IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 9, 2021, accepted January 8, 2022, date of publication January 11, 2022, date of current version January 24, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3142097

Prediction of Diabetes Empowered With Fused Machine Learning

USAMA AHMED[1,2], GHASSAN F. ISSA[3], MUHAMMAD ADNAN KHAN[1,4], SHABIB AFTAB[2,5], (Member, IEEE), MUHAMMAD FARHAN KHAN[6], RAED A. T. SAID[7], TAHER M. GHAZAL[3,8], (Member, IEEE), AND MUNIR AHMAD[5], (Member, IEEE)

[1]Riphah School of Computing and Innovation, Faculty of Computing, Riphah International University, Lahore 54000, Pakistan
[2]Department of Computer Science, Virtual University of Pakistan, Lahore 54000, Pakistan
[3]School of Information Technology, Skyline University College, University City, Al Sharjah, United Arab Emirates
[4]Pattern Recognition and Machine Learning Laboratory, Department of Software, Gachon University, Seongnam-si 13557, South Korea
[5]School of Computer Science, National College of Business Administration and Economics, Lahore 54000, Pakistan
[6]Department of Forensic Sciences, University of Health Sciences, Lahore 54000, Pakistan
[7]Faculty of Management, Canadian University Dubai, Dubai, United Arab Emirates
[8]Center for Cyber Security, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor 43600, Malaysia

Corresponding authors: Munir Ahmad (munir@ncbae.edu.pk) and Muhammad Adnan Khan (adnan@gachon.ac.kr)

ABSTRACT In the medical field, it is essential to predict diseases early to prevent them. Diabetes is one of the most dangerous diseases all over the world. In modern lifestyles, sugar and fat are typically present in our dietary habits, which have increased the risk of diabetes. To predict the disease, it is extremely important to understand its symptoms. Currently, machine-learning (ML) algorithms are valuable for disease detection. This article presents a model using a fused machine learning approach for diabetes prediction. The conceptual framework consists of two types of models: Support Vector Machine (SVM) and Artificial Neural Network (ANN) models. These models analyze the dataset to determine whether a diabetes diagnosis is positive or negative. The dataset used in this research is divided into training data and testing data with a ratio of 70:30 respectively. The output of these models becomes the input membership function for the fuzzy model, whereas the fuzzy logic finally determines whether a diabetes diagnosis is positive or negative. A cloud storage system stores the fused models for future use. Based on the patient's real-time medical record, the fused model predicts whether the patient is diabetic or not. The proposed fused ML model has a prediction accuracy of 94.87, which is higher than the previously published methods.

INDEX TERMS Diabetic prediction, fuzzy system, fused machine learning model, diabetic symptoms, disease prediction.

I. INTRODUCTION

Diabetes is one of the world's largest ongoing chronic metabolic disorders. There are two types of diabetes, Type-1, and Type-2. When the immune system damages pancreatic Beta cells (β-cells), Type-1 diabetes transpires inside the body, which leads to the release a tiny amount of insulin or no insulin. Type-2 diabetes is an autoimmune disease in which cells of the body fail to interact with insulin, or the pancreatic cells do not produce enough insulin to regulate blood glucose levels. An insufficient amount of insulin causes the blood glucose levels to rise and the metabolism of carbohydrates, fats, and proteins to weaken, resulting in Type-1 diabetes. Diabetes symptoms include (i) Polyuria (ii) Polydipsia (iii) Weakness (iv) Polyphagia (v) Obesity (vi) Sudden-Weight-Loss (vii) Genital-Thrush (viii) Visual Blurring (ix) Itching (x) Irritability (xi) Delayed-Healing (xii) Partial-Paresis (xiii) Muscle-Stiffness (xiv) Alopecia, etc. [1]. Diabetes is a metabolic disease and which causes millions of deaths around the world yearly due to various health complications. An increase of 70% death ratio from diabetes has been observed between 2000 to 2019 in all over the world [2]. An intelligent ML-based diagnostic system is required to detect these types of fatal diseases. An ML-based expert decision system can successfully diagnose diabetes patients at an early stage. Researchers used various different types of datasets for the prediction of diabetes. ML based framework need an appropriate dataset having necessary features for training, and validation. The selection of appropriate and concerned features from the

The associate editor coordinating the review of this manuscript and approving it for publication was Baozhen Yao.

dataset increases the abilities of the ML model to predict accurately. The dataset used in the proposed system comes from the (University of California Irvine) UCI Machine Learning repository [3], compiled by the hospital of Sylhet, Bangladesh.

Diabetic Mellitus (DM) occurs due to malabsorption of food which alteres glucose level in the body. Preventive measures against malnutrition or obesity that are sometimes primary causes of diabetes include healthy diet and change of lifestyle. Furthermore, these measures help to control the blood pressure, and lower the risk of health complications. Medical checkup makes it easier to diagnose the disease of diabetes. Some laboratory tests are also conducted to detect the disease. Type-2 DM patients need life-saving insulin for as long as they stay alive. Thus, if left unaddressed, this unhealthy condition drains individuals, families, and national resources. Early detection and symptomatic treatment are essential to ensure the healthy life and well-being of pre-diabetic patients. An intelligent medical diagnosis system based on symptoms, signs, laboratory tests, and observations will be helpful in disease detection and prevention. Artificial Intelligence (AI) has also been applied to medical diagnosis systems in several interesting ways for disease detection. This research proposes a framework for early detection of diabetic patients using machine learning fusion.

II. LITERATURE REVIEW

Recent literature has produced a significant amount of research to recognize diabetic patients based on symptoms by applying machine-learning techniques. Based on supervised learning, hybrid learning, or ensemble learning, Pradhan *et al.* [4] applied various algorithms for diagnosing diabetes mellitus to gain higher accuracy rate, but the ensemble approach performs better than the other two approaches. In an ensemble approach, Kumari *et al.* [5] improved classification accuracy by applying a soft voting classifier to the Pima-Diabetes dataset and Breast-Cancer dataset. According to the results, soft voting classifier achieved 79.08 % accuracy compared to the other machine-learning algorithms.

