contrast to gradient-based techniques, which compute feature importance based on gradients, DeepLIFT assigns importance scores to features by comparing A neuron's response to a reference activation. This approach overcomes some of the limitations of gradient-based methods, particularly in cases where gradients can be noisy or unreliable.

4. Anchor Explanations: Anchor Explanations are a model-agnostic technique that explains predictions by finding "anchors" or feature subsets that are highly indicative of a model’s prediction. These anchors are conditions that, when satisfied, are guaranteed to result in the same prediction regardless of other features. Anchor explanations are designed to be simple, intuitive, and locally faithful, provide lucid explanations of the model's decision-making procedure.

5. ICE (Individual Conditional Expectation): With other data held constant, ICE plots offer a means of seeing how a model's prediction changes as a particular feature changes. This approach facilitates comprehension of feature interactions and how each one affects model predictions. ICE is particularly useful in instances where feature interactions could be non-linear or complicated, as it illustrates how changes in one feature impact the model's conclusion, irrespective of others.

6. Counterfactual Explanations: Counterfactual Explanations focus on explaining a model’s decision by showing If the input characteristics had been different, what would have happened. For example, a counterfactual explanation could tell you that "if feature X had been 10% higher, the model would have predicted class Y instead of class Z." This method is especially helpful in decision-making scenarios where users want to understand what changes they can make to achieve a different outcome, providing actionable insights.

7. Rule-based Explanation Systems: The behavior of a model is explained by a set of if-then rules generated by rule-based explanation systems. Frequently, decision trees or rule induction methods are used to derive these rules. Presenting the model's decision-making process in a way that is simple for others to comprehend is the main objective. Decision trees and other naturally interpretable models benefit greatly from rule-based systems, which may be expanded to more complicated models using methods like rule extraction.

8. Saliency Maps (for CNNs): Convolutional neural networks are especially explained visually using saliency maps (CNNs). They draw attention to the areas of an input image—like pixels—that are most crucial for the model's prediction. Saliency maps show how sensitive the model is to particular aspects of the image by calculating the output's gradient in relation to the input image. Applications of computer vision such as object recognition and picture classification, this method is frequently employed.

9. Integrated Gradients: One technique for attributing deep learning models' predictions to their input characteristics is called Integrated Gradients. It computes the integral of the gradients of the model's output with respect to the input along a straight line from a baseline (such zero or mean) to the actual input. Compared to straightforward gradient-based techniques, integrated gradients offer a more consistent and dependable means of allocating significance to attributes.

10. XAI for Healthcare: Explainable AI in healthcare (XAI for Healthcare) is particularly critical, as healthcare models often impact critical decisions. XAI helps medical Patients and experts are aware of AI-driven diagnosis, predictions, and treatment recommendations. By using methods like SHAP, LIME, and saliency maps, Artificial intelligence (AI) models may offer clear insights into how patient data—like test results, genetic information, or medical images—influences choices. This openness is essential to fostering trust ensuring compliance with medical regulations, and facilitating shared decision-making between doctors and patients.

Evaluation preference:

1. Interpretability: The degree to which a person can comprehend the choices or forecasts produced by a machine learning model is known as interpretability. Models with high interpretability provide clear, understandable reasons for their predictions, often through simplified or intuitive explanations For instance, linear regression models or decision trees are considered interpretable due to their straightforward structure. Interpretability is crucial for trust, particularly in crucial areas where comprehending model behaviour is crucial, like healthcare or finance.

2. Accuracy of Explanations: The accuracy of explanations ensures that the provided justification for a model's decision correctly reflects the actual behavior of the model. This aspect is essential to ensure that the explanations genuinely represent how features influence predictions, without distorting the model’s logic. Techniques like SHAP and LIME, for instance, are prized for their capacity to offer precise, dependable explanations, assisting users in comprehending the true contributions of input elements.

3. User Trust: User trust in AI systems depends on the transparency and reliability of the explanations provided. If users can clearly understand and verify the rationale behind model predictions, their confidence in the system increases. In high-stakes areas like healthcare, trust is crucial since users (such as physicians or patients) must think that the model’s recommendations are sound and reliable.

4. Computational Complexity: The computational resources needed to produce explanations are referred to as computational complexity. When working with big datasets or intricate models like deep neural networks, some explanation approaches, like SHAP or LIME, might be computationally demanding. Efficient explanation methods balance complexity with the need for accurate, interpretable outputs.

5. Scalability: Scalability describes an explanation method's capacity to handle large datasets or complex models effectively. Scalable techniques can be applied to big data problems without significant loss of performance. Methods like LIME are scalable, though they may become computationally expensive with increasing model size.

6. Flexibility: The ability of an explanatory technique to be used with several kinds various models, including neural networks, decision trees, and ensemble approaches, is known as flexibility. More versatile approaches, such as LIME or SHAP, may be used to a variety of machine learning algorithms, which makes them useful in a variety of settings.

TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) : The TOPSIS technique takes into account a number of evaluation factors to identify the best option among a variety of possibilities. By reducing the distance from the worst-case situation and increasing the distance from other options within a set (the nadir point), it works well. The comparative weights of the TOPSIS criteria are significant and can be used into the analysis. This study looks at the various distance measuring and weighting strategies utilised in TOPSIS, as well as comparisons with other approaches and applications. [16] TOPSIS is the preferred method for requiring the least amount of subjective input from decision-makers because it just requires subjective input weights. Because of this, TOPSIS is a good choice for both increasing the distance to solutions and decreasing the distance to the Nadir point. .[17] Despite being widely used in many different fields, TOPSIS hasn't gained the same level of acceptance as attribute-based approaches. Its uses are numerous and include choosing cutters for flexible production, choosing processes for manufacturing and financial investment, and evaluating the effectiveness and success of financial ratios for certain businesses. Using neural network techniques for weighing and putting in place more intricate package extensions are two examples of method extensions. [18] TOPSIS, a popular decision-making tool, stands in opposition to the Analytic Hierarchy Process (AHP) to highlight its unique features. The tendency of TOPSIS to prioritise weights without offering a thorough judgement test is one of its drawbacks. On the other hand, AHP is limited by human information processing ability, which is usually just seven components, plus or minus two. [19] Finding the most advantageous alternative, which is distinguished by being remote, is the core idea of TOPSIS. keeps a closer greater separation between a negative and a positive ideal solution. Gelenbe's emphasis on this idea reveals the core of TOPSIS. Although TOPSIS is unable to deal with this kind of data directly, a-TOPSIS, a variation of TOPSIS, is used for algorithm ranking when benchmarks and alternatives are provided. [20] With an emphasis on carrying out a comprehensive examination, this paper tackles the problem of fairness in TOPSIS ranking indexes. In order to enhance the procedure, particularly when dealing with discrete components, Yang and Chou suggested adding multiple response simulations to the dopsys technique. It is crucial to remember that the design options produced by these simulations might not always be workable or useful in real-world situations. [21] In order to expedite the normalising procedure used in conventional TOPSIS and to reduce The issue is that in order to compare the scales, it makes advantage of a linear scale conversion. This paper proposes to expand TOPSIS in an ambiguous environment to manage the process of determining choice scenarios in uncertain circumstances when criteria comprise linguistic aspects. In order to account for uncertainty in decision-making data and teams, this extension employs estimates for each option to assess in accordance with each criterion. [22]

Step 1: The creation of the decision matrix X shows how different solutions perform in relation to specific criteria.

$$x_{ij} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (1)$$

Step 2: The criteria's weights are stated as

$$w_j = [w_1 \cdots w_n], \quad where\ , \sum_{j=1}^{n}(w_1 \ \cdots \ w_n) = 1 \quad (2)$$

Step 3: The matrix x_{ij}'s The calculated normalised values are

$$n_{ij} = \frac{x_{ij}}{\sqrt[2]{\sum_{i=1}^{m} x_{ij}^2}} \quad (3)$$

The following formula is used to calculate the weighted normalised matrix, N_{ij}.

$$N_{ij} = w_j \times n_{ij} \quad (4)$$

Step 4: Finding the optimal best and ideal worst values will be our first step: In this case, we have to decide if the influence is "+" or "-." The greatest value in a column with a "+" impact is the ideal best value for that column; the lowest value in a column with a "-" effect is the ideal worst value.

Step 5: We must now determine how each response differs from the optimal one.

$$S_i^+ = \sqrt{\sum_{j=1}^{n}(N_{ij} - A_j^+)^2} \quad for\ i \in \lceil 1, m \rceil\ and\ \ j \in \lceil 1, n \rceil \quad (5)$$

Step 6: We must now determine how each response differs from the best-case scenario.

$$S_i^- = \sqrt{\sum_{j=1}^{n}(N_{ij} - A_j^-)^2}\, for\ i \in \lceil 1, m \rceil\ and\ \ j \in \lceil 1, n \rceil \quad (6)$$

Step 7: The next step is to determine the alternative's closeness coefficient.

$$CC_i = \frac{S_i^-}{S_i^+ + S_i^-} \quad where, 0 \leq CC_i \leq 1, i \in \lceil 1, m \rceil \quad (7)$$

The value of the Closeness Coefficient shows the relative superiority of the options. A much better alternative is indicated by a larger,CC-i., whereas a significantly poorer alternative is indicated by a smaller,CC-i.

3. RESULT AND DISCUSSION

TABLE 1. Explainable in Artificial Intelligence (AI)

	Interpretability	Accuracy of Explanations	User Trust	Computational Complexity	Scalability	Flexibility
LIME (Local Interpretable Model)	8	7	9	6	8	9
SHAP (Shapley Additive Explanations)	7	9	9	8	7	8
DeepLIFT (Deep Learning Important Features)	6	8	8	7	6	7
Anchor Explanations	9	7	8	5	9	7
ICE (Individual Conditional Expectation)	7	8	7	5	8	8
Counterfactual Explanations	8	9	9	9	5	6
Rule-based Explanation Systems	10	6	8	4	9	6
Saliency Maps (for CNNs)	6	8	7	7	6	7
Integrated Gradients	7	9	8	8	7	9
XAI for Healthcare	9	9	10	7	8	5

The table 1 provided highlights various Explainable techniques in artificial intelligence (AI) and evaluates them based on six key attributes: interpretability, accuracy of explanations, user trust, computational complexity, scalability, and flexibility. LIME (Local Interpretable Model): LIME scores well in interpretability and user trust but faces challenges in computational complexity. Its scalability and flexibility are strong, making it useful in a variety of applications. SHAP (Shapley Additive Explanations): SHAP excels in accuracy and user trust, making it a highly reliable technique for providing clear explanations. It balances computational complexity and scalability, though its flexibility could be improved. DeepLIFT (Deep Learning Important Features): While DeepLIFT performs well in accuracy and user trust, it has lower flexibility and scalability, making it less versatile compared to other methods. Anchor Explanations: Anchor offers high interpretability and scalability but has limitations in computational complexity, which could affect its application in large-scale systems. ICE (Individual Conditional Expectation): ICE is strong in interpretability and scalability but has lower accuracy in explanations and faces computational challenges. Counterfactual Explanations: Known for high accuracy and user trust, counterfactual explanations face scalability and flexibility limitations, particularly in large datasets. Rule-based Explanation Systems: While offering high interpretability and scalability, rule-based systems are constrained by computational complexity, making them less efficient. Saliency Maps (for CNNs): Saliency maps provide moderate interpretability but are limited by scalability and flexibility, making them less suitable for complex tasks. Integrated Gradients: This technique achieves high accuracy and flexibility, but its scalability is not as high as some other methods. XAI for Healthcare: Specifically designed for healthcare applications, it excels in accuracy and user trust but lacks scalability and flexibility in broader contexts.

