

Diabetes and Heart Disease Prediction Using Ensemble Technique

Hemanth Krishnaprasad Bindu
School of Computing
Dublin City University
Dublin, Ireland

Sanjay Sathish Kumar
School of Computing
Dublin City University
Dublin, Ireland

Abstract—Diabetes and heart disease are two of the most common chronic illnesses in the world and have a major negative impact on people's health and well-being. In order to reduce the consequences and complications connected to these illnesses, detection and intervention are essential. In this study, we suggest applying machine learning approaches to the construction and assessment of predictive models for diabetes and heart disease. To create precise and understandable predictive models, we specifically investigate the use of the Random Forest, Gradient Boosting, and AdaBoost algorithms in conjunction with feature selection techniques like SelectKBest, which uses an ANOVA F-test value to identify the most relevant features for prediction. Our results demonstrated that the F-test method achieved the highest AUC-ROC scores of 0.814 for diabetes and 0.813 for heart disease, highlighting its effectiveness. Additionally, age-specific evaluation revealed significant variability in model performance.

I. INTRODUCTION

Heart disease and diabetes are major global health concerns, as the rate of both has been rising over the past few years. The World Health Organization (WHO) reports that over 422 million people¹ worldwide suffer from diabetes, and cardiovascular illnesses continue to be the top cause of death internationally. In order to effectively manage these chronic illnesses and lower the rates of mortality and complications that are associated with them, detection and management are essential. Still, it's difficult to anticipate diabetes and heart disease with any degree of accuracy, especially in at-risk groups.

This research aims to address the need for accurate predictive models for diabetes and heart disease by applying machine learning algorithms to health indicator data obtained from the Diabetes Health Indicators Dataset, BRFSS 2015. The Behavioural Risk Factor Surveillance System (BRFSS) is a health-related telephone survey that is collected annually by the Centre for Disease Control (CDC). The dataset contains over 250 thousand rows and 22 columns containing Diabetes, HighBP, HighChol, CholCheck, BMI, and other values. Classifiers like Random Forest, Gradient Boosting, and AdaBoost are selected for healthcare applications because of their flexibility, capacity to manage complex data patterns, and efficiency in ensemble learning. These attributes frequently result in enhanced predictive performance and understanding.

While previous studies have focused on predicting diabetes using various models, they often overlooked evaluating model performance across different age groups. Our research aims to fill this gap by investigating whether the efficacy of various machine learning models in predicting diabetes and heart disease differs significantly across age groups.

II. LITERATURE REVIEW

Different datasets were used in different articles. [1], [2], [3], [4], [5] has used Pima Indian Dataset, [6], [7], [8] has used UCI repository, [9], [10] has used BRFSS for the prediction of diabetes. Datasets like Kaggle, ADIPS, and others have also been used by authors. The Pima Indian dataset and UCI repository had a very small number of instances. Hence the results obtained from these datasets might not be the same while using other large datasets.

The datasets in the articles had some missing values. [11], [1], [4], [12] used replacement by mean method, while [5] used polynomial regression to fill the missing inputs. [9] removed the missing values and [6] applied both replacement by mean and removal technique. [13] estimated missing values using a dynamic linear method which outperformed Newton's divided difference method. The categorical data types were encoded into numerical values using label encoding by [1], [2], [14]. One hot encoding was used by [7] to do the same. [12], [4], [7] utilised Z-Score Normalisation and [1], [9] used min-max normalisation for standardisation. SMOTE [9], [10] and Cross fold validation techniques [4], [15] were used to overcome the data imbalances that occurred in the dataset.

Chollette C. Olisah et. al [5] proposed a supervised machine learning model which used a twice-growth deep neural network model (2GDNN) for classification. The model was optimized by tuning the hyper-parameters using a grid search. Aishwarya Mujumdara, Dr. Vaidehi V [11] imposed a pipeline model for the classification to improve the accuracy of the model compared to other models. Saloni Kumari et. al [1] published an article that used an ensemble soft voting classifier of three machine learning algorithms that are Linear regression, random forest, and naive bayes. It was done using the majority voting technique. Nazin Ahmed et. al [2] created a web application using various machine learning algorithms the model was analyzed using two datasets and the one with the highest accuracy for the two datasets was selected and used

¹<https://www.who.int/health-topics/diabetes>

in the web application. Victor Chang et. al [9] and KM Jyothi Rani [14] published articles that used five different machine learning classifiers to classify diabetes data. Zidian Xie et. al [10] also published a model similar to [9] but additionally focused on the risk factors associated with type-2 diabetes.