Sarwar *et al.* [6] used machine-learning algorithms for the detection of diabetes at an early stage by using Pima Diabetes dataset. Their accuracy rates achieved from KNN and SVM were 77%, which is higher than the other four algorithms. A limitation of this paper is the size of the dataset and the missing values. Dey *et al.* [7] used supervised machine learning algorithms: SVM, KNN, Naive Bayes, and ANN with Min-Max scaling (MMS) on the Pima dataset. The accuracy of the model ANN with MMS is 82.35 %, which is higher than the other four algorithms. In [8], the researchers used machine-learning algorithms including Naive Bayes, Random Forest, and Simple CART and used the Weka tool to predict diabetes. The SVM classifier performs better and achieved a 79.13% accuracy, which is higher than the other three algorithms. Saru *et al.* [9] predicted diabetes using a model based on Logistic Regression, SVM, Decision Tree, and KNN. They also compared their accuracy rates without and with Bootstrapping. The accuracy rate of the decision tree with bootstrapping is 94.4%, which is higher than the other two algorithms.

By using machine learning algorithms such as Decision Tree, ANN, Naive Bayes, and SVM, Sonar and Jaya Malini [10] constructed a model to predict diabetic patients. The accuracy rate of the decision tree is 85%, which is higher than the other two algorithms. Wei *et al.* [11], in their paper, designed a model using ML algorithms such as the Naive Bayes, Deep Neural Network (DNN), Logistic Regression, and Decision Trees. The accuracy rate of DNN is 77.86%, which is higher than the other four algorithms. Faruque *et al.* [12] proposed a model that uses four ML algorithms – Support Vector Machine (SVM), C4.5 Decision Tree, K-Nearest Neighbor (KNN), and Naive Bayes to predict diabetes. The accuracy rate of the C4.5 Decision Tree is 73.5%, which is higher than the other three algorithms. Jain *et al.* [13] predicted diabetes, uses various ML algorithms like Neural Network (NN), Fisher Linear Discriminant Analysis (FLDA), Random Forest, Chi-square Automatic Interaction Detection (CHAID), and SVM. The accuracy rate of NN is 87.88%, which is higher than the other four algorithms.

ML algorithms are currently useful for the detection of diseases but the previous research models are less accurate because they usually focused on pre-processing techniques, data balancing, and various types of supervised and semi-supervised learning models. Therefore, it is required to find new technique with decision level fusion which would be able to integrate the accuracy of multiple machine learning algorithms with high disease detection accuracy. For this purpose, a fused ML model is proposed which uses two supervised machine-learning approaches including ANN and SVM [14]–[16] followed by fuzzy logic for decision level fusion.

III. MATERIALS AND METHODS

This article proposes a Fused Model for Diabetes Prediction (FMDP). The proposed FMDP model consists of two main phases. The first phase consist of Training Layer while the second phase consists of Testing Layer. The Training Layer is divided into different stages, including data acquisition, preprocessing, classification, performance evaluation, and machine-learning fusion. The dataset used in this research is taken from the UCI Machine Learning Repository [3]. In the Data Acquisition stage, a dataset that has enough features can be used to predict diabetes. Data is cleaned, normalized, and divided in to training and test dataset during the preprocessing stage. Preprocessed data can be used to train Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) for the prediction. We can select several Machine-learning algorithms for the classification to achieve the required accuracy. However, in the proposed model, we used only two widely used ML algorithms (SVMs and ANNs) [14], [16], [19]. These algorithms are selected in this

research after some initial experiments where we have found these techniques more effective for this problem. We used various accuracy measures, including: accuracy, specificity, sensitivity, precision, and F1 score in the Performance Evaluation stage. If the proposed model does not meet the learning requirements, it will be retrained. When learning requirements are meet, the ANN and SVM outputs are used as inputs in machine-learning fusion. In the Machine-Learning Fusion stage, fuzzy rules are applied to the actual output of SVM and ANN results for final prediction. The fused trained model is then stored in the cloud.

The second phase of the proposed framework is reflected by the testing layer. The testing layer acquires dataset from medical database, and loads preprocessed training model from the cloud. A fused model is used to predict whether a diabetes diagnosis is positive or negative. Prediction accuracy is calculated by comparing the required output with the actual output.

The ANN model is trained with the preprocessed training dataset. We have divided the preprocessed data into training and test data with 70:30 ratio on the basis of class base split. For training the data we have used Bayesian regularization function with 5% is used for testing and 5% for validation, and the remaining 90% is used for training.

There are 16 hidden layers between input and output neurons. Where ω1, ω2, ω3… ω16 & ⍱1, ⍱2, ⍱3 … ⍱16 represents the input layer neurons and hidden layer neurons respectively. Output is represented as ''out ϖ''. A bias can be represented as ɦ1 and ɦ2. This produces an out ϑ and out ϱ based on the following equations 1 and 2.

$$\text{out}\vartheta = \frac{1}{1+\text{e}^{-(\text{ɦ}1\sum_{r=1}^{m}(\text{u}_{r,\vartheta}*\omega))}} \tag{1}$$

where, $\vartheta = 1,\ 2, \ldots,\ \text{n}$

$$\text{out}\varrho = \frac{1}{1+\text{e}^{-(\text{ɦ}2\sum_{\vartheta=1}^{n}(\text{ƥ}_{\vartheta,\varrho}*\ \text{out}\vartheta))}} \tag{2}$$

where, $\varrho = 1, 2, \ldots, \text{r}$

Using the squared error function, each output neuron's error can be calculated and summed to find the total error (E).

$$E = \frac{1}{2}\sum\nolimits_{\varrho}\left(\tau_{\varrho} - out_{\varrho}\right)^2 \tag{3}$$

Weights can be changed according to error using the formula in Equation.4

$$\Delta\omega\propto-\frac{\partial\text{E}}{\partial\omega} \tag{4}$$

Equation 5 updates the weight between a hidden layer and an output layer.

$$\Delta\text{ƥ}_{\vartheta,\varrho} = -\varepsilon\frac{\partial E}{\partial ⍱_{\vartheta,\varrho}} \tag{5}$$

As $⍱_{\vartheta,\varrho}$ cannot be calculated directly, so use the Equation. 6 formulae.

$$\Delta\text{ƥ}_{\vartheta,\varrho} = -\varepsilon\frac{\partial E}{\partial out_{\varrho}} \times \frac{\partial out_{\varrho}}{\partial net_{\varrho}} \times \frac{\partial net_{\varrho}}{\partial\text{ƥ}_{\vartheta,\varrho}} \tag{6}$$

where τ_{ϱ} represents the actual weight of ϱ as described in Equation. 7.

$$\Delta\text{ƥ}_{\vartheta,\varrho} = \varepsilon\left(\tau_{\varrho} - \text{out}_{\varrho}\right) \times \text{out}_{\varrho}\ \left(1 - \text{out}_{\varrho}\right)(\text{out}\vartheta) \tag{7}$$

Equations 8 and 9 describe how the weights b/w hidden-layer neurons and input-layer neurons are updated.

