Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Identification of High-Risk Factors and Advanced Detection of Diabetes Utilizing a Hybrid Conv-LSTM Model

Alhuseen Omar Alsayed[1,2,*], Nor Azman Ismail[1], Layla Hasan[1]

[1] Faculty of Computing, Universiti Teknologi Malaysia, Skudai, 81310 Johor Bahru, Johor, Malaysia
[2] Department of Research Affairs, Deanship of Scientific Research, King Abdul Aziz University, Jeddah 22254, Saudi Arabia

ARTICLE INFO

Article history:
Received 27 January 2025
Received in revised form 21 February 2025
Accepted 30 June 2025
Available online 20 July 2025

Keywords:
Diabetes prediction; risk factor identification; Conv-LSTM

ABSTRACT

Diabetes is a chronic disease that causes various damages to the human body, making early detection crucial. Hence, to address this issue, the current study utilizes hybrid convolutional long-short term memory (Conv-LSTM) Network which help to detect and classify diabetes at the early stages. The proposed Conv-LSTM enhances the model's prediction by allowing CNN for spatial extraction of feature and LSTM for temporal extraction of feature from the input data. The proposed approach is applied to BRFSS dataset through the implementation of a computerized system for early identification of diabetes. The data gathered from the BRFSS dataset undergoes pre-processing step to ensure that it is suitable for further processing. The pre-processed data is then fed into the Conv-LSTM model which is trained to identify diabetes based on the risk factor. The efficacy of the proposed CGRU framework has been proven by validating the experimental findings with the existing state-of-the-art approaches. Compared to existing methods like machine learning, the proposed framework exhibited better performance. This demonstrates the efficacy of the Conv-LSTM architecture for diabetes prediction achieving high accuracy rate of 98.5%. The approach successfully identifies people who are at high risk of acquiring diabetes and achieves high accuracy in early diabetes detection, allowing for prompt intervention and individualized healthcare treatment.

1. Introduction

Diabetes is one of the most prevalent and deadly chronic diseases that can cause several complications if left untreated or misdiagnosed [1]. As per the report of "International Diabetes Federation" during the year 2019, 463 million people worldwide suffer from diabetes and it is estimated that by 2045, the number is most likely to hike to 700 million. Furthermore, the number of individuals at increased risk of developing diabetes is 374 million [2]. Diabetes occurs when the blood contains more glucose level than the normal range. It can cause major harm to the kidneys, nerves, eyes, heart and blood vessels if left undiagnosed and untreated. Untreated diabetic condition can lead to severe health complications such as kidney disease, nerve issues and heart problems [3].

* *Corresponding author*
E-mail address: nuriy3@graduate.utm.my

https://doi.org/10.37934/ard.137.1.136150

There are two different types of diabetes: "type 1" and "type 2". Usually, type 1 diabetes affects young adults, primarily those under 30 [4]. Patients with type 1 diabetes mellitus experience a partial or complete loss of insulin. However, to date, there is still no effective method to prevent type 1 diabetes mellitus. On the other hand, type-2 diabetes is not entirely curable and is more frequent among middle-aged and older adults. Insulin therapy is necessary for optimal medication of type 1 diabetes. However, people with type-2 diabetes can live normal lives by managing their lifestyle and getting frequent checkups [5]. Individuals who have been suffering from diabetes for an extended period and have elevated blood sugar levels are at a greater risk of developing diabetic retinopathy [6]. Early detection and management of diabetes are essential to prevent blood sugar levels from rising to dangerous levels and causing serious complications [7]. Advanced computer methods are needed to quickly analyse different types of data, find hidden patterns and predict the risks and detect early signs of the disease accurately [8,9].

Numerous studies have examined diverse techniques to predict diabetes by examining large quantities of health records [10]. In recent times, various machine learning mechanisms like "Naïve Bayes, linear regression, support vector machines, artificial neural networks, decision trees and random forests" [11], have achieved improved accuracy in effectively predicting diabetes and identifying potentially related conditions. In recent years, deep learning techniques have emerged as a promising approach for predictive tasks in various domains. Deep learning (DL) model has been used in various domains like health care, agriculture, blockchain [12]. In this research, the DL algorithm is implemented to detect diabetic conditions. The first objective of the study is to compile a list of the primary factors responsible for causing diabetes and then prioritizing them from most significant to least significant. The second objective is to detect diabetic conditions according to the key factors identified. The key contributions of the research are mentioned as follows:

i. Identification of High-Risk Factors: The key risk factors for predicting diabetes include factors like age, BMI, cholesterol, heart disease and physical activity. Early identification of these high-risk factors is critical for identifying diabetic condition and preventing its development to the next stage.
ii. Hybrid Deep Learning Model: The novel approach, namely Conv-LSTM is proposed for early identification and detection of diabetes. The model incorporates both spatial and temporal correlations in medical data, providing a thorough knowledge of developing a model for disease prediction.
iii. Improved Diagnostic Accuracy: Compared to conventional techniques, the proposed Conv-LSTM method obtains higher accuracy in the early identification of diabetes. This development demonstrates the efficiency of hybrid deep learning framework in the field of healthcare analytics, particularly in assisting early identification of diabetes.

The recent literatures on existing diabetes prediction model along with the identified problem is discussed in Section II, while Section III provides detailed explanation of the proposed Hybrid Conv-LSTM mechanism for diabetes prediction. This is followed by the results and discussion in Section IV and finally, Section V concludes the research.

2. Related Works

Artificial intelligence (AI) and machine learning have demonstrated transformative potential across various fields for instance, AI enhances cultural heritage presentations through Batik pattern classification [13], supports holistic heritage retirement planning [14] and improves image quality assessment [15]. In education, AI enhances usability in gamified programming courses [16] and student recommendation [17], powers personalized services such as E-Hala restaurant recommendations [18]. Additionally, AI-driven diabetes prediction and severity grading [19] contribute to the advancement of mobile health solutions. These studies highlight AI's versatility in several fields including health care domain. Deep learning continues to drive advancements in AI by enabling more accurate predictions, improved pattern recognition and enhanced decision-making across multiple domains. Numerous studies have examined diverse techniques to forecast diabetes, highlighting the growing role of AI in enhancing predictive healthcare solutions.

Ramesh *et al.,* [20] proposed the approach on early identification of diabetes using machine learning approach. In order to optimize the hyperparameter for early detection and diagnosis of diabetes, the "Optimal Scrutiny Boosted Graph Convolutional LSTM" (O-SBGC-LSTM) is employed in this current study. The SBGC-LSTM is enhanced using the "Eurygaster Optimisation Algorithm" (EOA). The study offers temporal hierarchical structure which increases the top SBGC-LSTM's temporal responsive fields while lowering computation costs and improving the capacity to learn high-level semantic representations. The research obtains an accuracy just below 99.8% but is limited by insufficient computational technique, consequently making it more time consuming. Similarly, Qteat *et al.,* [21] suggested a technique for diabetes classification utilizing hybrid model of PSO and multi-layer perceptron neural network. The research used "DataPal" dataset to predict diabetes, where the dataset is applied to the machine learning technique to provide precise detection of diabetes. To diagnose the type of diabetes, the research uses 9 predictors across 314 instances involving females with diabetes. The KNN algorithm was employed to select the optimal features while SVM model was selected for diabetes prediction. The "Particle Swarm Optimization" was used to find the optimal solution and overall, the PSO-MLPNN was applied to pre-process the data. The model obtained an accuracy of 98.7%. However, the research is limited by the use of a small dataset and high computational time. In another research, Noviyanti *et al.,* [22] used random forest algorithm for early detection of diabetes. The framework was implemented on PIMA Indian Diabetes dataset. Early identification of diabetes using the Random Forest algorithm achieved an accuracy of 87%. The research stated that random forest algorithm can enhance the performance of early diabetes prediction but the model is limited in performance when compared to other machine learning models including feature selection and data balancing. However, a major drawback of the study is the use of a limited dataset. From the review, it can be noted that prior studies on early diabetes detection and monitoring used different machine learning models and faced various limitations. These include insufficient computational approach, high processing time, limited datasets and also a lack of real time monitoring capabilities. Therefore, to overcome this limitation, this current research utilizes the hybrid Conv-LSTM to improve accuracy and enable early detection of diabetes.