A. Sumathi, S. Meganathan [16] used diabetes detection tree ensemble clustering (DDTEC) for classification the model predicts type-1, type-2, and gestational diabetes. Nikos Fazakis et. al [12] has developed an ensemble model using weighted voting. Logistic regression and Random Forest were the two models that were used for the ensemble. Additionally, the authors integrated the model into an IOT-enabled framework. MD. Kamrul Hasan et. al [4] also used ensemble models which used a combination of different algorithms where AdaBoost along with XGBoost gave a better performance for the prediction. whereas, Daliya V.K, T.K. Ramesh [15] in this article proposed an optimized stacking ensemble model for regression and classification. The stacking ensemble model takes the prediction of different machine learning models and feeds as the input to the meta-model which is used for classification. Shirina Samreen [7] has developed a similar stacking ensemble model but additionally used ANOVA and Crow search algorithms for the feature selection and singular value decomposition for feature extraction. Hakim El Massari et. al [3] developed a model using the ontology method. Ontology is an open-source platform for building an intelligence system created by protege. The authors also compared the model with other algorithms and showed a better performance. Veena Vijayan V and Anjali C [6] Used AdaBoost with base classifier as support vector, Naive bayes, Decision Tree, and Decision Stump. The accuracy was boosted using the model. USAMA AHMED [8] fused machine learning approach where output from two models which are support vector machine and artificial neural network was fused to obtain higher accuracy. N. Yuvaraj, K.R SriPreethaal [17] have used information gain feature selection and have used various machine learning algorithms to predict diabetes which is integrated into the Hadoop cluster using R.J. Omana, M. Moorthi [13] used enhanced Markov chain model for the prediction of diabetes. The proposed models were evaluated by the authors using various evaluation metrics and received different results based on the models and dataset. [5] developed an ontology model that was able to achieve the highest accuracy of 97.248% Using the Pima Indian dataset. [7] produced a model using the UCI machine learning repository dataset and produced the highest accuracy of 98.46% using the stacked ensemble technique. BRFSS dataset has the highest amount of instances among the other datasets and [10], [9] was able to achieve an accuracy of 80% to 82%.

In the articles reviewed above the research focused on predicting diabetes using various models but failed to show the evaluation of models on different age groups. Our research is to find whether the performance of different machine learning models for predicting diabetes and heart disease varies significantly across different age groups. Additionally, we also aim to test our proposed model for the prediction of other

variables like heart disease, blood pressure, and cholesterol levels.

III. PROPOSED RESEARCH

A. Dataset

The dataset used in this study is a kaggle dataset from the Centers for Disease Control and Prevention's (CDC) Behavioral Risk Factor Surveillance System (BRFSS) survey conducted in 2015 [18]. This dataset comprises responses from 253,680 survey participants, each characterized by a set of health indicators and socio-demographic attributes. The target variable, 'Diabetes binary,' indicates the presence or absence of diabetes, with three classes: 0 (No diabetes or only during pregnancy), and 1 (Diabetes). There are 21 feature variables encompassing aspects such as high blood pressure, cholesterol levels, physical activity, dietary habits, and socio-demographic factors like age, sex, education, and income.

Feature	Data type	F-Value	Chi-Squared
Diabetes_binary	Binary (0,1)	—	—
HighBP	Binary (0,1)	18870.366	4491.800
HighChol	Binary (0,1)	10600.351	2804.501
CholCheck	Binary (0,1)	1068.397	145.025
BMI	Continuous	12516.719	10 312.876
Smoker	Binary (0,1)	940.878	122.939
Stroke	Binary (0,1)	2872.607	927.647
HeartDiseaseorAttack	Binary (0,1)	3310.962	2695.401
PhysActivity	Binary (0,1)	3590.290	1163.509
Fruits	Binary (0,1)	422.555	77.614
Veggies	Binary (0,1)	814.826	106.510
HvyAlcoholConsump	Binary (0,1)	828.525	104.701
AnyHealthcare	Binary (0,1)	67.047	2.942
NoDocbcCost	Binary (0,1)	250.886	37.458
GenHlth	Ordinal (1-5)	23924.565	5133.530
MentHlth	Continuous	1224.701	9491.555
PhysHlth	Continuous	7672.268	55 929.053
DiffWalk	Binary (0,1)	12699.342	3926.865
Sex	Binary (0,1)	250.842	1.349
Age	Ordinal (1-14)	8246.866	5204.870
Education	Ordinal (1-6)	3991.111	1569.601
Income	Ordinal (1-8)	7004.371	2956.984

TABLE I
FEATURES USED IN MODELS, THEIR DATA TYPE, AND VALUES FROM
VARIOUS FEATURE SELECTION METHODS

B. Data Preprocessing

In the dataset utilized for this study, there exists a significant class imbalance issue, with large number of instances belonging to the non-diabetic class compared to the diabetic class. Specifically, there are 218,334 instances of non-diabetic individuals and only 35,346 instances of diabetic individuals. Such a severe class imbalance can potentially lead to biased model performance, where the model may tend to favor the majority class and perform poorly on the minority class.

To overcome this impact of class imbalance on model training and performance, Synthetic Minority Over-sampling Technique (SMOTE) has been used as a preprocessing step on the training data. SMOTE technique addresses the class

imbalance by generating synthetic samples for the minority class. By artificially increasing the representation of the minority class in the dataset, SMOTE helps to balance the distribution of classes, thereby reducing the risk of model bias towards the majority class.

SMOTE is applied only to the training data after splitting it into training and testing sets. This ensures that synthetic instances are only generated based on the minority class instances in the training set, preventing any data leakage in the testing set. By generating synthetic instances of diabetic individuals, SMOTE effectively boosts the representation of the minority class in the training data, allowing an effective learning process for the machine learning models.

The application of SMOTE allows the models to learn from a more balanced dataset, where both classes are adequately represented. As a result, the trained models are expected to exhibit improved performance in accurately predicting instances from both the majority and minority classes. By addressing class imbalance with SMOTE during the data preprocessing stage, we aim to enhance the reliability and effectiveness of our predictive modeling approach for diabetes detection.

