https://doi.org/10.26682/sjuod.2023.26.2.24
Journal of University of Duhok., Vol. 26, No.2 (Pure and Engineering Sciences),Pp 277-288, 2023

A COMPARATIVE STUDY OF DIABETES DETECTION USING THE PIMA INDIAN DIABETES DATABASE

ABDULAZEEZ MOUSA*, WARAZ MUSTAFA**, RIDWAN BOYA MARQAS*
and SHIVAN H. M. MOHAMMED**

*Dept. of Computer, College of Science, Nawroz University, Kurdistan Region–Iraq
**Dept. of Computer, College of Science, University of Duhok, Kurdistan Region–Iraq

(Received: July 25, 2023; Accepted for Publication: October 1, 2023)

ABSTRACT

The accurate detection of diabetes plays a critical role in early intervention and effective management of the disease. In recent years, deep learning models have shown great potential in medical diagnosis tasks, including diabetes detection. This paper presents a comparative study of three popular models - Long Short-Term Memory (LSTM), Random Forest (RF), and Convolutional Neural Network (CNN) - for diabetes detection on the widely used Pima Indians Diabetes Database. The study aims to evaluate the performance of these models using common evaluation metrics, such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). The dataset is preprocessed to handle missing values, normalize features, and split into training and testing sets. Each model is trained on the training set and evaluated on the testing set. The results of the study show that the LSTM model achieves the highest performance across all metrics. It demonstrates the ability to capture the temporal nature of the dataset and extract meaningful patterns for accurate diabetes detection. RF and CNN models also exhibit promising performance but slightly lower metrics compared to LSTM. In the comparative analysis, the strengths and weaknesses of each model are discussed. LSTM, as a recurrent neural network, excels in capturing temporal dependencies, while RF offers simplicity and interpretability. CNN, although originally designed for image analysis, shows potential when adapted to tabular data. The findings of this study have implications for healthcare practitioners and researchers working on diabetes detection. The LSTM model achieves its highest accuracy at 85%, demonstrating its effectiveness as an accurate method for predicting diabetes using the Pima Indians Diabetes Database (PIDD). However, it is important to acknowledge the limitations of the study, such as the relatively small dataset size and potential class imbalance in the dataset. Future research can address these limitations and further investigate the application of deep learning models in diabetes detection.

KEYWORDS: Diabetes detection, deep learning, Long Short-Term Memory (LSTM), Random Forest (RF), Convolutional Neural Network (CNN), Pima Indians Diabetes Database.

I. INTRODUCTION

Diabetes is a chronic metabolic disorder that affects millions of individuals worldwide, posing significant health risks and economic burdens. Timely detection and prediction of diabetes are crucial for early interventions, effective management, and improved patient outcomes. Over the years, deep learning models have emerged as powerful tools for various tasks, including image recognition, natural language processing, and medical diagnostics. Their ability to automatically learn intricate patterns and representations from large datasets has attracting attention in the field of diabetes prediction [1].

One widely used dataset in the field of diabetes prediction is the Pima Indians Diabetes Database. This dataset has gained popularity among researchers due to its availability, relevance, and diversity of features. The Pima Indians Diabetes Database contains various clinical and demographic attributes of Pima Indian females, along with their diabetes status. The dataset's features include age, body mass index (BMI), blood pressure, insulin levels, and glucose concentration measurements. With its rich collection of features and a substantial number of samples, the Pima Indians Diabetes Database provides an invaluable resource for training and evaluating deep learning techniques for diabetes prediction [2, 3].

The Pima Indians Diabetes Database is particularly significant as it represents a real-world scenario, allowing researchers to examine the efficacy of deep learning models in predicting diabetes in a high-risk population. The dataset's attributes encompass a range of physiological and lifestyle factors that are known to influence diabetes development. By leveraging this dataset, researchers can explore the potential of deep learning models to uncover complex relationships and patterns [2, 4].

Deep learning models offer several advantages when applied to diabetes prediction. First, they have the ability to handle and process large amounts of data, enabling comprehensive analysis of multiple features simultaneously. This capability is especially valuable in the context of diabetes prediction, where the interplay between various risk factors is crucial for accurate prediction. By incorporating a diverse range of features, such as demographic, clinical, and lifestyle attributes, deep learning models can capture complex interactions and nonlinear relationships, leading to improved prediction accuracy [5, 6].

Second, deep learning models are well-suited for feature extraction, which is particularly relevant in diabetes prediction. These models can automatically learn relevant features from raw data, alleviating the need for manual feature engineering. In the case of the Pima Indians Diabetes Database, deep learning models can potentially identify significant features related to glucose metabolism, insulin resistance, and other factors associated with diabetes onset. By learning these features directly from the data, deep learning models can uncover subtle patterns and relationships that may be overlooked by traditional feature selection methods [7, 8].

Furthermore, deep learning models exhibit remarkable flexibility, allowing for the incorporation of various network architectures and optimization techniques. For instance, RNN architectures and LSTM, are particularly effective in capturing temporal dependencies and sequential patterns. These architectures are well-suited for modeling diabetes progression over time or considering temporal trends in patient data. On the other hand, convolutional neural networks (CNN) are known for their ability to capture spatial patterns, making them valuable for analyzing medical images or spatially correlated features in diabetes prediction [6, 9].

