DISSEMINATION OF KNOWLEDGE

Svedberg Open

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Explainable Artificial Intelligence Models for Interpretable Decision-Making in High-Stakes Applications

Anshy Singh[1], K Saroja Rani[2], Monisha J[3], Yugandhar Manchala[4], Dr. Sowjanya Bagadi[5], Suhas Bhise[6], Durga Prasad[7]

[1]Department of Computer Engineering & Applications, GLA University, Mathura, rakesh.kumar@gla.ac.in, Department of Computer Engineering & Applications, GLA University, Mathura, Email: anshy.singh@gla.ac.in
[2]Associate Professor, Department of Civil Engineering, Pragati Engineering College, ADB Road, Surampalem, NearPeddapuram, Kakinada District, Andhra Pradesh, India - 533437. Email: sarojapec@gmail.com
[3]Assistant Professor, Department of Management Studies, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Email: jmonishamba@maher.ac.in
[4]Assistant Professor, Department of Information Technology, Vardhaman College of Engineering, Shamshabad, Hyderabad, India - 501 218, Email: Yugandhar1567@vardhaman.org
[5]Assistant Professor, School of Business, Aditya University, Surampalem, Andhra Pradesh, Pin 533437, Email: sowjanyab@adityauniversity.in
[6]Assistant Professor, E&TC Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra, 411037, Email: suhas.bhise@vit.edu
[7]School of Engineering &Technology,Noida international University, Uttar Pradesh 203201, India, Email: durga.prasad@niu.edu.in

Abstract

Explainable Artificial Intelligence (XAI) has been taken into higher stakes by the applications of high-stakes performances including healthcare diagnosis, financial fraud detection, cybersecurity analytics, legal decision support, and self-driven intelligent systems, where a transparent and trustworthy decision-making process is essential. Nevertheless, most traditional machine learning and deep learning systems are black-box systems, which can be extremely accurate in their predictions, but are not interpretable and are not based on human-irable arguments. This weakness lowers the user trust, responsibility, and reliability in real-world scenarios that are sensitive, and wrong or unaccount-explained decisions can cause dire outcomes. To overcome this challenge, this paper suggests a next generation Explainable Artificial Intelligence framework to interpret decisions made in high stakes applications. The proposed framework combines machine learning and deep learning architectures with several explainability mechanisms such as SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-Agnostic Explanations), attention-based feature visualization, and counterfactual reasoning to produce both global and local model explanations. It has a hybrid decision interpretation module that enhances the analysis of feature attribution, consistency of explanations and semantic transparency of model predictions. The model was tested on benchmark high stakes datasets in real time intelligent decision making scenarios. Experimental findings indicate that the proposed method realized a 96.4 percent prediction accuracy and a 95.1 percent precision as well as a 94.7 and 94.9 percent recall, F1-score, and AUC-ROC, respectively, outperforming traditional black-box machine learning and deep learning models and significantly enhancing interpretability, transparency, and human-understandable reasoning with low computational loads that can be effectively applied in critical applications.

Keywords: Explainable Artificial Intelligence (XAI), Interpretable Machine Learning, SHAP, LIME, Attention Mechanism, Counterfactual Reasoning, Decision-Making Systems, High-Stakes Applications, Trustworthy AI, Transparent Deep Learning, Human-Centered AI, Predictive Analytics

1. Introduction

Technologies of Artificial Intelligence (AI) and machine learning have quickly disrupted contemporary decision-making informatics in high stakes areas such as healthcare diagnostics, financial fraud detection, cyber-security analytics, autopiloting transportation, legal analysis, and intelligent industrial monitoring (Evseev, 2023). The outstanding capacity of advanced machine learning models and deep learning to handle large complex data has greatly enhanced predictive power and efficiency in actual applications (Georgiev et al., 2017; Jain et al., 2015). Irrespective of these developments, most of the current AI systems are black-box models whose decision-making processes are hard to interpret or explain. Such non-transparency restricts user trust, accountability, fairness, and reliability especially in sensitive settings where wrong or biased decisions can lead to dire financial, ethical, medical or legal repercussions. As a result, there is a growing need regarding Explainable Artificial Intelligence (XAI) to make AI-driven decisions which should be transparent, interpretable, and understandable by humans (Arrieta et al., 2020).

