

## Research Article

# A Smart Healthcare Recommendation System for Multidisciplinary Diabetes Patients with Data Fusion Based on Deep Ensemble Learning

**Baha Ihnaini** <sup>1</sup>, **M. A. Khan** <sup>2</sup>, **Tahir Abbas Khan**,<sup>3</sup> **Sagheer Abbas** <sup>3</sup>,  
**Mohammad Sh. Daoud**,<sup>4</sup> **Munir Ahmad** <sup>3</sup>, and **Muhammad Adnan Khan** <sup>5</sup>

<sup>1</sup>Department of Computer Science, College of Science and Technology, Wenzhou-Kean University, Wenzhou 325060, China

<sup>2</sup>Riphah School of Computing and Innovation, Faculty of Computing, Riphah International University, Lahore Campus, Lahore 54000, Pakistan

<sup>3</sup>School of Computer Science, National College of Business Administration and Economics, Lahore 54660, Pakistan

<sup>4</sup>College of Engineering, Al Ain University, Abu Dhabi 112612, UAE

<sup>5</sup>Pattern Recognition and Machine Learning Lab, Department of Software, Gachon University, Seongnam 13557, Republic of Korea

Correspondence should be addressed to Munir Ahmad; [munir@ncbae.edu.pk](mailto:munir@ncbae.edu.pk) and Muhammad Adnan Khan; [adnan@gachon.ac.kr](mailto:adnan@gachon.ac.kr)

Received 12 June 2021; Revised 9 July 2021; Accepted 6 September 2021; Published 17 September 2021

Academic Editor: Ahmed A. Abd El-Latif

Copyright © 2021 Baha Ihnaini et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The prediction of human diseases precisely is still an uphill battle task for better and timely treatment. A multidisciplinary diabetic disease is a life-threatening disease all over the world. It attacks different vital parts of the human body, like Neuropathy, Retinopathy, Nephropathy, and ultimately Heart. A smart healthcare recommendation system predicts and recommends the diabetic disease accurately using optimal machine learning models with the data fusion technique on healthcare datasets. Various machine learning models and methods have been proposed in the recent past to predict diabetes disease. Still, these systems cannot handle the massive number of multifeatures datasets on diabetes disease properly. A smart healthcare recommendation system is proposed for diabetes disease based on deep machine learning and data fusion perspectives. Using data fusion, we can eliminate the irrelevant burden of system computational capabilities and increase the proposed system's performance to predict and recommend this life-threatening disease more accurately. Finally, the ensemble machine learning model is trained for diabetes prediction. This intelligent recommendation system is evaluated based on a well-known diabetes dataset, and its performance is compared with the most recent developments from the literature. The proposed system achieved 99.6% accuracy, which is higher compared to the existing deep machine learning methods. Therefore, our proposed system is better for multidisciplinary diabetes disease prediction and recommendation. Our proposed system's improved disease diagnosis performance advocates for its employment in the automated diagnostic and recommendation systems for diabetic patients.

## 1. Introduction

A recent development in biotechnologies and high throughout computing progressively contribute to quick and affordable e-healthcare data collections and disease diagnosis. The efficiency and reliability are dependent on accurate model building from e-healthcare big data. One of many life-threatening diseases is diabetes disease (DD) [1–3].

Diabetes disease arrests 422 million adults all over the world [4]. The death rate due to diabetes disease is 1.5 million, and 3.7 million deaths are due to diabetes and high blood pressure [4]. Diabetes disease is a multidisciplinary disease that arrests the human body's significant parts like the kidney, eyes, lungs, and heart. The diagnosis of this disease was done either manually by a medical practitioner or by any automatic device.

All of these types of measurements for diabetes disease have some benefits and some drawbacks. Any experienced medical expert cannot manually find the diabetes disease early due to some hidden side effects on the human body. With the intelligent recommendation system's help and application of deep machine learning (DML) and artificial intelligent methods, this disease can be predicted at the earlier stage [5–9] with a minimal error rate.

There are some healthcare automated systems for detection and recommendation of human diseases in recent researches. Myocardial infarction [10] is an acute disease for blood circulation in the heart. In this paper, deep CNN is applied for detection and to prevent humans from a heart attack. A computer-aided diagnosis (CAD) system [11] is used by applying the transfer learning technique for accurate and timely response to reduce the extensive calculation. In this paper, a CAD system works efficiently and accurately to detect and prevent heart attacks. Internet of Healthcare Things (IoHT) and Decentralization Interoperable Trust (DIT) [12] framework are a better healthcare system. In this paper, blockchain is used for data privacy and security. In this research, data is collected via IoHT at each point and transformed through blockchain for smooth and accurate healthcare data for better system accuracy.

Many ensemble learning models have been used in recent healthcare researches for better accuracy. For example, hepatocellular carcinoma [13] is a hazardous cancer disease in the human body. In this paper, an automated prediction system is developed using a stack learning approach for deep learning and examining healthcare data about this deadly disease. Stack learning is an ensemble learning technique. In this paper, evolutionary computational techniques are also used to examine the healthcare data about hepatocellular carcinoma disease. In cervical cancer [14] diagnosis, an ensemble machine learning approach is used. In this paper, two approaches are used for predicting disease on an images dataset.

In mobile edge computing [15], an automated recommendation system is proposed for the joint computation of multiuser offloading and task caching. In this paper, Q-learning and Deep-Q-Network-based algorithms are proposed for this system. Multilevel vehicular edge cloud computing [16], secure federated learning for 5G [17], and augmented Coronavirus disease detection [18] used an advanced ensemble deep learning approach for better results.

For detection of COVID-19 disease, a deep learning approach is adopted with an augmented approach [18] and it achieved 100% accuracy. In industrial mobile edge computing [19], the deep ensemble learning approach is used for resource allocation and data security.

The importance of a smart healthcare recommendation system is increasing day by day for better and timely prediction. To minimize the risk of life-threatening human diseases, we need an efficient system for diagnosing and effectively recommending life-threatening diseases such as diabetes. Electronic health records (EHRs) play an essential role in smart healthcare recommendation systems for predicting life-threatening diseases, especially for multidisciplinary and life-threatening diabetes diseases. However, the data collected from sensors and EHRs are unstructured. To

manage adequately such kinds of multisourced data for further examining is a challenging task.

Further, extracting the critical features and fusing them in a structured form is also a hectic and skills-demanding task. Therefore, this section is further divided into two parts: a wearable sensors-based diabetes prediction system and extracting information from EHRs textual data. Then, data fusion is essential for better results and accurate prediction of diabetes disease with DML.