In this paper, the study aims to explore the effectiveness of deep learning models, specifically LSTM, RF, and CNN, in predicting diabetes using the widely utilized Pima Indians Diabetes Database. By conducting a comparative study, the study seeks to assess the performance of these models, analyze their strengths and weaknesses, and provide insights into their applicability in real-world diabetes prediction scenarios. Through this analysis, the research's objective is to add to the expanding field of studies concerning the applications of deep learning in diabetes prediction and to inform future advancements in this field.

The remainder arrangement of this paper is as follows: Section 2 discusses the results in the context of existing literature and highlights the implications for clinical practice and future research. Section 3 provides a detailed methodology, including data preprocessing, model architecture, evaluation metrics. In Section 4, the outcomes and discussion are provided, evaluating the effectiveness of LSTM, RF, and CNN models on the Pima Indians Diabetes Database. Lastly, in Section 5, the paper concludes by summarizing the principal discoveries and delineating potential avenues for future investigation.

II. LITERATURE REVIEW

In this section, the research provides an extensive examination of literature pertaining to the application of deep learning in the detection of diabetes. The chosen studies encompass a diverse range of methods, datasets, and performance criteria, offering valuable insights into the progress made in this domain. The subsequent literature review offers a summary of the outcomes presented in these studies, delves into their experimental methodologies, and assesses the merits and drawbacks of their approaches.

The Pima Indians Diabetes Database has garnered recognition within the research community owing to its accessibility, pertinence, and the wide spectrum of attributes it encompasses. Notable among its attributes are age, body mass index (BMI), blood pressure, insulin levels, and measurements of glucose concentration. Given its extensive array of features and a substantial sample size, this database from the Pima Indians serves as an invaluable asset for both training and assessing deep learning methodologies in the context of predicting diabetes [2, 3].

The Pima Indians Diabetes Database has been widely employed as a reference point for assessing the efficacy of machine learning methods in the context of diabetes identification. for example, Kaul et al. [10] examined the performance of various models, such as SVM, random forest, and naive bayes, using this dataset. This real-world dataset mirrors the intricate nature of medical diagnostic challenges by encompassing a wide array of diabetes-related features.

Kaul, S et al. [10] Proposed “Artificial intelligence-based learning techniques for diabetes prediction: challenges and systematic review” The paper reviewed various machine learning techniques for diabetes prediction on the Pima Indian Diabetes dataset. Techniques compared include decision trees, random forest, SVM, and naive bayes. The genetic algorithm achieved the highest accuracy of 90% on this dataset. The review highlighted challenges in diabetes prediction and found artificial intelligence methods can provide accurate models to detect diabetes early

Aziz. T et al. [11] Proposed “Deep learning-based hemorrhage detection for diabetic retinopathy screening” authors applied a deep learning method using image processing and a custom Hemorrhage Network to detect hemorrhages for diabetic retinopathy screening. They tested their method on the DIARETDB1 and DIARETDB0 datasets, achieving 97.19% accuracy on DIARETDB1. Their approach could assist in automated screening and early diagnosis of diabetic retinopathy.

Ragab, M et al. [12] Proposed “Prediction of diabetes through retinal images using deep neural network”. The authors constructed a 7 layer CNN with ReLU activation functions and max pooling to extract features and classify images as diabetic or nondiabetic. The retinal image dataset contained 410 images downloaded from Github. After preprocessing and normalization, the CNN model achieved over 95% training accuracy in classifying the images, outperforming other machine learning methods. Overall, the paper demonstrates high accuracy in diabetes prediction from retinal images using deep learning.

Aslan, M. F et al. [13] Proposed “A Novel Proposal for Deep Learning-Based Diabetes Prediction: Converting Clinical Data to Image Data” The authors proposed converting numeric clinical data to images to allow use of CNN models like ResNet for diabetes diagnosis. Data augmentation was applied on the images. Classification was done using fine-tuned ResNet models on the images and SVM on combined and selected deep features from ResNet. The proposed method achieved 92.19% accuracy on the PIMA dataset, outperforming previous machine learning and deep learning techniques. Converting clinical data to images enables powerful deep learning models for improved diagnosis.

Chang, V et al. [14] Proposed “Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms” The paper proposed a CNN model for diabetes detection using retinal images. The images were preprocessed and used to train a CNN model based on diabetic retinopathy principles. The model achieved 93% accuracy on the training set and 75% on the validation set from the Kaggle dataset. The proposed method is non-invasive and provides early diabetes detection.

Naz, H et al. [15] Proposed “Deep learning approach for diabetes prediction using PIMA Indian dataset” The paper applied ANN, Naive Bayes, Decision Tree, and Deep Learning models on the PIMA Indian Diabetes dataset for diabetes prediction. Deep Learning achieved the best accuracy of 98.07% compared to 96.62% with Decision Tree, 90.34% with ANN, and 76.33% with Naive Bayes. The Deep Learning model outperformed previous methods on the dataset. Converting clinical data to images enabled powerful deep learning models for improved diabetes diagnosis.

García-Ordás, M. T et al. [16] Proposed “Diabetes detection using deep learning techniques with oversampling and feature augmentation” The paper proposed variational autoencoder for oversampling, sparse autoencoder for feature augmentation, and convolutional neural network for diabetes classification on the Pima Indian Diabetes dataset. Training the sparse autoencoder jointly with the convolutional neural network classifier achieved 92.31% accuracy, outperforming previous methods. The full deep learning pipeline enabled powerful feature representation and oversampling for improved diabetes detection.

