

Explainable Artificial Intelligence (XAI) for Heart Disease Prediction

Sharfuddin Shamshuddin Khan¹, Shakila Siddavatam²

¹Master's Student, Department of Computer Science, Abeda Inamdar Senior college, India

²Head of Department, Department of Computer Science, Abeda Inamdar Senior college, India

Abstract

This research focuses on building an Explainable Artificial Intelligence (XAI) system that helps make machine learning-based heart disease prediction more transparent and trustworthy for medical professionals. Although modern AI models can achieve impressive accuracy in healthcare, many of them operate as “black boxes,” making it difficult for clinicians to understand why a particular prediction is made. This lack of clarity often limits their acceptance and use in real clinical settings.

To bridge this gap, the study presents a web-based diagnostic platform that combines powerful machine learning models—such as Random Forest and XGBoost—with well-established explainability techniques, namely SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). The models are trained using the widely used UCI Heart Disease Dataset, which includes important clinical features like age, blood pressure, cholesterol levels, and chest pain type.

The platform is designed to give doctors clear, real-time explanations for every prediction. It allows them to see not only how the model behaves overall but also why it produced a specific result for an individual patient. Through an intuitive Streamlit-based interface, clinicians can enter patient details, view heart disease risk predictions, and explore interactive visualizations such as SHAP summary plots, feature dependency graphs, and LIME-based local explanations.

In addition to measuring standard performance metrics like accuracy, F1-score, and AUC-ROC, the system is evaluated for its practical usefulness and interpretability through simulated clinical cases and feedback from healthcare professionals. Early testing shows that adding explainability significantly improves transparency, builds clinician confidence, and supports better-informed medical decision-making.

Overall, the study highlights that explainability is not just an optional feature but a critical requirement for the responsible and effective use of AI in sensitive medical applications. Future improvements could include extending the system to predict multiple diseases, integrating it with Electronic Health Records (EHRs), supporting mobile health platforms, and deploying it on the cloud to enable scalable and collaborative clinical use.

Keywords: Explainable AI (XAI), Heart Disease Prediction, SHAP, LIME, Interpretable Machine Learning, Clinical Decision Support, Healthcare Transparency, Model Explainability, Random Forest.

1. Introduction

1.1 Problem Statement

The adoption of Artificial Intelligence in healthcare has enabled powerful predictive tools for identifying complex conditions such as heart disease. These systems can process large volumes of clinical data and uncover patterns that support early diagnosis and risk assessment. While their predictive performance is often impressive, many of these models function as “black boxes,” providing outcomes without clear explanations of how decisions are made.

In high-stakes medical environments, this lack of transparency presents a serious challenge. Clinical decisions demand accountability, reasoning, and trust, as they directly impact patient lives. When AI systems fail to explain the factors influencing their predictions, clinicians may hesitate to rely on them, raising ethical concerns related to responsibility, patient safety, and informed consent. Without interpretability, even accurate predictions can become difficult to validate or justify within established medical practice.

Current AI-driven healthcare solutions tend to prioritize accuracy over explainability, leaving a critical gap between model performance and human understanding. This disconnect limits the integration of AI into routine clinical workflows, as doctors require clear, interpretable insights to support decision-making. Bridging this gap is essential for building trustworthy AI systems that complement clinical expertise, enhance confidence in AI-assisted diagnoses, and ensure responsible deployment in real-world healthcare settings.

2. Literature Review

Explainable Artificial Intelligence (XAI) has emerged as a vital area of research aimed at making complex machine learning models more understandable to human users. Methods such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) are widely used to provide post-hoc explanations that clarify how individual features influence model predictions. In high-stakes fields like healthcare, prior studies emphasize that such transparency is not optional but essential, driven by ethical responsibility and regulatory requirements, including the European Union’s “Right to Explanation.”

Research consistently shows that interpretable models foster greater trust among users, particularly clinicians who must justify and stand behind diagnostic decisions. When healthcare professionals can understand why an AI system produces a specific outcome, they are more likely to integrate it into their clinical reasoning and workflows.

Explainability also supports accountability, error detection, and informed decision-making, all of which are critical in patient-centered care.

Despite these recognized benefits, most existing XAI solutions remain largely theoretical or disconnected from real-world clinical practice. Few systems provide seamless, real-time explanations through intuitive interfaces designed specifically for medical professionals. This gap highlights the need for practical, userfriendly XAI implementations that not only explain model behavior but also align with the demands and constraints of everyday clinical environments.