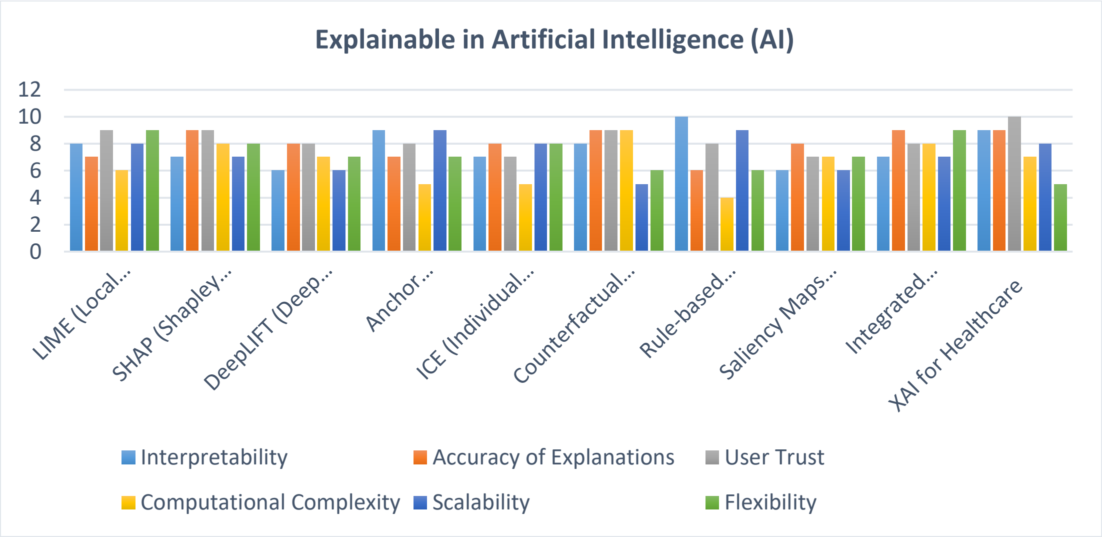


FIGURE 1. Explainable in Artificial Intelligence (AI)

The chart 1 titled " Explainable in Artificial Intelligence (AI)" compares various AI explanation methods based on five factors: Interpretability, Accuracy of Explanations, User Trust, Computational Complexity, Scalability, and Flexibility. A series of bars representing each approach is used, with the various colours signifying the various factors. Interpretable Local Model-agnostic Explanations, or LIME: This method shows a balanced performance in terms of interpretability, accuracy, and scalability, although it is slightly weaker on computational complexity. SHAP (Shapley Additive Explanations): SHAP is highly regarded for its interpretability and accuracy but faces some computational challenges, making it less efficient than other methods in terms of scalability. DeepLIFT (Deep Learning Important FeaTures): This method performs well on scalability and flexibility, making it suitable for deep learning applications, though it may sacrifice interpretability slightly. Anchor Explanations: It provides a high level of interpretability and user trust but has moderate scores for flexibility and scalability. Saliency Maps for CNNs (Convolutional Neural Networks): Known for its strong interpretability and scalability, this method works well in convolutional networks, offering good flexibility and trust but challenges in computational complexity. XAI for Healthcare: This method focuses on the healthcare sector, balancing interpretability and scalability but with some trade-offs in computational complexity.

TABLE 2. Square Root of Matrix

Square Root of Matrix					
49	81	36	64	81	64
81	81	64	49	64	49
64	64	49	36	49	36
49	64	25	81	49	81
64	49	25	64	64	49
81	81	81	25	36	64
36	64	16	81	36	100
64	49	49	36	49	36
81	64	64	49	81	49
81	100	49	64	25	81
49	81	36	64	81	64

The table 2 appears to represent the square roots of various matrix elements. The value that, when multiplied by itself, equals the original number is known as the square root of a number. In this case, the table shows values such as 49, 81, 36, etc., which are perfect squares, and their square roots are displayed as numbers like 7, 9, 6, and so on. For example, 49 has a square root of 7, 81 has a square root of 9, and 36 has a square root of 6. This matrix could be part of a mathematical exercise or a study on square roots.

TABLE 3. Normalized Data

Interpretability	Accuracy of Explanations	User Trust	Computational Complexity	Scalability	Flexibility
0.3242	0.2746	0.3409	0.2804	0.3414	0.3462
0.2837	0.3530	0.3409	0.3738	0.2988	0.3029
0.2431	0.3138	0.3030	0.3271	0.2561	0.2596
0.3647	0.2746	0.3030	0.2336	0.3841	0.3895
0.2837	0.3138	0.2651	0.2336	0.3414	0.3462
0.3242	0.3530	0.3409	0.4205	0.2134	0.2164
0.4052	0.2353	0.3030	0.1869	0.3841	0.3895
0.2431	0.3138	0.2651	0.3271	0.2561	0.2596
0.2837	0.3530	0.3030	0.3738	0.2988	0.3029
0.3647	0.3530	0.3788	0.3271	0.3414	0.3462

The table 3 represents normalized data across six factors: Interpretability, Accuracy of Explanations, User Trust, Computational Complexity, Scalability, and Flexibility. Normalization is a process that adjusts values to a common scale, making comparisons easier. In this table, each factor's values have been transformed into a range between 0 and 1. The rows represent different instances or observations, and the columns show the normalized values for each factor. For example, the normalized value for "Interpretability" in the first row is 0.3242, while "Scalability" has a value of 0.3414 in the same row. This normalized data helps in comparing performance across multiple dimensions.