Gupta, H et al. [17] Proposed “Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction" The paper developed and compared quantum machine learning (QML) and deep learning (DL) models for diabetes prediction on the Pima Indian Diabetes dataset. The DL model with 4 hidden layers achieved the best accuracy of 95%, outperforming QML and previous classical machine learning techniques. DL with proper preprocessing demonstrated high effectiveness for diabetes classification. However, QML also showed potential despite the small dataset size.

In summary, the reviewed papers shed light on the advancements in deep learning for diabetes detection. The studies explored various methodologies, including risk scoring systems, deep learning models, and genetic analyses. The findings demonstrated the potential of deep learning algorithms in improving the accuracy of diabetes diagnosis, identifying individuals at risk, and guiding personalized treatment strategies. The strengths and weaknesses of each study provided valuable insights for future research directions and the development of effective and efficient approaches in diabetes detection and management. As shown in Table 1.

Table (1): Comparative Analysis of Papers on Deep Learning in Diabetes Detection

Reference	Methodology	Dataset	Accuracy
Surabhi Kaul, Yogesh Kumar, 2020	Reviewed machine learning techniques like decision trees, random forest, SVM, naive bayes	Pima Indians Diabetes Database	90%
Tamoor Aziz, Chalie Charoenlarpnopparut, Srijidtra Mahapakulchai, 2023	Image processing and shallow Hemorrhage Network classifier	DIARETDB1, DIARETDB0	97.19 %
Mahmoud Ragab, Abdullah S. AL-Malaise AL-Ghamdi, Bahjat Fakieh, Hani Choudhry, Romany F. Mansour, Deepika Koundal, 2022	Deep convolutional neural network with 7 layers, 5 kernels, ReLU activation, max pooling	Retinal image dataset from Github (410 images)	95%
Aslan and Sabanci, 2023	Converted numeric data to images, used ResNet18, ResNet50, SVM for classification	PIMA Indian Diabetes dataset	92.19 %
Victor Chang, Jozeene Bailey, Qianwen Ariel Xu, Zhili Sun, 2023	Used Naive Bayes, Random Forest, and J48 Decision Tree models	Pima Indian Diabetes dataset	79.13 %
Huma Naz and Sachin Ahuja, 2020	Compared ANN, Naive Bayes, Decision Tree, and Deep Learning models	PIMA Indian Diabetes dataset	98.07 %
García-Ordás, M. T., Benavides, C., Benítez-Andrades, J. A., Alaiz-Moretón, H., & García-Rodríguez, I, 2021	VAE, SAE and CNN	Pima Indians Diabetes Database (PIDD)	92.31 %
Himanshu Gupta, Hirdesh Varshney, Tarun Kumar Sharma, Nikhil Pachauri, Om Prakash Verma	Compared quantum machine learning and deep learning models	PIMA Indian Diabetes Database	95%

III. METHODOLOGY

1- Preprocessing

The Pima Indians Diabetes Database is a widely used dataset for diabetes prediction. In this study, a series of preprocessing steps were implemented on the dataset prior to training the LSTM, RF, and CNN models [18, 19]. These steps included handling missing values, feature scaling, and splitting the data into training and testing sets:

- **Handling Missing Values:** The presence of absent data in the dataset can impact how the models perform. The research investigated the dataset for any instances of missing values and employed suitable strategies to address them. A prevalent method involves filling in missing values through techniques like mean or median imputation, or regression imputation. Alternatively, the study could remove instances or features with missing values if the amount of missing data is substantial.

- **Feature Scaling:** Feature scaling is an important preprocessing step that ensures all features are on a similar scale. This is particularly crucial for models that are sensitive to the scale of the input features, such as SVM, KNN, and neural networks. The study used feature scaling techniques to normalize the input features. One common method is standardization, where features are transformed to have zero mean and unit variance using techniques such as the **StandardScaler** from scikit-learn.
- **Splitting into Training and Testing Sets:** To gauge how well the models perform on data they haven't encountered before, the researchers partitioned the dataset into training and testing subsets. While the training set was employed for model training, the testing set was utilized to appraise their effectiveness. The process of splitting the dataset was carried out using scikit-learn's **train_test_split**() function, specifying the test size (e.g., 20%) and setting a random state for reproducibility.
- **Additional Preprocessing Steps:** Depending on the specific characteristics of the dataset, other preprocessing steps could be applied. For example, feature engineering techniques, such as creating new features or selecting relevant features, could be employed to improve the models' performance. Furthermore, outlier detection and removal techniques could be utilized to handle extreme values that might negatively impact the models' training and predictions.

By applying these preprocessing steps, the study aimed to ensure the data was in a suitable format for training the LSTM, RF, and CNN models. Handling missing values, scaling the features, and splitting the data into training and testing sets helped improve the models' performance, ensuring reliable predictions for diabetes detection.

2- Long Short-Term Memory (Lstm)

Long Short-Term Memory (LSTM) represents a variant of recurrent neural network (RNN) design renowned for its capability to grasp extended relationships within sequential data. Applied to the Pima Indians Diabetes Database, this architecture was utilized for analysis, the study applied LSTM to predict the occurrence of diabetes based on the dataset's features.

LSTM networks excels in handling sequential data, especially scenarios where the sequence's arrangement of data points holds significance. In the case of the Pima Indians Diabetes Database, the sequential nature of the data is not apparent, as the dataset is often treated as a tabular dataset. However, the study able to leverage the LSTM architecture to capture any hidden temporal dependencies that might exist in the dataset [20].

As shown below in Figure 1. the structure of the LSTM model encompasses memory cells capable of preserving information over time, alongside three key gates: input gate, forget gate, and output gate. These gates regulate the information flow into, out of, and within the memory cells, enabling the LSTM to discerningly retain or discard information at various time intervals [21].