3.Methodology (Development Process)

3.1 Research Design

This project adopts a design-and-development approach, gradually building a functional Explainable AI (XAI) system through a series of well-defined stages. These stages include preparing and cleaning the data, training reliable machine learning models, integrating explainability techniques, and designing an intuitive user interface. Each step is carefully refined to ensure the system works effectively as a whole rather than as isolated components.

The development process is strongly guided by user-centered design principles. Feedback from potential healthcare users is continuously incorporated to align the system with real clinical needs, workflows, and expectations. By keeping clinicians involved throughout the design process, the project aims to create a tool that is not only technically sound but also practical, meaningful, and easy to use in real-world medical settings.

3.2 Technical Research

An extensive evaluation of various machine learning models and explainability frameworks was conducted to determine the most appropriate components for the system. Random Forest and XGBoost were selected for their strong predictive performance, robustness, and ability to support meaningful interpretation of model behavior in clinical contexts. To enhance transparency, SHAP was chosen to provide global explanations by highlighting overall feature importance, while LIME was used to generate local, instancelevel explanations for individual predictions. The combination of these models and explainability techniques ensures a balanced approach that maintains high accuracy while delivering clear, clinically relevant insights.

4.Design and Implementation

4.1 System Architecture

- **Frontend:** A Streamlit-based web application that allows users to enter patient information, view heart disease risk predictions, and interact with visual explainability outputs.

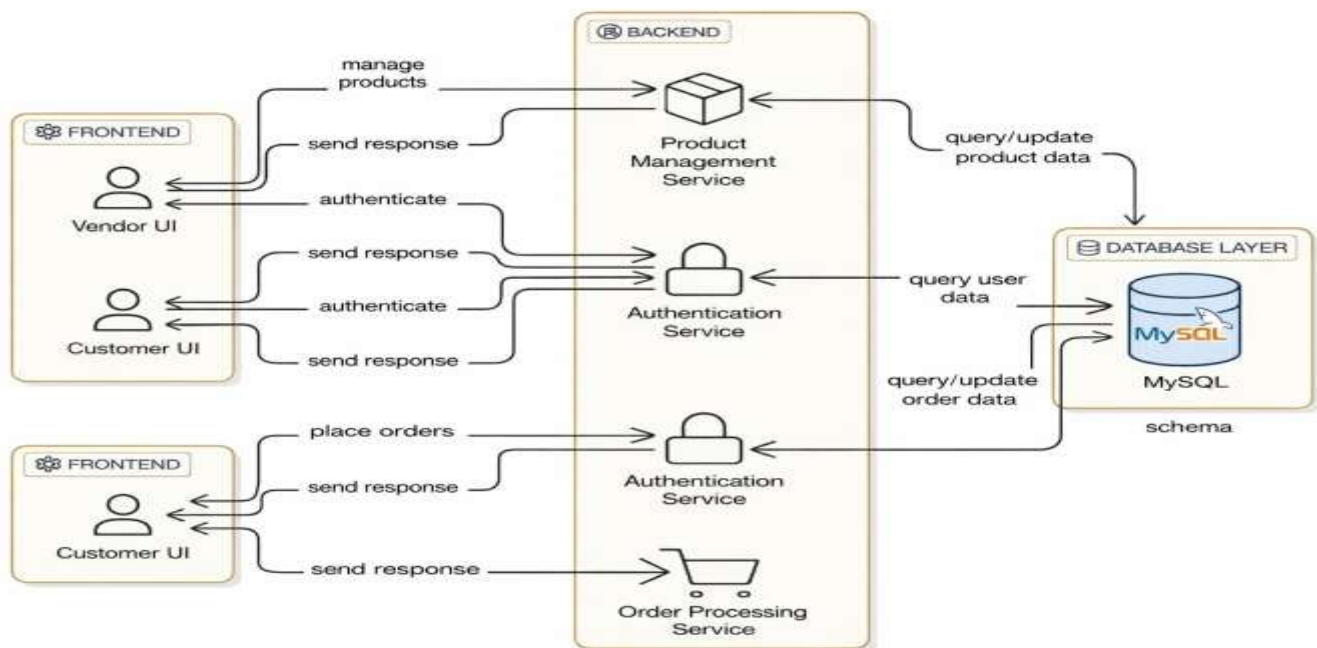
- **Backend:** Implemented in Python, utilizing Scikit-learn and XGBoost for model inference, along with SHAP and LIME to generate global and local explanations for predictions.
- **Database:** Uses SQLite or MySQL to store user session data, prediction results, and explanation logs for traceability and analysis.

