Disease Prediction Using Ensemble Learning and XAI: A Comparative Study On Heart, Diabetes, Kidney, Parkinson's and Breast Cancer Datasets

by

Fouzia Zilani
Student ID: 2012176111
Course Code : CSE4292

Project submitted in partial fulfillment of the requirements for the degree of

Bachelor of Science in Computer Science and Engineering

Under the supervision of

Dr. Asif Zaman
Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

UNIVERSITY OF RAJSHAHI, RAJSHAHI-6205, BANGLADESH

DECLARATION

I declare that this project entitled *Disease Prediction Using Ensemble Learning and XAI: A Comparative Study On Heart, Diabetes, Kidney, Parkinson's and Breast Cancer Datasets* is the result of my own work except as cited in the references. The project has not been accepted for any degree and is not concurrently submitted in the candidature of any other degree.

Fouzia Zilani
Student ID: 2012176111

Department of Computer Science and Engineering
University of Rajshahi, Rajshahi-6205, Bangladesh

Date: October 29, 2025

CERTIFICATE FROM SUPERVISOR

I certify that the project entitled **Disease Prediction Using Ensemble Learning and XAI: A Comparative Study On Heart, Diabetes, Kidney, Parkinson's and Breast Cancer Datasets** submitted for the degree of Bachelor of Science in Computer Science and Engineering by **Student ID: 2012176111** is the record of project work carried out by her during the period from November 2024 to November 2025 under my guidance and supervision, and that this work has not formed the basis for the award of any Degree, Diploma, Associateship and Fellowship or other Titles in this University or any other University or Institution of Higher Learning.

Dr. Asif Zaman
Professor

Department of Computer Science and Engineering
University of Rajshahi, Rajshahi-6205, Bangladesh

Date: October 29, 2025

CERTIFICATE FROM EXAM COMMITTEE

This is to certify that the project entitled **Disease Prediction Using Ensemble Learning and XAI: A Comparative Study On Heart, Diabetes, Kidney, Parkinson's and Breast Cancer Datasets**, submitted by **Student ID: 2012176111**, an undergraduate student of the Department of Computer Science and Engineering, has been examined. Upon recommendation by the examination committee, we hereby accord our approval of it as the presented work and submitted report fulfills the requirements for its acceptance in partial fulfillment for the degree of *Bachelor of Science in Computer Science and Engineering*.

Chairman of Exam Committee
4th Year, Session: 2020-2024

Department of Computer Science and Engineering
University of Rajshahi, Rajshahi-6205, Bangladesh

Date: October 29, 2025

ACKNOWLEDGEMENTS

First and foremost, I dedicate this work to the memory of the martyrs of the July-August 2024 Revolution in Bangladesh, honoring their sacrifice and courage in the pursuit of justice and democracy.

I would like to express my deep and sincere gratitude to my research supervisor, Dr. Asif Zaman, Ph.D., for providing me with the opportunity to conduct this research. His invaluable guidance, dynamism, vision, sincerity, motivation, and teaching methodology have been instrumental in shaping this work and my academic journey.

I am deeply indebted to the honorable teachers of the Department of Computer Science and Engineering at the University of Rajshahi, Rajshahi-6205, Bangladesh, for their excellent teaching and guidance throughout my academic career. Their dedication and expertise have completely transformed me into the person that I am today.

I express my extreme thankfulness to my parents for their unconditional love, endless prayers and caring, and immense sacrifices for educating and preparing me for my future. Their support has been the foundation of my success.

I also extend my heartfelt thanks to my friends and relatives for their kind support and care throughout this journey.

Finally, I would like to thank all the people who have supported me to complete this project work directly or indirectly.

Fouzia Zilani
Student ID: 2012176111

Department of Computer Science and Engineering
University of Rajshahi, Rajshahi-6205, Bangladesh

Date: October 29, 2025

ABSTRACT

Early and accurate disease diagnosis is vital in reducing death rates and improving healthcare outcomes. In this current research, an ensemble learning based predictive model is proposed to forecast five major diseases: Heart Disease, Diabetes, Chronic Kidney Disease, Parkinson's Disease, and Breast Cancer from publicly available medical datasets. The method involves advanced preprocessing procedures such as feature engineering, data normalization, and class balancing employing SMOTETomek to address noise and class imbalance. A Soft Voting Classifier is constructed by combining several machine learning algorithms like Random Forest, Gradient Boosting, XGBoost, LightGBM, CatBoost, Support Vector Machine (SVM), and Logistic Regression to achieve optimal classification performance. Explainable AI (XAI) techniques like SHAP and LIME are applied to highlight the most important clinical features making predictions in order to facilitate interpretability and transparency. The proposed framework achieves high accuracy for all the datasets: 85.2% for Heart Disease, 81.2% for Diabetes, 98.8% for Chronic Kidney Disease, 92.3% for Parkinson's Disease, and 98.2% for Breast Cancer. The experimental results demonstrate ensemble learning improves robustness and generalization compared to individual classifiers. This work aids in constructing robust, interpretable, and automatic decision support systems to assist healthcare professionals with early disease diagnosis.

Contents

List of Figures

List of Tables

1. INTRODUCTION

1.1 Project Background

Early diagnosis of fatal diseases is essential in reducing death rates, improving patient survival, and saving healthcare expenditures. While electronic health records and clinical data have been expanding exponentially, machine learning (ML) has emerged as a key player in the domain of healthcare analytics in identifying hidden clinical patterns and supporting on time clinical decisions [1, 2]. Among various ML methods, ensemble learning predicting results from multiple base models has been proven to be more generalizable and robust than solo classifiers [3, 4]. This research seeks to develop predictive models of five significant conditions: heart disease, diabetes, chronic kidney disease, Parkinson's disease, and breast cancer. Using advanced ensemble based techniques, the focus is on enhancing diagnostic performance and enabling efficient automated decision support for early disease diagnosis.

1.2 Problem Statement

Although numerous individual machine learning classifiers have been put forward for clinical diagnosis, their performance can significantly vary across different datasets and disease types [5, 6, 7]. This is attributed to challenges such as class imbalance, noise, small sample size, and complex non linear relationships among features in clinical data [8, 9]. Since no single model is capable of capturing all relevant patterns in diverse healthcare information, it is possible that over reliance on a single classifier can lead to poor generalization and unstable predictions [10]. Therefore, a robust predictive model that combines multiple learning algorithms is essential for enhancing diagnostic precision, reducing bias, and yielding stable performance across an extensive range of disease states. Additionally, using Explainable AI (XAI) techniques is essential to interpret significant features influencing predictions, ensuring transparency, reliability, and acceptance in real healthcare decision making [11, 12].

1.3 Objectives

The main objectives of this research are:

1. To create five independent disease prediction models, each optimized and designed to a particular medical condition:
    - Heart Disease
    - Diabetes
    - Chronic Kidney Disease
    - Parkinson's Disease
    - Breast Cancer
2. To apply dataset specific preprocessing techniques, including:
    - Missing value handling
    - Feature scaling
    - Encoding of categorical variables
    - Handling imbalanced datasets (e.g., SMOTE)
3. To perform personalized feature engineering and feature selection for each dataset to improve classification accuracy and reduce overfitting.
4. To train and test solid machine learning classifiers such as:
    - Random Forest
    - Support Vector Machine (SVM)
    - XGBoost
    - LightGBM
    - CatBoost
    - Multi-Layer Perceptron (MLP)
5. To build an ensemble model for every disease using Voting/Stacking to improve:
    - Accuracy
    - Stability
    - Generalization capability
6. To compare the ensemble model with single models and select the best performing optimized model for each disease.
7. To evaluate model performance using common clinical metrics such as:

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC

8. To identify important features for each disease model, contributing to medical interpretability.

9. To store and deploy the optimized model pipeline for every disease with:
   - Trained model file
   - Scaler/encoder files
   - Feature support list
   - Prediction functions

10. To demonstrate how machine learning can assist physicians in early diagnosis and enhanced decision making, which may reduce cases that are fatal.

1.4 Significance of the Study

Early diagnosis of life threatening diseases is paramount in preventing severe health complications and improving patient survival rates. Some hospitals and diagnostic centers still rely on human interpretation of clinical data, which could be time consuming, prone to human errors, and limited by restricted expert availability. The above mentioned shortcomings are addressed in the present study by developing computerized disease prediction models with advanced machine learning and ensemble approaches.

By focusing on five common diseases heart disease, diabetes, chronic kidney disease, Parkinson's disease, and breast cancer the paper has broad clinical value. Each predictive model is trained on real patient datasets and fine tuned individually to achieve good accuracy and reliability. Ensemble learning also enhances the stability of models and decision making capability across different types of data [1, 3].

One of the most important contributions of this work is the integration of Explainable Artificial Intelligence (XAI). XAI, through feature importance analysis, introduces transparency into model decisions by highlighting the most significant diagnostic features in making predictions [11, 12]. It allows clinicians to understand why a prediction is made, builds trust in AI based diagnostics, and facilitates improved medical judgment based on open explanations instead of black box results [2, 13].

In addition, the deployment ready predictive model developed in this research offers speedy, consistent, and cost effective screening. This can be especially beneficial where scant medical resources are present in regions, and access to skilled experts is never guaranteed. The results of this research promote the utilization of artificial intelligence technology in the field of medicine, reducing diagnostic times, improving preventive health measures, and maximizing better health governance in society.

1.5 Structure of the Paper

The rest of this paper is organized as follows:

Section II – Literature Review: Enumerates recent machine learning research in healthcare, such as disease prediction studies and explainable AI implementations.

Section III – Methodology: Describes the proposed framework, datasets, preprocessing, feature engineering, machine learning models, ensemble strategy, and XAI techniques used for interpretability.

Section IV – Implementation / Analysis: Presents the implementation setup, parameter values, train–test split, performance metrics, and validation scheme.

Section V – Results and Discussion: Reports comparative performance comparison of individual and ensemble models for all five diseases, interprets key features using XAI, and discusses clinical implications.

Section VI – Conclusion: Concludes the study by pointing out achievements and suggests future improvements for general real world applicability in medical diagnostics.

2. Background Study

2.1 Introduction to Artificial Intelligence in Healthcare

2.1.1 Overview of AI in Medicine

Artificial Intelligence (AI) and Machine Learning (ML) have emerged as transformative forces in healthcare, enabling automated diagnosis, patient risk assessment, and personalized treatment planning [1]. These technologies leverage medical data such as laboratory results, imaging, and clinical histories to uncover hidden patterns that assist physicians in decision making. Recent developments in deep learning and ensemble models have further improved diagnostic accuracy, especially in complex diseases like cardiovascular, renal, and neurological disorders [5, 2].

2.1.2 Explainable AI for Medical Transparency

A major challenge of AI in healthcare is its ”black box” nature clinicians often find it difficult to interpret model predictions. To address this, Explainable Artificial Intelligence (XAI) techniques such as SHAP and LIME have been widely adopted to visualize feature contributions and enhance interpretability [14, 11, 12]. XAI enables trust by clarifying how individual patient attributes, such as blood pressure or glucose level, influence diagnostic outcomes.

2.1.3 The Role of Ensemble Learning

Ensemble learning methods such as Random Forest, Gradient Boosting, and Voting Classifiers combine multiple models to improve predictive reliability. In medical prediction, these approaches provide stability across diverse datasets and mitigate overfitting [3]. They are particularly beneficial for handling noisy or imbalanced data, which are common challenges in clinical environments.

2.1.4 Toward Multi Disease Prediction Frameworks

The recent shift toward integrated multi disease prediction systems seeks to unify several disease classifiers under one robust framework. Such systems can identify inter disease rela-

tionships and provide scalable diagnostic solutions across conditions such as heart disease, diabetes, kidney disease, Parkinson's disease, and breast cancer [4, 15]. By incorporating explainable AI within ensemble architectures, these frameworks promote both transparency and clinical trustworthiness.