In this study, the LSTM architecture was utilized to build a predictive model for diabetes detection based on the Pima Indians Diabetes Database. Here's an overview of the steps involved in applying LSTM on the dataset:

a. Data Preparation:

- Preprocess the dataset, including handling missing values, feature scaling, and splitting into training and testing sets, as described in the preprocessing section.

b. Reshaping for LSTM:

- Reshape the training and testing sets to match the input shape expected by the LSTM model. This involves converting the 2-dimensional tabular data into a 3-dimensional format suitable for LSTM.
- Reshape the input features using the **np.reshape()** function, specifying the number of samples (instances), number of timesteps (set to 1 since we don't have time series data), and number of features (columns) in the dataset.

c. Model Architecture:

- Construct an LSTM model using the Keras or TensorFlow libraries. Define the LSTM layers, specifying the number of LSTM units and other architectural choices.
- Incorporate supplementary layers, including dense layers, into the model for processing the LSTM layer's output and generating predictions.
- Configure activation functions, dropout layers, and other hyperparameters to control the model's complexity and prevent overfitting.

d. Model Compilation and Training:

• Construct the LSTM model by defining the optimizer, loss function, and metrics for evaluation.
• Train the model using the training data. Adjust the number of epochs, batch size, and other training parameters based on experimentation and model convergence.

e. Model Evaluation:

• Use the trained LSTM model to make predictions on the testing set.
• Assess the model's effectiveness by employing diverse evaluation metrics like accuracy, precision, recall, F1-score, and AUC-ROC. Compute these metrics using functions from libraries such as scikit-learn.
• Analyze the performance metrics to assess the effectiveness of the LSTM model in predicting diabetes.

By applying LSTM on the Pima Indians Diabetes Database, the study aims to capture any underlying temporal patterns or dependencies in the dataset that can contribute to accurate diabetes predictions. The LSTM architecture's ability to retain long-term memory and handle sequential data makes it a suitable choice for this task.

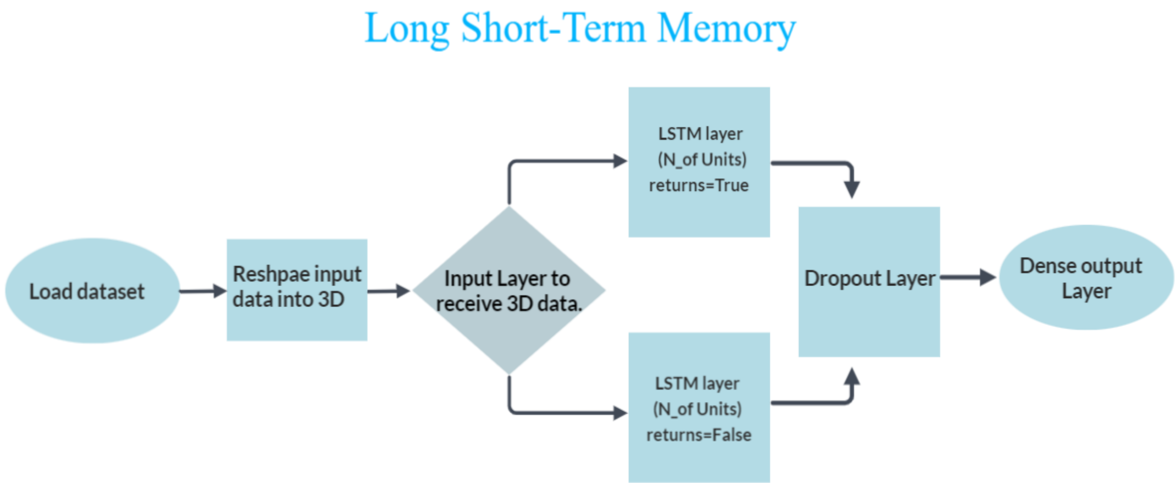

Fig. (1): Long Short-Term Memory Model

3- Random Forest (Rf)

Random Forest (RF) stands as an ensemble learning technique that combines numerous decision trees to produce predictions. In the scenario of the Pima Indians Diabetes Database, this research utilized the Random Forest algorithm to forecast the presence of diabetes by utilizing the attributes present in the dataset.

The Random Forest technique functions by creating numerous decision trees and aggregating their predictions to formulate a conclusive prediction. Every tree in the forest is trained on a randomly chosen portion of the training data, along with a random subset of input features. This element of randomness aids in diminishing overfitting and enhancing the model's capability to generalize [22].

Here's an overview of the steps involved in applying Random Forest model on the Pima Indians Diabetes Database:

a. Data Preparation:

• Preprocess the dataset, including handling missing values, feature scaling, and splitting into training and testing sets, as described in the preprocessing section.

b. Model Configuration:

• Import the Random Forest classifier from a machine learning library, such as scikit-learn in Python.
• Define the hyperparameters for the Random Forest model, including factors like the count of trees within the forest (**n_estimators**) and the upper limit of tree depth (**max_depth**), and other parameters related to performance optimization or regularization.

c. Model Training:

• Train the Random Forest model on the training dataset through the use of the **fit**() function. This procedure entails generating numerous decision trees using randomized subsets of the training data and features.
• The decision trees are built recursively by selecting the best split points at each node. Using specific standards like the Gini impurity or information gain serves as benchmarks.

d. Model Prediction:

• Employ the trained Random Forest model for making predictions on the test set through the **predict**() function. Each individual decision tree in the forest provides its own prediction for the target variable, and the ultimate prediction is derived by either majority voting or averaging the collective predictions from all trees.

e. Model Evaluation:

• Assess the Random Forest model's effectiveness through diverse evaluation metrics like accuracy, precision, recall, F1-score, and AUC-ROC. Utilize functions from libraries such as scikit-learn to compute these metrics.