The comparison of classification results with and without SMOTE clearly demonstrates the importance of addressing class imbalance. When SMOTE is applied, we observe improvements in various performance metrics such as precision, recall, and F1 score for both the majority and minority classes in the classification reports. Additionally, the AUC-ROC scores also indicate better overall performance of the models with SMOTE, further supporting the effectiveness of this technique in handling class imbalance. Therefore, based on the results, it is evident that SMOTE plays a crucial role in improving the predictive performance of the models for diabetes detection by addressing class imbalance.

C. Feature Selection

1) *SHAP*: In this study, we have utilized the SHAP (SHapley Additive exPlanations) feature selection technique to determine which features from the 2015 BRFSS dataset are most important in predicting diabetes and heart disease. The dataset is first preprocessed by removing features that are directly connected to heart disease, diabetes, and age. Using the Synthetic Minority Over-sampling Technique (SMOTE), the training data's class imbalance is addressed. The contribution of each feature to the model's predictions is measured by computing SHAP values.

The average SHAP value for each feature is obtained by combining the SHAP values for all samples and classes. The features are ranked according to their importance due to this aggregation. The ranking is shown visually, with the top features highlighted in a SHAP summary plot. Utilizing a vote classifier setup, the top 10 features that were determined to be the most significant are then utilized to train ensemble classi-

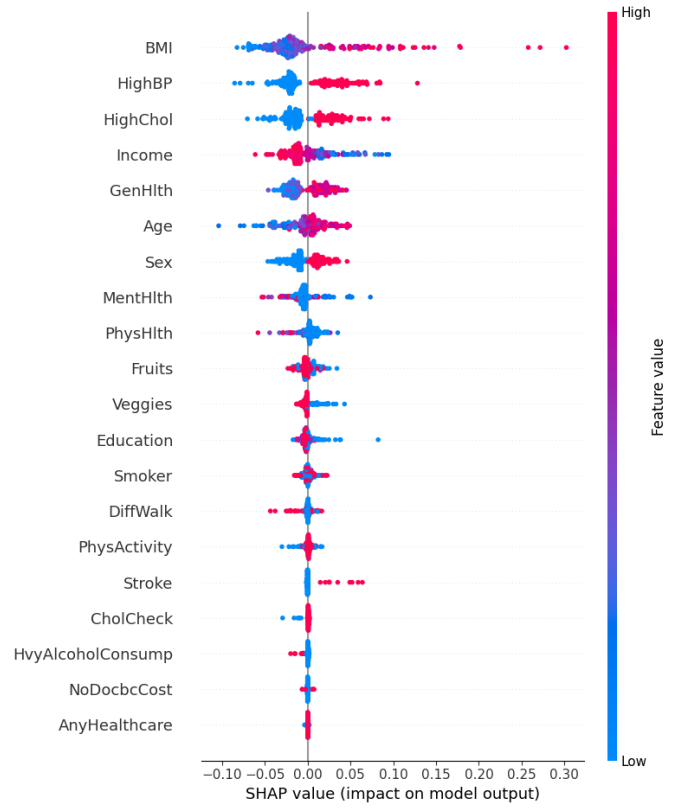


Fig. 1. SHAP Beeswarm Plot of Diabetes

fiers, such as Random Forest, Gradient Boosting, AdaBoost, and XGBoost.

The performance of these models is evaluated on test data, with the AUC-ROC score as the metric, achieving robust predictive performance for both diabetes and heart disease. This SHAP-based feature selection process enhances model interpretability and maintains high predictive accuracy, demonstrating the effectiveness of using SHAP values for feature selection in complex predictive modeling tasks.

2) *F-Test*: In this study, we employed the F-test to determine the significance of each feature in predicting diabetes and heart disease. The F-test evaluates the relationship between each feature and the target variable, assigning a score that shows their predictive power. Features with higher scores indicate stronger associations with the target, making them more valuable for prediction. We select features using F-test scores only from the training set to make sure there's no mixing of data. Then, we use these chosen features to test the model on new data.

By selecting features based on their F-test scores, we effectively streamline the dataset, retaining only the most informative attributes. This process not only improves the efficiency of the predictive model but also enhances its interpretability by focusing on the most influential factors. Consequently, the selected features contribute to more