3. Hybrid Conv-LSTM for Early Detection of Diabetes and their Risk Factor

The proposed Conv-LSTM model identifies the existence of diabetes by determining the risk factors. These factors are trained on the Conv-LSTM model to capture both the temporal and spatial and temporal patterns and classify the existence of diabetes. Figure 1 shows the workflow of the proposed Conv-LSTM model for detecting diabetes.

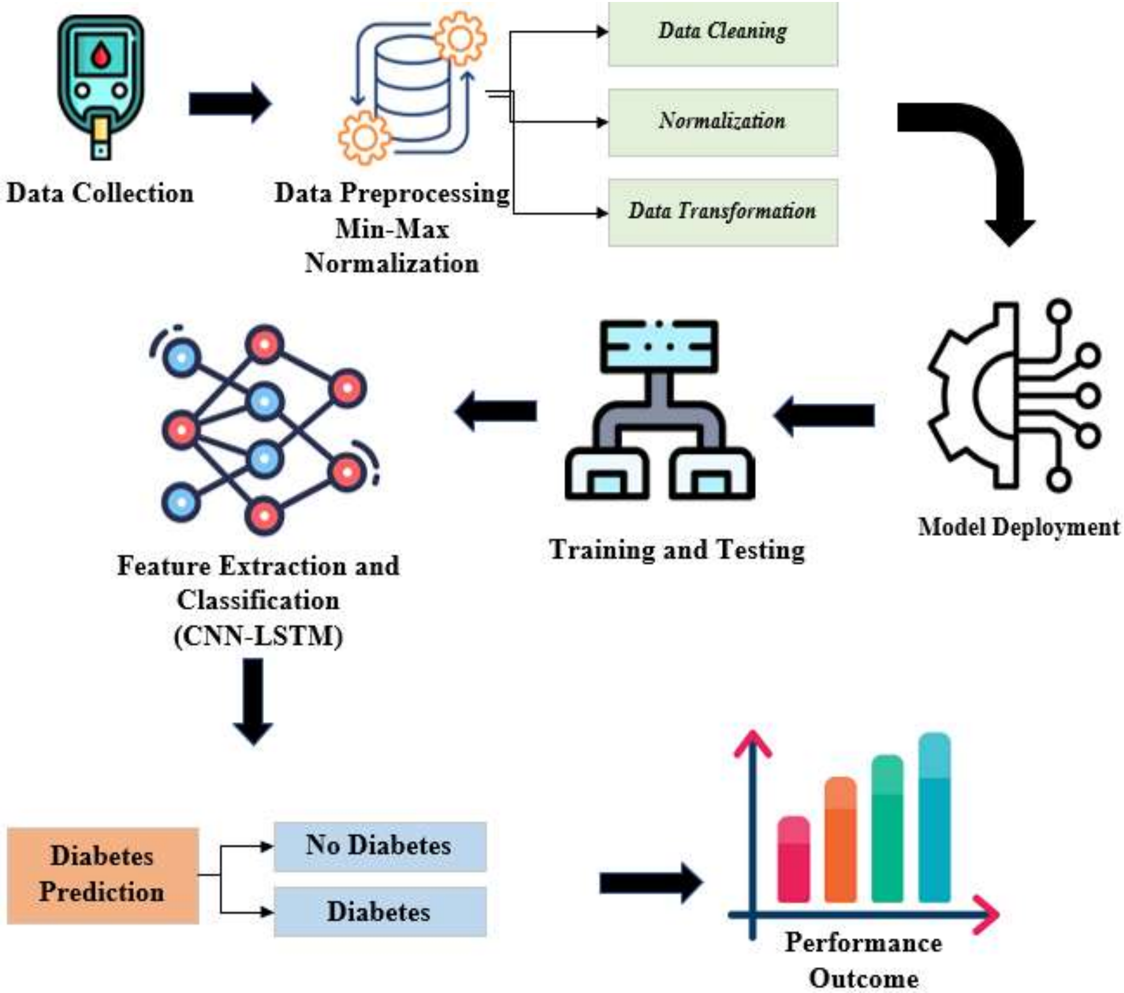

Fig. 1. Proposed hybrid Conv-LSTM for diabetes prediction

3.1 Data Collection

The Behavioural Risk Factor Surveillance System (BRFSS) dataset from the open source Kaggle database was used for diabetes prediction. A cleaned version of BRFSS 2015 survey comprising of 22 characteristics and 253,680 items was gathered.

3.2 Data Pre-Processing

Data pre-processing plays a crucial role in cleaning, transforming and preparing raw data for further analysis. It is used in both data analysis and deep learning. Data pre-processing ensures data integrity by removing the duplicating values, imputing data and handling the missing values by using suitable techniques like mean or mode imputation or by converting continuous data into discrete data [23]. Moreover, the collected data will be converted into a suitable form that can be used for further analysis. Data cleaning is accomplished by addressing missing values and removing the duplicated values. Data normalization was employed to rescale the numerical characteristics of datasets to a standard range. Z-normalization and batch normalization were employed in this work. The feature vector is updated with a zero-mean and a variance of one using the Z-normalization approach [24]. The normalization layer in the batch normalization approach is defined as follows in Eq. (1) and Z-normalization was executed on the output of the preceding layer below:

$$\check{z}_i = \frac{z_i - \gamma_a}{\sigma_a} \quad (1)$$

Where, "σ_a" represents the standard deviation of the current mini-batch and "γ_a" denotes its mean. Next, " γ " is used to shift the output values and the normalized value "$\check{z}_i$" is obtained by subtracting the batch mean and dividing by the batch standard deviation.

3.3 Hybrid Conv-LSTM for Feature Extraction and Classification

Deep learning model could be more effective to extract the intricate feature from the unprocessed data and it is widely used in the classification process. In deep learning model, CNN is one of the important subsets which is highly useful for both feature extraction and classification. CNNs improve diabetes prediction and classification accuracy by gradually extracting local characteristics from inputs, which help with early diagnosis and efficient treatment of the condition. Figure 2 depicts a hybrid Conv-LSTM model architecture that combines the advantages of LSTMs for capturing temporal dependency with CNNs for extracting spatial feature [25].

