International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Revue d'Intelligence Artificielle
Vol. 38, No. 3, June, 2024, pp. 957-968
Journal homepage: http://iieta.org/journals/ria

An Extensive Review on Significance of Explainable Artificial Intelligence Models in Discrete Domains for Informed Decisions Making

Renuka Agrawal*, Kanhaiya Sharma

Department of Computer Science and Engineering, Symbiosis Institute of Technology, Symbiosis International (Deemed University) (SIU), Pune 412115, India

Corresponding Author Email: renuka.agrawal@sitpune.edu.in

https://doi.org/10.18280/ria.380321

Received: 14 September 2023
Revised: 15 February 2024
Accepted: 21 March 2024
Available online: 21 June 2024

Keywords:
machine learning, artificial intelligence, Explainable Artificial Intelligence, feature visualization, contextual explanatory model

ABSTRACT

Machine learning approaches and models subset of Artificial Intelligence are becoming increasingly complex and incomprehensible. While domain specialists grasp the mathematical theory, they have difficulty communicating the reasoning to a broad audience. To solve this challenge, a new research topic called Explainable Artificial Intelligence (XAI) has arisen to develop a contextual explanatory model for practical deployment. Explainable Artificial Intelligence aims to make AI models interpretable and transparent by providing human-understandable explanations for their decisions. Techniques like feature visualization and attribution methods offer insights into AI decision-making, benefiting healthcare, finance, and autonomous vehicles applications. XAI enhances trust, accountability, and fairness in AI systems by allowing users to comprehend the reasoning behind predictions. However, striking a balance between interpretability and performance is challenging. Achieving this balance is crucial to leverage the potential of XAI in building trustworthy and ethical AI across various domains. This essential taxonomy describes the prospects and problems in the field of XAI and serves as a resource for future AI researchers. The study authors examined XAI's role in simplifying machine learning models, providing understandable explanations for AI decisions, promoting trust and accountability, and optimizing performance across diverse applications of military, healthcare and communications.

1. INTRODUCTION

Machine learning has revolutionized various aspects of our lives, and its applications extend prominently into the different realms of prediction or suggestions as required. These applications influence the power of machine learning to enhance user experiences, improve personalization, and optimize engagement. Widespread usage of AI tools signifies its ready acceptance but the Black Box problem of AI is a major challenge encountered by developers of Machine Learning Algorithms whenever an explanation is needed for the response generated by these models. In certain cases, like in Medical Domain such an explanation is necessary as per the predicted outcome a decision can be made for treatment of disease detected. Not only in medical but in different other domains, the clients or customers making use of the customized AI models demand an explanation of the outcome generated. Explainable Artificial Intelligence (XAI) which overcomes the major challenge of black box nature of AI models has been gaining attention. AI black box problem is the difficulty of interpreting the cognitive behind an AI system's outcomes or decisions [1, 2]. AI powered algorithms generate specific outcomes for the input being fed to it without providing any justified explanations. At times, it becomes difficult to provide a satisfactory explanation to the end user for the reasons behind the reasoning. In order to ensure Trust and Confidence in AI Systems, it is essential to be able to clarify and demonstrate to its users why a specific estimate is being done by the Machine Learning Tool. Explainable artificial intelligence (XAI) is a set of procedures and approaches that allows human users to understand and believe the outcomes generated by machine learning systems [3, 4]. Machine learning models are proving to be remarkably accurate on a variety of tasks and are now widely used in discrete domains [5]. The individuals who deal with these models, however, find that many of them are difficult to understand. This comprehension, also known as "explain ability" or "interpretability," enables users to comprehend how the computer decides to do a certain action [6, 7]. In the AI community explain ability is a gaining popularity from last one decade. With so much hype, AI is receiving greater scrutiny and criticism about its lack of transparency. How do these systems reach such important decisions? Regulators, official bodies, and users are seeking more transparency in every AI-based decision [8]. The more powerful the deep learning system, the opaquer it becomes. When algorithms make incorrect assumptions, the opaqueness makes it more difficult to determine what went wrong [9, 10]. Artificial Intelligence grounded systems are attaining a lead part in healthcare systems. Nevertheless, the black-box nature of these systems like neural networks challenges the users' reliance which is a major and basic requirement in the medical domain [11, 12]. In order to increase user understanding of AI predictions and judgements and increase system trustworthiness and dependability, XAI aims to create more transparent and interpretable AI. To be trusted, AI must not only classify objects correctly but also explain the logic behind its classification. The paper presents an overview of need of XAI along with its characteristic features. It also includes applications of XAI in different domains. The primary focus of this study is to assess how XAI can alleviate the Black Box problem in AI systems, especially in vital sectors such as healthcare, by offering clear explanations for machine learning model decisions, and subsequently Examine the impact of XAI on enhancing user understanding, trust, and confidence in AI predictions and judgments across diverse domains, aiming to increase system reliability and dependability through transparent and interpretable AI.

The paper is organized as follows: Section 1 discusses the importance and relevance of Explainable AI (XAI) in the contemporary landscape. Section 2 delves into the features and functionality of XAI, examining four fundamental aspects. Section 3 explores XAI applications across eight prominent domains. Lastly, Section 4 provides a concluding summary of the paper's findings.

2. XAI FEATURES AND WORKING

The purpose of XAI is to convey pragmatic reasoning to an end user who relies on an AI system's decisions or forecasts. Understanding the system's reasoning, or the reasons for a certain conclusion or recommendation, is thus required. For instance, an intelligence analyst upon receiving recommendations from a big data analytics system must understand the reasoning behind that specific recommendation to be confident of results. Similarly, to use an autonomous car effectively in the future, an operator who supervises it on a route must understand and intercept the system's decision-making procedure.

2.1 Features of XAI

XAI users are provided justifications with explanations that assist them comprehend the system's overall capabilities and limits, as well as a feel of how it will behave in practice.AI-based solutions are used by businesses to increase performance and make more sensible decisions. However, understanding how these tools function is just as important as using their output. The inability to explain something stops dealings from taking logical "what-if" scenarios and creating trust issues since the dealings or corporates are not able to provide justification behind the outcome. Prime principles describing the working of XAI are outlined as:

Expandability: All system outputs come with supporting data or an explanation. The Explanation principle necessitates that XAI arrangements provide evidence, argument, or explanation for every result. This theory asserts that a system must give an explanation for a certain event rather than that the evidence is right, instructive, or explanatory in and of itself.

Understanding: is the model's capacity to effectively convey to a human learner its most recent study and its understanding of creating the responses or answer. To comprehend the internal working of the model, the traits and behavioral elements of each individual component are examined. This concept is completely satisfied if the user comprehends the explanation and/or applies this information in executing a task or making a choice. A system may require various explanations for different types of users depending on the usage for which it is designed.

Fidelity and Accuracy: defines a model's reliability and correctness features by assessing the selected model internal performance activities. In other words, it brings transparency by illustrating the operational procedure of each system module. While Fidelity ensures dependability on model's outcomes, accuracy principle necessitates the explanations offered by a XAI system to be precise and accurate. Explanation is not the same as decision accuracy. In decision tasks, decision correctness refers to whether or not the system's judgment is right.

Transparency: If a user of model can have a clear idea of logic behind its working, then that model can be called as a transparent one. A model in XAI will be classified as a transparent model only if its internal working is known to its users so that they can train it as per their requirements whenever they wish to have a different set of results.

Adaptability: This characteristic specifies that the proposed model must be capable of responding accurately to new data set as well. When modest changes are required to meet system requirements, the deployed model must perform accurately and efficiently.

Knowledge Limitation: The system will only work in the conditions under which it was created, or if the system's output is sufficiently dependable. The system works within his knowledge. Based on this knowledge limit concept, the system detects conditions where its behavior is not designed or approved, or where its response is unreliable. The knowledge limitation principle can increase the reliability of the system by eliminating misleading, harmful or unfair decisions.