$$\Delta\tilde{\text{ʊ}}_{\text{i},\vartheta} \propto -\left[\sum\nolimits_{\varrho}\frac{\partial\text{E}}{\partial\text{out}_{\varrho}} \times \frac{\partial\text{out}_{\varrho}}{\partial\text{net}_{\varrho}} \times \frac{\partial\text{net}_{\varrho}}{\partial\text{out}_{\vartheta}}\right] \times[\frac{\partial\text{out}_{\vartheta}}{\partial\text{net}_{\vartheta}} \times \frac{\partial\text{net}_{\vartheta}}{\partial\tilde{\text{ʊ}}_{\text{i},\vartheta}}] \tag{8}$$

$$\Delta\tilde{\text{ʊ}}_{\text{i},\vartheta} = \xi\left[\sum\nolimits_{\varrho}(\tau_{\varrho} - out_{\varrho}) \times out_{\varrho}\,(1 - out_{\varrho}) \times \text{ƥ}_{\text{i},\vartheta}\right] \times out_{\varrho}\left(1 - out_{\varrho}\right) \times \omega_{\bar{\text{i}}} \tag{9}$$

The weights updating formula between hidden and output layer neurons is described in Equation.10.

$$\Delta\tilde{\text{ʊ}}_{\text{i},\vartheta}(\text{t}+1) = \tilde{\text{ʊ}}_{\text{i},\vartheta}(,\text{t}) + \lambda\ \Delta\tilde{\text{ʊ}}_{\text{i},\vartheta} \tag{10}$$

Once the training model has been successfully trained, it should be saved and validated with 30% of the remaining datasets. When results are saved, the output of the validation data is compared with the actual output and it is found that the prediction is 92.31%.

SVM generates a hyperplane that categorizes data based on classes. SVM categorizes diabetes symptoms into Positive and Negative [15], [17], [18]. Separating classes in a hyperplane begins by drawing a line. The line equation can be expressed in Equation.11.

$$\dot{\text{x}}_2 = \boldsymbol{a}\dot{\text{x}}_1 + \text{ƀ} \tag{11}$$

where $\boldsymbol{a}$ indicates the slope and b represents an intersecting point. Hence, it is written as follows:

$$\boldsymbol{a}\dot{\text{x}}_1 - \dot{\text{x}}_2 + \text{ƀ} = 0 \tag{12}$$

If $\dddot{\text{x}} = (\dot{\text{x}}_1, \dot{\text{x}}_2)^{\text{T}}$ & $\dddot{\omega} = (\boldsymbol{a}, -1)$, then Using the above expression, we can formulate an Equation. 13.

$$\dddot{\omega}.\dddot{\text{x}} + \text{ƀ} = 0 \tag{13}$$

Hyperplane equation can be used to analyze a three-dimensional vector. In Equation. 14, the vector of $\dddot{\text{x}} = (\dot{\text{x}}_1, \dot{\text{x}}_2)$ is represented by $\dddot{\omega}$.

$$\dddot{\omega} = \frac{\dot{\text{x}}_1}{\dot{\text{x}}} + \frac{\dot{\text{x}}_2}{\dot{\text{x}}} \tag{14}$$

In Equation.15, it is shown how n-dimensional vectors can be written.

$$\dddot{\omega}.\dddot{\text{x}} = \sum\nolimits_{i=1}^{\text{n}}\dot{\omega}_i\dot{\text{x}}_i \tag{15}$$

where, $i = 1, 2, \ldots, \text{n}$

Equation.15 enables to check whether the data has been classified correctly.

$$Đ_i = \acute{y}_i\left(\dddot{\omega}.\dddot{\text{x}} + \text{ƀ}\right)$$

Functional margins of datasets are referred to as $\dot{d}$ and are expressed as

$$\dot{d} = \min_{i=1\ldots m} \text{Đ}_i$$

The Geometric-Margin $\dot{d}$ of dataset provides the hyperplane that will be the optimal-hyperplane with the Lagrangian function

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

$$\nabla_{\dot{\omega}}\text{Ɣ}\left(\dot{\omega},\ \mathfrak{b}, \mathfrak{B}\right) = \dot{\omega} - \sum_{i=1}^{m} \mathfrak{B}_i y_i \dot{x}_i = 0 \tag{17}$$

$$\nabla_b \text{Ɣ}\left(\dot{\omega},\ \mathfrak{b}, \mathfrak{B}\right) = -\sum_{i=1}^{m} \mathfrak{B}_i y_i = 0 \tag{18}$$

After simplification, it can be written as

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

The Lagrangian function Ɣ is substituted.

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

Equation.20 can therefore also be used to define the above Equation.

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

where $i = 1, 2, 3, \ldots, m$

To avoid containment inequalities, apply the KKT (Karush-Kuhn-Tucker) condition to the Lagrangian multiplier procedure.

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

where $\dot{x}^*$ represents an optimum point, and the value for $\eth$ is positive; while for other points, it is nearly zero. Accordingly, the equation can be written as Equation.22.

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

The points nearest the hyperplane are also known as support vectors. Based on Equation 23,

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

In other words, it can be written as

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

Equation. 24 gets the value of b when we compute it.

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

Both sides of the equation are multiplied by y_i

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

It is known that $y_i{}^2$ equals 1.

$$\mathfrak{b} = y_i - \left[\dot{\omega}_i \cdot \dot{x}^*\right] \tag{27}$$

$$\mathfrak{b} = \left\{\frac{1}{\text{Ś}}\sum_{i=1}^{\text{Ś}} \left(y_i - \left[\dot{\omega}.\ \dot{x}\right]\right)\right. \tag{28}$$

Equation.27 determines no. of support vectors Ś, and predictions that are made based on the hyperplane.

In Equation.28, the hypothesis function is described where $\dot{\omega}_i$ represent the optimum weight.

$$H\left(\dot{\omega}_i\right) = \begin{cases} +\mathbb{1} & if\ \left(\dot{\omega}_i.\dot{x} + \mathfrak{b}\right) \geq \mathbb{0} \\ -\mathbb{1} & if\ \left(\dot{\omega}_i.\dot{x} + \mathfrak{b}\right) < \mathbb{0} \end{cases} \tag{29}$$

When points are above to the hyperplane, i.e. $+\mathbb{1}$, represents diabetes positive, and points are below to the hyperplane, i.e. $-\ \mathbb{1}$, represents diabetes negative. We used the same dataset with SVM as well as with ANN. The data is trained by using and optimizing all of the available parameters of SVM in Matlab R2020a. The five-fold cross-validation process splits data into five levels and validates them accordingly.

