CS550/DSL501: Machine Learning (2024–25–M)
Project Report Phase 3

Team name: MT67
1) R. C. Yajour Kichenamourty - M24MT006
2) Sagar Sudhir Pathak - M24MT007

Explainable AI Based Disease Prediction System

1 Problem Statement

People nowadays suffer from a variety of diseases because of environmental factors and their lifestyle choices. As a result, disease prediction at an earlier stage becomes a critical task. In the current healthcare landscape, disease prediction and diagnosis often rely on traditional methods, which may not always yield accurate or timely results, especially for complex diseases. With the growing availability of healthcare data from Electronic Health Records (EHRs), medical imaging, and wearable devices, the application of Machine Learning (ML) in healthcare has gained prominence. However, while ML models have shown great potential in disease prediction, they often operate as "black boxes", making it difficult for medical professionals and patients to trust and understand the predictions. This lack of transparency limits the adoption of ML system in critical healthcare environments, where explainability and trust are significant.

The problem lies in developing an ML-based healthcare application that can not only predict diseases with high accuracy but also provide clear explanations for its predictions. This will allow healthcare professionals to make informed decisions and gain trust in AI-powered medical tools.

2 Motivation

The motivation behind this project is to improve healthcare outcomes by leveraging Machine Learning techniques to enable earlier and more accurate disease diagnosis. Many diseases, such as diabetes, heart disease and Alzheimer's disease have better prognoses when detected early. Machine learning models have shown great potential in analyzing large datasets and identifying patterns that are not immediately visible to human practitioners. However, their adoption in real-world clinical practice has been slow due to the lack of interpretability in decision-making processes.

Explainable AI (XAI) offers the opportunity to bridge this gap by ensuring that the predictions made by the ML models are not only accurate but also understandable and transparent to healthcare providers. By combining ML with explainable AI, the project aims to increase trust in AI-based healthcare systems, making them more practical for real-world clinical use, improving patient outcomes, and assisting doctors in making better decisions.

3 Objectives

The **primary objective** of the proposed system is to bridge the gap between medical expertise and the general public by providing a user-friendly and transparent platform for self-diagnosis and medication guidance. The system integrates Explainable AI (XAI) to enhance user understanding of the reasoning behind disease predictions and medication recommendations.

The secondary objectives of this project are:

To improve accuracy and reliability: Achieve high predictive accuracy while maintaining the interpretability of the model's decisions.

To evaluate the system on real-world datasets: The model will be tested and validated on public healthcare datasets to ensure its effectiveness in disease prediction.

User-friendly interface: Design a user-friendly interface for medical professionals and general public to interact with the model's predictions and explanations.

4 Previous Work and Novelty

A large body of research highlights the effectiveness of machine learning models in disease prediction across multiple healthcare areas. Algorithms such as decision trees, neural networks, support vector machines (SVMs), and deep learning have been applied to predict diseases such as cancer, diabetes, and cardiovascular diseases using patient data from EHRs, genetic data, and imaging data. Recent advancements in Explainable AI (XAI) have led to methods like LIME (Local Interpretable Model-Agnostic Explanations), SHAP (SHapley Additive exPlanations), and Grad-CAM (Gradient-weighted Class Activation Mapping) being used to interpret complex models and provide human-interpretable insights.

However, many machine learning systems used in healthcare still face challenges related to model transparency and user trust, particularly in high-stakes environments like healthcare. Explainable AI has emerged as a key solution to address this issue, offering methods that can explain the predictions made by complex models in ways that are accessible to non-experts. Studies have shown that implementing XAI increases trust in AI systems and enhances their adoption in critical sectors like healthcare.

Moreover, research suggests that combining predictive analytics with interpretability can improve the decision-making process for healthcare providers, leading to better patient outcomes.

Sl. No	Paper Name and Year	Journal/ Paper Details	Paper Link
1	Explainable Artificial Intelligence Based Framework for Non Communicable Diseases Prediction (2023)	This Journal proposes a Deep Shapley Additive Explanations (DeepSHAP) based deep neural network framework equipped with a feature selection technique for NCDs prediction and explanation among the population in the United States.	Link1
2	Integration of Explainable Artificial Intelligence (XAI) in the Development of Disease Prediction and Medicine Recommendation System (2024)	The paper proposes to bridge the gap between medical expertise and the general public by providing a user-friendly and transparent platform for self-diagnosis and medication guidance. The system integrates Explainable AI (XAI) to enhance user understanding of the reasoning behind disease predictions and medication recommendations.	Link2
3	Comparison and Analysis of Various Machine Learning Algorithms for Disease Prediction (2023)	This paper presents a system wherein data can help the medical filed experts to detect the fatal diseases early and thus the survival rate of the victims will be increased. The Disease Prediction system is based on predicting the name of the disease of the user examining the symptoms that user gives as an input to the system. The system also predicts the risk due to the occurrence of the disease compared to the general disease which is lower or higher.	Link3