2.2 AI and explainable AI

The fundamental distinction between AI and explainable AI can be explained. XAI is a subset of AI that provides explanations and rationale behind its decisions and outcomes. XAI is concerned with constructing AI systems that are transparent, interpretable, and explainable to people, whereas AI is concerned with designing models that can accomplish tasks that need human intelligence. XAI refers to algorithms' ability to explain why they produce specific results. XAI employs simpler models that are easily understood by people, as opposed to standard AI, which use complicated algorithms to spot patterns and make choices. Explainable AI approaches are highly affected by how people make inferences and draw conclusions, allowing them to be duplicated within an explainable artificial intelligence system.

Figure 1 shows an Explainable Artificial Intelligence system that blends semantic technologies with deep learning models schematically. The blue color represents the standard AI system pipeline signifying Black Box nature of AI. The orange color represents the Knowledge Matching process of deep learning components with Knowledge Graphs (KGs) and ontologies. The red tint represents cross-disciplinary and interactive explanations facilitated by inquiry and reasoning techniques [13]. To provide a satisfactory and logical explanation of outcome is the prime objective that leads to development of XAI. The methods employed by XAI to provide explanations are, Decomposition, Visualization and

Mining for Explanations. XAI brings in features such as Transparency (expressing how the system arrived at a specific outcome), Justification (clarifying why the response delivered by the model is satisfactory), In formativeness, and Uncertainty estimations shown in Figure 2. The figure shows how explainability is achieved in XAI models. Several techniques have been developed in the field of explainable AI (XAI), mostly grouped according to three factors: complexity, scope, and model reliance. Complexity-related methods support algorithms that are naturally interpretable, including sparse linear models and Bayesian Rule Lists. But a significant obstacle is the intrinsic trade-off between interpretability and accuracy. A different paradigm includes natural language explanations and visuals for complex black-box models that are provided post hoc. The choice of either intrinsic interpretable models or post-hoc techniques depends on the particulars of the prediction task. Methodologies connected to scoop distinguish between interpretability on a global and local level. In order to make decisions at the population level, global interpretability aims to understand a model's overall logic. On the other hand, local interpretability focuses on explaining particular choices or forecasts, and is frequently enabled by techniques such as Local Interpretable Model-Agnostic Explanations (LIME). Model-specific and model-agnostic approaches are further distinguished by the model-related classification. Model-agnostic techniques, on the other hand, provide for flexibility across a variety of ML models, whereas model-specific methods are limited to specific model classes. Examples of model-agnostic techniques include visualization, knowledge extraction, influence methods, and example-based explanations.

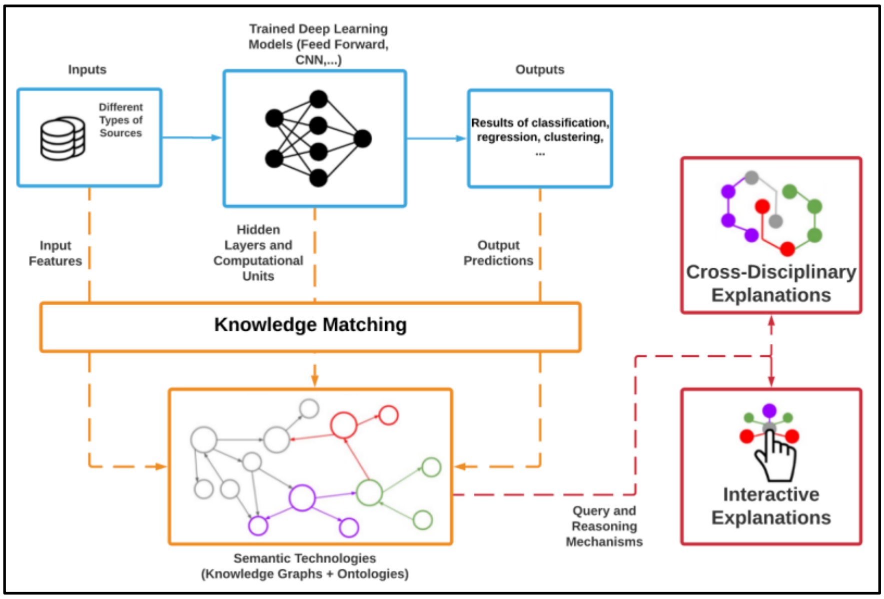

Trained Deep Learning Models (Feed Forward, CNN,...)
Inputs
Outputs
Different Types of Sources
Results of classification, regression, clustering, ...
Input Features
Hidden Layers and Computational Units
Output Predictions
Cross-Disciplinary Explanations
Knowledge Matching
Interactive Explanations
Query and Reasoning Mechanisms
Semantic Technologies (Knowledge Graphs + Ontologies)

Figure 1. AI (in blue color) and XAI system [13]

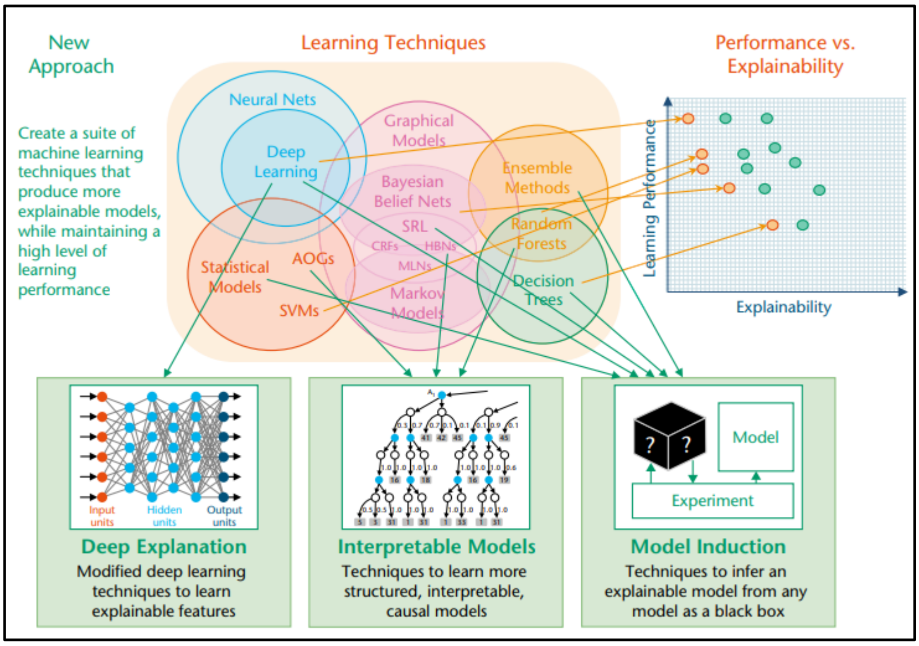

New Approach
Learning Techniques
Performance vs. Explainability
Create a suite of machine learning techniques that produce more explainable models, while maintaining a high level of learning performance
Neural Nets
Deep Learning
Graphical Models
Bayesian Belief Nets
SRL
CRFs
HBNs
MLNs
Markov Models
Ensemble Methods
Random Forests
Decision Trees
Statistical Models
AOGs
SVMs
Learning Performance
Explainability
Input units
Hidden units
Output units
Deep Explanation
Modified deep learning techniques to learn explainable features
Interpretable Models
Techniques to learn more structured, interpretable, causal models
Model
Experiment
Model Induction
Techniques to infer an explainable model from any model as a black box

Figure 2. Developing an XAI model [14]