Fuzzy logic uses membership functions. The fuzzy system uses SVM and ANN outputs as input variables. Membership functions define the set of rules that apply to both inputs and outputs. ANN and SVM used fuzzy logic to determine whether a patient's symptoms match a diabetes diagnosis or not. A mathematically fuzzy basis decision could be described as follows:

$$\zeta_{\text{ANN}} \cap \zeta_{\text{SVM}}\left(\text{ANN},\ \text{SVM}\right) = \min\left[\zeta_{\text{ANN}}\left(\text{ANN}\right),\ \ \zeta_{\text{SVM}}\left(\text{SVM}\right)\right] \tag{30}$$

The membership function of ANN is defined as ζ_{ANN} and that of SVM as ζ_{SVM}. According to the results, the outcome parameters for ANN and SVM are either 0 or 1.Two possible outcomes of each model produce four rules sets, which are given below.

- If the ANN model result is Positive (0) and SVM model result is Positive (0), then diabetes is Positive (0).
- If the ANN model result is Positive (0) and SVM model result is Negative (1), then diabetes is Positive (0).
- If the ANN model result is Negative (1) and SVM model result is Positive (0), then diabetes is Negative (1).
- If ANN model result is Negative (1) and SVM model result is Negative (1), then diabetes is Negative (1).

The individual prediction of ANN and SVM in terms of positive or negative is sent to the module of fuzzy logic, which consists of four rules (discussed above) reflected in fuzzy membership function in Table 1, where $\ddot{\mathfrak{Z}}$ reflects SVM membership function and $\mathfrak{Y}$ reflects ANN membership function. The fuzzy logic incorporates the decision level fusion for final prediction that whether the patient is diabetic or not. An algorithm for fuzzy inference can be expressed by $\bar{\text{R}}\,\ddot{\text{u}}^{\text{e}}$ which can be described as

$$\bar{\text{R}}\,\ddot{\text{u}}^{\text{e}} = \ddot{\mathfrak{Z}}^{\text{e}} \times \mathfrak{Y}^{\text{e}} \tag{31}$$

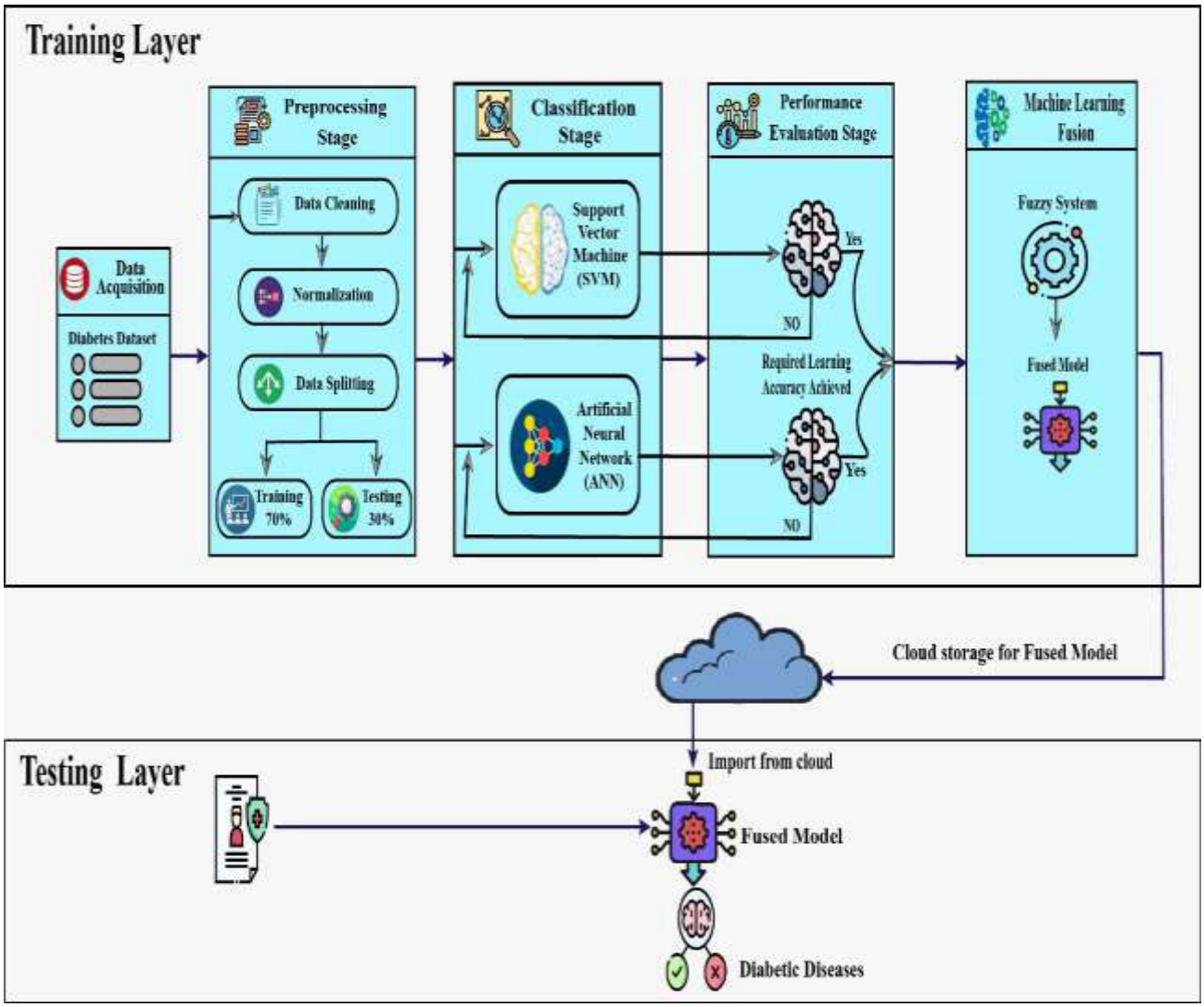


FIGURE 1. Proposed fused model for diabetes prediction (FMDP).

$$\zeta_{\text{ANN}\cap\text{SVM}} = \zeta_{SVM}\left(\breve{\mathfrak{Z}}\right) \cap \zeta_{ANN}\left(\mathfrak{H}\right) \tag{32}$$

Fuzzy relation $\mathcal{Q}_4$ is defined using the rules.

$$\mathcal{Q}_4 = \bigcup_{\dot{\mathfrak{e}}=1}^{4} \bar{\mathrm{R}}_{\cdot\,\cdot\,\cdot}^{\mathrm{e}}\,\mathrm{u}_{\cdot\,\cdot\,\cdot} \tag{33}$$

$$\zeta_{\ddot{\mathrm{R}}}(Decision) = \max_{1<\dot{\mathrm{x}}<4}\left[\prod_{\mathfrak{g}=1}^{4}\left(\zeta_{ANN_{\mathfrak{H}}}\zeta_{SVM_{\breve{\mathfrak{Z}}}}\right)\right] \tag{34}$$

A de-fuzzier can be applied using various types of methods like the center-of-area method (COA), the weighted average method; the mean of maxima method (MOM), and maximum-membership principle but the proposed model applies the centroid method de-fuzzier. The interface engine produces fuzzy output that is transformed using similar functionalities as the fuzzier to generate frangible output. Crisp points are discussed in Equation 35.