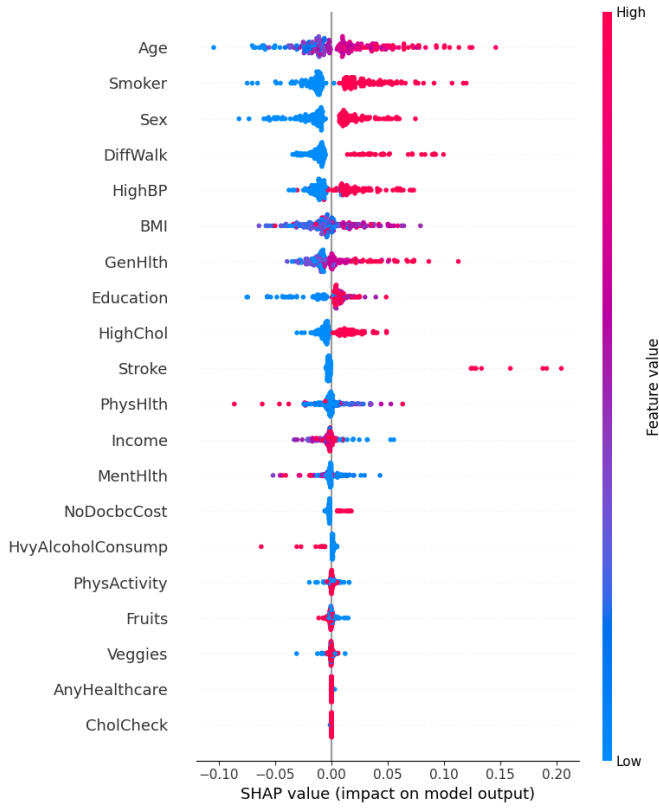


Fig. 2. SHAP Beeswarm Plot of Heart Disease

accurate predictions of health conditions, facilitating better decision-making in healthcare contexts.

3) *Correlation*: To extract the most relevant features from our dataset, we use a technique called Correlation Feature Selection (CFS). It functions by determining how strongly each feature is related to the presence of diabetes or heart disease. In essence, CFS helps in determining the most important variables for accurately predicting various medical disorders. When using Correlation Feature Selection (CFS), we pick features only from the training set to avoid mixing data. After that, we apply these selected features to test how well the model performs on new data. We improve the accuracy and comprehensibility of our prediction models by utilizing CFS. By concentrating on the important factors, this strategy simplifies our inquiry and guarantees that our study produces solid and insightful results.

IV. COMPARATIVE ANALYSIS OF FEATURE SELECTION TECHNIQUES

Three feature selection methods, namely SHAP values, F-test, and correlation analysis, were employed to identify significant features for predicting diabetes and heart disease.

Initially, utilizing SHAP values, the models achieved AUC-ROC scores of 0.802 for diabetes prediction and 0.795 for

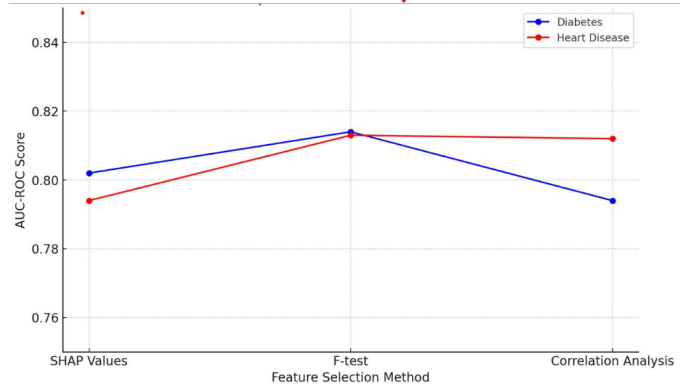


Fig. 3. AUC-ROC Comparison for Feature Selection Method

heart disease prediction. Subsequently, employing the F-test for feature selection, the AUC-ROC scores improved marginally to 0.814 for diabetes and 0.813 for heart disease. Notably, across different age groups, variations in model performance were observed, with some age categories exhibiting higher AUC-ROC scores. Specifically, for diabetes prediction, the AUC-ROC ranged from 0.529 for the youngest age group to 0.774 for the oldest age group, while for heart disease prediction, it ranged from 0.465 to 0.835, respectively.

Conversely, the correlation-based feature selection method resulted in models with AUC-ROC scores of 0.794 for diabetes and 0.793 for heart disease. Though slightly lower than the F-test results, the correlation-based approach still demonstrated competitive performance. These findings highlight the importance of selecting appropriate feature selection techniques tailored to the dataset and modeling objectives. Moreover, the disparity in model performance across age groups underscores the need for age-specific considerations in predictive modeling for chronic diseases like diabetes and heart disease.

The comparative analysis of feature selection techniques, including SHAP values, F-test, and correlation analysis, demonstrates the effectiveness of the F-test in identifying significant features for predicting diabetes and heart disease. The results indicate that models utilizing the F-test for feature selection achieved higher AUC-ROC scores for both diabetes (0.814) and heart disease (0.813) predictions compared to SHAP values and correlation analysis. Furthermore, the F-test exhibited consistency across different age groups. This consistency indicates that features selected through the F-test maintain their predictive power regardless of age, thereby enhancing the reliability and generalizability of the models.

Furthermore, the F-test offers a reliable feature selection technique that ensures that the features chosen have a visible impact on prediction accuracy. The F-test simplifies the dataset and enhances model interpretability by highlighting the most informative attributes. This makes it simpler to understand the underlying relationships in the data. With these benefits in mind, the F-test is the feature selection technique of choice because of its superior performance metrics,

statistical rigor, and capacity to improve the effectiveness and interpretability of predictive models. Because of this, the F-test is a crucial tool for creating precise and trustworthy models that forecast chronic illnesses like diabetes and heart disease.

General Health (GenHlth) stands out as the most influential feature with the highest F-Value, indicating its strong association with the outcome variable. High Blood Pressure (HighBP) follows closely, demonstrating a critical role in predicting the target variable. Body Mass Index (BMI) also shows substantial relevance, with a high F-Value reflecting its importance in the model.

Difficulty Walking (DiffWalk) and Age are among the top features, showing high F-Values and significant predictive power. Physical Health (PhysHlth) further contributes with an impressive F-Value, underscoring its impact. Income and Education, while lower on the list, still present notable F-Values, indicating their predictive strength.

