International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

An Ensemble Deep Learning Model for Diabetes Disease Prediction

Selma Aouamria[1], Djalila Boughareb*[2], Mohamed Nemissi[3], Zineddine Kouahla[4], Hamid Seridi[4]

Submitted:14/03/2024 **Revised**: 29/04/2024 **Accepted**: 06/05/2024

Abstract: Diabetes remains a significant health challenge with serious consequences if left undiagnosed or untreated. Addressing the issues of accurately labeled data, outliers, small number of samples and missing information in clinical datasets is crucial for effective diabetes prediction. Despite various efforts, there is still room for improvement in the accuracy of machine and deep learning methods for early diabetes detection. In this study, we propose a novel approach that integrates three proven deep learning models—Long Short Term Memory (LSTM), Deep Neural Networks (DNN), and Convolutional Neural Networks (CNN)—using a soft voting classifier to enhance predictive performance. Additionally, we employ data fusion to effectively address the challenge of small datasets. Our model demonstrated impressive accuracy rates when evaluated on the Pima Indian Diabetes Dataset (PIDD), the Frankfurt Hospital Germany Diabetes Dataset (FHGDD), and a combined dataset: 85.9% on PIDD, 98.0% on FHGDD, and 99.81% on the combined dataset. These results outperform those of individual classifiers, highlighting the effectiveness of our method in diabetes prediction.

Keywords: *Deep Learning, Diabetes, Ensemble Learning, Healthcare, LSTM*

1. Introduction

Diabetes, a metabolic disorder characterized by elevated blood sugar levels, poses significant health risks worldwide [1]. It manifests in three main forms: type 1, type 2, and gestational diabetes [2]. Type 1, often diagnosed in childhood, involves the body's immune system attacking the pancreas, resulting in insulin deficiency and potential organ damage. Type 2, also known as adult-onset diabetes, poses serious complications affecting various organs, particularly the kidneys, eyes, and nerves. Gestational diabetes occurs during pregnancy but typically resolves after childbirth, necessitating careful monitoring for both maternal and fatal health.

In diabetes prediction, the primary challenges include enhancing prediction accuracy, minimizing human error and bias, and handling large datasets. Traditional methods of diabetes detection frequently lack effectiveness; however, the emergence of advanced technologies, especially deep learning, has become essential in health research. Deep learning can address these challenges by automatically learning relevant features and modelling complex relationships. In addition, DL efficiently processes large datasets for real-time analysis, leading to better early detection of diabetes [3].

Convolutional Neural Networks (CNN), Long Short Term Memory (LSTM), and Multilayer Perceptron (MLP). The three baseline deep learning classifiers undertake training and testing, and the final prediction is determined through the soft voting method.

2. Recognizing the scarcity of diabetes-related data and the substantial data requirements of deep learning models, we explore the potential of data fusion techniques to mitigate this challenge.

The paper is structured to ensure a comprehensive exploration of the research domain. It begins with an Introduction, setting the stage by outlining the context and objectives of the study. Section 2 provides a thorough review of existing literature and studies relevant to the research topic. Section 3 elucidates the Materials and research methodology utilized; Section 4 introduces our novel methodology. Subsequently, Section 5 delves into Results, where the evaluation process is examined. Section 6 discusses the findings. Finally, Section 7, summarizes the key findings suggesting potential avenues for future investigation

2. Related works

Recent research has extensively explored the identification of diabetes patients through machine learning and deep learning models, aiming to boost prediction accuracy by combining individual models into ensemble systems ([8]; [9]; [10]; [11]; [12]; [13]; [14]; [15]; [16]; [17]; [18]; [19]; [20]; [21] [22]; [23]; [24]; [25]; [20]; [26]). For instance, A. Doğru et al. [9] recently investigated diabetes detection

[1] LabSTIC Laboratory, University of 8 Mai 1945 Guelma, BP 401 Guelma 24000, Guelma, Algeria
ORCID ID : 0009-0002-1035-5457
[2] LabSTIC Laboratory, University of 8 Mai 1945 Guelma, BP 401 Guelma 24000, Guelma, Algeria.
ORCID ID : 0000-0002-3432-2548
[3] LabSTIC Laboratory, University of 8 Mai 1945 Guelma, BP 401 Guelma 24000, Guelma, Algeria
ORCID ID : 0000-0001-9476-2513
[4] LabSTIC Laboratory, University of 8 Mai 1945 Guelma, BP 401 Guelma 24000, Guelma, Algeria
ORCID ID : 0000-0001-8105-9810
[5] LabSTIC Laboratory, University of 8 Mai 1945 Guelma, BP 401 Guelma 24000, Guelma, Algeria
ORCID ID : 0000-0002-0236-8541
** Corresponding Author Email: boughareb.djalila@univ-guelma.dz*

by introducing a novel super ensemble learner model consisting of four base learners (logistic regression, decision tree, random forest, and gradient boosting) along with a meta-learner (support vector machines). Their study utilized datasets related to early-stage diabetes risk prediction, including Pima and Diabetes 130-US hospital datasets. Experimental results demonstrated performance scores of 99.6% for the early-stage diabetes risk prediction dataset, 92% for the PIMA dataset, and 98% for the diabetes 130-US hospital dataset, respectively.

Meanwhile, L. Ismail et al. [10] explored 35 different classifiers, proposing a Bagging-based Logistic Regression (Bagging-LR) approach for Type 2 diabetes prediction. Bagging-LR emerged as the optimal detector, achieving diabetes mellitus detection in approximately 0.016 minutes while utilizing only five key features to boost classification accuracy to 99%. Additionally, S. Kumari et al. [11] introduced a soft voting ensemble classifier comprising three machine learning algorithms: Random Forest, Logistic Regression, and Naïve Bayes, for diabetes classification. On the Pima Indians diabetes dataset, their proposed ensemble classifier achieved the highest accuracy, precision, recall, and F1-score values, reaching 79.04%, 73.48%, 71.45%, and 80.6%, respectively.

