Multidiag-X: AI-Powered Multi-Disease Risk Prediction and Medical Report Analyzer

Dr. Anil Kumar Yadav[1], Samridhi Tiyagi[2], Taniya[3], Khushi Gupta[4], Tanvi Gupta[5]
[1]*Senior Assistant Professor, School of Computing Science Engineering and Artificial Intelligence VIT Bhopal University, Bhopal-Indore Highway, Kothrikalan, Sehore, Madhya Pradesh - 466114*
[2,3,4,5]*Student, School of Computing Science Engineering and Artificial Intelligence, VIT Bhopal University Bhopal-Indore Highway, Kothrikalan, Sehore, Madhya Pradesh - 466114*

***Abstract*—Understanding medical reports can be difficult for individuals without a medical background. MultiDiag-X is an AI-based system developed to simplify medical report interpretation and provide early risk awareness for diseases such as heart disease and diabetes. The system analyzes basic health data and medical reports to identify abnormal values and present them in simple, easy-to-understand language. It also uses explainable AI techniques to show the factors influencing predictions, improving transparency and trust. Based on the predicted risk levels, the system offers responsible guidance, helping users decide whether to maintain their lifestyle, monitor their health, or consult a doctor. This project focuses on improving health awareness and supporting informed decisions without replacing professional medical advice.**

***Index Terms*—Artificial Intelligence, Machine Learning, Disease Risk Prediction, Heart Disease, Diabetes, Medical Report Analysis, Explainable AI, Health Monitoring, Decision Support System**

I. INTRODUCTION

In recent years, people have become more conscious about their health, which has led to a significant increase in medical tests and diagnostic reports. However, understanding these re-ports is still a major challenge for individuals who do not have a medical background. Most reports include complex medical terms, numerical readings, and reference ranges that are not easy to interpret. Because of this, many people either ignore important warning signs or become unnecessarily worried due to confusion. In addition, these reports usually do not provide clear guidance, leaving users unsure about whether they need to seek medical attention.

At the same time, diseases like heart disease and diabetes are becoming more common and require early detection along with regular monitoring. Being aware of potential health risks at an early stage can help prevent serious complications. How-ever, access to simple and immediate health insights is still limited, especially for those who depend only on occasional medical checkups. With the rapid growth of Artificial Intelligence (AI) and Machine Learning (ML), it is now possible to analyze health-related data more efficiently and generate useful insights. These technologies can process large volumes of data, recognize patterns, and estimate disease risk levels with good accuracy. Moreover, Explainable AI (XAI) plays an important role in making these predictions more transparent by showing the key factors that influence the results, which helps in building user trust. MultiDiag-X is an AI-based system developed to make medical reports easier to understand and to provide early risk prediction for heart disease and diabetes. The system allows users to either enter basic health details or upload their medical reports, which are then analyzed to detect abnormal values and evaluate risk levels. It presents complex medical information in simple and clear language, making it easier for non-experts to understand.

In addition, the system provides explainable insights that highlight how different health parameters affect the predicted risk. Based on these results, it offers responsible suggestions, helping users decide whether to continue their current lifestyle, monitor their health more carefully, or consult a doctor. The main objective of MultiDiag-X is to improve health awareness, support better decision-making, and provide users with clear and useful information, while ensuring that

it does not replace professional medical advice or diagnosis.

II. LITERATURE REVIEW OF EXISTING SYSTEMS

In recent years, Artificial Intelligence (AI) and Machine Learning (ML) have been widely used in the healthcare field to improve disease prediction and overall patient care. Many studies have focused on building predictive models for diseases such as heart disease and diabetes using algorithms like Logistic Regression, Decision Trees, Support Vector Machines (SVM), and Random Forest. These models work by analyzing patient data such as age, blood pressure, glucose levels, and other important health parameters to estimate the chances of a disease. Several research works have shown that these approaches can achieve high accuracy and perform well under controlled conditions.

However, despite their strong performance, traditional dis-ease prediction systems still face some important challenges. One major issue is that many of these models function as "black-box" systems, meaning they provide predictions with-out clearly explaining how the result was obtained. This lack of transparency reduces user trust, especially for people without a medical background who may find it difficult to rely on such outputs. To address this problem, Explainable Artificial Intel-ligence (XAI) has been introduced. XAI focuses on making AI models more understandable and transparent. Techniques such as SHAP (SHapley Additive Explanations) are commonly used to show how different input features contribute to the final prediction, helping users better understand the reasoning behind the results.