• System workflow

1. The user (doctor) inputs patient health parameters via the web interface.
2. The backend processes the data through trained ML models (Random Forest/XGBoost).
3. SHAP and LIME generate feature importance scores and local explanations.
4. Results are displayed via interactive charts (summary plots, force plots, LIME tables).
5. All sessions and explanations are logged for review and audit.

• System Architecture Diagram

Figure 1: System Architecture of XAI



4.2 Technologies Used

The XAI utilizes a modern technology stack to achieve high performance, scalability, and secure operations, with each layer built using specialized technologies for optimal functionality.

Table 1: Technology Used

Component	Technology Used
Frontend	Streamlit, React.js, Next.js
Backend	Python, Flask, Scikit-learn, XGBoost
Database	MySQL
Explainability	SHAP, LIME
Visualization	Matplotlib, Plotly

4.3 User Interface (UI) & Screenshots

In everyday life, most people find medical reports difficult to understand, as they are usually written in technical language that only doctors can interpret. This application is being developed to bridge that gap, making heart disease predictions and explanations understandable not just for clinicians, but for ordinary users as well. By presenting insights in a simple, visual, and interactive way, the system helps anyone—patients or caregivers—grasp the results and understand the factors affecting their health.

4.3.1 User Interface Overview

There are a number of role-based interfaces in the system:

- **Homepage:** A welcoming landing page accessible to all users. Anyone can input health parameters and get real-time heart disease predictions immediately without registering. However, to save cases, view past prediction history, or access personalized dashboards, users need to register and log in. The homepage also highlights the platform's explainability features, helping users understand the predictions in a simple, visual way.
- **Registration Page:** Allows new users—whether medical professionals, researchers, or general users—to create an account by providing credentials and basic verification details. Registration is required to save predictions and track past cases.
- **Login Page:** A secure portal for all registered users to access their dashboards, saved predictions, and personalized insights.

User Dashboard:

- Input health parameters through a structured form.
- View **real-time heart disease risk predictions** with confidence scores.

- Explore interactive explainability visualizations, including SHAP summary plots, LIME local explanations, and feature importance charts.
- **Save and review past cases** for personal tracking, follow-up, or better understanding of their health patterns.

Screen

Figure 2: A web app dashboard that uses explainable AI to analyze uploaded medical reports and predict heart disease risk.