Also, A. Mahabub. [12] explored eleven renowned machine-learning algorithms, including K-Neighbors, Ada Boost, Decision Tree, Random Forest, Support Vector Machine, Gradient Boosting, Logistic Regression, Multi-Layer Perceptron, Multinomial Naïve Bayes, Extreme Gradient Boosting, and Gaussian Naïve Bayes, for early-stage diabetes detection on the Pima Indians Diabetes Database (PIDD). The top three machine-learning algorithms are selected and integrated into the Ensemble Voting Classifier to maximize accuracy. Furthermore, S. Bashir et al. [13] introduced an ensemble model comprising three types of decision trees (CART, ID3, and C4.5), achieving an accuracy of 76.5%. Singh et al. [27] employed an ensemble of various machine learning models, including SVM, NN, DT, XGBoost, and RF, on the PIDD Dataset, yielding an accuracy of 95% in predicting diabetes.

Using medical data from the Henry Ford Exercise Testing (FIT) project, M. Alghamdi et al. [14] developed an Ensembling-based predictive model utilizing the "Vote" technique, combining three decision tree classification methods (RF, NB Tree, and LMT) for predicting incident diabetes. They employed multiple methodologies to identify potential diabetes predictors, and the ensemble voting technique boosted prediction accuracy to an AUC of 0.922.

Similarly, H. B. Kibria et al. [15] applied six machine learning algorithms (ANN, RF, SVM, LR, AdaBoost, XGBoost) and a weighted soft voting classifier of the two best ML algorithms (RF, XGB) to identify potential diabetes predictors using the PIDD dataset, achieving the highest accuracy of 90%. In their pursuit of improved performance and accuracy, M. Alehegn et al. [16] embarked on further research to devise an ensemble hybrid model utilizing KNN, NB, RF, and the J48 approach to predict early-stage diabetes. H. Naz et al. [17] Conducted a comparative analysis of various machine learning and data mining techniques for early-stage diabetes prediction. Their evaluation on a diabetes dataset involved four ML techniques: ANN, NB, DT, and Deep Learning (DL), with DL achieving the highest accuracy of 98.07% among all classifiers.

In addition, E. P. Prakash et al. [18] aimed to forecast diabetes effectively in patients through online monitoring, employing a neural network-based ensemble voting classifier. This approach integrates several neural network models, including the artificial neural network model, recurrent neural network model, Deep Belief Network, Multilayer Perceptron, and Radial Basis Function. Their simulation results showcased the efficacy of the neural network ensemble voting method over existing bagging, stacking, and boosting ensemble models.

Z. Mushtaq et al. [19] Utilized hyper-tuned machine learning techniques in conjunction with an ensemble technique based on a voting method for diabetes prediction. They employed six classifiers, namely SVM, NB, KNN, LR, RF, and Gradient Boosting Classifier, for the classification of PIDD.

Following the dataset balancing process, the classifiers' accuracies were assessed. Random Forest exhibited superior performance compared to logistic regression, Support Vector Machine, k-nearest neighbors, Naive Bayes, and Gradient Boosting Classifier algorithms, achieving an accuracy of 80.7%. Furthermore, a voting classifier was examined, demonstrating accuracies of 81.7% on the original dataset and 81.5% on the balanced dataset.

S. Ramesh et al. [20] Proposed a deep neural network (DNN) integrated with Restricted Boltzmann Machine (RBM) for diabetes diagnosis, utilizing the PIMA Diabetes dataset. Their technique exhibited promising performance, achieving accuracy and recall values of 77% and 54%, respectively. Furthermore, their comparison revealed that deep learning models surpassed the rough set theory model in terms of precision. Furthermore, S. Ramesh et al. [28] presented a deep learning strategy for classifying diabetes using the Recurrent Deep Neural Network (RDNN) method. Their approach, applied to the Pima Indian diabetes dataset, yielded an accuracy of 81%.

In [21] , the authors conducted a comparative analysis of traditional machine learning and deep learning techniques for diabetes prediction.

K. Kannadasan et al. [22] Opted for a Deep Neural Network (DNN) framework for diabetes data classification. They utilized stacked autoencoders to extract features from the dataset and employed a softmax layer for classification. Additionally, the network underwent fine-tuning using back propagation with the training data. Their DNN methodology achieved a classification accuracy of 86.26% on the PIDD dataset.

S. K. Kalagotla et al. [23] Proposed a novel stacking strategy based on MLP, SVM, and LR for diabetes mellitus diagnosis. They combined intelligent models using the stacking approach, resulting in enhanced model performance. Their unique stacking method outperformed other models when compared to AdaBoost.

M. T. García-Ordás et al. [24] Developed a pipeline for forecasting diabetes patients using deep learning techniques. Their approach integrates a Variational AutoEncoder (VAE) for data augmentation, a sparse AutoEncoder (SAE) for feature enhancement, and a convolutional neural network (CNN) for classification.

In their study, G. Swapna et al. [25], a hybrid mellitus diagnosis system was proposed, combining two deep learning methods, LSTM and CNN. They achieved promising results, with an accuracy of 95.1%, demonstrating the effectiveness of deep learning techniques.

B. Ihnaini et al. [26] Introduced a smart recommendation system for diabetes prediction by employing an ensemble machine learning approach and fusing the Hospital Frankfurt Germany diabetes dataset with the Pima Indians diabetes dataset (PIDD). Their method leverages the fusion of two datasets to enhance predictive performance.

More recently, H. Qi et al. [29] introduced KFPredict, a novel soft-voting ensemble learning framework that combines multi-input models with diverse machine learning techniques for diabetes prediction. The framework utilizes a neural network model called KF_NN, which integrates essential features using a decision tree-based selection recursive feature elimination algorithm and correlation coefficient method. Three machine learning algorithms, including support vector machine, random forest, and k-nearest neighbors, are integrated with KF_NN using a soft voting technique. The system demonstrated notable performance on the Pima diabetes dataset, achieving an accuracy rate of 93.5%.

