



PERSONALISED HEALTHCARE RECOMMENDATION SYSTEM USING XGBOOST AND SHAP

Savitha. D¹, Sowmiya. V², Akshaya. C³

Assistant Professor, Department of Computer Applications, Vellalar College for Women, Erode, India¹

II-MCA, Department of Computer Applications, Vellalar College for Women, Erode, India^{2, 3}

Abstract: The advancement of Artificial Intelligence (AI) and Machine Learning (ML) has brought significant improvements to healthcare by supporting disease prediction, individual health assessment, and clinical decision-making. However, conventional healthcare systems often provide generalized recommendations without adequately considering individual differences in demographic information, lifestyle, physiological conditions, medical history, and health risks. This review proposes an AI-Based Personalized Healthcare Recommendation System that integrates Extreme Gradient Boosting (XGBoost) with SHapley Additive exPlanations (SHAP). The approach combines the strong predictive performance of XGBoost with Explainable Artificial Intelligence (XAI) to generate personalized and understandable healthcare insights. The system analyzes various healthcare attributes, including demographic details, body measurements, vital signs, lifestyle factors, medical history, family history, laboratory results, and existing health conditions to assess potential disease risks. XGBoost is effective for structured healthcare data as it can capture complex nonlinear relationships between different health-related features. SHAP provides explanations by identifying the contribution of individual features to each prediction, enabling users and healthcare professionals to understand important factors such as blood glucose, body mass index, cholesterol levels, physical activity, sleep duration, and family history.

Keywords: Personalized Healthcare, Machine Learning, XGBoost, SHAP

1. INTRODUCTION

Healthcare is rapidly evolving with the integration of digital health records, wearable devices, laboratory data, and Artificial Intelligence (AI). Conventional healthcare mainly provides generalized advice through periodic consultations and may not fully address individual patient characteristics. Personalized healthcare overcomes this limitation by analysing patient-specific demographic, lifestyle, physiological, and medical information to identify health risks and support preventive care. Machine Learning (ML) techniques can identify complex relationships within healthcare data. Among various algorithms, Extreme Gradient Boosting (XGBoost) is effective for structured and tabular healthcare datasets due to its strong predictive performance. However, prediction accuracy alone is insufficient in healthcare, where understanding the reasons behind a prediction is equally important.

SHapley Additive exPlanations (SHAP) provides interpretability by identifying the contribution of individual features to a model's prediction. The combination of XGBoost and SHAP therefore enables both health-risk prediction and understandable explanations. This study presents an AI-Based Personalized Healthcare Recommendation System using XGBoost and SHAP to analyse patient information, predict potential health risks, explain prediction factors, and provide personalized healthcare recommendations. The study also discusses the challenges, research gaps, applications, and future scope of AI-based personalized healthcare systems.

II. LITERATURE SURVEY

Personalized healthcare by leveraging artificial intelligence technology is becoming an interesting area of research due to the increasing amount of health-related data and the need for personalized health services. The list of AI-based personalized healthcare studies comprises disease prediction, personalized treatment suggestions, clinical decision support systems, explainable artificial intelligence, wearable health monitoring, and privacy-preserving healthcare systems.



Author & Year	Dataset	Proposed Technique	Feature Selection	Classification Method	Accuracy / Performance	Strength	Limitation	Inference
Chen & Guestrin (2016)	Structured benchmark datasets	Gradient boosting framework	Tree-based feature learning	XGBoost	High scalable performance	Efficient and scalable	Requires careful tuning	XGBoost is suitable for structured healthcare prediction
Allgaier et al. (2023)	Healthcare ML studies	Systematic review of XAI	SHAP, LIME and other XAI methods	Multiple ML models	Review-based	Improves understanding of predictions	Patient-level interpretability remains challenging	SHAP is highly relevant for tabular healthcare data (ScienceDirect)
Robust Diabetes Study (2023)	Diabetes dataset	SHAP-incorporated ML	SHAP-based feature importance	XGBoost and other ML models	XGBoost: 94.67%	Strong prediction and feature analysis	Limited dataset/population	Glucose, BMI and age were important factors (ScienceDirect)
Interpretable Heart Disease Study (2023)	Heart disease data	Explainable ML framework	SHAP and LIME	SVM, RF, XGBoost, KNN	Up to 88% F1-score	Provides interpretable predictions	Limited clinical validation	XAI improves trust in disease prediction (ScienceDirect)

III. RESEARCH GAP

Even though Artificial Intelligence has contributed to the advancement in the field of healthcare prediction and recommendation, there are still many research gaps in this area. The current literature shows that machine learning algorithms have achieved good results in the field of disease prediction and health risk classification. But many systems focus only on one aspect of the healthcare analysis, such as prediction, disease classification, feature analysis, and others. It is essential to create an integrated healthcare system, which involves prediction, explainability, recommendation, monitoring, security, and usage of various healthcare data sources. The next research gaps demonstrate the gaps in the current research in AI healthcare systems and the necessity to develop the AI-Based Personalized Healthcare Recommendation System using XGBoost and SHAP.