Many recommendation systems for healthcare are already proposed in recent researches. The significant contribution of this research is to enrich the healthcare dataset for the best prediction of multidisciplinary diabetes disease. We have collected the patients' data through wearable sensors and EHRs in the textual record form of each patient. After collecting the records of each patient, essential data from both ends are fused to enrich the healthcare dataset. The Ensemble deep learning approach works accurately and produces better results in larger healthcare datasets. Finally, we have developed a better recommendation system by collecting patients' records and applying an ensemble machine learning approach for accurate and timely prediction and recommendation of multidisciplinary diabetes disease patients.

The organization of the paper is as follows: Section 2 describes the most recent developments of diabetes disease detection and recommendation from the literature; Section 3 provides research methodology, data fusion, and proposed DML model; Section 4 presents the dataset selection, pre-processing, data fusion, and results and discussion; Section 5 describes the conclusion and future work; and the last section is devoted to references.

## 2. Related Work

Many researchers contributed to diagnosing diabetes disease. They have used machine learning (ML) classifier and artificial intelligence (AI) assistance for the prediction of diabetes disease. With the help of artificial intelligence, we can easily collect healthcare data. After collecting the big data from the healthcare center, we can easily predict human diseases, including multidisciplinary diabetes diseases.

In early detection of diabetes disease [20], the k-nearest neighbor (KNN) classifier model was used and the result was compared with that of the support vector machine (SVM) model achieving 85.6% accuracy. In this paper, the KNN machine learning classifier was used for predicting diabetes disease. The results comparison was made with another machine learning classifier called SVM to authenticate the work. For the detection of diabetes type II disease [21], the authors used a convolution neural network (CNN) and compared their work with the linear regression (LR) model and multilayer perceptron (MLP). In this paper, the neural network was used to diagnose diabetes type II disease. For results comparison, two machine learning classifiers were used for the authenticity of the work. An accuracy of 77.5% was achieved in the area under the curve (AUC). Analysis of early detection of diabetes disease with feature selection technique [22] was carried out using SVM classifier and their results were compared with random forest (RF), naïve Bayes (NB), decision tree (DT), and KNN classification models. The highest

accuracy achieved with SVM was 77.73%. In this paper, the feature selection technique was adopted. With the help of the feature selection technique, we can reduce the system's computational capability and improve accuracy. Multiple machine learning models were applied for comparison and authenticity of results. Bloodless techniques for the prediction of diabetes disease are used with computational tools [23]. The accuracy achieved through this technique was 91.67%. In diabetic retinopathy detection [24], the deep (DNN) technique was adopted. The accuracy achieved via CNN was 74.4%. Detection of multiclass retinal disease [25] was done with with AI. The CNN classifier was used and it achieved 92% accuracy.

A data-driven approach is used for predicting diabetes and cardiovascular diseases [26] with ML. This paper adopted an extreme gradient boost and compared it with the LR, SVM, RF, and weighted ensemble model. An accuracy of 95.7% was achieved in the area under the ROC curve. In type II diabetes disease prediction [27], an ensemble classification model was adopted. The accuracy achieved via this model was 82.2% in the AUC. A new methodology, smartphone-based diabetes detection [28], was presented. In this paper, image data was considered for diagnoses and further directions. A microcontroller-based agent [29] was used to measure the blood glucose level of patients. A sensor integrated therapy [30] for diabetes disease was used to monitor glucose levels in a diabetic patient. A self-recommendation smart app [31, 32] was used and trained on recorded health data like patients' daily physical activities and other important parameters related to diabetes.

The valuable information extraction from wireless sensor data and the patient's electronic medical record is also challenging for predicting multidisciplinary disease. To handle this challenge, different models have been presented for extracting the most valuable information from the healthcare textual data [33–35] for making a dataset for the prediction of diabetic disease. The textual dataset collected through EHR's was pre-processed and converted into a meaningful format as per smart healthcare recommendation system demands. The wireless sensor data of healthcare was also collected through wireless devices. After collecting data through a wireless communication device, data was preprocessed for removing noisy wireless data to make a rich and accurate healthcare dataset. In this way, we can easily apply machine learning algorithms for predicting multidisciplinary diabetes disease.

After collecting and converting meaningful data from textual data and fusing preprocessed wireless sensor data [36–38] for making a rich healthcare dataset of diabetes, Table 1 shows the comprehensive limitation of previously published approaches.

### 3. Smart Healthcare Recommendation System for Multidisciplinary Diabetes Patients

This section explains the overall structure of a smart healthcare recommendation system for multidisciplinary diabetes disease patients (SHRS-M3DP) in detail. The proposed approach is divided into distinct layers for an

accurate description of each layer working. In the end, the ensemble DML structure is presented, which is further deployed in the whole SHRS-M3DP to predict and recommend diabetes disease in the patients. It should deliver a concise and accurate representation of the experimental results, explanation, and the conclusion of experiments that can be drawn.

*3.1. Smart Healthcare Recommendation System for Multidisciplinary Diabetes Patients.* This part explains the overall structure of a smart healthcare recommendation system for multidisciplinary diabetes disease patients (SHRS-M3DP). Initially, the general structure of the proposed SHRS-M3DP is described. Then, a proposed system's structure is divided into distinct layers for an accurate description of each layer working. Finally, the ensemble deep learning model structure is presented, and it is further deployed in the whole SHRS-M3DP to predict and recommend diabetes disease in the patients.

The proposed structure of the SHRS-M3DP system is exhibited in Figure 1. It is divided into two main segments: (1) training phase and (2) validation phase. These phases are essential for the accurate prediction of multidisciplinary diabetes disease. These phases communicate via a cloud. The training phase comprises seven levels: (i) sensory layer, (ii) EHRs layer, (iii) raw feature layer-1, (iv) raw feature layer-2, (v) fused raw feature layer, (vi) preprocessing layer, and (vii) application layer. The sensory layer comprises input parameters, including age, family history, glucose, skin thickness, blood pressure (BP), pregnancies, insulin, and body mass index (BMI). The sensory layer's input values are collected and transferred to the database, raw feature layer-1, through the Internet of medical things (IoMT). Because wireless communication is applied, data collected from multiple feature nodes and stored in the database may be inaccurate. For that reason, we considered such kind of data as feature raw data. EHRs layer consists of lab reports, questions, observation, and the patient's medical history. All the data collected from the EHRs layer are reports and need some methodology to convert them into a structured format for further processing.

The Framingham risk factors (FRFs) methodology used in the smart healthcare monitoring system for heart disease prediction [39] is adopted to extract data from the EHRs, as shown in Figure 2, and stored in raw feature layer-2.

The data fusion approach is then applied for fusing the common features of both sensory data and EHRs to generate enhanced healthcare data on multidisciplinary diabetes disease.

These fused feature data are then stored in the fused feature layer for further processing to predict diabetes disease.

