AI-Driven Neural Networks for Early-Stage Diabetes Prediction

Sheetalrani R Kawale[1], Pooja Kallappagol[2]

[1,2]Dept. of Computer Science, Karnataka State Akkamahadevi Women University, Vijayapura-586105, India
Email: sheetalrani@kswu.ac.in[1], poojakallappagol30@gmail.com[2]

Received: 10.07.2024 Revised: 13.08.2024 Accepted: 23.09.2024

ABSTRACT
Glucose intolerance is a common catabolism ailment that can lead to serious consequences such as cardiovascular disease, renal failure, and blindness. Nearly 77 million people in India have type 2 diabetes, and another 25 million are at risk of getting it. India has the second-highest rate of diabetes in the world. A lot of people still don't know about the health risks they face, which shows how important early detection is to lower death rates and improve patient health. The proposed methods deal with how well CNN, LSTM, and SimpleRNN models can forecast the early phase of diabetes. For this research, we collected live (primary source) data and preprocessed, we are making it as a standard dataset and we will publish it, tentatively named "Southern India Diabetes Dataset (SIDD)". Our live dataset comprises 806 patient samples, of which 532 have diabetes patient samples and 274 are nondiabetes patient samples. We primarily considered demographic and clinical factors, such as age, gender, and diabetes symptoms. Various neural network models have trained the dataset. The models have achieved accuracy of 96% for the CNN, 95% for the LSTM, and 99.99% for the SimpleRNN. We evaluated the algorithms' performance based on the F1-score, recall, and precision. This indicates that modern deep learning models are proficient at distinguishing between individuals with diabetes and those who are not diabetic.

Keywords: Deep learning, Diabetes prediction, CNN, LSTM, SimpleRNN, Early-Stage diabetes

1. INTRODUCTION

Diabetes is a chronic form that stays frequently regarded as intractable and is primarily defined by the inability to regulate blood sugar. In this study, we collected live data and generated a standard dataset tentatively named "Southern India Diabetes Dataset (SIDD)" that includes 806 patient diabetic factors, focussing on demographic and clinical factors such as age, gender, and diabetes symptoms. It may result in lasting harm to different areas of the body, including the heart, blood vessels, eyes, and nerves among other regions of the body. The disease is caused by either inadequate insulin synthesis or cellular resistance to insulin [1]. This leads to inadequate glucose absorption by the body's cells, resulting in symptoms such as frequent urination, unexplained weight loss, excessive thirst, increased hunger, delayed wound healing, and dizziness. Neglecting diabetes may result in serious problems like cardiovascular disease, retinopathy, and renal failure [8]. The incidence of diabetes is rising rapidly, particularly in developing countries [4, 5, 6], where it significantly contributes to ailments such as renal disease, stroke, myocardial infarctions, blindness, and limb amputations. A considerable proportion of patients with diabetes stay undiagnosed, hence increasing the strain on healthcare systems when untreated problems arise [2, 11]. Diabetes is a substantial global public health concern, incurring considerable emotional, social, and economic burdens [9]. In India [3], the economic impact is especially alarming since diabetes is anticipated to rise among both younger and older demographics, exerting more pressure on the nation's healthcare system [13]. The World Health Organization specifies that beyond 422 million people widespread agonize from diabetes, resulting in 1.6 million deaths annually. Recent research indicates that the prevalence of diabetes is increasing at a more rapid pace than in prior years. India now holds the second position worldwide for the prevalence of diabetes, with more than 77 million persons suffering from type 2 diabetes, with another 25 million afflicted by prediabetes [15–17]. Numerous people remain oblivious to the health dangers they encounter, underscoring the need for early identification to mitigate mortality and enhance patient outcomes [18]. glucose intolerance, a catabolism disorder, impairs the physical potential to effectively translate subsistence into energy. Meals provide glucose, an essential energy source, while the pancreas produces insulin, a hormone that facilitates cellular glucose uptake. The rising occurrence of diabetes in countries like India underscores the need for public health measures focused on early detection, risk factor mitigation, and effective disease management to prevent complications [12]. Studies indicate that diabetes may be undiagnosed for a duration of 4 to 12 years prior to identification [22]. At the time of diagnosis, about fifty percent of people already exhibit diabetes-related problems [23]. Research demonstrates that early identification may mitigate serious consequences, including heart disease, blindness, vascular problems, stroke, renal failure, and amputations [24]. Therefore, timely diagnosis remains vital to improving the worth of lifespan for individuals through diabetes [11, 19, 20, 21]. Since their inception, deep learning algorithms have proven a formidable instrument in healthcare, especially for the early prediction of illnesses such as diabetes [25]. Some methods, like CNNs, LSTM, and SimpleRNN, can look at a lot of medical data and find patterns on their own, which can lead to more accurate predictions than regular ML methods [11, 26, 27]. Refat et al. [28] examined the early forecast of diabetes by assessing several ML and DL methodologies. They used a diabetic dataset from the UCI repository to evaluate nine different classification systems. Their findings demonstrated that all techniques were successful, with the XGBoost classifier attaining greater accuracy relative to the other algorithms and approaching perfect results in early-phase diabetes prediction. Laila, U. E., and colleagues erected a prognostic model for the early diabetes risk assessment via AdaBoost, bootstrap aggregation (Bagging), and random forests [29]. The field of healthcare makes extensive use of deep learning methods, particularly for predicting and evaluating the risk of diabetes. As a result, a number of researchers have developed unique methods for early diagnosis of this illness using a variety of ML and DL technologies. This literature review examines several research papers that employ ML and DL algorithms for diabetes prediction. They used a UCI dataset in their study to aid in the early prediction of diabetes. The random forest model demonstrated exceptional performance. Khafaga, D.S., et al. [30] created a model using the Apriority technique to analyse 12 features and generate association rules based on the lift matrix. The K Nearest Neighbour method in their study reached an accuracy of 97.36%, sensitivity of 99.21%, specificity of 95.94%, and precision of 98.22%. Atif, M. et al. [31] created a prediction algorithm for early diabetes using the UCI dataset of 520 samples. They used many methodologies, including k-NN, Decision Tree, SVM, Stochastic Gradient Descent, Random Forest, ANN, Naive Bayes, Logistic Regression, Gradient Boosting, and AdaBoost. Gradient boosting performed exceptionally well, achieving an accuracy of 97.2%, a ROC curve score of 98.8%, an F-measure of 97.2%, and precision and recall of 97.2%. Yurttakal, A. H., et al. [32] examined early diabetes prediction employing a deep neural network methodology referred to as stacked ensemble based on the UCI dataset. Their model attained an exceptional ROC curve score of 99.19%, sensitivity at 100%, and specificity at 98.39%. The integration of two deep neural networks (DNNs) had a success rate of 99.36%, surpassing configurations with three, four, or five DNNs and resulting in an F-measure of 99.47%. The proposed study recommends the following research directions: Section 2 is about impression of notable previous study towards early diabetes prediction via ML and DL methodologies. Section 3 outlines the methodology, including data collection, dataset characteristics, data processing, and the deep learning algorithms used in the study. Section 4 evaluates the algorithms accuracy, while the remainder of 5 provides the results and discussion.