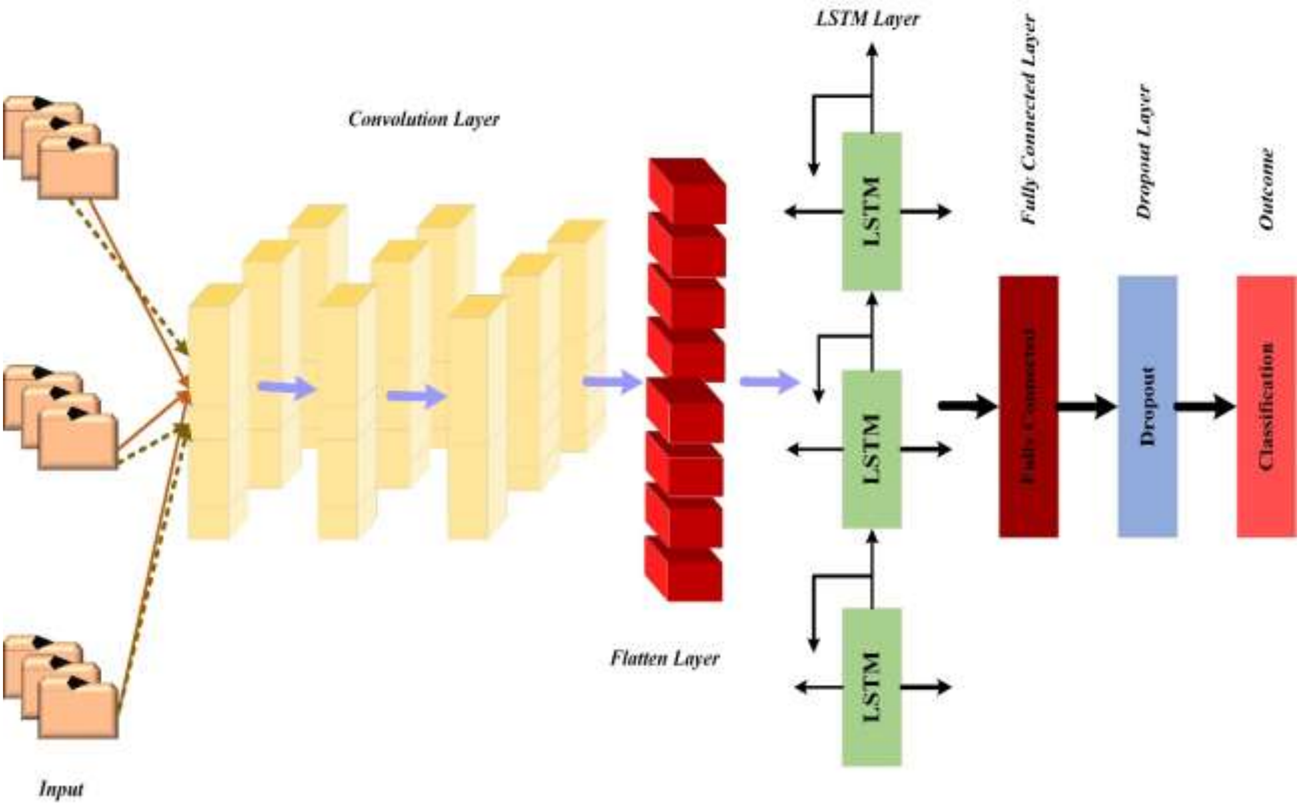

Fig. 2. Architecture of hybrid CNN-LSTM

3.3.1 Input layer

The process starts with an input layer that takes the initial data which is then sent into the convolutional layers for the extraction of crucial spatial characteristics. The input layer ensures that the data is prepared correctly for the advanced feature extraction processes carried out by the upcoming convolutional layers.

3.3.2 Convolutional layer

The data from the input layer is fed into the convolutional layer. To extract discrete elements and unique attributes from the data sequence, a one-dimensional filter is applied over the sensor signal in the context of this convolutional procedure. Considering a series of inputs, a convolutional layer's goal is to extract unique characteristics; characterized by Eq. (2):

$$a_z^{u,x} = \sigma\left(q_x + \sum_{y=1}^{Y} s_y^x \, g_{z+y-1}^{0,x}\right) \tag{2}$$

In Eq. (2) the index layer is denoted as "u", the non-linear activation function is denoted as "σ" the bias term is represented as q_x for the term "x" and the convolutional filter is represent as y. The weight of the feature map "x" is represented as s_y^x and y represents the feature index [26].

3.3.3 Max-pooling layer

In the max-pooling layer, the feature maps generated by the convolution layer is down sampled, which reduces the computational complexity. Moreover, the max-pooling process develops a pooled outcome by dividing the input feature map into non-overlapping region and choosing the maximum value from each region. This helps in conserving important spatial information while reducing the spatial dimensions of the feature maps, making them more controllable for subsequent layers. In diabetes prediction, this helps to efficiently extract specific feature from the given input data. This improved feature contributes to the overall predictive accuracy of the CNN framework which helps in early detection of risk assessment of diabetes.

3.3.4 Flatten layer

The flatten layer is essential for transiting from the feature extraction stage to the sequence processing stage. Convolutional layers provide multi-dimensional feature maps that record the patterns and spatial properties found during convolution after processing the input data. However, these feature maps remain in a multi-dimensional format that is not immediately compatible with LSTM networks' sequential processing architecture. In order to solve this, the flatten layer turns these multi-dimensional feature maps into a vector that is only one dimension. To complete this procedure, every value from the feature maps must be assembled into a single, continuous linear array. By reshaping the data into a format that can be fed into the LSTM layers, this transformation ensures that the learnt spatial properties are preserved in the data [27].

3.3.5 LSTM layer

The one-dimensional vector from the Flatten layer is fed into the LSTM layer. Since the standard neural networks cannot recall previous data, they must begin learning from scratch every time. LSTMs incorporate a cell state, which serves as the network's memory and endures through each time step. LSTMs employ gates to manage information flow, keeping only pertinent data and eliminating unnecessary data. The "forget gate" chooses the data from the previous cell state to keep, while the "input gate" selects new data to enter. The "output gate" determines the hidden state of the next phase. To update the cell state, the new candidate values from the input gate are combined with the output of the forget gate, which selects the information to retain from the previous cell state. The new hidden state, which combines the prior hidden state with the current input, is then computed using the updated cell state. This technique enables the LSTM to maintain long-term dependencies, which is crucial for accurately identifying diabetes based on historical health data.

The following Eq. (3) to (8) explain how the cell state (C^t) and hidden state (h^t) are updated:

i. Forget gate: Decide which information to discard from the cell state and Eq. (3) is used to calculate the forget gate:

$$f^t = \sigma(A_f.[h^{t-1}, x_t] + d_f) \tag{3}$$

ii. Input Gate: Decide which information to store in the cell state. Eq. (4) and Eq. (5) are used to calculate the input gate:

$$i^t = \sigma(A_f.[h^{t-1}, x_t] + d_i) \tag{4}$$

$$C'^t = \tanh(A_f.[h^{t-1}, x_t] + d_c) \tag{5}$$

iii. Cell State Update: Combine the previous cell state and new candidate values and calculated using Eq. (6):

$$C^t = f_t * C^{t-1} + i_t * C'^t \tag{6}$$

iv. Output Gate: The next hidden state is determined by the updated cell state and it is calculated using Eq. (7):

$$O^t = \sigma(A_f.[h^{t-1}, x_t] + d_o) \tag{7}$$

$$h_t = o^t * \tanh(C^t) \tag{8}$$

In the above Eq. (3) to (8) the sigmoid function is defined as "σ", the hyperbolic tangent activation function is defined as tanh and the weight matrices is denoted as "A", the bias term is denoted as "d", the input of the time step is denoted as "x_t". Based on historical and sequential medical data, LSTMs are well-suited for diabetes categorization and early diagnosis because they can efficiently capture long-term dependencies in patient health records.

3.3.6 Fully connected layer and drop out layer

The dropout layer and fully connected layer are essential elements that improve the generalization and performance of deep learning-based diabetes prediction models. The fully connected layer, which is frequently located at the end of convolutional neural networks, allows information taken from earlier layers to be integrated and interpreted to provide final predictions. On the other hand, the dropout layer is a regularization method that randomly sets a portion of the input neurons to zero throughout each training cycle in order to prevent overfitting problem [28].

4. Result and Discussion

The proposed model is implemented using python software in windows 10 platform. Various performance metrics like Accuracy, precision, recall, F1-score were used for the evaluation of the hybrid Conv-LSTM model's performance for the early detection of diabetes. The general health ratings by Heart Risk Status are shown in Figure 3.