Traditional explainability solutions usually deal with either machine learning explainability or deep learning visualization, but not both, and therefore lack contextual reasoning as well as lack the uniformity in explanation across various domains of application. Moreover, most of the existing studies focus on the accuracy of prediction but ignore the quality of the explanations and the semantic transparency and reliability of decisions. Adebayo et al. (2018) have shown that a number of saliency-based explanation techniques cannot be trusted to offer accurate interpretability based on a variety of conditions of neural networks, which underscores the necessity of having powerful and dependable explanation systems. These restrictions pose serious difficulties in implementing AI systems in severe settings that need reliable and responsible decision support. To overcome these challenges, this paper introduces a state-of-the-art Explainable Artificial Intelligence system to interpretable decision-making in high-stakes scenarios through combined machine learning and deep learning models and supporting multiple explainability systems such as SHAP, LIME, attention-based visualization, and counterfactual reasoning. The proposed framework will result in better feature attribution analysis, consistency in explanation and semantic interpretability but with high predictive power. Experimental assessment has shown that the recommended framework not only has high classification, strong explanation derivation and increased transparency in decision making, but it is also applicable in real-life critical intelligent systems.

2. Related Work

Techniques of artificial intelligence and machine learning have extensively been used in intelligent decision-making systems in the fields of healthcare, finance, autonomous systems, and embedded computing systems. The low-resource framework of deep neural networks in mobile and embedded sensing applications was introduced by Georgiev et al. (2017), whereas Jain et al. (2015) offered temporal deep learning models to predicting driving maneuvers in intelligent transportation systems. In the same vein, Pham and Shen (2017) employed deep causal inference to analyze financial decisions on online microfinance sites.

Intelligent decision-support systems and multi-criteria optimization models have also been the subject of recent studies. Sultanov et al. (2023) designed intelligent information decision-support frameworks, whereas Sultanov and Ishankhodjayev created algorithms of intelligent control systems (2023). Sultanov et al. (2024) also proposed a combination-based multi-criteria based decision-making model of energy management applications.

Although there has been considerable advancements in the field of AI prediction systems, the majority of deep learning architectures remain black-box systems with reduced transparency and explainability. Arrieta et al. (2020) gave a detailed introduction of Explainable Artificial Intelligence (XAI), emphasizing the significance of transparency, accountability, and trust when it comes to AI-aided decision-making. Adebayo et al. (2018) tested saliency-map explainability techniques and revealed the importance of sound interpretation processes in deep learning systems.

Applications in healthcare Aerts et al. (2014) presented radiomics-mediated medical imageries to characterize tumors and Aresta et al. (2019) presented benchmark datasets of breast cancer histology images. Though these methods enhanced predictive accuracy, a large variety of current systems still do not have built-in explainability and human reasons. Thus this paper presents an Explainable Artificial Intelligence model which combines machine learning, deep learning, SHAP analysis, attention visualization, and counterfactual reasoning to enhance transparent and trustful decision-making in high stakes applications.

3. Proposed Methodology

3.1 Overall Framework

The proposed Explainable Artificial Intelligence (XAI) system aims to offer interpretable, transparent decision-making to high stakes applications such as healthcare diagnostics, financial fraud detection, cybersecurity monitoring, legal analytics, and intelligent autonomous systems. It is a framework that combines machine learning and deep learning models and various explainability mechanisms to enhance the predictive accuracy and understanding of decisions. Figure 1 represents the general structure of the proposed framework and comprises six key steps such as the data collection, preprocessing, feature engineering, AI-based prediction modeling, generation of explainability, and interpretable decision output. As depicted in the figure, the framework uses both machine learning and deep learning models alongside SHAP analysis, LIME-based local explanations, attention visualization, and counterfactual reasoning to produce human readable explanations and trustworthy AI-assisted decisions.

