International Journal of Advance and Innovative Research
Volume 13, Issue 1 (IV): January - March 2026
ISSN 2394 - 7780
JVM's Mehta Degree College, Sector 19, Airoli
NAAC Re-accredited "A+" Grade
IQAC in association with Western Regional Centre, ICSSR Organized one day National Conference on "Integrating Multidisciplinary Approaches to Build a Resilient and Sustainable Future", held on 10th January 2026

EXPLAINABLE AI FOR DISEASE DIAGNOSIS

Shivani Sunil Pandey and Sarita Sarang
[1]M.Sc. IT Student and [2]Assistant Professor, Jnan Vikas Mandal's College, Airoli, Navi-Mumbai (MH), India

ABSTRACT
Artificial Intelligence (Al) has emerged as a powerful support system in modem healthcare particularly for early detection and diagnosis of diseases. While machine learning models often achieve high predictive accuracy, their decision-making processes are usually opaque, leading to a lack of trust among healthcare professionals. This study focuses on Explainable Artificial Intelligence (XAI), which aims to make AI-driven predictions transparent and understandable. The proposed work explores the use of SHAP (SHapley Additive exPlanations) to interpret disease diagnosis predictions, with a specific emphasis on heart disease. SHAP explains how individual medical attributes such as age, blood pressure, cholesterol level, and chest pain type influence the final prediction. By integrating explainability with predictive modeling, the proposed approach enhances transparency, reliability, and acceptance of AI systems in healthcare decision-making.

Keywords: *Artificial Intelligence, Explainable AI, XAI, SHAP, Disease Diagnosis, Healthcare*

1. INTRODUCTION

Healthcare institutions generate vast amounts of medical data that can be effectively utilized using artificial intelligence techniques. Machine learning models are increasingly applied in disease diagnosis due to their ability to detect complex patterns within data.

Despite their strong performance, many of these models function as black-box systems, offering predictions without revealing the reasoning behind them. In medical applications, such lack of interpretability can limit clinical adoption.

Explainable Artificial Intelligence (XAI) addresses this challenge by providing insights into how AI models reach their conclusions. XAI methods enable medical practitioners to understand, evaluate, and trust AI-assisted predictions. This research presents an explainable AI-based framework for disease diagnosis using SHAP to improve transparency and support informed decision-making in healthcare environments.

2. PROBLEM STATEMENT

Artificial inte11igence models are widely used for disease prediction because of their high accuracy. However, most traditional machine learning approaches do not provide explanations for their outputs, making them difficult to trust in critical domains such as healthcare. The absence of interpretability raises concerns related to trust, ethics, and reliability for both doctors and patients. Medical professionals require AI systems that not only deliver accurate predictions but also clearly explain the factors influencing those predictions, particularly in heart disease diagnosis.

3. OBJECTIVES

The key objectives of this research are:

1. To develop an efficient machine learning model for predicting heart disease.
2. To incorporate Explainable Artificial Intelligence (XAI) techniques for interpreting model outputs.
3. To utilize SHAP (SHapley Additive exPlanations) to determine feature importance.
4. To examine the influence of medical parameters such as age, cholesterol, blood pressure, and chest pain on disease prediction.
5. To enhance transparency, trust, and acceptance of AI-based healthcare systems.

4. LITERATURE REVIEW

Several studies have investigated the use of machine learning algorithms such as Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines for heart disease prediction. These techniques have demonstrated strong predictive capabilities but often fail to provide meaningful explanations for their decisions.

Recent research emphasizes the importance of Explainable AI in medical applications. Explanation methods such as LIME and SHAP have been introduced to interpret model behavior. Among these, SHAP is widely preferred due to its strong theoretical foundation in game theory and its ability to generate both global and local explanations. Previous studies confirm that SHAP effectively identifies critical risk factors in heart disease, assisting clinicians in understanding and validating AI-driven predictions. However, limited research has focused on integrating SHAP explanations into user-friendly healthcare decision support systems, which this study aims to address.

5. METHODOLOGY

The proposed methodology involves training a machine learning model using a medical dataset containing patient attributes such as age, blood pressure, cholesterol level, and chest pain type. A Random Forest classifier is employed for disease prediction due to its robustness and strong performance with structured medical data. SHAP is then applied to interpret the model's predictions by calculating the contribution of each feature toward the final output. This approach allows both global feature importance analysis and individual patient-level explanations.

