Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 46 (2025), No. 6, pp. 1861–1869
DOI : 10.47974/JIOS-2015

A comprehensive review of explainable artificial intelligence (XAI) in various domains : Applications and challenges

Richa Datta *
Department of Computer Science and Engineering
Maharishi Markendeshwar Institute of Computer Technology & Business Management (Hotel Management)
Maharishi Markendeshwar Engineering College
Maharishi Markandeshwar (Deemed To Be University)
Ambala 133203
Haryana
India

Vaishali Mehta †
Department of Computer Science and Engineering
Maharishi Markendeshwar Engineering College
Maharishi Markandeshwar (Deemed To Be University)
Ambala 133203
Haryana
India

Abstract

To make artificial intelligence, scientists look at how the brain works and how it learns. These investigations have led to the creation of software and smart systems. Artificial Intelligence (AI) is a way to solve everyday problems by leveraging technology and systems that act like people. We need to build trust and employ AI systems in important sectors since we rely on smart robots more and more in our daily lives. The short form of "Explainable Artificial Intelligence" is "XAI." AI that can explain its decisions or predictions to people. XAI wants to make AI systems more open, accountable, and trustworthy, especially when they are utilized in important fields like security, healthcare, or finance.

Subject Classification: *68T01, 68T07, 68T05.*

Keywords: *XAI, Machine learning, Health care, DL.*

* *ORCID-ID:* **0009-0003-0504-0837**
† *ORCID-ID:* **0000-0003-2017-0528**

* *E-mail:* **er.richacse@gmail.com (Corrseponding Author)**
† *E-mail:* **drvaishaliwashwa@gmail.com**

1. Introduction

The Defense Advanced Research Projects Agency (DARPA) came up with the term “explainable artificial intelligence.” It is called a “white box” because it shows how the model works [1]. There are two types of XAI approaches: data-driven and knowledge-driven. Not only this learning process provides the prediction but it also provides explanation for the prediction [2]. A data driven approach is a method that uses data to make decisions and inform actions. In order to find patterns, trends, and insights that can direct decisions, it requires gathering, evaluating, and interpreting data [3]. Explain ability highlights potentials factors that can change the predictions.

Explainable AI aims to convert opaque” black box” into a clear “glass box” (sometimes called a “white box”). Similar to a decision tree or linear regression model, a glass box model provides complete transparency since we know all of its parameters and how the model arrives at its conclusion [4]. Making a model understandable to its stakeholders is the primary goal of explanations. As a result, a number of techniques have recently been created to explain the choices made by sophisticated AI systems across a range of application domains Application areas of XAI [5, 6]

Explainable AI in agriculture Leaf Disease Classification was identified by XAI (See Figure 1) using DL models in the agricultural classification sector. The five- leaf dataset is used.

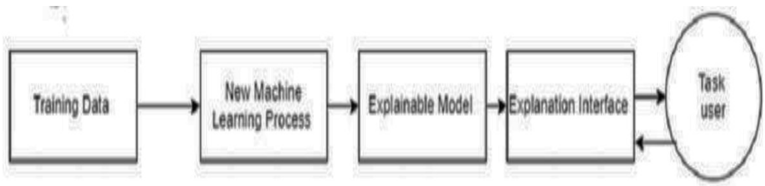

Figure 1

Typical XAI

Explainable AI in finance [7] was able to predict the credit card default application by using TREPAN decision tree clustering, network hidden layer representation clustering, and data sets from the UCI Machine Learning Repository.

Explainable AI in forecasting Using artificial intelligence (AI) and semantic technologies for the XAI, a demand forecasting detection system used in real-world applications. Computer vision with explainable AI [8]. In an effort to improve the model’s interpretability and comprehension, researchers has examined Multimodal Artificial Intelligence with XAI. For problems in NLP, computer vision, and many other domains, deep neural networks are essential. Through XAI methods, end users can better understand neural networks with the aid of three categories of visual analytics techniques: post-modelling explain ability, interpretable models, and pre-modelling explain ability [9].