While numerous researchers have proposed various ML algorithms and ensemble techniques such as SVM, DT, RF, K-NN, MLP, LR, AdaBoost, and Bagging for disease prediction,

their accuracy often remains suboptimal. It is evident that DL-based diabetes prediction generally outperforms traditional ML techniques in terms of performance. Some researchers have explored DL and hybrid deep learning neural networks, such as MLP [17], CNN [30], CNN-LSTM [31], and CNN-BiLSTM [32], for early-stage diabetes prediction.

Our work introduces novel contributions by assembling CNN, LSTM, MLP via soft voting techniques to enhance the results and accuracy of diabetes detection. Notably, this is the first instance of utilizing an ensemble of deep-learning models to address this issue. Moreover, many articles typically rely on a single small dataset, such as PIDD, to evaluate classification methods.

However, drawing conclusions solely from one dataset may not be applicable to others. Therefore, it is crucial to assess performance using multiple standard datasets. In our study, we utilized three datasets: PIDD, FGDD, and a fusion of these two datasets. By leveraging data fusion, we can extract complementary information, mitigate data insufficiency, and enhance the performance of our system in forecasting diabetes disease data into positive and negative classes.

3. Materials and Methods

This section outlines the materials and methods utilized in the experiment.

3.1. Dataset Description

We utilized two widely recognized public datasets: the "Pima Indians diabetes dataset" [33], and the "Hospital Frankfurt Germany diabetes dataset" [34]. These datasets collectively feature 9 attributes per sample, with 768 instances in the Pima Indians dataset and 2000 instances in the Hospital Frankfurt Germany dataset. Each instance includes a binary label indicating the patient's diabetic status (0 for non-diabetic, 1 for diabetic). To address the issue of small dataset size and meet the requirements of deep learning models, we amalgamate these two datasets into a unified dataset comprising 2768 instances.

Table 1 offers a comprehensive overview of the dataset

Table 1. Description of Attributes in Diabetes Datasets

r. no.	Attributes	Description	Range
1.	Pregnancy	Number of times a participant is pregnant	0-17
2.	Glucose	Plasma glucose concentration a 2 h in an oral glucose tolerance test It consists of Diastolic blood pressure	0-199
3.	Diastolic Blood pressure	(when blood exerts into arteries between heart) (mm Hg)	0-122
4.	Skin Thickness	Triceps skinfold thickness (mm). It concluded by the collagen content	0-99
5.	Serum Insulin	2-Hour serum insulin (mu U/ml)	0-846
6.	BMI	Body mass index (weight in kg/ (height in m) 2) Diabetes pedigree Function	0-67.1
7.	Diabetes pedigree Function	An appealing attributed used in diabetes prognosis	0.078-2.42
8.	Age	Age of participants Diabetes class variable, Yes represent	21-81
9.	Outcome	the patient is diabetic and no represent patient is not diabetic	Yes/No

attributes

alongside their respective value ranges, while a breakdown of the class distribution across the three datasets is presented in a separate table (table 2).

3.2. Multilayer Perceptron (MLP)

MLP, [35], are deep Artificial Neural Networks (ANN) with multiple hidden layers between input and output layers. Every layer of an MLP is fully connected and made up of a number of neurons and an activation function. When using binary classification, the input is a connection between feature spaces that only produces one output. The MLP employed in this approach utilizes the traditional back-propagation method with sigmoid and ReLU activation functions. The sigmoid function is

used in the last layer because this approach is developed to solve a binary classification problem. MLP techniques achieve promising accuracy in diabetes prediction although they face some challenges like gradient and vanishing problems, overfitting, high computational, and cost complex processes, etc. that have adverse effects on performance [36].

3.3. Convolutional Neural Networks (CNN)

CNN was originally developed for image processing [37], but the model is also very efficient for structured data.

CNN is a feed-forward deep neural network comprised of an input layer, convolutional layers, pooling layers, and a fully connected layer (output). The convolutional layer will convolute the input using a number of filters, where a smaller matrix represents each filter. During the pooling process, the size is reduced, while maintaining important features. Therefore, the prime task of convolutional and pooling layers is primarily

to extract features, whereas the major aim of fully connected layers is often to output the data from feature maps jointly and then provide them to the final layers. As a result, the performance of the network is increased and over-fitting is avoided [30]. The most notable benefit of CNN is that it extracts features automatically. It can additionally display the correlation between the items for the input data [38].

3.4. Long Short Term Memory Networks (LSTM)

LSTM is a type of RNN made up of feedback connections [39].

LSTM models can easily handle a lengthy input data series. A cell, an entry g e, an output gate, and a forgotten gate make up a typical LSTM system. The three gates keep track of data flow in and out of the cell, and the cell recalls values at any time [40].

Classifying, analysing, and estimating time series data, LSTM networks are well adapted due to the ability to delay the interval of uncertain events in a time series. LSTM has recently attracted much interest in disease forecasting and diabetes prediction such as in [31] where

Table 2 . The class distribution across the three datasets

Dataset	Class 1 (diabetic)	Class 2 (non-diabetic)
PIDD	268	500
FHGDD	1316	684
Their fusion	1584	1184

authors proposed an LSTM model to solve the gradient problem of the neural networks models.

The summary of the advantages and disadvantages of these deep learning algorithms is illustrated in Table 3. Thus, our main objective in this study was to merge these deep learning models by utilizing a weighted voting classifier, to take advantage of their complementary and to improve the accuracy of diabetes mellitus prediction. Their combination has been tested to forecast the COVID-19 outbreak [41] and news classification on Social Media in [42] where experimental findings show a significant increase in performance when using ensemble learning.

