IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 8 March 2024, accepted 23 May 2024, date of publication 1 August 2024, date of current version 12 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3436641

RESEARCH ARTICLE

An Innovative Ensemble Deep Learning Clinical Decision Support System for Diabetes Prediction

MANA SALEH AL RESHAN[1,2], (Member, IEEE), SAMINA AMIN[3], MUHAMMAD ALI ZEB[3], ADEL SULAIMAN[2,4], HANI ALSHAHRANI[2,4], (Senior Member, IEEE), ASADULLAH SHAIKH[1,2], (Senior Member, IEEE), AND MOHAMED A. ELMAGZOUB[2,5]

[1]Department of Information Systems, College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia
[2]Emerging Technologies Research Laboratory (ETRL), College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia
[3]Institute of Computing, Kohat University of Science and Technology, Kohat 26000, Pakistan
[4]Department of Computer Science, College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia
[5]Department of Network and Communication Engineering, College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia

Corresponding author: Adel Sulaiman (aaalsulaiman@nu.edu.sa)

ABSTRACT Diabetes is a significant global health concern, with an increasing number of diabetic people at risk. It is considered a chronic disease and leads to a significant number of fatalities annually. Early prediction of diabetes is essential for preventing its progression and reducing the risk of severe complications such as kidney and heart diseases. This study proposes an innovative Ensemble Deep Learning (EDL) clinical decision support system for diabetes prediction with high accuracy. The proposed EDL model uses Deep Learning (DL) architectures such as Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN), integrated with an ensemble learning-based stacking model. The EDL is implemented based on a stack ensemble model that applies meta-level models, including stack-ANN, stack-CNN, and stack-LSTM, to improve the prediction of diabetes. Three diabetes datasets, such as I. Pima Indian Diabetes Dataset (PIMA-IDD-I), II. Diabetes Dataset Frankfurt Hospital Germany (DDFH-G), and III. Iraqi Diabetes Patient Dataset (IDPD-I) are used to train the novel EDL models. The Extra Tree Classifier (ETC) approach is used to extract the relevant features from the data. The performance of the proposed EDL models is evaluated based on major evaluation metrics such as accuracy, precision, sensitivity, specificity, F-score, Matthews Correlation Coefficient (MCC), and ROC/AUC. Among the proposed EDL models, the stack-ANN achieved robust performance using DDFH-G, PIMA-IDD-I, and IDPD-I datasets with accuracy scores of 99.51%, 98.81%, and 98.45%, respectively. The overall results demonstrate that the proposed EDL models outperform previous studies in predicting diabetes.

INDEX TERMS Artificial neural networks, convolutional neural networks, diabetes mellitus, deep learning, ensemble learning, long short-term memory.

I. INTRODUCTION

Diabetes is a significant global health concern, with an increasing number of diabetic patients at risk. It leads to a substantial loss of lives [1]. Diabetes, also known as Diabetes Mellitus (DM), is a metabolic disorder characterized by prolonged elevated blood glucose levels due to the inability of the body to consume it effectively. Severe complications associated with DM include diabetic ketoacidosis, chronic renal failure, nonketotic hyperosmolar coma, foot ulcers, retinal damage, cardiovascular disease, stroke, and chronic renal failure. There are three primary types of DM, including Type 1 Diabetes (T1D), Type 2 Diabetes (T2D), and Gestational diabetes [1], [2].

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

T1D occurs due to insufficient insulin production, primarily affecting people under the age of 30. Excessive thirst, frequent urine, and high blood sugar levels are typical symptoms. T1D patients typically require medication for their management. T2D, on the other hand, is a prevalent type in which the body struggles to produce or use insulin effectively. It mainly distresses middle-aged and older adults and is often associated with factors like lifestyle choices, lack of physical activity, dietary habits, obesity, smoking, high cholesterol (hyperlipidemia), and hypertension (hyperglycemia). Additionally, gestational diabetes may be diagnosed during pregnancy [1], [2]. There is a serious public health risk associated with this disorder, requiring continued efforts in treatment, and prevention [3]. Figure 1 shows the basic types of diabetes.

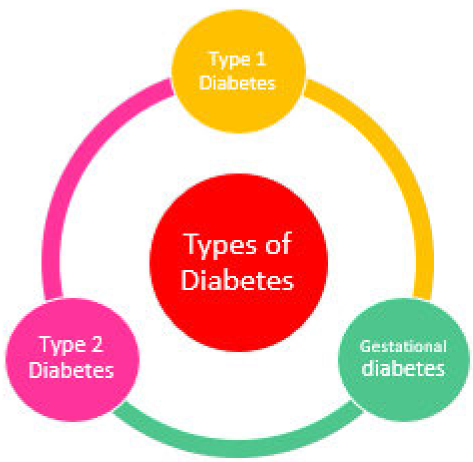


FIGURE 1. Types of diabetes.

According to the World Health Organization (WHO), over 420 million people worldwide suffer from DM, and over 650 million adults are classified as obese, with obesity rates having tripled since 1975. The chronic disease known as DM has become more common over time. The global rise in the number of diabetic patients is substantial, posing a significant health challenge globally. Table 1 lists the highest DM rates by the top 15 countries [4]. In the early stage of DM disease, many diabetic patients often underestimate the seriousness of their health situation [5], [6]. The development of techniques for early detection and prediction of DM is necessary, as delayed diagnosis leads to several health problems and results in a high annual mortality rate. Patients with DM are becoming more common, and this is considered a severe health concern [3], [7].Predicting DM in people of all ages is urgently necessary. Therefore, making appropriate lifestyle changes on time can help prevent the progression of DM and its related health concerns [8].

Currently, the scientific community has directed its attention toward the early and accurate prediction of DM through the utilization of strong computational approaches. Artificial Intelligence (AI) and soft computing methods play a vital role in transforming human concepts into practical applications. These systems find applications in medical diagnosis and various areas related to human health. Machine Learning (ML) and AI enable the early detection and prediction of DM through automated methods, offering advantages over manual diagnosis [9], [10]. Currently, there is a substantial body of literature on the automatic detection methods and self-management of DM using ML and AI techniques [11], [12].

TABLE 1. Highest diabetes rates by the top 15 countries [4].

S.NO#	Country	Diabetes rate
1.	Pakistan	30.8%
2.	French Polynesia	25.2%
3.	Kuwait	24.9%
4.	Nauru	23.42%
5.	New Caledonia	23.4%
6.	North Marina Islan	23.4%
7.	Marshall Islands	23.0%
8.	Mauritius	22.6%
9.	Kiribati	22.1%
10.	Egypt	20.9%
11.	American Samoa	20.3%
12.	Tuvalu	20.3%
13.	Solomon Islands	19.8%
14.	Qatar	19.5%
15.	Guam	19.1%

Conventional ML techniques have yielded promising results for the prediction and classification of DM into DM-positive and DM-negative cases. However, the accuracy of DM diagnosis can be further enhanced through the application of an EDL framework. To the best of our knowledge, none of the existing studies have used EDL techniques for this purpose. In this research, an innovative DM prediction framework is proposed that uses DL algorithms as base-level models, such as ANN, LSTM, and CNN. Then, EDL is implemented based on stack ensemble models that apply meta-level models, including stack-ANN, stack-CNN, and stack-LSTM to improve the prediction of DM. This combination of various neural network architectures allows the model to capture different patterns and features within the data. The EDL combines the predictions of each base-level model by improving the model's generalization to unseen data and decreasing overfitting.

The primary contributions of this study are summarized as follows:

1. To propose a clinical decision support system for diabetes prediction based on stacking ensemble model and DL approaches.
2. To combine different DL approaches, including ANN, LSTM, and CNN, as well as EDL models including stack-ANN, stack-LSTM, and stack-CNN, to improve feature extraction and prediction accuracy.

3. To train the EDL model using three diabetes datasets, including a small dataset (PIMA-IDD-I) with 768 instances, a large dataset (DDFH-G) with 2000 instances, and a multi-class (IDPD-I) dataset with 1000 instances.
4. To provide healthcare professionals with a tool for early predictions and personalized patient care in diabetes management with high accuracy.
5. To assess the efficiency of the proposed EDL, a comparative analysis has been conducted against the current state-of-the-art techniques. This evaluation ensures that the proposed EDL model outperforms existing methods and demonstrates its superiority in predicting DM.

The rest of the paper is arranged into six sections as follows: Section II reviews and highlights the importance of the related work, while the research gap is highlighted in Section III. Section IV illustrates the proposed EDL framework in detail. Section V presents the results of the proposed EDL models and Section VI discusses the results in comparison with the state-of-the-art techniques while acknowledging the limitations and providing future research developments. Finally, Section VII concludes the main findings of the proposed study.

II. RELATED WORK

Several research studies have been carried out in DM detection using ML and DL approaches concerning the importance of detecting diabetes [1], [6], [7]. For example, Du et al., [13] proposed an explainable ML-based Gestational DM prediction system. Ebrahim and Derbew [14] proposed an ML-based model for T2D detection and classification. Sarraju et al., [15] proposed BERT for DM detection using unstructured electronic health record data. Thotad et al., [16] developed Gaussian Naive Bayes (NB), linear discriminant analysis, Support Vector Classifier (SVM), Logistic Regression (LR), K-Nearest Neighbor (KNN), Decision Trees (DT), Extreme Gradient Boosting (XGB), and Random Forest (RF) for DM detection using health survey and Indian demographic data. Olickal et al., [17] studied comprehensive diabetes care in primary care setting. The study was carried out among PwDs aged $\geq$18 years at a tertiary care hospital in India. Aizu'bi et al., [18] proposed a DM prediction model based on SVM, KNN, DT, RF, adaptive boosting, LR, and DL. Pan et al., [19] developed XGB, RF with recursive feature elimination, and ANN for detecting diabetic retinopathy in the Chinese population. Another relevant research study has been conducted by Uddin et al., [20], who proposed a questionnaire-based survey addressing common risk factors of DM using SVM, DT, LR, gradient boost, XGB, and RF. Zhao et al., [21] proposed a soft voting technique that combines light gradient boosting XGB, and RF to detect T2D. To conduct this research, they used 9822 screening samples with 82 relevant features.

Bernardini et al., [22] developed an SVM approach for predicting T2D using electronic health records. Similarly, Eleftheriades et al., [23] proposed a regression tree and LR-based prediction model for the need for insulin therapy in females with gestational DM. They examined samples from 775 women who had been given an IADPSG-based diagnosis of gestational DM. Prasad et al., [3] proposed a KNN-based model for DM predcting using PIMA-IDD-I. Pyne et al., [24] proposed an ANN model for DM prediction, obtaining an accuracy of 80.79%. Chang et al., [25] developed RF, NB, and J48 DT using PIMA-IDD-I for predicting and classifying DM. Gupta et al., [26] implemented Adaptive Boosting, RF, XGB and ANN using a majority voting for DM prediction. Gupta et al., [27] designed KNN, DT, RF, and SVM models. Moreover, Alhalaseh et al., [28] proposed LR, NB, and RF techniques with a soft ensemble model using PIMA-IDD-I and IDPD-I datasets. Their proposed model achieved accuracy of 97% and 81% using IDPD-I and PIMA-IDD-I, IDPD-I D, respectively. Nasser and Dawood [29] developed a multi-layer perceptron using IDPD-I for predicting DM. Nuankaew [30] developed KNN, SVM, RF, and DL for T2D prediction. They trained the models across PIMA-IDD-I and IDPD-I datasets. Sivaranjani et al., [31] proposed RF and SVM using PIMA-IDD-I dataset, yielding accuracy scores of 83% and 81.4% respectively. Similarly, for DM prediction, in [32], [33], and [34] the RF model was proposed and yielded accuracy scores of 85.6%, 82%, and 82.26%, respectively.

