AI-Driven Chronic Diseases Prediction Model Based on Stacking Machine Learning Algorithms

Huda Naser Hussein and Firas Sabar Miften
Department of Computer Science and Artificial Intelligence, College of Education for Pure Sciences, University of Thi-Qar, 64001 Nasiriyah, Iraq
huda_naser_hussin@utq.edu.iq, firas@utq.edu.iq

Keywords: Chronic Diseases, Data Mining, Artificial Intelligence, Machine Learning.

Abstract: Chronic diseases are responsible for most deaths worldwide. These illnesses are increasingly prevalent, and there is a dire need for systems that can predict diseases before they develop. Early detection and prediction solutions could exploit powerful tools of artificial intelligence (AI), primarily methods of machine learning (ML) and data mining. In this study, a unified stacking framework was proposed to predict three significant chronic conditions, including heart disease, diabetes, and hypertension. The proposed model consists of Random Forest (RF), XGBoost, and LightGBM as base learners, while Logistic Regression (LR) is used as a meta learner. Three datasets on heart disease, diabetes, and hypertension from BRFSS 2015 are funded by the US CDC and were utilized to demonstrate implementation of the framework. An independent integrated framework was adopted across each dataset to have the same preprocessing, model settings, and validation scenarios across pipelines. For robust and reliable performance assessment, both hold out validation and stratified five-fold cross-validation were adopted. Using stratified five-fold cross-validation, the proposed framework has shown promising predictive performance with an overall accuracy of 99.20% for heart diseases, 97.90% for diabetes, and 99.93% for hypertension. In conclusion, these results indicate a competitive and stable ensemble model towards chronic disease prediction, which can be deployed in medical decision-making applications.

1 INTRODUCTION

Chronic diseases have various causes, and they can be permanent; chronic diseases are characterized by a prolonged course that has a major impact on what we call the quality of life. While these disorders are asymptomatic initially, they may develop later in life. As per the World Health Organization (WHO), heart disease is the leading cause of chronic illness, responsible for approximately 17.9 million deaths annually. In addition, uncontrolled hypertension is a cause of almost 10 million deaths each year.

An estimated 1.6 million people die from diabetes each year [1]. Predictive analytics in the healthcare industry is greatly improved by AI, especially ML, which makes predictions more precise. These forecasts facilitate the early identification of illnesses, enabling the development of treatment strategies before the onset of illness and thereby enhancing patient care [2]. Although previous studies in this field have come a great distance, the majority of research focuses on single disease rather than multiple diseases. Furthermore, most of these are small sample size studies, indicating that it is not feasible and generalizable in real clinical practice for those models. In this paper, we create a stacking ensemble model with logistic regression as the meta-learner and RF, and XGB, LGBM as base learners. It was trained and evaluated on three datasets pertaining to heart disease, diabetes, and hypertension.

These datasets are based on real data collected within the Behavioral Risk Factor Surveillance System (BRFSS) in 2015. The rest of this paper is structured as follows Section 2 summarizes the related works. Section 3 explains the methodology adopted, the data used, how to select features and balance the dataset to develop a proposed stacking model. Section 4 describes the evaluation metrics used to examine the proposed framework. Section 5 reports the results. Section 6 is the conclusion.

2 RELATED WORKS

There are many models based on ML and DL have used in the prediction of chronic disease over past few

years. Adaptive methods segmented from traditional single classifiers to hybrid optimization and multi-task learning frameworks. In this study, Previous studies were organized and divided based on the complexity of methods they employed moving from traditional single-classifier approaches through ensembles, hybrids, and eventually advanced deep learning models offering a structured and logical perspective. Although most studies concentrated on predicting accuracy, foundational aspects including generalization, robustness and validation strategies are still critical areas for exploration.