The following preprocessing layer plays a crucial role in the model. All deficiencies received through the sensory layer previously via wireless communication and EHRs layer are preprocessed in this phase. These missing values are managed by applying moving averaging and normalization methodology to mitigate noisy results. Subsequently, after

TABLE 1: Summary of existing literature reviews.

Paper	Classification methodology adopted	Limitations	Advantages
[12]	ML	<ol style="list-style-type: none"> <li>1. Single dataset</li> <li>2. No data fusion</li> <li>3. Only structured data</li> </ol>	Only the optimal feature selection technique was adopted
[13]	ML and AI	<ol style="list-style-type: none"> <li>1. Single dataset</li> <li>2. No data fusion</li> <li>3. Only DR image data</li> </ol>	The bloodless technique was adopted
[18]	DML and generalized linear model	<ol style="list-style-type: none"> <li>1. Single EHRs dataset</li> <li>2. No data fusion</li> </ol>	<ol style="list-style-type: none"> <li>1. Electronic health record</li> <li>2. Data fusion</li> <li>3. Feature selection</li> </ol>
[19]	AI	<ol style="list-style-type: none"> <li>1. Single dataset</li> <li>2. No data fusion</li> <li>3. Only DR image data</li> </ol>	<ol style="list-style-type: none"> <li>1. Automated software</li> <li>2. Smartphone-based DR and sight-threatening detection</li> </ol>
[20]	AI and ML	<ol style="list-style-type: none"> <li>1. Single dataset</li> <li>2. No data fusion</li> </ol>	Incorporating wearable devices and IoT to collect and manage big data
[29]	Supervised ML	<ol style="list-style-type: none"> <li>1. Single dataset</li> <li>2. No data fusion</li> </ol>	<ol style="list-style-type: none"> <li>1. Combine structured and unstructured data</li> <li>2. Feature selection</li> </ol>

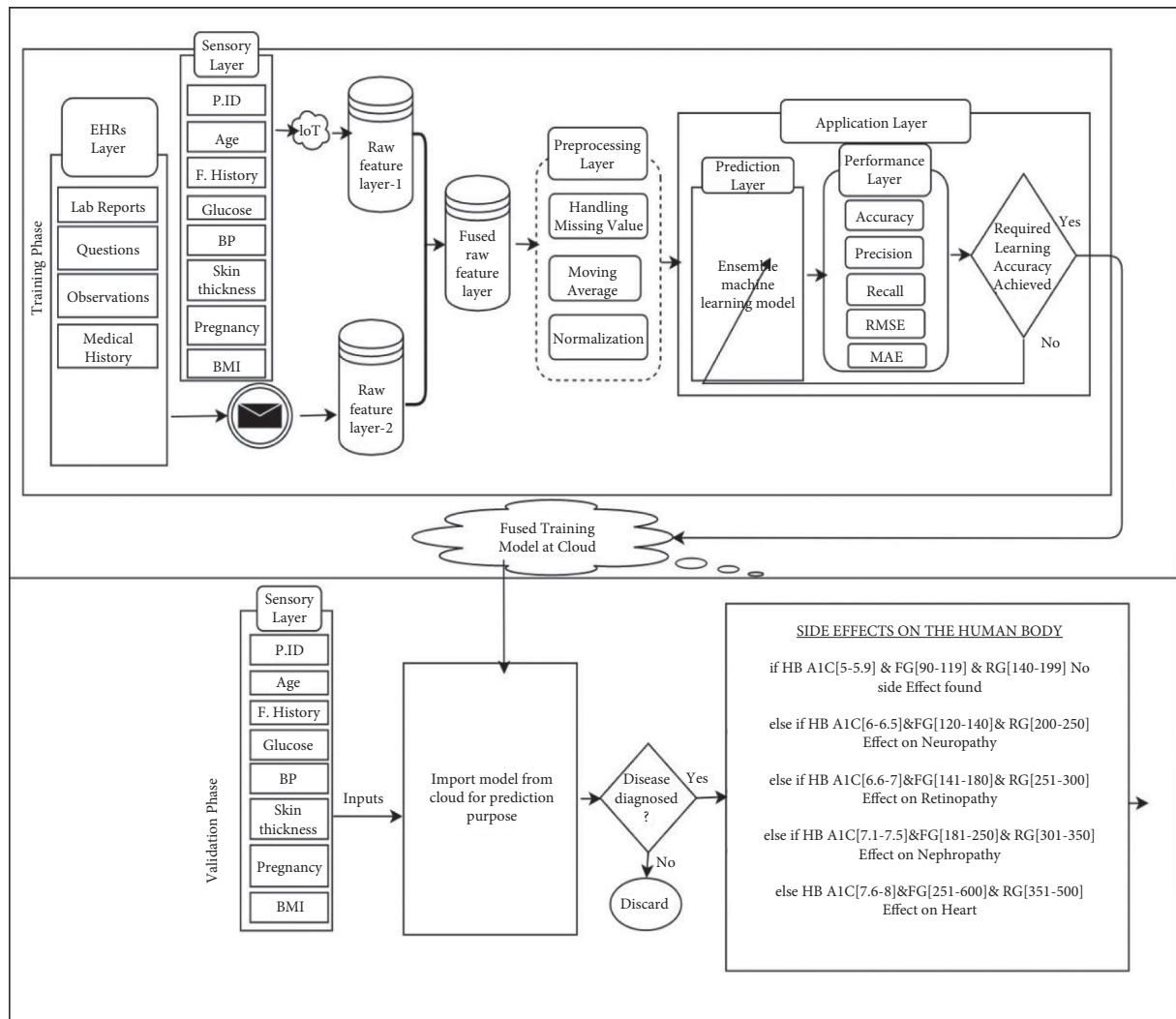


FIGURE 1: Proposed SHRS-M3DP model.