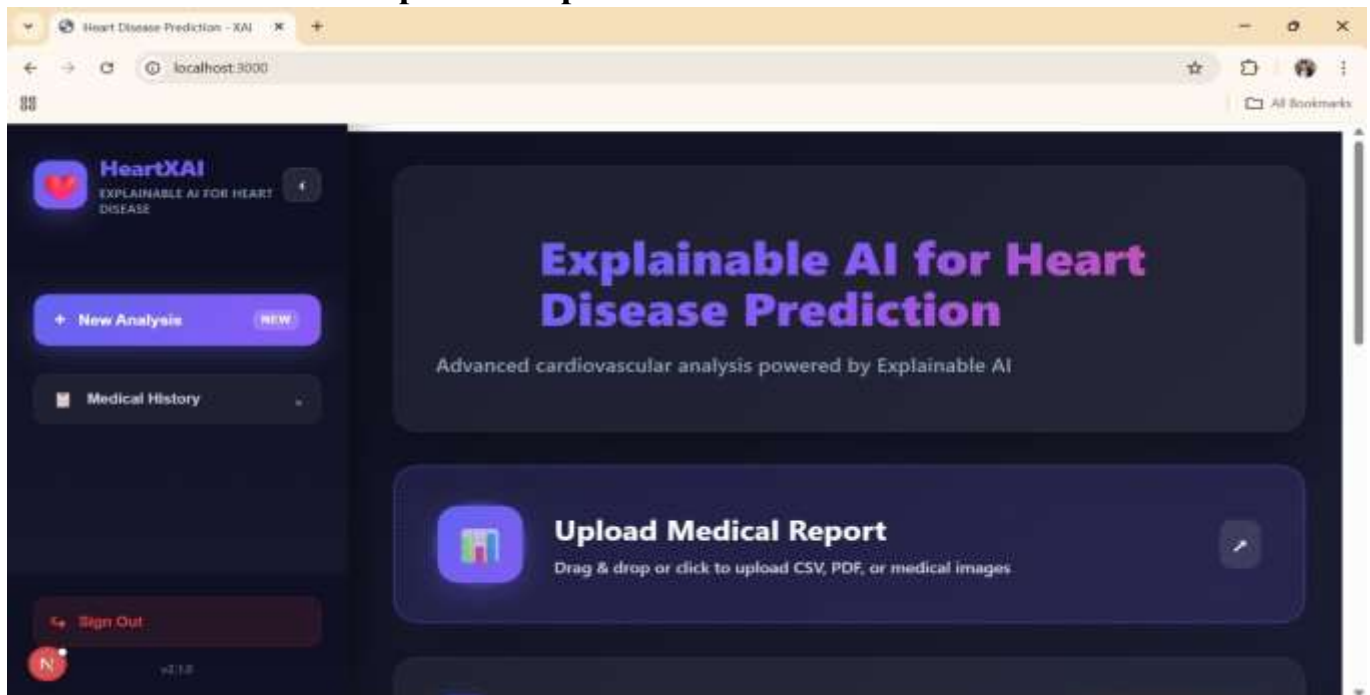


Figure 3: Web-based patient profile page for HeartXAI to collect health data for heart disease prediction.