5 Methodology

The Explainable AI-Based Disease Prediction System is designed to accurately predict diseases while ensuring the interpretability of its decisions. The process begins with data pre-processing, where raw data is cleaned by handling missing values, removing outliers, and normalizing the features. The dataset is then divided into a training set (80%) and a testing set (20%) for model evaluation. To optimize performance, the feature selection technique Elastic Net which combines Lasso (L1 regularization), and Ridge (L2 regularization) is applied to identify the most relevant features. The system incorporates n-fold cross-validation to validate the model using different splits of the dataset, ensuring robust evaluation. The machine learning models Deep Neural Networks (DNN), Random Forest, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN) are tested and hyperparameters are fine-tuned to maximize accuracy. For interpretability, the system employs five Explainable AI (XAI) techniques viz., Decision Trees, SHAP (SHapley Additive Explanations), LIME (Local Interpretable Model-Agnostic Explanations), and the Explainable Boosting Algorithm to provide clear insights into model predictions. The system allows users to input their data, generates a disease prediction, and offers an explanation for the outcome, ensuring trust and transparency. The output is a front-end dashboard that is developed for healthcare professionals and general public to interact with the system, view predictions, and understand the model's reasoning through visual and textual explanations.

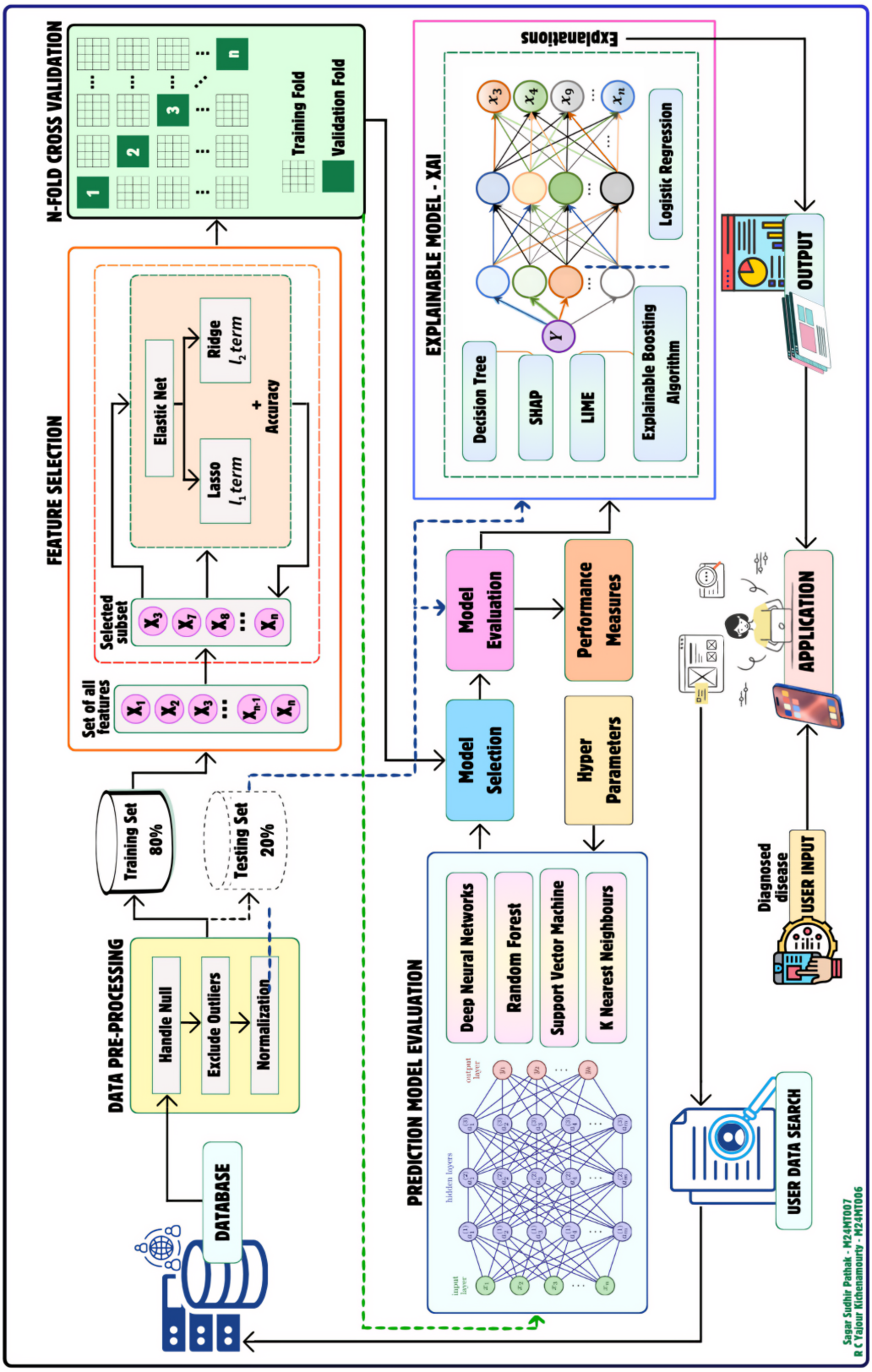

Figure 1: ARCHITECTURE

5.1 XAI MODELS USED:

5.1.1 SHAP (SHapley Additive exPlanations)

SHAP is a unified framework for interpreting machine learning predictions based on Shapley values from cooperative game theory. It quantifies the contribution of each feature to the prediction, making it highly interpretable.

How SHAP works in XAI:

- SHAP assigns a unique value to each feature for a particular prediction based on its marginal contribution when combined with other features.
- It provides both local explanations (for a single prediction) and global explanations (aggregating feature impacts).

Mathematical Representation:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)]$$

where:

- N: Set of all features.
- S: Subset of features excluding i.
- $f(S)$: Model output with only features in S.

Use in the System: SHAP explains why a disease prediction was made by evaluating the contribution of each input feature, such as age, symptoms, or test results.

5.1.2 LIME (Local Interpretable Model-agnostic Explanations)

LIME explains individual predictions by approximating the machine learning model locally with a simpler, interpretable model, such as a linear regression.

How LIME works in XAI:

- LIME perturbs the input data, generates synthetic samples, and evaluates the black-box model's predictions on them.
- It fits a simpler model (e.g., linear regression) on the perturbed samples to explain the prediction for the original input.

Mathematical Representation:

$$\min \left(L(f, g, \pi_x) + \Omega(g) \right)$$

where:

- f: Original black-box model.
- g: Simple interpretable model.
- π_x: Proximity measure (weighting function for local fidelity).
- $\Omega(g)$: Complexity of the interpretable model.

Use in the System: LIME provides a local explanation for why a specific disease prediction was made by identifying which features most influenced the outcome.

5.1.3 Decision Tree

A decision tree is an interpretable machine learning model that uses a tree-like structure of conditions (nodes) to split the data and predict outcomes.

How Decision Trees work in XAI:

- Decision trees inherently provide explanations through their structure. The path from the root to a leaf (prediction) shows how input features influence the decision.
- They are used as a baseline interpretable model in XAI.

Mathematical Representation: At each node, the data is split by minimizing the impurity measure, such as Gini impurity:

$$G = 1 - \sum_{i=1}^{C} p_i^2$$

or entropy:

$$H = -\sum_{i=1}^{C} p_i \log(p_i)$$

where p_i is the proportion of class i in the node.

Use in the System: Decision Trees explain predictions by showing the conditions (e.g., test values) that led to a specific disease classification.

5.1.4 Logistic Regression

Logistic regression is a statistical model that predicts the probability of a binary outcome based on input features. It is inherently interpretable as it shows the direct impact of each feature on the output.

How Logistic Regression works in XAI:

- Coefficients (β_i) of logistic regression directly explain the impact of each feature on the log-odds of the output.
- It provides both global (overall) and local (individual) explanations.

Mathematical Representation:

$$P(y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^{n} \beta_i x_i)}}$$

where:

- $P(y = 1|X)$: Probability of the positive class.
- β_i: Coefficients of features x_i.

Use in the System: Logistic Regression helps identify key features (e.g., specific symptoms) that significantly influence disease diagnosis probabilities.

5.1.5 Explainable Boosting Algorithm (EBA)

EBA is an interpretable machine learning algorithm combining boosting techniques and generalized additive models (GAMs). It models non-linear relationships while maintaining interpretability.

How EBA works in XAI:

- EBA represents predictions as the sum of feature functions, each capturing the contribution of a single feature or interaction.
- It uses boosting to optimize the accuracy of these functions iteratively.

Mathematical Representation:

$$f(x) = \sum_{i=1}^{n} g_i(x_i)$$

where $g_i(x_i)$ is a shape function representing the contribution of feature x_i.

Use in the System: EBA explains disease predictions by visualizing how each input feature (e.g., age, medical tests) contributes to the final output.

Summary of XAI Models in the System

- **SHAP and LIME:** Provide post-hoc explanations by showing feature importance for individual predictions.
- **Decision Tree:** Explains decisions through its interpretable structure.
- **Logistic Regression:** Highlights the impact of each feature on prediction probabilities.
- **EBA:** Visualizes complex relationships while remaining interpretable.