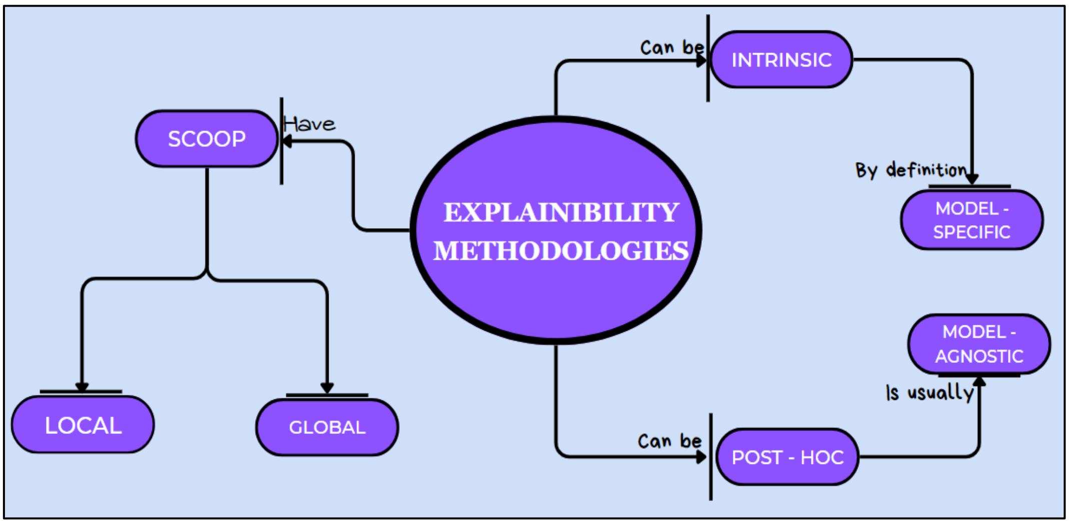


Figure 3. Techniques of explainability in XAI models

Visualization approaches like individual conditional expectation (ICE), partial dependency plots (PDP), and surrogate models are among the model-agnostic techniques that make complex models easier to understand. Rule extraction and model distillation are two aspects of knowledge extraction that are concerned with making internal representations understandable. Influence techniques, such as Layer-wise Relevance Propagation (LRP) and sensitivity analysis, measure the significance of features by varying inputs. Prototypes and critiques, or counterfactual explanations, are used in example-based explanations when choosing which examples to interpret. Although model-agnostic approaches are flexible, it is important to recognize that they may rely on approximations, which emphasizes the ongoing difficulty of balancing accuracy and interpretability in AI models.

Figure 3 covers in details how expandability is achieved in XAI models, which can be model specific or model agnostic. Further for ease the explainability strategies are further classified as complexity or intrinsic related methods, scoop related methods and model related methods.

3. APPLICATIONS OF XAI

Explainability is a new property that started to gain popularity in the AI community. Rather than just augmenting human judgment, AI-based systems are now making decisions on their own. Banks rely on this technology to grant loans. Transfer learning-based AI can even detect cancer autonomously. The persons affected with decisions taken on the basis of AI tools needs an explanation for outcome. XAI that reasons the outcome and provides an explanation for the outcome are found to outperform other peer technologies finds applications in multiple domains cloud, healthcare, social media, IoT, Military services, Service sector, communication, and manufacturing.

3.1 XAI in military services

Tabular representation of inclusion of XAI in military services is taken into consideration in done in Table 1. The developments and findings by different researchers are also included in this table.

Table 1. XAI in military services

Ref.#	Domain	Major Findings
[14]	Military	The XAI platform by DARPA is dedicated to developing and accessing a diverse array of new machine learning techniques, such as model induction for creating explainable models and adapting deep learning methods for interpretability.
[15]	Explainable Security	Applying XAI in cyber security models is crucial for creating explainable yet accurate defenses, enabling user comprehension, trust, and effective management.
[16]	Military Defense Services	This article reviews the objectives, organization, and research developments of the XAI program, it has provided a more accurate understanding of XAI. The program definitely acted as a promoter to motivate further researches in areas of XAI.
[17]	Cyber Security	A systematic review of Applications of XAI in cyber security is provided. Significance and need of XAI in defense and military is also impressed.
[18]	Military	The purpose of examining military cyber aspects of XAI is to outline and analyze critical elements and approaches relevant to the design and implementation of XAI models for targeting military cyber operations.
[19]	Security	A method for developing an XAI-based junior cyber analyst concept based on understanding the information requirements of both human and AI components is proposed. This is common for analysts in the military, where they are expected to explain why they make such predictions or recommendations.
[20]	Army	At the core of the Multi-Domain Operations (MDO) concept is the use of Intelligence, Surveillance and Reconnaissance (ISR) networks consisting of redundant remote and independent sensor systems and human intelligence shared between multiple partners. Achieving this concept requires the development of both Artificial Intelligence (AI) to improve distributed data analysis and Augmented Intelligence (IA) to improve human-machine perception.

Explainable AI has been utilized in military training to clarify the explanations that led an AI model chose a particular option. This is vital as it provides an explanation such as the reason why an object passed undetected or is not identified or failed to fire on a target [14]. Smart AI drones are capable of gathering much more information from enemy terrain with no human intervention. Military services are quite capable of utilizing and exploring these benefits, but they also intend to be aware of the justification of outcomes. The Defense Advanced Research Projects Agency (DARPA) launched the Explainable Artificial Intelligence (XAI) program in 2017, with the goal of developing a suite of new AI techniques that will allow end users to understand, trust, and effectively manage the next generation of AI systems [15-18]. The top Applications of AI in Military includes Cybersecurity, Logistics and Transport, Target recognition, Warfare healthcare, Monitoring of threats and safety of military personnel, combat simulation and training. An inclusion of explainability in all these applications by XAI will make them more trustworthy and interpretable.

3.2 XAI in healthcare

In health care and in medicines, XAI has wide applications. Patients often are eager to know the cause of anomaly, so that they can take measures to ensure that won't be able to catch the disease again. Besides this people also wish to know the reason behind a particular prescription. XAI can provide them both. Latest study on medical XAI emphases entirely on interpretability [20]. Greater interpretability is required due to the high level of responsibility and also for the need of transparency in the medical field.

Incorporating explanations from the medical field into legal and ethical AI is required to understand and interpret detailed decisions, outcomes and the current state of a person's illnesses [21]. Thus, medical XAI is very important to help implement AI in clinical decision support organizations. Medical professionals work more often with individual patients and need specific explanations adapted to each patient's situation to assess how XAI results apply or do not apply to individual patients [22]. When using AI-based systems, it is imperative that clinicians use the tools in a way that enhances the best possible outcome for patients. However, in order to provide patients with the most appropriate opportunities to promote their health and well-being, physicians must be able to take full advantage of the system's capabilities. To make reliable clinical decisions, doctors need explainability in the form of visual aids or natural language explanations, rather than relying solely on automated results. This allows them to modify estimates and approve appropriate course of treatment according to individual medical conditions as required [23]. Table 2 covers a description of work done by researchers in the field of XAI in medical domain.

Applications of AI with XAI have great promise for enhancing explainability and transparency in high-risk sectors like health care, where confidence is crucial. There is still room for improvement by exploring other deep learning models and variants of explainable artificial intelligence (XAI). Experts who can mediate between the worlds of informatics and medicine will become more and more in demand when using ML systems due to the more complexity of this sector.

Table 2. XAI in healthcare sector

Reference No.	Domain	Dataset	XAI Tech Used	Outcome and Limitations
[23]	Medical	8690 wound images from the eKare Inc. data repository	DNN, Transfer Learning LIME (Model Agnostic)	Four types of wounds are predicted by DNN using the transfer learning technique: diabetic, lymphovascular, pressure injury, and surgical. The model takes an image as input and outputs a prediction about the cause of a chronic wound. The classification's performance will be further improved by gathering more data. A thorough understanding of the selected wound type can be achieved by interpreting the XAI module's results.
[24]	Medical (Glaucoma)	Two different publicly available datasets, fundus images, coherence tomography images of the eyes and clinical medical records of glaucoma patients	Adaptive neuro-fuzzy inference system (ANFIS) and Sub-modular pick (SP-LIME)	The adoption of interpretable machine learning (IML) and explainable artificial intelligence (XAI) offers opportunities to bolster user confidence in decision-making processes. IML achieves coherent explanations of outcomes utilizing sub-modular pick local interpretable model-agnostic explanations (SP-LIME). SP-LIME is employed to interpret findings from Spike Neural Network (SNN). Results indicate that XAI and IML models provide patients and medical professionals with logical and compelling decisions.
[25]	Medical (Diabetic Retiniopathy)	21,576 CFPs from 9,734 eyes	ExplAIn	XAI effectively addresses a renowned classification issue, such as DR severity classification, by enabling the assessment of local patterns in DR lesions' significance. Its utility extends to novel categorization problems like disease progression prediction or diagnosing new diseases. However, a limitation arises in its applicability, as it's only suitable for binary or multilabel classification scenarios involving zero, one, or more labels per image rather than direct application to multiclass classification involving precisely one label per image.
[26]	Medical Images (Explainable Information Retrieval)	2621 images	CNN and LIME	The design and proposal for the categorization of WCE (wireless capsule endoscopy) frames into bleeding and non-bleeding categories involves the use of a convolutional neural network (CNN). Based on evaluation parameters, the performance of the suggested model is contrasted with that of