Physical Activity (PhysActivity) also show significant F-Values, highlighting their importance. These features, identified through their high F-Values, collectively provide a robust framework for understanding the underlying patterns and predicting outcomes effectively in the dataset.

V. MODEL SELECTION: PREDICTIVE MODELING FOR DIABETES AND HEART DISEASE

The selection of appropriate machine learning algorithms is crucial for developing an effective predictive model for diabetes and heart disease prediction. We have chosen AdaBoost, Random Forest, and Gradient Boosting as our primary algorithms due to their flexibility, robustness, and proven track record in addressing classification challenges effectively.

AdaBoost is a powerful boosting algorithm that continuously trains weak learners to gradually improve overall classification accuracy. It accomplishes this by adjusting the weights of misclassified instances in subsequent iterations, thereby focusing on cases that are more challenging to classify correctly. AdaBoost is particularly very effective at handling complex datasets and improving prediction accuracy by iteratively refining the model based on previous mistakes.

Random Forest is an ensemble learning technique that combines multiple decision trees to produce more accurate and robust predictions. Random Forest effectively captures the complex patterns present in the data and provides reliable predictions.

Gradient Boosting is another boosting algorithm that constructs a sequence of models iteratively, with each subsequent model focusing on the weaknesses of the previous one. By gradually reducing the errors made by earlier models, Gradient Boosting iteratively improves prediction accuracy and generalization performance. This technique help

in capturing the subtle relations between the target variable (diabetes) and the other features.

A. Ensemble Methods

We also make use of ensemble methods to further enhance predictive performance. Ensemble methods combine the predictions of multiple models to produce a final prediction that often outperforms any individual model. We employ majority voting, a common ensemble technique, to combine the predictions of AdaBoost, Random Forest, and Gradient Boosting models. In majority voting, each model's prediction carries equal weight, and the final prediction is determined by the most frequent prediction among the ensemble of models.

B. Transfer Learning

In this study, transfer learning was employed to improve heart disease prediction by utilizing a neural network model initially trained for diabetes prediction. Both the diabetes and heart disease models used the same dataset, which included health indicators relevant to both conditions. The process began with training an Artificial Neural Network (ANN) on the dataset to predict the presence of diabetes. The neural network was structured with three layers:

- **Input Layer:** This layer consisted of 128 neurons with ReLU (Rectified Linear Unit) activation function, which helps in learning complex patterns by applying a non-linear transformation.
- **Hidden Layer:** This layer included 64 neurons, also with ReLU activation, further transforming the input data and extracting features that are more abstract and relevant for the prediction task.
- **Output Layer:** The final layer had a single neuron with a sigmoid activation function, suitable for binary classification tasks, providing a probability score indicating the likelihood of diabetes presence.

Additionally, Dropout layers with a dropout rate of 0.5 were included after each dense layer to prevent overfitting by randomly setting a fraction of input units to zero during training.

Once the diabetes model was trained, the parameters from this model were transferred to a new model aimed at heart disease prediction.

By transferring these parameters, the heart disease model benefited from the initial training on diabetes, allowing it to start from an informed state. This transfer of knowledge leveraged the overlapping health indicators and underlying patterns relevant to both diabetes and heart disease. The heart disease model, initialized with these pre-trained weights, then continued training on the same dataset, but this time focusing on heart disease prediction. This approach effectively allowed the heart disease model to build upon the foundation set by the diabetes model, thus enhancing its ability to recognize patterns associated with heart disease and improving its

predictive performance.

TABLE II
DIABETES PREDICTION METRICS ON TEST SET (TRANSFER LEARNING)

Model	Accuracy	Precision	Recall	F1 Score	AUC-ROC
ANN	0.783	0.344	0.628	0.444	0.817

TABLE III
HEART DISEASE PREDICTION METRICS ON TEST SET (TRANSFER LEARNING)

Model	Accuracy	Precision	Recall	F1 Score	AUC-ROC
ANN	0.828	0.285	0.552	0.376	0.816

C. Selecting the Ensemble Model

Our quest for the most reliable predictive model involved a comprehensive evaluation of various machine learning algorithms. The ensemble model emerged as the top performer, surpassing individual classifiers such as Logistic Regression, Support Vector Classifier (SVC), K-Nearest Neighbors (KNN), Gradient Boosting, and XGBoost.

The ensemble model's superiority became evident when considering its balanced performance with both diabetes prediction and heart disease prediction. It achieved high accuracy, precision, recall, and F1 score for the majority class in diabetes and heart disease prediction tasks. It also maintained reasonable precision and recall for minority classes, ensuring comprehensive diagnostic capability.

The decision to opt for the ensemble model was further reinforced by its ability to mitigate the limitations of individual classifiers. By aggregating predictions from diverse algorithms, the ensemble model capitalized on the strengths of each model while compensating for their weaknesses. This amalgamation resulted in a versatile and resilient predictive tool capable of handling the intricacies of medical diagnostics with precision and accuracy.

In conclusion, the ensemble model emerged as the preferred choice due to its balanced performance, robustness, and versatility. Its ability to provide accurate predictions for both majority and minority classes makes it an invaluable asset in medical decision-making, ensuring optimal patient care and treatment outcomes.