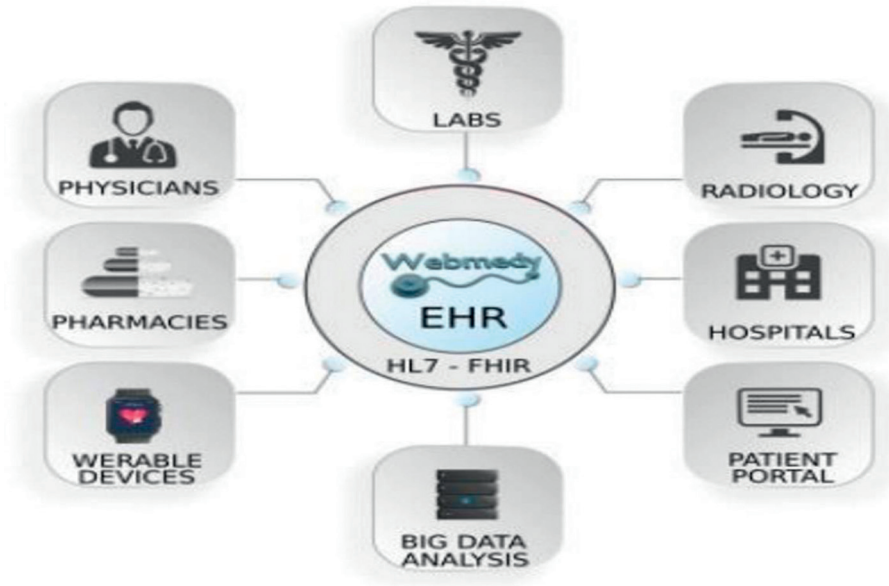


FIGURE 2: EHRs record conversion flow chart [40].

preprocessing, the fused feature dataset is forwarded to the application layer. This layer is further divided into two sublayers: (i) prediction layer and (ii) performance layer. In the prediction layer, an ensemble DML model is applied to predict multidisciplinary diabetes disease.

The ensemble deep learning combines several individual models to obtain better generalization performance of any predictive classification problem. The convergence process of the ensemble ML model is implemented in three ways: (1) max voting, (2) averaging, and (3) weighted average for classification.

The ensemble ML model used an advanced boosting technique for regularizing, limiting the overfitting issue, and producing better accuracy compared to other ML models. As a result, the response rate of the ensemble ML model is ten times faster compared to other ML models. In the ensemble ML model's boosting technique, the trained dataset is divided into multiple weak learners. The average error rate of one weak learner is updated in the next weak learner. Resultantly, the final strong learner was found to have a minimal error rate for prediction.

Ensemble deep machine learning classifier can be expressed as follows:

$$F(x) = \sum_m \alpha_m h_m(x), \quad (1)$$

where  $F(x)$  is a strong learner of ensemble classifiers,  $\alpha_m$  is weight calculated by considering the last iteration's error, and  $h_m(x)$  is a weak learner

In this way, we can achieve maximum prediction accuracy. The operational flow of the advanced boosting ensemble DML model is shown in Figure 3 (Algorithm 1).

The results are then sent to the performance layer. In this layer, data received from the previous layer is calculated. The performance layer results are evaluated based on accuracy, precision, recall, F1, root mean square error (RMSE), and

mean average error (MAE) achieved by the SHRS-M3DP model. After comparing results, "YES" indicates that our proposed SHRS-M3DP model successfully predicted diabetes disease, and "NO" means the prediction layer of the proposed SHRS-M3DP model will be modified till the learning criteria objectives are attained. After successfully training the proposed SHRS-M3DP model, the trained fused model moves to a cloud to further import and predict diabetes disease.

The validation phase comes in the last where trained fused SHRS-M3DP model is imported for prediction to authenticate whether the patient is affected with multidisciplinary diabetes disease based on the results.

The results are then sent to the next layer, called the performance layer. In this layer, the data received from the prediction layer is evaluated.

The performance layer results are then evaluated based on accuracy, precision, recall, F1, root mean square error (RMSE), and mean average error (MAE) achieved by the SHRS-M3DP model. After comparing the results, "YES" indicates that our proposed SHRS-M3DP model successfully predicted diabetes disease, and "NO" means the prediction layer of the proposed SHRS-M3DP model will be updated until the learning criteria are achieved. After successfully training the proposed SHRS-M3DP model, the trained fused model moves to a cloud to further import and predict diabetes disease.

The last phase is the validation phase. In this phase, the trained fused SHRS-M3DP model is imported for recommendation to validate whether the patient is affected with multidisciplinary diabetes disease based on the results.

## 4. Experiments

The data collected from EHRs and sensors were discussed previously in the proposed SHRS-M3DP model. In addition, the fused feature database was also discussed in the last

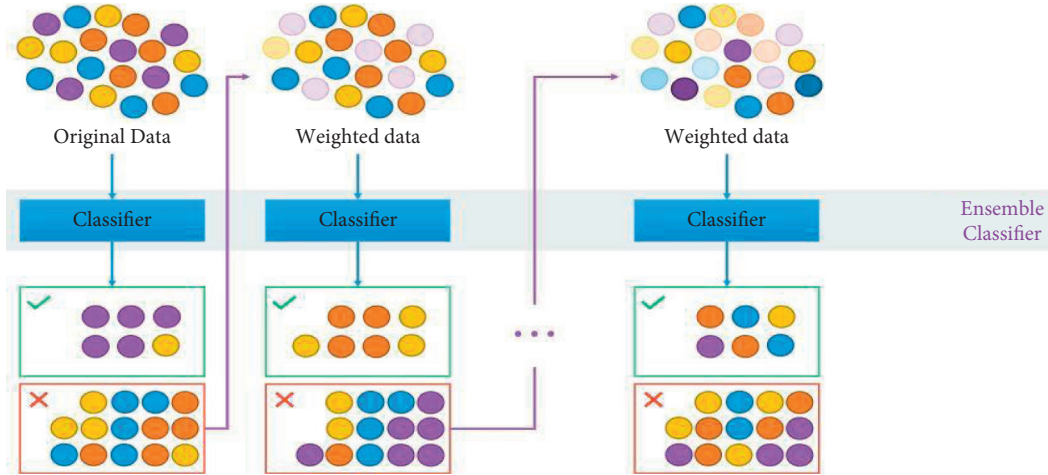


FIGURE 3: Flow chart of ensemble ML algorithm [41].

**Input:** training set  $\{(x_i - y_i)\}_{i=1}^n$  and a differentiable **loss Function**  $L(y_i, F(x))$ , number of iterations  $M$ .

**Output:** A targeted person affected by diabetes disease or safe from the diabetes disease

**Step 1:**

- (1) Initialize model with a constant value:  $F_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, \gamma)$
- (2) For  $m = 1$  to  $M$ :
  - (1) Compute so-called pseudo-residuals:  $r_{im} = -[\sigma L(y_i, F(x_i))/\sigma F(x_i)]_{F(x)=F(x_{m-1}(x))}$  for  $i = 1, \dots, n$
  - (2) Fit a base learner (or weak learner, i.e., tree)  $h_m(x)$  to (pseudo residuals), train it using the training set  $\{(x_i - y_i)\}_{i=1}^n$
  - (3) Compute multiplier  $r_m$  by solving the following one dimensional optimization problem :  $r_m = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$
  - (4) Update the model:  $F_m(x) = F_{m-1}(x) + r_m h_m(x)$
- (3) Output  $F_m(x)$

ALGORITHM 1: Ensemble ML-based diabetes prediction.

section. In this section, the proposed SHRS-M3DP model's performance is evaluated, and the results are discussed.