In the proposed EL model, the Multilayer Perceptron (MLP) architecture comprises three hidden layers with varying neuron counts (8, 32, 20, and 1), facilitating enhanced diagnostic accuracy despite the computationally intensive processing demands. For the Convolutional Neural Network (CNN), the input encompasses eight features with a binary target output. Leveraging a one-dimensional CNN algorithm, the input features undergo batch normalization (BN) coupled with Rectified Linear Unit (ReLU) activation to highlight essential features. Subsequently, a max-pooling layer with a pool size of 2 is applied, followed by a Flatten layer for vector transformation, elucidated mathematically in Equation (1).

Table 3. Summary of the advantages and disadvantages of the used DL algorithms.

Model	Advantages	Disadvantages
MLP	1) Good at mapping different features into independent spaces [43]. 2) Suitable for classification prediction problems [23]). 3) Can be applied to different types of data such as Images and text [18]. 4) Works well with large input data [23].	1) Over fitting will occur when there are too many hidden layers [27]. 2) The total number of parameters can be very large because it is fully connected [23, 44, 27]. 3) Vanishing gradient problem [36]
CNN	1) Appropriate for extracting features (feature learning) [21] 2) Weight sharing [24]. 3) Minimize learning parameters and complexity [25]. 4) Classification takes requires less time and produces higher accuracy, especially in image processing [24].	1) The computational cost is high [25] 2) Can't work well without a huge amount of data [38]. 3) Vanishing problem [31].
LSTM	1) Has several memory blocks that can overcome the vanishing gradient problem [32]. 2) Efficient at modeling complex sequential data [45].	1) computational time is high [45]). 2) LSTM is not suitable for extremely nonlinear or noisy data [42].

$$y(K) = f(\sum_1^N Xj * Wp(K) + Bp(K)) \quad (1)$$

Where: Wp: specified weight of p node, Xj: Batch Normalization of input feature, $f(.)$: the activation function ReLU, p: number of filers, and *Bp*: bias of p node.

Additionally, Long Short-Term Memory (LSTM) networks mitigate issues of vanishing and exploding gradients, with gates represented by Equations (2) to (6) [49].

$$i_t = \sigma_g(w_i x_t + u_i h_{t-1} + b_i) \quad (2)$$

$$f_t = \sigma_g\left(w_f x_t + u_f h_{t-1} + b_f\right) \quad (3)$$

3.5. Weighted voting classifier

Weighted voting or soft voting is one of the simplest and widely employed combiners [46]; [47]; [11]. It is a meta-classifier, that merges conceptually similar or dissimilar machine learning algorithms for classification and prediction via a majority vote [48]. A voting algorithm utilizes both hard and soft voting methods. The final prediction, in hard voting, is determined by majority voting, where the aggregator chooses the outcome that consistently appears among the basic models. In soft voting, the class label is predicated using the probability P given to the classes by the classifier. Because it integrates the results of many models, the voting classifier offers overall performance than other conventional base models [47].

4. Proposed approach

$$c_t = f_t * c_{t-1} + i_t * \sigma_h(w_c x_t + u_c h_{t-1} + b_c) \quad (4)$$

$$o_t = \sigma_g(w_o x_t + u_o h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t \quad (6)$$

Where: i_t, f_t, and c_t represent the input gate, forget gate, and output gate, respectively.

σ_g: specified the sigmoid function.

σ_h : specified hyperbolic tangent function.

h_{t-1}: Specified output of previous LSTM block at the time t-1.

x_t : Specified input at current time.

b_x: Specified biases for respective gate.

The weights are denoted by The symbols w and u. These weights generally keep the gradient problem from vanishing.

Employing a soft voting classifier, our model harmonizes predictions from MLP, CNN, and LSTM. Each model independently generates predictions, which are merged using a weighted voting approach. Equation (7) [50] defines the weighting equation.

$$\hat{y} = argmax_i \sum_{j=1}^{m} w_j * p_{ij} \quad (7)$$

Where p represents the predicted probability for each classifier, and w_j denotes the weight attributed to each classifier. Optimal weight determination for deep learning algorithms in the soft voting ensemble employs the grid search method.

The voting classifier capitalizes on the unique strengths of these foundational deep learning models to elevate the performance of diabetes prediction, as depicted in Fig.1. The advantages of LSTM, MLP, and CNN are mutually reinforcing. Previous research by [24], [25], [17], [30], [31], [32], and [23] has demonstrated their individual efficacy in achieving high accuracies. Specifically, MLP adeptly maps diverse features into independent spaces through multiple hidden layers, while LSTM effectively mitigates vanishing gradient issues common in traditional neural networks. Meanwhile, CNN excels at extracting

features and reducing complexity, thus combating overfitting.

Given the resource-intensive nature of training deep learning models and the necessity for extensive datasets, we employ a data

fusion technique to amalgamate two existing datasets into a larger, more comprehensive one. Our methodology comprises three main steps: firstly, preprocessing the data; secondly, individually training and testing the three neural networks; and finally, employing the soft voting classifier to integrate their predictions. This approach not only enhances accuracy but also yields superior results compared to employing individual classifiers.

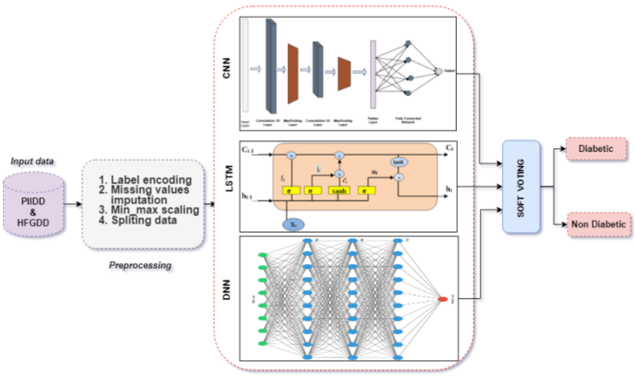

Fig.1. The flowchart diagram of the proposed ensemble deep learning approach.

4.1. Data pre-processing

Data pre-processing stands as a foundational step essential

for refining machine learning model performance by ensuring that input data is meticulously formatted and free from anomalies. This multifaceted process entails various critical tasks, including label encoding, outlier removal, missing value handling, and data standardization.