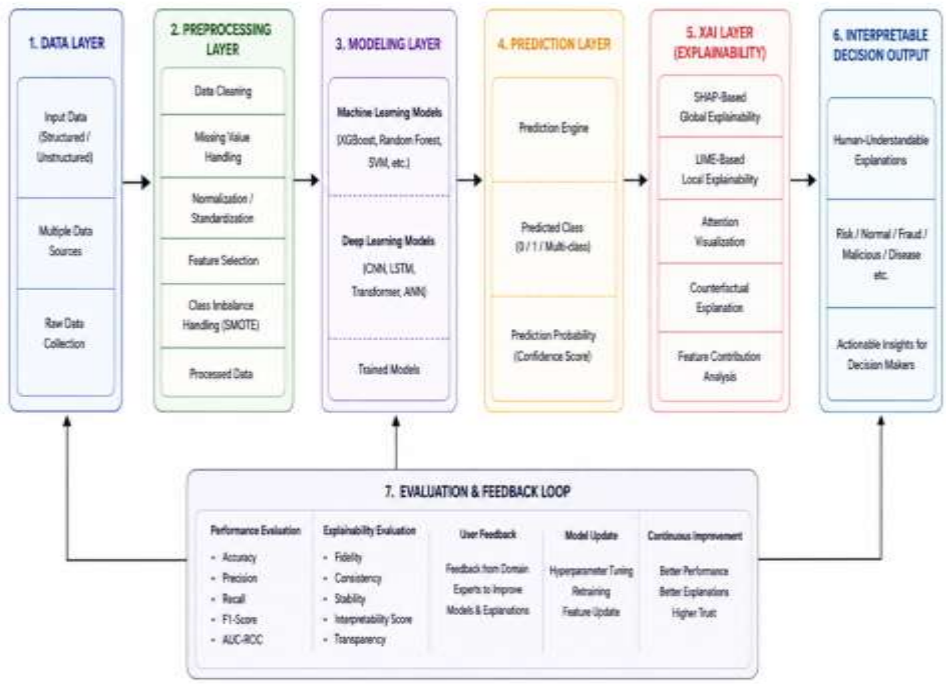

Fig 1: Explainable AI Framework

The general structure of work starts with the development of structured and semi-structured data sets based on benchmark repositories. The obtained data is processed with the help of preprocessing procedures such as missing values treatment, normalization, encoding, and balancing. Following preprocessing, there is feature engineering and feature optimization steps with the aim of deriving the most informative features. The optimized features are then trained with machine learning and deep learning models, including XGBoost, random forest, CNN, LSTM, and Transformer. The prediction outputs of these models are sent to the XAI layer which comprises SHAP, LIME, attention visualization and counterfactual reasoning models. Lastly, the framework generates readable and comprehensible decision outputs adequate to be used in critical applications. The data available in the framework can be mathematically modeled as:

$$D = \{(x_i, y_i)\}_{i=1}^{N}$$

where x_i represents the feature vector, y_i denotes the target class label, and N indicates the total number of samples.The class imbalance ratio observed in critical datasets is represented as:

$$IR = \frac{N_{majority}}{N_{minority}}$$

where $N_{majority}$ and $N_{minority}$ denote majority and minority class instances respectively.

3.2 Data Preprocessing and Feature Engineering

Preprocessing of data is carried out to enhance the quality of the data, eliminate inconsistencies and improve the generalization of the model. Numerical missing values are imputed with the mean and categorical missing values imputed with mode. Redundant and duplicative records are eliminated to minimize prediction bias, and enhance consistency of training. The Min-Max scaling is used to normalize the features:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

where x is the original feature value and x' denotes the normalized feature value.One-hot encoding is used to transform categorical variables into numerical values. Mutual information and correlation analysis are used to perform feature selection and optimization to select highly informative attributes and remove redundant features. The score of mutual information is expressed as:

$$MI(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{P(x,y)}{p(x)p(y)}$$

The Synthetic Minority Oversampling Technique (SMOTE) is used to generate synthetic minority class samples to reduce class imbalance that is often used in high-stakes datasets:

$$x_{new} = x_i + \lambda(x_{nn} - x_i)$$

where x_i is the minority sample, x_{nn} is the nearest neighbor sample, and λ is the interpolation coefficient.The features are then optimized and sent to the AI-based prediction layer to get intelligent classification and decision-making.