In recent years, DL approaches have been used in DM detection and prediction as follows. García-Ordás et al., [35] developed DL techniques to predict diabetic people using a CNN for classification and trained the model with PIMA-IDD-I.J. Khanam et al., [36] developed ML, and ANN methods for DM prediction. They found that the approach with RL and SVM works well on DM prediction. They designed the ANN method with various hidden layers and observed that the ANN with two hidden layers yielded an accuracy of 88.6%. Kannadasan et al., [37] aimed to classify the PIMA-IDD-I using a deep ANN with a stack encoder. Swapna et al., [38] designed a method for the classification of diabetic and normal HRV signals utilizing DL approaches. They built LSTM and CNN for obtaining complex temporal dynamic parameters from the HRV records. Shahin et al., [39] built a DL technique such as a deep belief network to classify and predict the progression of T2D. Compared to other ML techniques, the deep belief network produced results with a higher accuracy of 81.25%. Daanouni et al., [40] proposed ANN, ANN, DT, and KNN models for DM prediction using PIMA-IDD-I and DDFH-G datasets. Similarly, Rastogi and Bansal [41] applied ANN to detect DM and identify its type. For evaluation, they used the PIMA-IDD-I and achieved an accuracy of 85.09%.

III. RESEARCH GAP

DL approaches usually need lots of data, but regular data-driven methods have problems like finding enough labeled data and making the results easy to understand. The introduction of EDL as a guide and constraint in the existing

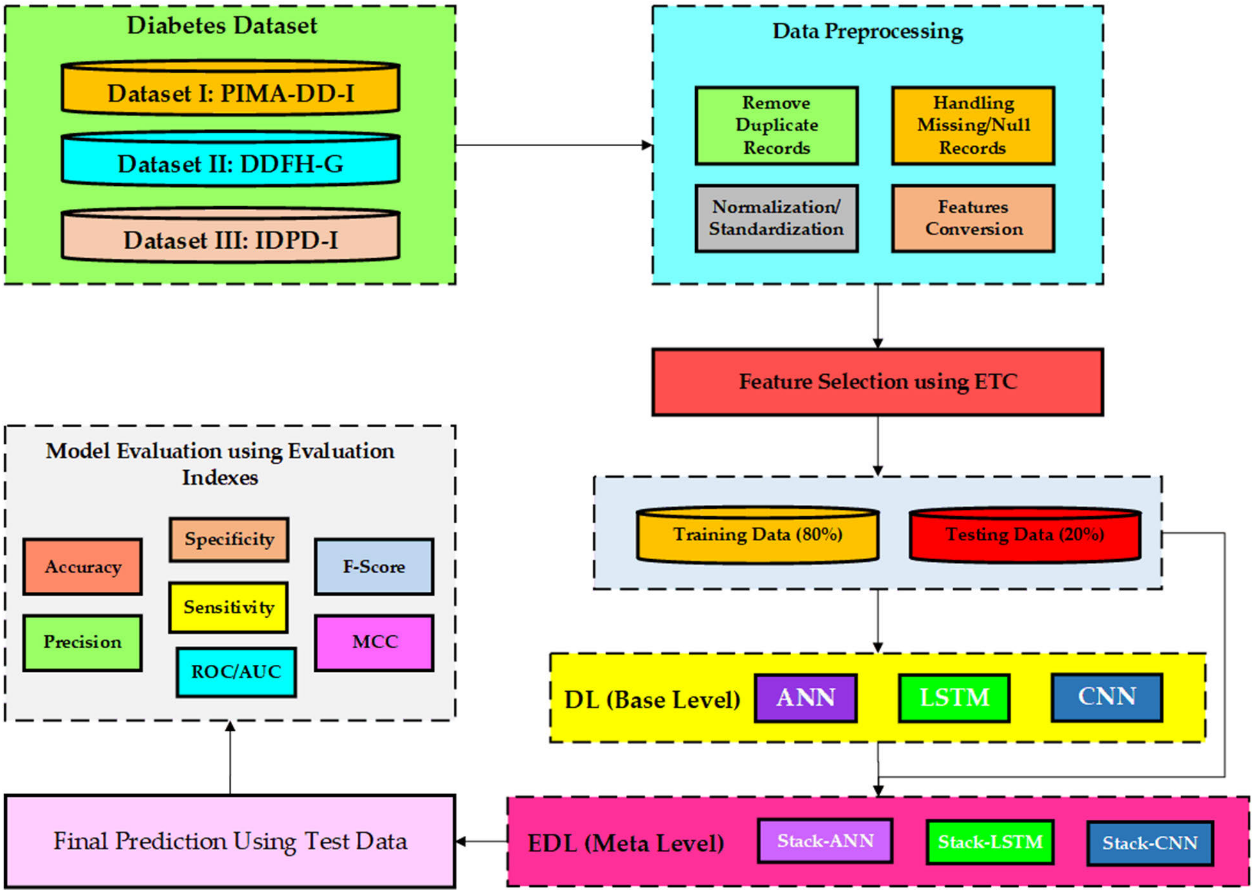


FIGURE 2. Proposed EDL framework.

data-driven models can effectively tackle these problems. For instance, there are not always enough electronic health records available to train DL models in developing healthcare decision support systems. At this time, the introduction of ensemble learning techniques in a given electronic medical dataset [42], [43] can greatly improve the effectiveness of DL approaches.

The extraction of knowledge from electronic health records through DL approaches is a valuable yet challenging task for diagnosis and prediction. Diabetes, if left undiagnosed, can adversely affect various organs in the human body, such as the kidneys and liver. This disease is prevalent across all age groups. In the literature, several researchers have attempted to predict and classify diabetes using ML and DL techniques. Nevertheless, the EDL method for predicting diabetes is missing from the existing classification and prediction approaches, highlighting a research gap. However, state-of-the-art techniques encounter a significant challenge marked by reduced accuracy when handling extensive and multi-class datasets. To address this challenge, the proposed EDL aims to introduce methodologies for feature optimization and classification in predicting DM with high accuracy. This study proposes a novel clinical decision support system based on DL architectures such as ANN, CNN, and LSMT, combined with a stacking ensemble model. These DL architectures are used as base-level models. EDL is developed using stack (i.e., meta-level) by introducing innovative models such as stack-ANN, stack-CNN, and stack-LSTM to improve the prediction of DM. In previous studies, another drawback is that none of the existing models have been trained on the different types of diabetes datasets. In this study, three different diabetes datasets are used to train the proposed EDL models. The datasets include a small PIMA-IDD-I with 768 samples, a large DDFH-G with 2000 samples, and a multi-class IDPD-I with 1000 samples. Moreover, the proposed study includes a comparison with current state-of-the-art studies and discusses the significance of the results obtained by the EDL system.

IV. PROPOSED ENSEMBLE DEEP LEARNING FRAMEWORK

The proposed EDL framework is depicted in Figure 2. The EDL framework consists of several stages, including datasets, data preprocessing, feature selection, data splitting, base-level models (ANN, LSTM, and CNN), meta-level models (stack-CNN, stack-LSTM, and stack-ANN, and evaluation using performances. To implement and execute the EDL framework, the Anaconda framework (Jupyter Notebook), and Python programming language were used. The implementation uses ML and DL toolkits such as Scikit-learn, NumPy, Keras, TensorFlow, etc. The different stages of the proposed EDL framework are discussed in detail as follows:

A. DIABETES DATASETS

The EDL is trained and experimented using three diabetes datasets, including PIMIA-IDD-I [44], DDFH-G [45], and IDPD-I [46]. The datasets consist of a binary classification and a multi-class classification, designed for predicting DM across two classes (DM-positive and DM-negative) and three classes (diabetic, non-diabetic, and prediabetes), respectively. The suggested datasets are categorized into small, large, and multi-class datasets, namely PIMIA-IDD-I, DDFH-G, and IDPD-I, respectively. The proposed datasets are explained in detail in the following subsections.

1) DATASET-I: PIMA INDIAN DIABETES DATASET (PIMA-IDD-I)

The PIMA-IDD-I derives from the UCI Repository [44]. There are 768 samples of diabetic patients, with nine different attributes in each sample. Participants without diabetes, those with prediabetes, and those with diabetes are distributed equally among these records. The final column serves as a binary target variable, where a value of 0 signifies the absence of diabetes in the patient, while a value of 1 signifies the presence of prediabetes or diabetes.

2) DATASET-II: DIABETES DATASET FRANKFURT HOSPITAL GERMANY (DDFH-G)

The DDFH-G dataset originates from the Hospital Frankfurt Germany [45], comprising 2000 cases, each with eight features, while the PIMA-IDD-I includes 768 patients, also with eight features. The DDFH-G dataset shares similar features with datasets utilized in studies by Pradhan et al. and Dwivedi; the key difference lies in the larger number of data points present in DDFH-G. The features encompass a spectrum ranging from the number of pregnancies to skin thickness. In the dataset, it is noted that specific features such as glucose, insulin, blood pressure, BMI, and skin thickness exhibit zero values, which is not practically feasible. Consequently, these instances are preserved as missing data and are substituted with the mean value of the respective feature column containing the missing values.

3) DATASET-III: IRAQI DIABETES PATIENTS DATASET (IDPD-I)

The IDPD-I dataset was gathered from the Iraqi society, specifically obtained from the Medical City Hospital laboratory and the Specialized Center for Endocrinology and Diabetes at Al-Kindy Teaching Hospital [46]. The IDPD-I has 1000 samples. The IDPD-I includes 565 males and 435 females, ranging from 20 to 79 years old. It is categorized into three classes: Diabetic (Y) with 837 samples, Non-Diabetic (N) with 103 samples, and Predicted Diabetic (P) with 53 samples. These samples are characterized by 11 physical examination indicators such as Patient Number, Age, Gender, Blood Sugar Level, Cholesterol, Creatinine Ratio (Cr), Urea, Body Mass Index (BMI), HBA1C, Fasting Lipid Profile (VLDL, LDL, Triglycerides, HDL Cholesterol), and Class (representing the diabetic patient's status as Diabetic, Non-Diabetic, or Pre-Diabetic).

B. DATA PREPROCESSING

To prepare the data for EDL training, any data entries including missing values were systematically eliminated from the dataset to ensure better results. Initially, the three district datasets were cleaned by removing all the records with null/missing and duplicate values. In the datasets, specific attributes such as glucose, blood pressure, skin thickness, insulin, and BMI exhibit zero values, which is not practically feasible. Consequently, these instances are preserved as missing samples and are substituted with the mean value of the respective attribute column containing the missing values. Additionally, the data values experienced normalization through the application of the MinMax Scaler before being used in the model. The initial dataset exhibited a significant class imbalance, with only 20% of the samples representing the positive class, while the majority of the samples were negative. To address this issue, SMOTE (Synthetic Minority Over-sampling Technique) was performed to oversample entries from the minority class, thus achieving a more balanced distribution of the dataset [47]. Given the categorical nature of the data and the binary target variable, we applied classification models to evaluate the accuracy of DM prediction. This comprehensive data preprocessing approach aims to enhance the performance and the ability of the proposed EDL to make promising predictions.

C. FEATURE SELECTION USING EXTRA TREE CLASSIFIER

The process of feature selection requires choosing relevant features from a distinct set to decrease computational power and complexity while enhancing accuracy. By utilizing the model characteristics property, the importance of each feature in PIMA-IDD-I can be assessed. Each feature is given a score based on its impact on the performance variable. This indicates the relevance of each feature. A higher score indicates greater importance. In this study, the ETC with the Gini relevance technique is applied to obtain the most key features from the data. The ETC, which comes with an integrated class designed for evaluating feature importance, proves especially helpful when examining the significance of features in tree-based models. Utilizing the ETC allows the model to locate key features that play an important role in improving the predictive performance and control over-fitting during the training of the EDL [42]. The ETC offers significant comprehensions for the analysis. The significance of various features may vary due to the randomness of feature samples, as reported in Table 2.

TABLE 2. Important features selection using the ETC approach.