D. Age Group Analysis

In order to determine how well the models predict diabetes and heart disease within particular age groups, the study evaluates the models' performance across several age groups. Training and testing sets are created from the dataset for each age group. To address class imbalance within these

subgroups, the Synthetic Minority Over-sampling Technique (SMOTE) is utilized. These age-specific subsets are then used to retrain and evaluate the ensemble classifiers, which were previously trained using the top features found through F-Values. The predictive accuracy of the models for each age group is determined by calculating the AUC-ROC scores. This thorough investigation shows how the performance of the model changes with age, offering insights into age-specific predicting skills and assisting in the identification of any potential age-related biases or variances in model effectiveness. This information is crucial for making health risk assessments and interventions to different age groups, enhancing the model's applicability in personalized healthcare.

VI. RESULTS

A. Overview

This section details the results obtained from our predictive modeling study on diabetes and heart disease using various machine learning algorithms and feature selection techniques. We evaluated the performance of models using metrics such as accuracy, precision, recall, F1 score, and AUC-ROC scores. Our primary goal was to identify the best combination of feature selection methods and predictive algorithms to enhance the accuracy and interpretability of the models.

B. Diabetes Prediction

1) *Model Performance Comparison:* We evaluated the performance of models for diabetes prediction with and without the application of the Synthetic Minority Over-sampling Technique (SMOTE). The key metrics for comparison included accuracy, precision, recall, and F1 score. The results are summarized in Table IV.

TABLE IV
COMPARISON OF DIABETES PREDICTION RESULTS

Metric	Without SMOTE	With SMOTE
Accuracy	0.866	0.858
Precision	0.550	0.476
Recall	0.150	0.260
F1 Score	0.236	0.336
AUC-ROC	0.816	0.814

From Table IV, it is evident that applying SMOTE leads to a significant improvement in recall and F1 score, which indicates a better balance between precision and recall. While the accuracy slightly decreased with SMOTE, the increase in recall (from 0.150 to 0.260) and F1 score (from 0.236 to 0.336) suggests a more robust model for identifying diabetic patients, particularly the minority class.

2) *Feature Selection Techniques:* To enhance model performance and interpretability, we used three different feature selection techniques: SHAP values, F-test, and correlation analysis. The performance of models with each feature selection technique is shown in Table V.

TABLE V
AUC-ROC SCORES FOR DIABETES PREDICTION

Feature Selection Method	AUC-ROC Score
SHAP Values	0.802
F-test	0.814
Correlation Analysis	0.794

The F-test method yielded the highest AUC-ROC score, indicating its superior ability to identify relevant features for diabetes prediction. SHAP values and correlation analysis also provided competitive AUC-ROC scores, demonstrating their utility in feature selection.

C. Heart Disease Prediction

1) *Model Performance Comparison:* Similarly, we compared the performance of models for heart disease prediction with and without SMOTE. The results are summarized in Table VI.

TABLE VI
COMPARISON OF HEART DISEASE PREDICTION RESULTS

Metric	Without SMOTE	With SMOTE
Accuracy	0.907	0.900
Precision	0.547	0.425
Recall	0.086	0.174
F1 Score	0.148	0.247
AUC-ROC	0.816	0.812

As seen in Table VI, the application of SMOTE improved the recall and F1 score for heart disease prediction, similar to the diabetes prediction results. The recall increased from 0.086 to 0.174, and the F1 score improved from 0.148 to 0.247, indicating a better performance in identifying patients with heart disease.

2) *Feature Selection Techniques:* The performance of models with different feature selection techniques for heart disease prediction is shown in Table VII.

TABLE VII
AUC-ROC SCORES FOR HEART DISEASE PREDICTION

Feature Selection Method	AUC-ROC Score
SHAP Values	0.794
F-test	0.813
Correlation Analysis	0.812

The F-test method again provided the highest AUC-ROC score for heart disease prediction, highlighting its effectiveness in feature selection. The SHAP values and correlation analysis also demonstrated strong performance, with AUC-ROC scores of 0.794 and 0.812, respectively.

D. Comparative Analysis of Feature Selection Techniques

The comparative analysis of feature selection techniques demonstrated the F-test's superiority in identifying significant features for both diabetes and heart disease prediction. The

models utilizing the F-test achieved higher AUC-ROC scores for both conditions, as shown in Table VIII.

TABLE VIII
COMPARATIVE AUC-ROC SCORES FOR FEATURE SELECTION TECHNIQUES

Feature Selection Method	Diabetes	Heart Disease
SHAP Values	0.802	0.794
F-test	0.814	0.813
Correlation Analysis	0.794	0.812

E. Ensemble Model Performance

Ensemble models, which combined predictions from AdaBoost, Random Forest, and Gradient Boosting, demonstrated superior performance compared to individual models. The ensemble approach capitalized on the strengths of each model, leading to balanced and accurate predictions. The performance metrics of individual and ensemble models are summarized in Table X.