Another limitation of existing systems is their narrow scope. Most current models are designed to predict only one disease at a time, which limits their usefulness in real-world situations where multiple health conditions may be related. In addition, many systems depend only on structured datasets and do not support real-world inputs like medical reports in PDF format. This forces users to manually enter data, which can be both time-consuming and error-prone.

Moreover, many existing solutions do not focus enough on user experience. The results are often presented using complex medical terms and numerical values that are difficult for non-experts to understand. These systems also usually lack clear guidance on what steps users should take after receiving the results. This creates a gap between prediction and actual decision-making, reducing the overall usefulness of such systems.

Recent research highlights the importance of combining multiple features into a single platform, such as prediction, explanation, and user guidance. However, very few systems successfully integrate all these aspects together. There is a clear need for a more complete solution that not only predicts disease risk but also simplifies medical reports, explains the outcomes clearly, and provides meaningful guidance to users.

To overcome these challenges, the proposed system, MultiDiag-X, is designed as an integrated health awareness platform. It combines multi-disease risk prediction (focused on heart disease and diabetes), medical report analysis, ex-plainable AI, and guided health advice into a single, user-friendly system. The main aim is to improve accessibility, transparency, and usability, making healthcare information easier to understand and more practical for everyday users.

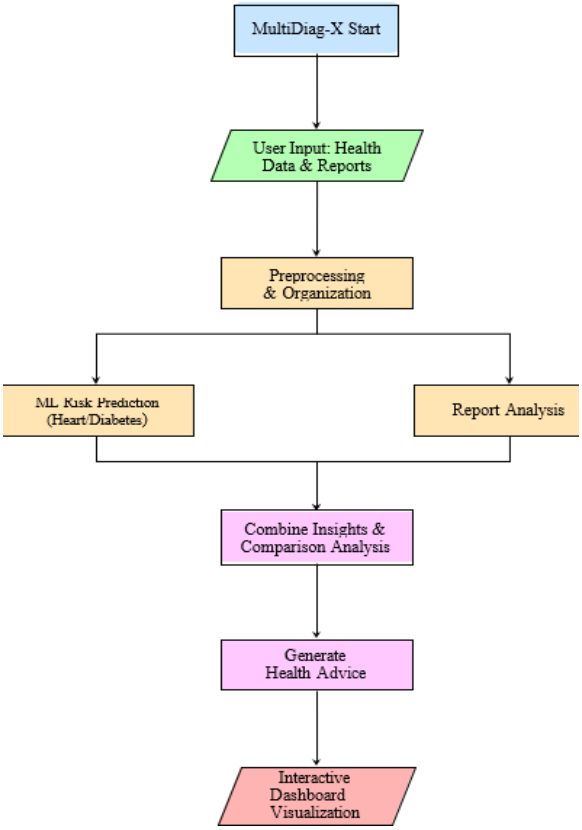

Fig. 1. MultiDiag-X flowchart with respect to existing system

III. PROPOSED SYSTEM DESIGN

The proposed system, MultiDiag-X, is an AI-based platform created to help people understand their medical reports more easily and to give early risk predictions for diseases like heart disease and diabetes. The main goal of the system is to make health information clear, easy to understand, and useful for users. It combines machine learning, explainable AI, and a simple user interface into one complete system. The system is designed with several connected modules that work together. It takes two types of input from users. The first is manual health data, such as age, blood pressure, and glucose level. The second is medical reports in PDF format. These inputs are first processed and organized so that the system can analyze them properly. At the center of the system is the multi-disease prediction model. This model uses machine learning techniques like Logistic Regression and Random Forest to estimate the risk of heart disease and diabetes. It studies the health data provided by the user and calculates a risk percentage for each disease. Based on this, the system groups the result into three levels: low, moderate, and high risk. It is important to understand that this system only predicts risk and does not give a medical diagnosis.

Another important part of the system is the medical report analysis module. Users can upload their reports in PDF format, and the system extracts key information using tools like pdf plumber or PyMuPDF. It then compares the values in the report with standard normal ranges to find any abnormal results. These findings are shown in a simple and easy-to-read format so that even non-medical users can understand them. To make the system more transparent, it includes an explainable AI feature. Using techniques like SHAP, the system shows which health factors had the most impact on the prediction. This helps users clearly see why a certain result was given and builds trust in the system.