Explainable AI in the healthcare domain This technique could be used in hospitals to diagnose MI (Myocardial Infarction, also known as a heart attack). The purpose of XAI is to win over AI models. In practical situations, it will increase the trust in predictive modeling. To identify the biomarkers that can be used to predict skin cancers, XG boost machine learning models have been trained using the Python SHAP (Shapley Additive explanations) programming language [10]. Using XAI, Parikh et al. [8] investigated the prediction of cognitive impairment and AD, also referred to as Alzheimer's disease. To predict the disease, Random Forest Classifiers and the ADNIMERGE R package have been utilized. LIME was used to explain the models, and relevant features needed for precise prediction were extracted using SHAP values.

Explainable AI in transportation- technique of Sensitive analysis is employed to distinguish between the categories by elucidating the essential components of genuine images. To improve the accuracy of picture classification in self-driving cars, the XAI was required. Following table represents limitations and applications areas of XAI models [11].

IV. XAI ALGORITHMS

These include Gradient-weighted Class Activation Mapping (Grad CAM), Shapely Additive explanations (SHAP), Layer-wise Relevance Propagation (LRP), Fuzzy Classifier, Partial Dependence Plots (PDP), and Local Interpretable Model-agnostic Explanations (LIME) [12].

LIME is used to give local explanations by modeling how the model would act around a certain data point and explaining why AI got to that place.

- Once a decision has been made, it searches for nearby points, constructs a model that represents the local choice, and then uses that model to describe each feature.
- Every image is broken down into super pixels by LIME as shown in Figure 2. After that, it asks the model about a random search space

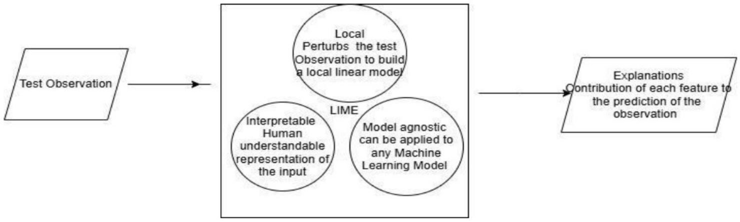

Figure 2

Shows Explainable AI with Lime

and modifies which super pixels are replaced with the selected color and which are left out.

SHAP- To demonstrate prediction in Shapley Values, consider each feature value as a player in a game where the payout is the prediction. Coalitional game theory's Shapley values technique explains how to divide the "payout" among the characteristics in a fair manner.

- The feature values that cooperate to obtain the gain (i.e., predict a specific value) are the players. The average marginal contribution of a feature value across all potential coalitions is known as the Shapley value.
- A useful technique for calculating Shapley values in machine learning models is SHAP as shown in Figure 3. Consequently, SHAP provides feature attributions and explanations for individual predictions using a sampling-based technique making it more scalable and applicable to complex models

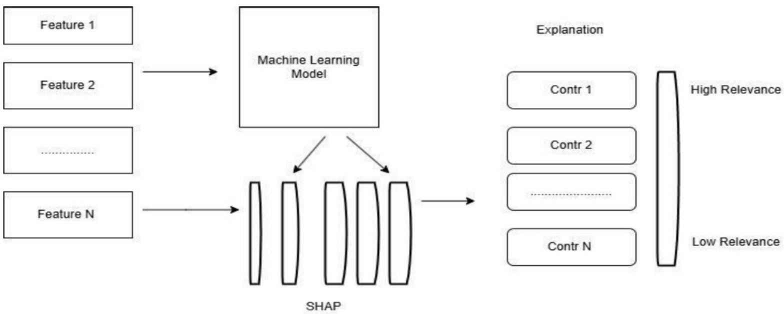

Figure 3

Shows Explainable AI with SHAP

PDP -Partially Dependence Plots show how much the projected response changes when the feature (the value of a given data column or component) changes. It can be helpful to find out if there is a simple or more complicated link between a feature and the response.