2. Proposed Method

The research will include essential phases such as live data collection, model training, model testing, and performance assessment. We divide the dataset into two predictive categories: "yes" (denoted by 1) and "no" (denoted by 0). Divided the dataset into an eighty percentage for training set and twenty percentage for testing set. We will use CNNs, LSTM, and SimpleRNN to forecast early-phase diabetes. We will train the models using the training set, assess their efficacy with the testing set, and further validate them using novel data. Fig. 2.1 delineates the procedure of the proposed method.

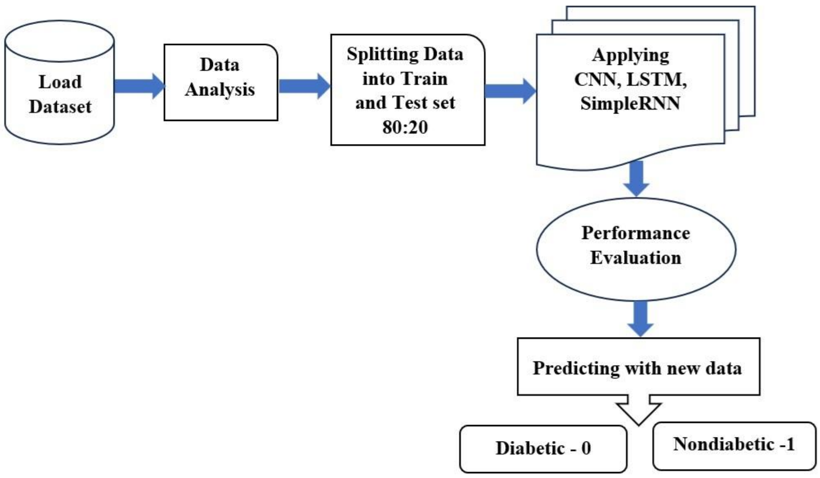

Fig 2.1: Flowchart of the Suggested Framework

2.1 Primary Data Collection

We have gathered real-time data from the Koppal Institute of Medical Sciences in Koppal for this study, which we have tentatively named the "Southern India Diabetes Dataset (SIDD)". This dataset includes 806 patient diabetic factors, primarily focusing on demographic and clinical factors such as age, gender, and symptoms associated with diabetes. We developed this dataset by leveraging the UCI early-stage diabetes risk prediction dataset and integrating its attributes. The dataset contains 17 characteristics, which are as follows: age, gender, polyuria, sudden weight loss, weakness, polyphagia, genital thrush, visual blurring, itching, irritability, delayed healing, partial paresis, muscle stiffness, alopecia, obesity, polydipsia, and class. Table 2.1 provides a comprehensive list of these characteristics.

Table 2.1. Characteristics of the Dataset

Sl.no	Features	Description
1	Age	21-80 years
2	Gender	Male and Female
3	Polyuria	Excessive Urination
4	sudden weight loss	Yes, No
5	weakness	Yes, No
6	Polyphagia	Excessive hunger [Yes, No]
7	Genital thrush	yeast infections in the genital area [Yes, No]
8	visual blurring	Yes, No
9	Itching	Yes, No
10	Irritability	Yes, No
11	delayed healing	Yes, No
12	partial paresis	Yes, No
13	muscle stiffness	Yes, No
14	Alopecia	Yes, No
15	Obesity	Yes, No
16	Polydipsia	Excessive Thirst [Yes, No]
17	class	[Diabetes -1 and Nondiabetes - 0]

The dataset used in this study has many significant attributes for diabetes prediction. The characteristics are age, ranging from 21 to 80 years, and gender, denoted as male (1) and female (0). Polyuria (excessive urination), sudden weight loss, weakness, polyphagia (excessive hunger), genital thrush (yeast infection), visual blurring, itching, irritability, delayed healing, partial paresis (partial muscle paralysis), muscle stiffness, alopecia (hair loss), obesity, and polydipsia (excessive thirst) are all categorised as Yes (1) or No (0). The goal variable, class, signifies the presence (1) or absence (0) of diabetes in the person. The purpose of this dataset is to aid in the creation and evaluation of diabetes prediction models, which encompass a range of clinical symptoms and demographic data.

2.2 Convolutional Neural Network

The CNN is utilized to perform binary classification of numerical data. We divided the dataset into features and target variables, resulting in the formation of training and testing sets. We executed the standardisation procedure for feature scaling and restructured the data to meet the requirements of the Conv1D layers. There was a Conv1D layer with 32 filters in the CNN design, then MaxPooling1D, Flatten, and Dense layers, and finally an output layer with a sigmoid activation function. The model endured training for one hundred epochs, used a batch size of sixteen, and employed the Adam optimiser with binary cross-entropy loss. The evaluation of the system's performance used a confusion matrix, a classification report, and visual representations of training and validation accuracy and loss. The Fig. 2.2 delineates the fundamental architecture of a Convolutional Neural Network (CNN).