6. COMPARATIVE ANALYSIS

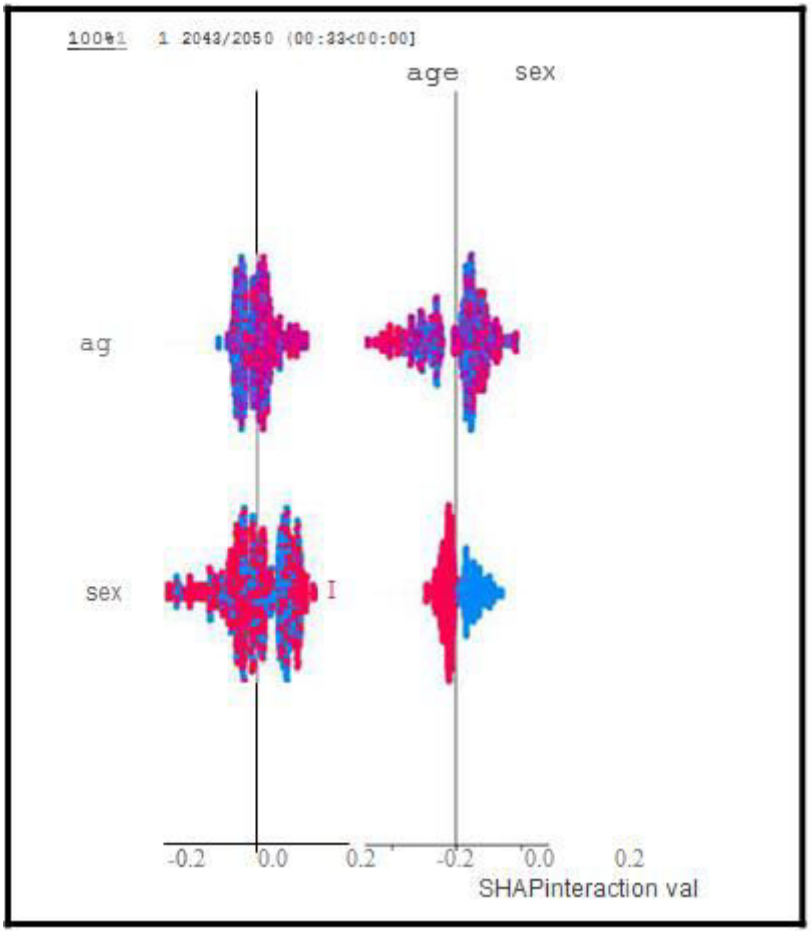
Technique	Accuracy	Interpretability	Reliability	Healthcare Suitability
Traditional ML (Black Box)	High	Low	Medium	Limited
LIME	Medium-High	Medium	Medium	Moderate
SHAP	High	High	High	Excellent

Traditional machine learning models offer strong accuracy but lack transparency. LIME provides local explanations but may produce inconsistent results. SHAP delivers stable, accurate, and comprehensive explanations, making it highly suitable for healthcare applications where interpretability is essential.

7. RESULTS AND DISCUSSION

The experimental evaluation demonstrates that the proposed model achieves high prediction accuracy while maintaining interpretability. SHAP analysis reveals that age, cholesterol level, and blood pressure are among the most influential factors in heart disease prediction. The explanations generated by SHAP allow clinicians to understand the reasoning behind each prediction, thereby increasing confidence in AI-assisted diagnostic systems. This combination of accuracy and transparency enhances the overall reliability of AI-based healthcare solutions.

```
0   import shap
    import pandas as pd
    from sklearn.ensemble import

    RandomForestclassifier data =pd.read

    :esv("heart.csv")

    X  data.drop(''target",  axis=1)
    y    data["target"]

    model= RandomForestclassifier(random_state=42)
    model.fit(XJ y)

    explainer= shap.Explainer(modelJ

    shap.summary_plot shap_values,
```



8. CONCLUSION

This study highlights the significance of explainable artificial intelligence in healthcare applications. By integrating SHAP with machine learning models, the proposed framework improves transparency and trust in disease diagnosis systems. Explainable AI not only supports accurate predictions but also enables healthcare professionals to make informed and reliable clinical decisions.

9. FUTURE WORK

Future research may extend this framework to other medical conditions and incorporate real-time explainability into clinical decision support systems. Additionally, combining multiple XAI techniques could further improve interpretability and robustness.

REFERENCES

[1] M. Mushtaq et al., "Causal and Explainable Machine Leaming Framework for Heart Disease Prediction using SHAP," Journal of Computing & Biomedical Informatics, 2025.

[2] "The role of explainable artificial intelligence in disease prediction: A systematic review," BMC Medical Informatics and Decision Making, 2025.

[3] "XAI-HD: An Explainable Artificial Intelligence Framework for Heart Disease Detection," Artificial Intelligence Review, 2025.

[4] P. Shah et al., "Predicting cardiovascular risk with hybrid ensemble learning and explainable AI," Scientific Reports, 2025.

[5] D. Tang et al., "Risk factor analysis for coronary artery disease using SHAP-based explainable machine learning," Scientific Reports, 2025.