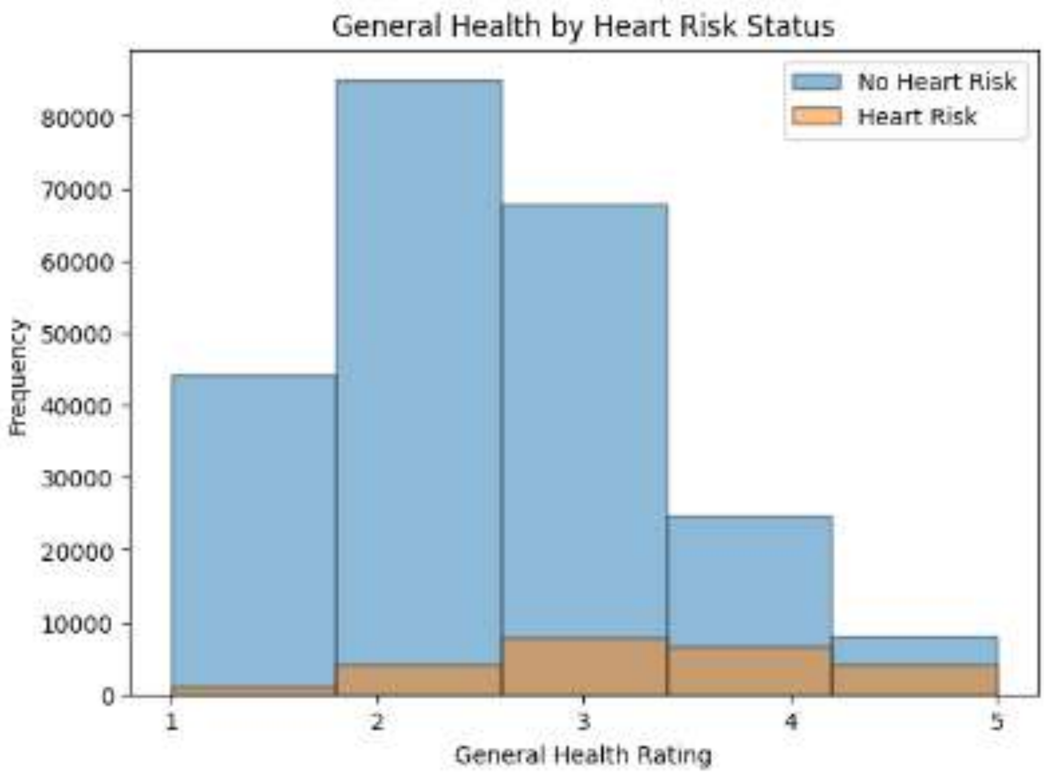

Fig. 3. Heart risk status

Figure 4 shows the scatter plot of age *vs* BMI with diabetes. This graph visually evaluates the link between age, BMI and diabetes in order to provide assistance for clinical studies by pointing out relevant patterns and correlations.

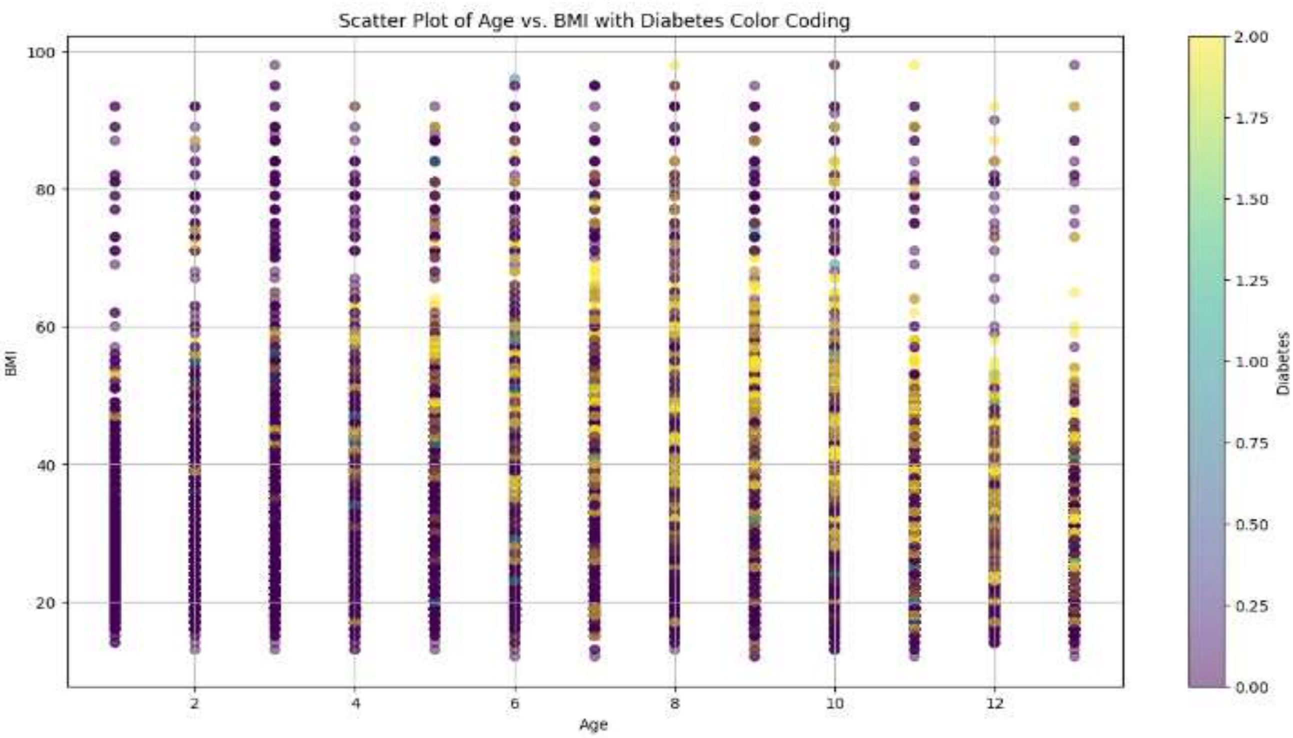

Fig. 4. Scatter plot of age *vs* BMI with diabetes

Figure 5 shows the sample of different health-related data. Each one illustrating information on specific health aspects such as the distribution of age, sex, BMI, the number of smokers and non-smokers and the percentage of fruits and vegetables consumed, physical health, mental health, education and income. These graphical representations aid in comprehending the distribution of data and demonstrating any relationships between variables.