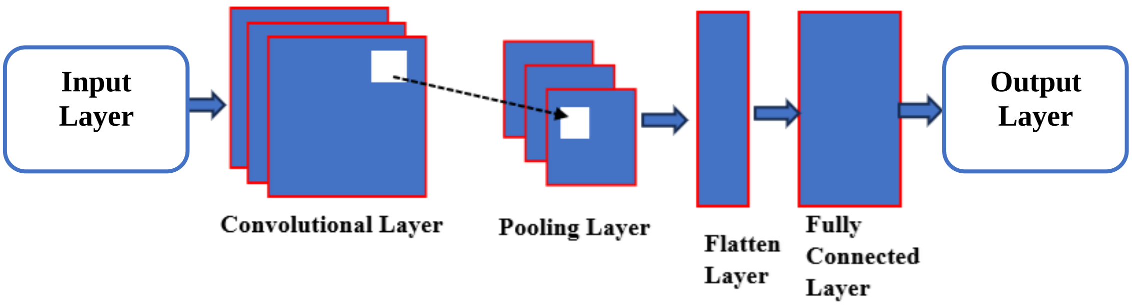

Fig 2.2: The fundamental architecture of a Convolutional Neural Network (CNN)

2.3 Long Short-Term Memory

Long Short-Term Memory networks, a kind of recurrent neural network, use memory cells and gating mechanisms to capture long-range dependencies in sequential input. Networks achieve this. This design enables LSTMs to excel in applications like sequence prediction and time-series forecasting. To complete this binary classification task, we had to change the shape of the LSTM's three-dimensional input features, make an LSTM model with fifty units and ReLU activation, and then add a dense layer with sigmoid activation. We constructed the model using the Adam optimiser with binary cross-entropy loss and trained it for one hundred epochs. The evaluation used a confusion matrix, a classification report, visualisations of accuracy and loss, and a heatmap of the confusion matrix. The Fig. 2.3 delineates the fundamental architecture of a Long Short-Term Memory networks.

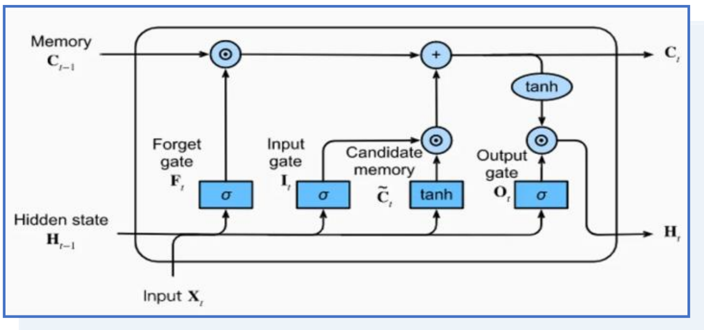

Fig 2.3: The fundamental architecture of a Long Short-Term Memory networks

2.4 Simple Recurrent Neural Network

A fundamental architecture for deep learning is the Simple Recurrent Neural Network, designed to manage sequential input by integrating temporal correlations. In contrast to feedforward networks, SimpleRNN is characterised by its internal memory, which evolves concurrently with the processing of incoming inputs. This facilitates the model's ability to identify patterns inside sequences more efficiently. This design includes a SimpleRNN layer with fifty units, triggered by ReLU, to provide a probability score for binary classification. The sigmoid function triggers a dense layer with a single unit, succeeding this layer. The dense layer generates the final classification outcome, whereas the RNN layer processes feature sequences. We construct the model using the Adam optimiser and binary cross-entropy loss, with accuracy serving as the performance measure. We trained the model for one hundred epochs with a batch size of sixteen, and then conducted validation on a distinct test set. The Fig. 4 delineates the fundamental architecture of a Simple Recurrent Neural Network.

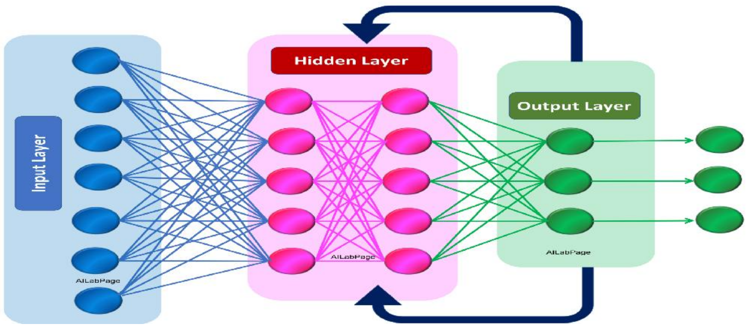

Fig 2.4: The fundamental architecture of a Simple Recurrent Neural Network

3. Experimental Results and Discussion

This research used three deep learning algorithms: CNN, LSTM, and SimpleRNN. Real-world situations provided the dataset for this study. Data visualisation methods included in the dataset analysis comprise bar charts, line graphs, scatter plots, and heat maps. Figures 3.1, 3.2, 3.3, and 3.4 display the visualisation approaches. Table 3.1 illustrates the separation of the dataset into two separate sets: the training set and the testing set. We conducted an exhaustive taxation of the effectiveness of the deep learning algorithms, presenting the findings in Tables 3.2 and 3.3, along with Figures 3.5, 3.6, and 3.7. Table 3.8 illustrates that the models underwent further assessment using fresh data to forecast diabetes or non-diabetes outcomes. Table 3.9 delineates a comparison between the findings derived from the proposed model and those from prior studies.

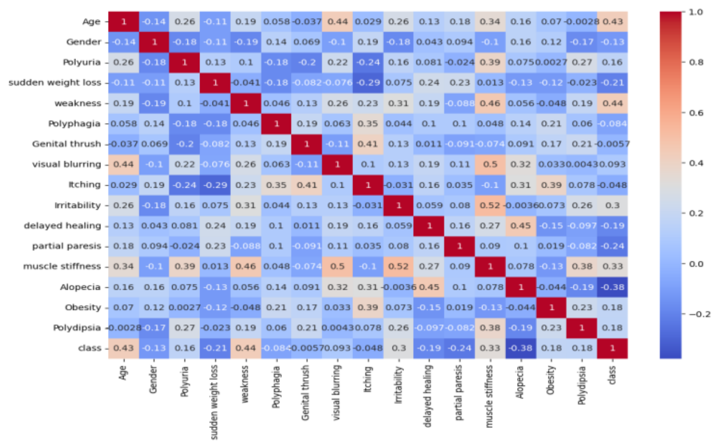

Fig 3.1: Correlation (Confusion) matrix

Illustration in the Fig. 3.1: Age (0.43) demonstrates a strong positive connection with the class. Weakness (0.44) and muscular stiffness (0.33) exhibit moderate associations with the class. Symptoms such as visual blurring, pruritus, irritation, and protracted healing have a modest relationship to the classification. Gender, baldness, obesity, and polydipsia have minimal positive associations, while polyuria, abrupt weight loss, and partial paresis reveal negligible negative relationships with the class.