				other conventional machine learning models. In order to help medical professionals make better and faster decisions when it comes to detecting gastroenterological bleeding, the authors have developed and prototyped an explainable machine learning tool.
[27]	Medical and Heathcare	454 articles	NA	By highlighting feature relevance and boosting system confidence, XAI algorithms like Shapley Additive explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) may explain ML models and enhance confidence in their predictions. Throughout most of the studies, LIME was the most discussed and utilized. also talked about LIME and its operation because it was being used a lot. The SHAP, CAM, and GradCAM followed LIME.
[28]	Medical	450 medical use cases	NA	Open-source and model-agnostic explanation methods, such as Shapley Additive explanations (SHAP) and Gradient Class Activation Mapping (Grad-CAM), are widely used in Explainable AI (XAI), especially for tabular and image data. ML pipelines have improved documentation, promoting clarity. XAI techniques are favored for their ease of implementation, while standardization of ML reports facilitates comparability and should be emphasized for future progress.
[29]	Medical (Skin Cancer)	2750 and 11527 images	Deep Neural Networks	The proposed model offers diagnostic insights and informative details regarding its decisions utilizing two dermoscopy datasets (ISIC 2017 and 2018), thereby enhancing safety. Additionally, the model asserts its capability to provide justifications for the outcomes derived from the dataset used for disease assessments.

3.3 XAI in service sectors

With the incorporation of technology and science, unbiased judicial outcomes can be assured so economic and social development of the nation can be made better. Commercial sectors as banks, that requires an in-depth understanding and information of their clients and products needs to be explainable upon demands. The sector is sure to attract more business if it includes customer satisfaction and transparency in its dealings. This can be achieved by adopting insights technology to support AI decision's [30] and also adopting to technologies that prevent misuse of finances from malicious users. The research [31] contributed to both Explainable AI (XAI) research community and practical business value to answer fraud detection business issues in the financial sector and banking services by applying five latest Explainable AI (XAI) approaches. The approach claims to improve model interpretability or transparency of present black-box fraud detection systems extensively used in Banking and Finance sector.

The advantages of applications of XAI in judicial system is many. They include improved competence and speed decision making besides justice delivery at a faster pace. Transparency of working leads to reduced error rates, a more consistent approach to judicial decision-making, and enhanced security and privacy [32]. By using a transparent and secure platform, XAI based system will create greater trust and confidence in the justice system in public. In multiple AI solicitation area, the result or outcome from a computerized predictive structure may definitely affect future course of action and the concerned authority needs to comprehend the reasoning behind the conclusion and also to assess the risks involved in it. Besides this, they also need to keep in mind the influence of the decision in the life of others [33, 34]. If we consider XAI based assisting autonomous insurance policy service providers, the persons insured need to know why, when, and how the policy they are planning to take will function before agreeing for it. It will be easier for the person going for insurance to make a decision if he is well aware of the pros and cons of it. Thus, a responsible XAI system becomes an imperative precondition for AI to be applied to any pragmatic realistic problem. Researches done impresses the need of explainability in AI for decision-making, elaborating the concepts needed for goal accomplishment, and future potentials for XAI based methods in Finance, insurance and banking sector [35-37].

3.4 XAI in communication sector

As we head towards 6G, communication networks are becoming more intricate. With this complexity, manual management by network operators is no longer a viable option. The networking community has extensively debated network automation as a practical solution to manage these complex communication networks [38, 39]. A comparative analysis of works done by different researchers involving XAI in communication sector is tabulated in Table 3.

Figure 4 shows the different domains where XAI can be utilized for secure and explainable end to end secured communication. Malicious agent's detection from AI Black box methods might leads to unpredictable results when these systems themselves undergo chances of attacks originating in network, application, encryption, physical, and transport layer attacks. Work done by different researchers and their findings as well as limitations is included here.

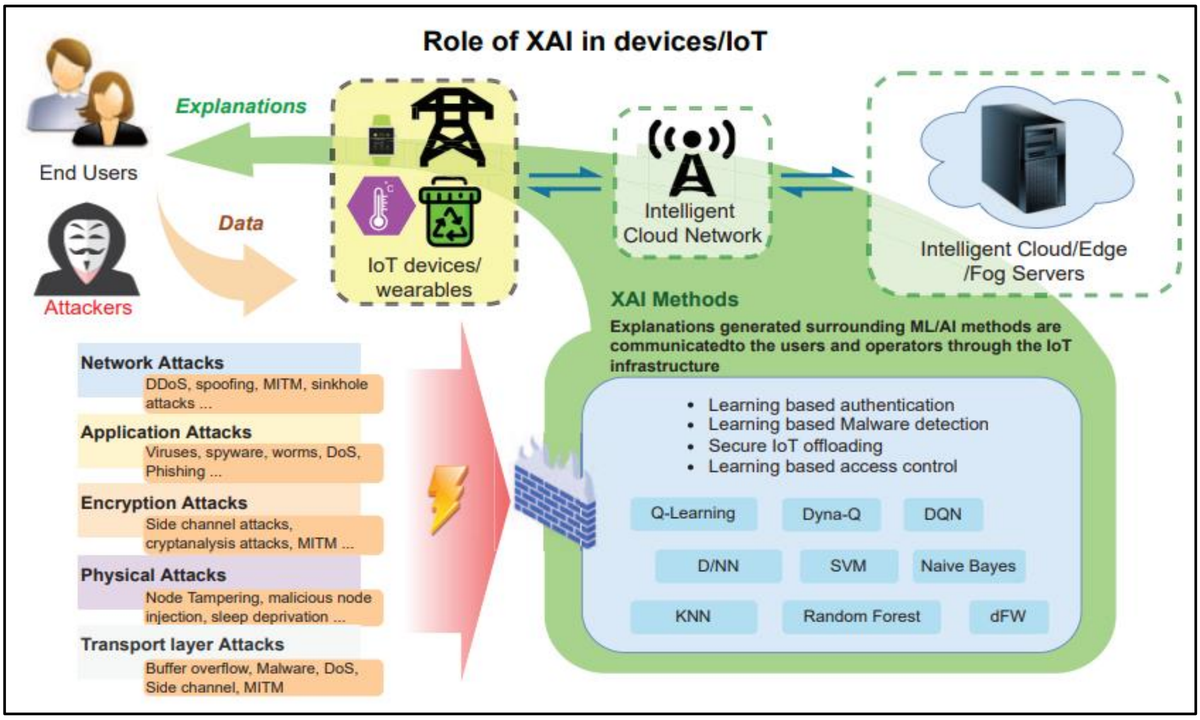
3.5 XAI in manufacturing industry

Understandable AI provides insights on the reasons for malfunctioning of a manufacturing unit and recommends necessary adjustments for improved performance over a period. This is of utmost importance for facilitating better communication and understanding between machines and humans, ultimately leading to increased situational awareness. The research [48] focuses on resolving the inadequacies of current XAI applications in 3D printing in ubiquitous computing through the introduction of four novel XAI approaches: (1) a gradient bar chart featuring a baseline, (2) a gradient bar chart for groups, (3) a gradient bar chart that can be manually adjusted, and (4) a scatterplot with bidirectional capabilities. To showcase its efficacy, the suggested methodology was employed in a case study. The findings from the bidirectional scatterplot experiment confirmed the appropriateness of the 3D printing facilities in terms of their closeness [49]. A technique for integrating XAI-derived findings into the Data Science process for building a highly accurate classifier. By employing Synthetic Minority Oversampling Technique (SMOTE) and medoid concepts, XAI tools such as Ceteris Paribus profiles, Partial Dependence, and Breakdown profiles have been utilized to obtain valuable insights [50, 51]. Inclusion of XAI in manufacturing sector leads to better performance of machines, quick fault detection and precise methods required for increased production. Complex nature of machinery involved in production and their location at critical positions as well as dynamic operating conditions are some of the features which can be taken care of by inclusion of XAI in manufacturing sector [52, 53]. When applied in cases like wind turbines [53], it can lead to overall system improved performance.