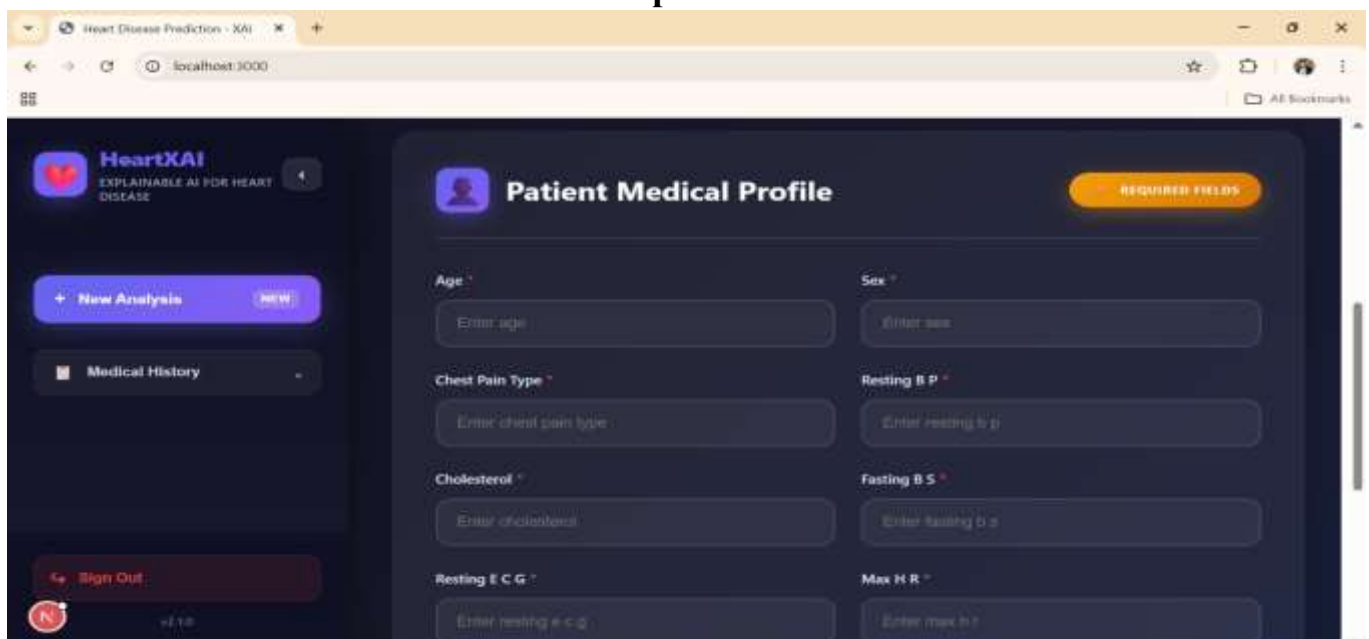
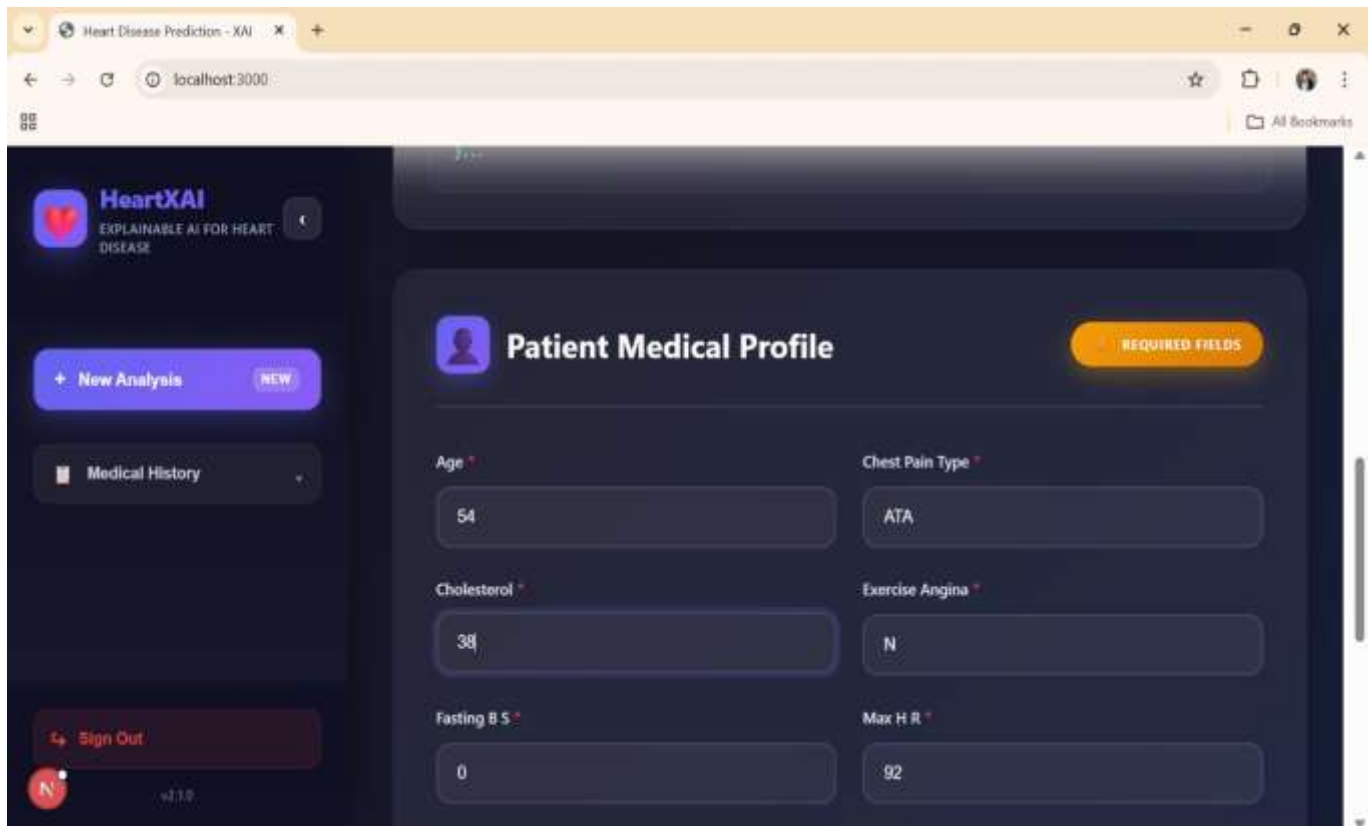


Figure 4: This screen shows the filled Patient Medical Profile, where extracted data is mapped into fields for final verification.