Feature	Ranking (PIMA-IDD-I)	Ranking (DDFH-G)
Glucose	0.23179060525382080	0.24127714276608095
Age	0.14556459190627932	0.14794762705465844
BMI	0.14275252575249722	0.14028730282586185
DiabetesPedigreeFunction	0.11871803451568212	0.11666415905835303
Pregnancies	0.10566878941654180	0.10634495645710182
BloodPressure	0.09837485112184098	0.09702644456837250
SkinThickness	0.08139176253489086	0.07773615953673563
Insulin	0.07573883949844681	0.07271620773283596

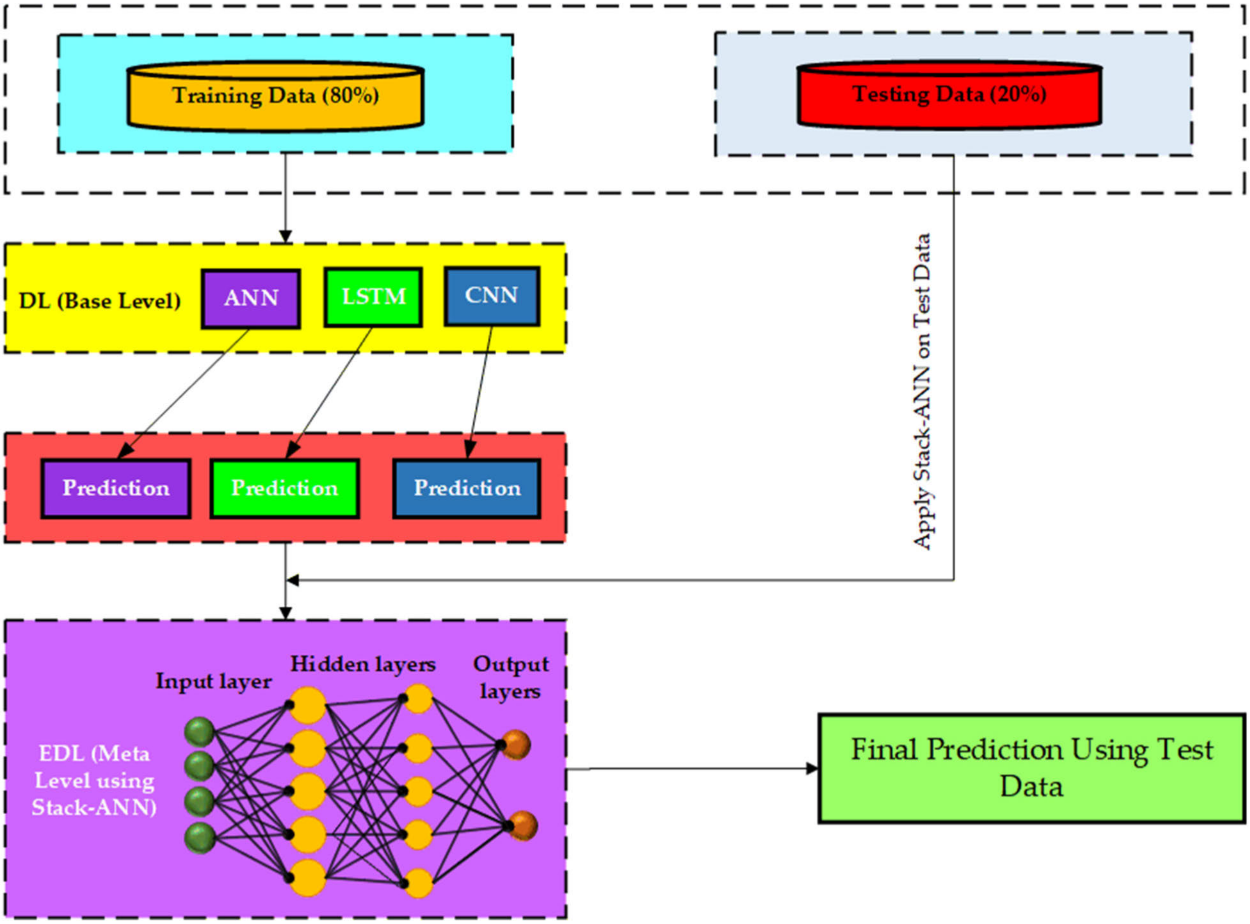


FIGURE 3. Proposed EDL-based Stack-ANN framework.

D. DATA SPLITTING

Data splitting is a method that requires distributing a dataset into smaller subsets, and in this case, the normalized pre-processed diabetes datasets are divided into two portions: the training set and the test set. The proposed datasets are divided into 80:20, where 80% of the data is used to train the proposed EDL model, and the remaining 20% is reserved for EDL model evaluation (testing). The training set comprises data samples used for learning and adjusting the model parameters, while the test set consists of data samples utilized to assess the efficiency of the EDL.

E. STACKING ENSEMBLE LEARNING

Ensemble learning is an ML technique where the predictions of multiple models can be combined to make more accurate predictions than each single or base-level model. In this study, a stacking ensemble model is used to combine the prediction results of the base-level models by evaluating a

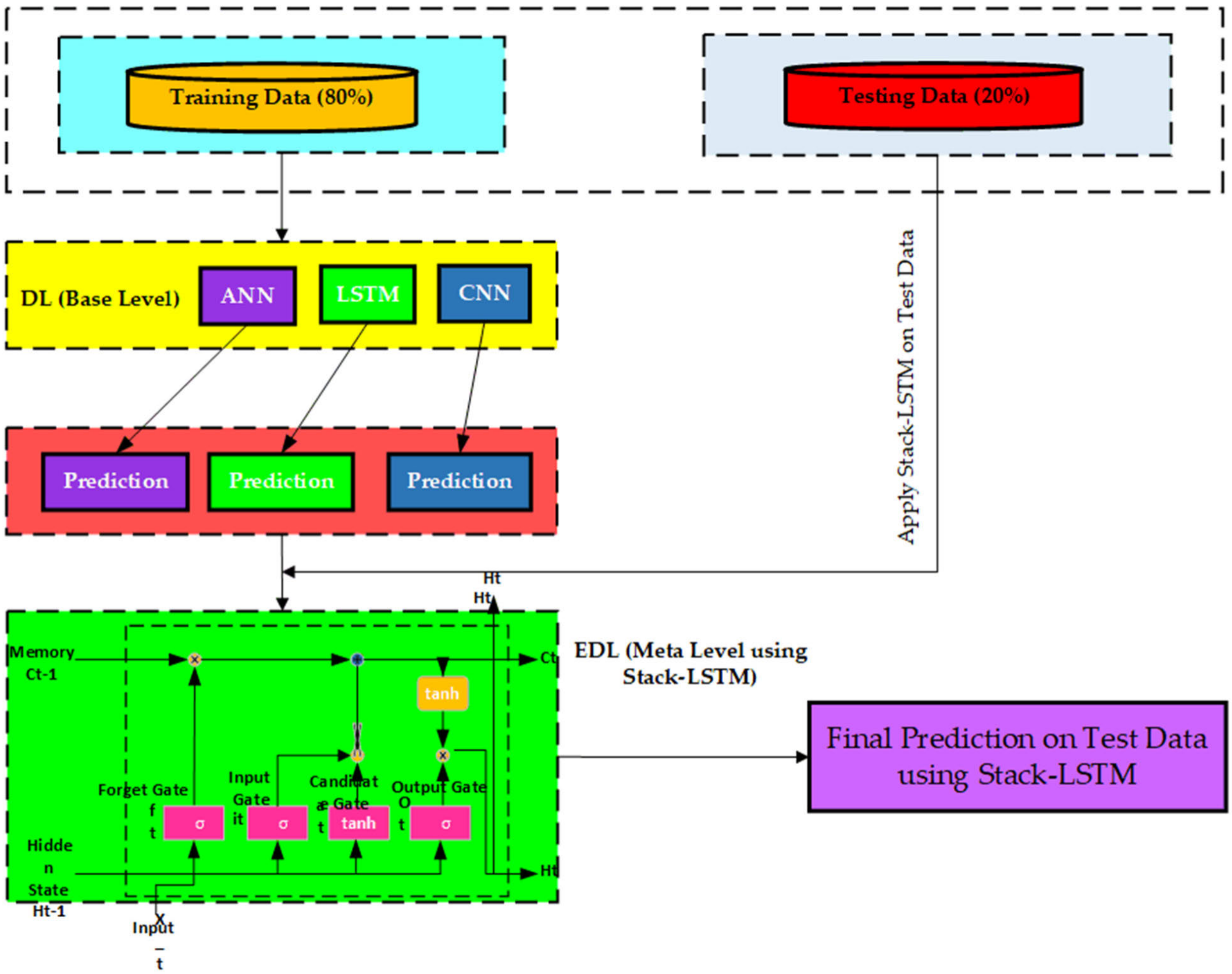


FIGURE 4. Proposed EDL-based Stack-LSTM framework.

new meta-level model. To develop EDL, ANN, LSTM, and CNN are proposed as base-level models, while stack-ANN, stack-LSTM, and stack-CNN are proposed as meta-level models. In the proposed study, the predictions generated by base-level (e.g., ANN, LSTM, and CNN) models serve as input parameters for meta-level (e.g., stack-ANN, stack-LSTM, and stack-CNN) models. The results of the base-level models are optimized through stack model. This helps in further enhancing the prediction result by taking advantage of base-level models. The EDL-based stacking models such as stack-ANN, stack-LSTM, and stack-CNN are graphically visualized in Figures 3, 4, and 5, respectively.

1) BASE MODELS

The proposed EDL framework is implemented based on three base-level models such as ANN, LSTM, and CNN. *ANN* is used in acquiring complex features in tabular data [47]. *LSTM* is used to process sequential data, making it suitable for time-series data such as patient health records [47]. *CNN* can be applied to image data but can also adapt to obtaining features from structured data [47]. Each of these base-level models independently analyzes and evaluates the input data and generates its predictions.

2) META MODELS

In the proposed EDL framework, three meta-level models are used, including stack-ANN, stack-LSTM, and stack-CNN. These models are explained briefly as follows:

a: STACK-ANN

Stack-ANN is a neural network that operates at a higher level than base-level models. Instead of directly predicting the target variable (e.g., DM positive and DM negative), the stack-ANN is designed to work on the predictions generated by the base-level models. Figure 3 depicts a framework of the EDL-based Stack-ANN.

During training, the base-level models (ANN, LSTM, and CNN) process the input data and make their respective predictions. The stack-ANN is then trained on the predictions (output) generated by the base-level models. The goal is for