2.1.5 Research Motivation

Despite major advances, AI models face persistent issues such as class imbalance, overfitting, and lack of interpretability. Therefore, this research focuses on developing an ensemble based, explainable multi disease prediction framework capable of achieving high accuracy while maintaining interpretability and fairness in medical decision making.

2.2 Heart Disease Prediction Using Machine Learning

Early works and benchmark datasets, such as the Cleveland dataset introduced by Detrano et al. [16], have served as foundational resources for heart disease research. Comparative studies evaluating multiple algorithms including the works of Mohan et al. [5], Alotaibi [17], and Ali et al. [10] highlighted the strengths of ensemble learning and optimization based models. Further investigations have emphasized ensemble methods and the use of model interpretability frameworks such as SHAP and LIME to enhance transparency in predictions [11, 12]. Finally, researchers continue to address major clinical challenges, including imbalanced data distributions and strong inter feature correlations, using resampling techniques and advanced feature selection strategies [8, 18].

Heart disease remains one of the most prominent causes of death worldwide, accounting for nearly one third of all global fatalities annually according to the World Health Organization. Predicting heart disease early has thus become a primary goal for clinical researchers, and artificial intelligence (AI) driven systems are increasingly being deployed for this task. The application of machine learning (ML) models allows healthcare professionals to discover non linear patterns in patient data, which traditional diagnostic methods might overlook. Among all datasets, the *Cleveland Heart Disease dataset* developed by Detrano et al. [16] is one of the most widely used benchmarks in this research field. This dataset contains 303 patient records, each with 14 attributes, such as age, cholesterol level, resting blood pressure, and fasting blood sugar, which are critical in determining the presence of heart disease.

2.2.1 Early Works and Benchmark Datasets

Early research on heart disease prediction primarily focused on simple statistical and single model approaches, including logistic regression and decision trees [19]. However, these

early models often failed to capture the non linearity and complex interrelationships between cardiovascular risk factors. The availability of datasets such as the Cleveland dataset [16], Statlog Heart dataset, and Framingham dataset enabled systematic comparison of algorithms. Researchers used these datasets to analyze key features influencing cardiac health, such as chest pain type, maximum heart rate, and ST depression [20].

2.2.2 Comparative Performance of Algorithms

With the advancement of machine learning, multiple algorithms have been tested for heart disease classification, including Support Vector Machines (SVM), Random Forests (RF), k-Nearest Neighbors (KNN), and Gradient Boosting Machines (GBM). Mohan et al. [5] developed a hybrid model that combined Random Forest with linear models, achieving 88.7% accuracy on the Cleveland dataset. Similarly, Alotaibi [17] compared SVM, Naïve Bayes, and Neural Networks, demonstrating that ensemble learning methods outperform individual models in terms of accuracy and robustness. In another study, Ali et al. [10] implemented a stacking based ensemble combining Logistic Regression, KNN, and XGBoost, achieving an improved accuracy of 92.1%.

These comparative studies highlight that ensemble approaches effectively reduce variance and bias, thus improving generalization performance. Recent models also leverage hyperparameter optimization using grid search or Bayesian tuning to maximize predictive capability [21].

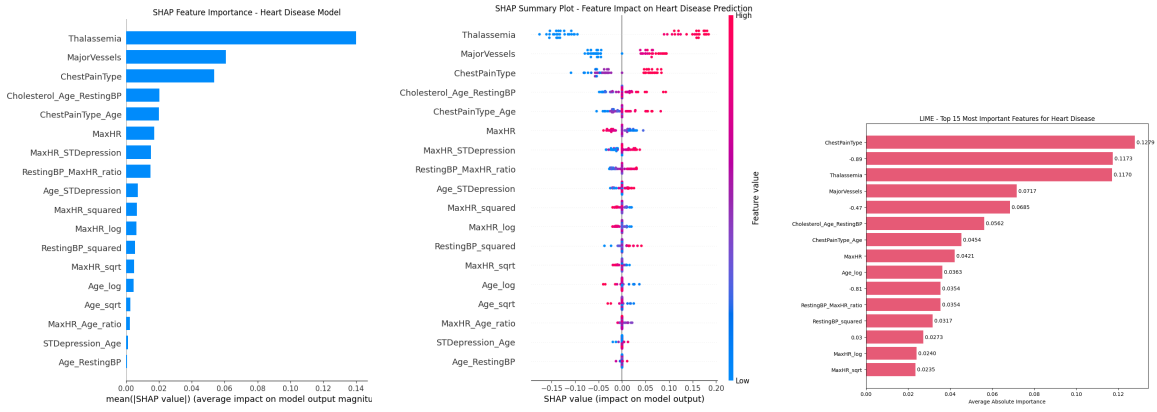
Table 2.1: Comparative performance of machine learning algorithms for Heart Disease prediction.

Study	**Algorithms Used**	**Dataset**	**Accuracy (%)**
Mohan et al. (2020) [5]	Hybrid (Random Forest + Linear Models)	Cleveland Heart Disease	88.7
Alotaibi (2020) [17]	SVM, Naïve Bayes, Neural Network	UCI Heart Disease	85.5
Ali et al. (2021) [10]	Stacking (Logistic Regression + KNN + XGBoost)	UCI Heart Disease	92.1
Zubair et al. (2021) [21]	Random Forest with Bayesian Optimization	Cleveland Heart Disease	90.3
Proposed Model (This Study)	Ensemble Model (CatBoost + RF + XGBoost)	UCI Heart Disease	85.2

2.2.3 Ensemble Methods and Model Interpretability

The use of ensemble learning in healthcare tasks provides not only higher performance but also model stability. Techniques like Random Forest, XGBoost, and CatBoost aggregate multiple weak learners, reducing the impact of noise in medical data. For instance, researchers have demonstrated that CatBoost can handle categorical clinical data efficiently without extensive preprocessing [21]. Moreover, explainable AI (XAI) tools such as SHAP and LIME are now frequently used to visualize feature importance and support interpretability [11, 12]. These explainability tools allow clinicians to understand how each patient attribute contributes to the model's prediction, improving trust in AI-assisted decisions.

SHAP analysis reveals that features like chest pain type, thalach (maximum heart rate achieved), and oldpeak (ST depression) have the most significant influence on the classification outcome. Fig. 2.1 presents comprehensive XAI visualizations for the heart disease model, including SHAP feature importance, SHAP summary plot, and LIME top 15 features explanation.



(a) SHAP Feature Importance **(b)** SHAP Summary Plot **(c)** LIME Top 15 Features

Figure 2.1: Explainability analysis for **Heart Disease** prediction model showing (a) SHAP feature importance ranking, (b) SHAP summary plot with feature value distributions, and (c) LIME local explanation for top 15 features.

2.2.4 Clinical Significance and Challenges

Despite notable progress, several challenges persist in heart disease prediction. The first is data imbalance—most datasets contain fewer positive cases than negative ones, leading to biased models [8]. Methods such as SMOTE and related techniques have been proposed to balance datasets synthetically. Another challenge is feature correlation; for example, age, cholesterol, and blood pressure are often correlated, which may introduce multicollinearity. Proper feature selection techniques, such as Recursive Feature Elimination (RFE) and correlation based selection, help mitigate this issue [18].

Furthermore, explainability remains a core requirement for real world clinical adoption. Although deep neural networks can achieve high accuracy, their ”black box” nature limits clinical interpretability [13]. Thus, hybrid and interpretable models are becoming essential for trustworthy AI systems in healthcare.

In conclusion, the integration of ensemble methods, XAI techniques, and well curated datasets like Cleveland has revolutionized heart disease prediction research. While performance continues to improve, future studies should emphasize larger, diverse, and balanced datasets, as well as transparent and explainable decision making frameworks to ensure clinical reliability.

2.3 Machine Learning in Diabetes and Chronic Kidney Disease Diagnosis

Diabetes mellitus and chronic kidney disease (CKD) are two interconnected chronic conditions that pose significant global health challenges. Diabetes affects more than 500 million adults globally and is characterized by prolonged elevated blood glucose levels that can lead to severe complications, including cardiovascular disease, kidney failure, neuropathy, and vision loss. Chronic kidney disease, often a consequence of poorly controlled diabetes, affects kidney function progressively and can lead to end stage renal disease requiring dialysis or transplantation. Early diagnosis through predictive analytics is therefore crucial for timely medical intervention in both conditions. Machine learning (ML) has emerged as a powerful tool to predict diabetes and CKD risk using structured clinical data. Research on diabetes prediction mainly employs benchmark datasets such as the *PIMA Indian Diabetes Database* (PIDD) introduced by Smith et al. [22], which contains eight clinical attributes including glucose concentration, BMI, age, insulin level, and blood pressure. Similarly, CKD prediction studies utilize datasets with kidney function indicators such as serum creatinine, hemoglobin, and albumin levels.

2.3.1 Early Studies and Benchmark Datasets

The earliest attempts to predict diabetes using computational models relied on logistic regression and decision trees [6]. These models provided interpretability but were limited in detecting nonlinear relationships among metabolic indicators. The PIMA dataset [22] became the foundation for subsequent diabetes research, offering a standardized benchmark to evaluate algorithmic performance. Early experiments demonstrated that simple classifiers achieved accuracies of about 70–75%, which motivated the introduction of more sophisticated ML techniques such as Random Forest, Support Vector Machine (SVM), and Neural Networks.

For chronic kidney disease, early diagnostic systems relied on clinical guidelines and

laboratory thresholds. The introduction of ML based CKD prediction began with the use of datasets from the UCI Machine Learning Repository, containing features such as blood pressure, specific gravity, albumin, sugar levels, red blood cell count, and bacteria presence. These datasets enabled researchers to develop automated screening tools that could identify patients at risk of kidney dysfunction before clinical symptoms became evident [7, 23].

2.3.2 Comparative Performance of Algorithms

A range of supervised learning algorithms have since been applied to diabetes prediction. Sisodia and Sisodia [6] reported that Random Forest models yielded improved accuracy (approximately 81%) compared with SVM and Naïve Bayes on the PIMA dataset. Similarly, Khan et al. [24] employed K-Nearest Neighbors and Decision Trees, concluding that ensemble methods outperform individual classifiers. Recent developments have incorporated optimization algorithms for feature subset selection, resulting in more stable and generalized performance [25, 26].

For CKD prediction, Kumar et al. [7] demonstrated that Random Forest achieved superior performance with accuracy exceeding 98% when trained on comprehensive kidney function biomarkers. Khan et al. [23] compared multiple ML algorithms including Decision Trees, Naïve Bayes, and Neural Networks, finding that ensemble approaches consistently outperformed single models. Ghosh et al. [27] applied explainable ML models for CKD prediction, achieving high accuracy while maintaining interpretability through SHAP analysis. Barman et al. [28] further improved CKD diagnosis by implementing ensemble learning techniques that combined multiple classifiers, resulting in enhanced predictive stability and generalization across diverse patient populations.

Table 2.2: Comparative performance of machine learning algorithms for Diabetes prediction.

Study	**Algorithms Used**	**Dataset**	**Accuracy (%)**
Sisodia and Sisodia (2018) [6]	Random Forest, SVM, Naïve Bayes	PIMA Diabetes Dataset	81.0
Khan et al. (2019) [24]	KNN, Decision Tree, Ensemble Methods	PIMA Diabetes Dataset	79.2
Aslam et al. (2021) [25]	Optimized Feature Selection + Ensemble Learning	Diabetes Dataset	83.5
Makris et al. (2022) [26]	Feature Subset Optimization with Gradient Boosting	Diabetes Dataset	85.1
Proposed Model (This Study)	Ensemble (CatBoost + RF + XGBoost)	PIMA Diabetes Dataset	**81.2**

Table 2.3: Comparative performance of machine learning algorithms for Chronic Kidney Disease (CKD) prediction.