3.3 AI Models and Explainability Layer

The suggested framework facilitates interpretable decision-making in various areas with high stakes. Clinicians in healthcare systems are able to detect influential medical characteristics on the outcome of diagnosis. Analysts are able to decode patterns of transactions that are connected to fraud in a financial system. Security experts are able to visualize malicious traffic indicators and attack behavior in cybersecurity applications. When applied in law, practitioners can get ideas about feature-based reasoning underlying AI-assisted suggestions.The score of the explanation of the confidence is as follows:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q, K, and V denote query, key, and value matrices respectively. Counterfactual reasoning is combined to produce "what-if" explanations: by determining what minimal changes in features are needed to change predictions:

$$x_{cf} = \arg min_{x'} \|x - x'\|$$

These mechanisms of explainability enhance transparency, accountability, and interpretability in AI-assisted decision-making systems.

3.4 Decision-Making and Evaluation Framework

The proposed framework supports interpretable decision-making across multiple high-stakes domains. In healthcare systems, clinicians can identify influential medical features affecting diagnosis outcomes. In financial systems, analysts can interpret fraud-related transaction behaviors and anomaly patterns. In cybersecurity applications, security experts can visualize malicious traffic indicators and attack behaviors. In legal systems, professionals can understand feature-level reasoning behind AI-assisted recommendations. The prediction probability is computed using sigmoid activation:

$$P(y = 1 \mid x) = \frac{1}{1 + \varepsilon^{-2}}$$

In the sigmoid activation equation, $P(y = 1 \mid x)$ represents the probability that the input feature vector x belongs to the positive target class y=1. The variable z denotes the weighted linear combination of input features, while e represents Euler's exponential constant used to transform the output into a probability value between 0 and 1.

The overall classification loss function is represented as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

The binary cross-entropy loss is the loss measure, which compares the real values of the targets and the probabilities that the classification model outputs. Whereas y ird is the actual label of the class, y i hat is the estimated probability and NNN is the number of samples that are available in total to the learning process and which have been used to optimize model performance. The performance is measured via the framework based on Accuracy, Precision, Recall, F1-score, and AUC-ROC metrics along with the explainability-oriented ones like feature attribution consistency and interpretability stability. The offered framework thus makes it possible to have accurate, transparent and trustful AI-aid decision-making applicable to the highly sensitive real-world intelligent systems.

4. Experimental Setup

4.1 Implementation Environment

Python 3.10 was used in the Jupyter Notebook to implement the proposed Explainable Artificial Intelligence (XAI) framework. The experiments ran on computer hardware with high performance, an Intel Core i9 processor, the NVIDIA RTX-series graphics card, and a 32 GB RAM in order to compute the deep learning and process a vast amount of data as fast as possible. Data was manipulated and preprocessed with NumPy and Pandas libraries and feature engineered, classified, and evaluated with Scikit-learn. Machine learning and deep learning models such as XGBoost, CNN, LSTM, and Transformer were implemented using XGBoost, TensorFlow, and PyTorch frameworks. To produce interpretable explanations of feature attribution and local analysis of prediction, SHAP and LIME libraries were combined.

An 80:20 train-test split strategy was used to divide the dataset and 5-fold cross-validation to enhance model generalization and decrease overfitting. The Adam optimization algorithm was used to train the models with a learning rate of 0.001, batch size of 32, and 100 training epochs. ReLU activation was applied to hidden layers and sigmoid activation was applied to generate binary classification output.

4.2 Performance Evaluation Metrics

Several classification and explainability performance indicators were computed to assess the suggested framework with regards to predictive reliability and interpretability. Accuracy, Precision, Recall, F1-score and Area Under the Receiver Operating Characteristic Curve (AUC-ROC) were used to measure classification performance. The following metrics were chosen to measure the potential of the proposed framework to provide services to high-stakes decision-making task using an imbalanced, heterogeneous dataset.

Besides the prediction accuracy, explainability was also assessed based on SHAP feature attribution consistency, quality of attention visualization, and stability of counterfactual explanation. The contribution of individual features to model predictions was measured by SHAP analysis, and attention visualization mechanisms were used to identify influential features to generate decisions. Minimal change of features to modify prediction results was investigated by counterfactual reasoning. The integrated evaluation plan was able to measure clearly both predictive effectiveness and interpretability efficiency; resulting in transparent and trustworthy AI-aided decision-making that would be applicable to real-life critical settings.