Data Visualization

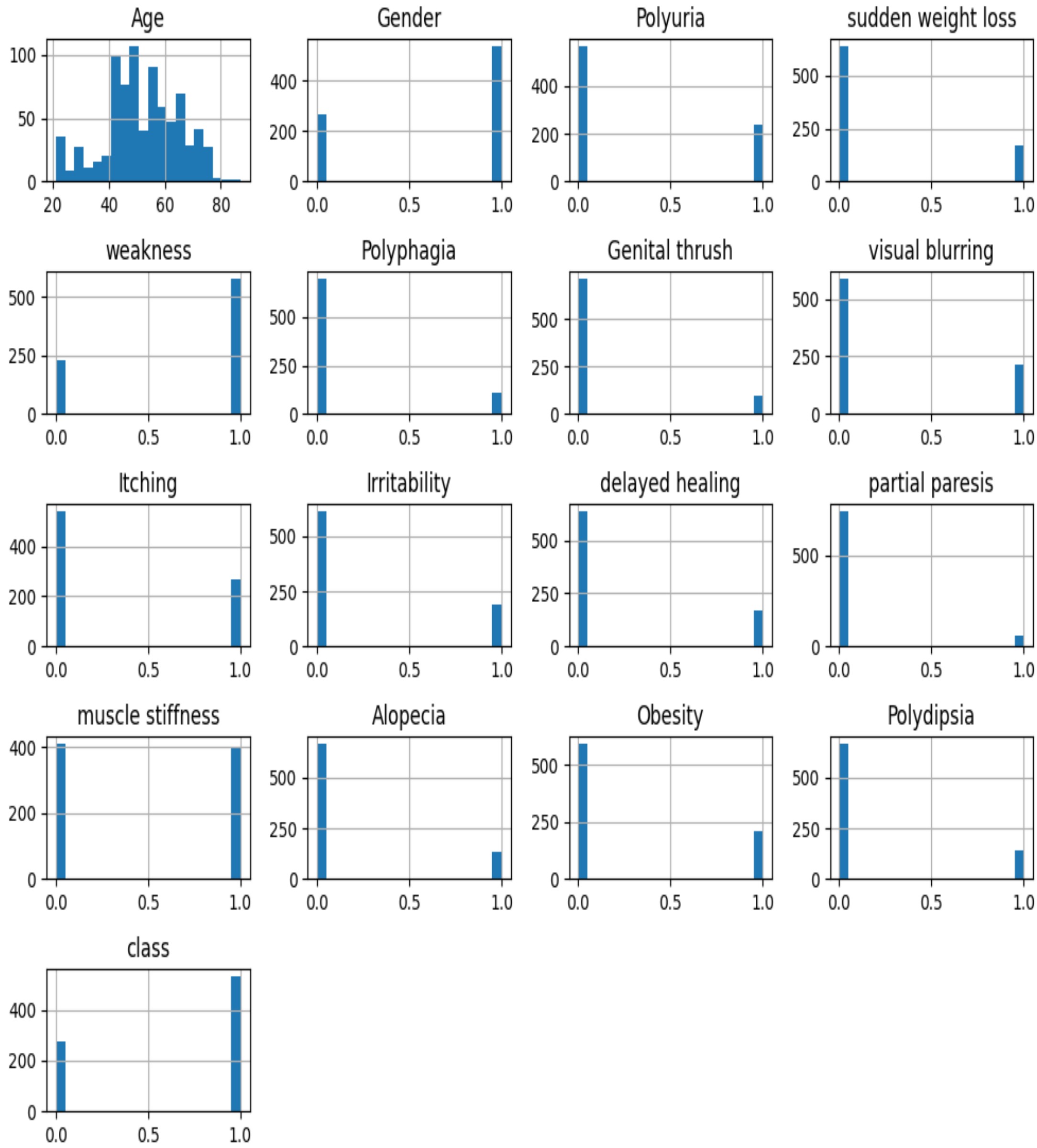

Fig 3.2: The Number of counts according to each feature in the dataset

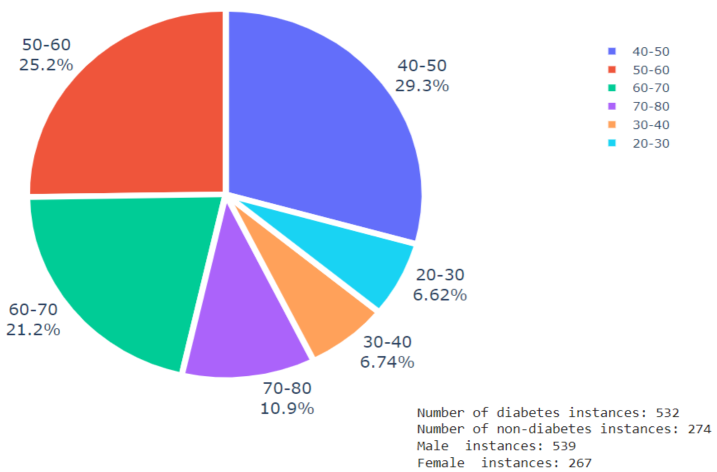

Fig 3.3: Diabetes distribution according to the ages

The pie chart shows the distribution of patients across many age categories, therefore stressing the diabetes incidence in the data of the whole patient group, the 40–50 age range is 29.3%; the 50–60 age range is 25.2%. The 50-60 age demographic is the second biggest group of patients. Individuals aged 60 to 70 comprise 21.2% of the population, while those aged 70 to 80 comprise 10.9%. A decreased proportion of persons falls into the younger age of 30-40 years and 20-30 years, representing 6.74% and 6.62%, respectively. The medical data indicate 532 instances of diabetes, far exceeding the 274 instances of non-diabetes. The gender distribution indicates that the sample comprises 539 male patients and 267 female patients. Acquiring demographic and health data is crucial for understanding age- and gender-specific trends in diabetes incidence, which may affect the deployment of focused healthcare interventions.