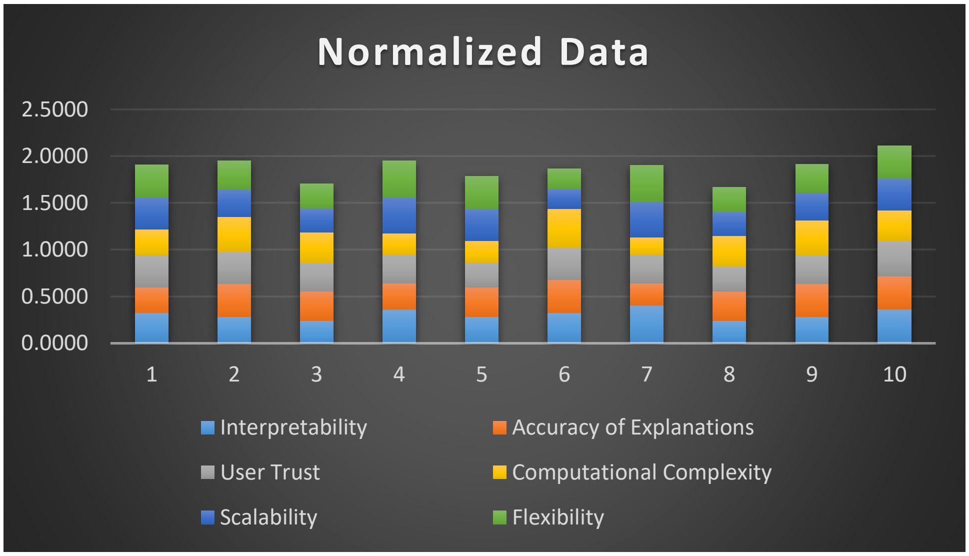


FIGURE 2. Normalized Data

The chart titled "Normalized Data" presents a comparison of different factors Interpretability, Accuracy of Explanations, User Trust, Computational Complexity, Scalability, and Flexibility across 10 items, with data normalized on a scale of 0 to 2. Each item is represented by stacked bars, with distinct colors indicating the different factors. Interpretability (Blue): This factor consistently ranks highly across all items, suggesting that interpretability is a key strength for these items. Accuracy of Explanations (Orange): This factor also performs well, though not as strongly as interpretability. It tends to be moderate across the data points. User Trust (Gray): User trust appears to have a lower value compared to interpretability and accuracy. The bars for this factor are generally shorter, indicating it may not be as prioritized as other factors. Computational Complexity (Yellow): The computational complexity is represented with varying lengths, showing that some items face more challenges in terms of efficiency. Scalability (Blue): Scalability has a moderate value across the chart, with a consistent height across items, highlighting its importance but balanced with other factors. Flexibility (Green): Flexibility ranks high, especially in certain items, suggesting that many of the evaluated methods are adaptable.

TABLE 4. Weight

0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25	0.25	0.25

A set of equal weights allocated to each dimension for assessing therapy modalities is illustrated in Table 4. Effectiveness, patient satisfaction, accessibility, side effects, scientific evidence, and affordability are all given equal weights of 0.25 in this approach, which guarantees a fair assessment across a range of factors.

TABLE 5. Weighted normalized decision matrix

Weighted normalized decision matrix					
0.0810	0.0686	0.0852	0.0701	0.0854	0.0865
0.0709	0.0883	0.0852	0.0935	0.0747	0.0757
0.0608	0.0784	0.0758	0.0818	0.0640	0.0649
0.0912	0.0686	0.0758	0.0584	0.0960	0.0974
0.0709	0.0784	0.0663	0.0584	0.0854	0.0865
0.0810	0.0883	0.0852	0.1051	0.0533	0.0541
0.1013	0.0588	0.0758	0.0467	0.0960	0.0974
0.0608	0.0784	0.0663	0.0818	0.0640	0.0649
0.0709	0.0883	0.0758	0.0935	0.0747	0.0757
0.0912	0.0883	0.0947	0.0818	0.0854	0.0865

The table 5 represents a weighted normalized decision matrix, where each entry is a normalized value adjusted by a corresponding weight. The rows represent different decision alternatives or instances, and the columns correspond to various decision criteria, such as Interpretability, Accuracy of Explanations, User Trust, Computational Complexity, Scalability, and Flexibility. The values in the table are the result of multiplying the normalized data from a previous table by the weights assigned to each criterion. This process allows for evaluating the alternatives based on their relative importance. For example, the value 0.0810 in the first row and first column is the weighted value for the "Interpretability" criterion for the first alternative.