Together, these models form a robust XAI framework that explains predictions at both local and global levels, ensuring transparency and trust in the disease prediction system.

6 Experimental Settings and Results

7 GitHub

GitHub Link

8 Individual Contributions

8.1 R. C. Yajour Kichenamourty - M24MT006

Yajour's primary responsibility was the initial data preprocessing and feature engineering for the dataset. This involved:

- **Data Collection:** Sourced data from publicly available health datasets such as Electronic Health Records (EHRs), wearable device data, and medical imaging repositories.
- **Data Cleaning:** Managed missing values, normalized data, and handled outliers using Python libraries such as `pandas` and `numpy`.
- **Feature Engineering:** Selected key features for disease prediction by using domain knowledge and statistical methods (e.g., correlation analysis) to filter out unimportant features.
- **Model Training:** Conducted experiments with several machine learning models like Random Forest, Gradient Boosting, Support Vector Machine(SVM), K Nearest Neighbours(KNN), XGBoost and Deep Neural Networks for initial disease prediction.
- **Performance Tuning:** Performed hyperparameter tuning using Grid Search and Random Search techniques to improve model performance and selected the best model for each of the Disease - Heart, Diabetes, Alzeimer's and General.
- **Documentation:** Handled documentation and reporting for the explainability aspect of the project, detailing how the system's predictions could be trusted based on the generated explanations.

8.2 Sagar Sudhir Pathak - M24MT007

Sagar was responsible for implementing the Explainable AI (XAI) components of the system, which included:

- **Exploration of XAI Techniques:** Investigated several XAI methods, such as LIME (Local Interpretable Model-agnostic Explanations), SHAP (SHapley Additive exPlanations), Decision Trees, Logistic Regression, Explainable Boosting and integrated for the given dataset.
- **Model Integration:** Integrated SHAP values with the disease prediction model to provide feature-level explanations for predictions. This ensured that healthcare professionals could understand the influence of different factors on predictions.

- **System Architecture:** Designed the overall architecture for the healthcare prediction system, ensuring that both the prediction and explanation components worked seamlessly.
- **Model Evaluation:** Evaluated the explainability of the system by conducting tests with sample predictions and explanations, ensuring transparency and accuracy in predictions.

In addition, Sagar contributed to developing an interface for integrating machine learning models into a healthcare application using Streamlit, including assisting with testing the system in a controlled environment.

9 Tasks and Milestones Achieved

The project invloved training of 3 different datasets viz., Heart Disease, Alzeimers' and Diabetes over 5 Explainable AI models viz., SHAP, LIME, Decision Tree, Logistic Regression and Explainable Boosting Algorithm.

- **Problem Understanding and Data Collection (Week 1-2):** Successfully identified a healthcare problem where AI-based disease prediction could be useful. Collected relevant healthcare datasets for training the machine learning model.
- **Data Preprocessing and Feature Engineering (Week 3-4):** Cleaned and preprocessed the dataset. Selected relevant features based on domain knowledge and statistical techniques.
- **Model Training and Evaluation (Week 5-6):** Trained multiple machine learning models for disease prediction and evaluated them on accuracy and precision. Finalized Random Forest as the best-performing model.
- **Integration of XAI Techniques (Week 7-9):** Integrated SHAP and LIME into the model to provide explainability for predictions. Successfully generated human-readable explanations for individual predictions.
- **Deployment and Testing (Week 10-11):** Deployed the system in a test environment and validated predictions with explanations. Worked on refining the user interface to display model explanations effectively.
- **Documentation (Week 12):** Prepared the Final Project Report along with the PowerPoint Presentation and compilation of all relevant codes, project files into the GitHub Link.

The RECALL values showed Explainable Boosting Algorithm performing best for Heart Disease and Diabetes dataset while Logistic Regression was efficient for Alzeimers' dataset. The Front end application was also made using the above said XAIs for respective datasets.

10 Further Works

- **Optimization for Real-Time Predictions:** The current system is capable of providing predictions and explanations but may not be optimized for real-time performance. Further work can involve optimizing the model's inference time to make it more suitable for real-time use in clinical settings.
- **User Interface Improvements:** Currently, the user interface for displaying explanations is functional but basic. Future work will focus on improving the visual representation of SHAP values and other XAI metrics to make them more intuitive for medical professionals.
- **Generalization for Other Diseases:** While the model is focused on a specific set of diseases, further work will focus on extending the model to cover a wider range of diseases by using more diverse datasets.
- **Model Calibration:** Explore model calibration techniques to improve the trustworthiness of the predicted probabilities, ensuring that predictions are not only accurate but also well-calibrated.

THANK YOU