Label encoding is a pivotal technique utilized to transform categorical or textual value labels into a numerical format intelligible to deep learning algorithms. By assigning a unique numerical value to each category, label encoding enables algorithms to process and analyze categorical data effectively.

Addressing missing values is another integral aspect of data pre-processing. In this step, missing values within the dataset are meticulously addressed to maintain data integrity. One common approach involves calculating the mean value of each attribute and replacing missing values with this calculated mean.

By ensuring that the dataset remains complete, this method facilitates accurate analysis and modeling.

Data standardization techniques, such as min-max normalization, are employed to scale numerical attribute values to a specific range, typically [0, 1]. This normalization fosters uniformity across attributes, preventing any single attribute from disproportionately influencing model training. Mathematically, min-max normalization is expressed using equation (8).

$$x_{norm} = \frac{x - min(X)}{max(X) - min(X)} \quad (8)$$

Where x : The original value.

x_{norm} : Represents the normalized value of x.

$\min(X)$: Represents the minimum value in the dataset.

$\max(X)$: Represents the maximum value in the dataset.

The Min-max algorithm rescales a variable within the training

sample range [$\min(X)$, $\max(X)$] to fit into the interval [0, 1]

through a linear mapping.

Upon completing pre-processing, the dataset is often divided into training and testing sets to facilitate model evaluation and performance assessment. This division allows for rigorous evaluation of model efficacy and generalization capabilities, ensuring robustness in real-world applications.

4.2. Training parameters

For all experiments some parameters were adapted: the training set is taken as 80%, and the testing set is taken as 20% batch size of 64, epoch = 300, loss function= binary cross-entropy, learning rate = 0.00004, and the optimizer used is Adam. These are chosen experimentally by trial and reducing error. The layered structure of baseline deep learning classifiers is shown in the Table 4.

Table 4. Training parameters

Models	Layers (Type)	Parameters
CNN	Sequential model with 1) Convolutionld (Conv1D) 2) Max pooling1d 3) conv1D_1 4) Max pooling1d1 5) Flatten 6) Dense 7) Dense Classification	1) F =64,Kernel =3,activation function =ReLU , padding =same, input shape = (8, 1). 2) Pool size = (2). 3) F =32, Kernel =3, activation function =ReLU , padding =same. 4) Pool size (2). 5) - 6) 16 neurons, activation function =ReLU. 7) 1 neuron, activation function =sigmoid.
LSTM	Sequential model with 1) LSTM 2) LSTM 3) Dense 4) Dense Classification	1) 32 neurons, input shape = (None, 1,8), dropout =0.1, recurrent dropout =0.1, return sequences =True. 2) 25 neurons, dropout =0.1, input shape =(1,8), recurrent dropout =0.1 3) 20 neurons, activation function =ReLU. 4) 1 neuron, activation function =sigmoid.
MLP	Sequential model with 1) Dense 2) Dense 3) Dense 4) Dense Classification	1) 32 neurons, activation =ReLU, input shape =(8,1). 2) .32 neurons, activation. 3) 20 neurons, activation =ReLU. 4) 1 neuron, activation =sigmoid.

5. Results

The performance parameters for the proposed model are taken as accuracy, precision, recall, F1-score, accuracy on the training set, and testing set. Accuracy is the rate of correct classification on the training set and testing set, and precision is the ratio of correct positive predictions to the number of predicted positive outcomes. At the same time, recall is simply the ratio of the

number of accurate positive predictions to the total actual positive

instances, and F1-score is the harmonic mean between precision and sensitivity. These metrics can be calculated by the following formulas:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{9}$$

$$Recall = \frac{TP}{TP+FN} \tag{10}$$

$$F1-score = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \tag{11}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{12}$$

Where $TP, TN,\ FP,\ and\ FN$ denote true positives, true

negatives, false positives, and false negatives, respectively. Fig.2. delineates the prediction accuracy of the suggested weighted voting classifier vis-à-vis other baseline deep learning models.

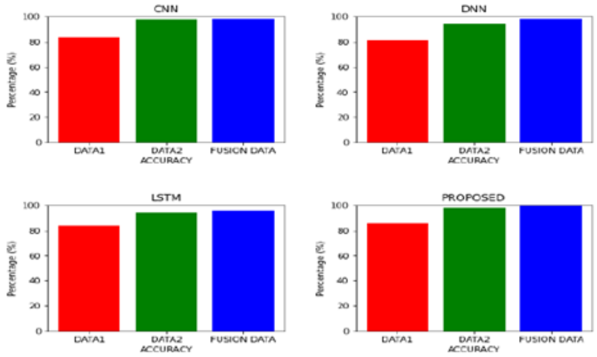

Fig.2. Accuracy of the individual and EL models.

The CNN model exhibited a notable increase in accuracy when applied to the combined PIDD and FHGDD datasets, rising from 83.11% on PIDD to 98.27%. Similarly, MLP saw significant improvement post-fusion, with accuracy rising from 81.16% on PIDD to 98.13%. LSTM, too, demonstrated enhanced accuracy after fusion, increasing from 83.76% on PIDD to 95.84%. Our proposed model, integrating multiple deep learning classifiers, showed superior performance, achieving 85.71% accuracy on PIDD and 98% on FHGDD with the soft voting classifier. Leveraging the fused dataset, our model excelled with an impressive accuracy of 99.81%. These results underscore the effectiveness of deep learning techniques and the weighted voting classifier in achieving heightened accuracy across diverse datasets, surpassing baseline classifiers with minimal error rates both pre- and post-fusion.The summary of performance metrics, including accuracy, precision, recall, and F1-score for the selected datasets, is presented in Tables 5 to 7.