Table 3. Applications of XAI in communications

Reference No.	Domain	Dataset	XAI Tech Used	Outcome and Limitations
[40]	IoT and 5G Communications	207 Research Articles	NA	The study offers a thorough analysis of Explainable Artificial Intelligence (XAI) elucidations tailored for Internet of Things (IoT) systems. It delves into the comprehensive examination of XAI explanations specifically designed for IoT systems. By leveraging emerging architectures built upon 5G services, cloud services, and big data management, the authors explore XAI techniques for adaptive explanations in IoT settings.
[41]	Microwave Communication	2513 Data Points	Model Agnostic-SHAP model.	Explainable Artificial Intelligence (XAI) significantly boosts trust in automated network administration. XAI finds application in two phases: during model development, where the machine learning model is designed, and during deployment, after the model is operational. While current research focuses solely on microwave communications, there's a pressing need to extend the exploration of explainable artificial intelligence techniques to other wireless communication domains.
[42]	IDS	NSL-KDD dataset	Adversarial Machine Learning	The methodology uses an adversarial approach to determine the least amount of alteration required to properly categorize the samples that were misclassified. The changes are employed to identify and display the pertinent characteristics that led to the incorrect classification. Perceptron classifiers, both linear and multilayer, were used in the experiments. The explanations were given through user-friendly graphs that are simple to understand. More work is required to incorporate the explanations and raise the model's accuracy.
[43]	IIoT security	NSL-KDD" and "UNSW"	TRUST XAI	Findings indicate that our model outperforms LIME by a factor of 25 in terms of speed and accuracy, making it a strong contender for critical and real-time applications. There are also certain restrictions. TRUST may overfit to the training set as a result of selecting the representatives based on knowledge gathering. This would result in poor performance with new data. Additional measures, such as the Gini index, could be considered when assessing models' success. The model can be extended to consider additional potential probability distributions, such as the Poisson.
[44]	Optical Networks	QoT Dataset Collection	SHAP	The main topic of discussion is the use of XAI for flightpath QoT estimates. How features' effects on the model were analyzed, how the model behaves was described, how features can be selected, and how information about the QoT estimation problem can be extracted and used to help domain experts in network design and decision-making. The work also demonstrated how to use SHAP for feature selection and examine instances incorrectly classified. The case study was limited to Optical networks only. It is considered a single-feature light path quality in transmission estimation.
[45]	B5G security	NA	NA	This manuscript comprehensively evaluates the most advanced AI, XAI, B5G technologies, and security aspects, including threat model and taxonomy. The technical aspects concerning the function of XAI in B5G security concerns were meticulously scrutinized. Facilitators such as IoT, RAN, Edge, core, backhaul, E2E slicing, and network automation were also discussed. Also, XAI association with security mechanisms such as encryption, anonymization, obfuscation, and federated learning were considered.

[46]	Network communication	--	--	This investigation explores the potential of utilizing XAI to advance accountability and resilience beyond the 5G period of AI-based security in communication. The analysis begins with assessing XAI's potential in the B5G period.
[47]	Communication	--	--	Explainable AI's role in making B5G network AI/ML models transparent and accountable, enhancing various B5G technologies.



Role of XAI in devices/IoT
End Users
Explanations
Data
Attackers
IoT devices/ wearables
Intelligent Cloud Network
Intelligent Cloud/Edge /Fog Servers
XAI Methods
Explanations generated surrounding ML/AI methods are communicatedto the users and operators through the IoT infrastructure
Learning based authentication
Learning based Malware detection
Secure IoT offloading
Learning based access control
Q-Learning
Dyna-Q
DQN
D/NN
SVM
Naive Bayes
KNN
Random Forest
dFW
Network Attacks
DDoS, spoofing, MITM, sinkhole attacks ...
Application Attacks
Viruses, spyware, worms, DoS, Phishing ...
Encryption Attacks
Side channel attacks, cryptanalysis attacks, MITM ...
Physical Attacks
Node Tampering, malicious node injection, sleep deprivation ...
Transport layer Attacks
Buffer overflow, Malware, DoS, Side channel, MITM

Figure 4. XAI in communication systems [45]

3.6 XAI in cloud

The intricate architecture of the cloud presents a challenging problem for XAI to accurately explain the factors that determine specific decisions. Furthermore, data quality is a major issue for XAI, as it is for most AI systems. XAI systems are needed to be trained on datasets which are of high quality, that encompass a wide range of scenarios to perform effectively. Additionally, developing XAI-based architectural evaluation tools is an expensive and hardware and software-intensive course, making cost a significant challenge [54-56]. The absence of standard XAI approaches and techniques makes it arduous to differentiate and assess different methods, posing another challenge [57].

As is clear from Figure 5, XAI is a valuable resource for cloud architecture evaluation, providing better insights into system architectures and enhancing transparency for organizations and cloud service providers to make informed decisions and adjust their operational policies [58, 59]. To reap the full benefits of XAI, best practices like usage of high-quality data, choosing suitable XAI systems, and rigorously validating and interpreting outcomes to safeguard correctness and deficiency of bias must be monitored.

3.7 XAI in social media

The vast amount of data available at social media platforms has much influenced decision-making process of common man. Social media platforms like Facebook and twitter are widely used by people not only to remain in contact with their friends and colleagues but also to keep themselves updated about the current change’s happenings around the globe. These platforms keep them informed and updated about the latest trends and technology. The widespread use and deep integration of social media in modern society has opened up new avenues for interaction, information exchange, group formation, and financial gain on an unprecedented scale. XAI can assist the people who make decisions based on recommendations received in these social media platforms [60, 61]. If a reasoning is provided by these social media platforms along with the proposed recommendations, then definitely it will be easier for the person to arrive at an effective decision. Despite the benefits, social media's popularity can be exploited by different anti-social entities to disseminate wrong information, also known as "Fake News," for profit or to influence society's behavior. Numerous counter measures have been devised to lessen the effects and dissemination of fake news. These include language-based methods that commonly utilize Deep Learning (DL) and Natural Language Processing (NLP) [62, 63]. XAI is also finding its usage in social media hate speech detection, as regulatory considerations and ethical concerns arise in social media. The model's great performance is no longer enough, according to current developments in the domain of artificial intelligence (AI). The system's decision-making process must also be explicable in realistic settings. The purpose of this [64, 65] study is to present a novel explainability method for BERT-based fake news identification.

