International Journal of Electrical and Computer Engineering (IJECE)
Vol. 16, No. 1, February 2026, pp. 321~332
ISSN: 2088-8708, DOI: 10.11591/ijece.v16i1.pp321-332

Hybrid machine learning framework for chronic disease risk assessment

Harini Shadaksharappa[1], Rashmi K. B.[2], Shreyas D. K.[3], Somanath Mikali[4], Vishesh P. Gowda[5], Uday Shankar C. A.[6], Siddarth B. Iyerr[7]
Department of Information Science and Engineering, B.M.S. College of Engineering, Bengaluru, India

Article Info

Article history:

Received Jan 11, 2025
Revised Oct 15, 2025
Accepted Nov 23, 2025

Keywords:

Artificial neural networks
Chronic diseases
Health metrics
Machine learning
Predictive model

ABSTRACT

Chronic diseases like asthma, diabetes, stroke, and heart disease are the major causes of morbidity globally, which emphasizes the need for efficient predictive models to facilitate early detection and precautionary measures. Previous studies have used machine learning approaches for single-disease prediction, where models are designed for specific diseases, such as diabetes or heart disease. However, very few attempts have been made to develop unified frameworks for predicting multiple diseases simultaneously. This work presents a novel, unified framework using an ensemble of extreme gradient boosting classifier (XGBClassifier) and artificial neural networks (ANN) as individual classifiers to concurrently predict the risk of developing asthma, diabetes, stroke, and heart disease. This work follows a questionnaire-based approach that utilizes demographic, lifestyle, health metrics, symptoms and exposure-related data to create personalized risk assessments. The model achieves satisfactory accuracy rates of 95.82% for asthma, 96.68% for diabetes, 94.91% for stroke, and 94.52% for heart disease. The findings highlight how this novel hybrid model serves as an effective approach to tackle the intricate interactions between chronic ailments. The research also includes a user-friendly website that comprises a questionnaire and makes use of the best performing model to predict the probabilities of developing different diseases.

Corresponding Author:

Rashmi K. B.
Department of Information Science and Engineering, B.M.S. College of Engineering
P.O. Box No.: 1908, Bull Temple Road, Bangalore-560019, Karnataka, India
Email: rashmikb.ise@bmsce.ac.in

1. INTRODUCTION

Over the last few decades, chronic diseases have reached unprecedented levels across the globe. The most prominent ones include asthma, diabetes, stroke, and cardiovascular disease. These conditions, taken together, form a huge burden on the health care systems around the world [1]. Heart diseases alone are responsible for almost 17.9 million deaths annually, as claimed by the World Health Organization (WHO) and over 500 million people suffer from chronic respiratory ailments, including asthma [2]. Diabetes affects almost 500 million people, and the figure is projected to increase by 25% in 2030 [3]. Preventive detection and prevention of these diseases have become one of the key challenges in modern healthcare systems, especially in areas where there is a lack of modern medical facilities and diagnostic equipment. These conditions are especially prevalent in developing countries, where healthcare access may be limited. Traditional diagnostic approaches generally need sophisticated laboratory testing, skilled medical personnel, and adequate resources, which hampers prompt identification and treatment. These issues are exacerbated in rural and impoverished communities, where access to specialized health services is limited or nonexistent.

The advent of computational methods in machine learning and the sphere of artificial intelligence, especially ensemble learning techniques, has changed the methodologies that are used in disease prediction. While the traditional methods used in medical diagnosis depend on a combination of clinical acumen and laboratory assessments, ensemble strategies like voting classifiers, in the presence of powerful algorithms such as extreme gradient boosting (XGBoost) and artificial neural network (ANN) interchangeably used as multi-layer perceptron (MLP) as used in this approach [4], are indeed valid alternatives for exact disease forecasting. Latest technological advancements greatly affect preventive care as early diagnosis of health problems has tremendously improved patient outcomes and decreased healthcare costs.

Comprehensive approaches to disease prediction are also vital because chronic diseases are often interrelated with common risk factors and coexisting symptoms. For instance, the risk factors related to environmental aspects can simultaneously increase respiratory and cardiovascular conditions, such as air quality or pollution level. However, diabetes is widely regarded as a factor that significantly elevates the potential for heart conditions and stroke [5]. The interaction among these chronic illnesses makes a call for holistic methodologies in predicting them, involving assessment of many interconnected risk factors and their complexities.

This research addresses these problems through an innovative multiple disease predictive system; the ensemble combines XGBClassifier and ANN to evaluate simultaneously the hazard of bronchial asthma, diabetes, stroke, and coronary heart disease. The integrated approach will use a system based on questions that will be demographic, lifestyle, health metrics, symptoms, and exposure-based to create personalized risk analysis. The uniqueness in methodology is reflected in its integrated framework designed for multiple disease prediction, thereby addressing a notable gap within the existing frame of research, which primarily focuses on models for the prediction of single diseases.

The rest of this paper is prepared as follows: section 2 affords an in-depth review of literature pertaining to the scope of machine learning in disease prediction. Section 3 addresses the methodologies that include data collection/pre-processing to the voting classifier with XGBClassifier and ANN. Furthermore, section 4 deals with the experiment results on different metrics. Lastly, section 5 has a summary of findings with a future perspective.

2. RELATED WORK

Recent research has shown substantial advances in the development of integrated systems for predicting numerous diseases at once. Gopisetti *et al.* [6] suggested a method to forecast several diseases using machine learning, demonstrating the ability to create user-friendly online applications in healthcare diagnostics utilizing frameworks such as Streamlit. Ray *et al.* [7] discussed different machine learning techniques for forecasting a variety of diseases and highlighted the importance of diagnostic tools in an integrated manner in modern healthcare conditions.

Significant advancements have been made in stroke prediction in recent years. The latest research carried out by Gupta *et al.* [8] achieved an accuracy of 95.16% in neural networks which is an important milestone in the prediction of strokes. Rahman *et al.* [9] took their findings a step forward to produce 99% accuracy with the aid of Random Forest ensemble algorithms. There is, however an important work by Mridha *et al.* [10] that emphasizes proper validation techniques wherein they proved how although random forest obtained 90.36% accuracy on the entire dataset, with the more realistic train-test split, it came out to be 82.23%, thereby proving the significance of avoiding data leakage in model evaluation. Elangovan *et al.* [11] significantly contributed to the literature by discussing the critical problem of imbalanced datasets in stroke prediction, offering very useful insights on how to deal with this ubiquitous problem in medical data analysis.

Several major studies have come out on diabetes prediction. Hasan *et al.* [12] investigated the application of ensemble algorithms in predicting diabetes, whereas Mujumdar and Vaidehi [13] reported significant outcomes utilizing different algorithms—specifically, their gradient boost model demonstrated an accuracy of 93%, whereas the logistic regression model achieved an outstanding accuracy of 96%. Diabetes prediction is always being improved, and Rani's research [14] claimed 99% accuracy using decision trees. Some studies, like Soni and Varma [15], showed more modest results, as random forest obtained 77% accuracy, showing the diversity of models' performance on different datasets and approaches. Yahyaoui *et al.* [16] suggested valuable insights despite using a smaller dataset of 768 samples, achieving 83.67% accuracy.

The research in asthma prediction has focused significantly on ecological factors. The trend in medical care usage due to environmental factors has been discussed by Jo *et al.* [17], and Hwang *et al.* [18] have applied deep learning methods for predicting the count of asthma patients through environmental information. Louisias *et al.* [19] studied the environmental determinants of asthma with regard to its symptoms, especially the role of pollen, allergens, and dust. A systematic review by Jayamini *et al.* [20] has analyzed an extensive range of machine learning techniques, which includes techniques such as logistic

regression, decision trees and ensemble methods like random forests, gradient boosting machines, and neural networks to predict asthma exacerbations.

Various research studies on the prediction of heart disease have proven to be effective using different methodologies. Dritsas and Trigka [21] achieved high accuracy with 87.8% and area under the curve (AUC) of 98.2% using a stacking ensemble model applied after synthetic minority over-sampling technique (SMOTE), utilizing 10-fold cross-validation. Bhatt *et al.* [22] showed success with MLP models at 87.23%. Kavitha *et al.* [23] proposed a hybrid model with accuracy 88.7%, and Sarra *et al.* [24] achieved better performance using an ANN model, which depicted up to 93.44% in accuracy and AUC of 0.95. Recent work by Yadav *et al.* [25] achieved 94.51% accuracy with AdaBoost and random forest feature selection, though their precision (48.33) and recall (39.52) metrics on test data highlight persistent challenges in clinical applicability. Overall, the set of studies represents the development and advancement of heart disease prediction models, which in several contexts, demonstrate promising progress.

3. METHODOLOGY

The current work attempts to develop a strong and integrated framework for predicting the developing risk of chronic conditions, including asthma, diabetes, stroke, and heart diseases. The motivation behind developing such a framework is to assist practitioners as well as patients during health care by providing timely warnings and suitable advice. For this purpose, we used multiple machine learning algorithms, applied preprocessing techniques to improve data quality, and adopted an ensemble-based approach to improve accuracy in prediction as highlighted by Figure 1.

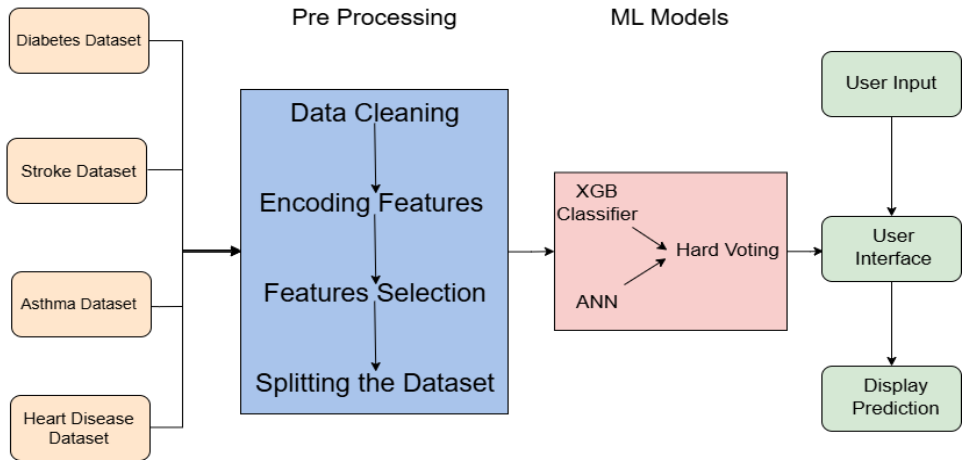

Figure 1. Flow diagram of model

3.1. Data source

The current study used datasets from free source to predict the likelihood of four chronic diseases: asthma, diabetes, stroke, and heart disease. These datasets included vast amounts of data containing demographic, medical, and lifestyle attributes relevant to each disease.

- Diabetes dataset (*https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset*): It contained 100,000 instances with feature descriptions including age, gender, body mass index (BMI), history of hypertension, factors such as a history of heart disease, smoking habits, Glycated hemoglobin (HbA1c) levels, and blood glucose measurements. The target variable for diabetes is binary.
- Stroke dataset (*https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset*): It comprises 5,110 records with features that include age, gender, marital status, hypertension, heart disease, smoking, glucose levels, BMI, and type of work, and residence with a binary target variable for the stroke.
- Asthma dataset (*https://www.kaggle.com/datasets/rabieelkharoua/asthma-disease-dataset*): There are 2,392 samples and 29 variables in the asthma dataset. The variables comprise demographics (age, gender, ethnicity), lifestyle factors (smoking, physical activity, diet, sleep), environmental exposures (pollution, pollen, dust), medical history (family asthma, allergies, eczema), clinical tests (forced expiratory volume in 1 second, forced vital capacity), and symptoms (wheezing, dyspnea, chest tightness, coughing). The target variable is the asthma diagnosis.

- Heart disease dataset (*https://www.kaggle.com/datasets/tarekmuhammed/patients-data-for-medical-field/data*): This dataset contained 237,630 records. It included demographics, along with height and weight, BMI, medical history such as a heart attack, angina, stroke, diabetes, or asthma, lifestyle factors like smoking or e-cigarettes, preventive care like vaccinations, chest scans, and recent health events, such as Coronavirus disease of 2019 (COVID-19). Heart disease history was the binary target variable.

3.2. Data pre-processing

3.2.1. Data cleaning

Every dataset was cleaned and freed of redundancy by removing duplicate entries and imputing missing values. Numerical features were handled using median imputation while categorical features were handled using mode imputation. Irrelevant attributes, such as patient IDs, were eliminated in order to achieve good accuracy.

3.2.2. Feature engineering

The feature engineering method was used to enhance the prediction ability of the model. Numerical features, including age, BMI, blood glucose, blood pressure, and HbA1c, were scaled using standard scaling to align them on the same scale. Binary features were subjected for label encoding such as gender, presence of certain disorder and one-hot encoding was implemented for multi-class ones like the medical conditions, the type of work, education and race/ethnicity. This allowed for proper interpretation of the categorical data. Using heatmap analysis, a feature selection method based on correlation was employed to determine which predictors are most pertinent and reduce the number of dimensions while keeping the ability to predict.

3.3. Modeling

3.3.1. XGBoost classifier

Extreme gradient boosting, or XGBoost for short, is a potent supervised machine learning technique for tasks involving regression and classification. It uses a boosting technique to expand on decision trees, where in relevant variables are given more weight and used in the subsequent decision tree in case the tree makes a false prediction. The outputs of each classifier or predictor are subsequently integrated to form a more robust and accurate model. By altering weights according to previous errors, XGBoost mixes the outputs of several trees additively, in contrast to Random Forest, which averages them. This enables more complex predictions. Regularization to avoid overfitting, parallel processing for performance, and a weighted quantile sketch technique for handling sparse data are important aspects. The loss function, algorithm of XGBoost classifier is as follows:

$$L_{XGB} = \sum_{i=1}^{N} L\left(y_i, F(x_i)\right) + \sum_{m=1}^{M} \Omega(h_x)$$

$$\Omega(h) = \gamma T + \frac{1}{2}\lambda\left||w^2|\right|$$

here, $L(y_i, F(x_i))$ is log loss function, $\Omega(h)$ is regularization term for each tree h, T is number of leaves of the tree, γ is parameter to control lowest loss reduction gain to split a node, and w is output values from the leaves.

Algorithm of XGBoost classifier

1. Model initialization
   Initialize $F_0(x)=0$
2. Iterative boosting process
   for t = 1 to T:
   - Calculate the gradient of the loss function $g_i = \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}$
   - Fits decision tree to predict these gradients
   - Computes tree predictions $h_t(x)$
   - Updating the model as $F_{t+1}(x)=F_t(x)+h_t(x)$
     where is η the learning rate $(0<\eta \leq 1)$
   - Application of regularization components as:
     L2 regularization on leaf weights $(\lambda \sum w_j^2)$
     Complexity penalty on number of leaves (T)

   Total objective: $Loss + \lambda \sum w_j^2 + \gamma T$
3. The final model is an ensemble of T trees combined additively and predictions based on cumulative tree outputs.

3.3.2. Multi-layer perceptron

Among the types of artificial neural networks is a multilayer perceptron (MLP). The architecture accommodates an input layer, one or more hidden layers, and an output layer. The weighted sum of inputs of each layer in a neural network is used by the neurons for performing activation functions. MLPs can learn complex patterns through the process of backpropagation wherein the weights are modified by reversing the propagation of errors. They do exceedingly well on tasks such as image recognition and natural language processing because they can mimic nonlinear interactions. A pictorial representation of an MLP architecture is shown in Figure 2. Performance requirements for MLPs require optimal settings of characteristic parameters like the learning rate and the count of hidden layers among others.

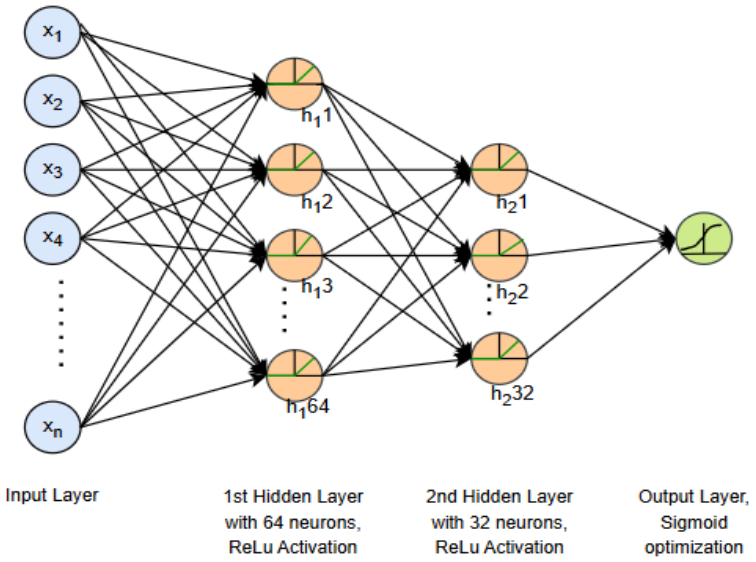

Figure 2. A multi-unit perceptron with 2 hidden layers of 64 and 32 neurons in each layer respectively

Algorithm 1. Multi-unit perceptron

1. Model initialization
   Weights $W^{(1)}$ and biases $b^{(1)}$ are initialized for each layer of MLP.
2. Defining the architecture of MLP namely:
   - L: Total number of layers in MLP.
   - h_l: Number of neurons in each layer l.
   - η : Learning rate.
   - E: Number of iterations/epochs.
   - Rectified Linear Unit (ReLU) activation for the hidden layers, $f(z) = \max(0, z)$
   - Sigmoid activation function for the output layer, (z) $\frac{1}{1+e^{-z}}$
3. Training the model
   for e = 1 to E:
   a. Forward propagation
      - Input to network $a^{(0)}=x$
      - For each layer l=1 to L-1 (all hidden layers):
        $z^{(1)}=W^{(1)}.a^{(1-1)}+b^{(1)}$
        $a^{(1)}=ReLU((z^{(1)}))$
      - When output layer reached l=L:
        $z^{(L)}=W^{(L)}.a^{(L-1)}+b^{(L)}$
        $a^{(L)}=$ Sigmoid(z(L))
        $a^{(L)}$ represents the predicted output $\hat{y}$.
   b. Loss computation
      Since classification loss is computed using binary cross-entropy function

$$L = -\frac{1}{m}\sum_{i=1}^{m}[y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i)]$$

   c. Backward propagation
      - Computation of gradient of loss with respect to output

$$\delta^{(L)} = \frac{\partial L}{\partial a^{(L)}} \odot \sigma'(z^{(L)})$$

      - Propagating the error through hidden layers backward

For $l = L - 1$ to 1:
$\delta^{(l)} = (W^{(l+1)})^T . \delta^{(l+1)} \odot ReLU'(z^{(l)})$
d. Updating parameters using Adam optimizer
4. The end model consists of computed weights and biases for all layers and output is predicted using forward propagation through the trained parameters given by formula

$$\hat{y} = Sigmoid(z^{(L)})$$

3.3.3. Voting classifier

It is an ensemble technique that uses several models to enhance the accuracy of classification. It uses majority voting or probability averaging to combine predictions from various classifiers (such as decision trees and support vector machines). By utilizing the advantages of various models, this method improves resilience and lowers the possibility of overfitting.

4. RESULT ANALYSIS

The models were tested to forecast health conditions such as Asthma, Diabetes, Stroke, and heart diseases. For that purpose, two primary models have been utilized: XGBoost classifier (XGBC) and ANN. Further refining of results was achieved using voting classifier by taking both models together.

The approach uses the XGBoost model with the XGBClassifier from the XGBoost library, optimized with $use\_label\_encoder = False$ and evaluated using logloss. The former is trained using default parameters and the results were robust across all target diseases. The model demonstrated high predictive performance, benefitting from XGBoost's ability to handle missing data and capture complex relationships.

The ANN was designed using the Keras library in TensorFlow, with a sequential architecture having three fully connected layers: 64 neurons with rectified linear unit (ReLU) activation in the first layer, 32 neurons with ReLU in the hidden layer, and a single neuron with sigmoid activation for binary classification running for 50 epochs. The model used the Adam optimizer, binary cross-entropy loss function with accuracy as the evaluation criterion. The accuracy of the model as evaluated on the datasets is as tabulated in Table 1.

Table 1. Accuracy of different models for disease prediction

Disease	Model	Accuracy (%)
Asthma	XGBC	95.82
	ANN	95.82
	Voting Classifier	95.82
Diabetes	XGBC	97.085
	ANN	96.76
	Voting Classifier	96.68
Stroke	XGBC	94.28
	ANN	94.91
	Voting Classifier	94.91
Heart disease	XGBC	94.45
	ANN	94.61
	Voting Classifier	94.52

The classification reports for stroke, diabetes, asthma, and heart disease predictions are illustrated in Figures 3 to 6. For stroke and asthma in Figures 3 and 5, the model achieved excellent precision, recall, and F1-scores for the negative class (all above 0.95), while positive class metrics were absent due to class imbalance; nonetheless, macro and weighted averages remained robust. In diabetes prediction in Figure 4, strong performance was observed for both classes, with positive class precision, recall, and F1-scores exceeding 0.75, and overall averages above 0.95, reflecting balanced detection. For heart disease in Figure 6, the model maintained high scores for the negative class, while positive class metrics were moderate, leading to solid macro and weighted averages. These results underscore the model's strong and consistent performance in accurately identifying both negative and positive cases across all disease categories and demonstrate the model's robust capability and positive impact in multi-disease prediction.

The user interface dashboard provides users with a clear and interactive means to input their data and view preliminary results. Data collection was facilitated by the structured questionnaire illustrated in Figure 7, which ensured comprehensive and standardized input from all users. The output of the predictive model, presented in Figure 8, visually displays the percentage risk of each disease for every participant, highlighting variations and enabling targeted analysis.

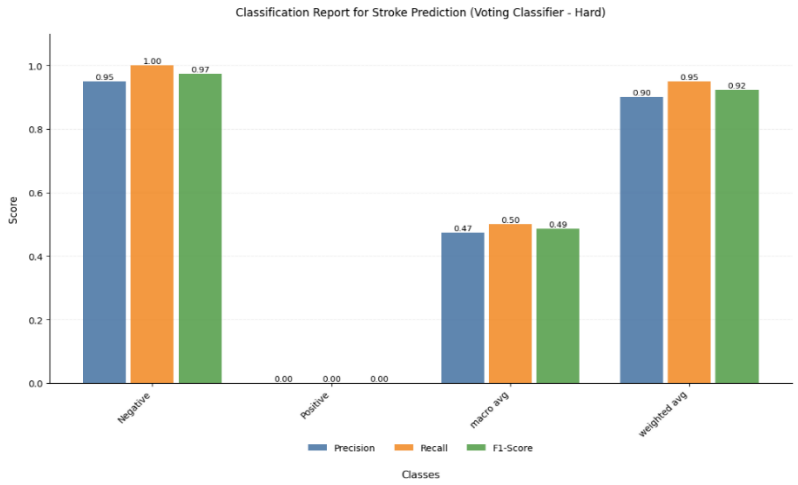

Figure 3. Classification report for stroke prediction representing precision, recall, F1-score for positive, negative classes along with macro and weighted average

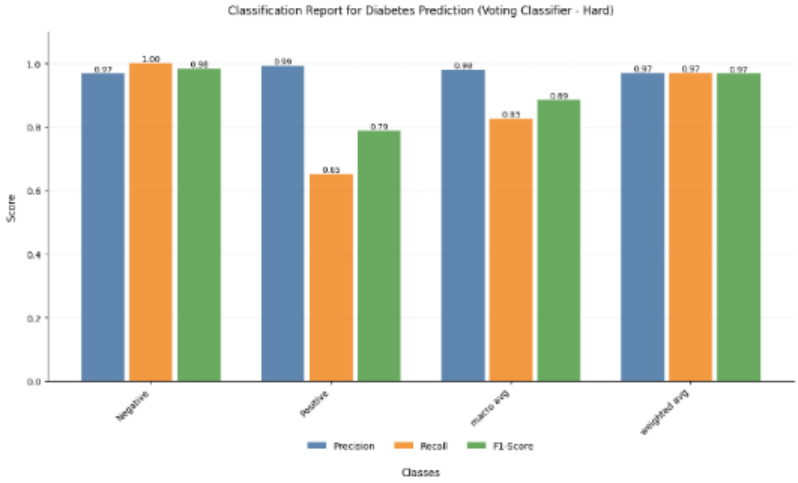

Figure 4. Classification report for diabetes prediction representing precision, recall, F1-score for positive, negative classes along with macro and weighted average

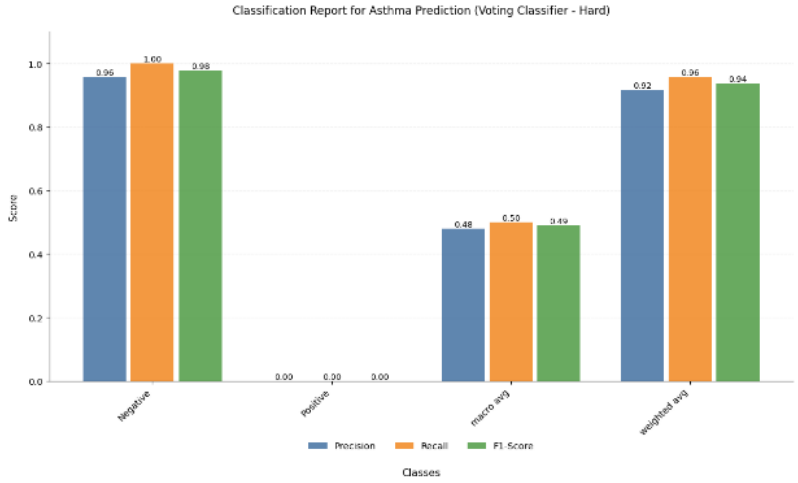

Figure 5. Classification report for asthma prediction representing precision, recall, F1-score for positive, negative classes along with macro and weighted average

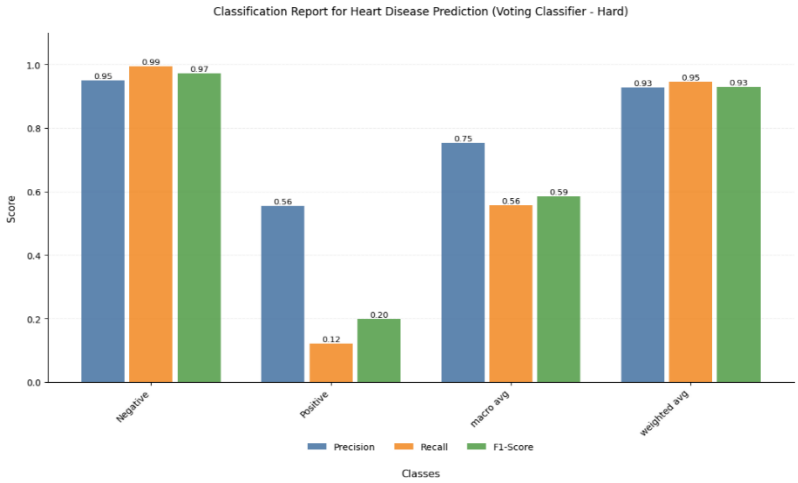

Figure 6. Classification report for heart disease prediction representing precision, recall, F1-score for positive, negative classes along with macro and weighted average

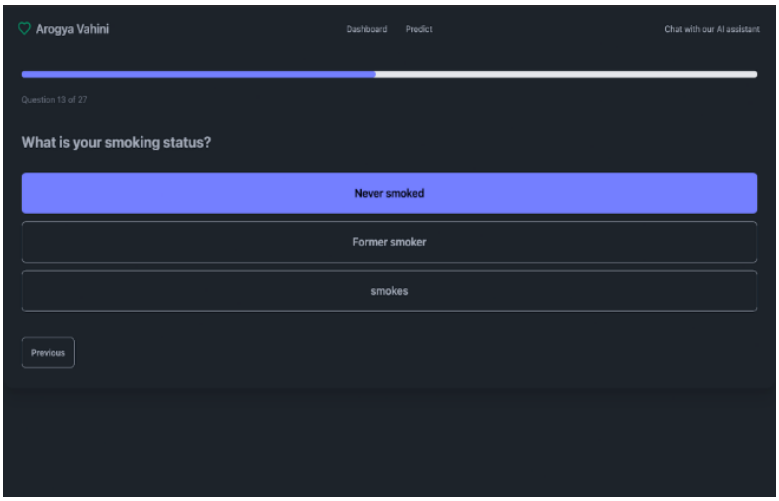

Figure 7. Questionnaire to collect data for analysis

Figure 8. Disease risk output of the model

5. CONCLUSION

This work introduces a system developed to predict the risk of four of the most significant chronic diseases namely asthma, diabetes, stroke, and cardiovascular disease, by making use of machine learning techniques. The model achieves remarkable accuracy for all four diseases via ensemble learning using XGBoost classifier and ANN with a hard voting classifier approach. The results are as follows: In the case of asthma prediction, XGBoost classifier and ANN achieved accuracies of 95.82% and 95.82% respectively, whereas the voting classifier maintained an accuracy of 95.82%. In diabetes prognosis, XGBoost classifier achieved an accuracy of 97.085%, ANN logged 96.76% while the voting classifier evaluated the metrics to 96.68% accuracy. Stroke prediction models also performed well, with XGBoost classifier at 94.28%, ANN at 94.91%, and the voting classifier remaining consistent at 94.91%. The prediction of heart disease produced an accuracy of 94.45% for XGBoost classifier, 94.61% for ANN, and 94.52% for the voting classifier. The findings emphasize the dependability of the proposed system, which consistently attains accuracies exceeding 94%. This accentuates the promise of ensemble learning approaches in the healthcare domain by harnessing the advantages of individual models while alleviating their limitations.

A web application that is designed to be user-friendly was built for this research to enhance the accessibility of this application to healthcare providers and patients. It includes a smooth interface for gathering real-time inputs through a questionnaire that can contribute in making accurate predictions of the four chronic conditions. In conclusion, this research calls attention to the promise of machine learning techniques in accurately predicting a range of chronic diseases. The creation of an ergonomic web application significantly improves its applicability, successfully connecting cutting-edge technology with the practical demands of healthcare in real-world settings. This research opens up new avenues for advancements in predictive healthcare and facilitates the ongoing digital evolution in medical diagnosis and prognosis.

There are numerous areas where the established system could be improved in the future. Incorporating data from real-time health monitoring would significantly enhance the system's capacity for dynamic prediction. Furthermore, inclusion of a wider variety of diseases and ailments could improve the system's efficacy for thorough health analysis. Also, studying the effect of demographic and regional differences on model performance would lead to large-scale implementation of the model. It could also be used to detect disease progression patterns and to make the prediction more personalized genomic data set can be input. To enhance the user experience while using the platform, interpretable Artificial Intelligence techniques can be used to provide explanations of the factors used during predictions.

FUNDING INFORMATION

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Harini Shadaksharappa	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	
Rashmi K. B.		✓		✓		✓	✓	✓	✓	✓	✓	✓	✓	
Shreyas D. K.	✓	✓		✓	✓		✓	✓	✓	✓				
Somanath Mikali	✓	✓	✓			✓		✓			✓			
Vishesh P. Gowda	✓	✓	✓		✓									
Uday Shankar C. A.	✓	✓	✓		✓			✓	✓	✓	✓			
Siddarth B. Iyerr		✓	✓	✓		✓			✓		✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest (financial, personal, or professional) in connection with manuscripts and have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author, RKB. The data, which contain information that could compromise the privacy of research participants, are not publicly available due to certain restrictions.

REFERENCES

[1] K. Hacker, "The burden of chronic disease," *Mayo Clinic Proceedings: Innovations, Quality \& Outcomes*, vol. 8, no. 1, pp. 112–119, 2024, doi: 10.1016/j.mayocpiqo.2023.08.005.

[2] WHO, "Cardiovascular diseases (CVDs)." World Health Organization, Accessed: Jan. 05, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[3] P. Saeedi *et al.*, "OP-0256 – Global and regional diabetes prevalence: estimates for 2019 and projections for 2030 and 2045," *International Diabetes Federation*, vol. 157, p. 2045, 2019, doi: 10.1016/j.diabres.2019.10784.

[4] B. Boehmke and B. Greenwell, *Hands-on machine learning with SKLerni, Keras and TensorFlow*, 2nd ed. Sebastopol, CA: O'Reilly Media, 2019.

[5] NIDDK, "Diabetes, heart disease, & stroke," National Institute of Diabetes and Digestive and Kidney Diseases, Accessed: Jan. 05, 2025. [Online]. Available: https://www.niddk.nih.gov/health-information/diabetes/overview/preventing-problems/heart-disease-stroke.

[6] L. D. Gopisetti, S. K. L. Kummera, S. R. Pattamsetti, S. Kuna, N. Parsi, and H. P. Kodali, "Multiple disease prediction system using machine learning and Streamlit," in *Proceedings - 5th International Conference on Smart Systems and Inventive Technology, ICSSIT 2023*, 2023, pp. 923–931, doi: 10.1109/ICSSIT55814.2023.10060903.

[7] S. Ray, S. Ali, and A. Katiyar, "Multiple disease prediction using ML," in *Proceedings - IEEE 2023 5th International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2023*, 2023, pp. 310–313, doi: 10.1109/ICAC3N60023.2023.10541406.

[8] A. Gupta *et al.*, "Predicting stroke risk: An effective stroke prediction model based on neural networks," *Journal of Neurorestoratology*, vol. 13, no. 1, p. 100156, 2025, doi: 10.1016/j.jnrt.2024.100156.

[9] S. Rahman, M. Hasan, and A. K. Sarkar, "Prediction of brain stroke using machine learning algorithms and deep neural network techniques," *European Journal of Electrical Engineering and Computer Science*, vol. 7, no. 1, pp. 23–30, 2023, doi: 10.24018/ejece.2023.7.1.483.

[10] K. Mridha, S. Ghimire, J. Shin, A. Aran, M. M. Uddin, and M. F. Mridha, "Automated stroke prediction using machine learning: an explainable and exploratory study with a web application for early intervention," *IEEE Access*, vol. 11, pp. 52288–52308, 2023, doi: 10.1109/ACCESS.2023.3278273.

[11] V. S. Elangovan, R. Devarajan, O. I. Khalaf, M. S. Sharif, and W. Elmedany, "Analysing an imbalanced stroke prediction dataset using machine learning techniques," *Karbala International Journal of Modern Science*, vol. 10, no. 2, pp. 246–259, 2024, doi: 10.33640/2405-609X.3355.

[12] 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. 76516–76531, 2020, doi: 10.1109/ACCESS.2020.2989857.

[13] A. Mujumdar and V. Vaidehi, "Diabetes prediction using machine learning algorithms," *Procedia Computer Science*, vol. 165, pp. 292–299, 2019, doi: 10.1016/j.procs.2020.01.047.

[14] K. J. Rani, "Diabetes prediction using machine learning," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 294–305, Jul. 2020, doi: 10.32628/CSEIT206463.

[15] M. Soni and D. S. Varma, "Diabetes prediction using machine learning techniques," *International Journal of Engineering Research & Technology*, vol. 09, no. 09, 2020.

[16] A. Yahyaoui, A. Jamil, J. Rasheed, and M. Yesiltepe, "A decision support system for diabetes prediction using machine learning and deep learning techniques," in *1st International Informatics and Software Engineering Conference: Innovative Technologies for Digital Transformation, IISEC 2019 - Proceedings*, 2019, pp. 1–4, doi: 10.1109/UBMYK48245.2019.8965556.

[17] E. J. Jo *et al.*, "Patterns of medical care utilization according to environmental factors in asthma and chronic obstructive pulmonary disease patients," *Korean Journal of Internal Medicine*, vol. 36, no. 5, pp. 1146–1156, 2021, doi: 10.3904/KJIM.2020.168.

[18] H. Hwang, J. H. Jang, E. Lee, H. S. Park, and J. Y. Lee, "Prediction of the number of asthma patients using environmental factors based on deep learning algorithms," *Respiratory Research*, vol. 24, no. 1, p. 302, 2023, doi: 10.1186/s12931-023-02616-x.

[19] M. Louisias, A. Ramadan, A. S. Naja, and W. Phipatanakul, "The effects of the environment on asthma disease activity," *Immunology and Allergy Clinics of North America*, vol. 39, no. 2, pp. 163–175, 2019, doi: 10.1016/j.iac.2018.12.005.

[20] W. K. D. Jayamini, F. Mirza, M. Asif Naeem, and A. H. Y. Chan, "Investigating machine learning techniques for predicting risk of asthma exacerbations: A systematic review," *Journal of Medical Systems*, vol. 48, no. 1, p. 49, 2024, doi: 10.1007/s10916-024-02061-3.

[21] E. Dritsas and M. Trigka, "Efficient data-driven machine learning models for cardiovascular diseases risk prediction," *Sensors*, vol. 23, no. 3, p. 1161, 2023, doi: 10.3390/s23031161.

[22] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective heart disease prediction using machine learning techniques," *Algorithms*, vol. 16, no. 2, p. 88, 2023, doi: 10.3390/a16020088.

[23] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart disease prediction using hybrid machine learning model," in *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021*, 2021, pp. 1329–1333, doi: 10.1109/ICICT50816.2021.9358597.

[24] R. R. Sarra, A. M. Dinar, and M. A. Mohammed, "Enhanced accuracy for heart disease prediction using artificial neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 1, pp. 375–383, 2023, doi: 10.11591/ijeecs.v29.i1.pp375-383.

[25] A. K. Yadav, G. K. Pal, and S. Gangwar, "Early stage prediction of heart disease features using AdaBoost ensemble algorithm and tree algorithms," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 3, pp. 545–551, 2024.

BIOGRAPHIES OF AUTHORS

Harini Shadaksharappa working as an assistant professor in the Department of Information Science and Engineering, B.M.S. College of Engineering. She obtained her M.Tech., in computer science and engineering from School of Engineering, Jain University, Bangalore in the year 2012, and registered for Ph.D. under V.T.U. in December 2017. Her research areas include image processing, high performance computing and machine learning. She can be contacted at email: hharinis8@gmail.com.

Rashmi K. B. working as an assistant professor, in the Department of Information Science and Engineering at B.M.S. College of Engineering. She holds an M.Tech. degree and has registered for Ph.D. from Visvesvaraya Technological University. Her research interests include natural language processing, information retrieval and machine learning. She can be contacted at email: rashmikb.ise@bmsce.ac.in.

Shreyas D. K. is currently pursuing a Bachelor of Engineering in Information Science and Engineering at B.M.S. College of Engineering, Bengaluru, India, and a Bachelor of Science in data science and applications at the Indian Institute of Technology Madras, India. He has been actively engaged in research projects involving machine learning, artificial intelligence, and healthcare analytics. He is currently working on Agentic AI systems and the development of next-generation AI models. His research interests include intelligent health systems, advanced machine learning frameworks, generative AI, predictive modeling, secure data-driven architectures, and the application of AI in public health. He can be contacted at email: shreyasdk28@gmail.com.

Somanath Mikali is a pre-final year undergraduate student in the Department of Information Science and Engineering at B.M.S. College of Engineering, Bengaluru, India. His research interests lie at the intersection of machine learning and real-world applications, particularly in healthcare, intelligent transport systems, and secure AI deployment. He has contributed to the development of ensemble learning frameworks for chronic disease risk prediction and is currently working on secure, VANET-based traffic optimization systems incorporating post-quantum cryptographic techniques. He can be reached at email: somanath.r.mikali@gmail.com.

Vishesh P. Gowda is currently pursuing a bachelor of engineering in information science and engineering at B.M.S. College of Engineering, Bengaluru, India. His work spans research and development across cloud computing, large language models (LLMs), and applied machine learning. He has contributed to projects involving chronic disease risk assessment using ensemble models and is passionate about building intelligent, scalable systems at the intersection of AI and healthcare. His current interests include cloud-native AI architectures, multi-modal LLM applications, agentic systems, and privacy-aware machine learning pipelines. He is also exploring the integration of generative AI and predictive analytics into real-world platforms. He can be reached at email: visheshhirrisave@gmail.com.

Uday Shankar C. A. is currently pursuing a bachelor of engineering in information science and engineering at B.M.S. College of Engineering, Bengaluru, India. He also holds a diploma in information science and technology from P.V.P. Polytechnic. He has been consistently involved in research projects focused on machine learning, artificial intelligence, and healthcare analytics. His current work includes the development of an AI-driven government scheme accessibility platform aimed at bridging the gap between citizens and public welfare schemes through intelligent automation. He can be contacted at email: udayshankarca2004@gmail.com.

Siddarth B. Iyerr is currently pursuing a bachelor of engineering in information science and engineering at B.M.S. College of Engineering, Bengaluru. He has been an enthusiast in the field of machine learning and aims to explore his ability in many researches in the field. He also extends his abilities in the field of artificial intelligence (AI), cloud computing and internet of things (IoT). His projects include credit card fraud detection system, demonstration of FastAPI and sentiment analyzer app. He can be contacted at email: siddarthbiyer@gmail.com.