INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 9, Issue 1 - V9I1-1171)
Available online at: https://www.ijariit.com

A fused machine learning technique for diabetes prediction.

Manyam Kishore Kumar Reddy
manyamkishore1121@gmail.com
Annamacharya Institue of Technology and Science, Rajampet, Andhra Pradesh

Gandlapati Bhavani
bhavanireddygandlapati@gmail.com
Annamacharya Institue of Technology and Science, Rajampet, Andhra Pradesh

Mulasthanam Venkata Moulika
moulikam22@gmail.com
Annamacharya Institue of Technology and Science, Rajampet, Andhra Pradesh

Addepalli Tharun Raju
tharunrockstar3@gmail.com
Annamacharya Institue of Technology and Science, Rajampet, Andhra Pradesh

Y Mahanandi
MAHANANDI0548@gmail.com
Annamacharya Institue of Technology and Science, Rajampet, Andhra Pradesh

ABSTRACT

Early disease diagnosis and prevention are crucial in the medical field. One of the world's most hazardous diseases is diabetes. Sugar and fat are commonly found in modern lifestyles. In our eating behaviours, which has elevated the risk of diabetes. It is crucial to comprehend the disease's signs in order to predict it. Machine-learning (ML) techniques are useful at the moment for disease identification. The model for predicting diabetes in this paper uses a fused machine learning technique. Support Vector Machine (SVM) and Artificial Neural Network (ANN) models make up the two different sorts of models that make up the conceptual framework. The dataset is examined by these models to assess if a diabetes diagnosis is accurate or not. The training data and testing data ratios for the dataset employed in this study are, respectively, 70:30. While the fuzzy logic ultimately decides whether a diabetes diagnosis is positive or negative, the output of these models serves as the input membership function for the fuzzy model. The fused models are saved in a cloud storage system for further usage. The fused model makes a prediction about the patient's diabetes status based on the patient's current medical record. The suggested fused ML model outperforms the previously published approaches with a prediction accuracy of 94.87.

Keywords: *Logistic Regression , SVM , ANN ,ML techniques.*

1. INTRODUCTION

One of the most widespread and chronic metabolic diseases in the world is diabetes. Diabetes comes in Type-1 and Type-2 varieties. Type-1 diabetes develops inside the body when the immune system harms pancreatic Beta cells (-cells), which results in the release of either very little or no insulin. An autoimmune condition known as type 2 diabetes occurs when the pancreas cells do not make enough insulin or the body's cells do not respond to insulin, causing blood glucose levels to become out of control.

Type-1 diabetes is a condition marked by an insufficiency of the hormone insulin, which raises blood glucose levels and weakens the metabolism of proteins, carbs, and lipids. Polyuria, Polydipsia, and Weakness are among the diabetes symptoms (iv) (v) Obesity and Polyphagia (vi) Sudden-Weight-Loss (vii) (vii) Genital-Thrush (viii) (viii) Itching (x), irritability (ix), and visual blurriness (xi) Delayed-Healing \s(xii) Partial-Paresis (xiii) (xiii) Muscle-Stiffness (xiv) (xiv) Alopecia, \setc. [1] . . Due to several health consequences, diabetes, a metabolic illness, is responsible for millions of deaths worldwide each year. Between 2000 and 2019 worldwide, there was a 70% increase in the fatality rate due to diabetes [2].

To identify these lethal diseases, a sophisticated ML-based diagnostic system is needed.Patients with diabetes can be successfully diagnosed at an early stage using an ML-based expert decision system. For the purpose of predicting diabetes, researchers used a variety of different datasets. An adequate dataset with the required features for training and validation is required for ML-based frameworks. The dataset's selection of pertinent and important features improves the ML model's capacity for precise prediction. The dataset utilized in the suggested system was assembled by the hospital in Sylhet, Bangladesh and is available in the (University of California Irvine) UCI Machine Learning repository [3].

When food is not properly absorbed by the body, it results in diabetic mellitus (DM), which changes the body's glucose levels. Healthy eating and a change in lifestyle are two diabetes prevention strategies against malnutrition or obesity, which are occasionally the disease's main causes. Additionally, by lowering the risk of health issues and controlling blood pressure, these steps aid.