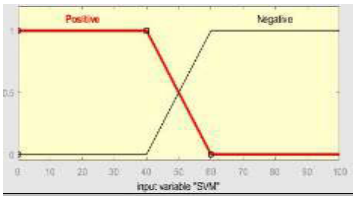
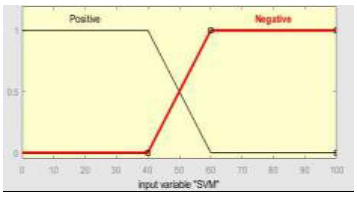
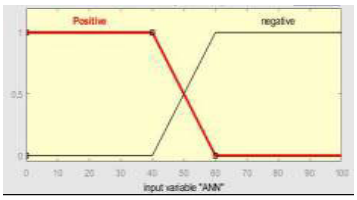
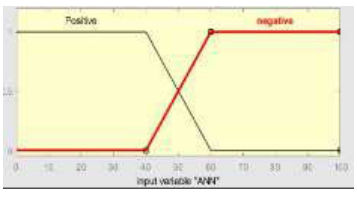
$$\mathcal{F} = \frac{\int_{\ddot{\mathrm{R}}} \zeta_{\ddot{\mathrm{R}}}\left(\dot{\ddot{\mathrm{R}}}\right) \mathrm{d}\ddot{\mathrm{R}}}{\int \zeta_{\ddot{\mathrm{R}}}\left(\dot{\ddot{\mathrm{R}}}\right) \mathrm{d}\ddot{\mathrm{R}}} \tag{35}$$

The graph in figure 2 describes that the x and y axes correspond to SVM and ANN, while z indicates the FMDP system. The FMDP System rule surface of diabetes can be seen in comparison to ANN and SVM results. The resultant FMDP System predicts no diabetes if both solutions predict no diabetes; And if both models predicts diabetes as yes then FMDP also predicts diabetes as yes.

Figure 3 shows that if ANN diagnoses yes (0) diabetes and SVM diagnoses no (1) diabetes, then the fused model also diagnoses yes (0) diabetes.

Figure 4 shows that if ANN diagnoses no (1) diabetes and SVM diagnose no (1) diabetes, then the fused model also diagnoses no (1) diabetes.

TABLE 1. Fuzzy membership functions of FMDP system.

$\mathrm{SVM}_{(\zeta_{\mathrm{SVM}(\ddot{\mathfrak{z}})})}$	$\zeta_{(\mathrm{Positive})}(\ddot{\mathfrak{z}}) = \left\{\max\left(\min\left(1, \frac{60-\ddot{\mathfrak{z}}}{20}\right), 0\right)\right\}$	
$\mathrm{SVM}_{(\zeta_{\mathrm{SVM}(\ddot{\mathfrak{z}})})}$	$\zeta_{(\mathrm{Negative})}(\ddot{\mathfrak{z}}) = \left\{\max\left(\min\left(\frac{\ddot{\mathfrak{z}}-40}{20}, 1\right), 0\right)\right\}$	
$\mathrm{ANN}_{(\zeta_{\mathrm{ANN}(\mathfrak{h})})}$	$\zeta_{(\mathrm{Positive})}(\mathfrak{h}) = \left\{\max\left(\min\left(1, \frac{60-\mathfrak{h}}{20}\right), 0\right)\right\}$	
$\mathrm{ANN}_{(\zeta_{\mathrm{ANN}(\mathfrak{h})})}$	$\zeta_{(\mathrm{Negative})}(\mathfrak{h}) = \left\{\max\left(\min\left(\frac{\mathfrak{h}-40}{20}, 1\right), 0\right)\right\}$	

In the proposed framework, the validation layer relates to the real-time diagnosis and classification of a diabetic. The proposed fused ML model can use real-time patient data as input and improve the disease detection system.

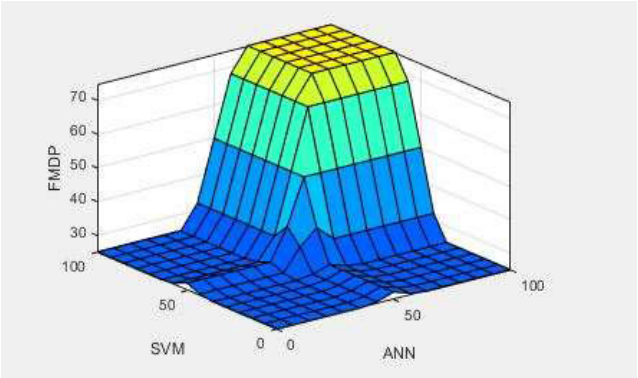


FIGURE 2. Proposed FMDP system rule surface.

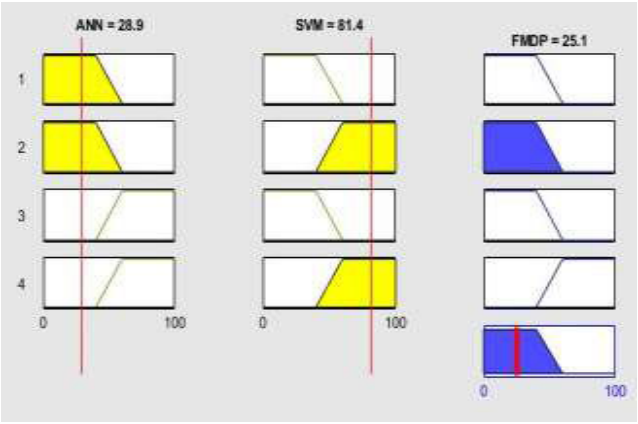


FIGURE 3. Proposed FMDP system result with diabetes (yes).