The system also includes a comparison feature that helps users understand changes in their health over time. It can compare current and past reports to show whether a person's condition is improving or getting worse. This allows users to track their health progress more effectively.

Finally, the system provides basic health guidance based on the risk level. If the risk is low, users are encouraged to continue their healthy habits. If the risk is moderate, they are advised to monitor their health and improve their lifestyle. If the risk is high, the system strongly suggests consulting a doctor. These suggestions are general and meant to guide users, not replace professional medical advice.

Overall, MultiDiag-X is designed as a simple and user-friendly platform that brings together prediction, explanation, and guidance in one place. It helps users understand complex medical information in a clear way while still encouraging them to seek proper medical care when needed.

IV. SYSTEM ARCHITECTURE DIAGRAM

The system architecture of MultiDiag-X follows a clear step-by-step process. First, the user provides input in the form of health details or medical reports. This input is then cleaned and organized so it can be used properly by the system. After that, the processed data is sent to machine learning models, which predict the risk of heart disease and diabetes. At the same time, explainable AI techniques are used to identify the key factors that affect these predictions. Alongside this, the report analysis module works on uploaded medical reports by extracting important lab values and checking whether they are normal or not.

All the results are then combined, and the system also performs comparisons to give better insights. Based on this, it generates simple explanations and provides basic health guidance for the user. Finally, everything is shown on an interactive dashboard with clear visuals, making it easy for users to understand their health information.

V. METHODOLOGY AND ALGORITHMS USED

The development of the MultiDiag-X system is based on a modular and AI-driven approach that focuses on improving health awareness and making medical reports easier to under-stand. The overall process is divided into several important stages.

In the first stage, data collection and preprocessing, datasets related to heart disease and diabetes are gathered. These datasets include important health parameters such as age, glucose level, blood pressure, and cholesterol. The data is then cleaned and prepared by removing missing values, correcting errors, and normalizing it to improve the performance of the models.

In the next stage, machine learning models are trained using labeled data to predict disease risk. Algorithms like Logistic Regression and Random Forest are used to analyze the input data and classify the risk into three categories: low, moderate, and high. The system also includes a medical report analysis feature. Users can upload their reports in PDF format, and tools such as pdfplumber or PyMuPDF are used to extract important lab values. These values are compared with standard normal ranges to detect any abnormalities and are then organized for further use.

To make the results easier to understand, the system uses explainable AI techniques. SHAP is applied to show how each health factor affects the prediction. This helps users clearly understand why a certain result was generated. Based on the predictions and report analysis, the system provides basic guidance to users. It categorizes them into different risk levels and suggests actions such as maintaining a healthy lifestyle, monitoring their condition, or consulting a doctor when needed.

The system uses a combination of different algorithms to improve accuracy and reliability. Logistic Regression is used to calculate the probability of diseases based on input data and is simple to interpret. Random Forest improves prediction accuracy by combining multiple decision trees and reducing errors. SHAP is used to explain the results by showing the impact of each feature. In addition, rule-based logic is applied to classify risk levels and provide safe and meaningful suggestions to users.

Overall, this approach ensures that the system is accurate, transparent, and easy to use, helping users better understand their health information.

VI. PROJECT FUNCTIONAL MODULES IMPLEMENTATION

The MultiDiag-X system is built using a modular design, where each part of the system handles a specific task. This approach helps in smooth processing and improves the accuracy of results. The system is divided into several functional modules.

The multi-disease risk prediction module is responsible for estimating the risk of heart disease and diabetes. It takes user inputs such as age, glucose level, blood pressure, and cholesterol. This data is processed using trained machine learning models like Logistic Regression and Random Forest. The system then calculates risk percentages and classifies them into low, moderate, or high levels.

The medical report analysis module manages the upload and processing of medical reports in PDF format. Using tools like pdfplumber or PyMuPDF, it extracts important lab values from the reports. These values are compared with standard normal ranges to find any abnormal results. The processed information is then prepared for further use in other parts of the system.