A portion of dependence Plots show how a particular feature, when all other features are held constant, marginally affects the model's predictions.

PDPs allow us to understand the impact of a particular feature on the complete data set by providing an overview of how changes to that feature impact the model's output [13].

In table 1 below shows various research problems their proposed approach, performance metrics and research findings and Gaps.

Table 1

Shows different problem statements and their research findings

Ref.	Problem statement	Proposed Approach	Performance Metrics	Research findings	Gaps
[11]	Classification of thyroid disease using XAI	Different thyroid prediction results are provided by the machine learning algorithms Support Vector Machine, Random Forest, and XG Boost.	F1-score, Ac-Curacy Precision Recall	The results show an enhanced accuracy of 99.1 for the ensemble classifier and 98.5 for the suggested XG Boost classifier.	Class bias and data preparation have little effect on the model's overall performance.
[12]	From video to vital signs: employing explainable AI to estimate blood pressure and measure pulse rate using personal device cameras	Feature Engineering	No of Participant, signal length, video collected ground truth	The ME and SD for the systolic model are 1.2 and 15.9 mmHg, respectively. The diastolic model's standard deviation (SD) is 13.1 mmHg, and its mean error (ME) is -3.2 mmHg.	Maintaining accuracy depends critically on the quality of the signal.

Contd...

[13]	An overview of XAI models For clinical practitioners.	LIME Shapley	Visual Textual Auxiliary Case based	Displayed image was classified as hand with prediction score of 0.98.Red areas denotes positive contribution, blue areas denote negative contribution	Because XAI is still not widely used in modern medical procedures, the majority of the survey questions were speculative
[14]	Liver Approach to Biomedical AI	Deep convolutional neural network and Grad-CAM.	Accuracy, precision, Recall, F-measure	The deep learning model, which was based using Keras-Tensor flow, has an accuracy score of 0.81 (0.82 for the hyper parameter-tuned model). This is because the target values are not evenly distributed.	In different medical settings different explainable strategies can be embedded with deep learning models in bio- health-care.

The table 2 below shows various application areas like agriculture, finance, for casting, computer vision, health care and transportation and techniques used in these problem statements.

Table 2

Limitations and applications areas ofxai models

Problem Statement	Domain	XAI Models /Technique used	Application	Outcomes
To study Leaf Disease Classification using Deep XAI	Agriculture	Gradient-weighted Class activation mapping, Local interpretable model-agnostic explanations Smooth Grad,	Classification of Leaf disease	With limited explanatory capacity, the LIME technique only describes how the leaves appear.

Contd...

To Provide a Human-Comprehensible Interpretation for Deep Learning Forecasting	Finance	TREPAN model	Business applications	Experimental results show that our "Cluster-TREPAN" approach outperforms the "Global TREPAN" approach
To study Rich and confidential Explainable Artificial Intelligence (XAI) based on knowledge graphs	Forecasting	Knowledge graph	Demand forecasting	To comprehend how such an approach can be used to models where features are not manually created.
Malware Detection Method for IoT Devices using Deep XAI	Computer vision	Gradient-Weighted class activation mapping	Detection of malware	Use of principal component analysis improves malware classification on IoT devices.
Analyzing X-ray Images using Explainable Artificial Intelligence.	Healthcare	Shapely additive explanations and class activation maps	Alzheimer's disease diagnosis using X-ray and image-related tasks	There is no actual use for the XAI models because they haven't been tested or improved based on explain ability results. Change the way these lines look too

Explainable AI is emerging and essential, particularly for research in medicine, which will boost confidence in AI.

VI. Conclusion

This study provides a comprehensive analysis of the literature on XAI techniques and their many applications. When research data sets are inaccurate or public datasets are used in place of real-world data, it highlights areas where research is lacking. In light of this, this review will be especially beneficial for understanding the principles and difficulties of XAI in a crucial scenario where users must make an informed choice based on information about the model's output and workings as provided by the Explanation Interface. However there are certain limitations found in this study such as most of the data set is noisy and could be biased by nature. The study did not include certain highly predictive prospective factors. Future directions implies to do research in delicate areas such as health care, forecasting, computer vision utilizing Explainable artificial intelligence.