Diabetes is easier to diagnose with a medical examination. The disease is also found using some laboratory techniques. For as long as they are alive, people with type-2 diabetes require life-saving insulin.

As a result, if left untreated, this ill condition depletes resources for people, families, and the country as a whole. For prediabetic individuals to live a healthy life and be in good health, early identification and symptomatic treatment are crucial. It will be possible to identify diseases and prevent their occurrence with the use of an intelligent medical diagnosis system based on symptoms, signs, laboratory tests, and observations. Intriguing applications of artificial intelligence (AI) for disease identification have also been made in medical diagnosis systems. This study suggests a framework for combining machine learning and traditional methods to identify diabetic patients early.

2. LITERATURE SURVEY

The use of machine learning approaches to identify diabetic patients based on their symptoms has generated a large amount of study in recent publications. Using different algorithms for diagnosing diabetes mellitus to increase accuracy rate, Pradhan et al. [4] utilized supervised learning, hybrid learning, or ensemble learning. However, the ensemble technique outperforms the other two approaches. By using a soft voting classifier on the Pima-Diabetes dataset and the Breast-Cancer dataset in an ensemble technique, Kumari et al. [5] increased classification accuracy. In comparison to other machine learning techniques, the results show that the soft voting classifier has an accuracy rate of 79.08%.

Utilizing the Pima Diabetes dataset, Sarwar et almachine.'s learning techniques were employed to identify diabetes early on. Compared to the other four algorithms, their accuracy rates from KNN and SVM were greater at 77%.

The dataset's size and the missing values are a drawback for this study. On the Pima dataset, SVM, KNN, Naive Bayes, and ANN with Min-Max scaling (MMS) were utilised by Dey et al. [7]. In comparison to the other four methods, the model ANN with MMS has a greater accuracy of 82.35%. In order to predict diabetes, the researchers in [8] employed the Weka tool together with machine-learning algorithms like Naive Bayes, Random Forest, and Simple CART.

In comparison to the other three algorithms, the SVM classifier performs better and attained an accuracy of 79.13%. A model built on Logistic Regression, SVM, Decision Tree, and KNN was used by Saru et al. [9] to predict diabetes. Additionally, they contrasted their accuracy rates between Bootstrapping and without it. The decision tree with bootstrapping outperforms the other two methods in terms of accuracy with a rate of 94.4%.

Compared to the other two methods, the decision tree's accuracy rate is greater at 85%. The Deep Neural Network (DNN), Logistic Regression, Decision Trees, and Naive Bayes ML methods were used by Wei et al. [11] to create a model in their article. DNN outperforms the other four algorithms in terms of accuracy with a rate of 77.86%.

Support Vector Machine (SVM), C4.5 Decision Tree, K-Nearest Neighbor (KNN), and Naive Bayes are four machine learning (ML) methods that Faruque et al. [12] suggested as a model to predict diabetes. Compared to the other three algorithms, the accuracy rate of the C4.5 Decision Tree is greater at 73.5%.

Diabetes was predicted by Jain et al. [13] using a variety of machine learning techniques, including Neural Network (NN), Fisher Linear Discriminant Analysis (FLDA), Random Forest, Chi-square Automatic Interaction Detection (CHAID), and SVM. In comparison to the other four algorithms, NN has a higher accuracy rate of 87.88%.

Since they often concentrated on pre-processing methods, data balancing, and different kinds of supervised and semi-supervised learning models, previous research models are less accurate than the ML algorithms that are currently effective for illness identification. Therefore, it is necessary to develop new techniques with decision level fusion that could combine the high illness detection accuracy of various machine learning algorithms with the accuracy of numerous machine learning algorithms. The employment of two supervised machine-learning techniques, ANN and SVM [14–16], along with fuzzy logic for decision level fusion, is proposed in a fused ML model for this purpose.

3. MATERIALS AND METHODS

A Fused Model for Diabetes Prediction is suggested in this article (FMDP). The FMDP model that has been suggested contains two key phases. Training Layer makes up the first phase, and Testing Layer makes up the second. Data collection, preprocessing, categorization, performance assessment, and machine learning fusion are some of the stages that make up the Training Layer. The UCI Machine Learning Repository's dataset was used in this study [3].