The explanation and insight module focuses on making the results easy to understand. It converts complex medical terms and numbers into simple language. It also uses explainable AI techniques like SHAP to show which health factors have the most impact on the prediction. Visual elements such as graphs are used to make the information clearer and easier to follow. The health interaction and comparison module help users understand how different health factors are related. It can also compare current data with previous reports to show whether a person's health is improving or getting worse. The results are presented in simple statements so users can easily track their progress over time.

The guided health advice module provides general suggestions based on the user's risk level. If the risk is low, users are advised to continue healthy habits. If it is moderate, they are encouraged to monitor their health and make lifestyle changes. If the risk is high, the system strongly suggests consulting a doctor. These suggestions are meant to guide users responsibly and support better decision-making.

Overall, the modular structure of MultiDiag-X allows each part of the system to work independently while still being connected. This makes the system efficient, easy to use, and reliable for improving health awareness.

VII. METHODOLOGY FOR DEVELOPING MULTIDIAG-X

The development of MultiDiag-X follows a clear and step-by-step approach to make the system accurate, easy to use, and reliable. The process begins with problem identification, where the need to simplify medical reports and provide early awareness of diseases like heart disease and diabetes was recognized. The system is designed to help users understand their health data in a simple and clear way.

The next step is data collection and preprocessing. In this stage, datasets related to heart disease and diabetes are collected. These include important health factors such as age, glucose level, blood pressure, and cholesterol. The data is then cleaned by removing missing values, fixing errors, and normalizing it so that the machine learning models can perform better.

After that, model development and training are carried out. Machine learning algorithms like Logistic Regression and Random Forest are trained using the processed data. These models predict the risk of diseases and give results in the form of probabilities. These probabilities are then grouped into three levels: low, moderate, and high risk, so that users can easily understand them.

At the same time, a medical report processing feature is developed. This allows users to upload their reports in PDF format. Tools like pdfplumber or PyMuPDF are used to extract important lab values from the reports. These values are compared with normal ranges to detect any abnormalities. This step ensures that both manual data and real medical reports can be analyzed properly.

To make the system more transparent, explainable AI is included using SHAP. This helps in showing how each health factor affects the prediction. The results are displayed with simple explanations and visual elements so that users can understand the reasoning behind them.

Next, the system provides guidance and decision support. Based on the risk level, it gives general suggestions such as maintaining a healthy lifestyle, monitoring health regularly, or consulting a doctor if needed. This ensures that the system is used responsibly.

Finally, all the modules are combined into a user-friendly interface using Streamlit. Users can enter their data, upload reports, and view results in an interactive and simple format. Overall, this approach makes the system accurate, transpar-ent, and easy to use, while keeping its purpose as a health awareness and support tool rather than a replacement for medical diagnosis.

VIII. CONTRIBUTION AND FINDINGS

The proposed system, MultiDiag-X, offers several useful contributions in the area of AI-based healthcare support. One of its main strengths is that it combines multiple features into a single platform. It brings together disease risk prediction, medical report analysis, explainable AI, and basic health guidance, instead of focusing on just one function. This makes the system more practical and helpful, especially for users without a medical background.

Another important contribution is its focus on simplicity and clarity. The system converts complex medical data and terms into easy-to-understand language, so that anyone can use it without confusion. It also uses explainable AI with SHAP to show how different health factors affect the prediction. This makes the results more transparent and helps build user trust. The system also includes a feature for analyzing medical reports in PDF format. Users can upload their reports, and the system automatically extracts key lab values and checks for any abnormal results. This removes the need for manual data entry and makes the process faster and more convenient.

In addition, MultiDiag-X provides guided suggestions based on the user's risk level. These suggestions help users decide whether to maintain their current lifestyle, monitor their health more closely, or consult a doctor. This supports better decision-making while keeping the system safe and responsible.

The results of the project show that machine learning models like Logistic Regression and Random Forest can effectively predict the risk of heart disease and diabetes using basic health data. The use of explainable AI makes these predictions easier to understand and more useful for users. Combining prediction, report analysis, and guidance makes the system more complete compared to many existing solutions.

Overall, the project shows that AI can be used not just for prediction, but also for improving health awareness and supporting better decisions. It also highlights the importance of making systems accurate, transparent, and user-friendly, while clearly maintaining that they are meant to assist users and not replace professional medical advice.