**4.1. Dataset.** The proposed SHRS-M3DP model is simulated with two different diabetes disease datasets: Hospital Frankfurt Germany diabetes dataset [42] and Pima Indians diabetes dataset. The Hospital Frankfurt Germany diabetes dataset consists of 2000 cases with eight features. The Pima Indians diabetes dataset consists of 768 patients with eight features. The fused features dataset for an experiment was made with the combination of features of both datasets. Both datasets' cases with some missing values are managed with proposed filtering, and normalization techniques were discussed earlier. The combined, fused dataset features are shown in Table 2. The fused features attributes, measuring units, and their ranges are also mentioned.

The fused features dataset consists of 2768 cases with eight features. A deep machine learning model cannot be utilized for the small dataset having nominal values. Therefore, all the nominal data is converted into numeric values for utilizing the ensemble deep learning model. Detailed features description of fused features is shown in Table 2.

TABLE 2: Feature information about the diabetes disease.

Sr. no.	Attribute	Unit	Ranges
1	Age	Year	01–120
2	Family history	Yes (1), no (0)	0, 1
3	Glucose	mg/Dl	37–380
4	Skin thickness	Mm	0–210
5	Blood pressure (BP)	mm Hg	90–190
6	Pregnancies	Number (0–9)	0–8
7	Insulin	uU/ml	0–764
8	BMI	Kg/m <sup>2</sup>	14–80.6
9	Diagnosis result	Positive (1), negative (0)	1, 1

**4.2. Performance Evaluation.** The experiment was carried out to indicate the proposed SHRS-M3DP model's performance for diagnosing diabetic disease. Initially, the data was collected from sensors, which were transferred through IoMT to the feature database. Similarly, the patients' data collected through lab reports, questions, observations, and medical history were converted from unstructured format to structure format for further preprocessing. After collecting the features from sensors and EHRs, both datasets' features were combined to make a rich health dataset for better prediction and recommendation of diabetes disease. Finally,

the processing module analyzed the final combined, fused feature dataset for further processing.

Furthermore, the Hospital Frankfurt Germany diabetes dataset and Pima Indians diabetes dataset were then utilized for training the diabetes disease prediction model. For evaluation purposes, the proposed ensemble deep learning model was compared with some other classifiers: SVM, LR, KNN, NB, RF, and DT. The proposed SHRS-M3DP model was used before and after the feature selection and performance was compared. The datasets were divided randomly into 80% and 20%, respectively, to train and test the models mentioned above in the proposed model.

**4.3. Evaluation Metrics.** Dissimilar evaluation metrics were used to conclude the model's overall efficiency, as shown in Table 3. With the accuracy metric's help, we can present the proposed deep learning model's overall predictive ability. In the confusion matrix, true positive (TP) and true negative (TN) determine the proposed classifier's capability to predict the absence and presence of diabetes disease. false negative (FN) and false positive (FP) identify the proposed model's total false prediction. Recall metric and precision metric calculate the sensitivity and success of the diabetes disease presented model individually.

The function measure (FM) metric is used for prediction accuracy. Root mean square error (RMSE) and mean absolute error (MAE) calculate the difference and absolute variations among the predicted and the actual values. The values of  $y_i$ ,  $\hat{y}_i$  denote the total numbers of observations of the predicted values and the actual values, respectively.

**4.4. Results.** This section presents the results of the above-mentioned proposed model and a comparison with other classifiers, respectively. The complete details of all classifiers for diabetes prediction are divided into three parts: prediction of diabetes disease, Pima Indians diabetes dataset consisting of 768 cases with eight features, Hospital Frankfurt Germany diabetes dataset consisting of 2000 patients with eight features, and finally with fused features dataset having 2768 cases with eight features as shown in Table 2, respectively.

The proposed SHRS-M3DP ensemble deep learning model prediction accuracy with other baseline classifiers is shown in Figure 4. The comprehensive explanation of each classifier before data fusion and after data fusion is as follows:

- (i) Learner regression classifier (LR): LR accuracy in dataset 1 is 74.6% for predicting diabetes disease with 786 cases. In dataset 2, LR performed better, with 77.7% accuracy with 2000 cases. Still, in the final data fusion dataset, the accuracy for LR's prediction of diabetes disease is decreased to 75.2% with 2786 cases.
- (ii) Naïve Bayes (NB): NB classifier's accuracy in dataset 1 is 72% for predicting diabetes disease with 786 cases. In dataset 2, NB performed better, with 76.5% accuracy with 2000 cases. Still, in the final data

fusion dataset, the accuracy for LR's prediction of diabetes disease is decreased to 74% with 2786 cases.

- (iii) Random forest (RF): RF classifier's accuracy in dataset 1 is 74.8% for predicting diabetes disease with 786 cases. In dataset 2, RF also performed better, with 81.2% accuracy with 2000 cases. Still, in the final data fusion dataset, the accuracy of random forest's prediction of diabetes disease is decreased to 80.5% with 2786 cases.
- (iv) K-nearest neighbor (KNN): KNN classifier's accuracy in dataset 1 is 73.3% for predicting diabetes disease with 786 cases. In dataset 2, KNN also performed better, with 77.7% accuracy with 2000 cases. In the final data fusion dataset, the accuracy of predicting diabetes disease with KNN is also increased up to 80.8% with 2786 cases.
- (v) Decision tree (DT): DT classifier's accuracy in dataset 1 is 74% for predicting diabetes disease with 786 cases. In dataset 2, DT performed better, with 83.7% accuracy with 2000 cases. In the final data fusion dataset, the accuracy of prediction of diabetes disease with DT is 84.3% with 2786 cases.
- (vi) Support vector machine (SVM): SVM classifier's accuracy in dataset 1 is 74.6% for predicting diabetes disease with 786 cases. In dataset 2, SVM performed better, with 84% accuracy with 2000 patients. In the final data fusion dataset, the accuracy for predicting diabetes disease with SVM is 84.3% with 2786 cases.

The proposed ensemble deep machine learning model performed outstandingly as compared to all the baseline classifiers. The proposed ensemble DML classifier's accuracy in dataset 1 is 72.7% for predicting diabetes disease with 786 cases. However, it is low due to the small dataset. In dataset 2, ensemble ML performed better and achieved 91% accuracy with 2000 cases, higher than all other classifiers. In the final data fusion dataset, the accuracy of prediction of diabetes disease with the proposed ensemble ML model is 99.6%, having a minimal error rate.

The summary of performance metrics, accuracy, precision, recall, F1, root mean square error, and mean absolute error of selected datasets individually and data fusion datasets, is presented in Tables 4–6.