Table 5. Comparison results of the proposed model with baseline classifiers on PIDD dataset

Model	Accuracy	Precision	Recall	F1_scors
CNN	83.11	82.97	68.42	75.00
MLP	81.16	76.92	70.17	73.39
LSTM	83.76	84.78	68.42	75.72
Proposed EL model	**86**	**84.91**	**75.43**	**79.62**

Table 6. Comparison results of the proposed model with baseline classifiers on FHGDD dataset

Model	Accuracy	Precision	Recall	F1_scors
CNN	97.5	97.08	95.68	96.37
MLP	94.25	90.27	93.52	91.87
LSTM	94.25	93.28	89.92	91.57
Proposed EL model	**98.0**	**97.81**	**96.40**	**97.10**

Table 7. Comparison results of the proposed model with baseline classifiers on the merged dataset.

Model	Accuracy	precision	recall	F1_scors
CNN	98.27	99.45	98.36	98.90
MLP	98.13	96.33	99	98.13
LSTM	95.84	94.47	92.93	93.69
Proposed EL model	**99.81**	**99.45**	**99.8**	**99.72**

6. Discussion

As depicted in Table 5. In this experiment, we focus solely on PIDD dataset. Our findings reveal that despite the dataset's limited size and the absence of feature selection techniques, the ensemble voting method we propose achieves significantly improved rates of accuracy, precision, recall, and F1-scores compared to single classifiers.

Similarly, in Table 6, we exclusively analyze the FHGDD dataset. Various evaluation metrics are utilized to compare CNN, MLP, and LSTM with our proposed model. Due to the dataset's extensive size, our model demonstrates superior overall performance over other deep learning models, as depicted in Fig.2. Therefore, our soft voting classifier yields more accurate results when provided with a larger dataset.

Our experiment results indicate that our proposed model achieves an outstanding accuracy of 98%, surpassing other DL classifiers by a significant margin. Furthermore, precision, recall, and F1-scores also exhibit remarkable performance compared to other DL classifiers. Table 6 displays the results after fusing both diabetes datasets. Our proposed weighted voting classifier is compared with baseline deep learning models in terms of accuracy, precision, recall, and F1-scores. As a result of data fusion, our model's performance exceeds that of other deep learning models by a considerable margin.

Ultimately, Table 8 presents a comparison of our proposed approach with state-of-the-art methods. The table presents a comparison of various research works in the field of diabetes prediction using different classifiers and ensemble techniques. The proposed ensemble deep learning model stands out with its utilization of CNN, MLP, and LSTM classifiers combined with a soft voting ensemble technique, resulting in impressive accuracies of 86.71% for

PIDD, 98.0% for FHGDD, and a remarkable 99.81% after data fusion. This model surpasses other works in terms of accuracy and the ensemble technique employed. Among the other studies listed, different classifiers and ensemble techniques were utilized, such as Random Forest (RF), Logistic Regression (LR), Naive Bayes (NB), and Support Vector Machine (SVM). Some studies incorporated stacking, boosting, or hybrid models to enhance predictive performance. Notably, [51]achieved high accuracies of 85.9% for PIDD and 99.5% for FHGDD using Random Forest (RF) and 99.8% after

data fusion. [43] Also achieved remarkable accuracy of 97.0% for FHGDD using various architectures of Deep Neural Network (DNN) techniques. However, the proposed ensemble deep learning model surpasses these previous works in terms of accuracy across both individual datasets and after data fusion, showcasing the advancements in predictive modeling techniques for diabetes detection.

7. Conclusion

Detecting early signs of diabetes, whether positive or negative, presents a considerable challenge. In recent years, numerous researchers have sought solutions by assessing the efficacy of deep learning models. Hence, this research introduces a weighted voting classifier that combines three deep learning algorithms—MLP, CNN, and LSTM—to create a more precise and resilient predictive model for diabetes. Additionally, we amalgamated datasets to form a larger, more comprehensive dataset. Results indicate that our proposed model outperforms other deep learning models, exhibiting superior accuracy of 99.81%, precision of

Table 8. comparison of our proposed approach with state-of-the-art works

Research Work	Year	Baseline classifiers	Ensemble Technique	PI-DD	FHG-DD	Data Fusion	Accu-racy
Proposed Ensemble Deep Learning Model	**2024**	**CNN, MLP, and LSTM**	**Soft voting**	+	+	+	**86.71 % 98.0 % 99.81 %**
[11]	2021	RF. LR. and NB	soft voting	+			79.08 %
[12]	2019	Naive Bayes, K-NN. SVM, RF. ANN. LR. GBoost. AdaBoost	Ensemble Voting of the three best classifiers	+			86%
[52]	2021	LR. LDA, KNN. SVM, DT.and SVM.	Super learner	+			86%
[19]	2022	LR. SVM Ns. GB, NB. and RF	Ensemble Voting of three best classifier	+			81.7 %
[43]	2021	Different architectures of DNN technique	-		+		97.0 %
[27]	2023	Birch-ANN	Hybrid model		+		97.25 %
[53]	2022	KNN, NB. LDA, and DT	Stacking technique		+		97.35 %
[51]	2022	RF, DT	RF	+	+	+	85.9 % 99.5 % 99.8 %
[26]	2021	LR, NB, DT, KNN, RF, and SVM.	boosting technique	+	+	+	72.7 % 91% 99.6 %

99.45 %, recall of 99.8%, and F1 scores of 99.72%. This model stands to aid physicians in diagnosing patients with greater certainty, thereby reducing the risk of human error. Looking ahead, our future endeavors involve exploring alternative deep learning techniques for similar objectives and testing the proposed model on diverse disease classification datasets. Additionally, our research endeavors to address additional challenges in healthcare-based machine learning models, including handling missing data, mitigating biases within datasets, and tackling class imbalances.

Author contributions

Selma. Aouamria., Djalila. Boughareb., Mohamed. Nemissi., Zineeldine. Kouhla., and Hamid.Seridi. contributed to the design and implementation of the research, to the analysis of the results, and to the writing of the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] J. Sreedharan, J. Muttappallymyalil, S. Al Sharbatti, S. Hassoun, R. Safadi, I. Abderahman, W. A. Hameed, A. M. Ibrahim, M. T. Takana and A. M. Fouda, "Incidence of type 2 diabetes mellitus among Emirati residents in Ajman, United Arab Emirates," Korean Journal of Family Medicine, vol. 36, p. 253, 2015.