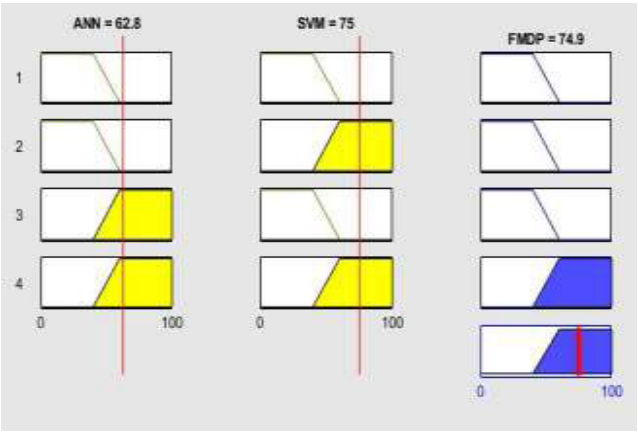


FIGURE 4. Proposed FMDP system result with diabetes (no).

IV. RESULTS AND DISCUSSION

To implement the proposed framework, we used a dataset [3] where the total number of instances is 520, and has 17 attributes based on diabetic symptoms. Sixteen features are independent, and one is the target feature (dependent). The target feature is labeled as the class, which has two values either 0 or 1. The class 0 represents that the person has diabetic symptoms (Positive) and class 1 represents that the person has no diabetic symptoms (Negative). The first feature of the dataset is Age in which 93 persons have age between 20 years to 35 years, 138 persons have age between 36 years to 45 years, 149 persons have age between 46 years to 55 years, 89 persons have age between 56 years to 65 years, and 51 persons are above 65 years. The second feature is Sex in which 382 are males and 192 are females. Male is reflected by "0" and Female is reflected by "1". The third feature is

TABLE 2. (Training) confusion-matrix for ANNs.

Input Values	Total= 364	Output Results: Positive $\dot{O}\dot{R}_0$	Output Results: Negative $\dot{O}\dot{R}_1$
	Positive ($\dot{E}\dot{R}_0$= 246)	236	10
	Negative ($\dot{E}\dot{R}_1$ = 118)	11	107

TABLE 3. (Testing) confusion-matrix for ANNs.

Input Values	Total= 156	Output Results: Positive $\dot{O}\dot{R}_0$	Output Results: Negative $\dot{O}\dot{R}_1$
	Positive ($\dot{E}\dot{R}_0$= 69)	61	8
	Negative ($\dot{E}\dot{R}_1$ = 87)	4	83

the Polyuria symptom, which has 258 values as ''yes'' and reflected by ''0'' and 262 values as ''no'', reflected by ''1''. The fourth feature is the Polydipsia symptom, which has 233 ''yes'' as ''0'' and 287 ''no'' as ''1''. The fifth feature is the Sudden weight loss symptom which has 217 ''yes'' as ''0'' and 303 ''no'' set to as ''1''. The sixth feature is the Weakness symptom, which has 305 ''yes'' set to ''0'' and 215 ''no'' set to ''1''. The seventh feature is the Polyphagia symptom, which has 237 ''yes'' set to ''0'' and 283 ''no'' set to ''1''. The eighth feature is the Genital Thrush symptom which has 116 ''yes'' set to ''0'' and 404 ''no'' set to ''1'' values. The ninth feature is the Visual Blurring symptom which has 233 values as ''yes'' and set to ''0'' and 287 as ''no'' set to ''1''. The tenth feature is the Itching symptom which has 253 as ''yes'' set to ''0'' and 267 has ''no'' set to ''1'' values. The eleventh feature is the Irritability symptom, which has 126 values as ''yes'' and set to ''0'' and 394 values as ''no'' set to ''1''. The twelfth feature is the Delayed healing symptom which has 239 ''yes'' values and set to ''0'' and 281 values as ''no'' set to ''1''. The thirteenth feature is the Partial paresis symptom which has 224 values as ''yes'' set to ''0'' and 296 values as ''no'' set to ''1'' v. The fourteenth feature is the Muscle stiffness symptom which has 195 values as ''yes'' set to ''0'' and 325 values ''no'' set to ''1''. The fifteenth feature is the Alopecia symptom, which has 179 values as ''yes'' set to ''0'' and 341 ''no'' set to ''1''. The sixteenth feature is the Obesity symptom, which has 88 ''yes'' set to ''0'' and 432 ''no'' set to ''1'' values.

In this article, MATLAB R2020a is used for simulation purposes. ANN and SVM are used for prediction, whereas fuzzy logic is used in decision-making. The dataset is divided into training and testing datasets with the ratio of 70:30. There are 364 instances, which are used in training of ANN and its confusion matrix is shown in Table 2.

Table. 2 describes the 246 positive cases, of which 236 cases were predicted accurately, whereas 10 cases were predicted incorrectly. However, there are 118 negative cases, of which 107 cases are predicted accurately, whereas 11 cases are predicted incorrectly.

There are 156 instances in testing data. The confusion matrix of ANN testing is shown in Table 3.

Table. 3 describes the 69 cases with positive diabetes, of which 61 cases predicted accurately, whereas 8 cases are predicted incorrectly. However, there are 87 cases of Negative diabetes, of which 83 cases are predicted accurately, whereas 4 cases are predicted incorrectly.

TABLE 4. (Training) confusion-matrix for SVMs.

Input Values	Total= 364	Output Results: Positive $\dot{O}\dot{R}_0$	Output Results: Negative $\dot{O}\dot{R}_1$
	Positive ($\dot{E}\dot{R}_0$ = 246)	227	19
	Negative ($\dot{E}\dot{R}_1$= 118)	13	105

TABLE 5. (Testing) confusion-matrix for SVMs.

Input Values	Total= 156	Output Results: Positive $\dot{O}\dot{R}_0$	Output Results: Negative $\dot{O}\dot{R}_1$
	Positive ($\dot{E}\dot{R}_0$ = 69)	59	10
	Negative ($\dot{E}\dot{R}_1$= 87)	7	80

TABLE 6. (Testing) confusion-matrix for FMDP.

Input Values	Total= 156	Output Results: Positive $\dot{O}\dot{R}_0$	Output Results: Negative $\dot{O}\dot{R}_1$
	Positive ($\dot{E}\dot{R}_0$= 69)	64	**5**
	Negative ($\dot{E}\dot{R}_1$ = 87)	3	84

We have used five-fold cross-validation for SVM. The confusion matrix of SVM training is shown in Table 4.