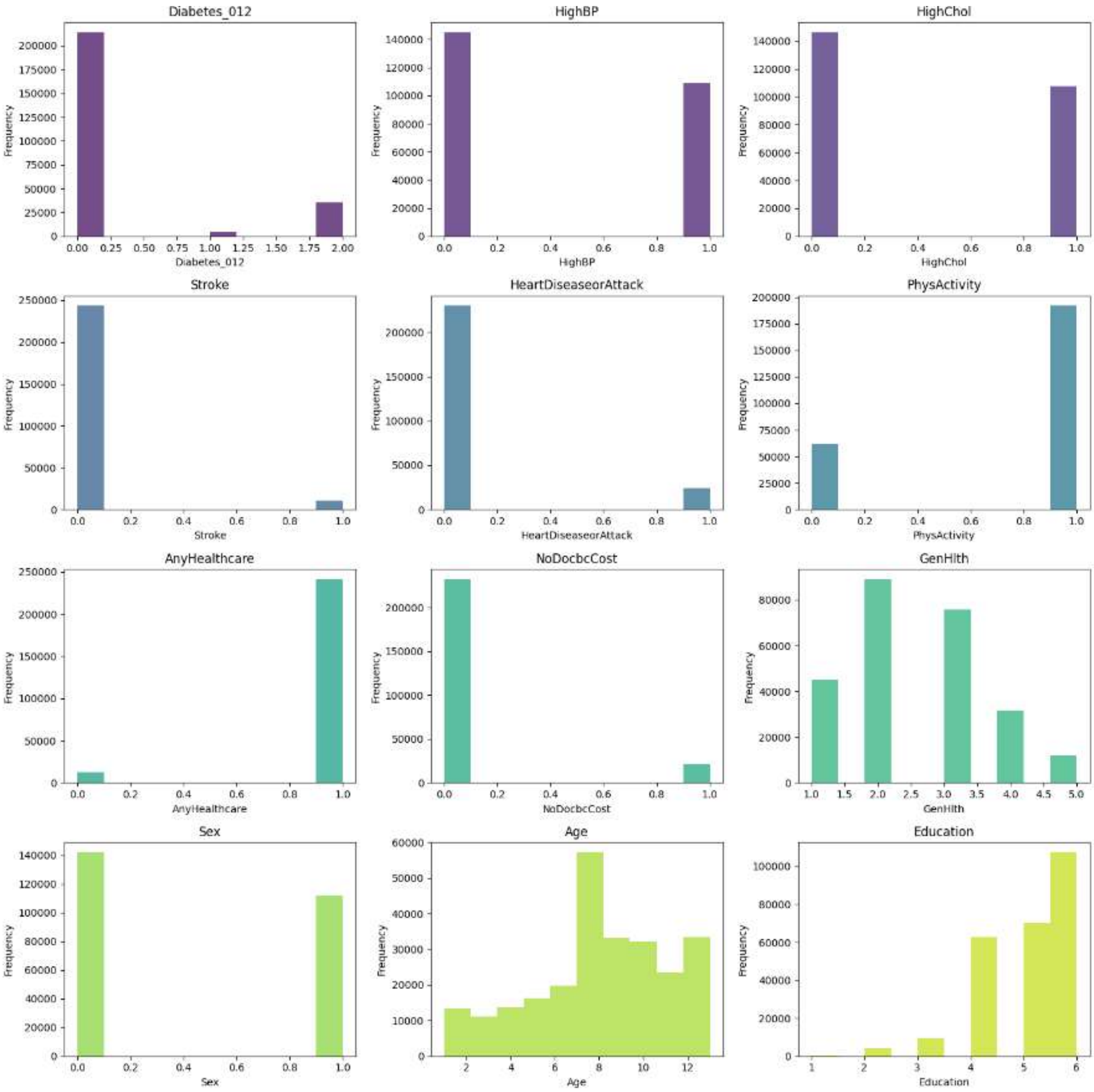

Fig. 5. Scatter types of health-related data

Figure 6 depicts a correlation matrix heatmap displaying the correlation coefficients among several factors. On the scale, which runs from -0.6 to 1.0, greater hues denote stronger positive or negative correlations. In the domains of statistics, data analysis and machine learning, this heatmap is highly beneficial for feature selection and understanding of data structures.

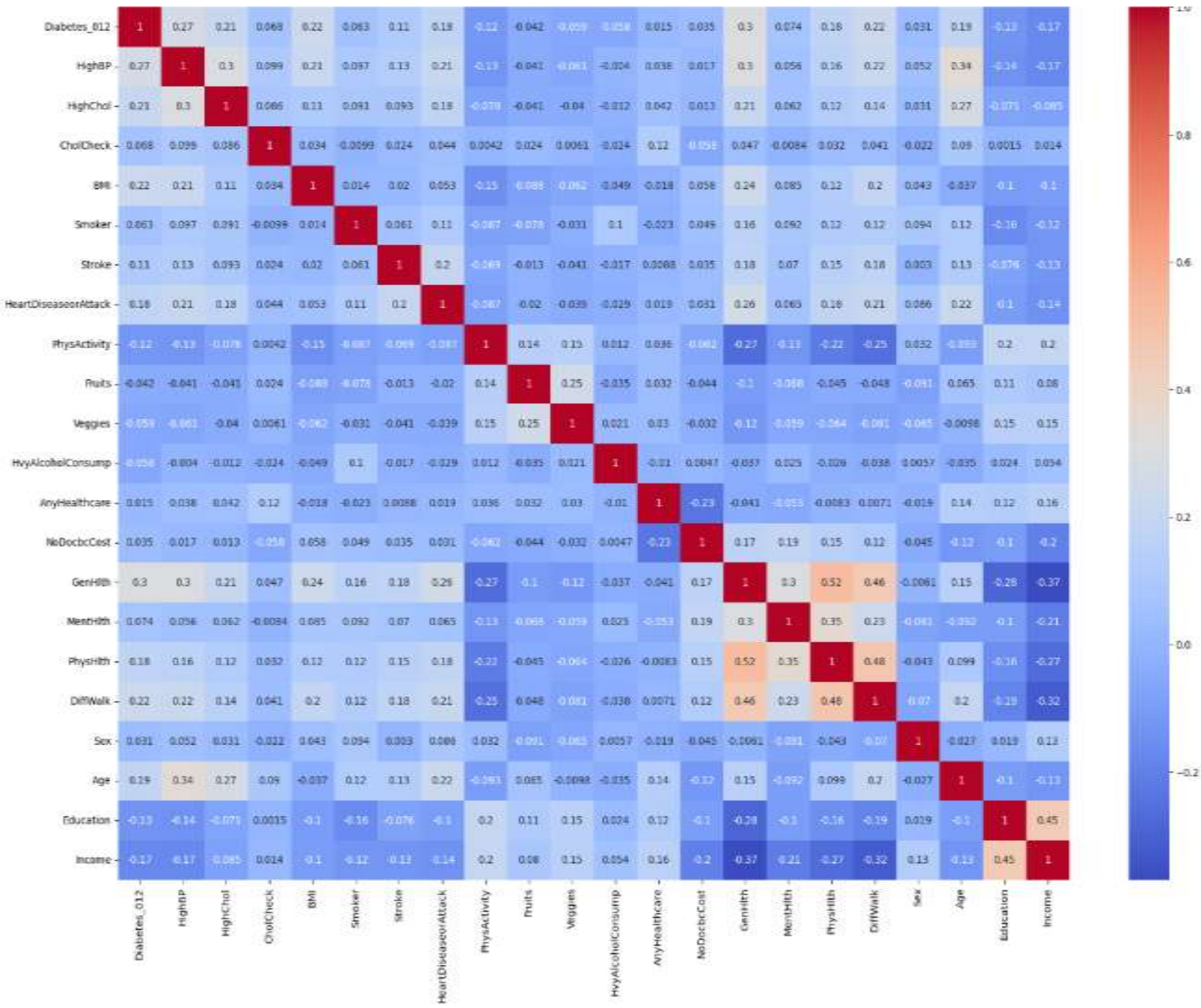

Fig. 6. Correlation matrix

Figure 7 displays two plots with the training and validation metrics for a machine learning model across 30 epochs. For both datasets, both metrics show great accuracy; they initially rise rapidly, then level out at 0.99 and 0.98, respectively.