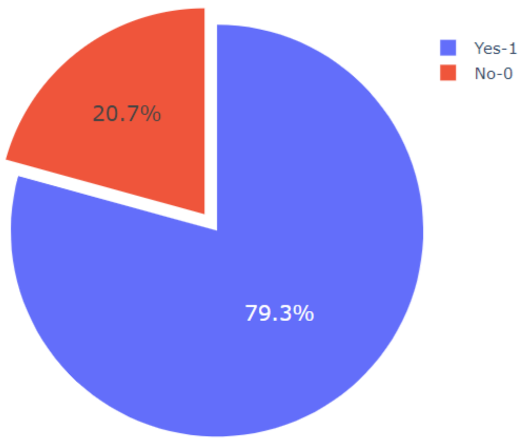

Fig 3.4: Number of Diabetes and nondiabetic in dataset

Illustration in Fig. 3.4: The Distribution of Diabetes and Nondiabetics Instances in the Dataset illustrates the proportions of diabetes (Yes-1) and nondiabetic (No-0) instances, with 79.3% classified as diabetic and 20.7% as nondiabetic.

Table 3.1: Allocation of Data for Diabetes Classification

Class	**Training Set**	**Validation Set**	**Test Set**
Diabetes	417	55	60
Nondiabetes	227	26	21
Total	**644**	**81**	**81**

Table 3.1 illustrates the configuration of validation, test and training data used in a diabetic classification problem. From the total of 806 samples, we split into 3 sets: a training set including 644 samples and a test and validation set consisting of 162 samples. We extracted the educational goals from the training dataset. There are 532 diabetic samples, with 417 designated for training and 115 for testing and validation. The non-diabetes group consists of 274 samples, with 47 allocated to the test and validation set and 227 to the training set.

Table 3.2: Performance of the CNN, LSTM and SimpleRnn classifiers

Algorithms	**Accuracy**	**Loss**	**Validation Accuracy**	**Validation Loss**	**Time**
CNN	93.35%	0.1633	97.53%	0.0841	6ms/step
LSTM	92.46%	0.1883	96.30%	0.0967	24ms/step
SimpleRNN	98.80%	0.0462	99.99 %	0.0051	11ms/step

We used three different deep learning classifiers CNN, LSTM, and SimpleRNN to determine which one was the best in diabetes classification. As shown in Table 3.2, all three models successfully aligned with the training data, achieving a training and validation accuracy of 93.35%, 92.46% and 98.80% and validation accuracy of 97.53%, 96.30% and 100%. While training, the CNN model had a loss of 0.1633, the LSTM model a loss of 0.1883, and the SimpleRNN model a loss of 0.0462. The validation losses are like 0.0841, 0.0967 and 0.0051. The losses are like of 6ms/step, 24ms/step and 11ms/step.

Table 3.3: Algorithms Classification Report Performance

Algorithms	Class	Precision	Recall	F1-score	Support
CNN	0	88%	100%	93%	21
	1	100%	95%	97%	60
LSTM	0	95%	86%	90%	21
	1	95%	98%	97%	60
SimpleRNN	0	100%	100%	100%	21
	1	100%	100%	100%	60

Table 3.3 demonstrate the classification report for each model, along with the F1-score, accuracy, and recall for the two categories of 0 (non-diabetes) and 1 (diabetic). All algorithms demonstrated perfect accuracy, recall, and F1-scores, suggesting that they were all very proficient in classifying patients into those with and without diabetes. Although the support for the nondiabetic class was much smaller (21 instances) than that for the diabetic class (60 samples), the classifiers managed this imbalance without any decline in performance.

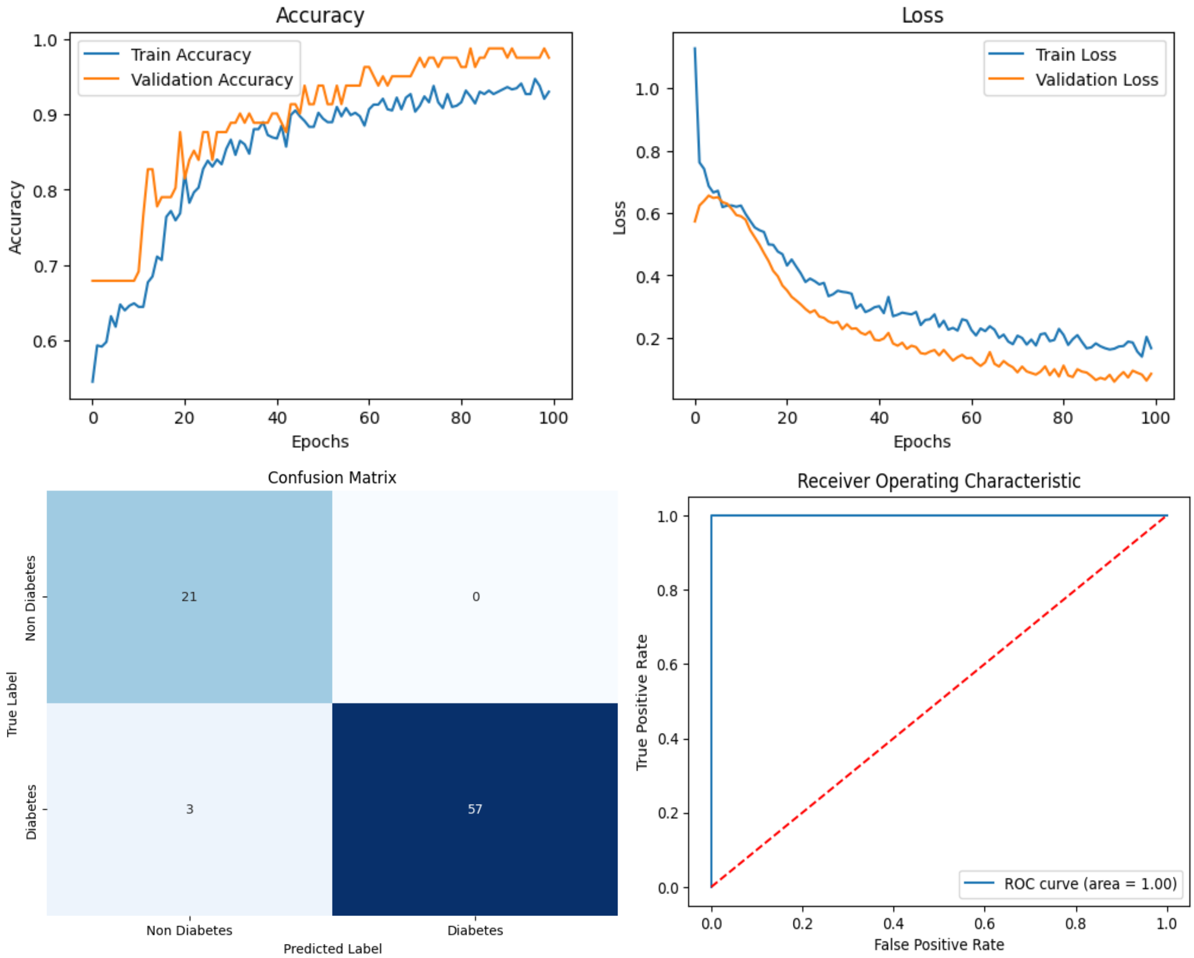

Fig 3.5: CNN Performance Metrics – Accuracy, Loss, and Confusion Matrix and Roc Curve