2.1 Traditional ML-Based Approaches

Many of the studies used traditional supervised learning algorithms to predict diseases. Bouqentar et al. tested six ML algorithms (LR, SVM, RF, DT, AdaBoost and KNN) for early heart disease prediction in the Cleveland and Statlog datasets; where SVM gave the highest accuracy results (92% and 91.18%, respectively) [3]. Similarly, Afolabi et al. Using electronic health records used LR, RF and KNN for diabetes prediciton with 96.09% accuracy using KNN [4]. Yüksel et al. used BRFSS 2015 dataset to develop a quick diabetes prediction model based on RF, achieving 75.2% accuracy [5]. These models had promising performance but their effectiveness is heavily reliant on data characteristics and feature richness.

2.2 Ensemble and Stacking-Based Models

Many researchers used ensemble and stacking methods to improve predictive performance. Yadav et al. proposed to stack multiple classifiers (KNN, NB, RF, DT, SVM, ANN, MLP and CNN) using logistic regression as a meta-classifier achieving an F1-score of 98.84% in predicting diabetes [6]. Abnoosian et al. used AUC values to build a weighted voting ensemble model with 99.87% accuracy [7]. Kaliappan et al. proposed a stacking classifier using RF, XGB, GB and SVM as base learners with LR as the meta-model which resulted in an accuracy of 98.1% [8]. These results imply that ensemble based approaches are better than single based classifiers but very high accuracy values may suggest towards overfitting if stringent validation techniques are not utilized.

2.3 Hybrid and Optimization-Based Approaches

Some hybrid frameworks which combine optimization techniques have been introduced to improve the performance. AL-JAMIMI proposed a Hybrid BO–XGBoost model with tuning through Bayesian Optimization and 100% accuracy for six datasets [9]. Yanes et al. employed a recursive feature elimination with cross validation (RFECV) based ensemble model using XGBoost, LightGBM and LR with hyperparameter optimisation through Optuna, obtaining balanced accuracy of 80.75% with interpretability via SHAP analysis [10]. Several optimization strategies can enhance the predictive capacity but could also increase overfitting risk in absence of appropriate validation methods.

2.4 Deep Learning and Multi-Task Learning Frameworks

Simultaneous predictions of multiple chronic diseases have been achieved with the help of deep learning and multi-task learning models. Kim et al. introduced a three-stage system that used BRITS and LASSO for preprocessing the data, then utilized two CNN-LSTM models and a multi-task CNN-LSTM model, yielding accuracies of 89.79% for diabetes prediction and 76.41% for hypertension [11]. Zhang et al. proposed MTL-COX which is an extension of Cox model with multi-task learning for chronic disease prediction [12]. Abualigah et al. proposed a hybrid framework, leveraging GBM for feature selection and DNN for prediction on large-scale datasets like MIMIC-IV and UK Biobank, reaching 92.4% accuracy [13]. Rankovic et al. assessed three AutoML frameworks (AutoGluon, H2O, MLJAR) with knowledge in synthetic data generation through TVAE and found that AutoGluon got the best solutions [14].

Despite the extensive advancements in machine learning applications for predicting chronic diseases, most previous studies still depend on single classifiers, restricted preprocessing strategies and evaluation based on a single dataset. Further, data partitioning approaches often inadequately address issues surrounding class imbalance and model generalization across them.

Therefore, this study customizes a unified AI-driven framework that combine between feature selection, data balancing, and ensemble learning approaches to improve both predictive accuracy and generalization in context to different datasets.

3 METHODOLOGY

In this research, a series of steps were applied; starting from the data preprocessing stage, which must be done to prepare the data for analysis. Moreover, dimensionality reduction was performed using Principal Components Analysis (PCA). Next balancing methods were applied and then next the stacking method was implemented. Figure 1 illustrates the proposed framework architecture. In this approach all preprocessing, dimensionality reduction, class balance and model architecture was identical across datasets.