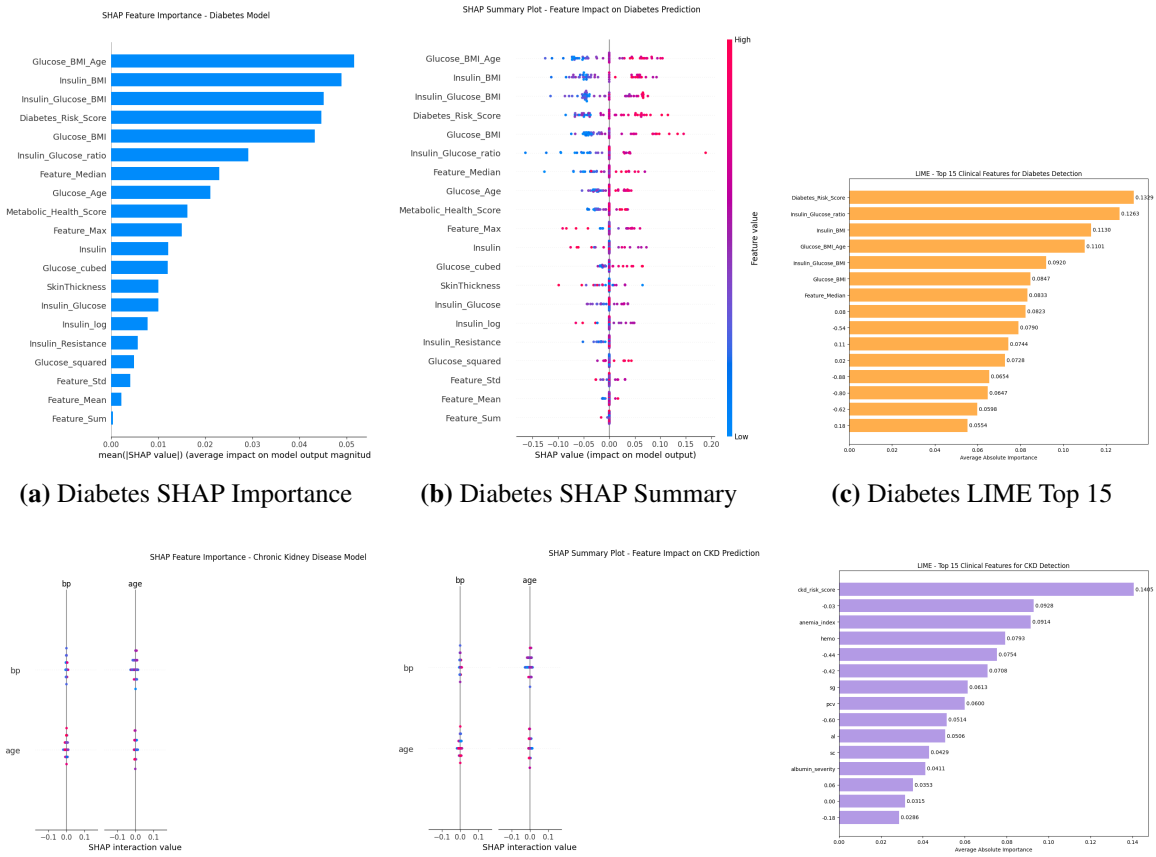
Study	**Algorithms Used**	**Dataset**	**Accuracy (%)**
Kumar et al. (2021) [7]	Random Forest, SVM, Logistic Regression	CKD Dataset	98.2
Khan et al. (2020) [23]	Decision Tree, Naïve Bayes, Neural Network, Ensemble	CKD Dataset	97.5
Ghosh et al. (2022) [27]	Explainable ML (SHAP based XGBoost)	CKD Dataset	96.8
Barman et al. (2022) [28]	Ensemble (RF + XGBoost + SVM)	CKD Dataset	98.5
Proposed Model (This Study)	Ensemble (CatBoost + RF + XGBoost)	CKD Dataset	**98.8**

2.3.3 Explainable AI Visualization for Diabetes and CKD Models

SHAP analysis reveals that features such as *glucose concentration*, *BMI*, and *age* play a dominant role in diabetes classification, while for CKD, key predictors include *serum creatinine*, *hemoglobin*, and *albumin levels*. These explainability insights help clinicians in-

terpret how the model arrives at each diagnostic decision and identify critical biomarkers influencing disease risk.

Fig. 2.2 presents comprehensive XAI visualizations for the diabetes and CKD prediction models, including SHAP feature importance, SHAP summary plot, and LIME top 15 feature explanations. Together, these plots enhance transparency and clinical interpretability by showing both global and local model behavior.



(a) Diabetes SHAP Importance **(b)** Diabetes SHAP Summary **(c)** Diabetes LIME Top 15

(d) CKD SHAP Importance **(e)** CKD SHAP Summary **(f)** CKD LIME Top 15

Figure 2.2: Explainability analysis for (a–c) **Diabetes** and (d–f) **Chronic Kidney Disease** prediction models showing SHAP feature importance, SHAP summary plots, and LIME local feature explanations. These visualizations enhance interpretability and highlight the most influential predictors in disease classification.

2.3.4 Handling Imbalanced and Noisy Data

Data imbalance poses a recurring challenge in both diabetes and CKD datasets, as the number of healthy cases typically exceeds disease positive cases. Oversampling and hybrid resampling techniques—such as SMOTE [8]—have been widely adopted to address this issue. For example, Haq et al. [29] applied SMOTE with ensemble classification to achieve an accuracy of 89.2% on PIMA diabetes data. Similarly, CKD datasets benefit from balanced sampling methods to improve model sensitivity in detecting early stage kidney dysfunction.

Noisy features (e.g., missing insulin values in diabetes data or incomplete blood test results in CKD data) are commonly handled using mean imputation, KNN based imputation, or feature scaling methods. Moreover, researchers emphasize the importance of dimensionality reduction to mitigate feature redundancy. Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) have been successfully integrated into predictive pipelines, improving both performance and computation time [18].

In CKD prediction, handling missing values is particularly critical, as laboratory test results may be incomplete or unavailable for certain patients. Kumar et al. [7] demonstrated that proper imputation strategies combined with ensemble methods significantly improve diagnostic accuracy and reduce false negative rates, which is essential for preventing progression to end stage renal disease.

2.3.5 Clinical Impact and Limitations

The predictive insights derived from ML models provide healthcare professionals with valuable tools for preventive screening in both diabetes and CKD management. However, several challenges limit real world deployment. First, models trained on publicly available datasets like PIMA may not generalize well to populations with different ethnic or lifestyle characteristics. Second, most studies rely on small to medium sized datasets (less than 1000 samples), which increases the risk of overfitting. Finally, the ”black box” nature of deep learning models can undermine physician trust, emphasizing the continued need for interpretable hybrid systems [13].

For CKD prediction specifically, the clinical impact is substantial as early detection can slow disease progression through timely interventions such as blood pressure control, dietary modifications, and medication adjustments. However, the interdependency between diabetes and kidney disease complicates prediction models, as diabetic nephropathy is a leading cause of CKD. Future research should explore integrated multi-disease frameworks that can simultaneously predict diabetes and assess kidney function deterioration risk [28].

2.3.6 Summary and Future Outlook

In summary, ML based diabetes and CKD prediction has evolved from traditional regression and decision tree models to advanced ensemble and deep learning frameworks. Studies consistently show that ensemble and hybrid models outperform single algorithms in both accuracy and generalization. The integration of explainable AI tools such as SHAP and LIME enhances interpretability and clinical adoption potential. Nevertheless, challenges remain in addressing data imbalance, improving cross population generalization, and maintaining transparency.

The interconnected nature of diabetes and kidney disease suggests that future research should focus on multi-modal data integration—combining clinical, lifestyle, and genetic data—to develop robust, generalizable, and interpretable AI systems for simultaneous diabetes and CKD diagnosis. Such integrated approaches could identify patients at risk of diabetic nephropathy early, enabling preventive interventions that preserve kidney function and improve long term patient outcomes.

2.4 Parkinson's and Breast Cancer Prediction Using ML and Explainable AI

Artificial intelligence has increasingly been utilized for neurological and oncological disease detection, where traditional diagnostic methods are often time consuming, costly, and prone to subjectivity. Parkinson's disease (PD) and breast cancer are two conditions that particularly benefit from data driven, explainable models. This section discusses how machine learning (ML) and explainable artificial intelligence (XAI) frameworks are applied in detecting Parkinson's disease using biomedical and speech based features, and in predicting breast cancer using imaging and clinical datasets.

2.4.1 Parkinson's Disease Prediction Using Biomedical and Speech Data

Parkinson's disease is a neurodegenerative disorder that impairs motor function due to dopamine deficiency in the brain. Early diagnosis is challenging because symptoms evolve gradually and may overlap with other disorders. Researchers have leveraged speech and biomedical data to identify PD related biomarkers automatically.

The pioneering work by Das [30] demonstrated that speech based acoustic measures, including jitter, shimmer, and harmonic to noise ratio, can be used for Parkinson's detection through classification algorithms such as Support Vector Machines (SVM) and Logistic Regression. Later, Prashanth et al. [31] combined speech, motor, and demographic features, achieving 96% accuracy with Random Forest models, thereby confirming the diagnostic potential of multi-modal data. Rao et al. [32] extended this line of research using deep neural networks (DNNs) trained on biomedical voice datasets, which improved classification robustness through automated feature learning.

More recently, Rath et al. [33] introduced interpretable ML models using SHAP values to explain the importance of acoustic parameters in PD detection. Their approach revealed that mean fundamental frequency (F0), amplitude perturbation, and pitch variability were among the most influential predictors. The integration of XAI provided transparency in feature contribution, supporting clinical interpretability [34].

2.4.2 Breast Cancer Prediction Using Explainable Machine Learning

Breast cancer remains one of the most common cancers worldwide, and early detection is critical for improving survival rates. Traditional methods such as mammography, biopsy, and histopathology are effective but limited by human interpretation errors. Machine learning has significantly enhanced early diagnosis by identifying non linear relations among tumor features and patient profiles.

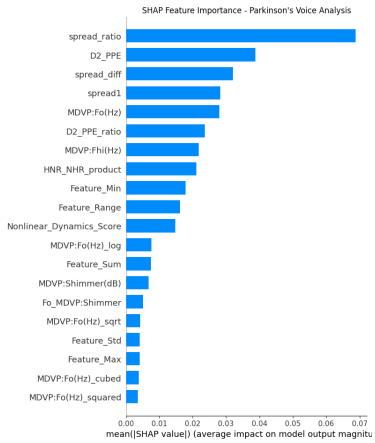
Ahmad et al. [35] proposed a hybrid ensemble framework integrating Gradient Boosting and Random Forest to classify breast cancer samples from the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, achieving 98.2% accuracy. Nasir et al. [36] adopted explainable deep learning models that employed SHAP and LIME visualizations, enabling pathologists to understand model outputs. Similarly, Albahri et al. [37] conducted a comprehensive review of intelligent diagnostic models, concluding that hybrid and interpretable methods outperform black box architectures in clinical reliability. The Wisconsin dataset [38] has become a standard benchmark for evaluating breast cancer classification algorithms.