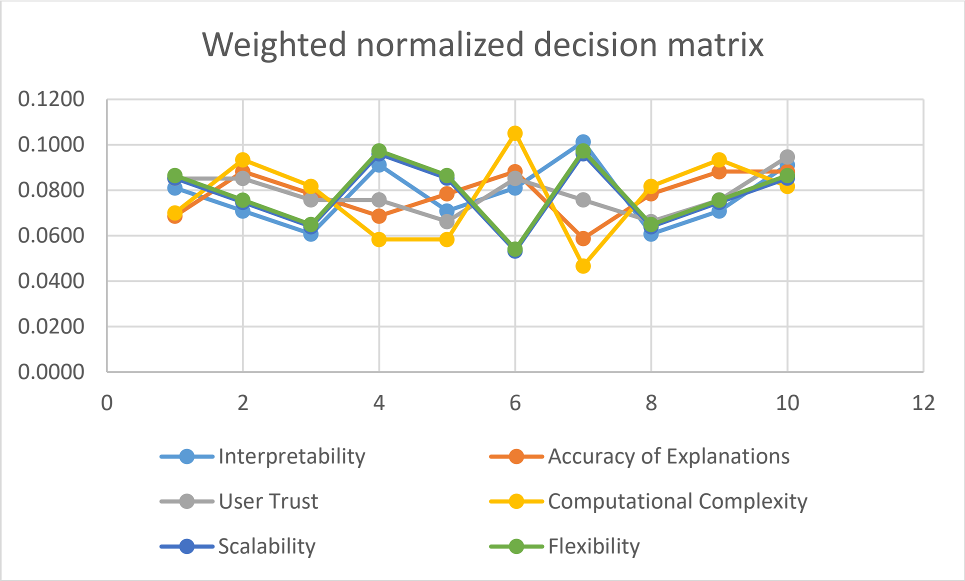


FIGURE 3. Weighted normalized decision matrix

The chart shown is a "Weighted Normalized Decision Matrix," likely used in decision-making processes to evaluate multiple alternatives across various criteria. The matrix compares the performance of options (represented by the X-axis) across five criteria: Interpretability (blue), User Trust (grey), Scalability (blue), Accuracy of Explanations (orange), Computational Complexity (yellow), and Flexibility (green). Each line represents how a particular criterion is weighted or normalized for each option. The Y-axis shows the normalized values, indicating the relative importance or performance of each criterion. This helps in visualizing trade-offs and making informed decisions based on the various factors.

TABLE 6. Positive Matrix

Positive Matrix					
0.1013	0.0883	0.0947	0.1051	0.0960	0.0974
0.1013	0.0883	0.0947	0.1051	0.0960	0.0974
0.1013	0.0883	0.0947	0.1051	0.0960	0.0974
0.1013	0.0883	0.0947	0.1051	0.0960	0.0974
0.1013	0.0883	0.0947	0.1051	0.0960	0.0974
0.1013	0.0883	0.0947	0.1051	0.0960	0.0974
0.1013	0.0883	0.0947	0.1051	0.0960	0.0974
0.1013	0.0883	0.0947	0.1051	0.0960	0.0974
0.1013	0.0883	0.0947	0.1051	0.0960	0.0974
0.1013	0.0883	0.0947	0.1051	0.0960	0.0974

The table 6 represents a "Positive Matrix," where each entry appears to be the same across all rows. This matrix likely indicates the best (or most favorable) values for each decision criterion—Interpretability, Accuracy of Explanations, User Trust, Computational Complexity, Scalability, and Flexibility—across the various alternatives being considered. The consistent values suggest that these factors are regarded as optimal or ideal, serving as a benchmark for comparing other alternatives. In decision analysis, the positive matrix helps to identify how far each alternative is from the best-case scenario, aiding in the evaluation process by comparison with the ideal outcomes.

TABLE 7. Negative matrix

Negative matrix					
0.0608	0.0588	0.0947	0.1051	0.0960	0.0974

0.0608	0.0588	0.0947	0.1051	0.0960	0.0974
0.0608	0.0588	0.0947	0.1051	0.0960	0.0974
0.0608	0.0588	0.0947	0.1051	0.0960	0.0974
0.0608	0.0588	0.0947	0.1051	0.0960	0.0974
0.0608	0.0588	0.0947	0.1051	0.0960	0.0974
0.0608	0.0588	0.0947	0.1051	0.0960	0.0974
0.0608	0.0588	0.0947	0.1051	0.0960	0.0974
0.0608	0.0588	0.0947	0.1051	0.0960	0.0974
0.0608	0.0588	0.0947	0.1051	0.0960	0.0974

The table 7 represents a "Negative Matrix," where each value is identical across all rows, similar to the Positive Matrix. In decision analysis, the negative matrix typically reflects the worst (or least favorable) values for each decision criterion—Interpretability, Accuracy of Explanations, User Trust, Computational Complexity, Scalability, and Flexibility—across different alternatives. These values are used as a benchmark to assess how far each alternative is from the worst-case scenario. The consistent values in this matrix suggest that these factors are considered suboptimal, helping to compare the alternatives by evaluating their distance from the worst outcomes and guiding the decision-making process.

TABLE 8. Si Positive & Si Negative & Ci & Rank

	Si Positive	**Si Negative**	**Ci**	**Rank**
LIME (Local Interpretable Model)	0.0484	0.1891	0.7962	2
SHAP (Shapley Additive Explanations)	0.0455	0.1778	0.7961	3
Deep LIFT (Deep Learning Important Features)	0.0687	0.1693	0.7113	9
Anchor Explanations	0.0550	0.2037	0.7873	4
ICE (Individual Conditional Expectation)	0.0651	0.1934	0.7481	7
Counterfactual Explanations	0.0648	0.1616	0.7139	8
Rule-based Explanation Systems	0.0681	0.2082	0.7535	6
Saliency Maps (for CNNs)	0.0719	0.1706	0.7035	10
Integrated Gradients	0.0484	0.1786	0.7867	5
XAI for Healthcare	0.0297	0.1905	0.8653	1

The table presents an analysis of different explanation methods used in machine learning, particularly those used in the context of explainable artificial intelligence (XAI). The columns represent three key components of each method: Si Positive, Si Negative, and Ci, followed by a Rank indicating the method's overall performance.

Si Positive and Si Negative: Si Positive represents the positive ideal performance value for each method. It is the value that shows how close the explanation method is to the best possible performance according to the selected criteria.