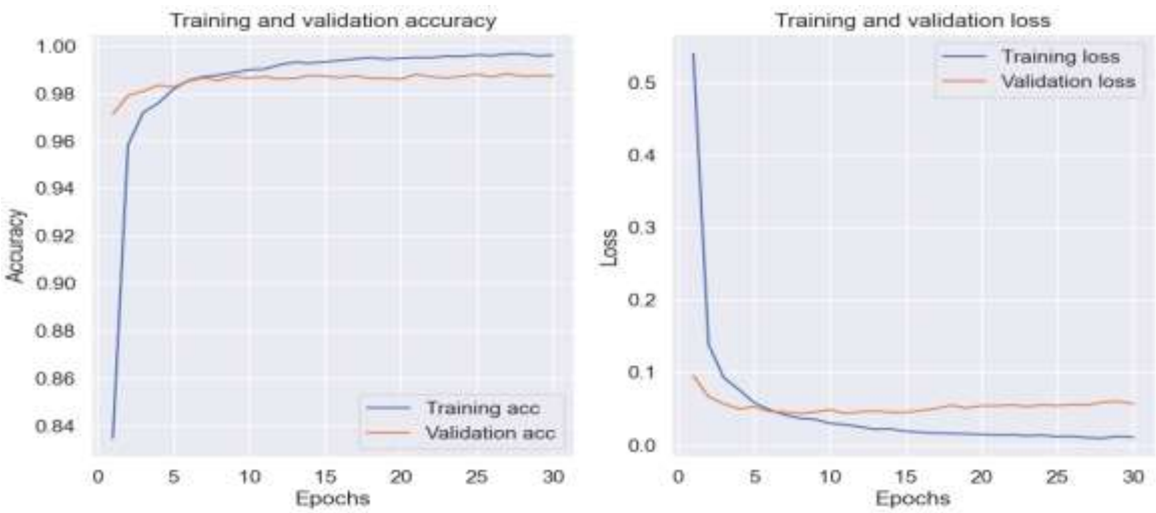

Fig. 7. Training and validation metrics

Figure 8 confusion matrix shows 357 true negatives, 117 true positives, 10 false positives and 10 false negatives, indicating the model's high accuracy in predicting diabetic and non-diabetic cases. The colour intensity represents the frequency of each outcome, with darker shades indicating higher counts. Figure 9 shows the Region of Convergence curve, demonstrating the efficacy of model with an accuracy of 98.5%.

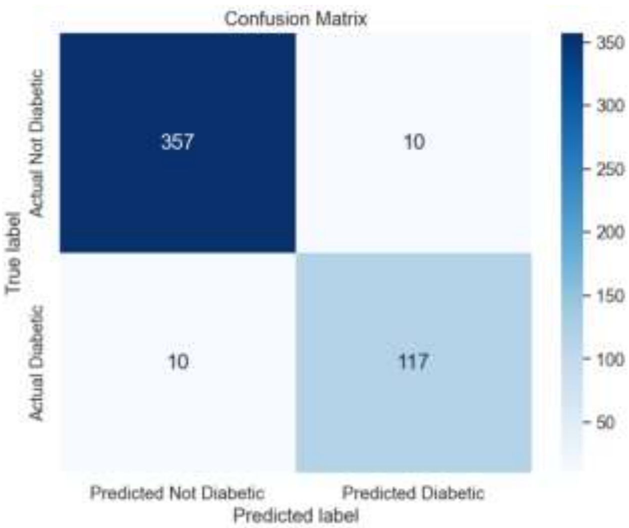

Fig. 8. Confusion matrix

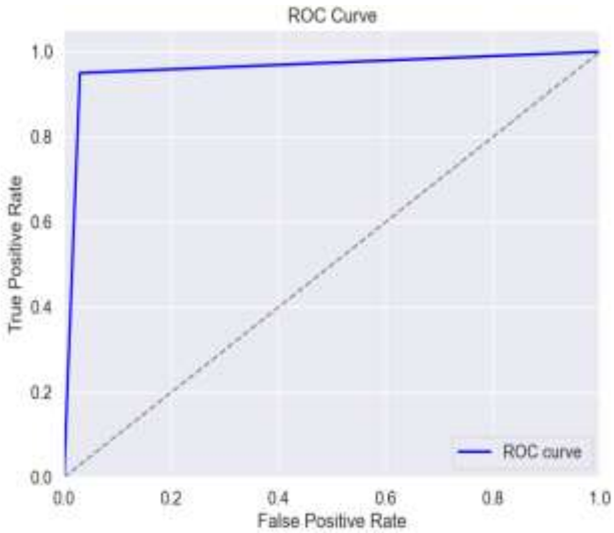

Fig. 9. ROC curve

4.1 Performance Evaluation

i. Accuracy: The percentage of accurate outcomes among all instances investigated is known as accuracy and it is calculated using Eq. (9):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{9}$$

In this case, TP stands for “true positive”, TN for “true negative”, FP for “false positive” and FN for “false negative”.

ii. Precision: Precision, which is computed using Eq. (10), is the percentage of positive findings that were properly identified.

$$Precision = \frac{TP}{TP+FP} \tag{10}$$

iii. Recall: The percentage of true positive instances that were accurately detected is known as recall or sensitivity and it is calculated using Eq. (11):

$$Recall = \frac{TP}{TP+FN} \tag{11}$$

iv. F1-Score: The F1-score, which can be computed using Eq. (12), is the harmonic mean of recall and accuracy:

$$F1\ Score = 2 \times \frac{Precison \times Recall}{Precison+Recall} \tag{12}$$

Figure 10 shows the metrics by class, demonstrating the performance of two classes namely, class 0 (Diabetes) and class 1 (No Diabetes). In both class 0 and class 1, the model achieved an accuracy of 98.5%. For class 0, the precision, recall and F1-score were 97%, while for class 1, the precision, recall and F1-score were 95%.

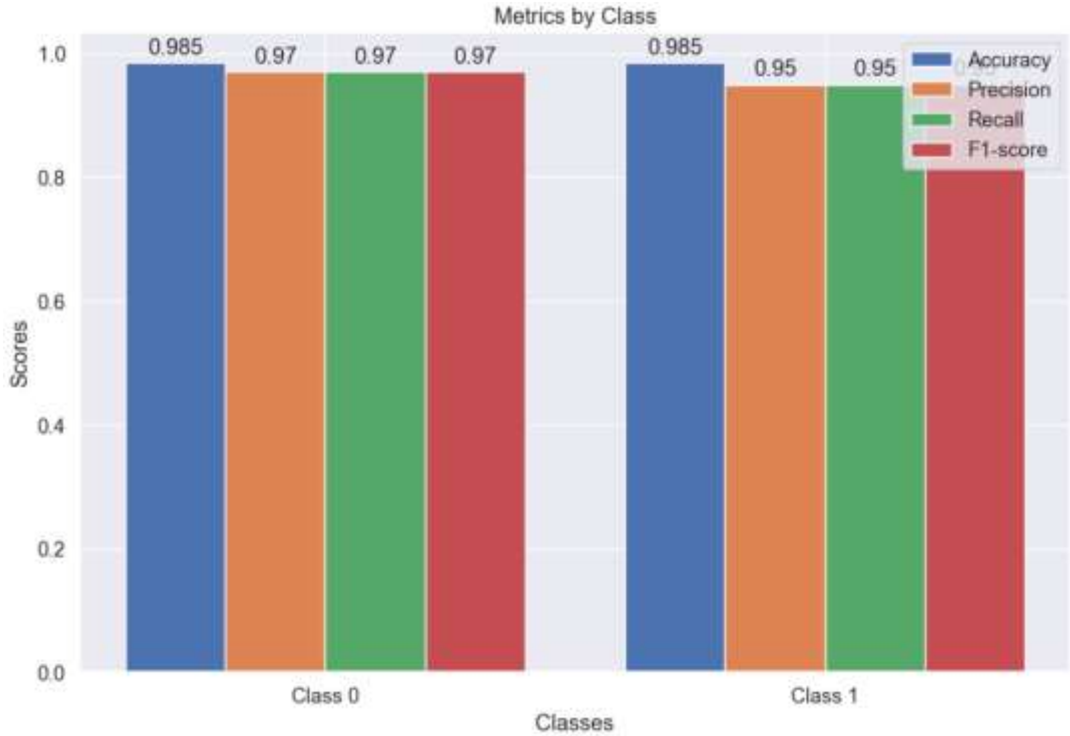

Fig. 10. Metrices by class

The performance comparison of existing and proposed model is illustrated in Figure 11 and Table 1. This shows that the proposed model has the higher accuracy of 98.5% compared to the neural network model which has lower accuracy of 82.4%.