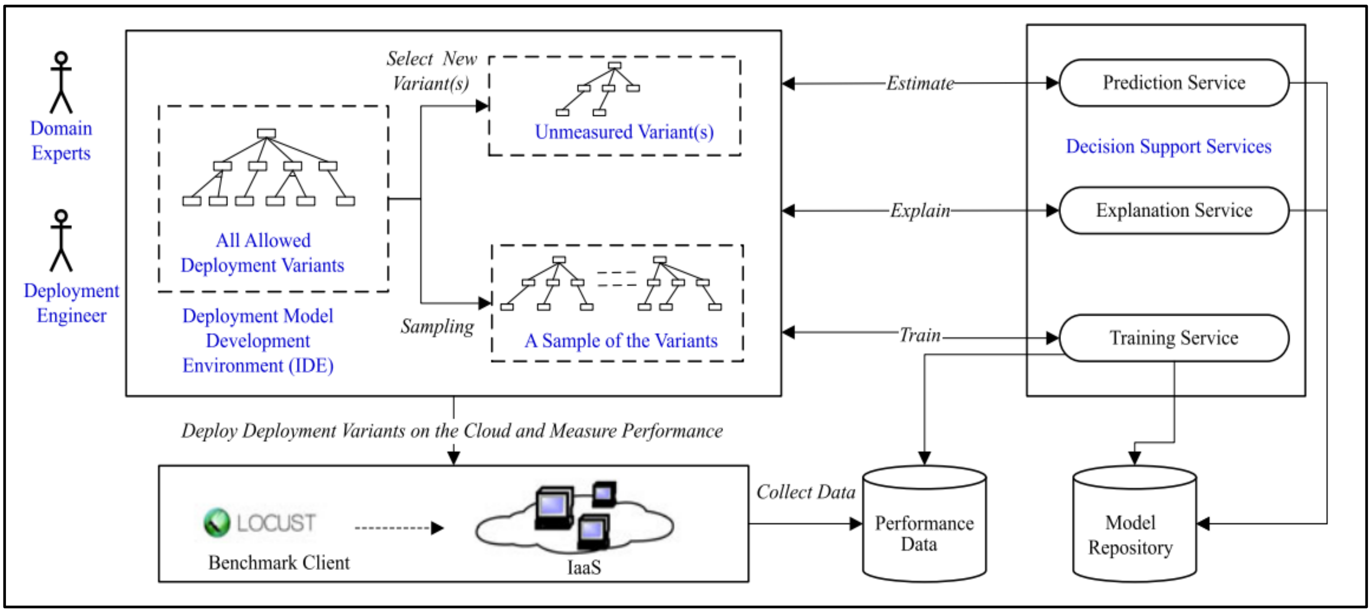


Figure 5. High level architecture of feature-oriented cloud architecture [58]

3.8 XAI in Internet of Things

The Internet of items (IoT) is a seeming domain of computers which basically links intelligent items and objects, enabling them to interact and offer people enhanced services [66]. The incorporation of XAI methodology into the IoT domain is expected to provide a substantial research potential in terms of the transparency, explicability, and interpretability of the AI and ML tools in IoT [66]. A thorough analysis of XAI models in AI-powered IoT applications has also been taken into consideration by a number of researchers, giving researchers a thorough view on the areas in which XAI approaches are used in the IoT domain. XAI models have significantly improved wireless communication services and IoT during the past few years [67, 68]. Applications of XAI in IoT systems are shown in Figure 6. IoT use cases that are dependent on the characteristics that can be explained are becoming more and more necessary in order to make sure that users can trust and understand these systems.

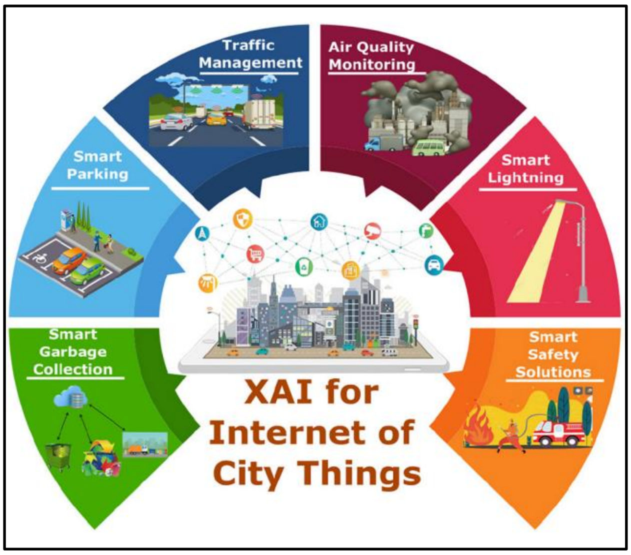


Figure 6. Important IoCT sector seeking XAI attention [68]

Figure 6 shows some of the significant areas where XAI can be utilized for designing and deployment of smart city. The openness and reliability of IoT frameworks may pose additional difficulties for efficient semantic communication [69]. In addition, reliability in the design possibilities of black-box attack mitigation is necessary to investigate and evaluate the accuracy and assurance of XAI approaches. Kök et al. [70] proposes an agent-based strategy for creating explainable Internet of Things (IoT) systems and gives an overview of the significance of XAI in smart home systems, a technology based on AI, which is getting popular and accessible to end users. The implementation of 35 city services in 27 cities between Europe and South Korea is considered in the study [71]. The resulting atomic services also generate a side market for smart city solutions, allowing expertise and know-how to be reused. Effective communication between AI models and its users depends on the ability of the former to have a faithful mental representation of the latter and a critical ability to assess the strengths and limitations of ML models.

4. CONCLUSION

Explainable Artificial Intelligence (XAI) offers a promising solution to the "black box" problem in AI, aiming to enhance transparency and interpretability of AI models. Its applications in critical domains like healthcare, finance, and autonomous vehicles can lead to better decision-making and increased user trust. Despite challenges in balancing interpretability with performance, XAI remains crucial for ensuring ethical and accountable AI deployment. Embracing XAI principles is essential for integrating AI technologies effectively into society while minimizing potential risks. In recent years, advancements in technology, data availability, and machine learning algorithms have fueled the acceptance of AI. However, the opaque nature of these models poses barriers to their widespread adoption, raising ethical concerns. To address these issues, this research proposes a pragmatic evaluation methodology for explainable machine learning models, comparing them with black-box models. While explanations are vital for building trust, the study highlights the importance of exploring methods to communicate model uncertainty effectively. These findings lay a strong foundation for future research, emphasizing the need for nuanced approaches to enhance trust and acceptance of AI models in various settings.

REFERENCES

[1] Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58: 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

[2] Von Eschenbach, W.J. (2021). Transparency and the black box problem: Why we do not trust AI. Philosophy & Technology, 34(4): 1607-1622. https://doi.org/10.1007/s13347-021-00477-0

[3] Samek, W., Wiegand, T., Müller, K.R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296. https://doi.org/10.48550/arXiv.1708.08296

[4] De, T., Giri, P., Mevawala, A., Nemani, R., Deo, A. (2020). Explainable AI: a hybrid approach to generate human-interpretable explanation for deep learning prediction. Procedia Computer Science, 168: 40-48. https://doi.org/10.1016/j.procs.2020.02.255

[5] Bishara, D., Xie, Y., Liu, W.K., Li, S. (2023). A state-of-the-art review on machine learning-based multiscale modeling, simulation, homogenization and design of materials. Archives of Computational Methods in Engineering, 30(1): 191-222. https://doi.org/10.1007/s11831-022-09795-8

[6] Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. International Journal of Human-Computer Studies, 146: 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

[7] Amann, J., Blasimme, A., Vayena, E., Frey, D., Madai, V.I. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. BMC Medical Informatics and Decision Making, 20: 1-9. https://doi.org/10.1186/s12911-020-01332-6

[8] Arreche, O., Guntur, T.R., Roberts, J.W., Abdallah, M. (2024). E-XAI: Evaluating black-box explainable AI frameworks for network intrusion detection. IEEE Access, 12: 23954-23988. https://doi.org/10.1109/ACCESS.2024.3365140

[9] Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1): 2053951715622512. https://doi.org/10.1177/2053951715622512

[10] Li, X.H., Cao, C.C., Shi, Y., Bai, W., Gao, H., Qiu, L., Chen, L. (2020). A survey of data-driven and knowledge-aware explainable AI. IEEE Transactions on Knowledge and Data Engineering, 34(1): 29-49. https://doi.org/10.1109/TKDE.2020.2983930

[11] Gilpin, L.H., Bau, D., Yuan, B.Z., Bajwa, A., Specter, M., Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. In 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA), Turin, Italy, pp. 80-89. https://doi.org/10.1109/DSAA.2018.00018