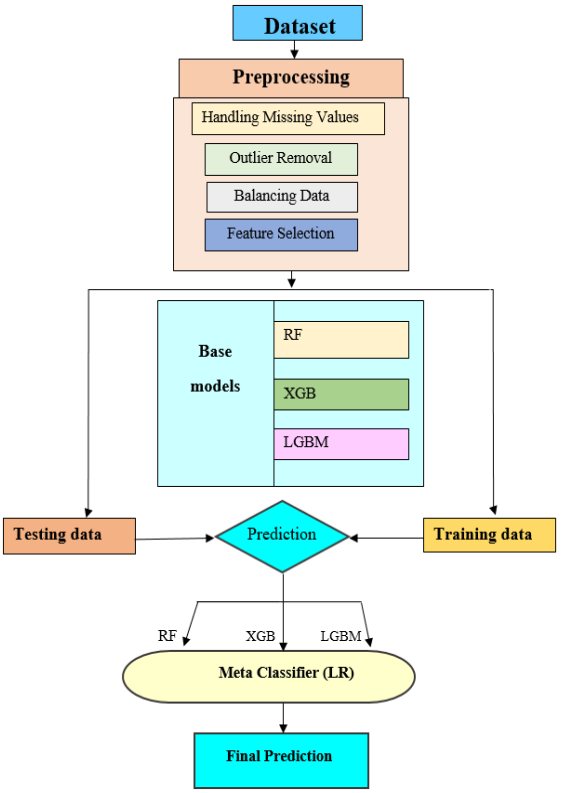

Figure 1: The architecture of the proposed framework.

Finally, we assessed the stacking model performance using numerous measures: accuracy, F1-score, recall, precision and AUC within the proposed framework. Then results were shown for each disease independently. This evaluation of our methodology helped us illustrate how the model is capable of making predictions for three separate sets of disorders, namely Heart diseases, diabetes and Hypertension.

3.1 Data Description

BRFSS 2015 dataset: The Behavioral Risk Factor Surveillance System (BRFSS), a United States government program, collects data on health-related risk behaviors, chronic health conditions, and use of preventive services from thousands of adults in the U.S. The simple random sample includes over four hundred thousand records and around three-hundred features [5]. This study utilized three publicly available subsets from the original BRFSS 2015 dataset. Chinese Heart Disease Dataset consists of 253681 samples and 22 features, with target variable [15]. The diabetes dataset has the same number of samples and features [16], and hypertension contains 26,084 records with 14 attributes [17]. The datasets differ in size/number of features and show significant class imbalance, especially in the subsets for heart disease and diabetes. The data preprocessing steps involved scaling, dimensionality reduction and class balancing to ultimately lead to prepared datasets for the model training process in a controlled experimental setting.

3.2 Data Preprocessing

The data was preprocessed to make it consistent and reliable. The median was used to impute missing values, and the Interquartile Range (IQR) method was applied to remove outliers. To address class imbalance, SMOTE technique was employed individually for each dataset to generate synthetic minority samples and achieve balanced distribution of classes.

The evaluation was performed using both the hold-out (80% training and 20% testing split) method as well with the stratified 5-fold cross-validation to guarantee robust performance assessment.

3.3 Feature Extraction

Dimensionality reduction is essential because with high-distributed digits, it takes too much time for the model to learn. Large medical datasets require efficient approaches for simpler, faster models to point out key predictors. PCA was thus applied for reduction in order to extract useful information and interparameter correlations. It leads to less correlated Principal Components which allows faster and accurate predictions, with better avoidance of overfitting. Principal component analysis was applied to all the datasets, except for hypertension which their feature space consisted of 14 features and removing any of them would not make sense in making decision.

3.4 Class Imbalance Handling

In order to overcome class imbalance issue, we applied Synthetic Minority Over-Sampling Technique (SMOTE). SMOTE generates new samples for the small class by taking instances that neighbour it and interpolating between their features, helping to achieve a better class distribution and improving generalization capacity of models.

3.5 Stacking Ensemble Approach

In this study, three machine learning models—RF, XGB, and LGBM—are integrated into a stacking ensemble framework to create a model for predicting chronic diseases. These models were selected because they performed well in challenging classification tasks and showed strong classification capabilities. The base learner layer is formed by each classifier learning independently and producing its own predictions. For the meta-learner layer, LR was employed, which receives the outputs from the base models as input and produces the final decision. Through investment in the combined strengths of these models, the stacking ensemble achieves better performance than using those models individually [8].

3.5.1 Base Learners Layer

In this section, we will explain briefly about the models were chosen as Base Learners.

3.5.1.1 Random Forest (RF)

It is an ensemble learning algorithm that constructs multiple decision trees during the training phase and then combines their outcomes through averaging or majority voting to produce the final decision. It is considered a powerful classifier, particularly effective for handling large-scale medical datasets that may contain noise, missing values, and outliers. The RF algorithm reduces overfitting due to randomness and grouping. In addition to enhancing the model's generalization capability, and transforms a collection of weak learners into a robust and reliable predictive model.

3.5.1.2 Extreme Gradient Boosting (XGB)

XGB is also an ensemble learning algorithm; however, it differs from Random Forest in the way it constructs its trees. XGB sequentially constructs trees, not independently, and the newly added tree tries to correct the residual error made by previous trees. It works on the principle of boosting and assigns weight to each tree based on how much it contributes to reducing the prediction error. Due to this approach, XGB performs very well when dealing with complex and unbalanced datasets such as the ones often present in predicting chronic disorders.

3.5.1.3 Light Gradient Boosting Machine (LGBM)

The Light Gradient Boosting Machine (LGBM) works based on the gradient boosting method. Unlike its predecessors, it develops decision trees one leaf at a time not level by level. This allows deeper trees to be built in those branches of the tree where more informative patterns exist, improving both the efficiency and predictive performance of the model. Since LGBM captures complex non-linearity among important health features with great precision, it becomes an ideal model for medical applications.

3.5.2 Meta-Learner Layer

The LR classifier that serves as the Meta-Learner model, which outputs probabilities, weighs predictions from Base Learner layer to deliver a final decision. LR acts as a final decision-maker, determining the predictive power of the base models. This is suitable for this due to its computational efficiency, ease of use and lesser overfitting. Three datasets for chronic diseases were used and preprocessing included balancing, missing values treatment and features selection with Principal Component Analysis. System efficiency and prediction accuracy were improved by implementing an ensemble model that used the stacking technique of four algorithms.

3.6 Hyperparameter Configuration

The hyperparameters of the models that were implemented are manually defined. The number of trees was fixed at 200 and the random state was fixed at 42 for Random Forest, XGB and LGBM classifiers. In the case of XGB and LGBM, 0.05 was chosen as learning rate, while the subsample ratio equals to 0.8, and for XGB, maximum tree depth was fixed at 6. In the stacking framework, LR was used with maximum of 1000 iterations. The other parameters were set to default. Hyperparameters selected for each machine learning model are displayed in Table 1. The hold-out method and stratified 5-fold cross-validation were

applied for performance evaluation. The dataset was divided into 80% training data and 20% testing data. And with 5-cross-validation came the stability of model and lower performance variance.

Table 1: Hyperparameter settings.

Model	Hyperparameters
RF	n_estimators = 200, random_state = 42
XGB	n_estimators = 200, learning_rate = 0.05, max_depth = 6, subsample = 0.8, random_state = 42
LGBM	n_estimators = 200, learning_rate = 0.05, subsample = 0.8, random_state = 42
LR	max_iter = 1000

4 EVALUATION METRICS

The performance of the proposed models was evaluated using five widely adopted classification metrics: Accuracy (ACC), Precision (PREC), Recall (REC), F1-Score (F1), and Area Under the Receiver Operating Characteristic Curve (AUC) [18]. These metrics provide complementary perspectives on model performance and are commonly used in machine learning applications.

Accuracy measures the overall proportion of correctly classified instances. Precision evaluates the reliability of positive predictions, whereas Recall assesses the model's ability to correctly identify actual positive cases. The F1-Score combines Precision and Recall into a single measure and is particularly useful when the class distribution is imbalanced. AUC reflects the ability of a model to discriminate between classes across different decision thresholds.

The results obtained using these evaluation metrics are presented in Table 2. Together, these measures provide a comprehensive assessment of the predictive performance and robustness of the examined models.

Table 2: Results on the heart disease dataset.

Model	ACC	PREC	REC	F1	AUC
RF	98.21	96.68	99.84	98.24	99.73
XGB	78.87	75.97	84.50	80.01	86.81
LGBM	78.31	75.41	84.06	79.50	86.23
Stacking	99.12	98.76	99.49	99.12	99.63

5 RESULTS

In this section, we present the experimental results of the proposed stacking model when evaluated on three benchmark datasets (diabetes, hypertension and heart disease). In order to provide a comprehensive and reliable evaluation of the predictive performance of the model, two validation strategies were used: a hold-out method (80/20 train–test split) and 5-fold cross-validation. The performance of the framework was evaluated through the chosen metrics. These evaluation metrics give us an in-depth picture of how well a model performs in classifying, its robustness and ability to generalize across distinct datasets.

5.1 Results on Heart Disease Dataset

In this section, we report experimental results for heart disease prediction based on the methods outlined by our proposed framework from both hold-out method and stratified 5-fold cross-validation.

5.1.1 Hold-Out Validation Results

The proposed stacking ensemble framework when applied on the heart disease dataset using the hold-out validation method (80/20 split) yielded high predictive performances for all evaluation metrics. The experimental performance shows that the composite model proposed outperforms the base learners in Accuracy, Precision, Recall, F1-score and AUC. The results of the model are shown in Table 2 on the heart disease data set. confusion matrix was created to understand the classification mistakes of the model, shown in Figure 2.

A The first 10 important features selected from the PCA technique were taken as the main variables of the heart disease dataset. Reducing the feature space to these key attributes lowered the computational complexity and enhanced the model's accuracy without compromising prediction quality. Figure 3 illustrates the selected features and their contribution to the prediction process. The stacking model achieved the highest overall performance on the heart disease dataset, exceeding the accuracy of all individual classifiers. This result confirms the ability and effectiveness of the proposed framework in predicting heart disease cases.

Table 3: 5-Fold cross-validation results (mean ± standard deviation) for the heart disease dataset.

Model	ACC	PREC	Recall	F1	AUC
RF	97.38 ± 0.07	95.03 ± 0.11	99.81 ± 0.05	97.36 ± 0.07	99.79 ± 0.02
XGB	76.85 ± 0.14	73.51 ± 0.14	81.48 ± 0.15	77.29 ± 0.13	84.62 ± 0.10
LGBM	76.67 ± 0.11	73.41 ± 0.10	81.12 ± 0.15	77.07 ± 0.11	84.32 ± 0.10
Stacking	99.20 ± 0.02	98.84 ± 0.04	99.50 ± 0.05	99.17 ±0.02	99.69 ± 0.03

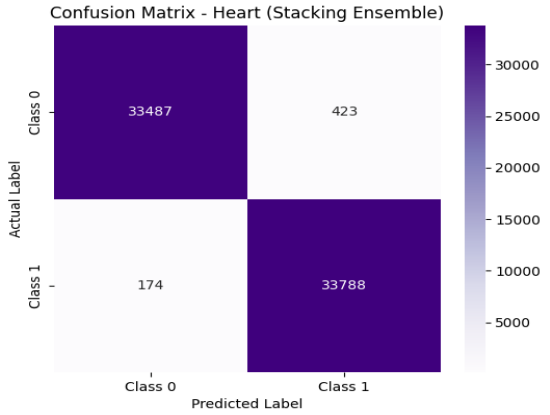

Figure 2: Heart disease dataset confusion matrix.

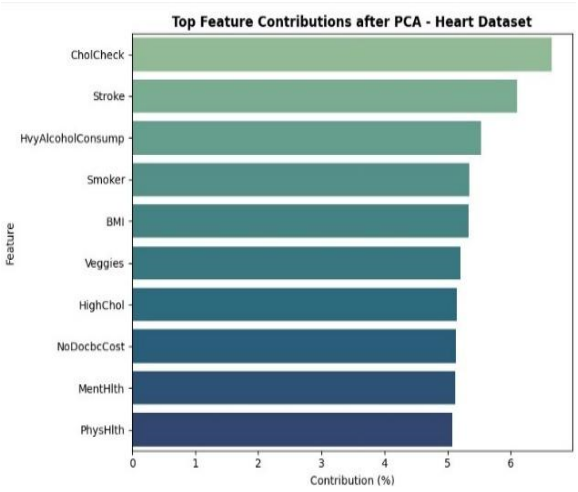

Figure 3: PCA-based feature contribution.

5.1.2 Cross-Validation Result

Five-fold stratified cross-validation was applied to evaluate the proposed model for robust and unbiased estimation. This method splits the data into five equal parts, then trains on four with one held out for validation, repeat five times to average performance. The model stacking performance on the balanced Heart Disease dataset is shown in Table 3. As observed from the evidence, it can be seen that the model performed quite well on average for all technical indicators with fewer standard deviations, which proves its high validity and accuracy across folds.

5.2 Results on Diabetes Dataset

This subsection summarizes the results of diabetes prediction using our stacking framework, with both hold-out evaluation and stratified 5-fold cross-validation separately for clear comparison.

5.2.1 Hold-Out Validation Results

For the diabetes dataset, the proposed framework also achieved strong results, but they are slightly lower than the results for heart disease. This is because the diabetes dataset is highly imbalanced, where the number of diabetic cases are less than non-diabetic cases. To address this imbalance, we reclassified class 2 diabetes patients to increase the number of diabetic cases, in addition to the balancing method used in the proposed framework. All of this contributed to this model achieving high results in diabetes, as shown in Table 4 and Figure 4 presents the confusion matrix for diabetes disease dataset.

The results showed that the proposed framework also achieved high performance in diabetes as well as heart disease, despite the data being imbalanced.

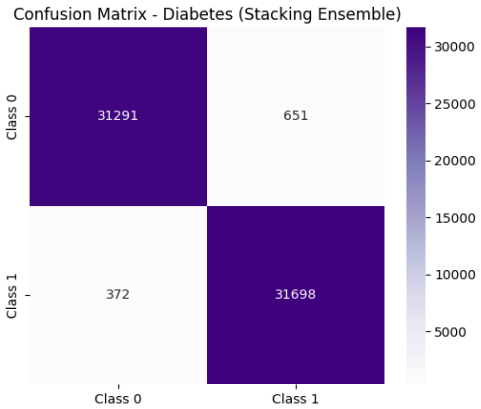

Figure 4: The confusion matrix for diabetes dataset.

5.2.2 Cross-Validation Result

As shown in the cross-validation results in Table 5, the proposed stacking model also achieved high performance on diabetes dataset. While the scores obtained here are somewhat lower than those from the hold-out evaluation, it is evident that our model nevertheless exhibited strong predictive power and stably uniform performance across all five folds.

5.3 Results on Hypertension Dataset

In this subsection, we present prediction results of hypertension based on the unified framework. Separate performance analysis results are reported for hold-out method and standard stratified 5-fold cross-validation.

5.3.1 Hold-Out Validation Results

Regarding hypertension, the proposed framework achieved very high performance on the hypertension dataset. The hypertension dataset was simple, lacking complexity and did not have high noise levels like diabetes. Furthermore, the number of outputs was very low. All of this contributed to a single classifier, such as RF, being sufficient for it. Therefore, the classifiers used in this proposed framework, such as Baseliner, all achieved complete results because they are powerful classifiers. Additionally, PCA was not applied, and the information remained complete in this dataset. Consequently, the model could easily learn patterns and implement full performance as shown in Table 6.

The following Figure 5 explains the confusion matrix of the proposed framework applied to the hypertension dataset.

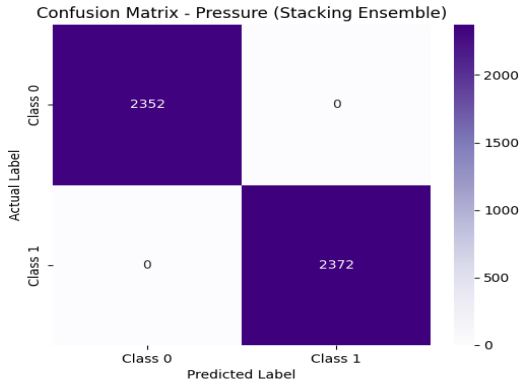

Figure 5: The confusion matrix for hypertension.

5.3.2 Cross-Validation Result

The stacking model has performed exceedingly well comparatively on 5-fold cross validation result on the Hypertension dataset as we see in Table 7. These scores are marginally lower than the hold-out evaluation, but the model is showing high predictive capabilities and stable performance across folds.

Table 4: Results on the diabetes dataset.

Model	ACC	PREC	Recall	F1	AUC
RF	95.72	92.60	99.41	95.88	99.33
XGB	75.32	73.38	79.63	76.37	83.29
LGBM	74.92	72.98	79.30	76.01	82.76
Stacking	98.40	97.99	98.84	98.41	99.22

Table 5: CV performance (mean ± standard deviation) for the diabetes dataset.

Model	ACC	Precision	Recall	F1-score	AUC
RF	94.12 ±0.07	89.63 ± 0.13	99.33 ±0.05	94.23 ± 0.07	99.34 ± 0.04
XGB	74.10 ±0.10	71.27 ± 0.12	77.76 ±0.04	74.37 ± 0.08	81.83 ± 0.13
LGBM	73.99 ±0.10	71.21 ± 0.13	77.54 ±0.10	74.24 ± 0.08	81.62 ± 0.13
Stacking	97.90 ±0.06	96.87 ± 0.07	98.85 ±0.07	97.85 ± 0.07	99.25 ± 0.06

Table 6: Results on hypertension dataset.

Model	ACC	PREC	Recall	F1	AUC
RF	100%	100%	100%	100%	100%
XGB	100%	100%	100%	100%	100%
LGBM	100%	100%	100%	100%	100%
Stacking model	100%	100%	100%	100%	100%

Table 7: CV performance for the hypertension.

Model	ACC	PREC	REC	F1	AUC
RF	99.92 ±0.02	99.91± 0.02	99.94 ±0.02	99.92 ± 0.02	100.00 ± 0.00
XGB	99.15 ±0.18	98.68 ± 0.27	99.78 ±0.09	99.23 ± 0.16	99.96 ± 0.02
LGBM	99.35 ±0.16	99.12 ± 0.36	99.70 ±0.12	99.41 ± 0.15	99.97 ± 0.02
Stacking	99.93 ±0.02	99.94 ± 0.06	99.94 ±0.03	99.94 ± 0.02	100.00 ± 0.00

The overall experimental results on heart disease, diabetes, and hypertension datasets explained the robustness and effectiveness of the proposed unified stacking framework. The stable performance obtained from both evaluation methods indicates the strong generalizability of the model and validates its practicality for early chronic disease prediction.

6 CONCLUSIONS

Detecting chronic illness early is important because these conditions continue to be a leading cause of global mortality. A total of different machine learning algorithms were implemented and assessed on three standard datasets in this study. The experimental findings exhibited that the proposed stacking based framework consistently outperformed the individual base classifier for all evaluation metrics. In the proposed framework, several key components are integrated together contributing to improved performance of the proposed model such as preprocessing steps, class imbalance handling, features selection and the chosen models. All of these pieces in combination helped to improve predictive stability, reduce overfitting risk and enhance generalizability. These properties of the model validated its reliability and also notating which data partitioning strategy gave it robust and consistent performance. With the relatively high accuracy, precision, recall, F1-score and AUC values achieved by the proposed framework it shows strong potential for real-world deployment on healthcare decision-support systems. Further research can validate the framework on larger-scale and real-world clinical datasets and also find ways to embed it in real-time healthcare monitoring systems.

REFERENCES

[1] Y. Liu and B. Wang, “Advanced applications in chronic disease monitoring using IoT mobile sensing device data, machine learning algorithms, and frame theory: A systematic review.”

[2] A. Al-Nafjan, A. Aljuhani, A. Alshebel, A. Alharbi, and A. Alshehri, “Artificial intelligence in predictive healthcare: A systematic review,” pp. 1-17, 2025.

[3] M. A. Bouqentar et al., “Early heart disease prediction using feature engineering and machine learning algorithms,” Heliyon, vol. 10, no. 19, pp. 1-23, 2024, doi: 10.1016/j.heliyon.2024.e38731.

[4] S. Afolabi, N. Ajadi, A. Jimoh, and I. Adenekan, “Predicting diabetes using supervised machine learning algorithms on e-health records,” Inform. Health, vol. 2, no. 1, pp. 9-16, 2025, doi: 10.1016/j.infoh.2024.12.002.

[5] A. K. Yüksel and M. S. Güzel, “Efficient diabetes prediction using random forests and minimal health indicators on the BRFSS dataset,” vol. 2, no. 1, pp. 34-43, 2025.

[6] P. Yadav, S. C. Sharma, R. Mahadeva, and S. P. Patole, “Exploring hyper-parameters and feature selection for predicting non-communicable chronic disease using stacking classifier,” IEEE Access, vol. 11, pp. 80030-80055, 2023, doi: 10.1109/ACCESS.2023.3299332.

[7] K. Abnoosian, R. Farnoosh, and M. H. Behzadi, “Prediction of diabetes disease using an ensemble of machine learning multi-classifier models,” BMC Bioinformatics, vol. 24, no. 1, pp. 1-24, 2023, doi: 10.1186/s12859-023-05465-z.

[8] J. Kaliappan et al., “Analyzing classification and feature selection strategies for diabetes prediction across diverse diabetes datasets,” Front. Artif. Intell., vol. 7, 2024, doi: 10.3389/frai.2024.1421751.

[9] H. A. Al-Jamimi, “Synergistic feature engineering and ensemble learning for early chronic disease prediction,” IEEE Access, vol. 12, pp. 62215-62233, 2024, doi: 10.1109/ACCESS.2024.3395512.

[10] N. Yanes, L. Jamel, B. Alabdullah, M. Ezz, A. Mohamed Mostafa, and H. Shabana, “Using machine learning for detection and prediction of chronic diseases,” IEEE Access, vol. 12, pp. 177674-177691, 2024, doi: 10.1109/ACCESS.2024.3494839.

[11] W. Vin et al., “Development of a machine learning model for precision prognosis of rapid kidney function decline in people with diabetes and chronic kidney disease,” Diabetes Res. Clin. Pract., vol. 217, p. 111897, 2024, doi: 10.1016/j.diabres.2024.111897.

[12] S. Zhang, F. Yang, L. Wang, S. Si, J. Zhang, and F. Xue, “Personalized prediction for multiple chronic diseases by developing the multi-task Cox learning model,” vol. 19, no. 9, 2023, doi: 10.1371/journal.pcbi.1011396.

[13] L. Abualigah et al., “Artificial intelligence-driven translational medicine: A machine learning framework for predicting disease outcomes and optimizing patient-centric care,” J. Transl. Med., vol. 23, no. 1, 2025, doi: 10.1186/s12967-025-06308-6.

[14] N. Rankovic, D. Rankovic, and I. Lukic, "Innovations in early detection of chronic non-communicable diseases among adolescents through an easy-to-use AutoML paradigm," Health Care Manag. Sci., vol. 28, no. 3, pp. 434-460, 2025, doi: 10.1007/s10729-025-09718-6.

[15] A. Teboul, "Heart disease health indicators dataset," Kaggle, [Online]. Available: https://www.kaggle.com/datasets/alexteboul/heart-disease-health-indicators-dataset.

[16] Teboul, "Diabetes health indicators dataset," Kaggle, [Online]. Available: https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset.

[17] P. Chuks, "Health dataset (hypertension)," Kaggle, [Online]. Available: https://www.kaggle.com/datasets/prosperchuks/health-dataset.

[18] A. Ogunpola, F. Saeed, S. Basurra, A. M. Albarrak, and S. N. Qasem, "Machine learning-based predictive models for detection of cardiovascular diseases," Diagnostics, vol. 14, no. 2, 2024, doi: 10.3390/diagnostics14020144.