3.1 Lack of Personalization

- ☐ Current systems often provide **standardised health predictions** for different individuals.
- ☐ They may not fully consider **individual health conditions, lifestyle, and medical background**.
- ☐ Differences in **diet, physical activity, sleep, and family history** are often overlooked.
- ☐ Generic healthcare advice may not effectively address **individual patient needs**.

3.2 Limited Model Explainability

- Machine learning models can identify complex relationships among health-related factors.
- Some models operate as black boxes, making their decisions difficult to interpret.
- In healthcare, accurate predictions alone are not enough.
- Users and healthcare professionals need to understand the reasons behind each prediction.



IV. CHALLENGING ISSUES

□ Quality of Healthcare Data

Patient datasets may include missing, inaccurate, inconsistent, or duplicate records. Such data-quality problems can negatively affect the training and prediction performance of the system.

□ Patient Data Privacy

Healthcare information is highly sensitive, making privacy and security important concerns. The system must protect patient data from unauthorised access, misuse, and potential security breaches.

□ Data Bias

If the training dataset does not adequately represent different populations, the model may produce less reliable results for certain groups. Maintaining balanced and representative data is therefore challenging.

V. MOTIVATION

The motivation for this study arises from the growing need for personalised and data-driven healthcare solutions that can address the specific needs of individual patients. Conventional healthcare recommendations are often based on general guidelines and may not adequately consider factors such as age, lifestyle, medical history, symptoms, vital signs, and other health indicators. Analysing these factors together can help identify individual health risks more effectively. Machine Learning can support this process by discovering meaningful patterns and relationships within healthcare data. In this study, XGBoost is employed to analyse structured patient information and predict potential health risks. However, accurate predictions alone may not be sufficient for healthcare applications, as users and healthcare professionals need to understand the factors influencing the results. Therefore, SHAP is incorporated to provide clear explanations of individual predictions, improving transparency and supporting personalised healthcare recommendations.

VI. SCOPE OF THE STUDY

- Focuses on analysing individual health profiles to support personalised healthcare.
- Considers information such as demographic details, lifestyle, symptoms, vital signs, medical history, and laboratory data.
- Applies XGBoost to identify health patterns and estimate potential health risks.
- Integrates SHAP to explain the major factors contributing to each prediction.
- Generates individualised healthcare recommendations according to the predicted risk.

VII. PROPOSED SYSTEM

The proposed system, Personalised Healthcare Recommendation System using XGBoost and SHAP, is designed to analyse individual health information, predict potential health risks, and provide personalised healthcare recommendations. The system considers patient details such as demographic information, lifestyle habits, symptoms, vital signs, medical history, and laboratory results. After preprocessing the collected data, the XGBoost algorithm is applied to identify health patterns and estimate potential risks. SHAP is integrated to explain the prediction by identifying the major features influencing the model's output. Based on the predicted risk and contributing factors, the system provides personalised guidance related to diet, physical activity, sleep, preventive care, and healthy lifestyle practices. This integrated approach combines health-risk prediction, explainable AI, and personalised recommendations to support preventive and patient-centred healthcare.

VIII. METHODOLOGY

7.1 Data Collection

Patient information such as demographic details, lifestyle factors, symptoms, vital signs, medical history, and laboratory results is collected for analysis.

7.2 Data Preprocessing

The collected data is cleaned by handling missing values, duplicate records, and inconsistencies. The data is then prepared in a suitable format for model analysis.

7.3 Feature Selection

Relevant health-related features are identified and selected to improve the effectiveness of health-risk prediction.

7.4 Health Risk Prediction using XGBoost

The processed patient data is provided to the **XGBoost** model to identify patterns and predict potential health risks.



7.5 Prediction Explanation using SHAP

SHAP is used to explain the XGBoost prediction by identifying the important features that contribute to increasing or decreasing the predicted risk.