[12] Emmert-Streib, F., Yli-Harja, O., Dehmer, M. (2020). Explainable artificial intelligence and machine learning: A reality rooted perspective. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(6): e1368. https://doi.org/10.1002/widm.1368

[13] Futia, G., Vetrò, A. (2020). On the integration of knowledge graphs into deep learning models for a more comprehensible AI—Three challenges for future research. Information, 11(2): 122. https://doi.org/10.3390/info11020122

[14] Gunning, D., Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. AI Magazine, 40(2): 44-58. https://doi.org/10.1609/aimag.v40i2.2850

[15] Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C.Y., Taher, F. (2022). Explainable artificial intelligence applications in cyber security: State-of-the-art in research. IEEE Access, 10: 93104-93139. https://doi.org/10.1109/ACCESS.2022.3204051

[16] Gunning, D., Vorm, E., Wang, Y., Turek, M. (2021). DARPA's explainable AI (XAI) program: A retrospective. Authorea Preprints. https://doi.org/10.22541/au.163699841.19031727/v1

[17] Mendes, C., Rios, T.N. (2023). Explainable artificial intelligence and cybersecurity: A systematic literature review. arXiv preprint arXiv:2303.01259. https://doi.org/10.48550/arXiv.2303.01259

[18] Maathuis, C. (2022). On explainable AI solutions for targeting in cyber military operations. In International Conference on Cyber Warfare and Security, The Netherlands, pp. 166-175.

[19] Holder, E., Wang, N. (2021). Explainable artificial intelligence (XAI) interactively working with humans as a junior cyber analyst. Human-Intelligent Systems Integration, 3(2): 139-153. https://doi.org/10.1007/s42454-020-00021-z

[20] Preece, A., Braines, D., Cerutti, F., Pham, T. (2019). Explainable AI for intelligence augmentation in multi-domain operations. arXiv preprint arXiv:1910.07563. https://doi.org/10.48550/arXiv.1910.07563

[21] Sarp, S., Kuzlu, M., Wilson, E., Cali, U., Guler, O. (2021). The enlightening role of explainable artificial intelligence in chronic wound classification. Electronics, 10(12): 1406. https://doi.org/10.3390/electronics10121406

[22] Kamal, M.S., Dey, N., Chowdhury, L., Hasan, S.I., Santosh, K.C. (2022). Explainable AI for glaucoma prediction analysis to understand risk factors in treatment planning. IEEE Transactions on Instrumentation and Measurement, 71: 1-9. https://doi.org/10.1109/TIM.2022.3171613

[23] Quellec, G., Al Hajj, H., Lamard, M., Conze, P.H., Massin, P., Cochener, B. (2021). ExplAIn: Explanatory artificial intelligence for diabetic retinopathy diagnosis. Medical Image Analysis, 72: 102118. https://doi.org/10.1016/j.media.2021.102118

[24] Singh, A., Pannu, H.S., Malhi, A. (2022). Explainable information retrieval using deep learning for medical images. Computer Science and Information Systems, 19(1): 277-307. https://doi.org/10.2298/csis201030049s

[25] Ali, S., Akhlaq, F., Imran, A.S., Kastrati, Z., Daudpota, S.M., Moosa, M. (2023). The enlightening role of explainable artificial intelligence in medical & healthcare domains: A systematic literature review. Computers in Biology and Medicine, 107555. https://doi.org/10.1016/j.compbiomed.2023.107555

[26] Allgaier, J., Mulansky, L., Draelos, R.L., Pryss, R.

(2023). How does the model make predictions? A systematic literature review on the explainability power of machine learning in healthcare. Artificial Intelligence in Medicine, 143: 102616. https://doi.org/10.1016/j.artmed.2023.102616

[27] Srinivasu, P.N., SivaSai, J.G., Ijaz, M.F., Bhoi, A.K., Kim, W., Kang, J.J. (2021). Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. Sensors, 21(8): 2852. https://doi.org/10.3390/s21082852

[28] Payrovnaziri, S.N., Chen, Z., Rengifo-Moreno, P., Miller, T., Bian, J., Chen, J.H., He, Z. (2020). Explainable artificial intelligence models using real-world electronic health record data: A systematic scoping review. Journal of the American Medical Informatics Association, 27(7): 1173-1185. https://doi.org/10.1093/jamia/ocaa053

[29] Barata, C., Celebi, M.E., Marques, J.S. (2021). Explainable skin lesion diagnosis using taxonomies. Pattern Recognition, 110: 107413. https://doi.org/10.1016/j.patcog.2020.107413

[30] Cui, Y., Cui, Y. (2020). Application of AI in judicial practice. Artificial Intelligence and Judicial Modernization, 21-31. https://doi.org/10.1007/978-981-32-9880-4_2.2020

[31] Hanif, A. (2021). Towards explainable artificial intelligence in banking and financial services. arXiv preprint arXiv:2112.08441. https://doi.org/10.48550/arXiv.2112.08441

[32] Sai, C.V., Das, D., Elmitwally, N., Elezaj, O., Islam, M.B. (2023). Explainable ai-driven financial transaction fraud detection using machine learning and deep neural networks. https://doi.org/10.2139/ssrn.4439980

[33] Demertzis, K., Rantos, K., Magafas, L., Skianis, C., Iliadis, L. (2023). A secure and privacy-preserving blockchain-based XAI-Justice system. Information, 14(9): 477. https://doi.org/10.20944/preprints202305.2017.v1

[34] Deeks, A. (2019). The judicial demand for explainable artificial intelligence. Columbia Law Review, 119(7): 1829-1850.

[35] Oconitrillo, L.R.R., Vargas, J.J., Camacho, A., Burgos, A., Corchado, J.M. (2021). RYEL System: A novel method for capturing and represent knowledge in a legal domain using explainable artificial intelligence (XAI) and granular computing (GrC). Interpretable Artificial Intelligence: A Perspective of Granular Computing, 369-399. https://doi.org/10.1007/978-3-030-64949-4_12

[36] Weber, P., Carl, K.V., Hinz, O. (2023). Applications of explainable artificial intelligence in finance—A systematic review of finance, information systems, and computer science literature. Management Review Quarterly, 1-41. https://doi.org/10.1007/s11301-023-00320-0

[37] Koster, O., Kosman, R., Visser, J. (2021). A checklist for explainable AI in the insurance domain. In International Conference on the Quality of Information and Communications Technology, Algarve, Portugal, pp. 446-456. https://doi.org/10.1007/978-3-030-85347-1_32

[38] Bora, A., Sah, R., Singh, A., Sharma, D., Ranjan, R.K. (2022). Interpretation of machine learning models using XAI-a study on health insurance dataset. In 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, pp. 1-6. https://doi.org/10.1109/ICRITO56286.2022.9964649.

[39] Mahbooba, B., Timilsina, M., Sahal, R. Serrano, M. (2021). Explainable artificial intelligence (XAI) to enhance trust management in intrusion detection systems using decision tree model. Complexity, 2021, pp.1-11. https://doi.org/10.1155/2021/6634811

[40] Nwakanma, C.I., Ahakonye, L.A.C., Njoku, J.N., Odirichukwu, J.C., Okolie, S.A., Uzondu, C., Kim, D.S. (2023). Explainable artificial intelligence (XAI) for intrusion detection and mitigation in intelligent connected vehicles: A review. Applied Sciences, 13(3): 1252. https://doi.org/10.3390/app13031252

[41] Ayoub, O., Di Cicco, N., Ezzeddine, F., Bruschetta, F., Rubino, R., Nardecchia, M., Tornatore, M. (2022). Explainable artificial intelligence in communication networks: A use case for failure identification in microwave networks. Computer Networks, 219: 109466. https://doi.org/10.1016/j.comnet.2022.109466

[42] Marino, D.L., Wickramasinghe, C.S., Manic, M. (2018). An adversarial approach for explainable ai in intrusion detection systems. In IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society, Washington, DC, USA, pp. 3237-3243. https://doi.org/10.1109/IECON.2018.8591457