Table. 4 describes the 246 cases of positive diabetes, of which 227 cases were predicted accurately, whereas 19 cases were predicted incorrectly. However, there are 118 cases of negative diabetes, of which 105 cases are predicted accurately, whereas 13 cases are predicted incorrectly. There are 156 instances, which are used in testing of SVM and its confusion matrix is shown in Table 5.

Table. 5 describes the 69 cases of positive diabetes, of which 59 cases were predicted accurately, whereas 10 cases were predicted incorrectly. However, there are 87 cases of negative diabetes, of which 80 cases predicted accurately, whereas 7 cases were predicted incorrectly.

TABLE 7. Results of ANN, SVM, and proposed FMDP.

	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

TABLE 8. Comparison of FMDP with state-of-the-art techniques.

Authors/Papers	Approach	Accuracy (%)	Miss Rate (%)
S. Kumari et al [5]	Ensemble soft voting classifier	79.08 %	20.92%
M. A. Sarwar et al [6]	KNN & SVM	77%	23%
S. K. Dey et al [7]	ANN with Min-Max scaling	82.35 %	17.65%
A. Mir et al [8]	SVM	79.13%	20.87%
S. Saru et al [9]	Decision Tree after bootstrapping	94.4%	5.6%
P. Sonar et al [10]	Decision Tree	85%	15%
S. Wei et al [11]	Deep Neural Network	77.86%	22.14%
M. F. Faruque et al [12]	Decision Tree	73.5%	26.5%
B. Jain et al [13]	Neural Network	87.88%	12.12%
Proposed Model	Fused ML Decision	94.87%	5.13%

Table. 6 reflects the confusion matrix of testing with proposed fused model. It reflects the 69 cases of positive diabetes, from which 64 were predicted accurately, whereas 5 were predicted incorrectly. On the other hand, there are total 87 cases of negative diabetes, of which 84 cases predicted accurately, and 3 cases were predicted incorrectly.

In the formulas given below, $\ddot{O}\dot{R}_0$, $\ddot{O}\dot{R}_1$, $\acute{E}\dot{R}_0$, and $\acute{E}\dot{R}_1$ reflect the predicted positive output, predicted negative output, expected positive results and expected negative results respectively.

$$\text{Ǎccuracy} = \frac{(\ddot{O}\dot{R}_0/\acute{E}\dot{R}_0 + \ddot{O}\dot{R}_1/\acute{E}\dot{R}_1)}{(\acute{E}\dot{R}_0 + \acute{E}\dot{R}_1)} \tag{36}$$

$$\text{M}'\text{iss Rate} = \frac{(\ddot{O}\dot{R}_1/\acute{E}\dot{R}_0 + \ddot{O}\dot{R}_0/\acute{E}\dot{R}_1)}{(\acute{E}\dot{R}_0 + \acute{E}\dot{R}_1)} \tag{37}$$

$$\text{'Positive 'Prediction Vlaue} = \frac{(\ddot{O}\dot{R}_1/\acute{E}\dot{R}_1)}{(\ddot{O}\dot{R}_1/\acute{E}\dot{R}_1 + \ddot{O}\dot{R}_0/\acute{E}\dot{R}_1)} \tag{38}$$

$$\text{Negative 'Prediction Vlaue} = \frac{(\ddot{O}\dot{R}_0/\acute{E}\dot{R}_0)}{(\ddot{O}\dot{R}_0/\acute{E}\dot{R}_0 + \ddot{O}\dot{R}_1/\acute{E}\dot{R}_0)} \tag{39}$$

$$\dot{\text{S}}\text{pecificity} = \frac{(\ddot{O}\dot{R}_0/\acute{E}\dot{R}_0)}{(\ddot{O}\dot{R}_1/\acute{E}\dot{R}_1 + \ddot{O}\dot{R}_0/\acute{E}\dot{R}_1)} \tag{40}$$

$$\text{Sensitivity} = \frac{(\ddot{O}\dot{R}_1/\acute{E}\dot{R}_1)}{(\ddot{O}\dot{R}_1/\acute{E}\dot{R}_0 + \ddot{O}\dot{R}_1/\acute{E}\dot{R}_1)} \tag{41}$$

$$\dot{\text{F}}\text{alse Discovery Rate} = \frac{(\ddot{O}\dot{R}_1/\acute{E}\dot{R}_0)}{(\ddot{O}\dot{R}_1/\acute{E}\dot{R}_0 + \ddot{O}\dot{R}_0/\acute{E}\dot{R}_0)} \tag{42}$$

$$\dot{\text{F}}\text{alse 'Positive Rate} = 1 - \dot{\text{S}}\text{pecificity} \tag{43}$$

$$\dot{\text{F}}\text{alse Ṅegative Rate} = 1 - \dot{\text{S}}\text{ensitivity} \tag{44}$$

$$\dot{\text{F}}1\dot{\text{S}}\text{core} = 2 * \frac{\text{'Positive 'Prediction Vlaue} * \dot{\text{S}}\text{ensitivity}}{\text{'Positive 'Prediction Vlau} + \dot{\text{S}}\text{ensitivity}} \tag{45}$$

The performance of both models (ANN and SVM) along with the proposed fused model is evaluated by using various accuracy measures as discussed above and reflected in Table.7. It can be seen that the proposed fused model performed better on testing data as compared to both of the used models (ANN, SVM).

The proposed fused model is also compared to previous published models and techniques in Table.8. It can be observed that the proposed fused technique outperformed al l of the other published techniques and achieved the accuracy of 94.87% and miss rate of 5.13%.

V. CONCLUSION

Though different models had been used for the prediction of diabetes, the accuracy of the proposed models in disease prediction has always been the main concern of researchers. Therefore, a new model is required in order to achieve higher prediction accuracy in diabetes prediction. This research

proposed a machine learning based diabetes decision support system by using decision level fusion. Two widely used machine learning techniques are integrated in the proposed model by using the fuzzy logic. The proposed fuzzy decision system has achieved the accuracy of 94.87, which is higher than the other existing systems. Through this diagnosis model, we can save several lives. Moreover, the death ratio of diabetes can be controlled if the disease is diagnosed and preventative measures are taken in early-stage.

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. Comput. Commun. Control Autom. (ICCUBEA)*, Aug. 2018, pp. 1–6, doi: 10.1109/ICCUBEA.2018.8697439.

[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 the machine 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.

[17] Y. Yang, J. Wang, and Y. Yang, "Improving SVM classifier with prior knowledge in microcalcification detection1," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep. 2012, pp. 2837–2840, doi: 10.1109/ICIP.2012.6467490.