HeartXAI
EXPLAINABLE AI FOR HEART DISEASE

Patient Medical Profile REQUIRED FIELDS

Age ^{*} 54 Chest Pain Type ^{*} ATA

Cholesterol ^{*} 38 Exercise Angina ^{*} N

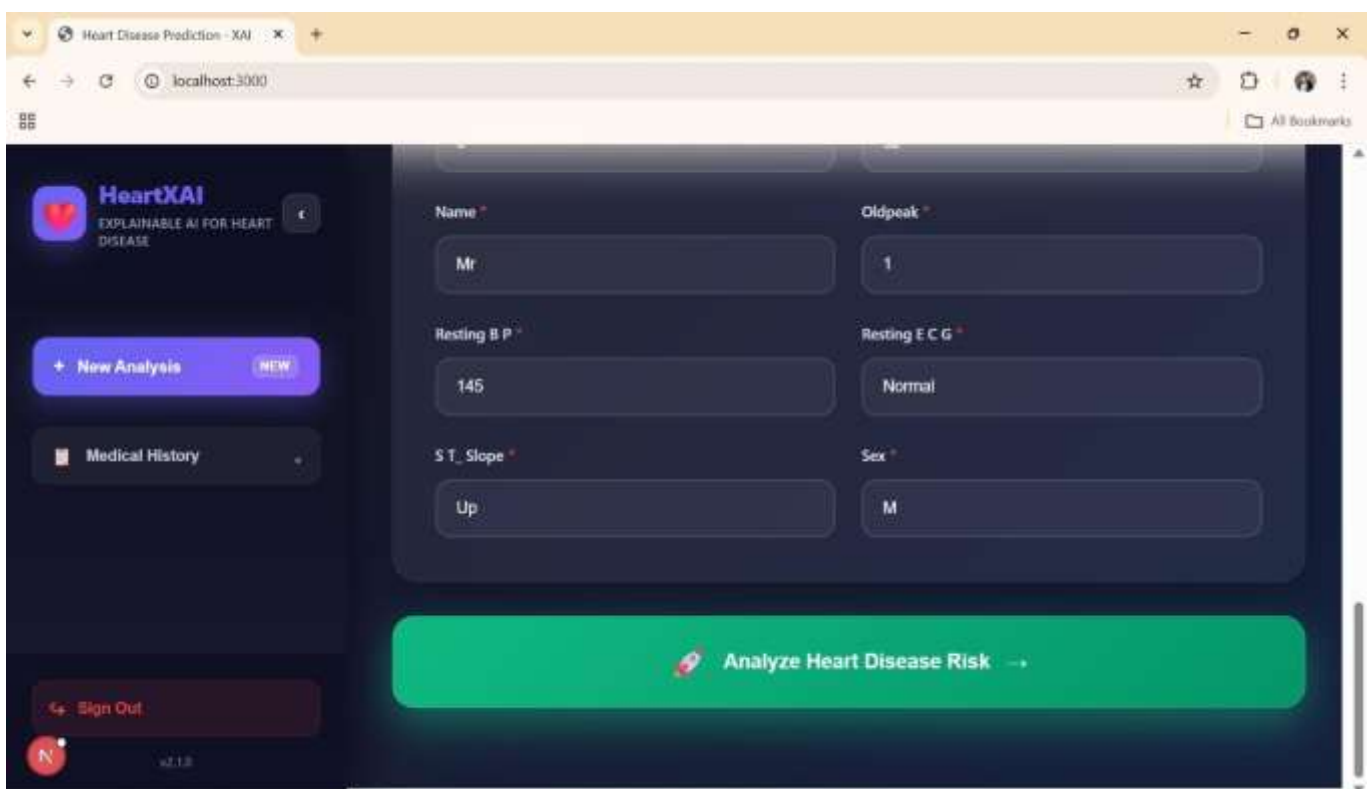
Fasting B.S ^{*} 0 Max H.R ^{*} 92

[New Analysis](#) NEW

[Medical History](#)

[Sign Out](#)

v2.1.0



HeartXAI
EXPLAINABLE AI FOR HEART DISEASE

Patient Medical Profile REQUIRED FIELDS

Name ^{*} Mr Oldpeak ^{*} 1

Resting B.P ^{*} 145 Resting E.C.G ^{*} Normal

S.T. Slope ^{*} Up Sex ^{*} M

[New Analysis](#) NEW

[Medical History](#)

[Sign Out](#)