In the data acquisition stage, diabetes can be predicted using a dataset with sufficient features. The preprocessing stage involves cleaning, normalizing, and dividing the data into a training and test dataset.

Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) can be trained for prediction using preprocessed data. To attain the necessary accuracy, we can choose from a variety of machine-learning methods for the categorization. However, we solely used SVMs and ANNs, two popular machine learning (ML) techniques, in the suggested model [14], [16], and [19]. Following some preliminary experiments, where we discovered that these strategies were more successful for this challenge, these algorithms were chosen for this study. In the Performance Evaluation step, we used a variety of accuracy metrics, including accuracy, specificity, sensitivity, precision, and F1 score. The proposed model will be retrained if it does not satisfy the learning requirements. The ANN and SVM outputs are used as inputs in machine learning fusion when learning conditions are satisfied.

The testing layer represents the second phase of the proposed structure. The testing layer downloads training model from the cloud after it has been preprocessed and loads data from medical databases. If a diabetes diagnosis is positive or negative, it is predicted using a fused model. Comparing the expected and actual outputs allows for the calculation of prediction accuracy.

In order to train the ANN model, the training dataset is first processed. In accordance with the class base split, we divided the preprocessed data into training and test data in a 70:30 ratio. With 5% utilized for testing, 5% for validation, and the remaining 90% for training, we employed the Bayesian regularization function on the data for this exercise.

$$\text{out}\vartheta = \frac{1}{1+e^{-(h1\sum_{r=1}^{m}(u_{r,\vartheta}*\omega))}} \qquad (1)$$

where, $\vartheta = 1, 2, \ldots, n$

$$\text{out}\varrho = \frac{1}{1+e^{-(h2\sum_{\vartheta=1}^{n}(p_{\vartheta,\varrho}*\text{out}\vartheta))}} \qquad (2)$$

Each output neuron's error can be calculated and added using the squared error function to determine the overall error (E).

$$E = \frac{1}{2}\sum_{\varrho}(\tau_{\varrho} - out_{\varrho})^2 \qquad (3)$$

Using formula, weights can be adjusted in accordance with error.

$$\Delta\omega \propto -\frac{\partial E}{\partial\omega} \qquad (4)$$

A hidden layer and an output layer's weight are updated by Equation 5.

$$\Delta p_{\vartheta,\varrho} = -\varepsilon\frac{\partial E}{\partial v_{\vartheta,\varrho}} \qquad (5)$$

$$\Delta p_{\vartheta,\varrho} = -\varepsilon\frac{\partial E}{\partial out_{\varrho}} \times \frac{\partial out_{\varrho}}{\partial net_{\varrho}} \times \frac{\partial net_{\varrho}}{\partial p_{\vartheta,\varrho}} \qquad (6)$$

$$\Delta p_{\vartheta,\varrho} = \varepsilon(\tau_{\varrho} - \text{out}_{\varrho}) \times \text{out}_{\varrho}(1 - \text{out}_{\varrho})(\text{out}\vartheta) \qquad (7)$$

The weights between hidden-layer neurons and input-layer neurons are changed according to equations 8 and 9.

$$\Delta\tilde{v}_{i,\vartheta} \propto -\left[\sum_{\varrho}\frac{\partial E}{\partial \text{out}_{\varrho}} \times \frac{\partial \text{out}_{\varrho}}{\partial \text{net}_{\varrho}} \times \frac{\partial \text{net}_{\varrho}}{\partial \text{out}_{\vartheta}}\right] \times \left[\frac{\partial \text{out}_{\vartheta}}{\partial \text{net}_{\vartheta}} \times \frac{\partial \text{net}_{\vartheta}}{\partial \tilde{v}_{i,\vartheta}}\right] \qquad (8)$$

$$\Delta\tilde{v}_{i,\vartheta} = \xi\left[\sum_{\varrho}(\tau_{\varrho} - out_{\varrho}) \times out_{\varrho}(1 - out_{\varrho})\right.$$

$$\Delta\tilde{v}_{i,\vartheta} = \xi\left[\sum_{\varrho}(\tau_{\varrho} - out_{\varrho}) \times out_{\varrho}(1 - out_{\varrho}) \times p_{i,\vartheta}\right] \times out_{\varrho}(1 - out_{\varrho}) \times \omega_{\tilde{i}} \qquad (9)$$