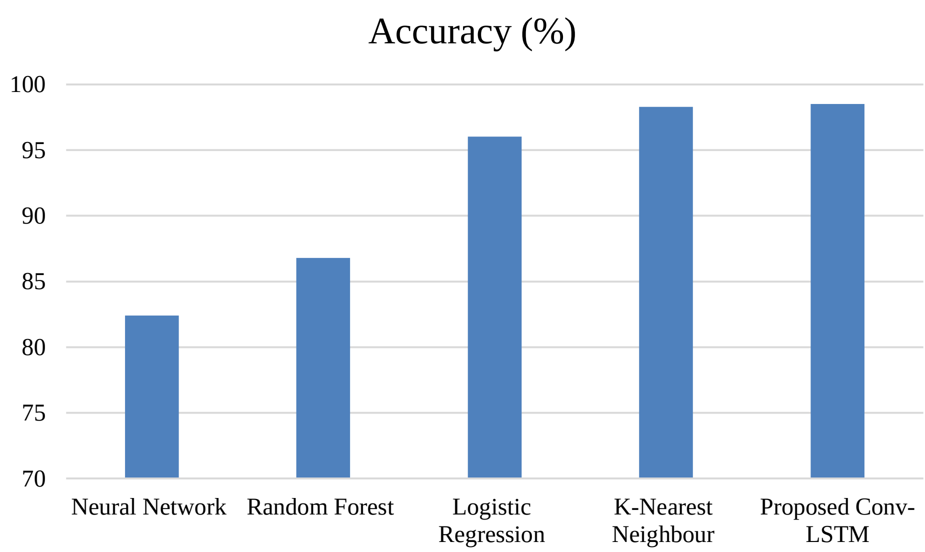

Fig. 11. Comparison of accuracy

Table 1
Performance comparison

Method	Accuracy
Neural Network [29]	82.4
Random Forest [30]	86.8
Logistic Regression [31]	96.02
K-Nearest Neighbour [32]	98.3
Proposed Conv-LSTM	98/5

4.2 Discussion

This study focuses on using a hybrid Conv-LSTM model to identify high-risk indicators and detect diabetes in its early stages. Diabetes is a chronic condition that requires accurate and timely diagnosis to prevent serious health complications. Conventional techniques for detecting diabetes sometimes entail laborious procedures and the manual extraction of features, which can be inaccurate and time-consuming [33]. High-quality data preparation is crucial, particularly when using the data for predictions. This discovery has significant ramifications for how healthcare is practiced. Diabetes patients have a higher quality of life and a lower chance of serious consequences when their condition is detected early and treated promptly. A reliable and automated method for detecting people at risk of diabetes is provided by the Hybrid Conv-LSTM model, enabling early diagnosis and individualized treatment regimens [34]. Moreover, feature significance analysis will be carried out to pinpoint important predictors that lead to the onset of diabetes, offering important information for efforts in preventive healthcare. The framework's efficacy in identifying high-risk factors and detecting diabetes early will be rigorously assessed, including a comparison with current techniques. Ultimately, the suggested framework aims to promote improvements in personalized medicine and healthcare analytics, enabling proactive diabetes treatment and lowering the risk of complications for those who are at-risk.

5. Conclusion

Diabetes is a disease that affects millions of individuals globally and this study is crucial for enabling early-stage detection to improve health outcomes and prevent complications. With an accuracy of 98.5%, this study's hybrid Conv-LSTM model for diabetes early detection outperforms other techniques including neural networks, random forests, logistic regression and K-Nearest Neighbour. The hybrid Conv-LSTM model improved the precision and dependability of diabetes prediction by utilizing deep learning techniques, demonstrating its efficacy. The model architecture may be modified to forecast other chronic diseases, increasing its usefulness in a range of healthcare settings. With its scalable and flexible solution for early disease identification and management, this technique has the potential to completely transform the analysis of health data. The goal of future work will be to improve the model's generalizability with the addition of more varied datasets and real-time data from wearables and other Internet of things medical instruments. The goal of this expansion is to create a complete prediction tool that can handle a range of health concerns, which will ultimately lead to a proactive and more individualized approach to treatment.

Acknowledgment
The authors would like to express gratitude to Universiti Teknologi Malaysia for the financial sponsorship of the research through the UTM Encouragement Research Grant (UTMER), grant reference number/no: PY/2022/03968; cost center: Q.J130000.3828.31J51.

References

[1] Naz, Huma, and Sachin Ahuja. "Deep learning approach for diabetes prediction using PIMA Indian dataset." *Journal of Diabetes & Metabolic Disorders* 19 (2020): 391-403. https://doi.org/10.1007/s40200-020-00520-5

[2] Rahman, Motiur, Dilshad Islam, Rokeya Jahan Mukti, and Indrajit Saha. "A deep learning approach based on convolutional LSTM for detecting diabetes." *Computational biology and chemistry* 88 (2020): 107329. https://doi.org/10.1016/j.compbiolchem.2020.107329

[3] Sarker, Iqbal H. "Machine learning: Algorithms, real-world applications and research directions." *SN computer science* 2, no. 3 (2021): 160. https://doi.org/10.1007/s42979-021-00592-x

[4] Li, Yukai, Huling Li, and Hua Yao. "Analysis and Study of Diabetes Follow-Up Data Using a Data-Mining-Based Approach in New Urban Area of Urumqi, Xinjiang, China, 2016-2017." *Computational and mathematical methods in medicine* 2018, no. 1 (2018): 7207151. https://doi.org/10.1155/2018/7207151

[5] Sneha, N., and Tarun Gangil. "Analysis of diabetes mellitus for early prediction using optimal features selection." *Journal of Big data* 6, no. 1 (2019): 1-19. https://doi.org/10.1186/s40537-019-0175-6

[6] Miklosik, Andrej, and Nina Evans. "Impact of big data and machine learning on digital transformation in marketing: A literature review." *Ieee Access* 8 (2020): 101284-101292. https://doi.org/10.1109/ACCESS.2020.2998754

[7] Alsayed, Alhuseen Omar, Nor Azman Ismail, and Layla Hasan. "Exploring user perspectives: empowering self-management strategies through semi-structured interviews in the context of mobile app-based approaches." In *Proceedings of the 2024 13th International Conference on Software and Computer Applications*, pp. 38-44. 2024. https://doi.org/10.1145/3651781.3651788

[8] Alsayed, Alhuseen Omar, Nor Azman Ismail, Layla Hasan, Asif Hassan Syed, Farhat Embarak, and Aminu Da'u. "A systematic literature review for understanding the effectiveness of advanced techniques in diabetes self-care management." *Alexandria Engineering Journal* 79 (2023): 274-295. https://doi.org/10.1016/j.aej.2023.08.026

[9] Alhuseen, Omar Alsayed, Nor Azman Ismail, Layla Hasan, and Farhat Embarak. "A comprehensive review of modern methods to improve diabetes self-care management systems." *International Journal of Advanced Computer Science and Applications* 9 (2023). https://doi.org/10.14569/IJACSA.2023.0140920

[10] Aslan, Muhammet Fatih, Kadir Sabanci, Akif Durdu, and Muhammed Fahri Unlersen. "COVID-19 diagnosis using state-of-the-art CNN architecture features and Bayesian Optimization." *Computers in biology and medicine* 142 (2022): 105244. https://doi.org/10.1016/j.compbiomed.2022.105244

[11] Oh, Wonsuk, Era Kim, M. Regina Castro, Pedro J. Caraballo, Vipin Kumar, Michael S. Steinbach, and Gyorgy J. Simon. "Type 2 diabetes mellitus trajectories and associated risks." *Big data* 4, no. 1 (2016): 25-30. https://doi.org/10.1089/big.2015.0029