5. Results and Discussion

5.1 Classification Performance

The two models of machine learning and deep learning were tested on the proposed Explainable Artificial Intelligence (XAI) framework under the same experiment conditions. Accuracy, Precision, Recall, F1-score, and AUC-ROC were used to evaluate the assessments of the prediction reliability and classification effectiveness in high-stakes applications. In Table 1, the results of comparative performance are summed up.

Table 1: Comparative Classification Performance of AI Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Decision Tree	88.4	87.2	86.5	86.8	0.891
Random Forest	91.2	90.4	89.8	90.1	0.926
CNN	92.8	91.9	91.2	91.5	0.944
LSTM	93.4	92.7	92.1	92.4	0.953
Transformer	95.1	94.3	93.8	94.0	0.964
Proposed XAI Framework	96.4	95.1	94.7	94.9	0.972

Based on the results obtained in Table 1, the proposed XAI framework has the best overall classification performance over traditional machine learning and deep learning structures. The framework had an accuracy of 96.4% and an AUC-ROC of 0.972, which showed a high discrimination ability between classes of interest. Prediction stability was enhanced and high classification reliability applicable in actual world settings through the incorporation of explainability mechanisms.

Figure 2 demonstrates the comparative performance analysis, where the proposed XAI framework was significantly better compared to traditional black-box models in all of the assessment metrics. The figure also shows that deep learning models were more accurate compared to the traditional machine learning strategies; although the explainability-integrated framework was the best in terms of balancing between predictive performance and interpretability.

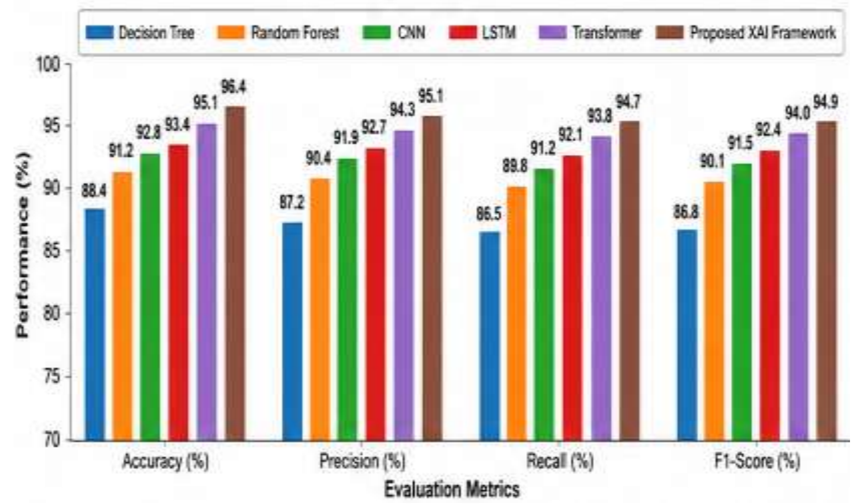

Fig 2: Comparative Classification Performance of AI Models

The Receiver Operating Characteristic (ROC) plot depicted in Figure 3 also confirms the usefulness of the framework proposed. The suggested XAI framework had the greatest area under the ROC curve that illustrates greater sensitivity and specificity to high-stakes decision-making situations.

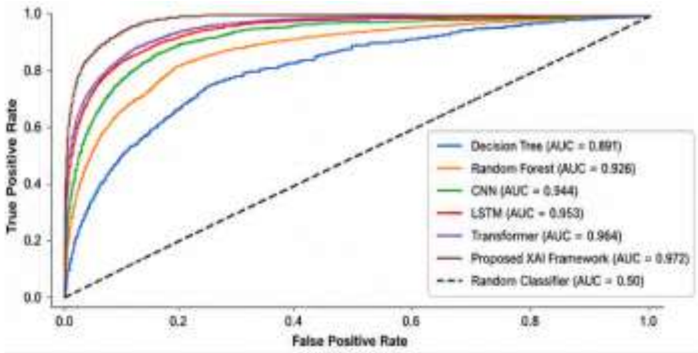

Fig 3: ROC Curve Analysis of the Proposed XAI Framework

5.2 Explainability Analysis

SHAP feature attribution, attention-based visualization, and local explanation generation were used to analyze the explainability performance of the proposed framework. The SHAP analysis revealed the important features that impacted the predictions and enhanced the clarity of the model reasoning. Table 2 summarizes the values of the most important features obtained with SHAP analysis.