Equation 10 provides information on the weights updating formula between neurons in the hidden and output layers.

$$\Delta\tilde{v}_{i,\vartheta}(t+1) = \tilde{v}_{i,\vartheta}(,t) + \lambda\,\Delta\tilde{v}_{i,\vartheta} \qquad (10)$$

30 percent of the remaining datasets should be used for validation after the training model has been successfully created. Result output from the validation is displayed when results are saved.

The prediction is determined to be 92.31% accurate when data is compared to the actual output.

Data is classified using classes by SVM, which creates a hyperplane. Positive and Negative symptoms of diabetes are classified by SVM [15], [17], and [18]. A line must first be drawn to divide classes in a hyperplane. Equation.11 allows the line equation to be represented.

$$\dot{x}_2 = a\dot{x}_1 + b \qquad (11)$$

Where a indicates the slope and b represents an intersecting point. Hence, it is written as follows:

$$a\dot{x}_1 - \dot{x}_2 + b = 0 \qquad (12)$$

If = (x˙1, x˙2) T & = (, −1), then Using the above expression, we can formulate an Equation. 13

$$\dddot{\omega}.\dddot{x} + b = 0 \qquad (13)$$

Three-dimensional vector analysis can be done using the hyperplane equation.

$$\dddot{\omega} = \frac{\dot{x}_1}{\vdots\dot{x}\vdots} + \frac{\dot{x}_2}{\vdots\dot{x}\vdots} \qquad (14)$$

N-dimensional vectors can be expressed in Equation 15.

$$\dddot{\omega}.\dddot{x} = \sum_{i=1}^{n} \dot{\omega}_i \dot{x}_i \qquad (15)$$

To determine if the data has been successfully categorized, use Equation 15

$$Đ_i = \ddot{y}_i \left(\dddot{\omega}.\dddot{x} + b\right)$$

Dataset's functional margins, or d, are stated as

$$\dot{d} = \min_{i=1\ldots m} Đ_i$$

The hyperplane that will be the best hyper plane for the Lagrangian function is provided by the dataset's Geometric-Margin d.

$$Y(\dot{\omega}, b, \mathfrak{B}) = \frac{1}{2}\dot{\omega}\cdot\dot{\omega} - \sum_{i=1}^{m} \mathfrak{B}_i \left[\ddot{y}\left(\dot{\omega}.\dot{x} + b\right) - 1\right] \qquad (16)$$

$$\nabla_{\dot{\omega}} Y(\dot{\omega}, b, \mathfrak{B}) = \dot{\omega} - \sum_{i=1}^{m} \mathfrak{B}_i y_i \dot{x}_i = 0 \qquad (17)$$

$$\nabla_b Y(\dot{\omega}, b, \mathfrak{B}) = -\sum_{i=1}^{m} \mathfrak{B}_i y_i = 0 \qquad (18)$$

It can be written as follows after being simplified:

$$\dot{\omega} = \sum_{i=1}^{m} \mathfrak{B}_i y_i \dot{x}_i \quad \& \quad \sum_{i=1}^{m} \mathfrak{B}_i y_i = 0 \qquad (19)$$

$$\dot{\omega}(\mathfrak{B}, b) = \sum_{i=1}^{m} \mathfrak{B}_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \mathfrak{B}_i \mathfrak{B}_j y_i y_j \dot{x}_i \dot{x}_j$$

Therefore, the aforementioned equation can also be defined using Equation.20.

$$\max_{\mathfrak{B}} \sum_{i=1}^{m} \mathfrak{B}_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \mathfrak{B}_i \mathfrak{B}_j y_i y_j \dot{x}_i \dot{x}_j \qquad (20)$$