IX. CONCLUSION

MultiDiag-X is an AI-based health awareness system developed to make medical reports easier to understand and to pro-vide early risk prediction for conditions like heart disease and diabetes. The system

brings together machine learning models, medical report analysis, and explainable AI in one platform to give users clear and meaningful health insights. One of its key strengths is its ability to convert complex medical data into simple language and visual explanations, allowing users to understand their health status without needing a medical background. The use of machine learning algorithms such as Logistic Regression and Random Forest helps the system predict disease risk with good accuracy based on user input data. At the same time, the integration of explainable AI through SHAP makes the predictions more transparent by showing which health factors influenced the results. This not only improves understanding but also increases user confidence in the system.

Another important feature of MultiDiag-X is its ability to analyze medical reports directly. By allowing users to upload reports in PDF format and automatically extracting key lab values, the system reduces manual effort and makes the process more convenient. It also highlights abnormal values and presents them in a simple format, helping users quickly identify potential concerns.

In addition, the system provides guided health suggestions based on the predicted risk levels. These suggestions encourage users to take appropriate steps, such as maintaining a healthy lifestyle, monitoring their condition, or seeking medical advice when needed. This makes the system more practical and supportive in real-life situations.

Overall, MultiDiag-X shows how AI can be used not only for prediction but also for improving health awareness and understanding. It helps bridge the gap between complex medical information and everyday users by making data more accessible and meaningful. At the same time, it clearly maintains its role as a supportive tool, emphasizing that it is not a replacement for professional medical diagnosis or consultation, but rather a system that assists users in making better-informed health decisions. [1]– [17].

REFERENCES

[1] R. Alkhanbouli, H. M. A. Almadhaani, F. Alhosani, and M. C. E. Simsekler, "The role of explainable artificial intelligence in disease prediction: A systematic literature review and future research directions," *BMC Medical Informatics and Decision Making*, vol. 25, art. no. 110, 2025.

[2] R. Alkhanbouli, M. C. E. Simsekler, *et al*., "multi-disease risk prediction models for population-scale personalised screening: A UK Biobank study," *medRxiv*, 2026.

[3] H. Chen and S. Liu, "AI-driven multimodal precision diagnosis and progression prediction: Data fusion mechanisms and clinical applications," *BMC Medical Informatics and Decision Making*, vol. 26, no. 1, 2026.

[4] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016.

[5] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart disease," *Bioinformatics*, vol. 36, no. 2, pp. 279–285, 2020.

[6] S. Deo, "Machine learning in medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015.

[7] R. Gupta and V. Sharma, "Personalized health monitoring using explainable AI: Bridging trust in predictive healthcare," *Scientific Reports*, vol. 15, art. no. 12397, 2025.

[8] J. Hou and L. Wang, "Interactive explainable AI for healthcare risk prediction: Integrating user feedback into model explanations," *Journal of Biomedical Informatics*, vol. 142, 2025.

[9] I. Kavakiotis *et al*., "Machine learning and data mining methods in diabetes research," *Computational and Structural Biotechnology Journal*, vol. 15, pp. 104–116, 2017.

[10] X. Li, Y. Zhang, *et al*., "VisionTrack: Multi-modal AI for multi-label retinal disease prediction using LLMs and GNNs," *Sensors*, vol. 25, no. 14, art. no. 4492, 2025.

[11] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[12] S. Park and H. Lee, "Clinical transformer: Predicting multi-morbidity risks from electronic health record textual reports," *Artificial Intelligence in Medicine*, vol. 158, 2025.

[13] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*,

2016.

[14] A. Shishehbori and M. U. Awan, "Web-based cardiovascular disease risk prediction using machine learning and SHAP-based visualizations," *Frontiers in Artificial Intelligence*, vol. 9, art. no. 1690664, 2026.

[15] Smith, B. Jones, *et al*., "AI-driven analysis of patient safety reports using large language models: An exploratory multiple methods study," *BMJ Quality & Safety*, 2026.

[16] Z. Wang and T. Miller, "Leveraging large language models and machine learning for success analysis in robust cancer crowdfunding predictions," *JMIR AI*, vol. 4, 2025.

[17] Q. Zhao and P. Kumar, "Multi-disease prediction based on deep learning: A systematic survey," *IEEE Access*, vol. 12, pp. 45120–45135, 2024.