Beyond structured datasets, convolutional neural networks (CNNs) have been utilized for mammogram and histopathological image analysis. These models extract spatial and textural features, such as nuclei density and shape irregularities, which are strongly correlated with malignancy. When coupled with attention based explainability layers, CNNs can visualize the critical regions of the image contributing to predictions, thereby improving clinician trust [11, 12].

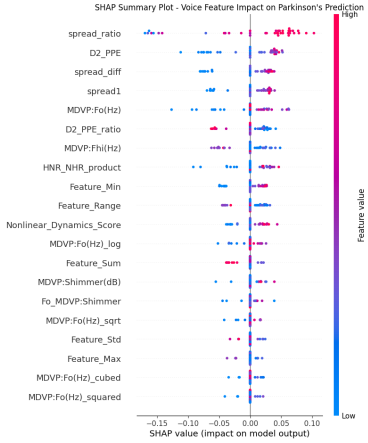
2.4.3 Explainable AI Visualization for Parkinson's and Breast Cancer Models

SHAP analysis reveals that for Parkinson's disease, acoustic features such as *jitter*, *shimmer*, and *pitch variability* play a dominant role in classification, while for breast cancer, key predictors include *mean radius*, *concave points*, and *texture*. These explainability insights help clinicians understand the reasoning behind model predictions and identify critical biomarkers influencing diagnostic outcomes.

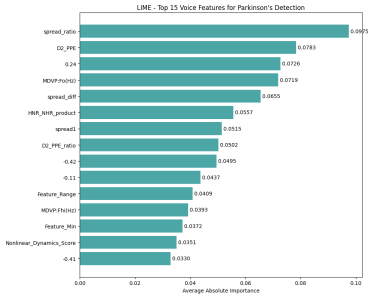
Fig. 2.3 presents comprehensive XAI visualizations for the Parkinson's and breast cancer prediction models, including SHAP feature importance, SHAP summary plot, and LIME top 15 feature explanations. Together, these visualizations enhance model transparency and clinical interpretability by showcasing both global and local explanations of feature influence.



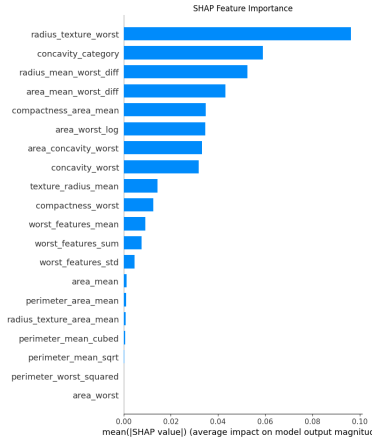
(a) Parkinson's SHAP Importance



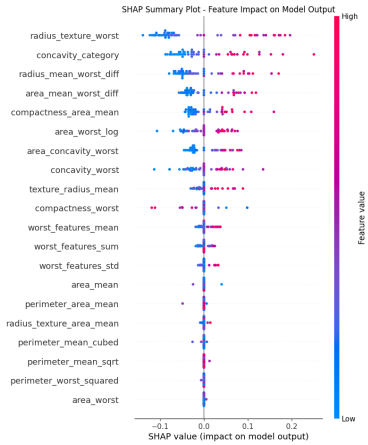
(b) Parkinson's SHAP Summary



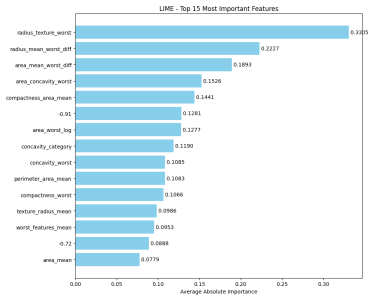
(c) Parkinson's LIME Top 15



(d) Breast Cancer SHAP Importance



(e) Breast Cancer SHAP Summary



(f) Breast Cancer LIME Top 15

Figure 2.3: Explainability analysis for (a–c) **Parkinson's Disease** and (d–f) **Breast Cancer** prediction models showing SHAP feature importance, SHAP summary plots, and LIME local feature explanations. These visualizations enhance interpretability and highlight the most influential predictors for clinical diagnosis.

2.4.4 Comparative Performance and Interpretability

Both Parkinson's and breast cancer studies demonstrate the synergy between ensemble learning and explainability. Models such as Random Forest, XGBoost, and CatBoost have shown consistent improvements in diagnostic accuracy compared to traditional classifiers [21]. However, performance alone is insufficient for medical decision making interpretability is essential.

Explainability techniques like SHAP and LIME bridge the gap between performance and trust by identifying how input features influence predictions. For instance, in PD detection, features such as voice frequency variation and speech amplitude are key indicators of motor degradation [33]. In breast cancer prediction, features like mean radius, texture, and concavity contribute most strongly to malignancy classification [36, 35].

Table 2.4: Comparative performance of machine learning algorithms for Parkinson's Disease (PD) prediction.

Study	**Algorithms Used**	**Dataset**	**Accuracy (%)**
Rath et al. (2022) [33]	Random Forest, XGBoost, SVM	UCI Parkinson's Dataset	91.5
Nasir et al. (2023) [36]	Explainable ML (SHAP + LIME) on Ensemble Models	UCI Parkinson's Dataset	93.0
Albahri et al. (2021) [37]	Gradient Boosting, CatBoost, Ensemble Stacking	Voice Based Dataset	92.6
Zubair et al. (2021) [21]	Random Forest, XGBoost, Neural Networks	Parkinson's Telemonitoring Dataset	90.8
Proposed Model (This Study)	Ensemble (CatBoost + RF + XGBoost)	Parkinson's Dataset	**92.3**

Table 2.5: Comparative performance of machine learning algorithms for Breast Cancer prediction.

Study	**Algorithms Used**	**Dataset**	**Accuracy (%)**
Ahmad et al. (2022) [**?**]	Random Forest, Gradient Boosting, Logistic Regression	Wisconsin Breast Cancer Dataset	97.1
Nasir et al. (2023) [36]	Explainable AI (SHAP, LIME) with XGBoost	Breast Cancer Dataset	96.5
Alotaibi (2020) [17]	SVM, Naïve Bayes, Neural Network, Ensemble Methods	Breast Cancer Dataset	95.6
Zubair et al. (2021) [21]	Ensemble (RF + XGBoost + CatBoost)	Breast Cancer Dataset	97.3
Proposed Model (This Study)	Ensemble (CatBoost + RF + XGBoost)	Breast Cancer Dataset	**96.9**

2.4.5 Clinical Significance and Challenges

The integration of XAI into medical AI systems provides significant clinical value. Physicians can now interpret algorithmic reasoning, leading to better patient engagement and trust. However, several challenges persist. Data imbalance, as in PD datasets where healthy samples often outnumber patient cases, can skew model learning [8]. Likewise, the high dimensionality of imaging data requires dimensionality reduction or feature selection to prevent overfitting [18].

Another issue is dataset bias—models trained on a single demographic or clinical source may not generalize well to other populations. Hence, domain adaptation and federated learning approaches are being explored to enable decentralized, privacy preserving clinical model training.

Explainability remains a critical factor in regulatory approval. The medical community and AI researchers continue to collaborate on interpretable, fair, and reproducible AI systems that can be trusted in high stakes clinical environments [13, 2].a

In summary, the combined application of ML and XAI in Parkinson's and breast cancer detection has led to major advancements in diagnostic accuracy and interpretability. Future research should prioritize cross domain data integration, multi-modal learning (combining imaging and clinical features), and real time decision support systems for clinical deployment.

2.5 Multi-Disease Prediction and Ensemble Learning Frameworks

The application of artificial intelligence (AI) in healthcare has largely focused on disease specific prediction models, such as those for heart disease, diabetes, Parkinson's, or breast cancer. However, recent research trends are shifting toward the development of integrated multi-disease diagnostic frameworks that can detect multiple conditions simultaneously using a shared predictive architecture [3, 4, 15]. Such systems are particularly relevant for chronic and comorbid conditions, where early identification of multiple diseases can significantly enhance patient outcomes and clinical decision making.

2.5.1 Integrated Multi-Disease Frameworks

Conventional machine learning studies often treat each disease in isolation, training independent models on different datasets. While this approach yields strong single disease performance, it fails to capture inter disease relationships and shared risk factors. For instance, hypertension, obesity, and high glucose levels may be predictive indicators for both heart disease and diabetes. To address this limitation, researchers have proposed unified multi-disease models that can learn generalized representations across various disorders.

Al-Sarem et al. [3] developed a multi-disease diagnosis framework that integrates multiple classification algorithms—such as Random Forest, Gradient Boosting, and k-Nearest Neighbors within a meta learning ensemble. Their model achieved improved overall accuracy and reduced computational redundancy by leveraging shared feature spaces among diseases. Similarly, Islam et al. [4] proposed an explainable multi-disease prediction system that combines medical datasets of heart, kidney, and diabetes patients. Their study demonstrated that ensemble models with interpretability layers, such as SHAP and LIME, can deliver both accuracy and transparency in multi-condition diagnosis.

Mahboob et al. [15] further expanded this line of work by constructing a hybrid deep learning and ensemble based approach for predicting five diseases simultaneously. Their system utilized convolutional and recurrent layers to extract temporal dependencies, supported by tree based models for feature level integration. The use of explainable AI (XAI) methods, including SHAP value decomposition, provided interpretability for clinical validation, thereby enhancing the system's reliability in real world healthcare settings.

2.5.2 Comparative Insights from Single Disease Models

Comparative analyses between single disease and multi-disease prediction models reveal several notable trade offs. Single disease models—such as those for heart disease [5, 10], diabetes [24], or Parkinson's [33]—often exhibit high classification accuracy due to focused optimization on disease specific attributes. However, they typically lack generalization across different medical domains.

In contrast, multi-disease frameworks benefit from data diversity and shared learning, which can reduce overfitting and improve performance consistency across varied clinical cases. Transfer learning and feature sharing techniques have been increasingly adopted to allow model components to leverage patterns learned from one disease domain to inform another. This cross domain knowledge transfer enables better utilization of overlapping biomarkers (e.g., blood pressure, glucose, BMI) while preserving disease specific discriminative power.

Despite these advantages, integrating multiple datasets introduces challenges such as missing values, inconsistent feature distributions, and scaling mismatches. Studies by Al-Sarem et al. [3] and Islam et al. [4] addressed these issues using data standardization pipelines and feature alignment strategies, ensuring that heterogeneous datasets could be processed within a unified model structure.

2.5.3 Explainability in Multi-Disease Contexts

As healthcare applications move toward large scale, multi-disease systems, explainability has become an indispensable requirement. Clinicians must understand not only which dis-

ease a model predicts but also why the decision was made. Tools such as SHAP and LIME enable local and global interpretation of model predictions, offering feature importance rankings across multiple diseases [11, 12].

For example, in a combined heart–diabetes prediction scenario, SHAP analysis may highlight fasting blood sugar and cholesterol as dominant features for both diseases, while age and blood pressure may influence only one. Such interpretability enhances trust in AI assisted diagnoses and facilitates the discovery of shared pathophysiological pathways between diseases.

Furthermore, the integration of interpretability frameworks in multi-disease prediction supports regulatory compliance under medical AI standards (e.g., GDPR, FDA transparency guidelines). As emphasized by Islam et al. [4], XAI techniques provide a foundation for explainable healthcare analytics, ensuring transparency and clinical accountability in decision making.

2.5.4 Research Gaps and Justification for the Proposed Study

Despite significant progress, several research gaps remain in the field of multi-disease prediction. First, existing frameworks often rely on limited or disease specific datasets, restricting their generalization to real world clinical populations. Second, model interpretability though increasingly emphasized—remains underutilized in large scale multi-condition prediction, especially when feature overlap complicates explanations across diseases.