TABLE IX
PERFORMANCE OF INDIVIDUAL AND ENSEMBLE MODELS FOR DIABETES

Model	Accuracy	Precision	Recall	F1 Score	AUC-ROC
AdaBoost	0.803	0.360	0.545	0.433	0.805
RF	0.847	0.401	0.219	0.284	0.768
GB	0.839	0.418	0.432	0.425	0.813
Ensemble	0.858	0.476	0.260	0.336	0.814

TABLE X
PERFORMANCE OF INDIVIDUAL AND ENSEMBLE MODELS FOR HEART DISEASE

Model	Accuracy	Precision	Recall	F1 Score	AUC-ROC
AdaBoost	0.857	0.312	0.432	0.362	0.799
RF	0.892	0.329	0.147	0.203	0.767
GB	0.881	0.358	0.333	0.345	0.808
Ensemble	0.900	0.425	0.174	0.247	0.813

F. Age-Specific Model Performance

The models' performance varied across different age groups, highlighting the importance of age-specific considerations in predictive modeling. The AUC-ROC scores for diabetes and heart disease prediction across different age groups are shown in Tables XI and XII.

TABLE XI
AUC-ROC SCORES FOR DIABETES PREDICTION ACROSS AGE GROUPS

Age Group	Ensemble Method AUC-ROC
18-24	0.616
25-29	0.820
30-34	0.810
35-39	0.831
40-44	0.832
45-49	0.814
50-54	0.824
55-59	0.812
60-64	0.802
65-69	0.786
70-74	0.744
75-79	0.725
80+	0.700

TABLE XII
AUC-ROC SCORES FOR HEART DISEASE PREDICTION ACROSS AGE GROUPS

Age Group	Ensemble Method AUC-ROC
18-24	0.489
25-29	0.867
30-34	0.811
35-39	0.822
40-44	0.797
45-49	0.848
50-54	0.816
55-59	0.822
60-64	0.801
65-69	0.770
70-74	0.751
75-79	0.714
80+	0.695

The varying AUC-ROC scores across age groups suggest that models may need to be calibrated or even developed separately for different age ranges to ensure optimal performance. The lower performance in the youngest age group for both diabetes and heart disease prediction indicates that more work is needed to identify relevant predictors and improve model accuracy for young adults. The models generally perform best in the middle age ranges (25-49 years for heart disease and 25-54 years for diabetes). The declining performance in older adults suggests that the models may not be capturing all relevant risk factors or that the variability in health outcomes increases with age, making prediction more challenging. In conclusion, these findings emphasize the importance of considering age-specific factors in predictive modeling of diabetes and heart disease. Tailoring models to different age groups could enhance their accuracy and reliability.

G. Summary

The results clearly demonstrate that the F-test method for feature selection, combined with ensemble modeling techniques, provides the most accurate and robust predictive performance for diabetes and heart disease. Addressing class imbalance with SMOTE improved recall and F1 scores, ensuring that minority class instances were better represented in the models. Age-specific evaluations further emphasized the

need for tailored models to enhance prediction accuracy across different demographic groups.

The metrics for diabetes and heart disease prediction using different models are summarized in Tables XIII and XIV.

TABLE XIII
DIABETES PREDICTION METRICS ON TEST SET

AUC-ROC	Accuracy	Precision	Recall	F1 Score
0.809	0.861	0.485	0.247	0.327

TABLE XIV
HEART DISEASE PREDICTION METRICS ON TEST SET

AUC-ROC	Accuracy	Precision	Recall	F1 Score
0.811	0.901	0.424	0.167	0.240

VII. DISCUSSION

Our study presents several novel contributions that advance the state of the art in predictive modeling for diabetes and heart disease. By addressing key gaps in the existing literature, our research enhances the accuracy and applicability of machine learning models in healthcare. Below, we discuss these innovative contributions in detail.

A. Ensemble Techniques

Our study highlights the effectiveness of ensemble methods in predictive modeling for chronic diseases. Ensemble methods, such as Random Forest, Gradient Boosting, and Adaboost, leverage the strengths of multiple individual models to enhance overall predictive accuracy. The comparative analysis of individual versus ensemble models, shown in Tables XI and XII, clearly demonstrates the superior performance of ensemble models. This approach mitigates the limitations of single models, such as overfitting or bias, by combining diverse perspectives and decision-making processes. Moreover, the application of ensemble techniques in conjunction with feature selection methods like the F-test has proven to be particularly beneficial. The F-test method, which identifies features with significant predictive power, enhances the model's ability to generalize across different age groups and conditions. This combination of robust feature selection and powerful ensemble methods results in highly accurate and reliable predictive models for both diabetes and heart disease.

B. Age-Specific Model Performance Analysis

One of the critical advancements in our study is the detailed analysis of model performance across different age groups. Previous research has largely focused on overall model accuracy and performance metrics without considering demographic variations. Our age-specific performance analysis reveals significant differences in predictive power among various age groups, underscoring the importance of tailoring predictive tools to specific demographic cohorts. This approach is vital for developing more personalized and accurate models that

can effectively address the healthcare needs of different age groups.

For instance, the performance metrics, as detailed in Tables XIII and XIV, illustrate that the predictive accuracy for both diabetes and heart disease varies across age groups. The models generally perform best in the middle age ranges (25-49 years for heart disease and 25-54 years for diabetes), while the youngest and oldest age groups show lower predictive performance. This suggests that health indicators and risk factors differ significantly across ages, and models trained on age-specific data can potentially yield better predictions.