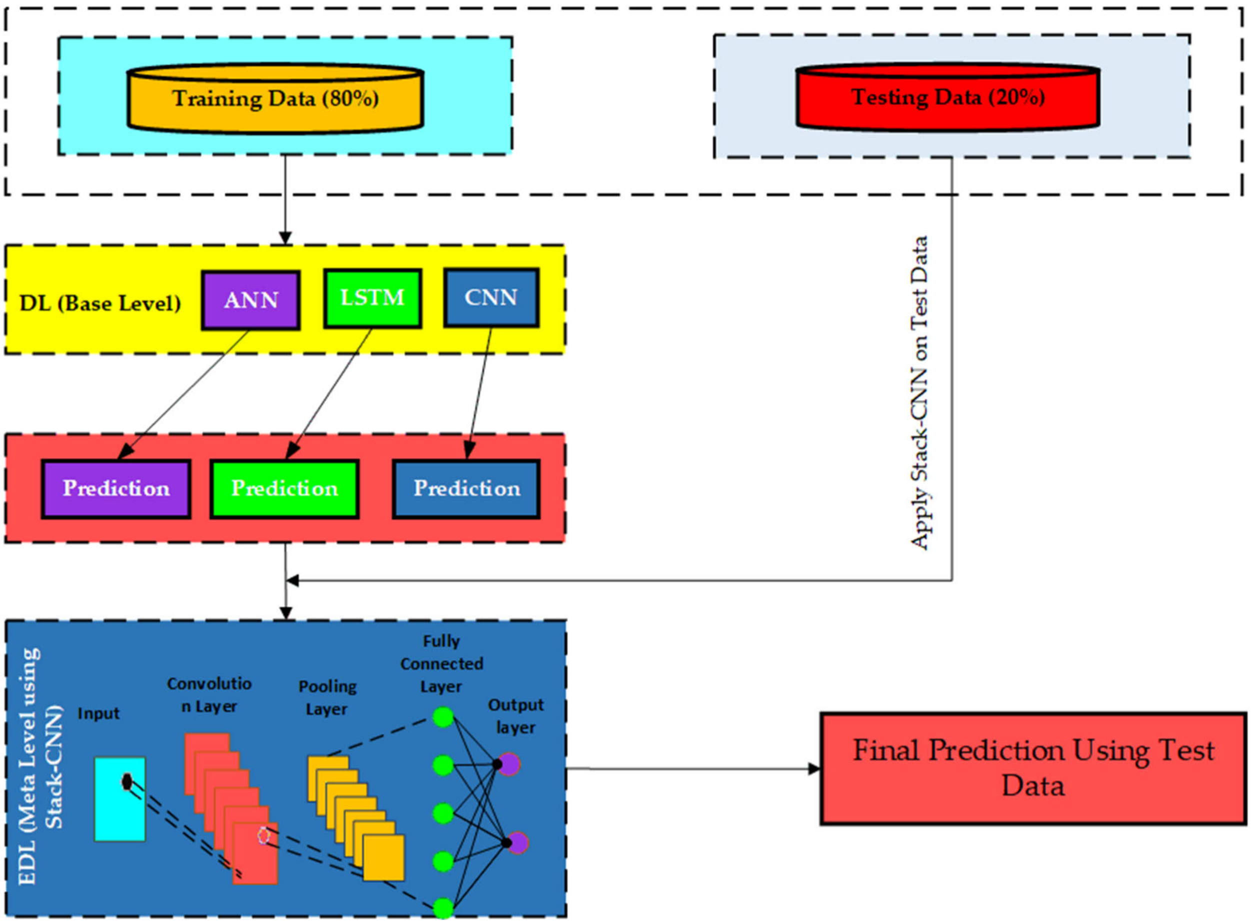


FIGURE 5. **Proposed EDL-based Stack-CNN framework.**

the stack-ANN to learn how to combine these predictions to make a more accurate and robust final prediction. Then various techniques are applied to combine these predictions, such as weighted averaging, a Dense layer with 32 nodes and Rectified Linear Unit (ReLU) function, a Dropout layer with a dropout rate of 0.5, a final Dense layer with 1 unit and a sigmoid activation function for binary classification and a SoftMax for multi-class classification. In addition, hyperparameters such as 20 epochs, a batch size of 32, a learning rate of 0.01, and a validation split of 0.2 (20% of the training data) are used for training the stack-ANN.

b: STACK-LSTM

Stack-LSTM is a neural network that operates at a higher level than base-level models. Instead of directly making predictions for the targeted task e.g., DM positive and DM negative, the stack-LSTM model processes the predictions generated by the base-level models. Figure 4 depicts a framework of the EDL-based Stack-ANN.

Throughout the training phase, the base-level models (ANN, LSTM, and CNN) process the input data and generate their respective predictions. The stack-LSTM is then trained on these predictions provided by the base-level models. The objective is for the stack-LSTM to learn how to integrate these predictions to produce a more accurate prediction result. In other words, the stack-LSTM uses the outputs from the base-level models as input features. It applies various techniques to merge these predictions, such as an LSTM layer with 64 units and ReLU activation, a Dense layer with 32 units and ReLU activation, and a final Dense layer with 1 unit and a sigmoid activation function for binary classification and a SoftMax for multi-class classification. The training is performed for 20 epochs with a batch size of 32 and a learning rate of 0.01.

c: STACK-CNN

Stack-CNN is a neural network that operates at a higher level compared to base-level models. Instead of directly making predictions for the DM prediction, the stack-CNN model processes the predictions made by the base-level models. Figure 5 depicts a framework of the EDL-based Stack-ANN.

During the training phase, the base-level models (ANN, LSTM, and CNN) process the input data and generate their respective predictions. The stack-CNN is then trained on these predictions provided by the base-level models. The stack-CNN uses the predictions from the base-level models

TABLE 3. Performance evaluation of the proposed base level and meta-level models using PIMA-IDD-I.

Evaluation Indexes	DL (Base models)			EDL (Meta models)		
	ANN	LSTM	CNN	Stack-ANN	Stack-LSTM	Stack-CNN
Accuracy	0.9277	0.9107	0.9073	0.9881	0.9723	0.9481
Precision	0.9265	0.9143	0.9067	0.9793	0.9734	0.9393
Specificity	0.9285	0.9101	0.9043	0.9840	0.9654	0.9388
Sensitivity	0.9276	0.9036	0.8954	0.9802	0.9644	0.9402
F-Score	0.9101	0.9051	0.9032	0.9780	0.9706	0.9378
MCC	0.9043	0.8953	0.8898	0.9428	0.9307	0.9228

as input features. Stack-CNN utilizes various techniques to merge these predictions, such as adding a 1D convolutional layer (Conv1D) with 32 filters, a kernel size of 3, a ReLU activation function, a 1D max-pooling layer (MaxPooling1D) with a pool size of 2, Flatten the output of the previous layers using Flatten(), a Dense layer with 64 units and ReLU activation, and a final Dense layer with 1 unit and a sigmoid activation function for binary classification and SoftMax for multi-classification tasks. The model is trained over 20 epochs. A batch size of 32 is used, meaning the weights were updated after processing each batch of 32 samples and a learning rate of 0.01. To merge the predictions from the base-level models, the stack-CNN captures complex patterns and relationships in the data that can be neglected by base-level models.

F. EVALUATION METRICS

The proposed EDL system is evaluated based on major evaluation metrics, namely accuracy, f-score, sensitivity, precision, specificity, ROC/AUC, and MCC. These indexes evaluate the EDL performance in a DM prediction task [47]. The following subsections elaborate on each evaluation index.

1) ACCURACY

Using Equation 1, accuracy is stated as the percentage of accurately predicted instances (including True Positives (TP) and True Negatives (TN)) relative to all instances (including True Positives (TP), True Negatives (TN), False Positive (FP), and False Negative (FN)). It is a common metric for evaluating classification models.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{FN} + \text{FP} + \text{TP} + \text{TNF}} \tag{1}$$

2) PRECISION

Equation 2 indicates that precision is the percentage of TP predictions among FP and TP predictions generated by the model. It assesses the ability of the EDL to make correct positive predictions while minimizing FP. Higher precision values indicate fewer FP.

$$\text{Preccision} = \frac{\text{TP}}{\text{FP} + \text{TP}} \tag{2}$$

3) SPECIFICITY

Equation 3 is used to calculate specificity, which is a TN rate that expresses the percentage of TN classes that the EDL accurately predicted.

$$\text{Specificity} = \frac{\text{TN}}{\text{FN} + \text{TN}} \tag{3}$$

4) SENSITIVITY

Sensitivity measures the proportion of TP instances correctly predicted by the EDL applying Equation 4. It assesses the EDL to capture positive instances.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{FN} + \text{TP}} \tag{4}$$

5) F-SCORE

F-Score is the harmonic mean of precision and sensitivity as determined in Equation 5. It provides a balanced measure of an EDL performance by considering both FP and FN.

$$\text{F} - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

6) MATTHEWS CORRELATION COEFFICIENT (MCC)

The quality of binary classifications is evaluated using MCC. It is especially helpful when handling unbalanced datasets because it considers TP and FP as well as TN and FN. Mathematically, it can be calculated by Equation 6.

$$\text{MCC} = \frac{\text{TN} \times \text{TP} - \text{FN} \times \text{FP}}{\sqrt{(\text{TP} + \text{FN})\ \text{x}\ (\text{TN} + \text{FP})\ \text{x}\ (\text{TP} + \text{FP})\ \text{x}(\text{TN} + \text{FN})}} \tag{6}$$

V. RESULTS

This section presents experimental results on the performance of the proposed EDL in terms of accuracy, f-score, sensitivity, precision, specificity, ROC/AUC, and MCC using PIMA-IDD-I, DDFH-G, and IDPD-I for predicting DM. Tables 3, 4, and 5 outline a performance comparison of the DL-based base-level models (ANN, LSTM, CNN) and EDL-based meta-level models (stack-LSTM, stack-CNN, stack-ANN), based on various evaluation metrics. The evaluation metrics include accuracy, precision, sensitivity, specificity, f-score, and MCC.

TABLE 4. Performance evaluation of the proposed base level and meta-level models using DDFH-G.

Evaluation Indexes	DL (Base models)			EDL (Meta models)		
	ANN	LSTM	CNN	Stack-ANN	Stack-LSTM	Stack-CNN
Accuracy	0.9477	0.9324	0.9273	0.9951	0.9836	0.9744
Precision	0.9365	0.9365	0.9367	0.9892	0.9834	0.9704
Specificity	0.9285	0.9285	0.9143	0.9902	0.9877	0.9688
Sensitivity	0.9402	0.9332	0.9251	0.9897	0.9854	0.9685
F-Score	0.9401	0.9301	0.9232	0.9922	0.9743	0.9606
MCC	0.9364	0.9323	0.9399	0.9728	0.9620	0.9507

TABLE 5. Performance evaluation of the proposed base level and meta-level models using IDPD-I.

Evaluation Indexes	DL (Base models)			EDL (Meta models)		
	ANN	LSTM	CNN	Stack-ANN	Stack-LSTM	Stack-CNN
Accuracy	0.9462	0.9087	0.9025	0.9845	0.9788	0.9689
Precision	0.9343	0.9067	0.9122	0.9893	0.9803	0.9690
Specificity	0.9274	0.9132	0.9013	0.9788	0.9775	0.9722
Sensitivity	0.9388	0.9055	0.9054	0.9802	0.9785	0.9662
F-Score	0.9424	0.9024	0.9132	0.9878	0.9706	0.9670
MCC	0.9343	0.8923	0.8902	0.9602	0.9544	0.9487

Table 3 lists a comprehensive evaluation of different models on the PIMA-IDD-I dataset, categorized into DL (base-level) and EDL (meta-level) models. Among the base-level models, the ANN exhibits the highest accuracy (0.9277), precision (0.9265), and sensitivity (0.9276), while the LSTM and CNN also yield competitive results. The meta-level models, which are ensembles of the base models, consistently outperform each base model's overall evaluation metrics. Specifically, the stack-ANN achieves significant accuracy (0.9881), precision (0.9793), and specificity (0.9840), indicating superior overall performance. The EDL models, including stack-LSTM, achieved promising results based on accuracy, precision, specificity, sensitivity, f-score, and MCC of 0.9723, 0.9734, 0.9654, 0.9644, 0.9706, and 0.9307, respectively, using PIMA-IDD-I. Using PIMA-IDD-I, stack-CNN also performed better achieving accuracy, precision, specificity, sensitivity, f-score, and MCC scores of 0.9481, 0.9393, 0.9388, 0.9402, 0.9378, and 0.9228, respectively. This demonstrates the efficiency of combining base models to enhance predictive capabilities. The MCC values further highlight the ability of the proposed EDL in binary classification tasks. Overall, the results underscore the benefits of EDL in improving the robustness and accuracy of the proposed EDL models for the PIMA-IDD-I dataset, which can also be seen in Figure 6.