Fig. 3.5 shows the model's gain and loss in accuracy over 100 epochs of training and evaluation. The validation accuracy is consistently improving, which is very similar to the training accuracy. The confusion matrix shows that the model correctly identified most cases of diabetes and non-diabetes. The area under the curve (AUC) of the ROC curve is 100%, which means that the classification works very well and there are no fake positives or negatives.

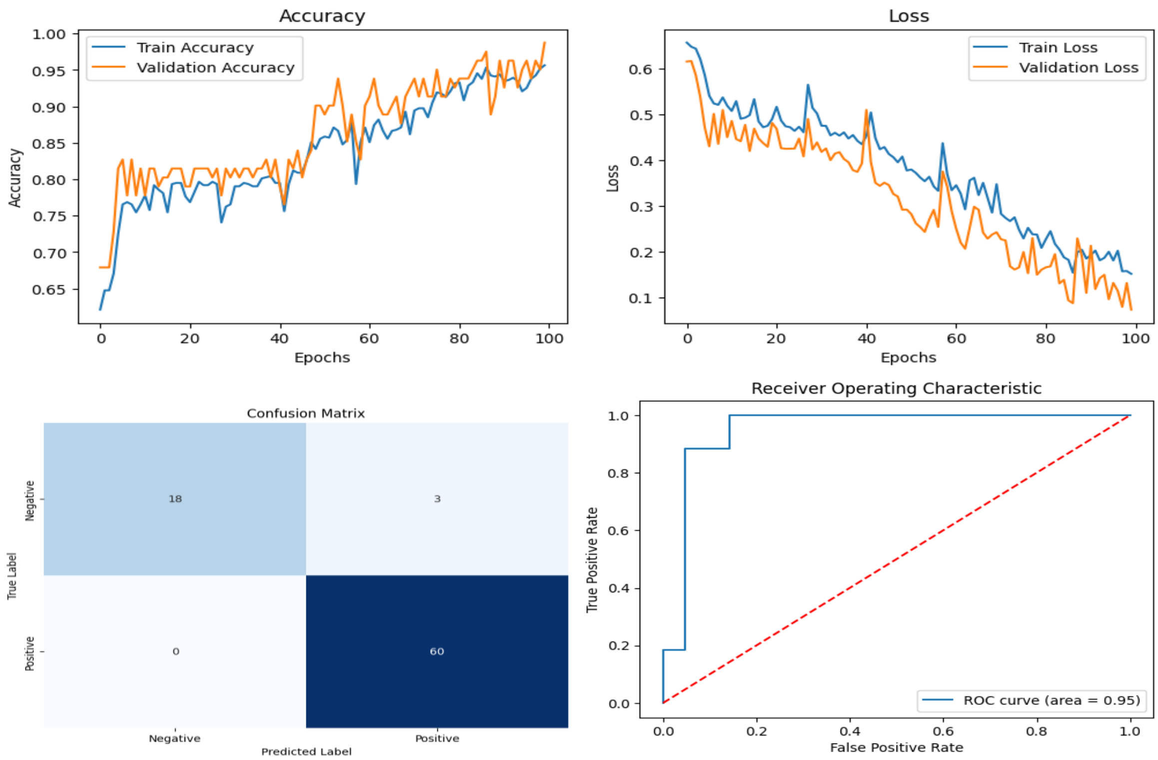

Fig 3.6: LSTM Performance Metrics – Accuracy, Loss, and Confusion Matrix and Roc Curve

Fig 3.6 shows the model's gain and loss in accuracy over 100 epochs of training and evaluation. The validation accuracy is consistently improving, which is very similar to the training accuracy. The confusion matrix shows that the model correctly identified most cases of diabetes and non-diabetes. The area under the curve (AUC) of the ROC curve is 95%, which means that the classification works very well and there are 3 false negatives.