Use the KKT (Karush-Kuhn-Tucker) condition to prevent confinement inequalities when using the Lagrangian multiplier method.

$$\mathfrak{B}_i \left[y_i \left(\dot{\omega}_i \cdot \dot{x}^* + b\right) - 1\right] = 0 \qquad (21)$$

$$\left[y_i \left(\dot{\omega}_i \cdot \dot{x}^* + b\right) - 1\right] = 0 \qquad (22)$$

Support vectors also refer to the points closest to the hyperplane. on the basis of Equation 23

$$\dot{\omega} - \sum_{i=1}^{m} \mathfrak{B}_i y_i \dot{x}_i = 0 \qquad (23)$$

It can be expressed as follows:

$$\dot{\omega} = \sum_{i=1}^{m} \mathfrak{B}_i y_i \dot{x}_i \qquad (24)$$

$$y_i[(\dot{\omega}_i \cdot \dot{x}^* + b) - 1] = 0 \qquad (25)$$

$$y_i^2[(\dot{\omega}_i \cdot \dot{x}^* + b) - 1] = 0 \qquad (26)$$

$$b = y_i - [\dot{\omega}_i \cdot \dot{x}^*] \qquad (27)$$

$$b = \{\frac{1}{S} \sum_{i=1}^{S} (y_i - [\dot{\omega}. \dot{x}]) \qquad (28)$$

The number of support vectors S and predictions made using the hyperplane are determined by equation 27.Positive diabetes is represented by points above the hyperplane, or +1, and negative diabetes is represented by points below the hyperplane, or 1. We employed the same data set using both ANN and SVM. All of the SVM's available parameters are used to train the data in Matlab R2020a, with each parameter being optimized. In the five-fold cross-validation procedure, data are divided into five levels and validated in accordance with each level.

Membership-based logic makes use of these. Input variables for the fuzzy system come from SVM and ANN outputs. The set of guidelines that apply to inputs and outputs are defined by membership functions. If a patient's symptoms are consistent with a diabetes diagnosis, ANN and SVM can identify this. Following are some possible descriptions of a mathematically fuzzy base decision:

The four rule sets that result from each model's two potential outcomes are listed below.
Diabetes is Positive if both the SVM and ANN models yield Positive results of 0. (0).
Diabetes is Positive if the results of the ANN model and the SVM model are both Positive (0) or Negative (1). (0).
Diabetes is Negative if the results of the ANN model and the SVM model are both Negative (Negative) (1). (1).
If both the SVM and ANN model results are negative (1), diabetes is considered negative (1).

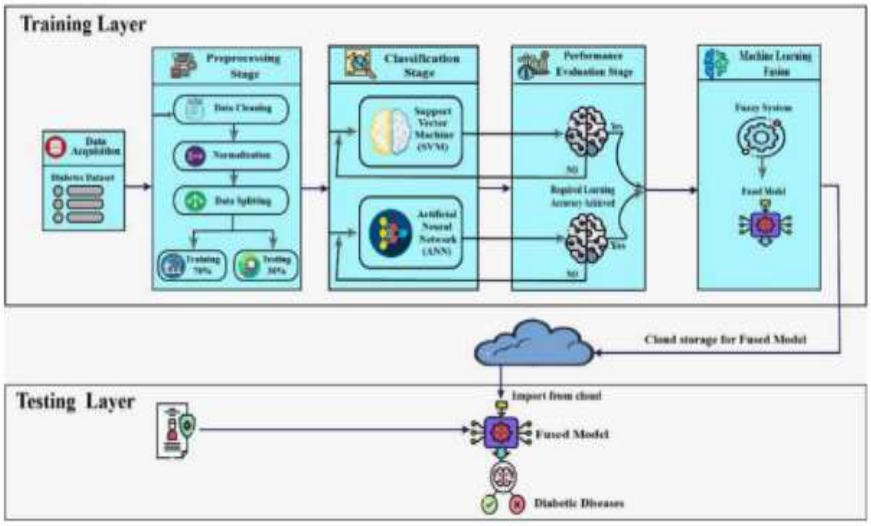

FIGURE 1. Proposed fused model for diabetes prediction (FMDP).