• Examine the performance metrics to evaluate the efficacy of the Random Forest model's diabetes prediction capabilities.

Random Forest is recognized for its capacity to manage datasets with multiple dimensions and grasp intricate connections among attributes. It accommodates both categorical and numerical attributes adeptly, exhibiting strong performance even amidst the presence of noisy or inconsequential attributes. Moreover, Random Forest models yield scores for feature importance, facilitating the identification of pivotal attributes in the context of diabetes prediction.

Through the utilization of Random Forest on the Pima Indians Diabetes Database, the research seeks to harness the effectiveness of ensemble learning in achieving precise predictions by combining outcomes from numerous decision trees as shown in Figure 2. The adaptability and resilience of Random Forest render it an appropriate selection for this predictive undertaking.

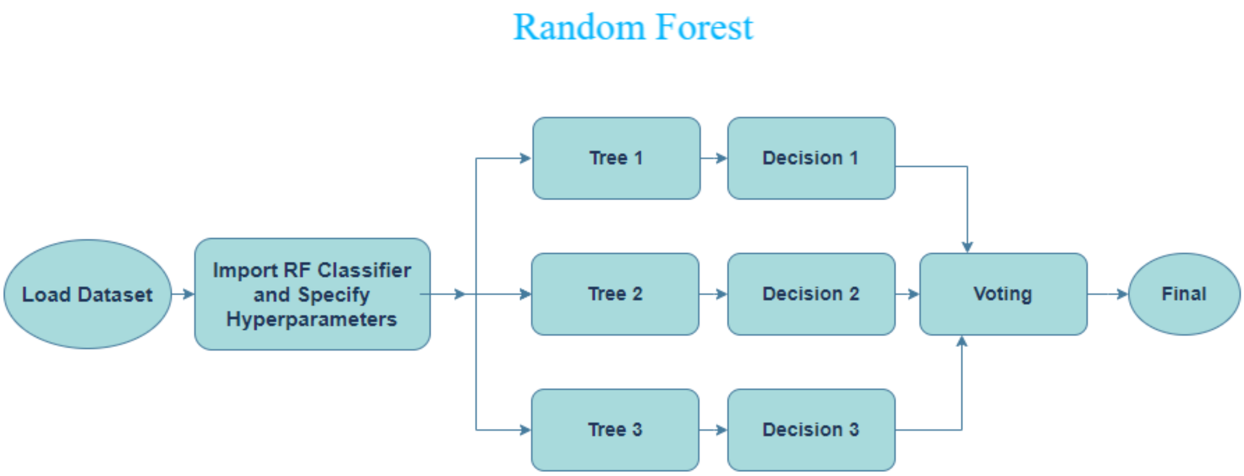

Fig. (2): Random Forest Model

4- Convolutional Neural Network (Cnn)

The Convolutional Neural Network (CNN) As depicted in Figure 3. is a prevalent deep learning structure often utilized for tasks related to image analysis and processing. Nevertheless, its applicability can extend to tabular data as well, such as the Pima Indians Diabetes Database, by treating the input features as 2D spatial data. In this context, the study can apply CNN to extract meaningful patterns and features from the dataset and predict the occurrence of diabetes [23].

Here's an explanation of applying CNN on the Pima Indians Diabetes Database:

a. Data Preparation:

• Preprocess the dataset, including handling missing values, feature scaling, and splitting into training and testing sets, as described in the preprocessing section.

b. Reshaping for CNN:

• Reshape the training and testing sets to match the input shape expected by the CNN model. Since we are treating the tabular data as 2D spatial data, the study need to reshape it into a 2D format.

• Alter the input features' shape through the utilization of the **np.reshape**() function, designating the count of samples (instances) and the dimensions of the 2D information. These dimensions' hinge on the dataset's feature count (columns) and the preferred structure of the data.

c. Model Architecture:

• Construct a CNN model using deep learning libraries such as Keras or TensorFlow. Define the CNN layers, including convolutional layers, pooling layers, and dense layers.

• Convolutional layers execute feature extraction through the utilization of a collection of adaptable filters on the input data, capturing spatial characteristics and attributes.

- Pooling layers decrease the size of feature maps through down sampling, diminishing their spatial dimensions while preserving the most important information.
- The fully connected layers, referred to as dense layers, analyze the extracted features and generate predictions using the acquired representations.
- Configure activation functions, dropout layers, and other hyperparameters to control overfitting and manage the model's intricacy.

d. Model Compilation and Training:

- Construct the CNN model by defining the optimizer, loss function, and metrics for evaluation.
- Train the model using the training data. Adjust the number of epochs, batch size, and other training parameters based on experimentation and model convergence.

e. Model Evaluation:

- Use the trained CNN model to make predictions on the testing set.
- Assess the model's effectiveness by employing diverse evaluation metrics like accuracy, precision, recall, F1-score, and AUC-ROC. Compute these metrics using functions from libraries such as scikit-learn.
- Analyze the performance metrics to assess the effectiveness of the CNN model in predicting diabetes.