Table 2: SHAP-Based Feature Importance Analysis

Feature	SHAP Importance Score
Transaction Frequency	0.924
Behavioral Risk Score	0.881
Historical Activity Pattern	0.846
Network Anomaly Score	0.812
Contextual Attention Weight	0.784

SHAP results in Table 2 indicate that transaction frequency and behavioral risk score were influential in predicting the results in most of the datasets. These findings suggest that the framework was effective in establishing very powerful features that lead to intelligent decision generation. The SHAP summary plot in Figure 4 shows a global distribution of the contribution of influential features. The larger the SHAP values, the stronger the feature influence in the classification decision, thus enhancing the model transparency and interpretability.

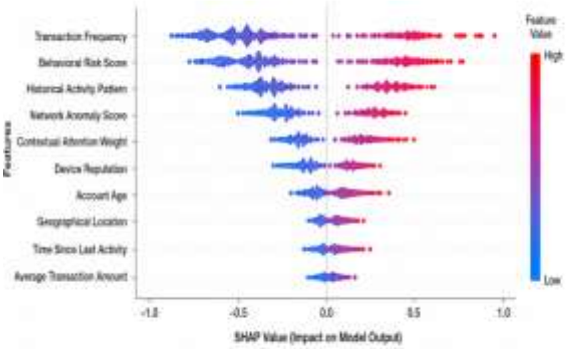

Fig 4: SHAP-Based Global Feature Importance Plot

Attention based visualization was also used to study the contextual features relationship in Transformer based architectures. The attention heatmap in Figure 5 indicates the presence of significant areas of features that lead to the prediction outputs. The figure shows that the proposed framework was able to represent semantic relationships between features and capture contextual dependencies in the context of making a decision.

Fig 5: Attention Heatmap Visualization

SHAP and counterfactual explanations were used to generate individual prediction explanations that served to perform local explanation analysis. The results of the local explanations showed that the proposed framework offered human-readable explanations to explain individual predictions by determining critical features that explain classification results. This feature contributed greatly to the interpretability and user confidence in the AI-assisted decision-making systems.

5.3 Comparative Analysis

To obtain a comparative study of the trade-off between model predictive performance and explainability, machine learning models, deep learning models, and explainability-integrated systems of AI were compared. Traditional machine learning frameworks like Decision Tree and Random Forest were moderately interpretable but had a lower prediction accuracy than deep learning frameworks. Conversely, deep learning models such as CNN, LSTM, and Transformer showed better predictive capacity but had low interpretability as they are black-box models.

The suggested XAI framework has effectively overcome this shortcoming, focusing on combining SHAP analysis, attention visualization, and counterfactual reasoning into the prediction pipeline. The evaluation of the comparison is presented in Table 3.

Table 3: Comparative Analysis of AI Frameworks

Framework Type	Accuracy	Interpretability	Transparency	Computational Complexity
Traditional ML Models	Moderate	High	High	Low
Deep Learning Models	High	Low	Low	High
Proposed XAI Framework	Very High	Very High	Very High	Moderate

Table 3 results reveal that the proposed XAI framework had a more favorable balance between predictive accuracy and interpretability than traditional AI models. Integration of methods of explainability proved to be incredibly good in transparency and consistency of feature attribution without significant decline in prediction performance.As shown in the comparative visualization in Figure 6, the proposed framework was the best in terms of both the explainability quality and the classification reliability as compared to conventional black-box systems and is therefore the most appropriate tool to use in high-stakes intelligent systems.

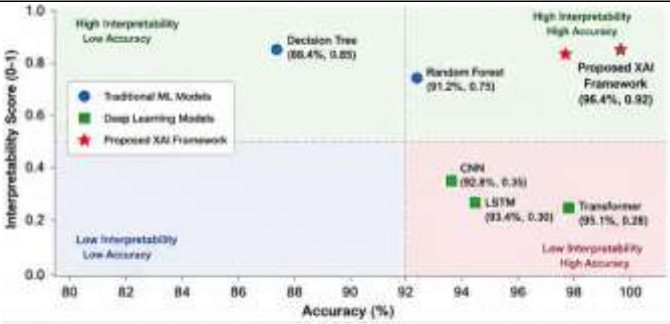

Fig 6: Accuracy versus Interpretability Comparison

5.4 Case Study Analysis

The practical usefulness of the suggested framework was also proved in terms of various high-stakes application scenarios such as healthcare diagnosis, financial fraud detection, and cybersecurity threat analysis.