Table 4 and Figure 7 provide a detailed assessment of EDL performance on the DDFH-G dataset, distinguishing between DL (base-level) and EDL (meta-level) models. The comparison is also graphically depicted in Figure 7. In the realm of base models, the ANN stands out with the highest accuracy (0.9477), precision (0.9365), sensitivity (0.9402), and F-Score (0.9401), while the LSTM and CNN models also exhibit competitive results. Notably, the EDL (meta-level) models, representing ensembles of the base models, consistently surpass base models across all evaluation metrics. The stack-ANN demonstrates exceptional performance, achieving a notable accuracy of 0.9951 and high values for precision, specificity, sensitivity, and F-Score. Using DDFH-G, the stack-LSTM obtained robust performance with accuracy, precision, specificity, sensitivity, F-score, and MCC scores of 0.9836, 0.9834, 0.9877, 0.9854, 0.9743, and 0.9620, respectively. Similarly, using DDFH-G, the stack-CNN showed robust performance with accuracy, precision, specificity, sensitivity, F-score, and MCC scores of 0.9744, 0.9704, 0.9688, 0.9685, 0.9606, and 0.9507, respectively. These results highlight the potential of the

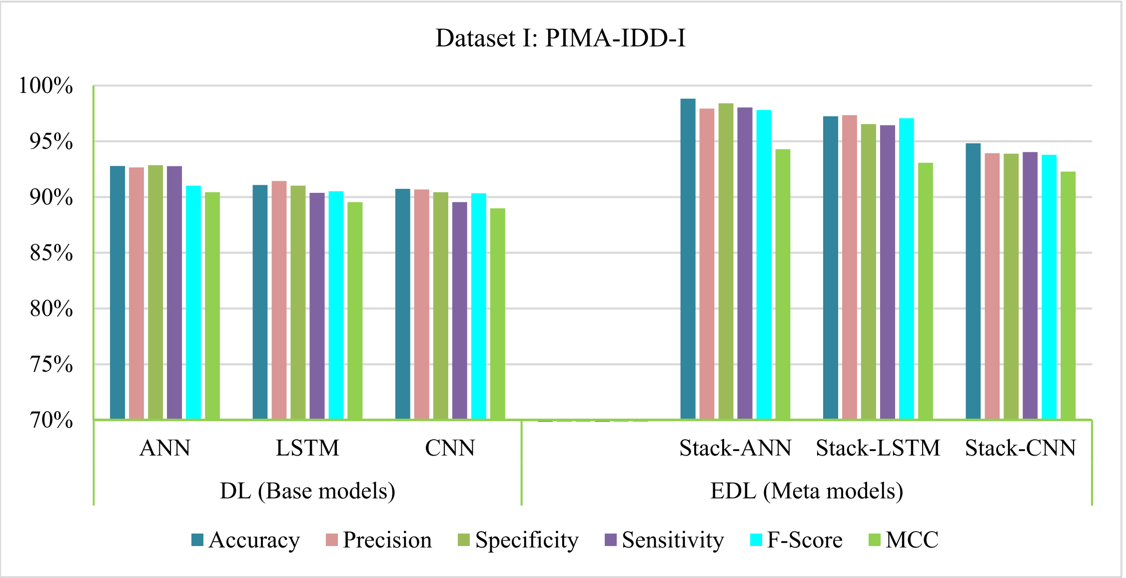


FIGURE 6. Performance of the proposed DL (base level) and EDL (meta-level) models using PIMA-IDD-I.

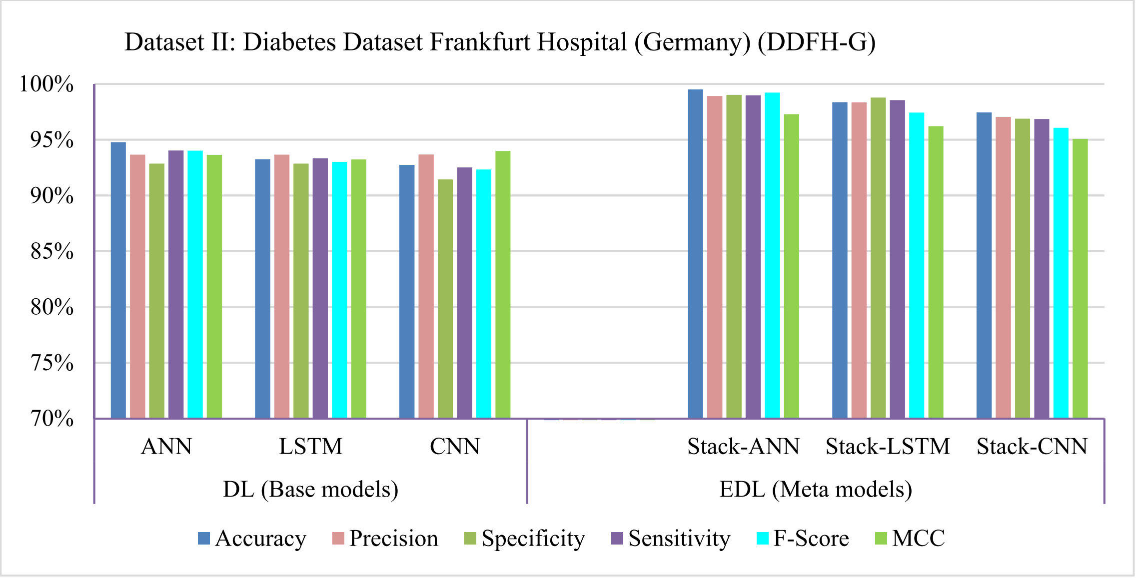


FIGURE 7. Performance of the proposed DL (base level) and EDL (meta-level) models using DDFHG.

proposed EDL by combining the predictions from base models.

Table 5 provides a comprehensive evaluation of EDL performance on the IDPD-I dataset, classifying models into two categories: base level and meta-level as outlined in Figure 8. Among the base models, the ANN emerges as the top performer with high accuracy (0.9462), precision (0.9343), and sensitivity (0.9388). The LSTM and CNN models also exhibit competitive results, showcasing diversity in the base-level architectures. Notably, the meta-level models, representing ensembles of the base models, outshine base models across multiple evaluation metrics. The stack-ANN stands out with exceptional accuracy (0.9845), precision (0.9893), and sensitivity (0.9802), demonstrating the effectiveness of combining diverse base models. Using IDPD-I, the stack-LSTM also performed better, achieving accuracy, precision, specificity, sensitivity, F-score, and MCC scores of 0.9788, 0.9803, 0.9775, 0.9785, 0.9706, and 0.9544, respectively. Similarly, the stack-CNN yielded robust performance, achieving accuracy, precision, specificity, sensitivity, F-score, and MCC scores of 0.9689, 0.9690, 0.9722, 0.9662, 0.9670, and 0.9487, respectively. These findings show the robust performance of the proposed EDL models in enhancing predictive accuracy using IDPD-I. The achieved results strengthen the power

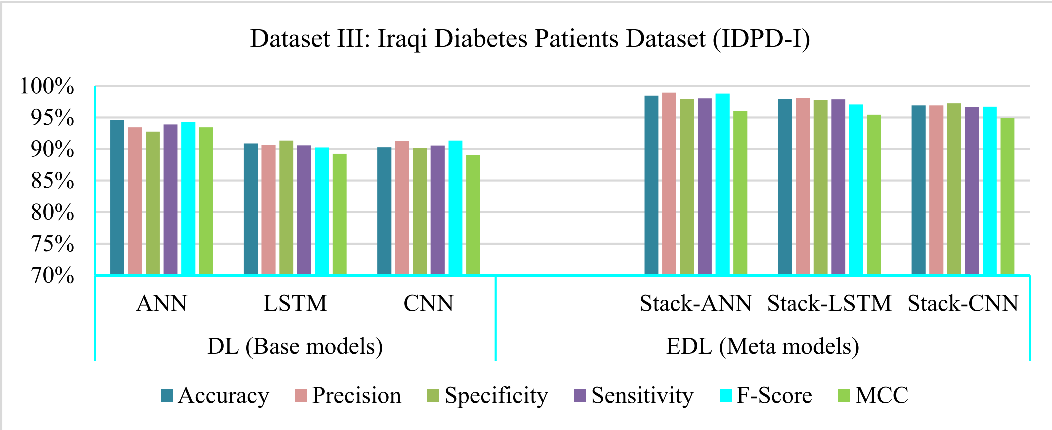


FIGURE 8. **Performance of the proposed DL (base level) and EDL (meta-level) models using IDPD-I.**

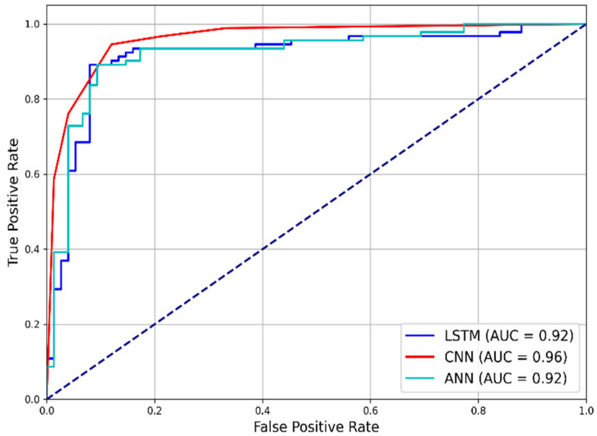


FIGURE 9. **ROC/AUC score for the proposed DL (base models): LSTM, CNN, and ANN using PIMA-IDD-I.**

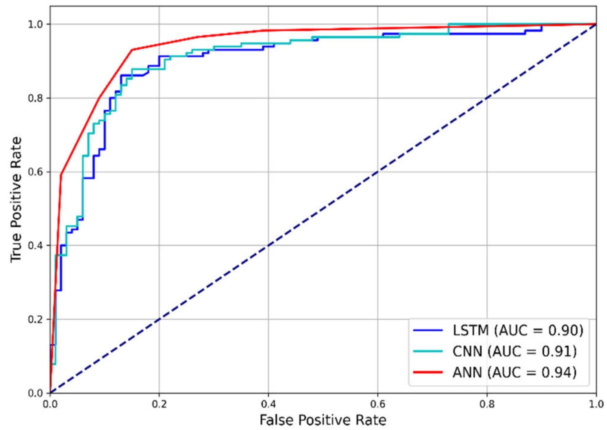


FIGURE 10. **ROC/AUC score for the proposed DL (base models): LSTM, CNN, and ANN using IDPD-I.**

of the EDL models in multi-class classification tasks. This highlights the potential of EDL to improve model robustness and performance in predicting diseases, thereby contributing to healthcare decision support systems.

Figures 9, 10, and 11 present an evaluation of the performance of three DL (base-level models), including ANN, CNN, and LSTM on three different datasets: PIMA-IDD-I, DDFH-G, and IDPD-I. In Figure 9, using PIMA-IDD-I, the ROC curve and AUC values illustrate the discriminative capabilities of each model in classifying DM as positive and negative. The ANN emerges as the standout performer with an AUC of 0.92, signifying robust accuracy for predicting DM. The CNN follows closely with an AUC of 0.96, and the LSTM falls in between with an AUC of 0.92. Similarly, Figure 10 depicts the evaluation of these models using the IDPD-I. The LSTM, CNN, and ANN yield AUC values of 0.90, 0.91, and 0.94, respectively. Figure 11 reveals the ANN as the top performer with an AUC of 0.97, followed by the LSTM with 0.96 and the CNN with 0.94 using DDFH-G. These findings collectively underscore the performance

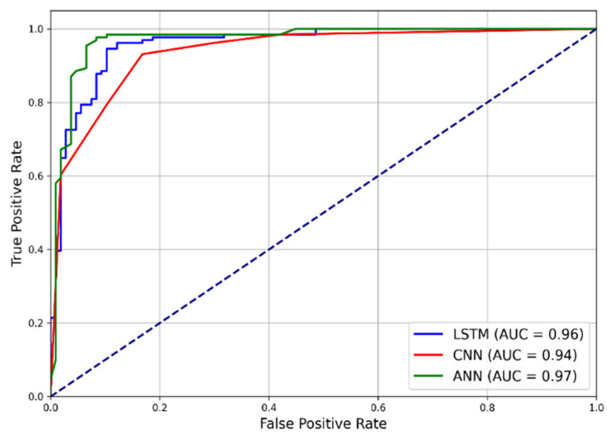


FIGURE 11. **ROC/AUC score for the proposed DL (base models): LSTM, CNN, and ANN using DDFH-G.**

variations among the base-level models, highlighting the efficiency of base-level models in predicting DM.