Si Negative represents the negative ideal performance value, reflecting how close the method is to the worst possible outcome. In decision-making, these values help assess each method's effectiveness relative to the ideal and worst-case scenarios. A higher Si Positive value means the method performs better in terms of achieving the ideal goal, while a lower Si Negative value indicates that the method stays closer to the best-case scenario.

Ci (Composite Score): The Ci value is the composite score, which is a weighted evaluation combining the Si Positive and Si Negative values. This score provides a balanced view of a method's overall performance. A higher Ci value means that the method is closer to the ideal solution, taking both the positive and negative aspects into account. In this case, XAI for Healthcare has the highest Ci value of 0.8653, indicating it is considered the most effective among the listed methods based on the criteria used for evaluation.

Rank: The Rank column shows the final rank of each method based on its composite score (Ci). The rank is determined by comparing the Ci values across all methods, with the method having the highest composite score ranked first and the lowest ranked last.

Interpretation of Results: XAI for Healthcare ranks first with a Ci of 0.8653, indicating that it performs the best overall across the methods compared. This suggests that the explanation approach used in healthcare is highly valued, possibly due to its strong interpretability, user trust, and relevance in a critical domain like healthcare.

LIME (Local Interpretable Model), SHAP (Shapley Additive Explanations), and Integrated Gradients follow closely with high Ci values of 0.7962, 0.7961, and 0.7867, respectively, placing them in the top ranks. These methods are well-known in the XAI community for providing robust explanations, with LIME and SHAP being popular for their interpretability and adaptability in different machine learning models.

Deep LIFT ranks 9th with a Ci of 0.7113, which suggests that it is less effective than other methods, although it still offers valuable insights, particularly in deep learning contexts.

Methods like Anchor Explanations, ICE (Individual Conditional Expectation), Counterfactual Explanations, and Rule-based Explanation Systems rank in the middle (4th to 7th), showing that while these methods are effective, they may not be as optimal as the top-ranking methods like LIME and SHAP.

Saliency Maps (for CNNs) ranks lowest at 10th with a Ci of 0.7035. This indicates that, while Saliency Maps are useful in some contexts (especially convolutional neural networks), they may not be as broadly effective or reliable in providing comprehensive explanations compared to other methods.

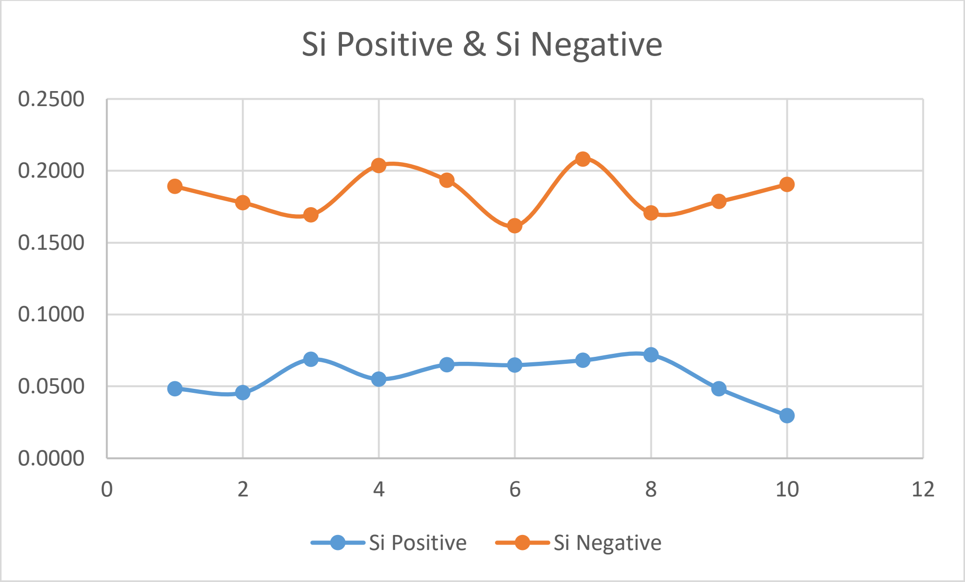


FIGURE 4. Si Positive & Si Negative

Figure 4 shows the trends of Si Positive and Si Negative values across 10 data points. The Si Positive series is represented by a blue line and generally fluctuates within the range of 0.03 to 0.07. The trend shows slight increases and decreases but remains relatively low and stable throughout the dataset. The Si Negative series, indicated by the orange line, is consistently higher, ranging from approximately 0.16 to 0.21. This trend shows moderate fluctuations, with noticeable peaks around points 4 and 7 and dips around points 5 and 6. The figure highlights that Si Negative values are consistently higher than Si Positive values.