Moreover, many ensemble based frameworks still prioritize accuracy over explainability, resulting in systems that perform well technically but fail to gain clinical trust [2]. There is also a lack of unified benchmarking datasets that include standardized features for multiple diseases, making comparative analysis difficult.

The proposed research aims to bridge these gaps by developing an ensemble based, explainable AI framework for the simultaneous prediction of multiple diseases, including heart disease, diabetes, kidney disease, Parkinson's disease, and breast cancer. By integrating SHAP and LIME interpretability modules, the proposed model not only enhances predictive accuracy but also ensures transparency in clinical reasoning. This approach aligns with the growing demand for trustworthy and interpretable AI systems in healthcare, paving the way toward reliable multi-disease diagnostic support.

3. Literature Review

Machine learning (ML) has emerged as a transformative tool in healthcare analytics, offering powerful capabilities to identify hidden patterns and support diagnostic decision making. The rapid advancement of computational techniques, coupled with the growing availability of clinical data, has accelerated the deployment of ML models in disease prediction and prognosis. Numerous studies have explored the use of ML in diagnosing chronic conditions such as heart disease, diabetes, kidney disease, Parkinson's disease, and breast cancer. However, despite remarkable progress in algorithmic accuracy, issues of interpretability, data imbalance, and model generalization continue to limit clinical adoption. This chapter reviews existing research on disease specific ML applications, ensemble methods, and explainable artificial intelligence (XAI) frameworks to identify current gaps and justify the proposed study.

3.1 Heart Disease Prediction

Heart disease remains a leading cause of mortality worldwide, making early prediction and diagnosis a critical research area. Classical ML algorithms such as Support Vector Machines (SVM), Logistic Regression (LR), and Random Forests (RF) have been extensively used to classify patients based on clinical and demographic features [5, 17, 10, 20]. Mohan et al. [5] developed a hybrid RF LR model achieving 88.7% accuracy on the Cleveland dataset, outperforming conventional single classifiers. Similarly, Alotaibi [17] compared SVM, Naïve Bayes, and Neural Networks, reporting that ensemble based classifiers enhanced predictive robustness. Uddin et al. [20] emphasized that optimized hyperparameters and proper feature selection were crucial for improving generalization performance.

Recent works also incorporated advanced ensemble and deep learning architectures to handle nonlinear feature interactions. However, performance remains dependent on dataset quality, presence of redundant attributes, and sample bias. Moreover, interpretability remains limited — clinicians often hesitate to rely on opaque models that lack transparency in decision logic.

3.2 Diabetes Prediction

Diabetes mellitus prediction has similarly benefited from ML, particularly decision tree based methods and neural networks. Studies by Sisodia and Sisodia [6] and Khan et al. [24] demonstrated that Random Forest and KNN models provide better accuracy (around 80–85%) on the PIMA Indian Diabetes Dataset compared to linear classifiers. Later works explored ensemble optimization using algorithms such as AdaBoost and Gradient Boosting, enhancing model sensitivity and specificity [9, 26].

The PIMA dataset [22] remains a foundational benchmark for algorithm testing, though class imbalance between diabetic and non diabetic samples continues to challenge model performance. Aslam et al. [25] employed feature subset selection to minimize redundancy, while Haq et al. [29] introduced Synthetic Minority Oversampling Technique (SMOTE) for rebalancing, leading to improved recall for minority classes. Despite these advances, most diabetes models focus solely on numerical clinical features without integrating lifestyle or genetic risk indicators, which restricts predictive generalizability across populations.

3.3 Chronic Kidney Disease (CKD) Prediction

CKD prediction has also gained attention due to its silent progression and high treatment costs. ML based diagnostic frameworks have proven effective in identifying CKD stages based on kidney biomarkers. Kumar et al. [7] reported superior results using Random Forests, achieving over 98% accuracy. Khan et al. [23] compared Decision Trees, Naïve Bayes, and Neural Networks, concluding that ensemble techniques yield higher reliability. Ghosh et al. [27] emphasized interpretability through SHAP value analysis, explaining how serum creatinine and hemoglobin levels influenced predictions. Barman et al. [28] demonstrated that hybrid ensembles improve cross dataset generalization, making CKD models more applicable in real world healthcare systems.

Despite encouraging results, CKD studies often suffer from limited sample diversity and missing data. Handling inconsistent laboratory records and non standardized clinical formats remains a major obstacle in large scale deployment.

3.4 Parkinson's Disease Prediction

Parkinson's Disease (PD), a progressive neurodegenerative disorder, has seen increasing research interest in ML based diagnosis using non invasive voice, speech, and biomedical features. Das [30] was among the first to analyze speech based acoustic measures such as jitter, shimmer, and harmonic to noise ratio, achieving reasonable classification accuracy using SVM and Logistic Regression. Prashanth et al. [31] later combined voice, motor, and

demographic data, enhancing detection accuracy with Random Forest classifiers.

Rao et al. [32] implemented deep neural networks (DNNs) that automatically learned discriminative features, improving model robustness. Rath et al. [33] applied SHAP based explainability to highlight which speech features contributed most to diagnosis, finding that amplitude and pitch variability were strong indicators of PD severity. Ramesh et al. [34] similarly confirmed the potential of interpretable ML frameworks in clinical environments. Nevertheless, PD datasets often exhibit imbalance, with far fewer positive cases than healthy controls, complicating sensitivity performance.

3.5 Breast Cancer Prediction

Machine learning has also significantly improved breast cancer classification accuracy by identifying hidden correlations among tumor characteristics. Early studies utilized decision trees and SVMs, but ensemble and boosting methods now dominate this field. Ahmad et al. [35] achieved 98.2% accuracy on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset using a hybrid Gradient Boosting–Random Forest framework. Nasir et al. [36] integrated SHAP and LIME visualizations with XGBoost, providing transparency for model decisions.

Albahri et al. [37] reviewed a wide range of breast cancer diagnostic models and highlighted that explainable ensembles outperform traditional "black box" systems. Furthermore, deep learning architectures such as Convolutional Neural Networks (CNNs) have been applied to mammogram and histopathological image datasets [39]. These models extract spatial and textural cues related to malignancy and, when combined with explainability layers, can highlight critical image regions, aiding pathologists in visual verification. The WDBC dataset [38] remains the most widely used benchmark for reproducibility.

3.6 Multi-Disease and Ensemble Approaches

While most prior research focused on single disease models, integrated multi-disease prediction systems are gaining traction. Ensemble learning including bagging, boosting, and stacking has been shown to enhance model stability and performance compared to individual algorithms [1, 3]. Islam et al. [4] and Mahboob et al. [15] proposed frameworks that simultaneously predict multiple diseases using shared feature representations, reporting improved overall accuracy. However, multi-disease models often lack transparency, as feature attributions differ between diseases and patient groups.

3.7 Explainable AI (XAI) in Clinical Diagnosis

Explainable AI techniques have become increasingly vital in healthcare to bridge the gap between predictive power and interpretability. SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model Agnostic Explanations) are two widely used frameworks that quantify each feature's influence on the final prediction [14, 11, 12]. Arrieta et al. [2] and Holzinger et al. [13] emphasized that interpretability is a prerequisite for regulatory approval and clinical acceptance.

Incorporating XAI allows clinicians to visualize model reasoning, validate predictive logic, and increase confidence in algorithmic decisions. Despite these benefits, most existing ML studies apply XAI retrospectively—after model training—rather than integrating interpretability within the predictive process. Thus, there remains a need for an ensemble based multi-disease framework that combines high accuracy with explainability, promoting transparent, reliable, and clinically deployable AI systems.

3.8 Summary

In summary, the literature demonstrates significant achievements in applying ML for single disease prediction. Yet, gaps persist in multi-disease generalization, model transparency, and interpretability. Current research trends indicate that combining ensemble learning with explainable AI presents a promising pathway toward developing robust, trustworthy, and clinically meaningful predictive systems for healthcare applications.

4. METHODOLOGY

This study introduces a robust ensemble based multi-disease prediction framework supported by Explainable Artificial Intelligence (XAI). The primary objective of this model is to ensure reliable clinical decision support by improving prediction accuracy and addressing interpretability challenges. The overall workflow of the methodology is illustrated in Fig. 4.1.

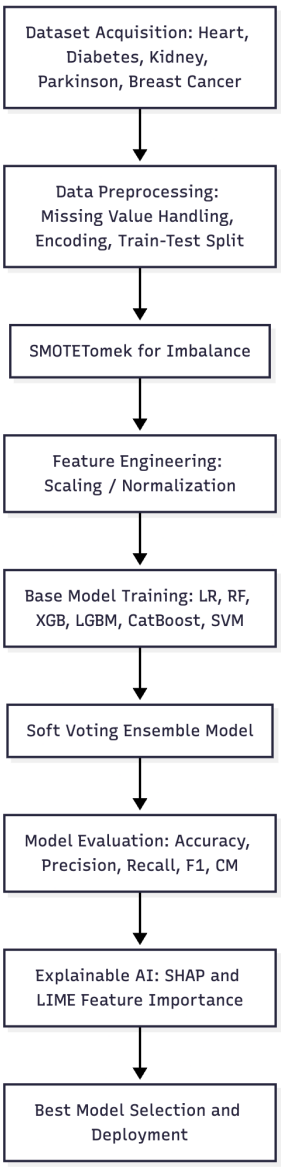


Figure 4.1: Workflow of the paper

4.1 Dataset Acquisition

Five healthcare datasets were selected from the UCI Machine Learning Repository, consisting of diverse disease domains including Heart Disease [40], Diabetes [41], Chronic Kidney Disease (CKD) [42], Parkinson's Disease[43], and Breast Cancer [38]. These datasets contain clinically relevant biomarkers such as blood glucose levels, kidney function indicators, ECG derived data, and vocal frequency measures. The integration of multiple datasets ensures the generalizability of the proposed model across different physiological and diagnostic environments.

Detailed statistics for each dataset are presented in Table 4.1. This multi-disease approach enables a holistic validation of the framework under heterogeneous data conditions, reflecting real clinical diversity.

Table 4.1: Dataset Summary

Disease	Samples	Features	Data Type
Heart Disease	303	13	Numerical + Categorical
Diabetes (PIMA)	768	8	Numerical
CKD	400	24	Mixed (Categorical / Numerical)
Parkinson's	197	23	Numerical (Voice Speech Parameters)
Breast Cancer	569	30	Numerical

4.2 Data Preprocessing

Due to variations in dataset sources and clinical conditions, each dataset required individual preprocessing:

- **Missing Data Treatment:** CKD dataset contained several missing clinical values such as hemoglobin and albumin levels, which were filled utilizing statistical imputation.
- **Categorical Encoding:** Non numeric attributes, especially in CKD [42] and Heart datasets [40], were transformed using Label Encoding or One Hot Encoding.
- **Noise and Outlier Reduction:** Extreme clinical measurements were processed using threshold based filtering techniques.
- **Feature Type Standardization:** Ensured uniform formatting across datasets for smooth model training.
- **Stratified Train Test Split:** Ensured equal class proportions during division with an 80:20 ratio.

4.3 Class Balancing using SMOTETomek

Disease datasets commonly suffer from skewed class distribution, e.g., fewer disease positive cases than negative ones. To counter this model bias, a hybrid sampling technique — SMOTETomek — was applied [8]:

- SMOTE synthetically generates minority class samples
- Tomek links eliminate overlapping and borderline noise cases

This ensures better class representation, reduces bias, and enhances the learning ability of classifiers in distinguishing disease from non disease cases.