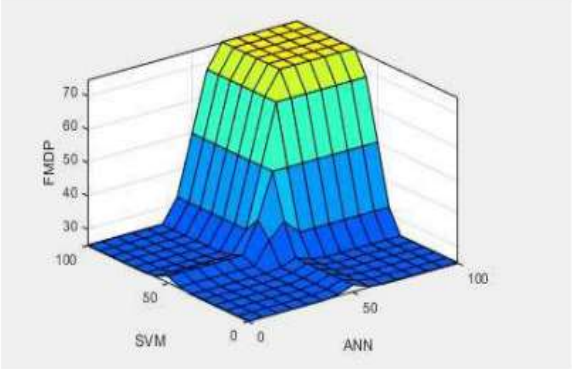

FIGURE 2. Proposed FMDP system rule surface.

4. RESULTS AND DISCUSSION

Utilizing a dataset [3] with 520 total cases and 17 variables based on diabetes symptoms, we were able to put the suggested framework into practice. Seventeen features both separate, with the objective feature being one of them (dependent).The class, which can have either a value of 0 or 1, is the name of the feature that is being targeted. Classifications range from 0 (indicating diabetes symptoms) to 1 (indicating no symptoms of diabetes) (Negative). The dataset's first feature is Age, which includes 51 individuals over the age of 65 and 89 individuals between the ages of 56 and 65.

Additionally, 138 individuals between the ages of 36 and 45, 149 individuals between the ages of 46 and 55, and 93 individuals between the ages of 20 and 35 make up the dataset. A total of 382 men and 192 women make up the second feature, which is sex. "0" represents the male and "1" represents the female reflection. The Polyuria symptom, which contains 258 "yes" values and is represented by "0", and 262 "no" values and is represented by "1", is the third feature.

The Polydipsia symptom, with 287 "no" responses and 233 "yes" responses, is the fourth feature. With 217 "yes" responses set to "0" and 303 "no" responses set to "1," the fifth feature is the symptom of Sudden weight loss. With 215 "no" values set to "1" and 305 "yes" values set to "0," the sixth feature, the Weakness symptom, is the most extreme. With 237 "yes" answers set to "0" and 283 "no" answers set to "1," the seventh feature is the Polyphagia symptom. With 116 "yes" values converted to "0" and 404 "no" values set to "1," the eighth feature is the symptom of genital thrush. The Visual Blurring symptom is the ninth feature, with 287 "no" values set to "1" and 233 "yes" values.

. The Itching symptom is the eighth characteristic, with 253 "yes" values set to "0" and 267 "no" values set to "1". The symptom of irritability is the eleventh feature, with 394 values for "no" set to "1" and 126 values for "yes" set to "0." The Delayed healing symptom, which has 239 "yes" values set to "0" and 281 "no" values set to "1," is the eleventh feature. The partial paresis symptom, which is the twelfth feature, contains 296 values for "no" set to "1" and 224 values for "yes" set to "0." The fourteenth characteristic, the Muscle stiffness symptom, has 325 "no" values assigned to "1" and 195 "yes" values set to "0." The symptom of alopecia is the fifteenth feature, with 341 "no" values set to "1" and 179 "yes" values set to "0." The symptom of obesity, which is the sixteenth feature, has 432 "no" values assigned to "1" and 88 "yes" values set to "0."

In order to conduct the simulations in this study, MATLAB R2020a is employed. When making decisions, fuzzy logic is used as opposed to ANN and SVM for prediction. A 70:30 split is used to separate the dataset into training and testing halves. An ANN is trained using 364 examples, and Table 2 displays the confusion matrix for this training.

According to Table , of the 246 positive instances, 236 were correctly anticipated, while 10 were wrongly forecasted. But out of the 118 negative occurrences, 107 are correctly anticipated, while 11 are wrongly forecasted.

In the testing data, there are 156 occurrences. In Table , the ANN testing confusion matrix is displayed.

The 69 cases of positive diabetes are listed in Table 3; 61 of these cases had accurate predictions, while 8 had inaccurate predictions. Negative diabetes, on the other hand, occurs in 87 cases, 83 of which can be accurately predicted, while the other four cannot.