7.6 Personalised Recommendation

Based on the predicted risk and contributing factors, the system generates personalised recommendations related to diet, physical activity, sleep, preventive care, and healthy lifestyle practices.

7.7 System Output

The system presents the predicted health risk, SHAP-based explanation, and personalised healthcare recommendations in an understandable form.

IX. APPLICATIONS OF PERSONALISED HEALTHCARE RECOMMENDATION SYSTEM USING XGBOOST AND SHAP

- ☐ **Disease Risk Prediction** – Identifies potential health risks using patient data.
- ☐ **Personalised Healthcare** – Provides recommendations based on individual health conditions.
- ☐ **Diet and Fitness Guidance** – Suggests suitable diet and physical activity practices.
- ☐ **Preventive Healthcare** – Helps identify risk factors and encourages preventive measures.
- ☐ **Patient Monitoring** – Supports continuous assessment of health-related information.
- ☐ **Clinical Decision Support** – Assists healthcare professionals with data-driven insights.

X. CONCLUSION

The Personalised Healthcare Recommendation System using XGBoost and SHAP offers an intelligent and transparent method for analysing patient health information, identifying potential health risks, and providing personalised healthcare recommendations. The XGBoost algorithm analyses various patient attributes, including demographic information, lifestyle patterns, symptoms, vital signs, medical history, and other health indicators, to identify relevant patterns and predict health risks. SHAP improves the interpretability of these predictions by showing the contribution of individual features to the model's output. Using the predicted risk level and significant contributing factors, the system can generate personalised recommendations related to diet, physical activity, lifestyle changes, preventive care, and health monitoring.

This approach can help individuals gain better awareness of their health conditions while providing healthcare professionals with useful data-driven insights for decision support. Overall, integrating machine learning, explainable AI, and personalised recommendations creates a practical framework for supporting preventive and patient-centred healthcare. Future enhancements may include the use of larger and more diverse datasets, real-time health monitoring, wearable-device integration, multi-disease prediction, privacy-preserving techniques, and clinical validation to further improve the system's effectiveness and practical applicability.

XI. FUTURE RESEARCH DIRECTIONS

- ☐ **Integration with Wearable Devices** – Future systems can be connected with smartwatches and fitness devices to collect real-time data such as heart rate, physical activity, sleep patterns, and other health indicators.
- ☐ **Larger and Diverse Datasets** – Research can focus on using larger datasets collected from different age groups, regions, and healthcare backgrounds to improve the reliability and generalisation of the model.
- ☐ **Real-Time Health Monitoring** – The system can be enhanced to continuously monitor patient health information and provide timely risk alerts and recommendations.
- ☐ **Advanced Machine Learning Models** – Future research can compare XGBoost with advanced models such as LightGBM, CatBoost, deep learning, and ensemble approaches to improve prediction performance.
- ☐ **Advanced Explainable AI** – SHAP can be combined with other explainability techniques to provide more detailed and user-friendly explanations of healthcare predictions.

REFERENCES

- [1] Rathod, V. N., Goudar, R. H., & Hukkeri, G. S. (2026). AI-augmented personalized treatment in healthcare: A comprehensive survey of models, benchmarking and deployment challenges.
- [2] Khakpaki, A., Javid, S. A., Farshbaf-Khalili, A., et al. (2026). AI-integrated clinical decision support systems for precision medicine in real-world healthcare.



- [3] Towards integration of artificial intelligence into medical devices as a real-time recommender system for personalised healthcare: State-of-the-art and future prospects. *Health Sciences Review*, 2024.
- [4] Shiddik, et al. (2026). Explainable Artificial Intelligence in Healthcare: Current Landscape, Challenges, and Future Directions.
- [5] Explainable Artificial Intelligence for Tabular Data in Healthcare: A Systematic Review of Methods, Evaluation, and Applications. *Computers*, 2026.
- [6] Explainable AI Applications in Healthcare: A Systematic Review. *Algorithms*, 2026.
- [7] Overview of RAG-based and LLM-based approaches to personalization in healthcare AI applications. *Frontiers in Artificial Intelligence*, 2026.
- [8] Application of AI and digital health tools in public health management of T2DM: From mechanism prediction to personalized treatment. *Frontiers in Public Health*, 2026.
- [9] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- [10] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*
- [11] eting model predictions. *Advances in Neural Information Processing Systems*