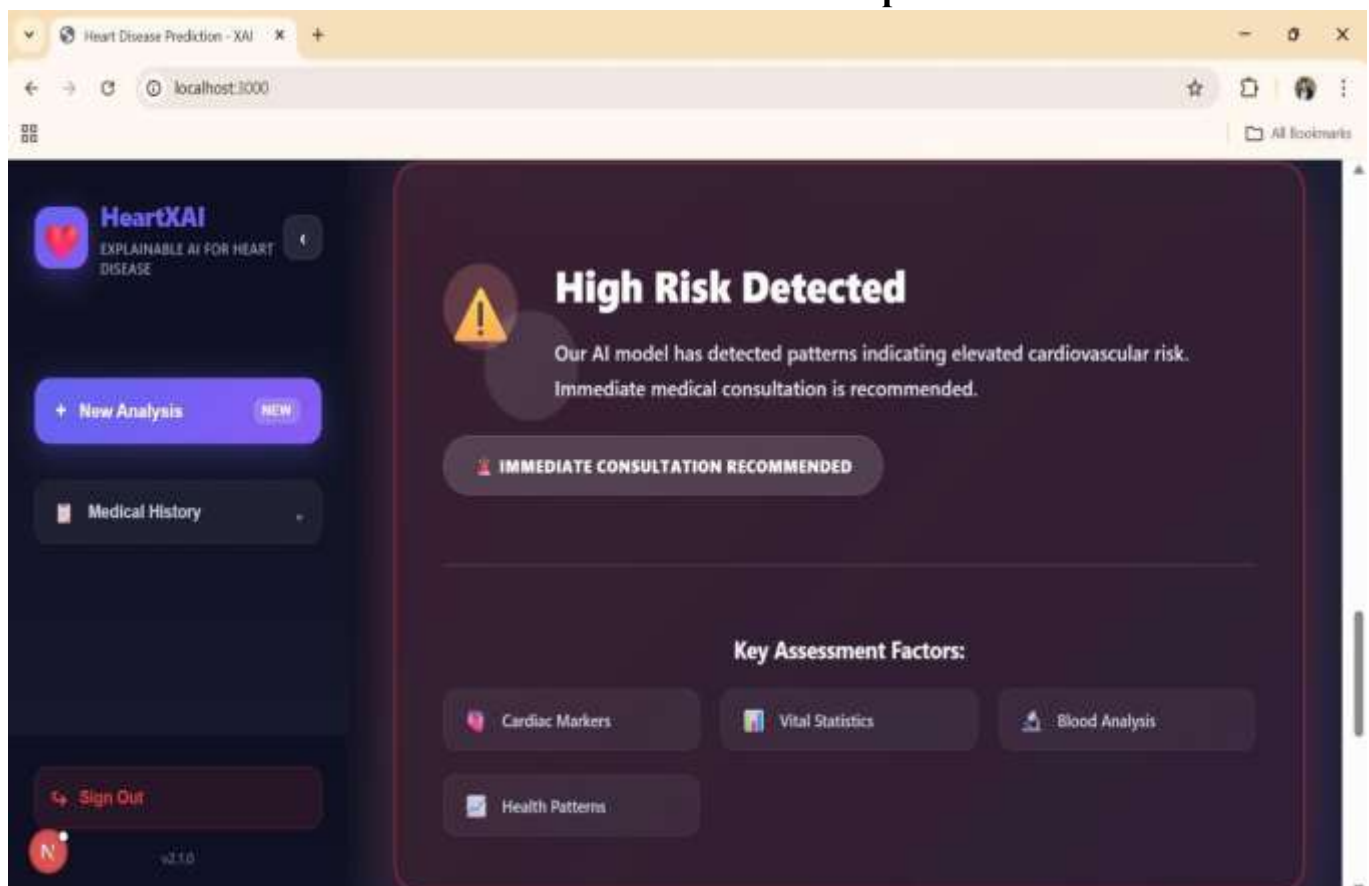
[Analyze Heart Disease Risk](#)

v2.1.0

Figure 5: This interface automatically extracts patient medical data from reports into JSON format for HeartXAI analysis.



Figure 6: This results screen shows a High Risk alert from the AI model with urgent medical consultation advice based on patient data.



6. Discussion

6.1 Strengths of the System

- **Transparency in Predictions:** The system provides clear, visual explanations for every prediction, helping users understand why a certain risk score is generated.
- **User-Friendly Interface:** Designed to be intuitive, the platform can be used easily by healthcare professionals and general users alike, even with minimal technical knowledge.
- **Real-Time Interpretation:** Users receive immediate feedback, with insights available both at the **global level** (overall feature importance) and **local level** (individual prediction explanations).
- **Ethical Compliance:** The platform supports responsible AI use by aligning with regulatory guidelines and best practices for explainable AI in healthcare.