SVM has been validated five times in total. In Table 4, the SVM training's confusion matrix is displayed.

The 246 instances of positive diabetes are listed in Table 4, 227 of which were correctly predicted, whereas 19 were wrongly predicted. There are 118 cases of negative diabetes, of which 105 are successfully diagnosed, whereas 13 are wrongly anticipated. The 69 cases of positive diabetes are listed in Table 5, 59 of which had accurate predictions whereas the remaining 10 had inaccurate predictions. But there are 87 examples of negative diabetes, and 80 of them had accurate predictions, while 7 had inaccurate predictions.

	SVMs Training	**SVMs** Testing	**ANNs** Training	**ANNs** Testing	**FMDP** Testing
Ǎccuracy	0.9121	0.8910	0.9423	0.9231	0.9487
Miss Rate	0.0879	0.109	0.0577	0.0769	0.0513
Ṡensitivity	0.9458	0.8939	0.9555	0.9385	0.9552
Ṡpecificity	0.8468	0.8889	0.9145	0.9121	0.9438
'Positive 'Prediction Vlaue	0.9228	0.8551	0.9593	0.8841	0.9275
Ṅegative'Prediction Vlaue	0.8898	0.9195	0.9068	0.9540	0.9655
Ḟalse 'Positive Rate	0.1532	0.1111	0.0855	0.0879	0.0562
Ḟalse Ṅegative Rate	0.0542	0.1061	0.0445	0.0615	0.0448
Ḟ1 Ṡcore	0.9342	0.8741	0.9574	0.9104	0.9412

The testing confusion matrix using the suggested fused model is shown in Table 6. It depicts the 69 positive diabetes cases, 64 of which had accurate predictions, while only five had inaccurate predictions. However, there are 87 cases of negative diabetes in all, 84 of which were correctly predicted, and 3 of which were wrongly anticipated.

The numbers 0, 1, 0, and 1 in the formulas below represent, in that order, expected positive results, expected positive output, expected negative output, and expected results.

$$\dot{F}\text{alse } \text{`Positive Rate} = 1 - \dot{S}\text{pecificity}$$

$$\dot{F}\text{alse } \dot{N}\text{egative Rate} = 1 - \dot{S}\text{ensitivity}$$

$$\dot{F}1\dot{S}\text{core} = 2 * \frac{\text{`Positive `Prediction Vlaue} * \dot{S}\text{ensitivity}}{\text{`Positive `Prediction Vlau} + \dot{S}\text{ensitivity}}$$

Utilizing several accuracy measures, as stated previously and shown in Table, the performance of both models (ANN and SVM) as well as the suggested fused model is assessed. 7. It is clear that, in comparison to both models that were utilised, the proposed fused model outperformed them in testing (ANN, SVM).Table includes comparisons between the suggested fused model and earlier published models and methodologies. It is apparent that the suggested fused technique beat all other reported procedures, with an accuracy of 94.87% and a miss rate of 5.13%.

5. CONCLUSION

Despite the fact that a variety of models have been employed to predict diabetes, researchers have always been more concerned with the efficacy of the suggested models.

Consequently, a new model is necessary to increase diabetes prediction accuracy. In this study, a decision level fusion-based machine learning decision support system for diabetes was developed. The fuzzy logic is utilised in the proposed model to combine two commonly used machine learning techniques. In comparison to other current systems, the suggested fuzzy decision system has a superior accuracy score of 94.87. We can save many lives by using this diagnostic model. Additionally, if diabetes is identified early on and preventative steps are implemented, the fatality rate from the condition can be reduced.

6. REFERENCES

[1] F. Islam, R. Ferdousi, S. Rahman, and H. Y. Bushra, Computer Vision and Machine Intelligence in Medical Image Analysis. London, U.K.: Springer,2019.