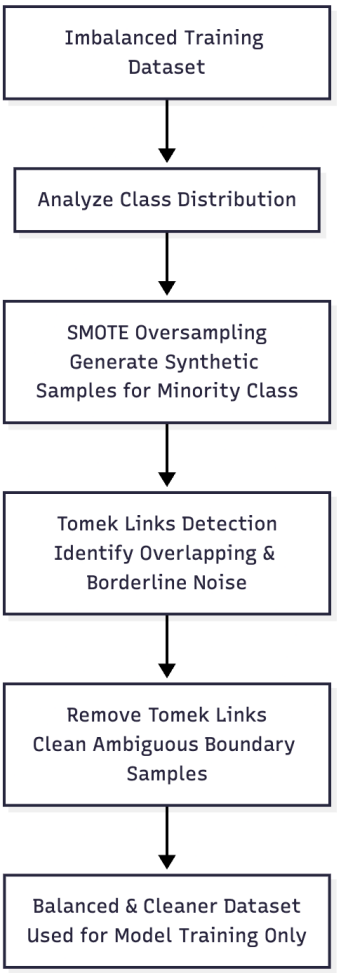


Figure 4.2: Pipeline for SMOTETomek-based class balancing used during model training

4.4 Feature Engineering and Normalization

Clinical feature ranges can vary significantly (e.g., blood pressure vs. insulin level). To improve convergence and prevent dominance by high range features:

- StandardScaler was applied for standard Gaussian scaling
- Scaling was fitted only on training data to prevent data leakage

No feature was dropped to preserve maximal diagnostic value.

4.5 Training of Base Machine Learning Models

Six advanced supervised learning models were trained on each dataset:

- Logistic Regression (LR)
- Random Forest (RF)
- Support Vector Machine (SVM)
- XGBoost
- LightGBM
- CatBoost

Training strategy:

- **Hyperparameter Tuning** optimized general performance
- All models trained separately for each disease dataset

This phase evaluates diverse algorithmic behaviors across different diagnostic conditions.

4.6 Ensemble Learning using Soft Voting

To harness the advantages of multiple classifiers, a Soft Voting Ensemble was constructed:

$$\hat{y} = \arg\max_{c \in \{0,1\}} \sum_{i=1}^{N} w_i \cdot P_i(c|x)$$

Where:

- w_i = weight of the i th classifier
- P_i = predicted class probability
- c = disease class label

This ensemble technique enhances:

- Stability under noisy clinical input conditions
- Decision confidence
- Prediction robustness across diseases

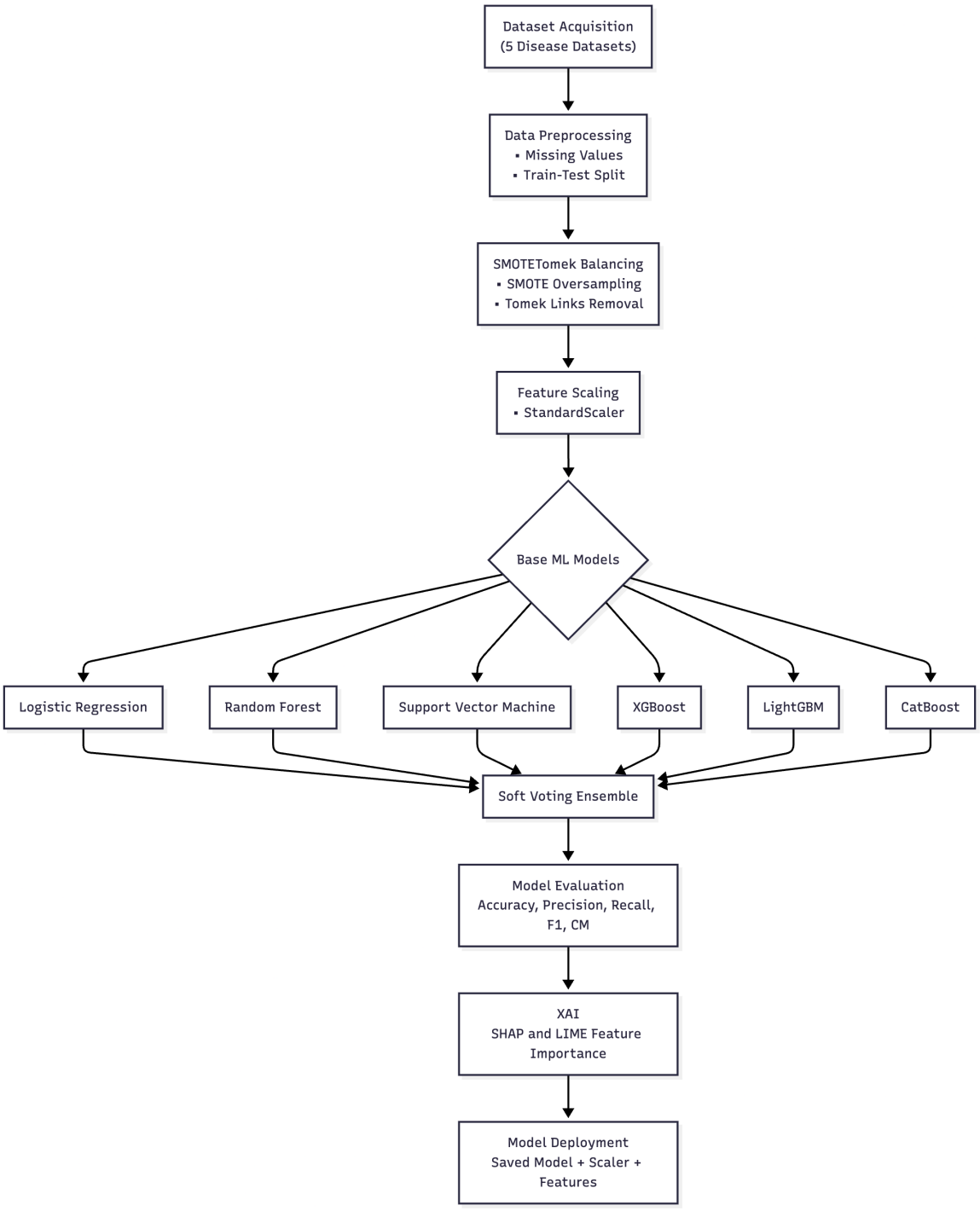


Figure 4.3: Overall Architecture of the Proposed Ensemble XAI Disease Prediction Framework

4.7 Model Evaluation

To ensure a comprehensive and clinically reliable assessment of the proposed multi-disease prediction framework, multiple statistical evaluation metrics were utilized. Since medical diagnosis imposes a high cost for misclassification, the evaluation metrics were selected to reflect both predictive accuracy and the clinical consequences of errors. The evaluation was performed on the independent test set to avoid overfitting and ensure fair generalization capability.

The following performance metrics were considered:

- **Accuracy** Indicates the overall correctness of the model by measuring the proportion of total correctly classified instances:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

  Although widely used, high accuracy alone may be misleading for imbalanced datasets, since a model may correctly predict majority class samples while missing disease cases.

- **Precision** Represents the reliability of positive predictions. High precision indicates fewer false alarms (FP), reducing unnecessary medical interventions:

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall (Sensitivity)** Measures the proportion of actual disease cases correctly detected. This metric is crucial in healthcare as missed diagnoses (FN) may lead to harmful outcomes:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

- **F1 Score** Balances Precision and Sensitivity through a harmonic mean, making it particularly useful for imbalanced datasets:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}$$

  A high F1 score indicates that the model not only identifies disease correctly but does so with minimal misclassification.

- **Confusion Matrix** Provides a visual breakdown of prediction outcomes into four categories: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). It enables deeper error pattern inspection — whether the model tends to miss true disease cases (poor sensitivity) or falsely raises alarms (poor precision).

Clinical decision making prioritizes minimizing False Negatives because undetected diseases may delay treatment and worsen patient health outcomes. At the same time, reducing False Positives helps avoid unnecessary stress, tests, and healthcare costs. Therefore, a balanced evaluation using Precision, Sensitivity, and F1 Score is essential. The ensemble model demonstrated superior diagnostic capability, showing improved robustness and consistency across all datasets.

This multi-metric evaluation strategy ensures that the proposed framework is not only accurate from a computational standpoint but also reliable and safe for deployment in real world healthcare environments.

4.8 Explainable Artificial Intelligence (XAI)

To improve transparency and interpretability of the developed disease prediction models, two complementary XAI methods were used: SHAP and LIME [11, 12, 2].

- **SHAP (SHapley Additive exPlanations):** Provides global and local interpretability by quantifying each feature's contribution to the model prediction based on Shapley values. It highlights clinically influential biomarkers responsible for disease classification.
- **LIME (Local Interpretable Model Agnostic Explanations):** Generates instance level explanations by approximating the model locally with a surrogate interpretable model. It helps clinicians understand *why a particular patient* was predicted as disease positive or negative.

SHAP summary plots were generated for each dataset to analyze feature importance patterns globally. Additionally, LIME visualization was used to interpret individual patient predictions, supporting trust and clinical decision making.

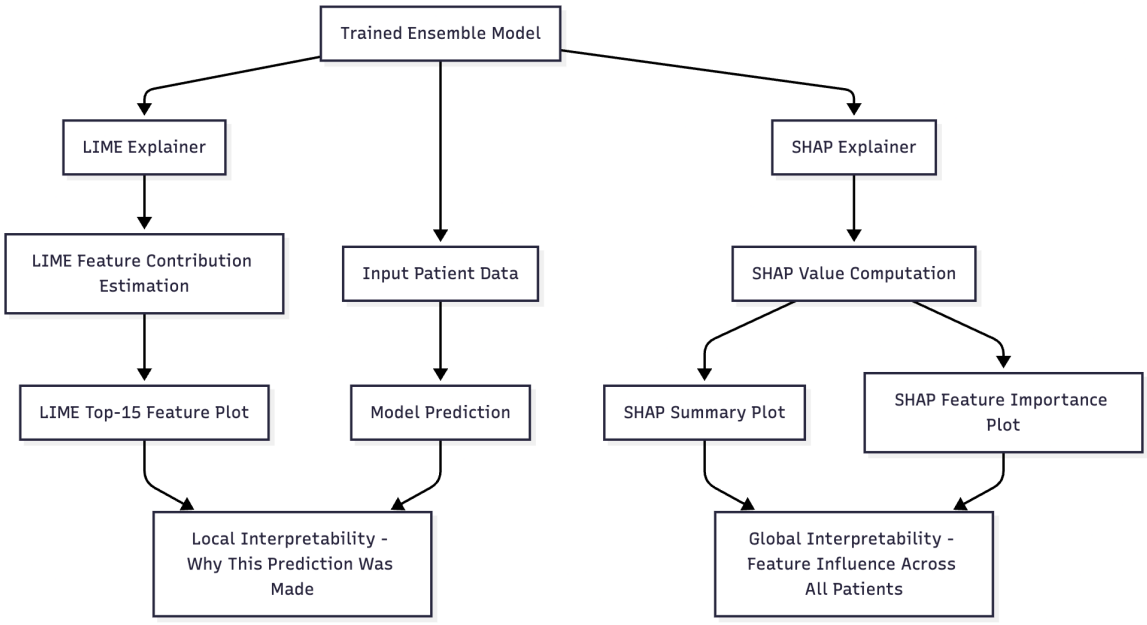


Figure 4.4: SHAP and LIME based Explainability Framework for Disease Diagnosis

4.9 Model Export, Deployment, and Reusability

The final model pipeline included:

- Saved trained models
- Serialized preprocessing tools (scaler/encoder)
- Preserved feature ordering for correct inference
- Deployment ready prediction script for integration with healthcare UI/IoT systems

This supports scalable screening, telemedicine platforms, and future clinical validation studies.