Different DML classifiers are compared with the proposed model by various evaluation metrics, as shown in Table 4. In this experiment, only Pima Indians diabetes dataset is considered, without feature selection technique. The performance of each metric on a given dataset is precisely shown in Table 4. The proposed model's overall performance is less compared to the other classifiers due to the small dataset and the absence of feature selection technique.

In Table 5, only the Hospital Frankfurt Germany diabetes dataset is considered, without feature selection technique. Different DML classifiers are compared with the proposed model by various evaluation metrics, as shown in Table 5.

TABLE 3: Performance metrics.

Name of metric	Description	Equation #
Accuracy (Acc)	$(TP + TN) / (TP + TN + FP + FN)$	(a)
Precision (Pre)	$TP / (TP + FP)$	(b)
Recall (Rec)	$TP / (TP + FN)$	(c)
F1-measure (FM)	$((2 * Pre * Rec) / (Pre + Rec))$	(d)
RMSE	$\sqrt{1/N \sum_{i=1}^N (y_i - \hat{y}_i)^2}$	(e)
MAE	$\sqrt{1/N \sum_{i=1}^n  y_i - \hat{y}_i }$	(f)

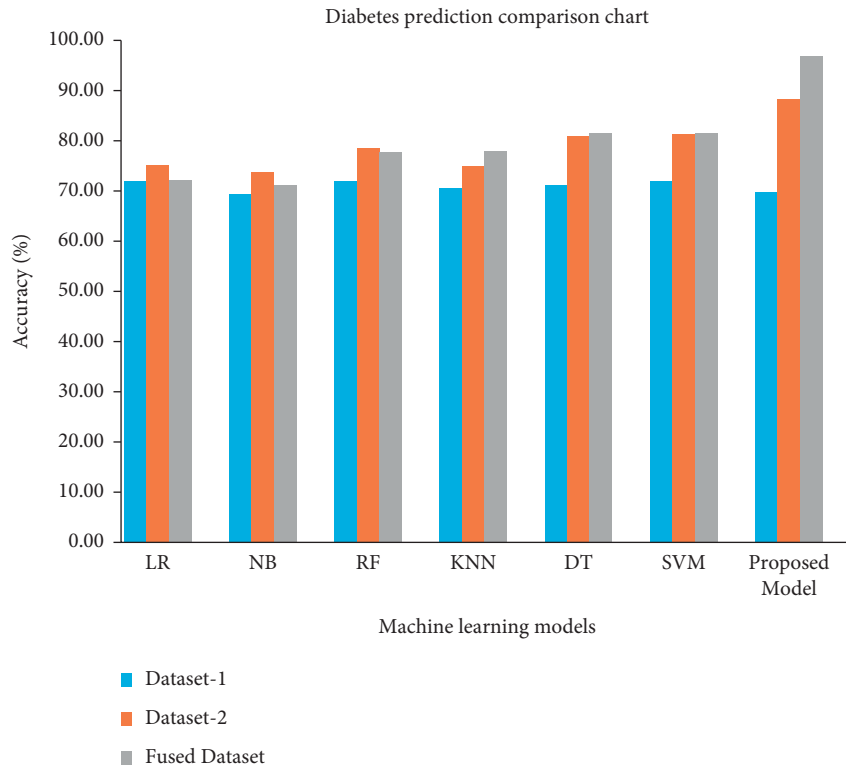


FIGURE 4: Accuracy of models before and after data fusion.

TABLE 4: Comparison results of the proposed model with other classifiers before data fusion on the Pima Indians diabetes dataset.

Classifier model	Acc (%)	Pre (%)	Rec (%)	FM (%)	RMSE	MAE
Logistic regression	74.68	0.68	0.52	0.59	0.25	0.50
Naïve Bayes	72.08	0.62	0.52	0.57	0.28	0.53
Random forest	74.68	0.69	0.50	0.58	0.25	0.50
K-nearest neighbors	73.38	0.67	0.48	0.56	0.27	0.52
Decision tree	74.03	0.63	0.63	0.63	0.26	0.51
Support vector machine	74.68	0.70	0.48	0.57	0.25	0.50
<b>*Proposed model</b>	72.73	0.63	0.56	0.59	0.27	0.52

TABLE 5: Comparison results of the proposed model with other classifiers before data fusion on the Hospital Frankfurt Germany diabetes dataset.

Classifier model	Acc (%)	Pre (%)	Rec (%)	FM (%)	RMSE	MAE
Logistic regression	77.75	0.71	0.58	0.64	0.22	0.47
Naïve Bayes	76.50	0.67	0.61	0.64	0.24	0.48
Random forest	81.25	0.75	0.69	0.71	0.19	0.43
K-nearest neighbors	77.75	0.71	0.59	0.65	0.22	0.47
Decision tree	83.75	0.73	0.83	0.78	0.16	0.40
Support vector machine	84.00	0.79	0.73	0.76	0.16	0.40
<b>*Proposed model</b>	91.00	0.89	0.84	0.86	0.09	0.30



TABLE 6: Comparison results of the proposed model with other classifiers after data fusion on a fused dataset.

Classifier model	Acc (%)	Pre (%)	Rec (%)	FM (%)	RMSE	MAE
Logistic regression	75.27	0.67	0.55	0.61	0.25	0.50
Naïve Bayes	74.01	0.64	0.55	0.59	0.26	0.51
Random forest	80.51	0.75	0.65	0.70	0.19	0.44
K-nearest neighbors	80.87	0.77	0.63	0.70	0.19	0.44
Decision tree	84.30	0.77	0.78	0.77	0.16	0.40
Support vector machine	84.30	0.81	0.71	0.76	0.16	0.40
<b>*Proposed model</b>	99.64	1.00	0.99	0.99	0.00	0.06

TABLE 7: Summary of comparison using diabetes datasets with existing methods.

Sr. no.	Author/year	Fusion method	Classifier/compared with classifier	Overall accuracy
1	Mohebbi et al. [10]/2017	—	CNN/LR, MLP	77.5%
2	Sneha and Gangil [11]/2019	Yes/—	SVM/RF, NB, KNN, DT	77.7%
3	Nguyen et al. [16]/2019	—	Ensemble ML/—	82.2%
4	Aminah and saputro [9]/2019	—	KNN/SVM	85.6%
5	Sah and sarma [12]/2018	—	AI	91.6%
6	Dinh et al. [15]/2019	—	XGB/LR, SVM, RF, weight ensemble model	95.7%
7	<b>Proposed method</b>	Yes/yes	Ensemble deep machine learning classifier/LR, NB, DT, KNN, RF, SVM	99.6%

The proposed model’s overall performance is more outstanding compared to the other classifiers due to the large dataset, as shown in Figure 4. Therefore, our proposed deep machine learning model can achieve more accurate results concerning a higher dataset ratio. In this experiment, the proposed model achieved 91% accuracy, much higher compared to other DML classifiers. On the other hand, in this experiment, RMSE and MAE are also very low compared with different DML classifiers’s RMSE and MAE.