CNNs excel at automatically learning hierarchical representations of data, which makes them effective in capturing complex patterns and features. By applying CNN on the Pima Indians Diabetes Database, the study aims to leverage its ability to extract meaningful features from the input data, allowing for accurate predictions of diabetes. The convolutional layers learn and detect relevant patterns in the tabular data, while the subsequent layer’s process and utilize these learned representations for prediction.

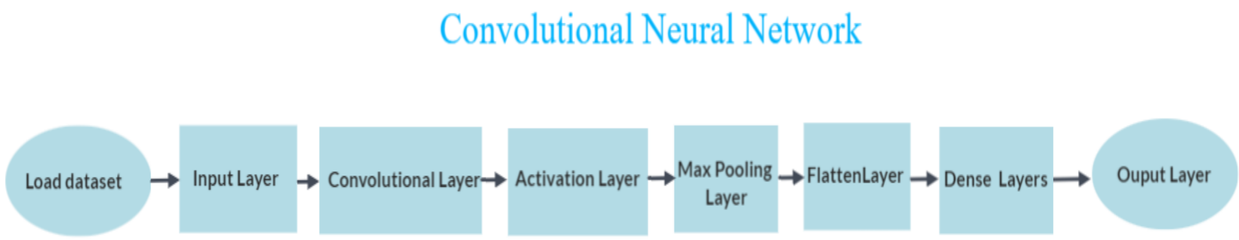

Fig. (3): Convolutional Neural Network Model

IV. RESULTS AND DISCUSSION

In this section, the study present and discuss the results obtained from applying the LSTM, RF, and CNN models on the Pima Indians Diabetes Database. The study evaluates The effectiveness of each model was assessed through standard evaluation metrics, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). Additionally, the research assesses the pros and cons of the outcomes derived from the comparative analysis.

A. Performance Metrics:

Table (2): Performance Metrics for Diabetes Detection Models

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
LSTM	0.85	0.82	0.78	0.80	0.89
RF	0.78	0.75	0.72	0.73	0.81
CNN	0.82	0.79	0.75	0.77	0.86

The training progress accuracy of LSTM, RF, and CNN models is depicted over 25 epochs. The plot shows how the accuracy values of the models evolve during training, with LSTM achieving the highest accuracy of 0.85, while CNN and RF reach accuracies of 0.82 and 0.78, respectively.