Figure 12 presents a graphical comparison using the ROC curve and AUC values for the proposed EDL models applied

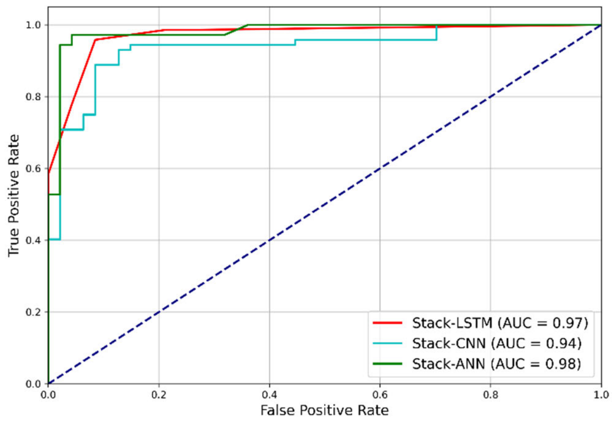


FIGURE 12. ROC/AUC score for the proposed EDL models: stack-LSTM, stack-CNN, and stack-ANN using PIMA-IDD-I.

to the PIMA-IDD-I dataset. Figure 12 illustrates the performance of each EDL model (stack-ANN, stack-CNN, and stack-LSTM) in terms of their respective AUC values. The stack-ANN achieves the highest AUC value (0.98), representing strong performance in classifying and predicting both positive and negative cases of DM. The stack-CNN model follows closely with a slightly lower AUC value (0.94), while the stack-LSTM model falls in between with an AUC of 0.97. Using IDPD-I, Figure 13 exhibits the performance of these models, where the stack-ANN, stack-LSTM, and stack-CNN achieved an AUC score of 0.98, 0.98, and 0.97, respectively. Similarly, using DDFH-G, Figure 14 shows AUC scores for the stack-ANN, stack-LSTM, and stack-CNN as 0.99, 0.98, and 0.97, respectively. These results show the efficiency of the proposed EDL models in classifying and predicting outcomes related to DM. The AUC values serve as a quantitative measure, with higher values indicating better performance in distinguishing between DM positive and DM negative cases.

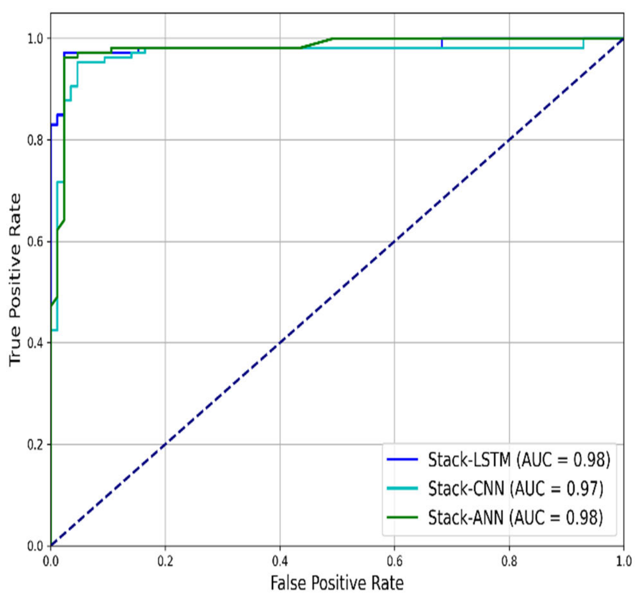


FIGURE 13. ROC/AUC scores for proposed EDL models (stack-LSTM, stack-CNN, and stack-ANN) using IDPD-I.

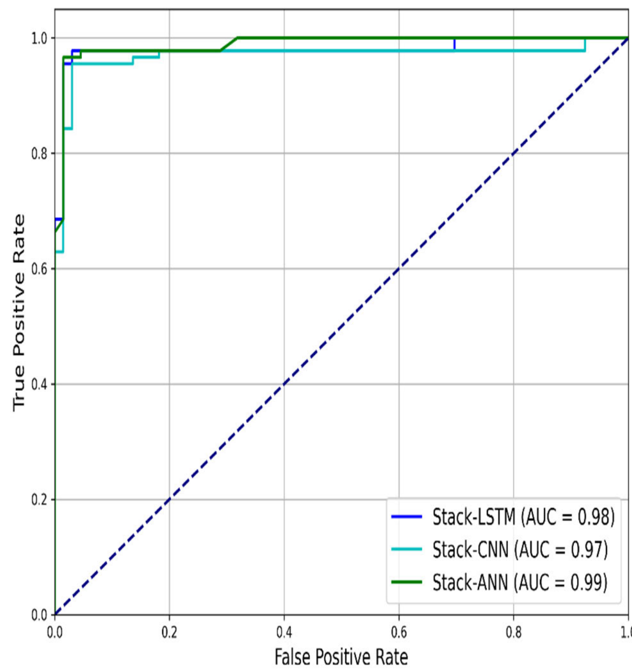


FIGURE 14. ROC/AUC score for the proposed EDL models: stack-LSTM, stack-CNN, and stack-ANN using DDFH-G.

Figure 15 highlights the performance evaluation of the proposed EDL (meta-level) models using PIMA-IDD-I, DDFH-G, and IDPD-I. The experimental results demonstrate the performance of three different DM datasets, namely PIMA-IDD, DDFH-G, and IDPD-I, each evaluated under three EDL configurations: stack-ANN, stack-LSTM, and stack-CNN. The accuracy values show the predictive competencies of the proposed EDL models. Stack-ANN achieved the highest accuracy of 98.81%, stack-LSTM, and stack-CNN also yielded promising accuracy scores of 97.23% and 94.81% respectively, using PIMA-IDD. Similarly, DDFH-G demonstrates strong performance, achieving 99.51% accuracy with stack-ANN, 98.36% with stack-LSTM, and 97.44% with stack-CNN. Finally, using the IDPD-I, stack-ANN, stack-LSTM, and stack-CNN exhibited accuracy results of 98.45%, 97.88%, and 96.89%, respectively. These findings reveal a thorough assessment of the proposed EDL models in accurately predicting DM.

VI. DISCUSSION

The experimental results show that the EDL models significantly improve DM prediction accuracy, compared to state-of-the-art ML and DL techniques. The proposed EDL framework offers significant support to healthcare professionals in early prediction and intervention.

Table 6 reports a comparison between ML and DL methods for DM prediction applying different diabetic datasets. ML techniques include RF [27], [34], LR [48], ANN [24],

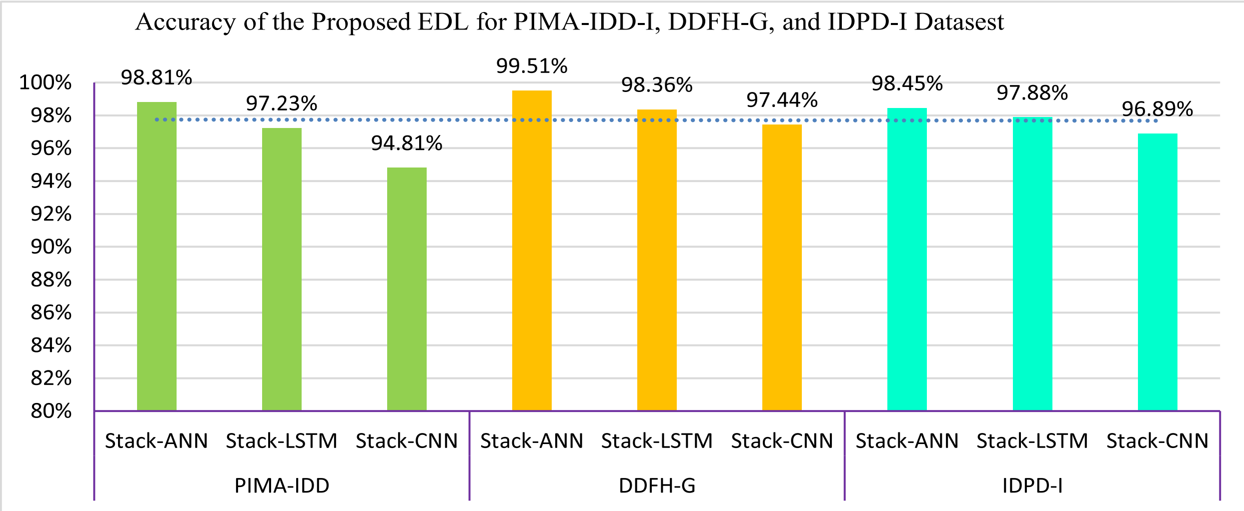


FIGURE 15. Accuracy performance for each EDL (meta-models) on three diabetes datasets.

TABLE 6. Comparing the proposed EDL models with state-of-the-art approaches.

Ref.	Dataset	Model	Accuracy
[32]	PIMA-IDD-I	RF	85.6%
[28]	PIMA-IDD-I, IDPD-I	Soft voting (LR, RF, NB)	81.0%, 97.0%
[29]	IDPD-I	Multilayer Neural Network	98%
[52]	Taipei Municipal Medical Center	LR, NN, Decision Jungle, DT	80.2%, 90.8%, 93.1%, 95.3%
[50]	DDFH-G	KNN, LR, SVM	78%, 78%, 77%
[49]	DDFH-G	ANN	96.5%
[48]	PIMA-IDD-I	LR	82.46%,
[36]	PIMA-IDD-I	Adaptive Boosting, ANN	79.42%, 88.6%
[35]	PIMA-IDD-I	CNN, BiLSTM	92.31%, 94.0%
[30]	PIMA-IDD-I, IDPD-I	DL	74.74%, 94.72%
[40]	PIMA-IDD-I	ANN, Deep ANN, DT KNN	87.76%, 80.99%, 85.68%, 95.96%
[40]	DDFH-G	ANN, Deep ANN, DT KNN	89.0%, 98%, 95.5%, 95.50%
[51]	DDFH-G, PIMA-IDD-I	RF, SVM	97.6%, 83.1%
[24]	PIMA-IDD-I	ANN	87.79%
[25], [53]	PIMA-IDD-I	NB	79.13%, 76.30%
[26]	PIMA-IDD-I	Deep Dense Layer ANN	84.42%
[27], [34]	PIMA-IDD-I	RF	88.61%, 82.26%.
[37]	PIMA-IDD-I	Deep ANN	86.26%
[39]	PIMA-IDD-I	Deep Belief Network	81.25%.
[41]	PIMA-IDD-I	ANN	85.09%
[31]	PIMA-IDD-I	RF, SVM	83%, 81.4%
[54]	PIMA-IDD-I	Light Gradient Boosting	86%
Proposed EDL Models	PIMA-IDD	Stack-ANN, Stack-LSTM, Stack-CNN	98.81%, 97.23%, 94.81%
	DDFH-G	Stack-ANN, Stack-LSTM, Stack-CNN	99.51%, 98.36%, 97.44%
	IDPD-I	Stack-ANN, Stack-LSTM, Stack-CNN	98.45%, 97.88%, 96.89%

KNN [49], SVM [50], [51], and Adaptive Boosting [36]. In addition, DL techniques include CNN and BILSTM [35], Deep belief network [39], and ANN [37], [40]. The accuracy results demonstrate the performance of these techniques on diabetes datasets such as PIMA-IDD-I, IDPD-I, and DDFH-G. Following Table 6, it can be observed that the traditional ML and DL techniques have also yielded promising results for the prediction and classification of DM [49], [40], [51]. However, the accuracy of DM prediction has been further enhanced through the application of an EDL framework. Furthermore, the proposed EDL models, including stack-ANN, stack-LSTM, and stack-CNN obtained promising accuracy of 99.51%, 98.36%, and 97.44% respectively, using DDFH-G.

It is demonstrated that among all models, the proposed EDL models, which include stack-LSTM (97.23%), stack-CNN (94.81%), and stack-ANN (98.81%) utilizing PIMA-IDD, achieved promising accuracy. Finally, EDL achieved accuracies of 98.45%, 97.88%, and 96.89% for stack-ANN, stack-LSTM, and stack-CNN, respectively, using the IDPD-I. This suggests that the proposed EDL system is demonstrated to be highly efficient for DM prediction tasks. Moreover, traditional ML models such as RF [27], [32], [34], NB [25], RF, SVM [31], and LR [48] have performed reasonably well, with accuracies ranging from 76.30% to 88.61%. It can be observed that the traditional ML models, including RF, LR, and SVM achieved reasonably good accuracy but fall short of the highest accuracy levels reached by the EDL models.