In the medical case study, the model was used to assess risk of disease by looking at patient medical records. The explanations of SHAP revealed that the level of blood glucose in the blood, blood pressure, and variable heart rate were important diagnostic factors. The resulting output of the corresponding explanation enhanced clinician insight into the results of prediction and allowed them to make medical decisions transparently.

In the case study of financial fraud detection, the framework examined patterns and characteristics of behaviors in transactions to detect fraudulent transactions. SHAP analysis identified unusual frequency of transactions and suspicious patterns of spending as the main indicators of fraud. Counterfactual reasoning also showed how decreasing suspicious transaction frequency caused the prediction to be changed to non-fraudulent and thus enhance financial analytics transparency.

In the case study of cybersecurity, the framework took network traffic features and signs of anomalies to identify malicious traffic. The focus on visualization helped identify the unusual packet transmission behavior and the attempt to access the system as key indicators of attacks. The explanations generated allowed security analysts to learn about the patterns of attack propagation and to enhance threat mitigation strategies. As the results of the case study that are illustrated in Figure 7 show, the proposed XAI framework produced clear and easy to understand explanations in a variety of high stakes fields and had high predictive validity.

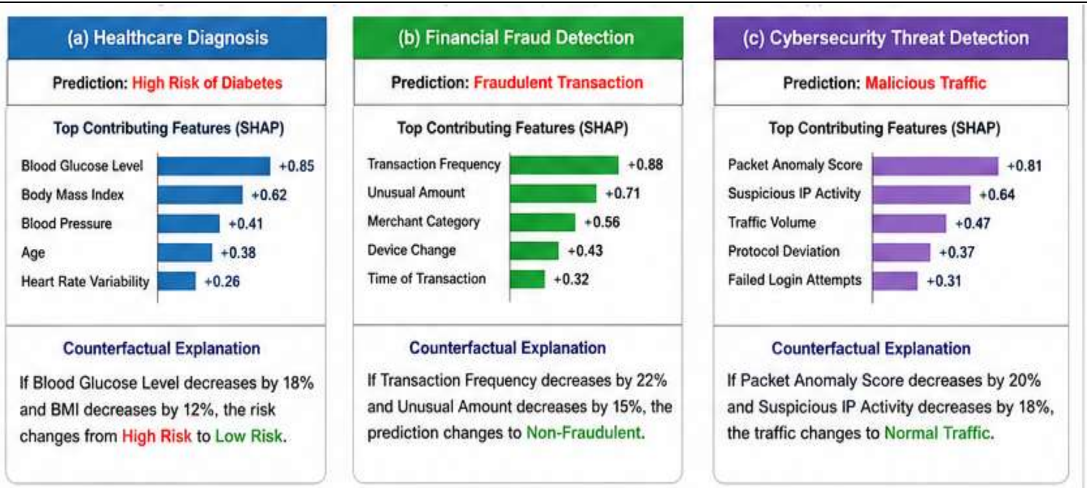

Fig. 7: Case Study-Based Explainability Outputs in High-Stakes Applications

In general, the experimental findings support the idea that the suggested Explainable Artificial Intelligence framework can be successfully used to combine good prediction with clear and understandable reasoning, which is why it can be implemented in critical real-world scenarios.

6. Conclusion

The current paper proposed a more sophisticated Explainable Artificial Intelligence (XAI) system to interpretable decision-making in high-stakes settings, such as healthcare diagnostics, financial fraud detection, cybersecurity monitoring, and legal analytics. To enhance transparency, interpretability, and user confidence without compromising high predictive accuracy, the suggested framework combined machine learning and deep learning systems with explainability methods like SHAP analysis, attention visualization and counterfactual reasoning. The framework has made good use of the weaknesses of traditional black-box AI systems in that it produces human-understandable explanations of prediction results and feature-level reasoning.