6.2 Challenges and Limitations

- **Model Complexity:** Some explanations, particularly SHAP outputs, may still appear technical or difficult for non-expert users to interpret fully.
- **Data Dependency:** The accuracy and reliability of predictions depend heavily on the quality, diversity, and completeness of the training data.
- **Computational Overhead:** Certain explainability calculations, especially SHAP for large datasets, can be resource-intensive and may slow performance.

6.3 Future Scope

- **Integration with Electronic Health Records (EHRs):** To streamline clinical workflows and allow automatic data input.
- **Mobile Application Development:** Enabling on-the-go diagnostics and accessibility for users outside clinics.
- **Extended Disease Models:** Expanding the platform to predict other conditions such as diabetes, stroke, or kidney disease.
- **Cloud Deployment:** For better scalability, faster processing, and collaborative use across multiple locations.
- **Multilingual Support:** Making the platform accessible to users worldwide by supporting multiple languages.
- **Multi-Tasking Support:** Allowing users to run multiple patient predictions and analyses simultaneously, improving efficiency in busy clinical or research settings.

7. Conclusion

This research successfully developed an Explainable AI (XAI) system for heart disease prediction that combines high accuracy with clear, interpretable insights. By integrating SHAP and LIME into an intuitive Streamlit interface, the platform allows both healthcare professionals and general users to understand why a particular prediction is made, rather than just presenting a risk score. This transparency helps build trust in AI, enhances diagnostic confidence, and supports informed decision-making in clinical and personal health contexts. The system demonstrates that AI in healthcare can be both powerful and accountable, providing visual explanations that highlight the factors influencing predictions and enabling users to track changes over time. Overall, the project lays a foundation for future AI tools in healthcare that are accurate, ethical, and user-friendly, with potential enhancements such as multi-disease prediction, mobile access, and integration with electronic health records further bridging the gap between technology and human insight.

8. References

1. Ribeiro, Marco Tulio, Singh, Sameer, & Guestrin, Carlos (2016). "Why Should I Trust You?" *Explaining the Predictions of Any Classifier*. Proceedings of KDD. (Original paper introducing LIME). ([ResearchGate](#))
2. Lundberg, Scott M., & Lee, Su-In (2017). *A Unified Approach to Interpreting Model Predictions*. Advances in Neural Information Processing Systems (NeurIPS). (Original SHAP paper). ([ResearchGate](#))
3. Dey, Sudipta, & Roy Chowdhury, Tathagata (2024). *A Comparative Survey of SHAP and LIME: Explaining Machine Learning Models for Transparent AI*. International Journal of Innovative Research in Education, 11(6), 827–835. ([ResearchGate](#))
4. Roshinta, Trisna Ari, & Szűcs, Gábor (2024). *A Comparative Study of LIME and SHAP for Enhancing Trustworthiness and Efficiency in Explainable AI Systems*. IEEE International Conference on Computing (ICOCO). ([ResearchGate](#))
5. Hasan, Md. Mahmudul (2024). *Understanding Model Predictions: A Comparative Analysis of SHAP and LIME on Various ML Algorithms*. Journal of Scientific and Technological Research, 5(1), 17–26. ([jstr.bousst.edu.bd](#))
6. Ahmed, Shamim et al. (2024). *A Comparative Analysis of LIME and SHAP Interpreters with Explainable ML-Based Diabetes Predictions*. IEEE Access. ([ResearchGate](#))
7. Elsevier (2024). *Enhancing Histopathological Medical Image Classification for Early Cancer Diagnosis using Deep Learning and Explainable AI – LIME & SHAP*. Biomedical Signal Processing and Control. ([ScienceDirect](#))
8. Expert Systems with Applications (2025). *AEC: A Novel Adaptive Ensemble Classifier with LIME and SHAP-Based Interpretability for Fake News Detection*. ([ScienceDirect](#))
9. Souza, Vitor Amadeu (2026). *Comparative Analysis of LIME and SHAP in the Explanation of Sentiment Classification Models on Text Data*. ([ResearchGate](#))
10. Rao, Sannidhi et al. (2022). *A Study of LIME and SHAP Model Explainers for Autonomous Disease Predictions*. IEEE Bombay Section Signature Conference. ([ResearchGate](#))