In addition, DL approaches yielded reliable accuracy results in classification as compared with existing approaches [24], [26]. The early prediction of DM is important and obtaining a higher accuracy rate in DM prediction is definitive. Hence the researchers have proposed several ML and DL approaches for DM prediction such as BiLSTM [35], ANN [48], Deep ANN [37], [40], and Deep Belief Network [39], also demonstrated competitive performance. CNN and BILSTM [35], in particular, achieved high accuracy, with 92.31% and 94.0%, respectively. These results highlight the capacity of DL models to capture complex patterns in electronic medical data.

A. IMPLICATIONS

The most significant theoretical implication of this study is the design of a novel EDL framework for DM prediction outperforms state-of-the-art methods. These models outperform most other models due to their unique architecture and a specific combination of ensemble learning approaches. This helps address the risk of overfitting and enhances the generalization capabilities of EDL, enabling it to extract valuable information from the electronic medical records leading to more accurate clinical predictions and decision support systems.

To assist medical professionals and specialists, in the initial diagnosis of diabetic patients, the EDL aims to identify emerging trends in electronic medical data, which can include new patterns and factors related to DM. This can be a significant contribution to the medical community, as it allows for early prediction and intervention. The incorporation of the improved EDL into clinical decision support systems can enhance early prediction and personalized treatment plans, thereby enhancing patient outcomes and reducing the burden on healthcare organizations.

B. LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

Despite the benefits of the proposed EDL, we also acknowledge challenges related to data quality, EDL complexity, validation, and generalization to other diseases. The accuracy of predictions heavily depends on the quality of input data. Inaccurate or incomplete medical records can lead to unreliable predictions and affect the performance of the prediction.

To address these drawbacks, future research shall focus on introducing a novel automatic clinical decision-making system using reinforcement learning techniques for DM prediction and prevention. We plan to design a reinforcement learning-based smart system that will recommend personalized treatment and food to T1D and T2D patients. The proposed EDL can further be extended to predict other chronic diseases. In the future, we also plan to use other diabetes datasets to validate the proposed EDL.

Additionally, in the future, hybrid approaches can be explored [47], which combine the strengths of DL and traditional techniques to further enhance the DM prediction system aiding health organizations. For instance, further optimization techniques can be applied to design DL approaches more computationally efficiently.

VII. CONCLUSION AND FUTURE WORK

In this study, we introduced a clinical decision support system that explored the power of EDL to predict diabetes. The proposed EDL utilized a combination of various DL models, including CNN, LSTM, and ANN. By merging the predictions of DL approaches, the ensemble approach with meta-models, including stack-ANN, stack-CNN, and stack-LSTM was implemented. This research introduced a novel EDL approach for binary classification and multi-class classification designed for predicting diabetes across two classes (DM-positive and DM-negative) and three classes (diabetic, non-diabetic, and prediabetes), respectively. The proposed EDL was trained using small, large, and multi-class diabetes datasets including, I. PIMA-IDD-I, II. DDFH-G, and III. IDPD-I, respectively. For feature selection, the ETC technique was used to obtain the important features from these datasets. This allows the model to identify key features that play an important role in enhancing the predictive performance and control over-fitting during the training of the EDL. The EDL was evaluated using major evaluation techniques such as accuracy, precision, sensitivity, specificity, F-score, MCC, and ROC/AUC.

Among the proposed models, the stack-ANN performed exceptionally well on three datasets DDFH-G, PIMA-IDD-I, and IDPD-I with an accuracy of 99.51%, 98.81%, and 98.45%, respectively. The performance of EDL in stack-ANN underscores its potential for clinical application. The result demonstrated that the predictive EDL approach can predict diabetics with high accuracy. The stack-LSTM and stack-CNN models also performed well, with high accuracy, precision, MCC, ROC/AUC, and sensitivity scores. This study helps overcome the risk of overfitting and enhances the generalization capabilities of EDL, leading to more accurate clinical predictions for DM and other chronic diseases. In the future, the EDL can be used to predict various chronic diseases, including heart disease, kidney disease, lung cancer, breast cancer, etc. contributing to healthcare organizations.

REFERENCES

[1] S. Binhowemel, M. Alfakhri, K. AlReshaid, and A. Alyani, "Role of artificial intelligence in diabetes research diagnosis and prognosis: A narrative review," *J. Health Inform. Developing Countries*, vol. 17, no. 2, pp. 1–12, Aug. 2023. [Online]. Available: https://www.jhidc.org/index.php/jhidc/article/view/410

[2] R. M. Alamoudi et al., "Fasting Ramadan in patients with T1DM—Saudi Arabia versus other countries during the COVID-19 pandemic," *Diabetes Metabolic Syndrome, Clin. Res. Rev.*, vol. 17, no. 1, Jan. 2023, Art. no. 102676, doi: 10.1016/j.dsx.2022.102676.

[3] B. V. V. S. Prasad, S. Gupta, N. Borah, R. Dineshkumar, H. K. Lautre, and B. Mouleswararao, "Predicting diabetes with multivariate analysis an innovative KNN-based classifier approach," *Preventive Med.*, vol. 174, Sep. 2023, Art. no. 107619, doi: 10.1016/j.ypmed.2023.107619.

[4] World Population Rev. *Diabetes Rates By Country 2023*. Accessed: Feb. 10, 2023. [Online]. Available: https://worldpopulationreview.com/country-rankings/diabetes-rates-by-country

[5] V. Jaiswal, A. Negi, and T. Pal, "A review on current advances in machine learning based diabetes prediction," *Primary Care Diabetes*, vol. 15, no. 3, pp. 435–443, Jun. 2021, doi: 10.1016/j.pcd.2021.02.005.

[6] A. Mujumdar and V. Vaidehi, "Diabetes prediction using machine learning algorithms," *Proc. Comput. Sci.*, vol. 165, pp. 292–299, Jan. 2019.

[7] R. Gupta, S. Lodha, K. K. Sharma, S. K. Sharma, J. S. Makkar, A. Bana, V. Natani, S. Kumar, S. Bharati, and S. K. Sharma, "Association of type 2 diabetes with coronary risk factors, clinical presentation, angiography, coronary interventions and follow-up outcomes: A single centre prospective registry," *Diabetes Metabolic Syndrome, Clin. Res. Rev.*, vol. 17, no. 2, Feb. 2023, Art. no. 102709, doi: 10.1016/j.dsx.2023.102709.

[8] J. Chaki, S. T. Ganesh, S. K. Cidham, and S. A. Theertan, "Machine learning and artificial intelligence based diabetes mellitus detection and self-management: A systematic review," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3204–3225, Jun. 2022, doi: 10.1016/j.jksuci.2020.06.013.

[9] R. K. Nadesh and K. Arivuselvan, "Type 2: Diabetes mellitus prediction using deep neural networks classifier," *Int. J. Cogn. Comput. Eng.*, vol. 1, pp. 55–61, Jun. 2020, doi: 10.1016/j.ijcce.2020.10.002.

[10] H. Lai, H. Huang, K. Keshavjee, A. Guergachi, and X. Gao, "Predictive models for diabetes mellitus using machine learning techniques," *BMC Endocrine Disorders*, vol. 19, no. 1, pp. 1–9, Dec. 2019, doi: 10.1186/s12902-019-0436-6.

[11] H. L. Tong, H. Ng, and H. Arul Ananthan, "Predicting diabetes mellitus with machine learning techniques," *J. Eng. Technol. Appl. Phys.*, vol. 6, no. 1, pp. 91–99, Mar. 2024.

[12] M. Bhattacharya and D. Datta, "Development of predictive models of diabetes using ensemble machine learning classifier," in *Proc. 1st Int. Conf. Advancements Smart Comput. Inf. Secur. (ASCIS)*, Rajkot, India, 2022, pp. 377–388.

[13] Y. Du, A. R. Rafferty, F. M. McAuliffe, L. Wei, and C. Mooney, "An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus," *Sci. Rep.*, vol. 12, no. 1, p. 1170, Jan. 2022, doi: 10.1038/s41598-022-05112-2.

[14] O. A. Ebrahim and G. Derbew, "Application of supervised machine learning algorithms for classification and prediction of type-2 diabetes disease status in afar regional state, Northeastern Ethiopia 2021," *Sci. Rep.*, vol. 13, no. 1, p. 7779, May 2023, doi: 10.1038/s41598-023-34906-1.

[15] A. Sarraju, A. Zammit, S. Ngo, C. Witting, T. Hernandez-Boussard, and F. Rodriguez, "Identifying reasons for statin nonuse in patients with diabetes using deep learning of electronic health records," *J. Amer. Heart Assoc.*, vol. 12, no. 7, Apr. 2023, Art. no. e028120.

[16] P. N. Thotad, G. R. Bharamagoudar, and B. S. Anami, "Diabetes disease detection and classification on Indian demographic and health survey data using machine learning methods," *Diabetes Metabolic Syndrome, Clin. Res. Rev.*, vol. 17, no. 1, Jan. 2023, Art. no. 102690, doi: 10.1016/j.dsx.2022.102690.

[17] J. J. Olickal, P. Chinnakali, B. S. Suryanarayana, S. Rajanarayanan, T. Vivekanandhan, G. K. Saya, K. Ganapathy, and D. K. S. Subrahmanyam, "Down referral and assessing comprehensive diabetes care in primary care settings: An operational research from India," *Diabetes Metabolic Syndrome, Clin. Res. Rev.*, vol. 17, no. 1, Jan. 2023, Art. no. 102694, doi: 10.1016/j.dsx.2022.102694.

[18] S. AlZu'bi, M. Elbes, A. Mughaid, N. Bdair, L. Abualigah, A. Forestiero, and R. A. Zitar, "Diabetes monitoring system in smart health cities based on big data intelligence," *Future Internet*, vol. 15, no. 2, p. 85, Feb. 2023, doi: 10.3390/fi15020085.

[19] H. Pan et al., "A risk prediction model for type 2 diabetes mellitus complicated with retinopathy based on machine learning and its application in health management," *Frontiers Med.*, vol. 10, 2023, Art. no. 1136653.

[20] M. J. Uddin, M. M. Ahamad, M. N. Hoque, M. A. A. Walid, S. Aktar, N. Alotaibi, S. A. Alyami, M. A. Kabir, and M. A. Moni, "A comparison of machine learning techniques for the detection of type-2 diabetes mellitus: Experiences from Bangladesh," *Information*, vol. 14, no. 7, p. 376, Jul. 2023, doi: 10.3390/info14070376.

[21] M. Zhao, J. Wan, W. Qin, X. Huang, G. Chen, and X. Zhao, "A machine learning-based diagnosis modelling of type 2 diabetes mellitus with environmental metal exposure," *Comput. Methods Programs Biomed.*, vol. 235, Jun. 2023, Art. no. 107537, doi: 10.1016/j.cmpb.2023.107537.

[22] M. Bernardini, L. Romeo, P. Misericordia, and E. Frontoni, "Discovering the type 2 diabetes in electronic health records using the sparse balanced support vector machine," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 1, pp. 235–246, Jan. 2020, doi: 10.1109/JBHI.2019.2899218.

[23] M. Eleftheriades, C. Chatzakis, E. Papachatzopoulou, V. Papadopoulos, I. Lambrinoudaki, K. Dinas, G. Chrousos, and A. Sotiriadis, "Prediction of insulin treatment in women with gestational diabetes mellitus," *Nutrition Diabetes*, vol. 11, no. 1, p. 30, Oct. 2021, doi: 10.1038/s41387-021-00173-0.

[24] A. Pyne and B. Chakraborty, "Artificial neural network based approach to diabetes prediction using pima Indians diabetes dataset," in *Proc. Int. Conf. Control, Autom. Diagnosis (ICCAD)*, May 2023, pp. 1–6, doi: 10.1109/ICCAD57653.2023.10152382.

[25] V. Chang, J. Bailey, Q. A. Xu, and Z. Sun, "Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms," *Neural Comput. Appl.*, vol. 35, no. 22, pp. 16157–16173, Aug. 2023, doi: 10.1007/s00521-022-07049-z.