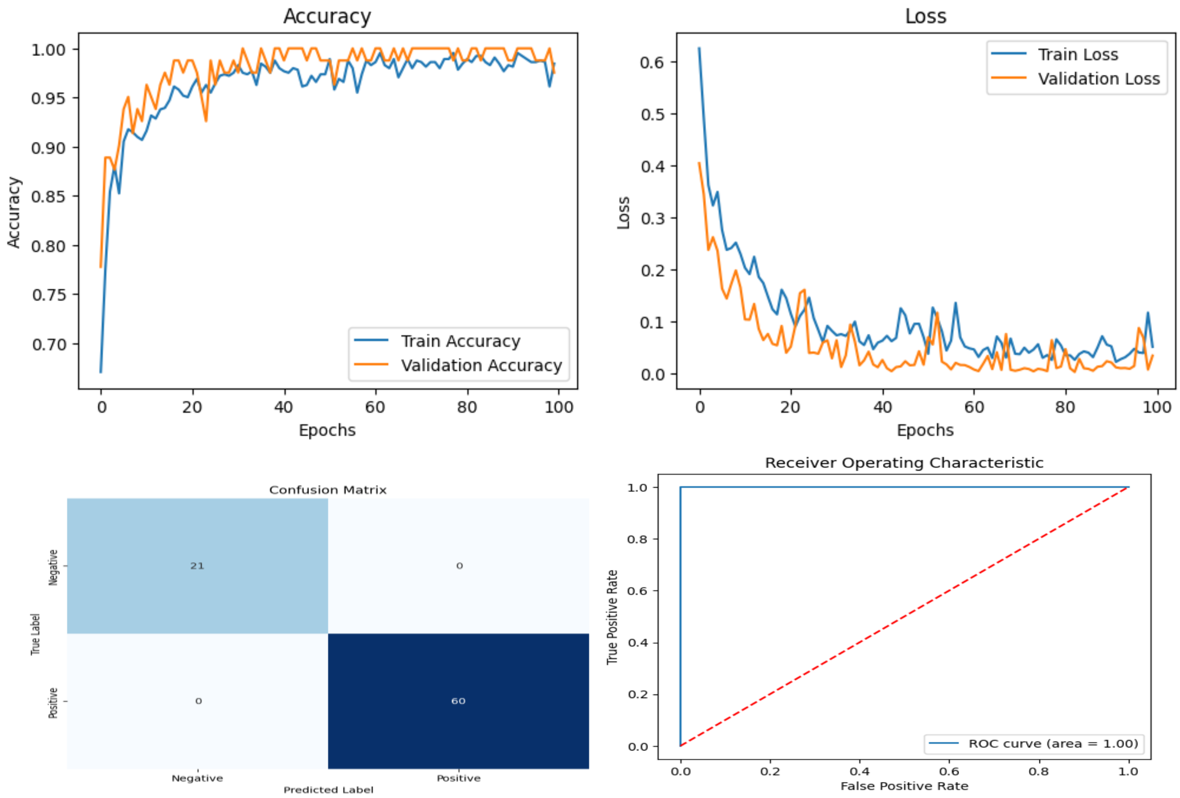

Fig 3.7: SimpleRNN Performance Metrics – Accuracy, Loss, and Confusion Matrix and Roc Curve

Fig. 3.7 depicts the performance measures of the SimpleRNN model, together with accuracy, loss, confusion matrix, and ROC curve. The model attained almost flawless accuracy, with training and validation accuracy rapidly converging. The confusion matrix indicates 100% accuracy in classification, while the ROC curve reveals an AUC of 100%, signifying outstanding performance. Both training and validation losses decreased markedly, indicating successful model tuning.

Table 3.4: Testing of algorithms prediction with test set

CNN			**LSTM**			**Simple RNN**		
Actual	Predicted	Confidence	Actual	Predicted	Confidence	Actual	Predicted	Confidence
0	0	0.530593	0	0	0.834472	0	0	0.996326
1	1	0.976719	1	1	0.875242	1	1	0.999300
1	1	0.801012	1	1	0.863130	1	1	0.999303
1	1	0.995636	1	1	0.883799	1	1	0.998057
1	1	0.993566	1	1	0.465355	1	1	0.988812

Table 3.4 illustrates the efficacy of CNN, LSTM, and SimpleRNN algorithms in forecasting diabetes using test set data. With high confidence, all three models accurately classified both "non-diabetic" and "diabetic" categories, with a projected probability of around 1 for the right classifications. The findings indicate that all evaluated neural network models can successfully differentiate between diabetes and non-diabetic patients. This illustrates the efficacy of these algorithms in medical predictive tasks, particularly with empirical datasets.

Table 3.5: Comparison of the results for proposed algorithms for UCI dataset

Authors	**Methods**	**Dataset**	**Accuracy**
Karthikeyini, S., et al. [3]	KNN, DT, RF, SVM, NB, XGBoost, ANN, CNN	UCI Dataset	99.04%
Teju, V., et al. [8]	Stacked ensemble-based deep neural network approach	UCI Dataset	99.36%
Proposed Method	**CNN, LSTM, SimpleRNN**	**SIDD-806**	**99.99%**

Table 3.5 illustrates a relative investigation of the performance of several algorithms on the UCI dataset. The suggested technique, including CNN, LSTM, and SimpleRNN, has surpassed prior methodologies, attaining 99.99% accuracy on a primary dataset. Conversely, conventional ML techniques used on the UCI dataset, including KNN, SVM, and ANN, produced 99.04% accuracy and stacked ensemble deep learning methodology enhanced performance to 99.36%. The results illustrate the efficacy of sophisticated neural networks for precise predictions on fluctuating datasets.

CONCLUSION

Diabetes is becoming more common in countries like India, which shows that early diagnosis, lowering risk factors, and effective disease control are important for public health to avoid problems. In this study, we gathered real time data and created a standard dataset, tentatively named "Southern India Diabetes Dataset (SIDD)". This article aimed to evaluate the accuracy of deep learning algorithms, including CNN, LSTM, and SimpleRNN, in premature-phase diabetes forecast. Since diabetes is crucial for efficient treatment and the avoidance of severe consequences, since diabetes constitutes a global health issue. Appraised the models using an 806-patient records dataset and all achieved remarkable levels of accuracy. Concurrently the SimpleRNN achieved 99.99% accuracy. These results will be verified by the medicos. After the verification, we will check how well deep learning approaches are accurately forecasting diabetes with new samples, particularly in its earliest stages. This would provide medical professionals with strong tools to help quickly identify disease and helps in healthcare industry to facilitate prompt diagnosis, thereby enhancing patient outcomes and potentially reducing the incidence of diabetes-related comorbidities. In the future, we aim to broaden our SIDD dataset by incorporating we are not only focusing on diabetic data but also will focus on non-diabetes data, ultimately transforming it into a SIDD standard dataset, an open resource for researchers to explore and advance innovations in this field.

REFERENCES

[1] Uddin, M. A., Islam, M. M., Talukder, M. A., Hossain, M. A. A., Akhter, A., Aryal, S., & Muntaha, M. (2024). Machine learning based diabetes detection model for false negative reduction. Biomedical Materials & Devices, 2(1), 427-443.

[2] Retal, S., Sahbani, H., Kharmoum, N., Rhalem, W., & Ezziyyani, M. (2022, May). Machine learning for diabetes prediction: a systematic review and a conceptual framework for early prediction. In International Conference on Advanced Intelligent Systems for Sustainable Development (pp. 75-83). Cham: Springer Nature Switzerland.

[3] Karthikeyini, S., Rupa, M., Athira, S., & Ravikumar, M. (2023, January). Forecast the Early Stage of Diabetes Mellitus Using Machine Learning. In International Conference on Communication and Computational Technologies (pp. 569-578). Singapore: Springer Nature Singapore.