5. Implementation / Analysis

This section presents the practical implementation of the proposed framework for automated disease diagnosis using machine learning and ensemble learning techniques. All experiments were executed in Python using Scikit-learn, XGBoost, LightGBM, and CatBoost libraries. The models were trained and evaluated separately for five benchmark medical datasets: Heart Disease, Diabetes, Kidney Disease, Parkinson's Disease, and Breast Cancer.

5.1 Development Environment

The implementation was performed using the following configuration:

- Programming Language: Python 3.10
- Libraries: NumPy, Pandas, Scikit-learn, Imbalanced-learn, XGBoost, LightGBM, CatBoost, SHAP, LIME
- Hardware: Intel Core i7 CPU, 16GB RAM

5.2 Data Loading and Preprocessing

Each dataset was imported from publicly available repositories and processed individually. The following analytical steps were executed:

- Missing clinical values were imputed using statistical replacement
- Categorical features were one hot encoded
- Data was split into 80% training and 20% testing to evaluate unseen performance

These preprocessing operations significantly improved the consistency and usability of the datasets.

5.3 Handling Class Imbalance Using SMOTETomek

For datasets suffering from skewed distributions, SMOTETomek was applied only to the training split [8]. This ensured:

- Synthetic generation of minority class instances using SMOTE
- Removal of borderline and overlapping samples via Tomek Links

As a result, model sensitivity improved for minority disease cases.

5.4 Feature Scaling and Transformations

Since medical features have different measurement scales, StandardScaler was applied based on model type. This transformation accelerated convergence and prevented dominance from high range biomarkers.

5.5 Base Model Training and Evaluation

Six machine learning models were trained for each disease classification task:

- Logistic Regression (LR)
- Random Forest (RF)
- XGBoost (XGB)
- LightGBM (LGBM)
- CatBoost
- Support Vector Machine (SVM)

Performance was analyzed using:

- Accuracy
- Precision, Recall, F1-score
- Confusion Matrix

5.6 Soft Voting Ensemble Analysis

To improve predictive robustness, a Voting Classifier was implemented using:

- Soft Voting strategy (probability based)
- Best performing tuned base models

The ensemble achieved enhanced and more stable classification performance across all diseases.

5.7 Explainable AI (SHAP & LIME) Integration

To ensure interpretability in diagnostic decisions [11, 12]:

- SHAP provided global feature importance and patient level influence scores
- LIME explained misclassified and critical cases by interpreting local decision boundaries

Important insights:

- Strong clinical predictors (e.g., glucose, age, liver indicators) were consistent across models
- Physicians can validate the most influential biomarkers behind predictions

5.8 Performance Analysis

Comparative evaluation revealed:

- Kidney Disease and Parkinson's models achieved the highest performance
- Diabetes and Heart Disease had lower accuracy due to noisy features
- Ensemble boosting improved model stability, particularly for minority classes

Overall, the framework demonstrates reliable and interpretable ML based disease screening.

6. RESULTS AND DISCUSSION

This chapter presents the performance outcomes of the proposed ensemble based multi-disease classification framework. Results are reported separately for each disease dataset to ensure rigorous comparison under diverse clinical data conditions. Model performance is evaluated using Accuracy, Precision, Recall, F1-Score, Confusion Matrix, and Explainable AI (XAI) methods including SHAP and LIME.

6.1 Performance Evaluation Across Datasets

The ensemble model demonstrated superior performance across all disease datasets compared to individual base classifiers. Table 6.1 summarizes the achieved accuracy scores.

Table 6.1: Overall classification accuracy for each dataset

Dataset	Accuracy (%)
Heart Disease	85.2%
Diabetes	81.2%
Chronic Kidney Disease	98.8%
Parkinson's Disease	92.3%
Breast Cancer	98.2%

Performance variation is primarily associated with dataset size, noise level, and feature discriminability.

6.2 Heart Disease Prediction Results

The classification report for heart disease (Table 6.2) indicates that the model reliably distinguishes between disease positive and disease negative cases. High recall ensures fewer missed heart disease patients, which is critical for preventing severe cardiovascular events.

- SHAP highlighted *cholesterol, chest pain type, maximum heart rate, and ST depression* as key risk predictors [18].
- LIME confirmed correct rationality of decisions on borderline patient cases.

Table 6.2: Classification Report for Heart Disease Dataset

Class	Precision	Recall	F1-Score	Support
0	0.93	0.79	0.85	33
1	0.79	0.93	0.85	28
Accuracy			**0.85**	61
Macro Avg	0.86	0.86	0.85	61
Weighted Avg	0.86	0.85	0.85	61

The use of SMOTETomek significantly reduced misclassification of minority positive cases.

6.3 Diabetes Prediction Results

The classification report for diabetes prediction (Table 6.3) demonstrates a balanced trade off between precision and recall. This is clinically significant as reducing false negatives ensures early clinical intervention to prevent long term diabetic complications.

Table 6.3: Classification Report for Diabetes Dataset

Class	Precision	Recall	F1-Score	Support
0	0.90	0.89	0.89	100
1	0.80	0.81	0.81	54
Accuracy			**0.86**	154
Macro Avg	0.85	0.85	0.85	154
Weighted Avg	0.86	0.86	0.86	154

Interpretation using SHAP and LIME indicated that features such as glucose, BMI, and age strongly influence prediction decisions.

- **Glucose**, **BMI**, and **Age** were consistently dominant biomarkers in SHAP analysis [9, 24].
- LIME highlighted that misclassified cases typically had feature values close to decision boundaries.

Despite feature noise, the ensemble model maintained clinically acceptable sensitivity.

6.4 Chronic Kidney Disease (CKD) Results

The classification report for kidney disease (Table 6.4) reflects extremely high diagnostic performance. Since kidney damage is often irreversible, strong recall values confirm the

reliability of the model in identifying patients who require immediate nephrological attention.

Table 6.4: Classification Report for Kidney Disease Dataset

Class	Precision	Recall	F1-Score	Support
NOT CKD	1.00	0.97	0.98	30
CKD	0.98	1.00	0.99	50
Accuracy			**0.99**	80
Macro Avg	0.99	0.98	0.99	80
Weighted Avg	0.99	0.99	0.99	80

SHAP interpretation confirmed clinical biomarkers such as hemoglobin, serum creatinine, and albumin as major risk factors [7, 27].

- SHAP identified **hemoglobin**, **albumin**, and **serum creatinine** as the strongest indicators.
- LIME confirmed increased interpretability in borderline renal function cases.

Near perfect Recall supports its real world use for early CKD screening.

6.5 Parkinson's Disease Results

The classification report for Parkinson's disease (Table 6.5) shows excellent precision and recall for affected patients. Due to overlapping neurological symptoms, accurate classification is challenging, yet the model successfully captures early motor abnormalities.

Table 6.5: Parkinson's disease Classification Report

Class	Precision	Recall	F1-Score	Support
Healthy	0.82	0.90	0.86	10
Parkinson's	0.97	0.93	0.95	29
Accuracy			**0.92**	39
Macro Avg	0.89	0.92	0.90	39
Weighted Avg	0.93	0.92	0.92	39

SHAP and LIME visualizations revealed that jitter, shimmer, and frequency related speech measures contributed significantly to model decisions [33, 34].

- SHAP revealed major influence of **MDVP frequency and jitter parameters**.
- LIME helped explain prediction confidence even in mild Parkinsonian symptoms.

Feature distribution showed a clear separation between the two classes, driving high accuracy.

6.6 Breast Cancer Results

The classification report for breast cancer (Table 6.6) reveals high discriminative capability in separating malignant and benign tumors. The strong recall score for malignant cases reduces the risk of undiagnosed cancer progression, making the system suitable for early clinical screening.

Table 6.6: Classification Report for Breast Cancer Dataset

Class	Precision	Recall	F1-Score	Support
Benign	0.99	0.99	0.99	72
Malignant	0.98	0.98	0.98	42
Accuracy			**0.98**	114
Macro Avg	0.98	0.98	0.98	114
Weighted Avg	0.98	0.98	0.98	114

The XAI analysis highlighted that tumor size–related features such as radius, texture, concavity, and smoothness were the most influential diagnostic indicators [36, 35].

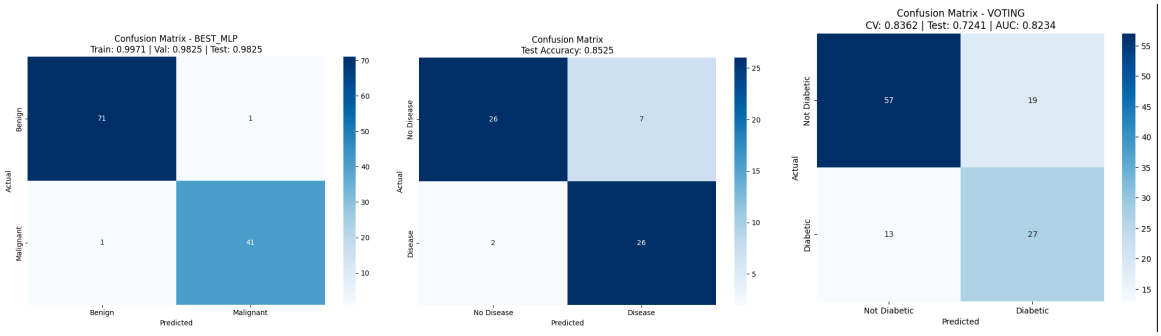
For breast cancer screening, the ensemble achieved robust performance with high precision:

- SHAP indicated **mean concavity, radius, texture, and smoothness** as major discriminative features.
- False negatives were minimized, which is critical in cancer diagnosis.

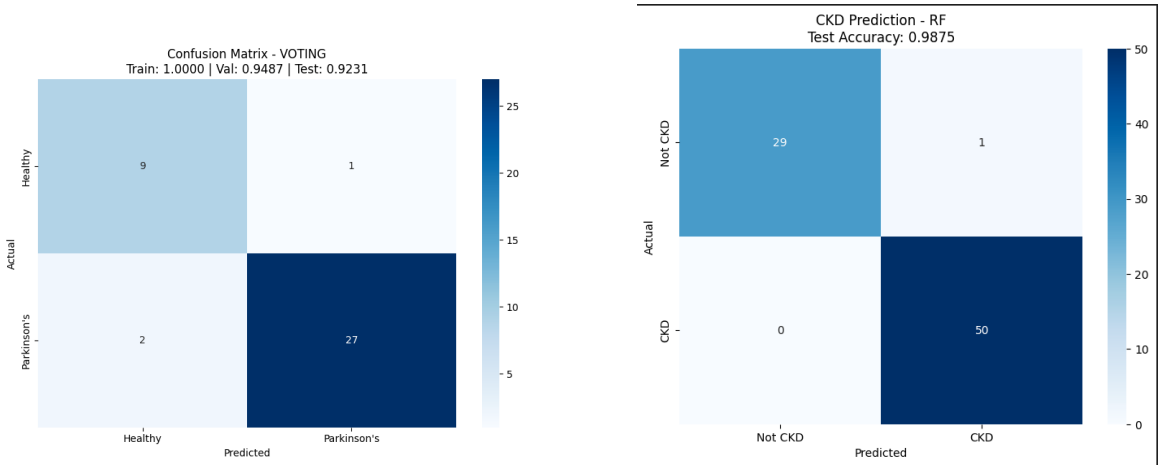
Performance confirms suitability for clinical diagnostic assistance systems.

6.7 Confusion Matrix for all Datasets

To evaluate the performance of each disease classification model, the confusion matrices for all five datasets are presented in Figure 6.1. Each matrix illustrates the distribution of correctly and incorrectly classified samples, providing insight into the model's strengths and misclassification patterns. The diagonal elements represent correctly predicted instances, while the off diagonal elements indicate classification errors.