[12] Wee, Boon Feng, Saaveethya Sivakumar, King Hann Lim, Wei Kitt Wong, and Filbert H. Juwono. "Diabetes detection based on machine learning and deep learning approaches." *Multimedia Tools and Applications* 83, no. 8 (2024): 24153-24185. https://doi.org/10.1007/s11042-023-16407-5

[13] Shaubari, Ezak Fadzrin Ahmad, Luluk Elvitaria, Noor Azah Samsudin, Shamsul Kamal Ahmad Khalid, Ira Puspita Sari, and Zul Indra. "A Data-Driven Approach for Batik Pattern Classification using Convolutional Neural Networks (CNN)." *Semarak International Journal of Electronic System Engineering* 4, no. 1 (2024): 22-30. https://doi.org/10.37934/sijese.4.1.2230a

[14] Ibrahim, Dayang Kartini Abang, Mohamad Hardyman Barawi, and Ahmad Shauffi Mohamad Sharkawi. "Harnessing AI for Comprehensive Retirement Planning and Preparation (AIRP2): A Multidimensional Approach to Financial, Career, Health, Social, Leisure and Spiritual Well-being." *International Journal of Computational Thinking and Data Science* 4, no. 1 (2024): 40-52. https://doi.org/10.37934/ctds.4.1.4052

[15] Verak, Nutveesa, Phaklen Ehkan, Ruzelita Ngadiran, Suwimol Jungjit, Siraya Sitthisarn, Mohd Nazri Mohd Warip, Mohd Zaizu Ilyas, and Fazrul Faiz Zakaria. "Deep Learning-based Blind Image Quality Assessment using Extreme Learning Machine." *Pena Journal of Computer Science and Informatics* 1, no. 1 (2025): 1-12. https://doi.org/10.37934/pjcsi.1.1.112

[16] Azzali, Fazli, Azizi Abas, and Roshidi Din. "Enhancing Usability in Gamified Computer Programming Courses for Higher Education." *Progress in Computers and Learning* 2, no. 1 (2025): 1-11. https://doi.org/10.37934/picl.2.1.111

[17] Alsayed, Alhuseen Omar, Mohd Shafry Mohd Rahim, Ibrahim AlBidewi, Mushtaq Hussain, Syeda Huma Jabeen, Nashwan Alromema, Sadiq Hussain, and Muhammad Lawan Jibril. "Selection of the right undergraduate major by students using supervised learning techniques." *applied sciences* 11, no. 22 (2021): 10639. https://doi.org/10.3390/app112210639

[18] Mahadi, M. I., Nurulhuda Zainuddin, Norzatul Bazamah Azman Shah, Nur Asyira Naziron, and Siti Fatimah Mohd Rum. "E-halal restaurant recommender system using collaborative filtering algorithm." *Journal of Advanced Research in Computing and Applications* 12, no. 1 (2018): 22-34.

[19] Alsayed, Alhuseen Omar, Nor Azman Ismail, Layla Hasan, Muhammad Binsawad, and Farhat Embarak. "Leveraging a hybrid convolutional gated recursive diabetes prediction and severity grading model through a mobile app." *PeerJ Computer Science* 11 (2025): e2642. https://doi.org/10.7717/peerj-cs.2642

[20] Ramesh, B., and Kuruva Lakshmanna. "A novel early detection and prevention of coronary heart disease framework using hybrid deep learning model and neural fuzzy inference system." *IEEE Access* 12 (2024): 26683-26695. https://doi.org/10.1109/ACCESS.2024.3366537

[21] Qteat, Haneen, and Mohammed Awad. "Using Hybrid Model of Particle Swarm Optimization and Multi-Layer Perceptron Neural Networks for Classification of Diabetes." *International Journal of Intelligent Engineering & Systems* 14, no. 3 (2021). https://doi.org/10.22266/ijies2021.0630.02

[22] Noviyanti, Cindy Nabila, and Alamsyah Alamsyah. "Early detection of diabetes using random forest algorithm." *Journal of Information System Exploration and Research* 2, no. 1 (2024). https://doi.org/10.52465/joiser.v2i1.245

[23] Sun, Hanqing, Zheng Liu, Guizhi Wang, Weimin Lian, and Jun Ma. "Intelligent analysis of medical big data based on deep learning." *IEEE Access* 7 (2019): 142022-142037. https://doi.org/10.1109/ACCESS.2019.2942937

[24] Tan, Y. Nguyen, Vo Phuc Tinh, Pham Duc Lam, Nguyen Hoang Nam, and Tran Anh Khoa. "A transfer learning approach to breast cancer classification in a federated learning framework." *IEEe Access* 11 (2023): 27462-27476. https://doi.org/10.1109/ACCESS.2023.3257562

[25] Saba, Tanzila, Ahmed Sameh Mohamed, Mohammad El-Affendi, Javeria Amin, and Muhammad Sharif. "Brain tumor detection using fusion of hand crafted and deep learning features." *Cognitive Systems Research* 59 (2020): 221-230. https://doi.org/10.1016/j.cogsys.2019.09.007

[26] Ordóñez, Francisco Javier, and Daniel Roggen. "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition." *Sensors* 16, no. 1 (2016): 115. https://doi.org/10.3390/s16010115

[27] Aslan, Muhammet Fatih, and Kadir Sabanci. "A novel proposal for deep learning-based diabetes prediction: converting clinical data to image data." *Diagnostics* 13, no. 4 (2023): 796. https://doi.org/10.3390/diagnostics13040796

[28] Maniruzzaman, Md, Md Jahanur Rahman, Benojir Ahammed, and Md Menhazul Abedin. "Classification and prediction of diabetes disease using machine learning paradigm." *Health information science and systems* 8 (2020): 1-14. https://doi.org/10.1007/s13755-019-0095-z

[29] Xie, Zidian, Olga Nikolayeva, Jiebo Luo, and Dongmei Li. "Building risk prediction models for type 2 diabetes using machine learning techniques." *Preventing chronic disease* 16 (2019): E130. https://doi.org/10.5888/pcd16.190109

[30] Bhola, Geetanjali, Aman Garg, and Manisha Kumari. "Comparative study of machine learning techniques for chronic disease prognosis." In *Computer Networks and Inventive Communication Technologies: Proceedings of Third ICCNCT 2020*, pp. 131-144. Springer Singapore, 2021. https://doi.org/10.1007/978-981-15-9647-6_10

[31] Kalyankar, Gauri D., Shivananda R. Poojara, and Nagaraj V. Dharwadkar. "Predictive analysis of diabetic patient data using machine learning and Hadoop." In *2017 international conference on I-SMAC (IoT in social, mobile, analytics and cloud)(I-SMAC)*, pp. 619-624. IEEE, 2017. https://doi.org/10.1109/I-SMAC.2017.8058253

[32] Ullah, Zahid, Farrukh Saleem, Mona Jamjoom, Bahjat Fakieh, Faris Kateb, Abdullah Marish Ali, and Babar Shah. "Detecting High-Risk Factors and Early Diagnosis of Diabetes Using Machine Learning Methods." *Computational Intelligence and Neuroscience* 2022, no. 1 (2022): 2557795. https://doi.org/10.1155/2022/2557795

[33] Alex, Suja A., J. Jesu Vedha Nayahi, H. Shine, and Vaisshalli Gopirekha. "Deep convolutional neural network for diabetes mellitus prediction." *Neural Computing and Applications* 34, no. 2 (2022): 1319-1327. https://doi.org/10.1007/s00521-021-06431-7

[34] Yahyaoui, Amani, Akhtar Jamil, Jawad Rasheed, and Mirsat Yesiltepe. "A decision support system for diabetes prediction using machine learning and deep learning techniques." In *2019 1st International informatics and software engineering conference (UBMYK)*, pp. 1-4. IEEE, 2019. https://doi.org/10.1109/UBMYK48245.2019.8965556