[43] Zolanvari, M., Yang, Z., Khan, K., Jain, R., Meskin, N. (2021). Trust XAI: Model-agnostic explanations for AI with a case study on IIOT security. IEEE Internet of Things Journal, 10(4): 2967-2978. https://doi.org/10.1109/JIOT.2021.3122019

[44] Ayoub, O., Troia, S., Andreoletti, D., Bianco, A., Tornatore, M., Giordano, S., Rottondi, C. (2022). Towards explainable artificial intelligence in optical networks: The use case of lightpath QoT estimation. Journal of Optical Communications and Networking, 15(1): A26-A38. https://doi.org/10.1364/JOCN.470812

[45] Sharma, S., Nag, A., Cordeiro, L., Ayoub, O., Tornatore, M., Nekovee, M. (2020). Towards explainable artificial intelligence for network function virtualization. In Proceedings of the 16th International Conference on Emerging Networking Experiments and Technologies, New York, USA, pp. 558-559. https://doi.org/10.1145/3386367.3431673

[46] Wu, Y., Lin, G., Ge, J. (2022). Knowledge-powered explainable artificial intelligence for network automation toward 6G. IEEE Network, 36(3): 16-23. https://doi.org/10.1109/MNET.005.2100541

[47] Senevirathna, T., La, V.H., Marchal, S., Siniarski, B., Liyanage, M., Wang, S. (2022). A survey on XAI for beyond 5G security: technical aspects, use cases, challenges and research directions. arXiv preprint arXiv:2204.12822. https://doi.org/10.48550/arXiv.2204.12822

[48] Ferguson, W., Batra, D., Mooney, R., Parikh, D., Torralba, A., Bau, D., Lee, S. (2021). Reframing explanation as an interactive medium: The EQUAS (Explainable QUestion Answering System) project. Applied AI Letters, 2(4): e60. https://doi.org/10.1002/ail2.60

[49] Wang, Y.C., Chen, T. (2023). New XAI tools for selecting suitable 3D printing facilities in ubiquitous manufacturing. Complex & Intelligent Systems, 9(6): 6813-6829. https://doi.org/10.1007/s40747-023-01104-5

[50] Kharal, A. (2020). Explainable artificial intelligence based fault diagnosis and insight harvesting for steel

plates manufacturing. arXiv preprint arXiv:2008.04448. https://doi.org/10.48550/arXiv.2008.04448

[51] Hanchate, A., Bukkapatnam, S.T., Lee, K.H., Srivastava, A., Kumara, S. (2023). Explainable AI (XAI)-driven vibration sensing scheme for surface quality monitoring in a smart surface grinding process. Journal of Manufacturing Processes, 99: 184-194. https://doi.org/10.1016/j.jmapro.2023.05.016

[52] Askr, H., Elgeldawi, E., Aboul Ella, H., Elshaier, Y.A., Gomaa, M.M., Hassanien, A.E. (2023). Deep learning in drug discovery: An integrative review and future challenges. Artificial Intelligence Review, 56(7): 5975-6037. https://doi.org/10.1007/s10462-022-10306-1

[53] Naumets, S., Lu, M. (2021). Investigation into explainable regression trees for construction engineering applications. Journal of Construction Engineering and Management, 147(8), 04021084. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002083

[54] Astolfi, D., De Caro, F., Vaccaro, A. (2023). Condition monitoring of wind turbine systems by explainable artificial intelligence techniques. Sensors, 23(12): 5376. https://doi.org/10.3390/s23125376

[55] Forti, S., Ferrari, G.L., Brogi, A. (2020). Secure cloud-edge deployments, with trust. Future Generation Computer Systems, 102: 775-788. https://doi.org/10.1016/j.future.2019.08.020

[56] Guerrero, C., Lera, I., Juiz, C. (2019). A lightweight decentralized service placement policy for performance optimization in fog computing. Journal of Ambient Intelligence and Humanized Computing, 10: 2435-2452. https://doi.org/10.1007/s12652-018-0914-0

[57] Chen, M.H., Dong, M., Liang, B. (2016). Joint offloading decision and resource allocation for mobile cloud with computing access point. In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, pp. 3516-3520. https://doi.org/10.1109/ICASSP.2016.7472331

[58] Kumara, I., Ariz, M.H., Chhetri, M.B., Mohammadi, M., Van Den Heuvel, W.J., Tamburri, D.A. (2022). FOCloud: feature model guided performance prediction and explanation for deployment configurable cloud applications. IEEE Transactions on Services Computing, 16(1): 302-314. https://doi.org/10.1109/TSC.2022.3142853

[59] Ye, K., Shen, H., Wang, Y., Xu, C.Z. (2020). Multi-tier workload consolidations in the cloud: Profiling, modeling and optimization. IEEE Transactions on Cloud Computing, 10(2): 899-912. https://doi.org/10.1109/TCC.2020.2975788.

[60] Rai, A. (2020). Explainable AI: From black box to glass box. Journal of the Academy of Marketing Science, 48: 137-141. https://doi.org/10.1007/s11747-019-00710-5

[61] Cirqueira, D., Almeida, F., Cakir, G., Jacob, A., Lobato, F., Bezbradica, M., Helfert, M. (2020). Explainable sentiment analysis application for social media crisis management in retail. In 4th International Conference on Computer-Human Interaction Research and Applications - Volume 1: WUDESHI-DR, Budapest, Hungry. https://doi.org/10.5220/0010215303190328

[62] Farrow, R. (2023). The possibilities and limits of XAI in education: A socio-technical perspective. Learning, Media and Technology, 48(2): 266-279. https://doi.org/10.1080/17439884.2023.2185630

[63] Lim, Perrault, S.T. (2022). Explanation preferences in XAI Fact-checkers. In Proceedings of 20th European Conference on Computer-Supported Cooperative Work. European Society for Socially Embedded Technologies (EUSSET).

[64] Szczepański, M., Pawlicki, M., Kozik, R., Choraś, M. (2021). New explainability method for BERT-based model in fake news detection. Scientific Reports, 11(1): 23705. https://doi.org/10.1038/s41598-021-03100-6

[65] Ciampaglia, G.L. (2018). Fighting fake news: a role for computational social science in the fight against digital misinformation. Journal of Computational Social Science, 1(1): 147-153. https://doi.org/10.1007/s42001-017-0005-6 (2018).

[66] Sobin, C.C. (2020). A survey on architecture, protocols and challenges in IoT. Wireless Personal Communications, 112(3): 1383-1429. https://doi.org/10.1007/s11277-020-07108-5

[67] Jagatheesaperumal, S.K., Pham, Q.V., Ruby, R., Yang, Z., Xu, C., Zhang, Z. (2022). Explainable AI over the Internet of Things (IoT): Overview, state-of-the-art and future directions. IEEE Open Journal of the Communications Society, 3: 2106-2136. https://doi.org/10.1109/OJCOMS.2022.3215676

[68] Hou, B., Gao, J., Guo, X., Baker, T., Zhang, Y., Wen, Y., Liu, Z. (2021). Mitigating the backdoor attack by federated filters for industrial IoT applications. IEEE Transactions on Industrial Informatics, 18(5): 3562-3571. https://doi.org/10.1109/TII.2021.3112100

[69] Dobrovolskis, A., Kazanavičius, E., Kižauskienė, L. (2023). Building XAI-based agents for IoT systems. Applied Sciences, 13(6): 4040. https://doi.org/10.3390/app13064040

[70] Kök, I., Okay, F.Y., Muyanlı, Ö., Özdemir, S. (2023). Explainable artificial intelligence (XAI) for Internet of Things: A survey. IEEE Internet of Things Journal, 10(16): 14764-14779. https://doi.org/10.48550/arXiv.2206.04800

[71] Cirillo, F., Gómez, D., Diez, L., Maestro, I.E., Gilbert, T.B.J., Akhavan, R. (2020). Smart city IoT services creation through large-scale collaboration. IEEE Internet of Things Journal, 7(6): 5267-5275. https://doi.org/10.1109/JIOT.2020.2978770