[2] J. F. Ndisang, A. Vannacci and S. Rastogi, Insulin resistance, type 1 and type 2 diabetes, and related complications, vol. 2017, Hindawi, 2017.

[3] T. Sharma and M. Shah, "A comprehensive review of machine learning techniques on diabetes detection," Visual Computing for Industry, Biomedicine, and Art, vol. 4, p. 1–16, 2021.

[4] B. F. Wee, S. Sivakumar, K. H. Lim, W. K. Wong and F. H. Juwono, "Diabetes detection based on machine learning and deep learning approaches," Multimedia Tools and Applications, vol. 83, no. 8, pp. 24153-24185, 2024.

[5] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald and E. Muharemagic, "Deep learning applications and challenges in big data analytics," Journal of big data, vol. 2, p. 1–21, 2015.

[6] T. Zhu, K. Li, P. Herrero and P. Georgiou, "Deep learning for diabetes: a systematic review," IEEE Journal of Biomedical and Health Informatics, vol. 25, p. 2744–2757, 2020.

[7] M. W. Nadeem, H. G. Goh, V. Ponnusamy, I. Andonovic, M. A. Khan and M. Hussain, "A fusion-based machine learning approach for the prediction of the onset of diabetes," in Healthcare, 2021.

[8] Sathurthi, S., "An analysis of parallel ensemble diabetes decision support system based on voting classifier for classification problem," Electronic Government, an International Journal, vol. 16, p. 25–38, 2020.

[9] Doğru, S. Buyrukoğlu and M. Arı, "A hybrid super ensemble learning model for the early-stage prediction of diabetes risk," Medical & Biological Engineering & Computing, p. 1–13, 2023.

[10] L. Ismail, H. Materwala, M. Tayefi, P. Ngo and A. P. Karduck, "Type 2 diabetes with artificial intelligence machine learning: methods and evaluation," Archives of Computational Methods in Engineering, p. 1–21, 2021.

[11] S. Kumari, D. Kumar and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier," International Journal of Cognitive Computing in Engineering, vol. 2, p. 40–46, 2021.

[12] Mahabub, "A robust voting approach for diabetes prediction using traditional machine learning techniques," SN Applied Sciences, vol. 1, p. 1667, 2019.

[13] S. Bashir, U. Qamar, F. H. Khan and M. Y. Javed, "An efficient rule-based classification of Diabetes using ID3, C4. 5, & CART ensembles," in 2014 12th International Conference on Frontiers of Information Technology, 2014.

[14] 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, p. e0179805, 2017.

[15] H. B. Kibria, M. Nahiduzzaman, M. O. F. Goni, M. Ahsan and J. Haider, "An ensemble approach for the prediction of diabetes mellitus using a soft voting classifier with an explainable AI," Sensors, vol. 22, p. 7268, 2022.

[16] M. Alehegn, R. Joshi and M. Alehegn, "Analysis and prediction of diabetes diseases using machine learning algorithm: Ensemble approach," International Research Journal of Engineering and Technology, vol. 4, p. 426–436, 2017.

[17] H. Naz and S. Ahuja, "Deep learning approach for diabetes prediction using PIMA Indian dataset," Journal of Diabetes & Metabolic Disorders, vol. 19, p. 391–403, 2020.

[18] E. P. Prakash, S. Karthik, M. V. Kamal, B. Reddy S, M. A. Mukunthan, K. Sahile and others, "Implementation of Artificial Neural Network to Predict Diabetes with High-Quality Health System," Computational Intelligence and Neuroscience, vol. 2022, 2022.

[19] Z. Mushtaq, M. F. Ramzan, S. Ali, S. Baseer, A. Samad and M. Husnain, "Voting classification-based diabetes mellitus prediction using hypertuned

machine-learning techniques," Mobile Information Systems, vol. 2022, p. 1–16, 2022.

[20] S. Ramesh, H. Balaji, N. C. S. N. Iyengar and R. D. Caytiles, "Optimal predictive analytics of pima diabetics using deep learning," International Journal of Database Theory and Application, vol. 10, p. 47–62, 2017.

[21] Yahyaoui, A. Jamil, J. Rasheed and M. Yesiltepe, "A decision support system for diabetes prediction using machine learning and deep learning techniques," in 2019 1st International informatics and software engineering conference (UBMYK), 2019.

[22] K. Kannadasan, D. R. Edla and V. Kuppili, "Type 2 diabetes data classification using stacked autoencoders in deep neural networks," Clinical Epidemiology and Global Health, vol. 7, p. 530–535, 2019.

[23] S. K. Kalagotla, S. V. Gangashetty and K. Giridhar, "A novel stacking technique for prediction of diabetes," Computers in Biology and Medicine, vol. 135, p. 104554, 2021.

[24] M. T. García-Ordás, C. Benavides, J. A. Benítez-Andrades, H. Alaiz-Moretón and I. García-Rodríguez, "Diabetes detection using deep learning techniques with oversampling and feature augmentation," Computer Methods and Programs in Biomedicine, vol. 202, p. 105968, 2021.

[25] G. Swapna, R. Vinayakumar and K. P. Soman, "Diabetes detection using deep learning algorithms," ICT express, vol. 4, p. 243–246, 2018.

[26] B. Ihnaini, M. A. Khan, T. A. Khan, S. Abbas, M. S. Daoud, M. Ahmad and M. A. Khan, "A smart healthcare recommendation system for multidisciplinary diabetes patients with data fusion based on deep ensemble learning," Computational Intelligence and Neuroscience, vol. 2021, 2021.