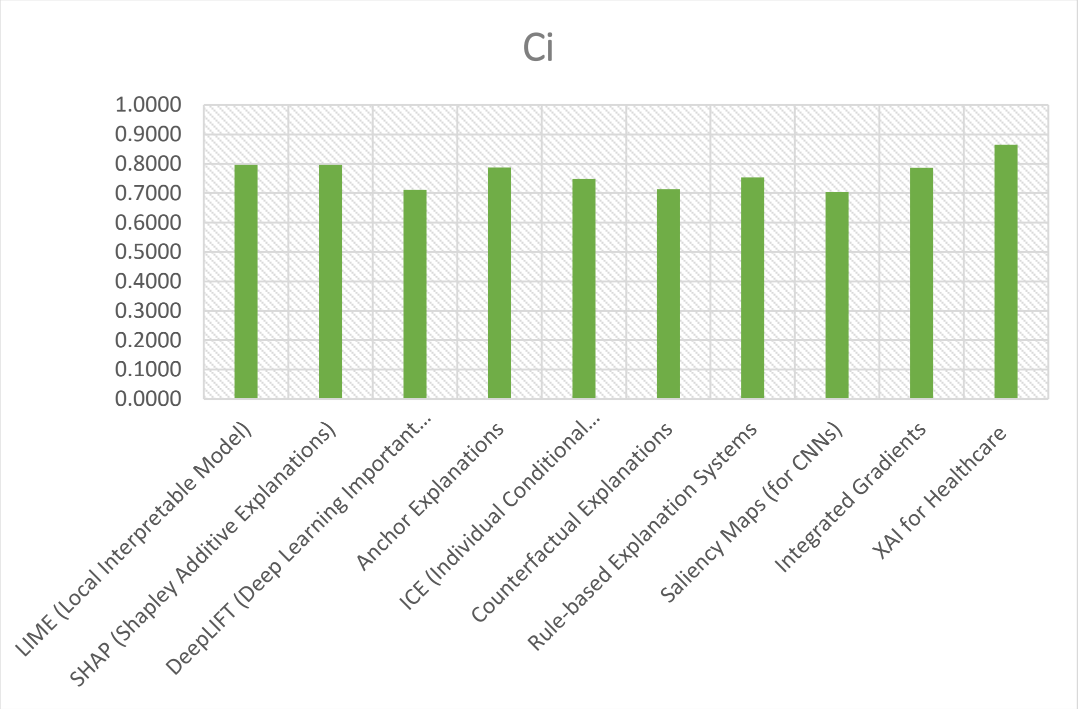


FIGURE 5. Ci

Figure 5 presents a bar chart illustrating Ci values for various explainable AI (XAI) techniques. Each bar represents a method, such as LIME, SHAP, DeepLIFT, and others, with their corresponding Ci scores on the y-axis, ranging from 0 to 1. Most methods achieve Ci values between 0.7 and 0.9, indicating high consistency or effectiveness. XAI for Healthcare shows the highest value, near 0.9, while DeepLIFT and Rule-based Explanation Systems have relatively lower scores. The results suggest that these methods perform similarly in their ability to provide consistent explanations, though slight variations exist depending on the method applied.

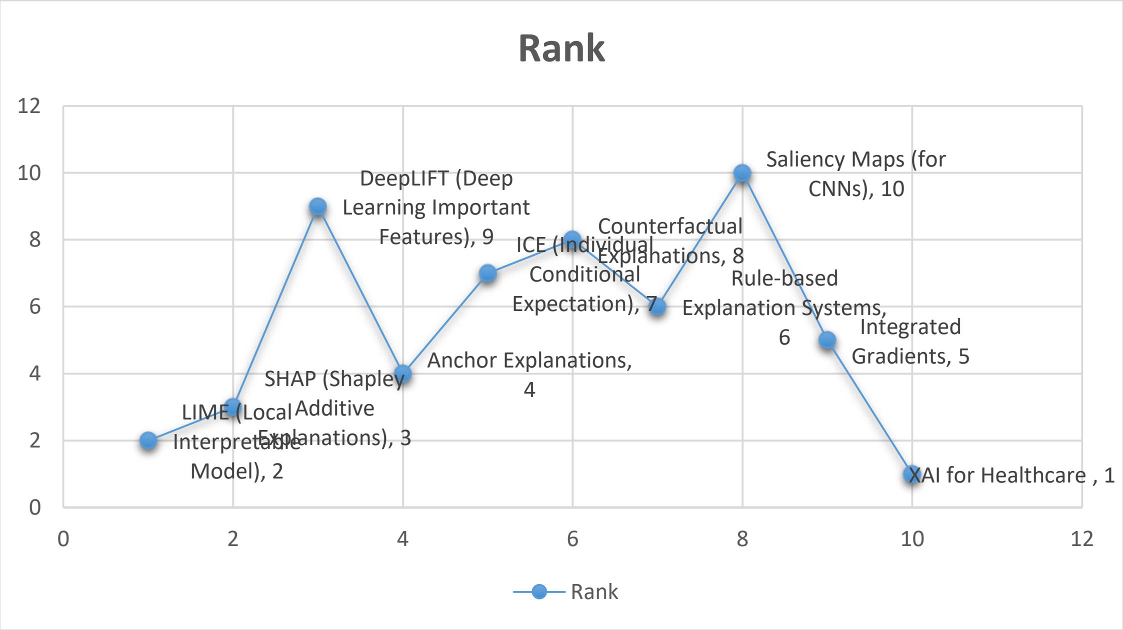


FIGURE 6. Rank

Figure 6 displays the ranking of various explainable AI (XAI) techniques. The x-axis represents different methods, while the y-axis indicates their corresponding ranks, ranging from 1 (highest) to 10 (lowest). XAI for Healthcare ranks 1st, indicating top performance, followed by LIME (2nd) and SHAP (3rd). Methods such as DeepLIFT (9th) and Saliency Maps (10th) are ranked lower. The chart shows notable fluctuations, with techniques like Anchor Explanations (4th) and Integrated Gradients (5th) performing moderately well. These rankings reflect the comparative effectiveness or applicability of these methods in the given context.