[2] World Health Organization (WHO). (2020). WHO Reveals Leading Causes of Death and Disability Worldwide: 2000–2019. Accessed: Oct. 22, 2021. [Online]. Available: https://www.who.int/news/item/09-12-2020-who-reveals-leading-causes-of-death-and-disability-worldwide-2000-2019

[3] A. Frank and A. Asuncion. (2010). UCI Machine Learning Repository.Accessed: Oct. 22, 2021.[Online].Available:http://archive.ics.uci.edu/ml

[4] G. Pradhan, R. Pradhan, and B. Khandelwal, ''A study on various machine learning algorithms used for prediction of diabetes mellitus,'' in Soft Computing Techniques and Applications (Advances in Intelligent Systems and Computing), vol. 1248. London, U.K.: Springer, 2021, pp. 553–561,doi: 10.1007/978-981-15-7394-1_50.

[5] S. Kumari, D. Kumar, and M. Mittal, ''An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier,'' Int. J. Cogn. Comput. Eng., vol. 2, pp. 40–46,Jun.2021, doi:10.1016/j.ijcce.2021.01.001.

[6] M. A. Sarwar, N. Kamal, W. Hamid, and M. A. Shah, ''Prediction of diabetes using machine learning algorithms in healthcare,'' in Proc.24th Int. Conf. Autom. Comput. (ICAC), Sep. 2018, pp. 6–7, doi:10.23919/IConAC.2018.8748992.

[7] S. K. Dey, A. Hossain, and M. M. Rahman, ''Implementation of a web application to predict diabetes disease: An approach using machine learning algorithm,'' in Proc. 21st Int. Conf. Comput. Inf. Technol. (ICCIT),Dec. 2018, pp.21–23,doi:10.1109/ICCITECHN.2018.8631968.

[8] A. Mir and S. N. Dhage, ''Diabetes disease prediction using machine learning on big data of healthcare,'' in Proc. 4th Int. Conf. Com-put. Commun. Control Autom. (ICCUBEA), Aug. 2018, pp. 1–6, doi:10.1109/ICCUBEA.2018.8697439.

[9] [9] S. Saru and S. Subashree. Analysis and Prediction of Diabetes Using Machine Learning. Accessed: Oct. 22, 2022. [Online]. Available:https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3368308

[10] P. Sonar and K. JayaMalini, ''Diabetes prediction using different machine learning approaches,'' in Proc. 3rd Int. Conf. Comput. Methodologies Commun. (ICCMC), Mar. 2019, pp. 367–371, doi:10.1109/ICCMC.2019.8819841.

[11] S. Wei, X. Zhao, and C. Miao, ''A comprehensive exploration to themachine learning techniques for diabetes identification,'' in Proc. IEEE 4th World Forum Internet Things (WF-IoT), Feb. 2018, pp. 291–295, doi:10.1109/WF-IoT.2018.8355130.

[12] M. F. Faruque and I. H. Sarker, ''Performance analysis of machine learning techniques to predict diabetes mellitus,'' in Proc. Int. Conf. Electr., Comput. Commun. Eng. (ECCE), Feb. 2019, pp. 7–9, doi:10.1109/ECACE.2019.8679365.

[13] B. Jain, N. Ranawat, P. Chittora, P. Chakrabarti, and S. Poddar, ''A machine learning perspective: To analyze diabetes,'' Mater. Today: Proc., pp. 1–5,Feb. 2021, doi: 10.1016/J.MATPR.2020.12.445.

[14] N. B. Padmavathi, ''Comparative study of kernel SVM and ANN classifiers for brain neoplasm classification,'' in Proc. Int. Conf. Intell. Comput., Instrum. Control Technol. (ICICICT), Jul. 2017, pp. 469–473,doi: 10.1109/ICICICT1.2017.8342608.

[15] J. Liu, J. Feng, and X. Gao, ''Fault diagnosis of rod pumping wells based on support vector machine optimized by improved chicken swarm optimization,'' IEEE Access, vol. 7, pp. 171598–171608, 2019, doi:10.1109/ACCESS.2019.2956221.

[16] P. Chinas, I. Lopez, J. A. Vazquez, R. Osorio, and G. Lefranc, ''SVM and ANN application to multivariate pattern recognition using scatter data,''IEEE Latin Amer. Trans., vol. 13, no. 5, pp. 1633–1639, May 2015, doi: 10.1109/TLA.2015.7112025.