[27] Y. Singh and M. Tiwari, "A Hybrid Approach for Prediction of Type 2 Diabetes Using Birch Clustering and Artificial Neural Network," Intelligent Systems and Smart Infrastructure: Proceedings of ICISSI 2022, pp. 421-433, 2023.

[28] S. Ramesh, R. D. Caytiles and N. C. S. Iyengar, "A deep learning approach to identify diabetes," Advanced Science and Technology Letters, vol. 145, p. 44–49, 2017.

[29] H. Qi, X. Song, S. Liu, Y. Zhang and K. K. L. Wong, "KFPredict: An ensemble learning prediction framework for diabetes based on fusion of key features.," Computer methods and programs in biomedicine, vol. 231, p. 107378, April 2023.

[30] R. Kamalraj, S. Neelakandan, M. R. Kumar, V. C. S. Rao, R. Anand and H. Singh, "Interpretable filter based convolutional neural network (IF-CNN) for glucose prediction and classification using PD-SS algorithm," Measurement, vol. 183, p. 109804, 2021.

[31] M. Rahman, D. Islam, R. J. Mukti and I. Saha, "A deep learning approach based on convolutional LSTM for detecting diabetes," Computational biology and chemistry, vol. 88, p. 107329, 2020.

[32] A. P. Ratna, P. D. Purnamasari, N. K. Anandra and D. L. Luhurkinanti, "Hybrid Deep Learning CNN-Bidirectional LSTM and Manhattan Distance for Japanese Automated Short Answer Grading: Use case in Japanese Language Studies," in 2022 the 8th International Conference on Communication and Information Processing, 2022.

[33] U. M. Learning, "Pima Indians Diabetes Database," October 2016. [Online]. Available: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database. [Accessed 17 4 2023].

[34] J. Dasilva, "Diabetes Dataset," 2018. [Online]. Available: https://www.kaggle.com/datasets/johndasilva/diabetes. [Accessed 17 4 2023].

[35] S. Pal, N. Mishra, M. Bhushan, P. S. Kholiya, M. Rana and A. Negi, "Deep learning techniques for prediction and diagnosis of diabetes mellitus," in 2022 International Mobile and Embedded Technology Conference (MECON), 2022.

[36] R. Lohiya and A. Thakkar, "Intrusion detection using deep neural network with antirectifier layer," in Applied Soft Computing and Communication Networks: Proceedings of ACN 2020, 2021.

[37] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai and others, "Recent advances in convolutional neural networks," Pattern recognition, vol. 77, p. 354–377, 2018.

[38] M. Jogin, M. S. Madhulika, G. D. Divya, R. K. Meghana, S. Apoorva and others, "Feature extraction using convolution neural networks (CNN) and deep learning," in 2018 3rd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT), 2018.

[39] R. Pascanu, C. Gulcehre, K. Cho and Y. Bengio, "How to construct deep recurrent neural networks," arXiv preprint arXiv:1312.6026, 2013.

[40] M. F. Aslan and K. Sabanci, "A Novel Proposal for Deep Learning-Based Diabetes Prediction: Converting Clinical Data to Image Data," Diagnostics, vol. 13, p. 796, 2023.

[41] N. B. Yahia, M. D. Kandara and N. B. B. Saoud, "Deep ensemble learning method to forecast COVID-19 outbreak," ResearchSquare: Durham, NC, USA, 2020.

[42] Rahman, "A Deep Ensemble Model for News Classification on Social Media," in 2021 International Conference on Cyber Warfare and Security (ICCWS), 2021.

[43] T. Beghriche, M. Djerioui, Y. Brik, B. Attallah and S. B. Belhaouari, "An efficient prediction system for diabetes disease based on deep neural network," Complexity, vol. 2021, p. 1–14, 2021.

[44] Singh, A. Dhillon, N. Kumar, M. S. Hossain, G. Muhammad and M. Kumar, "eDiaPredict: an ensemble-based framework for diabetes prediction," ACM Transactions on Multimidia Computing Communications and Applications, vol. 17, p. 1–26, 2021.

[45] S. Larabi-Marie-Sainte, L. Aburahmah, R. Almohaini and T. Saba, "Current techniques for diabetes prediction: review and case study," Applied Sciences, vol. 9, p. 4604, 2019.

[46] S. R. Sannasi Chakravarthy and H. Rajaguru, "Ensemble-Based Weighted Voting Approach for the Early Diagnosis of Diabetes Mellitus," in Sustainable Communication Networks and Application: Proceedings of ICSCN 2021, Springer, 2022, p. 451–460.

[47] J. Cao, S. Kwong, R. Wang, X. Li, K. Li and X. Kong, "Class-specific soft voting based multiple extreme learning machines ensemble," Neurocomputing, vol. 149, p. 275–284, 2015.

[48] H. T. X. Doan and G. M. Foody, "Increasing soft classification accuracy through the use of an ensemble of classifiers," International Journal of Remote Sensing, vol. 28, p. 4609–4623, 2007.

[49] P. B. K. Chowdary and R. U. Kumar, "An Effective Approach for Detecting Diabetes using Deep Learning Techniques based on Convolutional LSTM Networks," International Journal of Advanced Computer Science and Applications, vol. 12, 2021.

[50] N. Rai, N. Kaushik, D. Kumar, C. Raj and A. Ali, "Mortality prediction of COVID-19 patients using soft voting classifier," International Journal of Cognitive Computing in Engineering, vol. 3, p. 172–179, 2022.

[51] Azbeg, "Diabetes emergency cases identification based on a statistical predictive model," Journal of Big Data, vol. 9, p. 1–25, 2022.

[52] R. Saxena and others, "Role of K-nearest neighbour in detection of Diabetes Mellitus," Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, p. 373–376, 2021.

[53] M. Ali, M. N. Haider, S. A. Lashari, W. Sharif, A. Khan and D. A. Ramli, "Stacking Classifier with Random Forest functioning as a Meta Classifier for Diabetes Diseases Classification," Procedia Computer Science, vol. 207, p. 3459–3468, 2022.