In Table 6, both diabetes datasets are being considered with the data fusion technique. Different DML classifiers are compared with the proposed model by various evaluation metrics, as shown in Table 3. The performance of each metric on a given dataset is precisely shown in Table 6. The proposed model’s overall performance is much higher compared to the other classifiers due to the fused technique by making a rich healthcare dataset. Our proposed model performed outstandingly due to the data fusion technique and produced extraordinary results. In this experiment, the proposed model achieved 99.64% accuracy, much higher compared to other DML classifiers. In this experiment, RMSE is 0%, and MAE is only 0.06%. As shown in Table 6, all other classifiers produce results less than 85%.

Our proposed model accurately performed well with the help of data fusion technique. Our proposed ensemble DML model has achieved higher accuracy compared to other studies done in the recent past. The details of recent studies on diabetes with their authors are also summarized and shown in Table 7. The significant contribution for higher accuracy in our model is due to data fusion. In this way, we have made a rich healthcare dataset for the prediction of multidisciplinary diabetes disease. In this way, we have achieved higher accuracy compared to other studies, which is 99.6%.

## 5. Conclusion

The prediction of human diseases, particularly multidisciplinary diabetes, is challenging for better and timely treatment. A multidisciplinary diabetes illness is a life-threatening disease worldwide which attacks major essential human body parts. A proposed SHRS-M3DP model is presented to predict and recommend multidisciplinary diabetes disease in the patients quickly and efficiently. The ensemble deep ML model and data fusion technique are used for fast response and better accuracy rate. The proposed model efficiently predicted and recommended whether the patient is a victim of multidisciplinary diabetes disease or not. The proposed SHRS-M3DP model can also identify the effect of human body parts: Neuropathy, Retinopathy, Nephropathy, or Heart. The proposed SHRS-M3DP model simulation is made by using Python language. Finally, the study of this research concluded that the proposed SHRS-M3DP model’s overall performance is 99.6%, which is outstanding compared to previously published approaches.

*5.1. Contribution.* Many recommendation systems for healthcare have already been proposed in recent researches. The significant contribution of this research is to enrich the healthcare dataset for the best prediction of multidisciplinary diabetes disease. We have collected the patients’ data through wearable sensors and EHRs in the textual record form of each patient. After collecting the records of each patient, essential data from both ends are fused to enrich the healthcare dataset. The ensemble deep learning approach works accurately and produces better results in larger healthcare datasets. Finally, we have developed a better recommendation system by collecting patients’ records and applying an ensemble machine learning approach for

accurate and timely prediction and recommendation for multidisciplinary diabetes disease patients. The overall performance of our recommendation system is 99.6%. In this way, future academic research and practices will be helpful for new researchers in this medical field, especially for automated prediction and recommendation systems for human diseases.

**5.2. Future Work.** The proposed SHRS-M3DP recommendation system achieved overall good performance. However, there is still a need to work for a better generalized efficient prediction and recommendation system for all human diseases. The complexity of the deep ensemble algorithm will also be considered in the near future for accurate and quick results of this algorithm.

### Data Availability

The data used in this paper can be requested from the corresponding author.

### Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this work.

### Authors' Contributions

B. I., M. A. K., and T. A. K. collected data from different resources. M. A. K. and T. A. K. performed formal analysis and Simulation, M. A. K., B. I., M. A., and M. S. D. contributed to writing—original draft preparation; M. A. K. and M. A. contributed to writing—review and editing. M. A. K. and S. A. supervised the paper. B. I., M. S. D., and T. A. K. drafted pictures and tables. M. A. K. and M. A. performed revision and improved the quality of the draft. All authors have read and agreed to the published version of the manuscript.

### References

- [1] N. S. Artzi, S. Shilo, E. Hadar et al., "Prediction of gestational diabetes based on nationwide electronic health records," *Nature Medicine*, vol. 26, no. 1, pp. 71–76, 2020.
- [2] V. Raman, P. Then, and P. Sumari, "Proposed retinal abnormality detection and classification approach: computer-aided detection for diabetic retinopathy by machine learning approaches," in *Proceedings of the 2016 8th IEEE International Conference on Communication Software and Networks*, Beijing, China, June 2016.
- [3] J. Ramsingh and V. Bhuvaneshwari, "An efficient map reduce-based hybrid NBC-TFIDF algorithm to mine the public sentiment on diabetes mellitus—a big data approach," *Journal of King Saud University Computer and Information Sciences*, 2018, In press.
- [4] WHO, *Global Health Risks Report 2016*, <https://www.who.int/diabetes/global-report>, World Health Organisation, Geneva, Switzerland, 2020, <https://www.who.int/diabetes/global-report>.
- [5] N. Sharma and A. Singh, "Diabetes detection and prediction using machine learning/IoT: a survey," *Communications in Computer and Information Science*, vol. 18, pp. 471–479, 2018.
- [6] S. Afzali and O. Yildiz, "An effective sample preparation method for diabetes prediction," *The International Arab Journal of Information Technology*, vol. 15, no. 6, 2018.
- [7] N. Theera-Umpon, I. Poonkasem, S. Auephanwiriyaikul, and D. Patikulasila, "Hard exudate detection in retinal fundus images using supervised learning," *Neural Computing & Applications*, vol. 32, 2020.
- [8] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting diabetes mellitus with machine learning techniques," *Frontiers in Genetics*, vol. 9, 2018.
- [9] M. Alghamdi, M. Al-Mallah, S. Keteyian, C. Brawner, J. Ehrman, and S. Sakr, "Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: the Henry Ford exercise testing (FIT) project," *PLoS One*, vol. 12, 2017.
- [10] M. Hammad, M. H. Alkinani, B. B. Gupta, and A. A. A. El-Latif, "Myocardial infarction detection based on the deep neural network on imbalanced data," *Multimedia Systems*, vol. 27, 2021.
- [11] A. Alghamdi, M. Hammad, H. Ugail et al., "Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities," 2020, <https://arxiv.org/abs/1906.09358>.
- [12] E. M. Abou-Nassar, A. M. Iliyasu, P. M. El-Kafrawy, O.-Y. Song, A. K. Bashir, and A. A. A. El-Latif, "DITrust chain: towards blockchain-based trust models for sustainable healthcare IoT systems," *IEEE Access*, vol. 8, pp. 111223–111238, 2020.
- [13] W. Książek, M. Hammad, P. Pławiak, U. Rajendra Acharya, and R. Tadeusiewicz, "Development of novel ensemble model using stacking learning and evolutionary computation techniques for automated hepatocellular carcinoma detection," *Biocybernetics and Biomedical Engineering*, vol. 40, pp. 1512–1524, 2020.
- [14] V. Chandran, M. G. Sumithra, A. Karthick et al., "Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images," *BioMed Research International*, vol. 2021, Article ID 5584004, 2021.
- [15] I. A. Elgendy, W.-Z. Zhang, H. He, B. B. Gupta, and A. A. Abd El-Latif, "Joint computation offloading and task caching for multi-user and multi-task MEC systems: reinforcement learning-based algorithms," *Wireless Networks*, vol. 27, no. 3, pp. 2023–2038, 2021.
- [16] M. Khayyat, I. A. Elgendy, A. Muthanna, A. S. Alshahrani, S. Alharbi, and A. Koucheryavy, "Advanced deep learning-based computational offloading for multilevel vehicular edge-cloud computing networks," *IEEE Access*, vol. 8, pp. 137052–137062, 2020.
- [17] Y. Liu, J. Peng, J. Kang, A. M. Iliyasu, D. Niyato, and A. A. A. El-Latif, "A secure federated learning framework for 5G networks," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 24–31, 2020.
- [18] A. Sedik, M. Hammad, F. E. Abd El-Samie, B. B. Gupta, and A. A. A. El-Latif, "Efficient deep learning approach for augmented detection of Coronavir disease," *Neural Computing & Applications*, vol. 8, 2021.
- [19] I. A. Elgendy, A. Muthanna, M. Hammoudeh, H. Shaiba, D. Unal, and M. Khayyat, "Advanced deep learning for resource allocation and security aware data offloading in industrial mobile edge computing," *Big Data*, vol. 9, pp. 1–14, 2021.