4. CONCLUSION

In artificial intelligence (AI), Explainable is a crucial and developing component that aims to increase the understandability, transparency, and accountability of AI systems. The increasing integration of AI into a growing number of industries, including healthcare, banking, and autonomous systems, has resulted in a sharp rise in demand for explainable AI (XAI). AI systems are sometimes viewed as "black boxes" because of their inability to be understood, particularly those that are built on intricate models like deep learning. One of the difficulties caused by this opacity is the inability to comprehend the decision-making process, which can erode confidence, impede adoption, and result in unfair or biassed conclusions. Explainable's main objective is to close the gap between the human stakeholders who must trust and communicate with AI models and their complex, high-performance capabilities. Transparent models allow for a better understanding of decision-making processes, which can help in identifying errors, improving the model, and ensuring its ethical use. For example, in the medical field, explainable AI can assist doctors by providing insights into how a diagnosis was made, supporting better clinical decisions. Similarly, in finance, Explainable can make automated credit scoring systems more reliable and transparent, enhancing fairness. One of the critical advantages of explainable AI is that it facilitates accountability. When the reasoning behind AI decisions is clear, It gets simpler to recognise and deal with biases, inconsistencies, or ethical concerns that might arise. For instance, if a self-driving car makes a poor decision, understanding the model’s reasoning can lead to better safeguards and corrective actions. Explainable is also essential for compliance with legal and regulatory standards. As AI becomes more embedded in society, governments and organizations are increasingly focusing on policies that demand transparency in AI systems. However, finding a balance between performance and Explainable is still a major obstacle. Deep neural networks and other high-performing models are frequently complicated and not always explicable. Simplified models may be more interpretable but often lack the accuracy of their more complex counterparts. Researchers are actively exploring methods to provide Explainable without compromising model performance, such as post-hoc explanations, surrogate models, and hybrid approaches.

REFERENCES

[1]. Confalonieri, Roberto, Ludovik Coba, Benedikt Wagner, and Tarek R. Besold. "A historical perspective of explainable Artificial Intelligence." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 11, no. 1 (2021): e1391.

[2]. Angelov, Plamen P., Eduardo A. Soares, Richard Jiang, Nicholas I. Arnold, and Peter M. Atkinson. "Explainable artificial intelligence: an analytical review." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 11, no. 5 (2021): e1424.

[3]. Došilović, Filip Karlo, Mario Brčić, and Nikica Hlupić. "Explainable artificial intelligence: A survey." In *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)*, pp. 0210-0215. IEEE, 2018.

[4]. Vilone, Giulia, and Luca Longo. "Notions of Explainable and evaluation approaches for explainable artificial intelligence." *Information Fusion* 76 (2021): 89-106.

[5]. Adadi, Amina, and Mohammed Berrada. "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)." *IEEE access* 6 (2018): 52138-52160.

[6]. Tjoa, Erico, and Cuntai Guan. "A survey on explainable artificial intelligence (xai): Toward medical xai." *IEEE transactions on neural networks and learning systems* 32, no. 11 (2020): 4793-4813.

[7]. Arrieta, Alejandro Barredo, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García et al. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." *Information fusion* 58 (2020): 82-115.

[8]. Jiménez-Luna, José, Francesca Grisoni, and Gisbert Schneider. "Drug discovery with explainable artificial intelligence." *Nature Machine Intelligence* 2, no. 10 (2020): 573-584.

[9]. Speith, Timo. "A review of taxonomies of explainable artificial intelligence (XAI) methods." In *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*, pp. 2239-2250. 2022.

[10]. Yang, Christopher C. "Explainable artificial intelligence for predictive modeling in healthcare." *Journal of healthcare informatics research* 6, no. 2 (2022): 228-239.

[11]. Meske, Christian, Enrico Bunde, Johannes Schneider, and Martin Gersch. "Explainable artificial intelligence: objectives, stakeholders, and future research opportunities." *Information Systems Management* 39, no. 1 (2022): 53-63.

[12]. Vassiliades, Alexandros, Nick Bassiliades, and Theodore Patkos. "Argumentation and explainable artificial intelligence: a survey." *The Knowledge Engineering Review* 36 (2021): e5.

[13]. Chamola, Vinay, Vikas Hassija, A. Razia Sulthana, Debshishu Ghosh, Divyansh Dhingra, and Biplab Sikdar. "A review of trustworthy and explainable artificial intelligence (xai)." *IEEe Access* (2023).

[14]. Samek, W. "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models." *arXiv preprint arXiv:1708.08296* (2017).

[15]. Islam, Mir Riyanul, Mobyen Uddin Ahmed, Shaibal Barua, and Shahina Begum. "A systematic review of explainable artificial intelligence in terms of different application domains and tasks." *Applied Sciences* 12, no. 3 (2022): 1353.

[16]. Shih, Hsu-Shih, Huan-Jyh Shyur, and E. Stanley Lee. "An extension of TOPSIS for group decision making." Mathematical and computer modelling 45, no. 7-8 (2007): 801-813.

[17]. Krohling, Renato A., and André GC Pacheco. "A-TOPSIS–an approach based on TOPSIS for ranking evolutionary algorithms." Procedia Computer Science 55 (2015): 308-317.

[18]. Jahanshahloo, Gholam Reza, F. Hosseinzadeh Lotfi, and Mohammad Izadikhah. "Extension of the TOPSIS method for decision-making problems with fuzzy data." Applied mathematics and computation 181, no. 2 (2006): 1544-1551.

[19]. Chen, Pengyu. "Effects of the entropy weight on TOPSIS." Expert Systems with Applications 168 (2021): 114186.

[20]. Kuo, Ting. "A modified TOPSIS with a different ranking index." European journal of operational research 260, no. 1 (2017): 152-160.

[21]. Lin, Ming-Chyuan, Chen-Cheng Wang, Ming-Shi Chen, and C. Alec Chang. "Using AHP and TOPSIS approaches in customer-driven product design process." Computers in industry 59, no. 1 (2008): 17-31.

[22]. Chen, Chen-Tung. "Extensions of the TOPSIS for group decision-making under fuzzy environment." Fuzzy sets and systems 114, no. 1 (2000): 1-9.