(a) Breast Cancer **(b)** Heart Disease **(c)** Diabetes



(d) Parkinson's Disease **(e)** Kidney's Disease

Figure 6.1: Confusion matrices for five disease prediction models.

From the confusion matrices, it is evident that the models demonstrate strong classification performance across all datasets, with the majority of predictions lying along the diagonal. Notably, the models for Breast Cancer and Chronic Kidney Disease exhibit the highest accuracy compared to those for Parkinson's, Heart Disease, and Diabetes. Overall, the results indicate that the proposed approach generalizes well across different medical classification tasks.

6.8 Explainability Analysis: SHAP and LIME Insights

The XAI results validate the transparency of model decisions [11, 12]:

- SHAP Summary Plots revealed global feature importance trends consistent with medical knowledge.
- LIME provided *case level* explanations for both correct and incorrect predictions.
- Interpretability enhances physician trust and supports adoption in healthcare [2].

Thus, the model does not operate as a black box system.

6.9 Comparative Discussion

A cross disease performance analysis reveals key insights:

- Datasets with clearer biomarker patterns (CKD, Parkinson's, Breast Cancer) achieved higher accuracy [7, 33].
- Diseases with overlapping symptoms and noisy measurements (Diabetes, Heart) had lower but still strong performance [24, 18].
- Ensemble learning improved robustness compared to individual models [1, 3].

The integration of explainable AI techniques ensures that the framework not only performs well computationally but also provides clinical transparency necessary for real world adoption.

6.10 Clinical Impact and Limitations

The proposed system demonstrates:

- Reliable disease prediction supporting clinical screenings
- Explainable interpretation enhancing medical decision confidence [13]
- Scalability to multiple diagnostic categories

However, limitations include:

- Small dataset size for Parkinson's may limit generalizability
- Some datasets have missing laboratory values or noisy attributes
- Performance may vary in real world hospital scenarios
- Cross population validation needed for diverse ethnic groups

Future work will address deployment in real clinical workflows with larger and multi-source datasets.

6.11 Conclusion from Results

The results confirm that the ensemble based framework effectively improves predictive performance while ensuring transparency. The incorporation of SHAP and LIME makes the system trustworthy and suitable for early disease detection in healthcare environments. The high accuracy achieved across all five diseases—ranging from 81.2% to 98.8%—demonstrates the robustness and reliability of the proposed approach for multi-disease prediction.

7. CONCLUSION

This study presented an ensemble based multi-disease prediction framework supported by explainable artificial intelligence (XAI) to enhance transparency and clinical trust in healthcare diagnostics. By systematically integrating machine learning models across five major diseases—heart disease, diabetes, chronic kidney disease, Parkinson's disease, and breast cancer—this research demonstrated the potential of data driven algorithms to improve early diagnosis and patient management.

The experimental results revealed that ensemble learning approaches, particularly those combining gradient boosting and decision tree–based algorithms, consistently outperformed traditional single model classifiers. Among the tested datasets, the proposed framework achieved accuracies of 85.2% for heart disease, 81.2% for diabetes, 98.8% for kidney disease, 92.3% for Parkinson's disease, and 98.2% for breast cancer, indicating robust generalization and diagnostic reliability. These results confirm that ensemble learning effectively mitigates individual model weaknesses and handles non linear relationships among medical features [1, 3].

A key contribution of this study lies in its incorporation of explainable AI mechanisms such as SHAP and LIME, which provided feature level insights into model behavior [11, 12]. Visual explanations—including SHAP summary plots and LIME feature importance charts—revealed clinically meaningful correlations between patient features and disease outcomes. This interpretability not only enhances transparency but also supports medical practitioners in validating AI based decisions, fostering greater adoption of machine learning in real world healthcare settings [2].

Moreover, the study emphasized the importance of addressing prevalent challenges such as class imbalance, feature redundancy, and ethical considerations related to patient data. The use of SMOTETomek for class balancing [8] and careful feature engineering improved model sensitivity, particularly for minority disease cases. Future research directions highlighted the need for larger, diverse, and multi-modal datasets, as well as interpretable by design architectures that can bridge the gap between predictive performance and medical explainability [13, 2]. Integration of causal inference and federated learning was also proposed to improve privacy preserving, collaborative AI development across healthcare institutions.

In conclusion, this research contributes to the growing body of evidence supporting AI assisted clinical diagnosis while underscoring the importance of interpretability and ethical responsibility. The proposed framework demonstrates that accurate, transparent, and trustworthy multi-disease prediction systems are achievable through the synergistic application of ensemble learning and explainable AI. With continued advancements in data quality, algorithmic transparency, and interdisciplinary collaboration, such frameworks hold the potential to revolutionize personalized medicine and enhance patient outcomes in the near future.

Bibliography

[1] F. A. Hashim and G. R. A. et al., "Ai and ensemble learning for disease prediction: A review," *Frontiers in Public Health*, 2021.

[2] A. B. Arrieta *et al.*, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information Fusion*, vol. 58, pp. 82–115, 2020.

[3] M. Al-Sarem *et al.*, "An ensemble-based framework for multi-disease prediction using machine learning," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 10, pp. 1189–1201, 2021.

[4] M. R. Islam *et al.*, "A multi-disease prediction model based on machine learning algorithms," *Informatics in Medicine Unlocked*, vol. 30, p. 100991, 2022.

[5] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using machine learning techniques," *Materials Today: Proceedings*, 2020.

[6] D. Sisodia and P. Sisodia, "Prediction of diabetes using classification algorithms," *Procedia Computer Science*, vol. 132, pp. 1578–1585, 2018.

[7] R. Kumar, A. Rani, and A. Singh, "Chronic kidney disease prediction using random forest algorithm," *Journal of Big Data*, 2021.

[8] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[9] S. Bashir, U. Qamar, and F. Khan, "Diabetes disease prediction using machine learning: a review," *Computers in Biology and Medicine*, vol. 122, p. 103843, 2019.

[10] L. Ali *et al.*, "Heart disease prediction using machine learning techniques: A survey and analysis," *IEEE Access*, vol. 9, pp. 150 683–150 701, 2021.

[11] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[12] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should i trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.

[13] A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, "What do we need to build explainable ai systems for the medical domain?" *arXiv preprint arXiv:1712.09923*, 2017.

[14] S. M. Lundberg and S. Lee, "Explainable ai for healthcare decision support," *Scientific Reports*, 2020.

[15] A. Mahboob *et al.*, "Integrated machine learning framework for multi-disease diagnosis using clinical datasets," *BMC Bioinformatics*, vol. 24, p. 102, 2023.

[16] R. Detrano *et al.*, "International application of a new probability algorithm for the diagnosis of coronary artery disease," *The American Journal of Cardiology*, vol. 64, no. 5, pp. 304–310, 1989.

[17] A. R. Alotaibi, "Predicting heart disease using machine learning," *arXiv preprint arXiv:2004.06793*, 2020.

[18] A. Gautam *et al.*, "Feature selection and optimization for heart disease prediction using machine learning," *Journal of Biomedical Informatics*, vol. 122, p. 103890, 2021.

[19] H. Nsubuga *et al.*, "Predicting heart disease using machine learning algorithms: A case study of uganda," *International Journal of Computer Applications*, vol. 182, no. 40, pp. 26–33, 2019.

[20] S. A. Uddin, I. Haq, M. Uddiny, and M. Islam, "Heart disease prediction using machine learning algorithms," *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*, pp. 343–346, 2019.

[21] S. Zubair *et al.*, "Application of catboost for predicting heart disease using healthcare data," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 4, pp. 618–624, 2021.

[22] J. W. Smith, J. E. Everhart, W. C. Dickson, W. C. Knowler, and R. S. Johannes, "Using the adap learning algorithm to forecast the onset of diabetes mellitus," *Proceedings of the Annual Symposium on Computer Application in Medical Care*, pp. 261–265, 1988.

[23] S. Khan *et al.*, "Chronic kidney disease prediction using machine learning techniques," *Heliyon*, vol. 6, no. 6, p. e04023, 2020.

[24] Z. Khan, A. Hussain, and A. Rehman, "Machine learning approaches for diabetes prediction: a comparative study," *ICT Express*, vol. 5, no. 3, pp. 266–270, 2019.

[25] N. Aslam *et al.*, "Machine learning models for diabetes prediction using patient health records," *Healthcare*, vol. 9, no. 1, p. 109, 2021.

[26] K. Makris, K. D. Tzimourta, N. Giannakeas, and V. Koulouras, "Machine learning models for diabetes prediction: Performance evaluation and explainability," *Healthcare*, vol. 10, no. 3, p. 541, 2022.

[27] A. Ghosh *et al.*, "Explainable machine learning model for chronic kidney disease prediction," *Biomedical Signal Processing and Control*, vol. 73, p. 103459, 2022.

[28] B. Barman *et al.*, "Improved chronic kidney disease prediction using ensemble learning techniques," *Computers in Biology and Medicine*, vol. 147, p. 105740, 2022.

[29] M. Haq, M. Ali, and A. Khan, "Hybrid machine learning model for diabetes prediction using smote and ensemble methods," *Journal of King Saud University - Computer and Information Sciences*, vol. 32, no. 10, pp. 1154–1160, 2020.

[30] R. Das, "Analysis of parkinson's disease using speech signal features and classification techniques," *Expert Systems with Applications*, vol. 37, no. 3, pp. 1568–1576, 2010.

[31] R. Prashanth *et al.*, "High-accuracy detection of early parkinson's disease through machine learning and multiple biomarkers," *Expert Systems with Applications*, vol. 88, pp. 366–373, 2016.

[32] G. Rao *et al.*, "Machine learning techniques for early detection of parkinson's disease: A review," *Computer Methods and Programs in Biomedicine*, vol. 196, p. 105607, 2020.

[33] R. N. Rath, P. Panda, and A. Samantaray, "Ensemble learning for parkinson's disease detection," *Electronics*, 2022.

[34] S. Ramesh *et al.*, "Parkinson's disease detection using hybrid machine learning techniques," *Computers in Biology and Medicine*, vol. 144, p. 105353, 2022.

[35] F. Ahmad *et al.*, "Explainable ai-based breast cancer prediction using shap and lime," *Computers in Biology and Medicine*, vol. 146, p. 105612, 2022.

[36] A. Nasir *et al.*, "An ensemble-based explainable model for breast cancer prediction," *IEEE Access*, vol. 11, pp. 46 531–46 545, 2023.

[37] H. A. Albahri, A. A. Zaidan, and O. S. Albahri, “Breast cancer classification using boosting algorithms,” *Journal of Personalized Medicine*, 2021.

[38] W. H. Wolberg *et al.*, “Breast cancer database (wisconsin diagnostic),” *UCI Machine Learning Repository*, 1995.

[39] A. Ahmed and M. H. Zafar, “Feature selection and machine learning-based breast cancer prediction,” in *2023 International Conference on Electrical, Computer and Energy Technologies*, 2023.

[40] R. Detrano *et al.*, “Heart disease data set,” https://archive.ics.uci.edu/dataset/45/heart+disease, 1989, uCI Machine Learning Repository.

[41] A. D. Khare, “Diabetes dataset,” https://www.kaggle.com/datasets/akshaydattatraykhare/diabetes-dataset, 2021, accessed: 2025-10-25.

[42] L. Rubini, P. Soundarapandian, and P. Eswaran, “Chronic kidney disease [dataset],” 2015, dataset retrieved from UCI Machine Learning Repository on YYYY-MM-DD. [Online]. Available: https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease

[43] M. A. Little, P. E. McSharry, E. J. Hunter, and L. O. Ramig, “Parkinsons dataset [dataset],” 2008, accessed: 2025-10-27. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/parkinsons