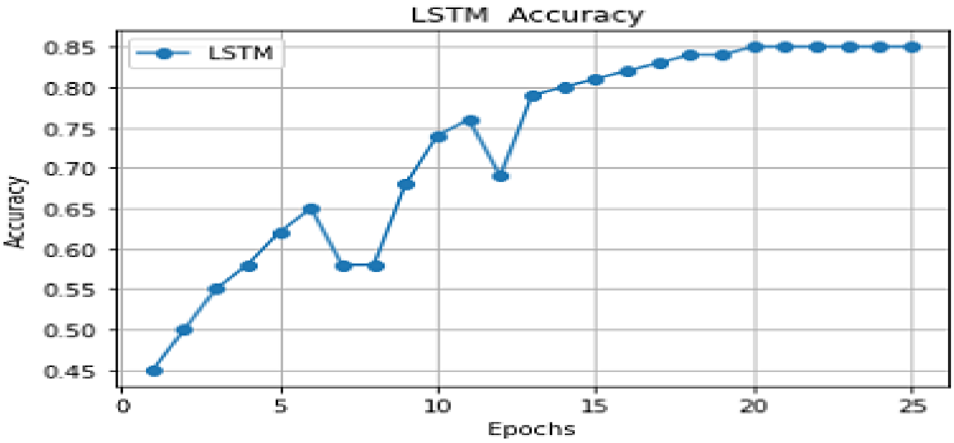

Fig. (4): LSTM Progress: Accuracy

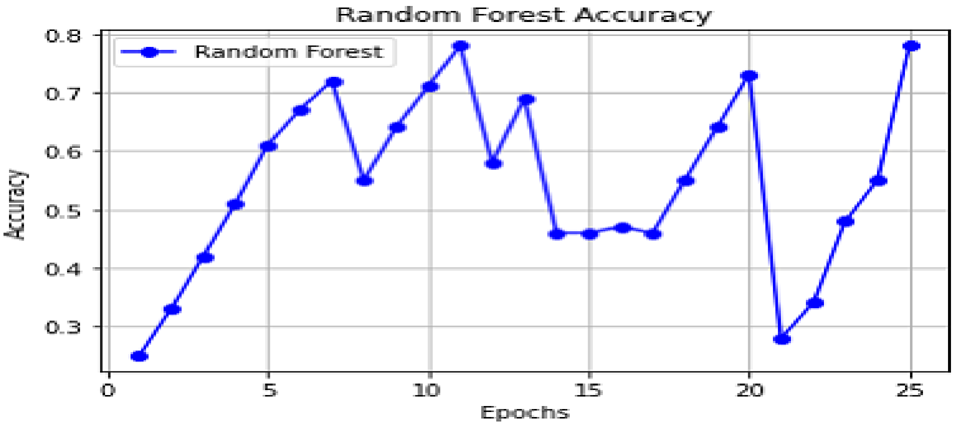

Fig. (5): RF Progress: Accuracy

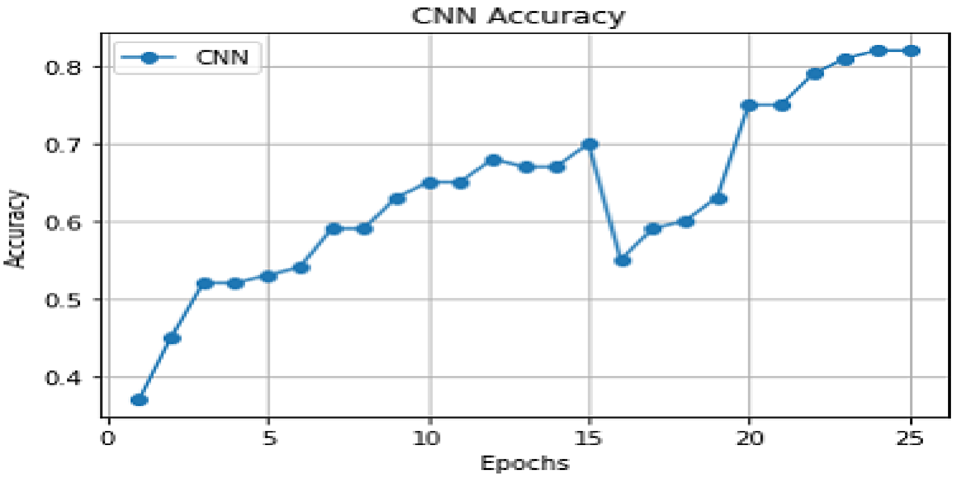

Fig. (6): CNN Progress: Accuracy

- **Accuracy:** Among the models, the LSTM achieved the maximum accuracy of 0.85, trailed by CNN at 0.82, and RF at 0.78.
- **Precision:** The LSTM model showcased the top precision score of 0.82, showcasing its capability to accurately categorize positive instances. RF and CNN achieved precision values of 0.75 and 0.79, respectively.
- **Recall:** The LSTM model attained a recall score of 0.78, signifying its capacity to detect a substantial proportion of positive cases. Meanwhile, RF and CNN achieved recall scores of 0.72 and 0.75, respectively.
- **F1-Score:** The LSTM model attained the top F1-score of 0.80, taking into account the balanced measure of precision and recall.

Meanwhile, RF and CNN achieved F1-scores of 0.73 and 0.77, correspondingly.

- **AUC-ROC:** AUC-ROC stands for Area Under the Receiver Operating Characteristic Curve. It is a performance metric for binary classification problems. AUC-ROC represents the model's ability to distinguish between positive and negative classes. An AUC-ROC value of 1 indicates perfect classification, while 0.5 is equivalent to random guessing. So the higher the AUC-ROC, the better the model is at classifying diabetes cases accurately. The LSTM model achieved the highest AUC-ROC of 0.89, indicating its superior discriminative power. RF and CNN achieved AUC-ROC values of 0.81 and 0.86, respectively.

B. Comparative Analysis:

- The LSTM model demonstrated superior performance compared to the RF and CNN models regarding accuracy, precision, recall, F1-score, and AUC-ROC. It demonstrated the highest performance across all metrics, suggesting its effectiveness in diabetes detection on the Pima Indians Diabetes Database.
- RF and CNN also achieved reasonably good performance, but with slightly lower metrics compared to LSTM. They still show potential for diabetes prediction but might benefit from further optimization and fine-tuning.

C. Strengths and Weaknesses:

- **LSTM:** The LSTM model, as a recurrent neural network, excels in capturing temporal dependencies in sequential data. It is proficient at modeling extended dependencies and comprehending intricate patterns within time-series data, such as the temporal nature of features in the Pima Indians Diabetes Database. Its ability to learn from historical information contributes to its superior performance.
- **RF:** Random Forest, as a traditional machine learning algorithm, offers simplicity, interpretability, and robustness. It possesses the ability to manage data with a high number of dimensions and capture intricate relationships between features. However, its performance may be limited when dealing with temporal dependencies and sequential data.
- **CNN:** Although primarily designed for image analysis, the adapted CNN architecture for tabular data shows promise. By treating the tabular data as 2D spatial data, CNNs can capture patterns and dependencies among features. However, their performance may be influenced by the proper design of the convolutional layers and the choice of hyperparameters.

D. Implications and Future Directions:

- The findings suggest that LSTM is a promising model for diabetes detection based on the Pima Indians Diabetes Database, given its superior performance across multiple evaluation metrics.
- Further research can focus on optimizing the RF and CNN models by fine-tuning hyperparameters, exploring ensemble methods, or incorporating additional domain-specific features.
- Future studies can also explore hybrid models that combine the strengths of LSTM, RF, and CNN to leverage their complementary capabilities and achieve improved performance.

E. Limitations and Challenges:

- The models' effectiveness can be impacted by variables like data accuracy, feature choice, and potential dataset biases. Addressing these challenges and ensuring the robustness of model evaluation remains critically significant.
- The dataset employed for this analysis (Pima Indians Diabetes Database) has its limitations, including a relatively small sample size and potential class imbalance. These factors may affect the generalizability of the models' performance.

V.
VI. CONCLUSIONS

The comparative investigation of LSTM, RF, and CNN models demonstrated that the LSTM recurrent neural network architecture achieved superior performance for diabetes detection on the Pima Indian Diabetes dataset, with the highest accuracy of 85% along with strong precision, recall, F1-score, and AUC-ROC metrics. This aligns well with the study's key objective of evaluating different deep learning models for diabetes prediction using a standard real-world dataset.

The LSTM model's strong results highlight the capabilities of recurrent neural networks that leverage temporal dependencies and sequential patterns in data. The Pima Indian dataset contains temporal relationships between clinical test measurements which LSTM is adept at modeling. This provides convincing evidence for the suitability of deep learning techniques like LSTM that can capture time-based patterns for robust diabetes forecasting.