[18] J. Liu, J. Feng, Q. Xiao, S. Liu, F. Yang, and S. Lu, "Fault diagnosis of rod pump oil well based on support vector machine using preprocessed indicator diagram," in *Proc. IEEE 10th Data Driven Control Learn. Syst. Conf. (DDCLS)*, May 2021, pp. 120–126, doi: 10.1109/DDCLS52934.2021.9455702.

[19] P. Patil, N. Yaligar, and S. M. Meena, "Comparision of performance of classifiers–SVM, RF and ANN in potato blight disease detection using leaf images," in *Proc. IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC)*, Dec. 2017, pp. 1–5, doi: 10.1109/ICCIC.2017.8524301.

USAMA AHMED received the B.S.C.S. degree from the University of Sargodha. He is currently pursuing the M.S. degree in computer science with the Riphah School of Computing and Innovation, Faculty of Computing, Riphah International University, Lahore, Pakistan. Currently, he is serving as an Instructor/Tutor of the Computer Science Department, Virtual University of Pakistan. His research interests include machine learning, medical diagnosis, data mining, and image processing.

GHASSAN F. ISSA received the M.S. and Ph.D. degrees in computer science/artificial intelligence from Old Dominion University, VA, USA, in 1987 and 1992, respectively. He was a Faculty Member and the Department Chair of computer science at the Pennsylvania College of Technology (Penn State), USA, from 1992 to 1995. He also served as the Dean for computer science at Applied Science University, Amman, Jordan, from 2003 to 2005, and the Dean for information technology at the University of Petra, Amman, from 2008 to 2018. He is a Professor of computer science. Currently, he is a Professor and the Dean of the School of Information Technology, Skyline University College, Al Sharjah, United Arab Emirates. His research interests include AI, and machine learning with work on deep neural networks fine-tuning, learning by analogy, and associative classification algorithms.

MUHAMMAD ADNAN KHAN received the B.S. and M.Phil. degrees from the International Islamic University, Islamabad, Pakistan, and the Ph.D. degree from ISRA University, Pakistan. He is currently working as an Associate Professor with the Riphah School of Computing and Innovation, Faculty of Computing, Riphah International University, Lahore, Pakistan, and an Assistant Professor with the Pattern Recognition and Machine Learning Laboratory, Department of Software, Gachon University, South Korea. Before joining Gachon University and Riphah International University, he has worked in various academic and industrial roles in Pakistan. He has been teaching graduate and undergraduate students in computer science and engineering for the past 13 years. Currently, he is guiding five Ph.D. scholars and seven M.Phil. scholars. He has published more than 200 research articles with Cumulative JCR-IF of more than 370 in international journals as well as reputed international conferences. His research interests include machine learning, MUD, image processing and medical diagnosis, and channel estimation in multi-carrier communication systems using soft computing. He received the Scholarship Award from the Punjab Information and Technology Board, Government of Punjab, Pakistan, for his B.S. and M.Phil. degrees, and the Scholarship Award from the Higher Education Commission, Islamabad, in 2016, for his Ph.D. degree.

SHABIB AFTAB (Member, IEEE) received the M.S. degree in computer science from the COMSATS Institute of Information Technology, Lahore, Pakistan, and the M.Sc. degree in information technology from the Punjab University College of Information Technology (PUCIT) Lahore. He is currently pursuing the Ph.D. degree in computer science with the School of Computer Science, National College of Business Administration and Economics, Lahore. Currently, he is serving as a Lecturer of computer science at the Virtual University of Pakistan. His research interests include data mining and software process improvement.

MUHAMMAD FARHAN KHAN received the M.Phil. degree in forensic sciences from the University of Veterinary and Animal Sciences (UVAS), Lahore. He is currently pursuing the Ph.D. degree in forensic sciences with the University of Heath Sciences Lahore, Lahore, Pakistan. Currently, he is serving as a Lecturer at the Forensic Sciences Department, University of Health Sciences Lahore. His research interests include medical sciences, behavioral sciences, medical diagnosis, forensic sciences, molecular biology, data mining, and image processing.

RAED A. T. SAID received the B.Sc. degree in statistics from Yarmouk University, Irbid, Jordan, the M.Sc. degree in population studies from The University of Jordan, and the Ph.D. degree in statistics from the University of Leeds, U.K.

He has worked in several universities in United Arab Emirates for more than 20 years. Before joining Canadian University Dubai (CUD), in 2018, he was an Assistant Professor at the Business Administration College, Al Ain University, where he also served as the Director for the Quality Assurance and Institutional Research Center for three years. Prior to that, he has taught research methodologies and different statistics and mathematics courses for the business, engineering, and education students at Ajman University. He is particularly interested in developing enhanced data mining methods for uncovering patterns in spatiotemporal data. He has published in refereed journals and presented papers at numerous international conferences on topics related to data mining applications for global sustainability and contributed to developments in science, technology, and innovation. He is also involved in research activities regarding information technologies and their applications to enhance the quality of education and to improve the assessment methods and the delivery of courses in higher educational institutions.

TAHER M. GHAZAL (Member, IEEE) received the B.Sc. degree in software engineering from Al Ain University, in 2011, the M.Sc. degree in information technology management from The British University in Dubai, associated with The University of Manchester and The University of Edinburgh, in 2013, and the Ph.D. degree in IT/software engineering from Damascus University, in 2019. He is currently pursuing the Ph.D. degree in information science and technology with Universiti Kebangsaan Malaysia. He has more than ten years of extensive and diverse experience as an Instructor, a Tutor, a Researcher, a Teacher, an IT Support/Specialist Engineer, and a Business/Systems Analyst. He served in engineering, computer science, ICT, the Head of STEM, and innovation departments. He was also involved in quality assurance, accreditation, and data analysis in several governmental and private educational institutions under KHDA, Ministry of Education, and Ministry of Higher Education and Scientific Research, United Arab Emirates. His research interests include the IoT, IT, artificial intelligence, information systems, software engineering, web developing, building information modeling, quality of education, management, big data, quality of software, and project management. He is actively involved in community services in the projects and research field.

MUNIR AHMAD (Member, IEEE) received the Master of Computer Science from the Virtual University of Pakistan, in 2018. He is currently pursuing the Ph.D. degree in computer science with the School of Computer Science, National College of Business Administration and Economics. He has spent several years in industry. He is currently working as the Executive Director/Head of the IT Department, United International Group, Lahore, Pakistan. He has vast experience in data management and efficient utilization of resources at multinational organizations. He has conducted many research studies on sentiment analysis and utilization of AI for prediction on various healthcare issues. His research interests include data mining, big data, and artificial intelligence.