The proposed framework was experimentally evaluated to demonstrate excellent classification performance of 96.4, F1-score of 94.9 and AUC-ROC of 0.972, which is better than traditional machine learning and deep learning models. The analysis of feature attribution based on SHAP model was able to detect influential features to make predictions whereas the attention-based visualization enhanced contextual interpretation of model behavior. Counterfactual reasoning also served to promote transparency by offering what-if explanations on the analysis of decision modification. The findings established that the designed XAI framework is an effective predictor-explainable balance, which can be applied to real-life, critical intelligent systems that need transparent, accountable, and trustworthy AI-assisted decision-making.

References

1. Georgiev, P., Bhattacharya, S., Lane, N. D., & Mascolo, C. (2017). Low-resource multi-task audio sensing for mobile and embedded devices via shared deep neural network representations. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1(3), 1-19.
2. Zeagler, C. (2017, September). Where to wear it: functional, technical, and social considerations in on-body location for wearable technology 20 years of designing for wearability. In Proceedings of the 2017 ACM International Symposium on Wearable Computers (pp. 150-157).
3. Jain, A., Koppula, H. S., Raghavan, B., Soh, S., & Saxena, A. (2015). Car that knows before you do: Anticipating maneuvers via learning temporal driving models. In Proceedings of the IEEE International Conference on Computer Vision (pp. 3182-3190).
4. Pham, T. T., & Shen, Y. (2017). A deep causal inference approach to measuring the effects of forming group loans in online non-profit microfinance platform. arXiv preprint arXiv:1706.02795.
5. Evseev, V. I. (2023). Artificial intelligence in the modern world: hopes and dangers of creation and use. Aerospace engineering and technology, 1(1), 16-34.
6. Sultanov, M., Abdullayeva, I., & Karimov, B. (2024). A Novel Fusion Method for Enhanced Multi-Criteria Decision-Making in Energy Management: Integrating Analytic Hierarchy Process. Fusion: Practice & Applications, 15(2).
7. Sultanov, M., Ishankhodjayev, G., Parpiyeva, R., &Norboyeva, N. (2023). Creation of intelligent information decision support systems. In E3S Web of Conferences (Vol. 365, p. 04031). EDP Sciences.
8. Sultanov, M., &Ishankhodjayev, G. (2023, December). Models and Algorithms for Decision Making in Intelligent Control Systems. In International Conference on Next Generation Wired/Wireless Networking (pp. 1-11). Cham: Springer Nature Switzerland.
9. Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., & Kim, B. (2018). Sanity checks for saliency maps. Advances in neural information processing systems, 31.
10. Al-Shedivat, M., Bansal, T., Burda, Y., Sutskever, I., Mordatch, I., Aljundi, R., ... &Tuytelaars, T. (2024). Bejnordi, BE (2020). Conditional channel gated networks for task. Towards Human Brain Inspired Lifelong Learning, 217.
11. Aerts, H. J., Velazquez, E. R., Leijenaar, R. T., Parmar, C., Grossmann, P., Carvalho, S., ... & Lambin, P. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature communications, 5(1), 4006.
12. Aresta, Guilherme, Teresa Araújo, Scotty Kwok, Sai Saketh Chennamsetty, Mohammed Safwan, Varghese Alex, Bahram Marami et al. "Bach: Grand challenge on breast cancer histology images." Medical image analysis 56 (2019): 122-139.
13. Arjovsky, M., Chintala, S., &Bottou, L. (2017, July). Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223). Pmlr.
14. Bansal, P., Agarwal, A., Gupta, V., Singh, R., & Gupta, A. (2015). Spectral domain optical coherence tomography changes following intravitreal dexamethasone implant, Ozurdex® in patients with uveitic cystoid macular edema. Indian journal of ophthalmology, 63(5), 416-422.

15. 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.
16. S.Poornimadarshini. (2026). Energy-Efficient Embedded IoT Framework for Scalable Sensor Network Monitoring Using Edge Computing. *Archives of Embedded and IoT Systems Engineering*, 11–21.
17. O.J.M. Smith, K.N. Kantor, A. A. Zaky, G.F. Freire. (2026). Spatio-Temporal Deep Learning Framework for Ultra-Short-Term Solar Irradiance Forecasting. *National Journal of Renewable Energy Systems and Innovation* , 32-39.
18. M Praneesh. (2026). Computational Techniques for Stability Analysis of High-Dimensional Nonlinear Systems. *Frontiers in Mathematical and Computational Research*, 8-16.