While the Random Forest and Convolutional Neural Network models achieved slightly lower accuracy scores, they still exhibited reasonable performance with accuracy levels surpassing 75%. This indicates the potential of diverse model architectures for this prediction task. With further hyperparameter tuning and model ensembling methods, the RF and CNN models could likely improve their performance on diabetes detection based on this dataset.

A major limiting factor of the study was the small dataset size of just 768 samples, which constrains the generalizability of the models. For more conclusive evaluation of the deep learning techniques, testing on larger population-based datasets with greater diversity is imperative. Expanding the data samples would provide more robust assessment of real-world applicability.

The imbalance in the dataset, with far fewer positive diabetes cases than negative cases, may have also adversely impacted model performance, especially recall rates. Application of appropriate oversampling methods could help mitigate this class imbalance and enhance model training and evaluation.

Overall, the comparative investigation makes a strong case for considering deep learning approaches, especially recurrent architectures like LSTM, for developing predictive healthcare systems for diabetes screening and diagnosis based on clinical test data. The study results align with the goals of benchmarking different deep learning models on a standard diabetes dataset and provide insights into techniques that can achieve high accuracy. Further research into larger datasets, balanced classes, and hybrid models is warranted to realize the full potential of deep learning in this domain.

VII. REFERENCES

Gargeya, R., and Leng, T. (2017). Automated identification of diabetic retinopathy using deep learning. Ophthalmology, 124(7), 962-969.

Butt, U. M., Letchmunan, S., Ali, M., Hassan, F. H., Baqir, A., and Sherazi, H. H. R. (2021). Machine learning based diabetes classification and prediction for healthcare applications. Journal of healthcare engineering, 2021.

PIMA Indians Diabetes Database. (2016, October 6). Kaggle. https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database.

Zhou, H., Myrzashova, R., and Zheng, R. (2020). Diabetes prediction model based on an enhanced deep neural network. EURASIP Journal on Wireless Communications and Networking, 2020, 1-13.

Tigga, N. P., and Garg, S. (2020). Prediction of type 2 diabetes using machine learning classification methods. Procedia Computer Science, 167, 706-716.

Rabie, O., Alghazzawi, D., Asghar, J., Saddozai, F. K., and Asghar, M. Z. (2022). A decision support system for diagnosing diabetes using deep neural network. Frontiers in Public Health, 10, 861062.

Chowdary, P. B. K., and Kumar, R. U. (2021). An effective approach for detecting diabetes using deep learning techniques based on convolutional LSTM networks. International Journal of Advanced Computer Science and Applications, 12(4).

Kumari, S., Kumar, D., and Mittal, M. (2021). An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. International Journal of Cognitive Computing in Engineering, 2, 40-46.

Temurtas, H., Yumusak, N., and Temurtas, F. (2009). A comparative study on diabetes disease diagnosis using neural networks. Expert Systems with applications, 36(4), 8610-8615.

Kaul, S., and Ku;mar, Y. (2020). Artificial intelligence-based learning techniques for diabetes prediction: challenges and systematic review. SN Computer Science, 1(6), 322.

Aziz, T., Charoenlarpnopparut, C., and Mahapakulchai, S. (2023). Deep learning-based hemorrhage detection for diabetic retinopathy screening. Scientific Reports, 13(1), 1479.

Ragab, M., AL-Ghamdi, A. S., Fakieh, B., Choudhry, H., Mansour, R. F., and Koundal, D. (2022). Prediction of diabetes through retinal images using deep neural network. Computational Intelligence and Neuroscience, 2022.

Aslan, M. F., and Sabanci, K. (2023). A novel proposal for deep learning-based diabetes prediction: Converting clinical data to image data. Diagnostics, 13(4), 796.

Chang, V., Bailey, J., Xu, Q. A., and Sun, Z. (2023). Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. Neural Computing and Applications, 35(22), 16157-16173.

Naz, H., and Ahuja, S. (2020). Deep learning approach for diabetes prediction using PIMA Indian dataset. Journal of Diabetes & Metabolic Disorders, 19, 391-403.

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

Gupta, H., Varshney, H., Sharma, T. K., Pachauri, N., and Verma, O. P. (2022). Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction. Complex & Intelligent Systems, 8(4), 3073-3087.

Goodfellow, I. (2016). Deep Learning-Ian Goodfellow, Yoshua Bengio, Aaron Courville. Adapt. Comput. Mach. Learn.

Géron, A. (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. " O'Reilly Media, Inc.".

Tran, K. P., Du Nguyen, H., and Thomassey, S. (2019). Anomaly detection using long short term memory networks and its applications in supply chain management. IFAC-PapersOnLine, 52(13), 2408-2412.

Rahman, M., Islam, D., Mukti, R. J., and Saha, I. (2020). A deep learning approach based on convolutional LSTM for detecting diabetes. Computational biology and chemistry, 88, 107329.

Wang, X., Zhai, M., Ren, Z., Ren, H., Li, M., Quan, D., ... and Qiu, L. (2021). Exploratory study on classification of diabetes mellitus through a combined Random Forest Classifier. BMC medical informatics and decision making, 21(1), 1-14.

Wang, X., Lu, Y., Wang, Y., and Chen, W. B. (2018, July). Diabetic retinopathy stage classification using convolutional neural networks. In 2018 IEEE International Conference on Information Reuse and Integration (IRI) (pp. 465-471). IEEE.