[4] Skyler, J. S., Bakris, G. L., Bonifacio, E., Darsow, T., Eckel, R. H., Groop, L., ... & Ratner, R. E. (2017). Differentiation of diabetes by pathophysiology, natural history, and prognosis. Diabetes, 66(2), 241-255.

[5] Falvo, D., & Holland, B. E. (2017). Medical and psychosocial aspects of chronic illness and disability. Jones & Bartlett Learning.

[6] Pandeeswari, L., Rajeswari, K., & Phill, M. (2015). K-means clustering and Naïve Bayes classifier for categorization of diabetes patients. Eng Technol, 2(1), 179-185.

[7] Sahoo, P., & Bhuyan, P. (2021). Primitive diabetes prediction using machine learning models: an empirical investigation. Turkish Journal of Computer and Mathematics Education, 12(11), 229-236.

[8] Teju, V., Sowmya, K. V., Yuvanika, C., Saikumar, K., & Krishna, T. B. D. S. (2021, December). Detection of diabetes melittus, kidney disease with ML. In 2021 3rd international conference on advances in computing, communication control and networking (ICAC3N) (pp. 217-222). IEEE.

[9] Goodman, R. A., Posner, S. F., Huang, E. S., Parekh, A. K., & Koh, H. K. (2013). Peer reviewed: defining and measuring chronic conditions: imperatives for research, policy, program, and practice. Preventing chronic disease, 10.

[10] Casey, R., & Ballantyne, P. J. (2017). Diagnosed chronic health conditions among injured workers with permanent impairments and the general population. Journal of Occupational and Environmental Medicine, 59(5), 486-496.

[11] Shah, K., Punjabi, R., Shah, P., & Rao, M. (2020). Real time diabetes prediction using naïve bayes classifier on big data of healthcare. International Research Jounral of Engineering and Technology (IRJET), 7(5), 102-107.

[12] Mathur, P., Leburu, S., & Kulothungan, V. (2022). Prevalence, awareness, treatment and control of diabetes in India from the countrywide national NCD monitoring survey. Frontiers in public health, 10, 748157.

[13] Shankar, A. (2023). A Narrative Review Addressing the Gap in Diabetes Diagnosis: Challenges and Pathways in India. Student's Journal of Health Research Africa, 4(12), 10-10.

[14] Ahmad, R., & Haque, M. (2021). Oral health messiers: diabetes mellitus relevance. Diabetes, Metabolic Syndrome and Obesity, 3001-3015.

[15] World Health Organization. Non-Communicable Diseases. 2019. Available online: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases (accessed on 14 Sep 2024).

[16] World Health Organization. NCD Mortality and Morbidity. 2016. Available online: https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/ncd-mortality#:~:text=Noncommunicable%20diseases%20(NCDs)%20kill%2041,%2D%20and%20middle%2Dincome%20countries. (accessed on 14 Sep 2024).

[17] World Health Organization. Projections of Mortality and Causes of Death, 2016 to 2060. 2016. Available online: https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/ncd-mortality (accessed on 14 Sep 2024).

[18] https://mdrf.in/index.html.

[19] Halpern, A., Mancini, M. C., Magalhães, M. E. C., Fisberg, M., Radominski, R., Bertolami, M. C., ... & Nery, M. (2010). Metabolic syndrome, dyslipidemia, hypertension and type 2 diabetes in youth: from diagnosis to treatment. Diabetology & metabolic syndrome, 2, 1-20.

[20] Chaudhury, A., Duvoor, C., Reddy Dendi, V. S., Kraleti, S., Chada, A., Ravilla, R., ... & Mirza, W. (2017). Clinical review of antidiabetic drugs: implications for type 2 diabetes mellitus management. Frontiers in endocrinology, 8, 6.

[21] Alam, T. M., Iqbal, M. A., Ali, Y., Wahab, A., Ijaz, S., Baig, T. I., ... & Abbas, Z. (2019). A model for early prediction of diabetes. Informatics in Medicine Unlocked, 16, 100204.

[22] Harris, M. I., Klein, R., Welborn, T. A., & Knuiman, M. W. (1992). Onset of NIDDM occurs at least 4–7

yr before clinical diagnosis. Diabetes care, 15(7), 815-819.
[23] Ambady, R., & Chamukuttan, S. (2008). Early diagnosis and prevention of diabetes in developing countries. Reviews in Endocrine and Metabolic Disorders, 9, 193-201.
[24] Ergün, Ö. N. (2021). Early stage diabetes prediction using machine learning methods. Avrupa Bilim ve Teknoloji Dergisi, (29), 52-57.
[25] Naz, H., & Ahuja, S. (2020). Deep learning approach for diabetes prediction using PIMA Indian dataset. Journal of Diabetes & Metabolic Disorders, 19, 391-403.
[26] Chowdary, P. B. K., & 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), 519-525.
[27] Gurkan, C., Kozalioglu, S., & Palandoken, M. (2021). Development of Decision Support System for Early Stage Diabetes Diagnosis Using Machine Learning and Deep Learning Methods: A Comprehensive Analysis. Academic Perspective Procedia, 4(1), 186-194.
[28] Refat, M. A. R., Al Amin, M., Kaushal, C., Yeasmin, M. N., & Islam, M. K. (2021, October). A comparative analysis of early stage diabetes prediction using machine learning and deep learning approach. In 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC) (pp. 654-659). IEEE.
[29] Laila, U. E., Mahboob, K., Khan, A. W., Khan, F., & Taekeun, W. (2022). An ensemble approach to predict early-stage diabetes risk using machine learning: An empirical study. Sensors, 22(14), 5247.
[30] Khafaga, D. S., Alharbi, A. H., Mohamed, I., & Hosny, K. M. (2022, October). An Integrated Classification and Association Rule Technique for Early-Stage Diabetes Risk Prediction. In Healthcare (Vol. 10, No. 10, p. 2070). MDPI.
[31] Atif, M., Anwer, F., Talib, F., Alam, R., & Masood, F. (2023). Analysis of machine learning classifiers for predicting diabetes mellitus in the preliminary stage. Int J Artif Intell, 2252(8938), 1303.
[32] Yurttakal, A. H., & Baş, H. (2021). Possibility Prediction of Diabetes Mellitus at Early Stage Via Stacked Ensemble Deep Neural Network. Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi, 21(4), 812-819.