References

[1] T. K. Mohammed, D. N. Vasundhara, S. H. Mehanoor, E. Sreedevi, P. R. Kumar, C. Manihass, and S. F. Baba, "A novel fusion approach for advancement in crime prediction and forecasting using hybridization of ARIMA and recurrent neural networks," *Journal of Information Systems Engineering and Management*, vol. 10, no. 38, pp. 404–420 (2025).

[2] G. L. Saini, P. D. Singh, K. D. Singh, and P. Narooka, "Enhancing data privacy through cryptographic innovations in discrete mathematical systems," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 28, no. 5-A, pp. 1461–1472 (2025).

[3] P. R. Kumar, R. K. Jha, and P. A. Kumar, "BrainHyperintensities: Automatic segmentation of white matter hyperintensities in clinical brain MRI images using improved deep neural network," *The Journal of Supercomputing*, vol. 80, pp. 15545–15581 (2024).

[4] B. Goutam, M. F. Hashmi, Z. W. Geem, and N. D. Bokde, "A comprehensive review of deep learning strategies in retinal disease diagnosis using fundus images," *IEEE Access*, vol. 10, pp. 57 796–57 823 (2022).

[5] P. R. Kumar, R. K. Jha, and A. Katti, "Brain tissues segmentation in neurosurgery: A systematic analysis for quantitative tractography approaches," *Acta Neurologica Belgica*, vol. 124, pp. 1–15 (2023).

[6] R. K. Moje, B. Tiple, S. M. Patil, T. Jadhav, A. P. Munshi, and A. Revekar, "A framework for multi-task learning optimization in deep neural networks: Balancing task priorities for improved performance," *Journal of Information and Optimization Sciences*, vol. 46, no. 4-B, pp. 1141–1151 (2025).

[7] M. Bellucci, N. Delestre, N. Malandain, and C. Zanni-Merk, "Towards a terminology for a fully contextualized xai," *Procedia Computer Science*, vol. 192, pp. 241–250 (2021).

[8] J. Parikh, A. Bhargava, S. M. Antony, Y. Puri, U. Maniar, D. Singhal, M. Singh, and M. Sharma, "An edge AI based real-time spatial monitoring system," *Journal of Information and Optimization Sciences*, vol. 46, no. 1, pp. 33–42 (2025).

[9] B. Shilpa, P. R. Kumar, and R. K. Jha, "LoRa DL: A deep learning model for enhancing the data transmission over LoRa using autoencoder," *The Journal of Supercomputing*, vol. 79, pp. 17079–17097 (2023).

[10] M. N. Nikhate, K. L. Bondar, and S. B. Kiwne, "Fixed point theorems for self-mappings in generalized metric spaces," *Journal of Dynamical Systems and Geometric Theories*, vol. 21, no. 2, pp. 121–126 (2023).

[11] P. R. Kumar and T. Ananthan, "Machine vision using LabVIEW for label inspection," *Journal of Innovation in Computer Science and Engineering*, vol. 9, no. 1, pp. 58–62 (2019).

[12] Z. S. Alsham, E. Bahçekapılı, and A. Ayaz, "Trends in IoT applications in smart campuses: A topic modeling approach," *COLLNET Journal of Scientometrics and Information Management*, vol. 19, no. 1, pp. 21–40 (2025).

[13] B. Shilpa, P. R. Kumar, and R. K. Jha, "Spreading factor optimization for interference mitigation in dense indoor LoRa networks," in *Proc. IEEE IAS Glob. Conf. Emerg. Technol.* (GlobConET), pp. 1–5 (2023).

[14] H. Siriwardana, K. N. Singh, P. R. Kumar, and C. Misra, "Automated-road crossing system using real-time object tracking," in Proc. 10th Int. Conf. Rel., Infocom Technol. Optim. (ICRITO), pp. 1–5 (2022).

Received December, 2024