C. Dual-Focus Predictive Modeling

Unlike many studies that focus on predicting a single health condition, our research uniquely addresses both diabetes and heart disease using machine learning techniques. This dual-focus approach not only broadens the applicability of predictive models but also allows for a comparative analysis between two major health conditions. By utilizing a shared dataset and similar feature selection methods, we ensure a consistent evaluation framework, making our findings more robust and generalized.

VIII. CONCLUSION

In conclusion, this study underscores the importance of selecting appropriate feature selection techniques tailored to the dataset and modeling objectives. Our analysis demonstrated that the F-test method outperformed the other methods, achieving the highest AUC-ROC scores of 0.814 for diabetes and 0.813 for heart disease. This highlights the effectiveness of the F-test in identifying statistically significant features that enhance model accuracy. However, SHAP values and correlation analysis also proved to be valuable, with AUC-ROC scores of 0.802 and 0.794 for diabetes, and 0.794 and 0.812 for heart disease, respectively. These methods provided competitive results, showing their utility in capturing feature importance from different perspectives. Additionally, the age-specific evaluation revealed significant variability in model performance, with AUC-ROC scores ranging from 0.529 to 0.774 for diabetes and 0.465 to 0.835 for heart disease across various age groups. This variability highlights the necessity of age-specific models to address the distinct risk profiles of different age cohorts. Overall, our findings emphasize the critical role of feature selection methods and the incorporation of age-specific considerations in predictive modeling, which can lead to more effective and personalized healthcare strategies for managing chronic diseases like diabetes and heart disease. The variability in model performance across different age groups suggests that a one-size-fits-all approach may not be the most effective strategy. Instead, developing age-specific models can lead to better-targeted interventions and treatments, thus enhancing the overall effectiveness of healthcare delivery. Future work could build on these findings by exploring additional machine learning techniques and feature selection methods to further improve model performance. Additionally,

incorporating more diverse datasets and external validation could enhance the generalizability of the models.

REFERENCES

- [1] S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 40–46, 2021.
- [2] N. Ahmed, R. Ahammed, M. M. Islam, M. A. Uddin, A. Akhter, M. A. Talukder, and B. K. Paul, "Machine learning based diabetes prediction and development of smart web application," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 229–241, 2021.
- [3] H. El Massari, Z. Sabouri, S. Mhammedi, and N. Gherabi, "Diabetes prediction using machine learning algorithms and ontology," *Journal of ICT Standardization*, vol. 10, no. 2, pp. 319–337, 2022.
- [4] M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, "Diabetes prediction using ensembling of different machine learning classifiers," *IEEE Access*, vol. 8, pp. 76 516–76 531, 2020.
- [5] C. C. Olisah, L. Smith, and M. Smith, "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective," *Computer Methods and Programs in Biomedicine*, vol. 220, p. 106773, 2022.
- [6] V. V. Vijayan and C. Anjali, "Prediction and diagnosis of diabetes mellitus—a machine learning approach," in *2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*. IEEE, 2015, pp. 122–127.
- [7] S. Samreen, "Memory-efficient, accurate and early diagnosis of diabetes through a machine learning pipeline employing crow search-based feature engineering and a stacking ensemble," *IEEE Access*, vol. 9, pp. 134 335–134 354, 2021.
- [8] U. Ahmed, G. F. Issa, M. A. Khan, S. Aftab, M. F. Khan, R. A. Said, T. M. Ghazal, and M. Ahmad, "Prediction of diabetes empowered with fused machine learning," *IEEE Access*, vol. 10, pp. 8529–8538, 2022.
- [9] V. Chang, M. A. Ganatra, K. Hall, L. Golightly, and Q. A. Xu, "An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators," *Healthcare Analytics*, vol. 2, p. 100118, 2022.
- [10] L. D. Xie ZiDian and Nikolayeva, O. and Luo JieBo and, "Building risk prediction models for type 2 diabetes using machine learning techniques," *Preventing Chronic Disease*, vol. 16, p. E130, 2019.
- [11] A. Mujumdar and V. Vaidehi, "Diabetes prediction using machine learning algorithms," *Procedia Computer Science*, vol. 165, pp. 292–299, 2019.
- [12] N. Fazakis, O. Kocsis, E. Dritsas, S. Alexiou, N. Fakotakis, and K. Moustakas, "Machine learning tools for long-term type 2 diabetes risk prediction," *IEEE Access*, vol. 9, pp. 103 737–103 757, 2021.
- [13] J. Omana and M. Moorthi, "Predictive analysis and prognostic approach of diabetes prediction with machine learning techniques," *Wireless Personal Communications*, vol. 127, no. 1, pp. 465–478, 2022.
- [14] K. J. Rani, "Diabetes prediction using machine learning," *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, vol. 6, pp. 294–305, 2020.
- [15] D. VK and T. Ramesh, "Optimized stacking ensemble models for the prediction of diabetic progression," *Multimedia Tools and Applications*, pp. 1–25, 2023.
- [16] A. Sumathi and S. Meganathan, "Machine learning based pattern detection technique for diabetes mellitus prediction," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 6, p. e6751, 2022.
- [17] N. Yuvaraj and K. SriPreethaa, "Diabetes prediction in healthcare systems using machine learning algorithms on hadoop cluster," *Cluster Computing*, vol. 22, no. Suppl 1, pp. 1–9, 2019.
- [18] A. Teboul, "Diabetes health indicators dataset, brfss," 2015. [Online]. Available: <https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset>