Hindawi

Research Article

Blended Ensemble Learning Prediction Model for Strengthening Diagnosis and Treatment of Chronic Diabetes Disease

T. R. Mahesh [ID],[1] **Dhilip Kumar** [ID],[2] **V. Vinoth Kumar** [ID],[1] **Junaid Asghar** [ID],[3] **Banchigize Mekcha Bazezew** [ID],[4] **Rajesh Natarajan** [ID],[5] and **V. Vivek** [ID][1]

[1]Department of Computer Science and Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bangalore, India
[2]Department of Computer Science and Engineering,
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India
[3]Faculty of Pharmacy, Gomal University, Dera Ismail Khan 29050, Khyber Pakhtunkhwa, Pakistan
[4]Department of Electrical and Computer Engineering, Wollo University, Ethiopia
[5]Department of Information Technology, University of Technology and Applied Science, Shinas. Sultanate of Oman, Oman

Correspondence should be addressed to Banchigize Mekcha Bazezew; banchigizeb@kiot.edu.et

Diabetes mellitus (DM), commonly known as diabetes, is a collection of metabolic illnesses characterized by persistently high blood sugar levels. The signs of elevated blood sugar include increased hunger, frequent urination, and increased thirst. If DM is not treated properly, it may lead to several complications. Diabetes is caused by either insufficient insulin production by the pancreas or an insufficient insulin response by the body's cells. Every year, 1.6 million individuals die from this disease. The objective of this research work is to use relevant features to construct a blended ensemble learning (EL)-based forecasting system and find the optimal classifier for comparing clinical outputs. The EL based on Bayesian networks and radial basis functions has been proposed in this article. The performances of five machine learning (ML) techniques, namely, logistic regression (LR), decision tree (DT) classifier, support vector machine (SVM), *K*-nearest neighbors (KNN), and random forest (RF), are compared with the proposed EL technique. Experiments reveal that the EL method performs better than the existing ML approaches in predicting diabetic illness, with the remarkable accuracy of 97.11%. The proposed ensemble learning could be useful in assisting specialists in accurately diagnosing diabetes and assisting patients in receiving the appropriate therapy.

1. Introduction

Diabetes mellitus (DM) is a collection of metabolic disorders indicated by abnormal insulin production as well as its action. This insulin deficiency leads to elevated blood levels of glucose and decreased carbohydrate and protein metabolism [1]. DM is a prevalent endocrine disease that affects roughly 200 million people across the world. Diabetes affected 30.3 million persons in the United States in 2015, accounting for 9.4% of the population. More than one-fourth of them were unaware that they had this condition. One out of four people over the age of 65 has diabetes [2]. Gradually, the human living environment changes dramatically due to increased demand on

the health of human beings, human lifestyle, changes in climate, and other environmental changes. Infections and diseases cause millions of individuals to fall unwell and face other harmful health issues, and these problems are quickly increasing every day.

DM is a chronic condition that happens when the pancreas stops producing insulin or when the body's insulin is not used properly. Insulin is a hormone produced by the pancreas that functions as a key to allowing glucose from our diet to move from the bloodstream into our cells for energy production. In the blood, all carbohydrate foods are broken down into glucose. Insulin aids glucose absorption into cells. When the body is unable to make or utilize insulin properly,

blood glucose levels rise (known as hyperglycaemia). High glucose levels are linked to long-term harm to the body and the failure of many organs and tissues. Diabetes that is not well controlled can have catastrophic repercussions, including damage to a variety of organs and tissues in the body, including the heart, kidneys, eyes, and nerves. The different symptoms of the DM include frequent hunger, frequent thirst, frequent urination, urinary and genital infections, extreme tiredness, wounds that will not heal, and unexplained weight loss especially in case of children. Because many people do not experience any of these symptoms, diabetes is frequently diagnosed by chance. Those who are at risk of acquiring diabetes should have regular medical examinations to avoid late discovery of the disease. The different symptoms of diabetes are shown in Figure 1.

There are three main kinds of diabetes. Due to the complexity of diabetes, diagnosing and distinguishing among these different kinds are difficult [3–5].

1.1. Type-1. An autoimmune reaction occurs when the body's defense system assaults the cells that make insulin, resulting in type-1 diabetes. As a result, the body produces very little insulin. The specific causes are unknown although they are thought to be connected to a combination of genetic and environmental factors. Type-1 diabetes can strike anyone at any age although it strikes children and adolescents the most commonly. When a person has type-1 diabetes, his or her body generates very little or no insulin, necessitating daily insulin injections to keep blood glucose levels in check.

1.2. Type-2. Type-2 diabetes is the most common form of diabetes, accounting for approximately 90% of all cases. Insulin resistance is a condition in which the body's response to insulin is inadequate. Because insulin is unable to function effectively, blood glucose levels rise, causing more insulin to be released. This can gradually fatigue the pancreas, resulting in the body producing less and less insulin, further raising blood sugar levels (hyperglycaemia). Type-2 diabetes is most typically diagnosed in older persons although it is becoming more common in children, adolescents, and younger adults as obesity, physical inactivity, and poor diet become more prevalent. The major complications that might be caused by type-2 diabetes are depicted in Figure 2.

Diabetes has a dangerous effect on the eyes' small blood vessels. As a result, it may cause problems such as glaucoma (a condition in which fluid builds up in the eyes), cataracts (clouding of the lens of the eyes), and diabetes retinopathy (a condition in which the retina at the back of the eye is damaged). Diabetes can cause visual loss or blur over time. As a result, regular eye examinations by a qualified ophthalmologist are required. Early detection can help to avoid serious consequences. According to polls, early identification can prevent blindness in patients with diabetes by 90 percent. Figure 3 shows the normal eye and eye with retinopathy.

1.3. Gestational Diabetes Mellitus (GDM). GDM is a serious and underappreciated threat to the health of mothers and

FIGURE 1: Diabetes' symptoms.

children. High blood pressure, large birth weight babies, and obstructed labour are common pregnancy problems for women with GDM. Within five to ten years following delivery, around half of women with a history of GDM acquire type-2 diabetes. The prevalence of high blood glucose during pregnancy rises dramatically with age, peaking in women over 45. However, diabetes mellitus, most usually type 2, is discovered in roughly 5–10% of women with GDM after pregnancy. GDM is completely curable, but it necessitates close medical monitoring during pregnancy.

2. Related Work

The utilization of ML approaches to improve the precision of DM risk prediction has been studied in a variety of ways [6]. The authors [7] have done a comparison of the accuracy of numerous algorithms on the dataset of diabetes. They discovered that the J48 algorithm has 73.82% accuracy before data preprocessing, which is superior to others. Following preprocessing, both KNN and RF showed better performance. In this work [8], J48, KNN, and LR were compared on the diabetes dataset in this work. J48 was found to be the most accurate, with a classification accuracy of 78.27%. Based on its accuracy of 80%, this work [9] created web-based software for DM detection. In this work, the authors compared prediction techniques such as decision trees (DT), neural networks, Naive Bayes (NB), LR, and RF, and ensemble techniques. The authors found that RF performed the best in terms of accuracy and ROC score of 75.558% and 0.912, respectively.

The authors in this [10] looked at different ML algorithms on a medical dataset. They compared different performance indicators. In this research, ML and data mining approaches used in DM research were identified and reviewed in a systematic manner. Diabetes mellitus is quickly becoming a very important and pressing worldwide health issue of the twenty-first century. At present, significant research has been conducted in practically every facet of DM research. When compared to other ML algorithms, authors [11] employed
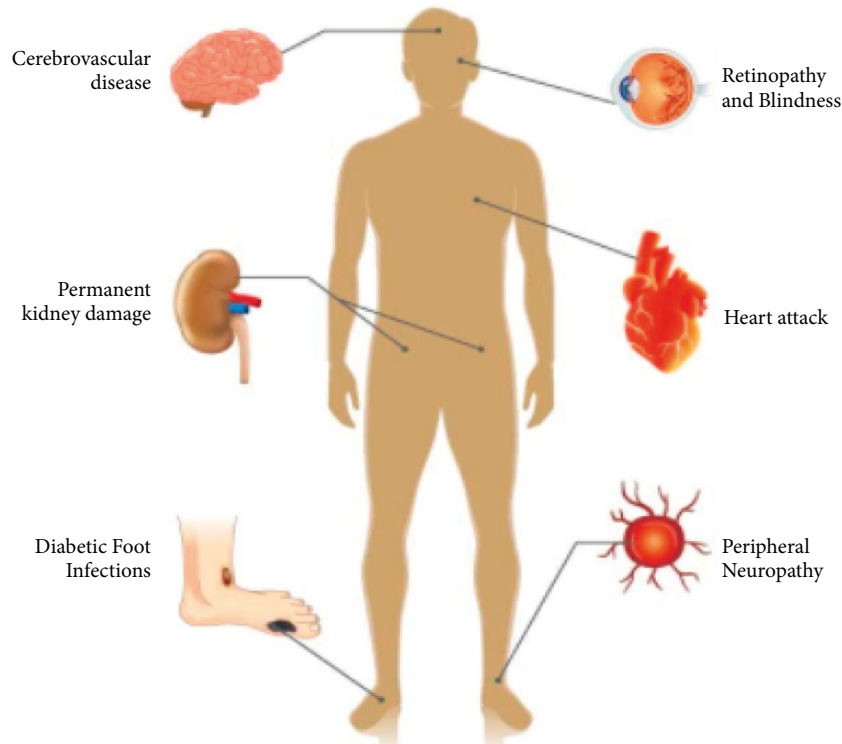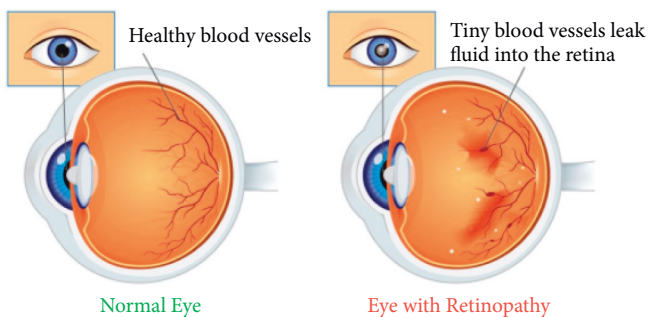
FIGURE 2: Type-2 diabetes complications.

FIGURE 3: Diabetes retinopathy.

C4.5 with 78.25% accuracy. According to them, J48 is famous for its correctness.

The authors of this article [12] present an overview of how DM is linked to cardiovascular autonomic neuropathy that impairs HRV. As a result, diabetes was discovered using HRV spectral analysis. Using DWT features retrieved from HRV signals, the authors demonstrated an automated DM detection technique. Using the DT classifier, they were able to get sensitivity, accuracy, and specificity of 92%, 92.59%, and 91.46%, respectively, using their proposed method. In their diabetes forecasting experiments, the authors looked into NB, ANN, as well as SVM classifiers [7]. They conducted a research study that was weighted. The majority vote method was used in this investigation. The findings show that a combination of classifiers outperformed any individual classifier.

The authors of this study [13] examined different ML algorithms on the Pima Indian dataset. Feature selection was applied to the dataset. According to their findings, attribute selection improved the performance of diabetes prediction. This study's retrospective cross-sectional methodology prevents the authors from establishing a cause-and-effect link. This strategy may simply be utilized in studies to find the optimal type-2 diabetes phenotype or prediction in different nations. However, because the study group was limited to Korean women and men, the findings could not be extended to other populations. There can be disparities in socioeconomic class, race, gender, and nationality within a single patient cohort.

Using the J48 DT, the authors [6] emphasized the application of Adaboost ensemble methods for diagnosing the disease. According to the findings of the trial, Adaboost ensemble approaches outperform J48 and even bagging. In this study, the scientists employed NB, J48, and RF for the prediction of DM. In terms of accuracy, they discovered that RF outperforms naive Bayes and J48 techniques. According to the authors [14], the accuracy achieved by implementing PCA was not good, but the outcome of using all features along with MRMR was superior. The results showed that by using fasting glucose made the performance better, particularly in the Luzhou dataset. This implies that, while fasting glucose is the most significant predictor, it cannot produce the best results on its own, so multiple indices were required to effectively predict. Furthermore, they discovered that there is little difference between RT, DT, and neural network results when comparing the three classifications; however, RFs are definitely better compared to the other three classifiers. The highest performance for the Luzhou dataset was 80.84%, while that of the Pima Indians dataset was 77.21%,

indicating that ML can be used to predict diabetes, but selecting appropriate features, classifiers, and data mining methods is crucial. According to the results, the Adaboost ensemble technique outperforms bagging and the stand-alone J48 DT in this study [15].

Several ML algorithms for diabetes mellitus prediction were utilized in the studies listed above, and they have been evaluated for getting good results. In this article, the ML techniques DT, RF, LR, SVM, and KNN classifiers are utilized as base learners, and their performance is compared to that of the ensemble classifier using various performance measures, namely, accuracy, precision, sensitivity, and f1-measure in the prediction of diabetes.

3. Methodology

The likelihood of survival and the likelihood of diabetes recurrence are highly dependent on medical treatment and the accuracy of the diagnosis. The various phases of the prediction process of diabetes are depicted in Figure 4.

The dataset taken from Kaggle consists of 693 diabetes-affected people's records. The different symptoms along with their descriptions are shown in Table 1. Arbitrarily extracted information was used in this investigation, with a ratio of 75 : 25 split between training and testing data. The model was trained to utilize training sets (520 records), and its effectiveness was tested via test data. The dataset consists of 16 features or attributes whose values will determine whether the person is likely to be affected by diabetes or not. The target or output variable has a value of 1 for the presence of diabetes and 0 for its absence.

3.1. Logistic Regression (LR).
This approach has been modified for binary classification problems. The fundamental goal of LR is to figure out what the coefficients are worth. The LR reduces the value to a number between 0 and 1. The LR model chooses a probability of 0 or 1 for the provided data instance of the class to predict. This method can be used to solve problems where we have various reasons to forecast. Equation (1) is the definition of the LR standard function:

$$h\theta(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}. \tag{1}$$

The constants are represented in β_1 and β_0 by the data label X.

3.2. K-Nearest Neighbor (KNN).
KNN is also a supervised ML algorithm [11]. It is mostly used to solve categorization challenges. The object is categorized using K-neighbors. Before running the procedure, the positive number K must be defined. Euclidean distance is frequently used to determine various object measurements [16]. The Euclidean distance and Manhattan distance are calculated using the following equations:

$$\text{Euclidean} = \sqrt{\sum_{i=1}^{k} (a_i - b_i)^2}, \tag{2}$$

$$\text{Manhattan} = \sum_{i=1}^{k} |a_i - b_i|. \tag{3}$$

From the above equations, Euclidean as well as Manhattan of KNN is computed with a and b data upto I variables.

3.3. Support Vector Machine (SVM).
The SVM algorithm [17] is a supervised ML approach. For a small data set with a few outliers, this model is ideal. The goal is to determine the hyperplane that will be used to split the data points. The hyperplane will be used to divide two spaces into different domains. Similar forms of data will be found in such a domain. The support vector machine's decision state is represented by the following equation:

$$||Y|| = \sqrt{y_1^2 + y_1^2 + \cdots + y_n^2}. \tag{4}$$

3.4. Decision Tree (DT).
DT is one technique to show an algorithm that is made up of completely conditional control statements. The DTs are a collection of well-known supervised classification algorithms. They do well on classification problems; the decisional path is straightforward to interpret, and the algorithm for building (training) them is quick and very simple. A single DT is depicted in Figure 5.

3.5. Random Forest (RF).
RF is an ML technique that is part of the supervised ML model. The RF classifier is made up of numerous decision trees representing various subjects [18]. It takes the average of each tree's subset to improve predictive accuracy. RF, rather than depending on a single decision tree, uses the majority vote prediction technique from all the trees and then predicts the result [19]. Every node in the decision tree answers a query about the situation. RF example is depicted in Figure 6.

3.6. Proposed Ensemble Learning (EL) Technique.
For the categorization of diabetes mellitus, an ensemble classifier based on Bayesian Networks and Radial Basis Functions is proposed in this article [20]. Ensemble classifiers are modelled using Bayesian networks (BNs) as well as Radial Basis Function (RBF). Two individual classifiers and the suggested ensemble classifier will be discussed in the subsections. Figure 7 depicts the suggested approach's process flow diagram.

3.6.1. Bayesian Network (BN).
Traditional statistical models do not allow for the incorporation of past knowledge into the model. The probabilistic associations between items or objects are represented using a BN [21]. BNs are directed acyclic graphs with nodes indicating variables and missing edges representing conditional independence between
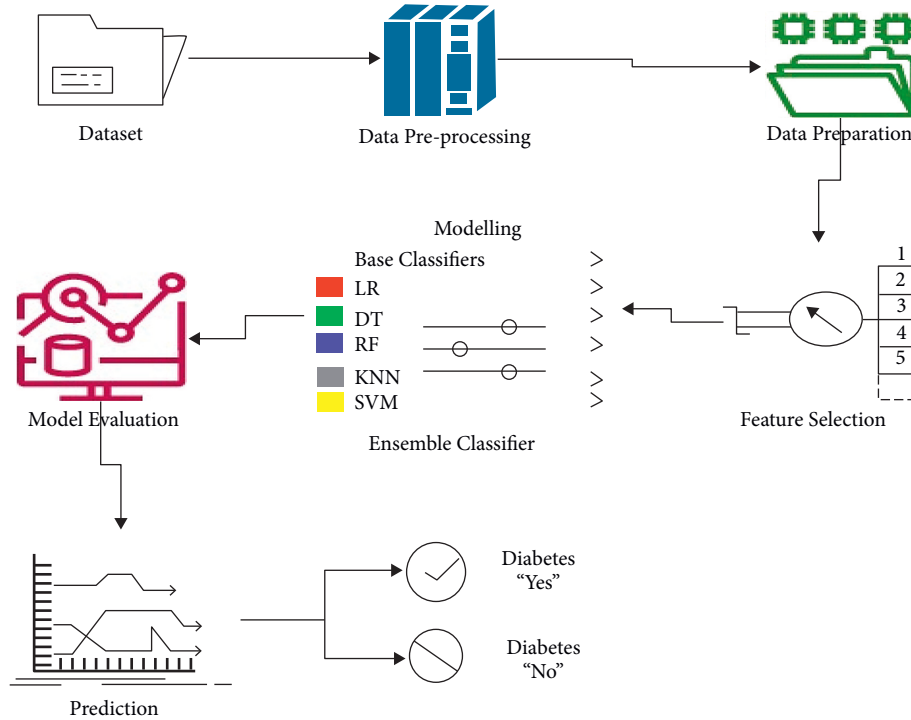
FIGURE 4: Different phases of the prediction process.

TABLE 1: Feature descriptions.

Attribute/feature	Description
Age	Between 20 years and 90 years
Gender	1 denotes male, 0 denotes female
Polyuria	1 denotes male, 0 denotes female
Polydipsia	1 denotes male, 0 denotes female
Sudden weight loss	1 denotes male, 0 denotes female
Weakness	1 denotes male, 0 denotes female
Polyphagia	1 denotes male, 0 denotes female
Genital thrush	1 denotes male, 0 denotes female
Visual blurring	1 denotes male, 0 denotes female
Itching	1 denotes male, 0 denotes female
Irritability	1 denotes male, 0 denotes female
Delayed healing	1 denotes male, 0 denotes female
Partial paresis	1 denotes male, 0 denotes female
Muscle stiffness	1 denotes male, 0 denotes female
Alopecia	1 denotes male, 0 denotes female
Obesity	1 denotes male, 0 denotes female
Class	1 denotes positive, 0 denotes negative

FIGURE 5: A single DT.

P (nodes|parents (node)) for all nodes

$$P(X_1, \ldots\ldots, X_n) = \prod_{i=1}^{n} P(X_i | X_1, \ldots\ldots, X_{i-1})$$

$$= \prod_{i=1}^{n} P(X_i | \text{Parents}(X_{1i})).$$

(5)

When huge networks are used, this combined distribution will reduce computation. The use of a Bayesian network is motivated by the fact that they record high accuracy for complicated and uncertain domains.

variables. The probability function is assigned to each node. Bayesian networks are used to represent knowledge in an area that is uncertain. Statistics, probability theory, graph theory, and computer science are all combined in Bayesian Networks [22].

BN depicts a joint probability distribution over a set of random variables "X." The BN is explained as $Y = \langle G, \theta \rangle$ where G is DAG. It contains the nodes $N_1, N_2, \ldots, N_n$, and edges depict dependencies among the variables. Joint distribution for a BN is defined as shown in equation (5).
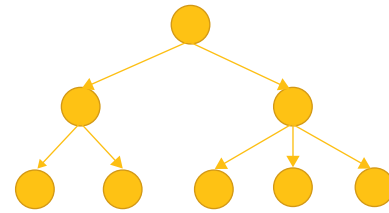
3.6.2. Radial Basis Function (RBF). The supervised algorithm RBF is derived from function approximation. RBFs have three layers: an I/P layer, an O/P layer, and a hidden layer [23]. Each hidden unit in the input space has its own receptive field. These units are known as radial centers, and they are denoted by the letters $V_1, V_2, \ldots, V_n$. The transformation from hidden units to output is linear but not from the input layer to the hidden unit. RBF will create a local
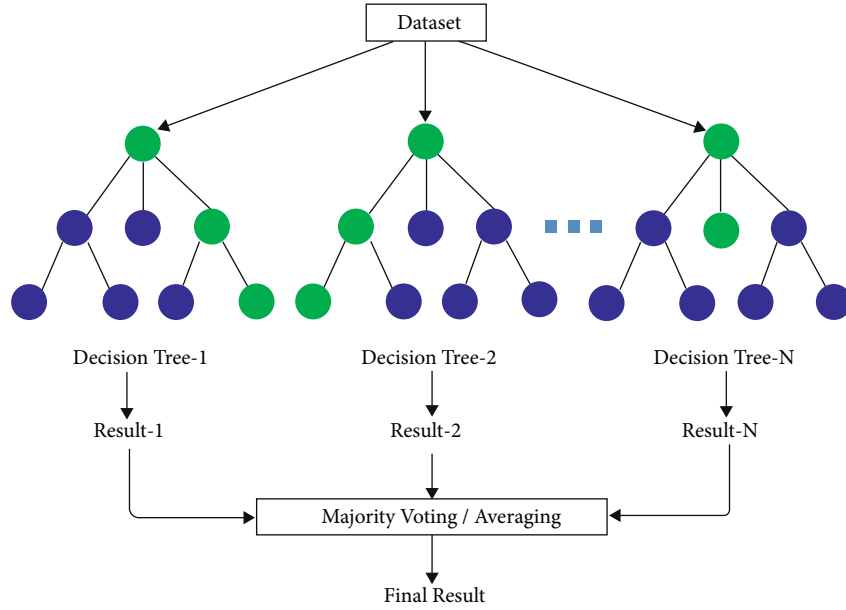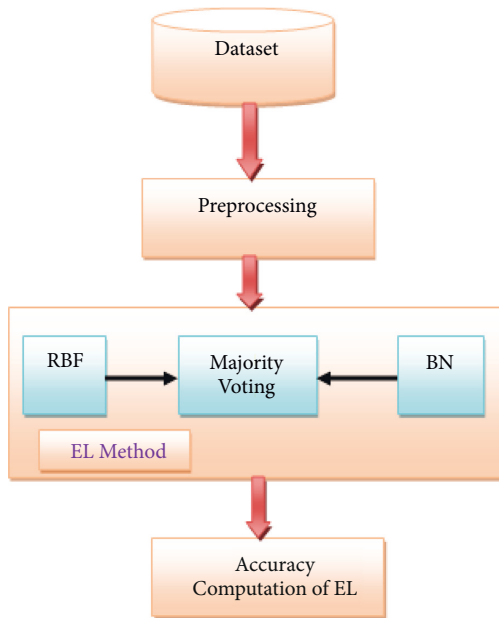
FIGURE 6: RF example.

FIGURE 7: Proposed process diagram.

map, as a result of which it will learn quickly. RBF employs two-step learning, requiring the acquisition of both weight and centers. By calculating the input's similarity to instances from the training set, RBF achieves supervised learning. Each neuron computes the distance between the input and its prototype to classify a fresh sample. The Gaussian equation is shown as follows:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\left((x-u)^2/2\sigma^2\right)}, \qquad (6)$$

where x is the input, u is the mean, and $\sigma\sigma$ is the standard deviation.

The neuron activation function of RBF is defined as shown in the following equation:

$$\varnothing(x) = e^{-\beta\|x-\mu\|^2}. \qquad (7)$$

The RBF is motivated by the fact that it learns faster than a simple feed-forward network because of the following advantages: training is quick, MLP output nodes implement linear summation, they are good at interpolation, and radial basis functions are implemented by hidden nodes.

3.6.3. Ensemble Learning (EL) Method. Ensemble learning is an effective machine learning technique for improving model performance. The model's prediction capability is improved by combining the varied sets of learners (Base Learners). The importance of selecting the appropriate ensemble for diabetes prediction cannot be overstated.

RBF and Bayesian Network have been employed as base learners in the proposed methodology, which was then integrated with the EL method. Majority voting can be represented as shown in the following equation:

$$\sum_{t=1}^{T} d_{t,J} = \frac{C}{\max_{j=1}} \sum_{t=1}^{T} d_{t,j}, \qquad (8)$$

where T is the dataset and C is the class label.

4. Results and Discussion

A confusion matrix is a matrix that comprises data on actual and expected classifications and is used to assess the algorithm's performance using the matrix data. We get both correct and incorrect predictions from each classifier. False Positive and False Negative predictions are two types of incorrect predictions. When we forecast

```
 (i) Input required: Diabetes Dataset
(ii) Output expected: Prediction of Ensemble Technique
     Step_1: Preprocessing on the Pima dataset of diabetes
     Step_2: Separate the dataset as test and training data
     Step_3: Construct EL (Majority voting) using BN and RBF with k-Fold-cross-Validation (k = 10)
     Step_4: Let us call each new sample "S" Test data
     Step_5: EL (test sample) = majority voting (class (RBF), class (BN))
     Step_6: Keep track of the suggested approach's accuracy and evaluate it using a variety of measures.
```

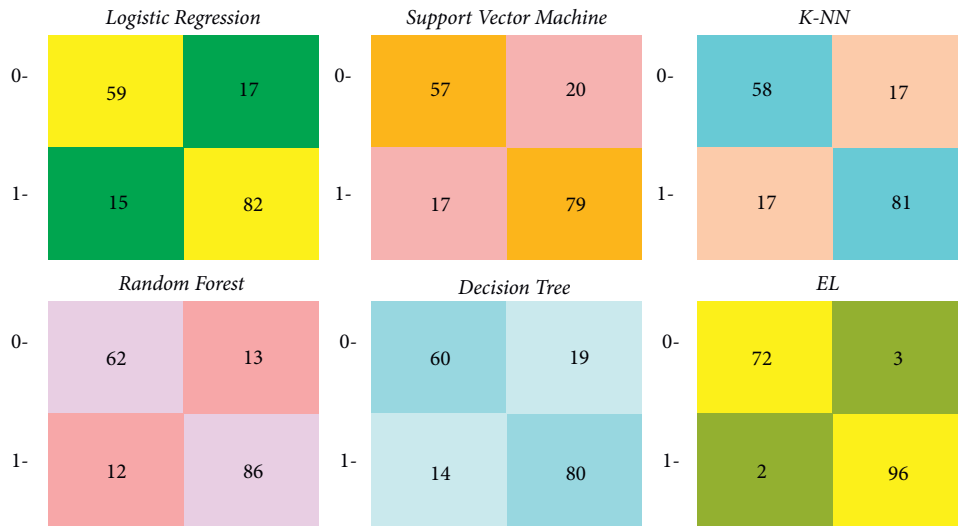Algorithm 1: Stages in the proposed ensemble learning (EL) model.

Figure 8: Confusion matrix.

Table 2: Performance metrics of different ML techniques in %.

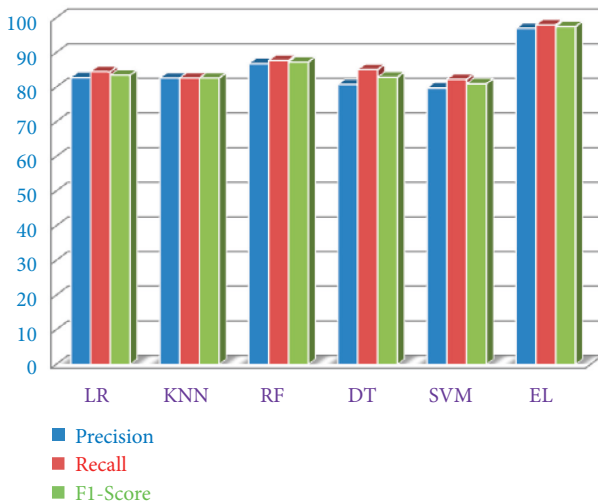| ML techniques	Precision	Recall	F1-score
LR	82.82	84.53	83.51
KNN	82.65	82.65	82.65
RF	86.86	87.75	87.30
DT	80.80	85.10	82.89
SVM	79.79	82.29	81.02
EL (BN + RBF)	**96.96**	**97.96**	**97.46**

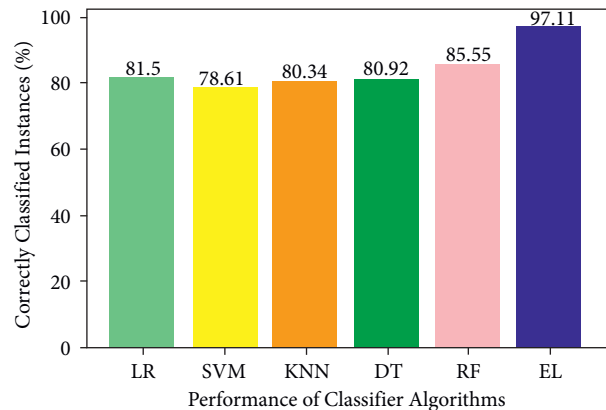Figure 9: Performance of various ML techniques.
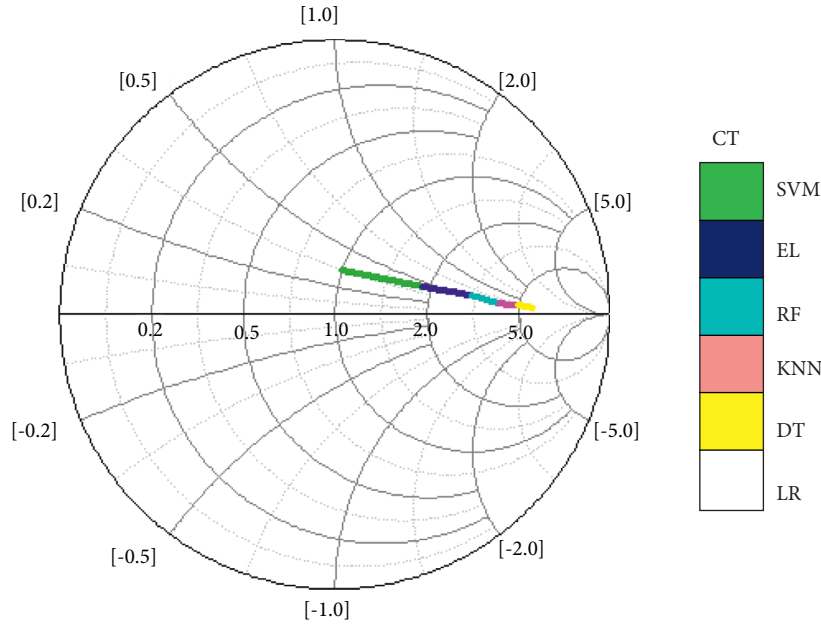

Figure 10: Accuracy of algorithms.

FIGURE 11: Computation analyses of ML techniques.

that something will happen/occur but it does not, we call it a False Positive also called as type I error (a valid null hypothesis is rejected). For example, we predicted that diabetes would develop, but it did not. When we forecast that something will not happen/occur, but it does, we call it a False Negative also called as type II error (the failure to reject a false null hypothesis). For example, we might anticipate that there will be no diabetes, but diabetes does exist. Type I errors are typically thought to be less dangerous than type II faults. In disciplines, such as medicine, both errors may appear to be fatal. The confusion matrix of all the ML algorithms is depicted in Figure 8.

The entire aforementioned models have been evaluated against precision, F1-measure, and recall, and the values obtained are depicted in Table 2.

The precision is computed using the following equation:

$$precision = \frac{TPs}{TPs + FPs}. \quad (9)$$

The precision of LR, KNN, RF, DT, SVM, and EL classifiers is 82.82%, 82.65%, 86.86%, 80.80%, 79.79%, and 96.96%, respectively.

Recall is the proportion of TPs out of TPs and FNs. The recall is computed using the following equation:

$$recall = \frac{TPs}{TPs + FNs}. \quad (10)$$

The recall of LR, KNN, RF, DT, SVM, and EL classifiers is 84.53%, 82.65%, 87.75%, 85.10%, 82.29%, and 97.96%, respectively.

The F1-score which is also called as harmonic mean that considers both precision as well as recall is computed using the following equation:

$$F1 - score = 2 * \frac{precision * recall}{precision + recall}. \quad (11)$$

The F1-score of LR, KNN, RF, DT, SVM, and EL classifiers is 83.51%, 82.65%, 87.30%, 82.89%, 81.02%, and 97.46%, respectively. The recall, precision, and F1-scores of all the techniques are shown in Figure 9.

The number of correctly classified data instances divided by the total number of data instances is known as accuracy. The accuracy is computed using the following equation:

$$accuracy = \frac{TNs + TPs}{TNs + TPs + FPs + FNs}. \quad (12)$$

The accuracy of the different ML algorithms is also computed. The accuracy of LR, SVM, KNN, DT, RF, and EL classifiers is 81.50%, 78.61%, 80.34%, 80.92%, 85.55%, and 97.11%, respectively. The EL method has outperformed with an accuracy of 97.11% in comparison to other ML algorithms. Figure 10 shows the accuracy performance of the aforementioned ML algorithms.

Figure 11 illustrates the processing time required for several classification techniques. In comparison to other methods, EL techniques take longer to process data due to the blending process of base learners. CG (conjugate gradient) optimization was utilized for primary LR with compliance. L2 regularization and Alpha = 0.0001 were being used by SVM to get the best learning rate. Using Gini impurity, a DT was divided into smaller trees, which consumed optimal time for processing. It was enlarged until all leaves were pure or included fewer than two nodes, whichever came first.

5. Conclusion

If identified early enough, diabetes has the potential to save thousands of lives. The main purpose of this article is to

summarize all of the recent and ongoing investigations using machine learning algorithms for diabetic illness prediction. In addition, this study provides all of the relevant and required information for researchers who are new to deep learning and wish to examine ML methods. In conclusion, EL demonstrates its power in terms of efficacy and efficiency, as well as precision, recall, accuracy, and $F1$-score. The blended EL method provides the best accuracy of 97.11% compared to other models. More study is needed in this field to increase the performance of categorization systems so that they can forecast a larger number of variables. The goal is to figure out how to parameterize our categorization algorithms in order to attain high accuracy. A number of datasets are continuously being investigated to explore how ML techniques might be used to better predict diabetes.

Data Availability

The Irvine ML Repository data used to support the findings of this study are available at https://archive.ics.uci.edu/ml/datasets/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] I. Rish, "An empirical study of the naive Bayes classifier," *IJCAI*, vol. 3, 2001.

[2] T. R. Mahesh, V. Vinoth Kumar, and V. Vivek, "Early predictive model for breast cancer classification using blended ensemble learning," *International Journal of System Assurance Engineering and Managament*, 2022.

[3] U. Fayyad, G. P. Piatetsky-Shapiro, and P. Smyth, "The KDD process for extracting useful knowledge from volumes of data," *Communications of the ACM*, vol. 39, no. 11, pp. 27–34, 1996.

[4] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, no. 3, pp. 37–54, 1996.

[5] W. S. Mcculloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biology*, vol. 52, no. 1-2, pp. 99–115, 1990.

[6] M. Abadi, P. Barham, J. Chen et al., "Tensorflow:A system for large-scale machine learning," *OSDI*, vol. 16, pp. 265–283, 2016.

[7] U. R. Acharya, K. S. Vidya, D. N. Ghista, W. J. E. Lim, F. Molinari, and M. Sankaranarayanan, "Computer-aided diagnosis of di-abetic subjects by heart rate variability signals using discretewavelet transform method," *Knowledge-Based Systems*, vol. 81, pp. 56–64, 2015.

[8] X. X. Chen, H. Tang, W. C. Li, H. Wu, W. Chen, and H. Ding, H Lin, Identification of bacterial cell wall lyases via pseudo amino acid composition," *BioMed Research International*, vol. 20168 pages, 2016.

[9] S. Habibi, M. Ahmadi, and S. Alizadeh, "Type 2 diabetes mellitus screening and risk factors using decision tree: results of data mining," *Global Journal of Health Science*, vol. 7, no. 5, pp. 304–310, 2015.

[10] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," *Computational and Structural Biotechnology Journal*, vol. 15, pp. 104–116, 2017.

[11] B. J. Lee and J. Y. Kim, "Identification of type 2 diabetes risk factors using phenotypes consisting of anthropometry and triglycerides based on machine learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 1, pp. 39–46, 2016.

[12] G. S. Gowramma, T. R. Mahesh, and G. Gowda, "An automatic system for IVF data classification by utilizing multilayer perceptron algorithm," *ICCTEST*, vol. 2, pp. 667–672, 2017.

[13] K. Sharma, T. R Mahesh, and J Bhuvana, "Big data technology for developing learning resources," *Journal of Physics: Conference Series*, vol. 1979, no. 1, Article ID 012019, 2021.

[14] A. A. Deshmukh, M. Dubal, T. R Mahesh, and C. R. Chauhan, "Data security analysis and security extension for smart cards using java card," *International Journal of Advanced Information Technology (IJAIT)*, vol. 2, no. 2, 2012.

[15] T. R. Mahesh, S. Prabhanjan, and M. Vinayababu, "Noise reduction by using fuzzy image filtering," *Journal of Theoretical and Applied Information Technology*, vol. 15, 2010.

[16] H. K. Shashikala, T. R. Mahesh, V. Vivek, M. G. Sindhu, C. Baig, and T. Z. Baig, "Early detection of spondylosis using point-based image processing techniques," in *Proceedings of the 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*, pp. 655–659, Bangalore, India, August 2021.

[17] M. J. Dubai, T. R. Ghosh, and P. A. Ghosh, "Design of new security algorithm: using hybrid Cryptography architecture," in *Proceedings of the 2011 3rd International Conference on Electronics Computer Technology*, pp. 99–101, Kanyakumari, India, April 2011.

[18] S. Perveen, M. Shahbaz, A. Guergachi, and K. Keshavjee, "Performance analysis of data mining classification techniques to predict diabetes," *Procedia Computer Science*, vol. 82, pp. 115–121, 2016.

[19] S. Mahesh, T. R. Mahesh, and M. Vinayababu, "Using data mining techniques for detecting terror-related activities on the web," *Journal of Theoretical and Applied Information Technology*, vol. 16, 2010.

[20] P. Ghosh and T. R. Mahesh, "A privacy preserving mutual authentication protocol for RFID based automated toll collection system," in *Proceedings of the 2016 International Conference on ICT in Business Industry & Government (ICTBIG)*, pp. 1–5, Indore, India, November 2016.

[21] P. K. Shukla, J. Kaur Sandhu, A. Ahirwar, D. Ghai, P. Maheshwary, and P. K. Shukla, "Multiobjective genetic algorithm and convolutional neural network based COVID-19 identification in chest X-ray images," *Mathematical Problems in Engineering*, vol. 2021, Article ID 7804540, 2021.

[22] G. Pinaki and T. R. Mahesh, "Smart city: concept and challenges," *International Journal on Advances in Engineering Technology and Science*, vol. 1, p. 1, 2015.

[23] T. R. Mahesh, V. Dhilip Kumar, V. Vinoth Kumar et al., "AdaBoost ensemble methods using K-fold cross validation for survivability with the early detection of heart disease," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–11, Article ID 9005278, 2022.