[26] N. Gupta, B. Kaushik, M. K. I. Rahmani, and S. A. Lashari, "Performance evaluation of deep dense layer neural network for diabetes prediction," *Comput., Mater. Continua*, vol. 76, no. 1, pp. 347–366, 2023, doi: 10.32604/cmc.2023.038864.

[27] S. C. Gupta and N. Goel, "Predictive modeling and analytics for diabetes using hyperparameter tuned machine learning techniques," *Proc. Comput. Sci.*, vol. 218, pp. 1257–1269, Jan. 2023, doi: 10.1016/j.procs.2023.01.104.

[28] R. Alhalaseh, D. A. G. Al-Mashhadany, and M. Abbadi, "The effect of feature selection on diabetes prediction using machine learning," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jul. 2023, pp. 1–7, doi: 10.1109/iscc58397.2023.10218243.

[29] E. S. Nasser and F. A. A. Dawood, "Diagnosis and classification of type II diabetes based on multilayer neural network," *Iraqi J. Sci.*, vol. 63, no. 10, pp. 3744–3758, 2021.

[30] P. Nuankaew, S. Chaising, and P. Temdee, "Average weighted objective distance-based method for type 2 diabetes prediction," *IEEE Access*, vol. 9, pp. 137015–137028, 2021.

[31] S. Sivaranjani, S. Ananya, J. Aravinth, and R. Karthika, "Diabetes prediction using machine learning algorithms with feature selection and dimensionality reduction," in *Proc. 7th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, vol. 1, Mar. 2021, pp. 141–146, doi: 10.1109/ICACCS51430.2021.9441935.

[32] S. Naveen, S. Mishra, Bhavana, P. Pujitha, J. Chandana, and Priyanka, "Machine learning based diabetic prediction using random forest," in *Proc. 2nd Int. Conf. Vis. Towards Emerg. Trends Commun. Netw. Technol. (ViTECoN)*, May 2023, pp. 1–5, doi: 10.1109/ViTECoN58111.2023.10157881.

[33] N. Abdulhadi and A. Al-Mousa, "Diabetes detection using machine learning classification methods," in *Proc. Int. Conf. Inf. Technol. (ICIT)*, Jul. 2021, pp. 350–354, doi: 10.1109/ICIT52682.2021.9491788.

[34] V. Chang, M. A. Ganatra, K. Hall, L. Golightly, and Q. A. Xu, "An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators," *Healthcare Anal.*, vol. 2, Nov. 2022, Art. no. 100118, doi: 10.1016/j.health.2022.100118.

[35] 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," *Comput. Methods Programs Biomed.*, vol. 202, Apr. 2021, Art. no. 105968, doi: 10.1016/j.cmpb.2021.105968.

[36] J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *ICT Exp.*, vol. 7, no. 4, pp. 432–439, Dec. 2021, doi: 10.1016/j.icte.2021.02.004.

[37] K. Kannadasan, D. R. Edla, and V. Kuppili, "Type 2 diabetes data classification using stacked autoencoders in deep neural networks," *Clin. Epidemiol. Global Health*, vol. 7, no. 4, pp. 530–535, Dec. 2019, doi: 10.1016/j.cegh.2018.12.004.

[38] G. Swapna, R. Vinayakumar, and K. P. Soman, "Diabetes detection using deep learning algorithms," *ICT Exp.*, vol. 4, no. 4, pp. 243–246, Dec. 2018, doi: 10.1016/j.icte.2018.10.005.

[39] O. R. Shahin, H. H. Alshammari, A. A. Alzahrani, H. Alkhiri, and A. I. Taloba, "A robust deep neural network framework for the detection of diabetes," *Alexandria Eng. J.*, vol. 74, pp. 715–724, Jul. 2023, doi: 10.1016/j.aej.2023.05.072.

[40] O. Daanouni, B. Cherradi, and A. Tmiri, "Predicting diabetes diseases using mixed data and supervised machine learning algorithms," in *Proc. 4th Int. Conf. Smart City Appl.*, 2019, pp. 1–6, doi: 10.1145/3368756.3369072.

[41] N. Pradhan, G. Rani, V. S. Dhaka, and R. C. Poonia, "Diabetes prediction using artificial neural network," in *Deep Learning Techniques for Biomedical and Health Informatics*, B. Agarwal, V. E. Balas, L. C. Jain, R. C. Poonia, and H. I. Manisha, Eds., New York, NY, USA: Academic, 2020, ch. 14, pp. 327–339.

[42] M. S. A. Reshan, S. Amin, M. A. Zeb, A. Sulaiman, H. Alshahrani, A. T. Azar, and A. Shaikh, "Enhancing breast cancer detection and classification using advanced multi-model features and ensemble machine learning techniques," *Life*, vol. 13, no. 10, p. 2093, Oct. 2023.

[43] H. Zhou, Y. Xin, and S. Li, "A diabetes prediction model based on Boruta feature selection and ensemble learning," *BMC Bioinf.*, vol. 24, no. 1, p. 224, Jun. 2023, doi: 10.1186/s12859-023-05300-5.

[44] Kaggle, UCI Mach. Learn. (2023). *Pima Indians Diabetes Database*. Accessed: Feb. 1, 2023. [Online]. Available: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

[45] Kaggle. (2023). *Dataset of Diabetes, Taken From the Hospital Frankfurt, Germany*. Accessed: Feb. 1, 2023. [Online]. Available: https://www.kaggle.com/datasets/johndasilva/diabetes

[46] A. Rashid. (2020). *'Diabetes Dataset', Mendeley Data, V1*. Accessed: Feb. 1, 2023. [Online]. Available: https://data.mendeley.com/datasets/wj9rwkp9c2/1

[47] M. S. A. Reshan, S. Amin, M. A. Zeb, A. Sulaiman, H. Alshahrani, and A. Shaikh, "A robust heart disease prediction system using hybrid deep neural networks," *IEEE Access*, vol. 11, pp. 121574–121591, 2023, doi: 10.1109/ACCESS.2023.3328909.

[48] R. Rastogi and M. Bansal, "Diabetes prediction model using data mining techniques," *Meas., Sensors*, vol. 25, Feb. 2023, Art. no. 100605, doi: 10.1016/j.measen.2022.100605.

[49] M. M. Nishat, "Performance assessment of different machine learning algorithms in predicting diabetes mellitus," *Bioscience Biotechnol. Res. Commun.*, vol. 14, no. 1, pp. 74–82, Mar. 2021.

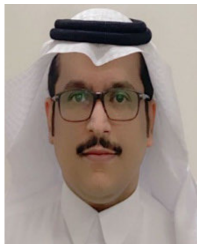
[50] K. M. J. Rani, "Diabetes prediction using machine learning," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 6, pp. 294–305, Jan. 2020.

[51] M. O. Edeh et al., "A classification algorithm-based hybrid diabetes prediction model," *Frontiers Public Health*, vol. 10, 2022, Art. no. 829519.

[52] C.-Y. Chou, D.-Y. Hsu, and C.-H. Chou, "Predicting the onset of diabetes with machine learning methods," *J. Personalized Med.*, vol. 13, no. 3, p. 406, Feb. 2023, doi: 10.3390/jpm13030406.

[53] D. Sisodia and D. S. Sisodia, "Prediction of diabetes using classification algorithms," *Proc. Comput. Sci.*, vol. 132, pp. 1578–1585, Jan. 2018, doi: 10.1016/j.procs.2018.05.122.

[54] K. Roy, M. Ahmad, K. Waqar, K. Priyaah, J. Nebhen, S. S. Alshamrani, M. A. Raza, and I. Ali, "An enhanced machine learning framework for type 2 diabetes classification using imbalanced data with missing values," *Complexity*, vol. 2021, pp. 1–21, Jul. 2021, doi: 10.1155/2021/9953314.



MANA SALEH AL RESHAN (Member, IEEE) received the B.S. degree in information systems from King Khalid University, Abha, Saudi Arabia, in 2007, the M.S. degree (Hons.) in computer, information, and network security from DePaul University, Chicago, USA, in 2011, and the Ph.D. degree in computer science from The Catholic University of America (CUA), Washington, DC, USA, in 2019. He was a Teaching Assistant with the College of Computer Science and Information Systems, Najran University, Saudi Arabia, from 2007 to 2009. He was a Lecturer with the College of Computer Science and Information Systems, Najran University, in 2012, where he is currently an Associate Professor with the Department of Information Systems. His current research interests include computer network and security, system security, wireless and mobile security, body area networks, health informatics, software systems and security, and cloud security.

SAMINA AMIN received the M.S. degree in computer science from the Institute of Computing, Kohat University of Science and Technology, Kohat, Pakistan, in 2020, where she is currently pursuing the Ph.D. degree. She has published several articles in reputed journals. Her research interests include machine learning, deep learning, data science, data mining, natural language processing, social media analysis, neural networks, recurrent neural networks, convolutional neural networks, image processing, health informatics, recommender systems, sentiment analysis, and reinforcement learning. She has received the Gold Medal from Kohat University of Science and Technology. She serves as a reviewer for different journals.

MUHAMMAD ALI ZEB is currently pursuing the Ph.D. degree with the Institute of Computing, Kohat University of Science and Technology, Kohat, Pakistan. He is also a Lecturer with the Institute of Computing, Kohat University of Science and Technology. He has published several papers in reputed journals and conferences. His research interests include machine learning, deep learning, social media analysis, natural language processing, sentiment analysis, data science, visualization, topic modeling, neural networks, image processing, and temporal graphs neural networks. He serves as a reviewer for different journals.

ADEL SULAIMAN received the master's degree in software engineering from the University of St. Thomas, USA, and the Ph.D. degree in computer science from Iowa State University, USA. He is currently an Associate Professor with the College of Computer Science and Information Systems, Najran University. He is also the Vice Dean of Quality and Development in the Deanship of E-Learning and leads the research unit with the College of Computer Science, Najran University. His research interests include human–computer interaction, large-scale simulations, mixed reality, scientific visualization, and data science.

HANI ALSHAHRANI (Senior Member, IEEE) received the bachelor's degree in computer science from King Khalid University, Abha, Saudi Arabia, the master's degree in computer science from California Lutheran University, Thousand Oaks, CA, USA, and the Ph.D. degree from Oakland University, Rochester, MI, USA. He is currently an Associate Professor of computer science and information systems with Najran University, Najran, Saudi Arabia. His current research interests include smartphones, the IoT, crowdsourcing security, and privacy.

ASADULLAH SHAIKH (Senior Member, IEEE) received the B.Sc. degree in software development from the University of Huddersfield, U.K., the M.Sc. degree in software engineering and management from Goteborg University, Sweden, and the Ph.D. degree in software engineering from the University of Southern Denmark. He was a Researcher with UOC, Barcelona, Spain. He is currently a Full Professor with the College of Computer Science and Information Systems, Najran University, Najran, Saudi Arabia. He has more than 200 publications in the area of software engineering in international journals and conferences. His publications have been cited more than 2000 times (H-index of 21 and i10 index of 62) according to Google Scholar. He has vast experience in teaching and research. His current research interests include software engineering, health informatics, artificial intelligence using machine learning, cloud computing, E-learning, and mobile technologies. He is an editor and an associate editor of several SCIE journals and an International Advisory Board of several conferences. Further details can be obtained using www.asadshaikh.com.

MOHAMED A. ELMAGZOUB was born in Riyadh, Saudi Arabia, in 1985. He received the B.S. degree (Hons.) in electrical and electronic engineering from Omdurman Islamic University, Sudan, in 2007, and the master's degree in electrical communication engineering and the Ph.D. degree in electrical engineering (communications) from Universiti Teknologi Malaysia (UTM), Malaysia, in 2012 and 2016, respectively. He was a Telecommunication Engineer with the Communication Department, Dams Implementation Unit, Sudan, from 2008 to 2009. He is currently an Associate Professor with the Department of Network and Communication Engineering, Najran University. He has authored ten ISI articles, two of them highly cited review articles in the area of passive optical networks and hybrid optical and wireless communication systems. His research interests include passive optical networks, hybrid optical and wireless communication systems, optical OFDM communications, and MIMO and next-generation access networks.