- [20] R. Aminah and A. H. Saputro, "Diabetes prediction system based on iridology using machine learning," in *Proceedings of the 2019 6th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*, Semarang, Indonesia, September 2019.
- [21] A. Mohebbi, T. B. Aradottir, A. R. Johansen, H. Bengtsson, M. Fraccaro, and M. Mørup, "A deep learning approach to adherence detection for type 2 diabetics," in *Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jeju, Korea, July 2017.
- [22] N. Sneha and T. Gangil, "Analysis of diabetes mellitus for early prediction using optimal features selection," *Journal of Big Data*, vol. 6, 2019.
- [23] P. Sah and K. K. Sarma, "Bloodless technique to detect diabetes using the soft computational tool," in *Ophthalmology IGI Global*, Hershey, PA, USA, 2018.
- [24] M. Arora and M. Pandey, "Deep neural network for diabetic retinopathy detection," in *Proceedings of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pp. 189–193, Faridabad, India, February 2019.
- [25] S. Karthikeyan, P. Sanjay Kumar, R. J. Madhusudan, S. Sundaramoorthy, and P.-K.-K. Namboori, "Detection of multi-class retinal diseases using artificial intelligence: an expeditious learning using deep CNN with minimal data," *Biomedical and Pharmacology Journal*, vol. 12, 2019.
- [26] A. Dinh, S. Miertschin, A. Young, and S. D. Mohanty, "A data-driven approach to predicting diabetes and cardiovascular disease with machine learning," *BMC Medical Informatics and Decision Making*, vol. 19, 2019.
- [27] B. P. Nguyen, H. N. Pham, H. Tran et al., "Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records," *Computer Methods and Programs in Biomedicine*, vol. 182, 2019.
- [28] R. Rajalakshmi, R. Subashini, R. M. Anjana, and V. Mohan, "Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence," *Eye*, vol. 32, 2018.
- [29] M. Chen, J. Yang, J. Zhou, Y. Hao, J. Zhang, and C.-H. Youn, "5G-smart diabetes: toward personalized diabetes diagnosis with healthcare big data clouds," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 16–23, 2018.
- [30] P. Choudhary, S. De Portu, A. Arrieta, J. Castañeda, and F. M. Campbell, "Use of sensor-integrated pump therapy to reduce hypoglycemia in people with type 1 diabetes: a real-world study in the UK," *Diabetic Medicine*, vol. 36, 2019.
- [31] A. Steinert, E. Steinhagen-Thiessen, and M. Haesner, "App-basiertes selbstmonitoring bei Typ-2-diabetes," *Zeitschrift für Gerontologie und Geriatrie*, vol. 50, no. 6, pp. 516–523, 2017.
- [32] R. Davoodi and M. H. Moradi, "Mortality prediction in intensive care units (ICUs) using a deep rule-based fuzzy classifier," *Journal of Biomedical Informatics*, vol. 79, 2018.
- [33] L. B. Fazlic, A. Hallawa, A. Schmeink, A. Peine, L. Martin, and G. Dartmann, "A novel NLP-fuzzy system prototype for information extraction from medical guidelines," in *Proceedings of the 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, Croatia, May 2019.
- [34] C. W. Song, H. Jung, and K. Chung, "Development of a medical big-data mining process using topic modeling," *Cluster Computing*, vol. 22, 2019.
- [35] E. A. Bernal, X. Yang, Q. Li et al., "Deep temporal multimodal fusion for medical procedure monitoring using wearable sensors," *IEEE Transactions on Multimedia*, vol. 20, 2018.
- [36] M. Muzammal, R. Talat, A. H. Sodhro, and S. Pirbhulal, "A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks," *Information Fusion*, vol. 53, pp. 155–164, 2020.
- [37] H. F. Nweke, Y. W. Teh, U. R. Alo, and G. Mujtaba, "Analysis of multi-sensor fusion for mobile and wearable sensor-based human activity recognition," in *Proceedings of the International Conference on Data Processing and Applications 2018*, Guangzhou, China, May 2018.
- [38] H. Kaur and V. Kumari, "Predictive modelling and analytics for diabetes using a machine learning approach," *Applied Computing and Informatics*, vol. 17, 2019.
- [39] F. Ali, S. El-Sappagh, S. M. R. Islam et al., "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Information Fusion*, vol. 63, 2020.
- [40] "Electronic Health Record," 2021, <https://webmedy.com/ehr/>.
- [41] G. S. Engine, "Boosting (machine learning)," 2020, [https://en.wikipedia.org/wiki/Boosting\(machine\\_learning\)#/media/File:Ensemble\\_Boosting.svg](https://en.wikipedia.org/wiki/Boosting(machine_learning)#/media/File:Ensemble_Boosting.svg).
- [42] H. F. Germany, "Diabetes data set," 2020, <https://www.kaggle.com/johndasilva/diabetes>.