ISSN PRINT 2319 1775 Online 2320 7876

*Research paper*  UGC CARE Listed ( Group -I) Journal Volume 11,S Iss 3, Dec 2022

# A Novel Multiheaded Convolution Neural Network with Multilevel SVM Architecture for classification of Diabetes

**Sofia Goel[1], Sudhansh Sharma[2]**

*1 School of Computer and Information Sciences (SOCIS), IGNOU, New Delhi, India.
Email: sofiagoel@gmail.com
[2] School of Computer and Information Sciences (SOCIS), IGNOU, New Delhi, India.
sudhansh@ignou.ac.in
Email: sudhansh@ignou.ac.in

**ABSTRACT**

Classification of diabetes is an area of major concern in the field of medical science. Diabetes is not a single disease rather a group of health disorders that is hazardous to human health if not detected on time. Diabetes can be classified under three classes based on blood glucose level(BG): hyperglycemia(BG≥180mg/dL), hypoglycaemia (≤70 mg/dL) and euglycaemia (>70 and <180). BG level is detected through (CGM) techniques which generate time-series data. Various deep learning approaches like Convolution Neural Network have been employed for the classification of diabetes due to their multi-layered architecture and efficient feature extraction. However, the learning capacity of CNN model is limited due to the multidimensional times series nature of the data. Simply, adding more layers to the model makes it more complex and time-consuming and reduces the efficiency of classification. In this study, we have tried to overcome these limitations. This research proposes a hybrid model of Multiheaded CNN and multilevel SVM(MHCNN-3-SVM). In the proposed model, three multiple heads extract the features of each time series independently and the fully connected layer is replaced with multilevel SVM which is a powerful classifier. Also, the proposed model is tuned with different variants of head size and kernel size for performance optimization. UVA/PADOVA dataset of 30 patients (10 adults,10 adolescents,10 children) has been used in this study for multilevel classification. The results of the experiment depict that the proposed hybrid model obtains the highest accuracy of 99.49% and outperformed the state of the art approaches.

**Keywords:** Convolution Neural Network, Diabetes, Deep Learning, Multi-headed, Support VectorMachines.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper  UGC CARE Listed ( Group -I) Journal Volume 11,S Iss 3, Dec 2022

## 1. INTRODUCTION

Type 2 diabetes is a life-threatening condition that happens when the body generates insulin but does not use it properly. Insulin resistance is the term used to describe people who are insulin resistant. Insulin is produced by the pancreas, which aids glucose entry into body cells and maintains appropriate blood sugar levels. When the amount of glucose produced isn't properly utilized, blood sugar levels fluctuate, either too low or too high hypoglycemia or hyperglycemia arises as a result of this. When T2DM is diagnosed, it is common to find that the majority of the serious harm to the body has already occurred. This is a major scientific concern because the disease is a chronic metabolic ailment that can only be managed, not cured [1]. Diabetes type 2 is not simply a medical condition, but it's also a suite of medical disorders. Type 2 diabetes is more than just a metabolic issue; it's a debilitating disease that necessitates glucose and insulin management. To do so, you'll need to keep track of a series of observations in chronological order, such as blood glucose levels, electrocardiograms (ECGs), and so on. Because of their nonstationary and nonlinear characteristics, glucose and insulin dynamics are difficult to capture. It also varies from one person to the next [2]. Therefore, blood glucose forecasting with high accuracy plays a vital role in diabetes management. This can aid in monitoring the correct insulin dose of the patient relieving him from mental, social and economic pressure.

In recent times, efforts have been undertaken towards the expansion of a deep learning model for the detection and forecasting of time series data. Several predictive algorithms, most of which built on machine learning, were created to identify the glycemic trends of persons with diabetes based on the readings acquired by the CGM [3] [4] [5].These algorithms, in particular, employ deep learning models such as CNN and RNN to recognise the structure of CGM and its temporal correlation, and then treat it using time series analysis approaches [6] [7].In addition, research circulated developing hybrid of CNN and LSTM for forecasting BG levels [8] [9].The main objective around all research for using neural networks was featuring longer sequences and feature extraction. For which CNN was found to be highly competitive and even better than LSTM [10] [11].

The above review proves that deep learning models made a remarkable performance in all types of prediction problems. One of the most noteworthy work was proposed by Kezhi Li et al. [12]who significantly performed glucose forecasting by introducing a deep learning framework named Glunet.It was based on the concept of multilayer dilated convolutions and gated activations. Their results for short-term (30-60) minutes forecasting showed improvements over existing approaches. A different approach was performed by by Maxime De Bois et al. [13] and by Eleni I. Georga et al. [14]where they treat the machine learning algorithms from two different perspectives, one for short term prediction of hypoglycemic events and another for long term

for prediction of hyperglycemic events.

Despite their great results, one of the primary issues remains the prediction of sudden changes in blood glucose readings caused by insufficient feature extraction in time series data. Another major challenge is in deep learning models is overfitting. Simply adding multiple layers to the neural network makes the model complicated. Another noticeable factor in diabetes data is there exists a temporal relationship between all samples of different features. Thus to capture all features optimization of hyperparameters such as kernel size is a matter of concern.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper  UGC CARE Listed ( Group -I) Journal Volume 11,S Iss 3, Dec 2022

In this research, a real-time glucose forecasting model centered on time series algorithms to aid in the prevention of diabetic issues is proposed. The proposed work is based on a deep learning architecture of CNN with multiple kernel sizes, multiple convolution layers with multiple head sizes. In this study, we are performing long and short-term forecasting of blood glucose using various deep learning models. The dataset used in silico data UVA/PADOVA of 15 days for 30patients.The subjects fall into three categories namely adults, adolescents and children.

Blood glucose data is highly nonlinear and non-stationary, thus to forecast the data multiheaded CNN model is employed for feature learning and is compared with various deep learning LSTM models [15]. The reason behind the selection of LSTM and CNN for forecasting is LSTM perform sequence to sequence time series forecasting and has proven to perform long term trend analysis and CNN performs auto feature selection, extraction and forecasting all in one model [15] [16]. For each sort of forecasting problem based on time series data, there are enormous LSTM models to choose from. In this study, we have compared vanilla, layered and bidirectional LSTM models of deep learning. LSTM models used for forecasting problems of blood glucose time series data are well suited to discover the long-term dependencies in sequential data due to its potential of internal memory [10].

CNN is a sort of deep neural network applied for two purposes: feature extraction component and classification part. Unlike the traditional model, in this model independent CNNs are used, which are better known as convolution heads, to deal with the prediction of blood glucose levels. Here, data is addressed individually thus avoiding the need for preprocessing of data and delivering a more customised architecture for each type of observation. This architecture is referred to as Multiheaded CNN. It is implemented as a multiheaded model to capture the correlations of blood glucose levels of the past and future. The aim of the paper is to provide the best suited model forecasting model for sequential non-stationary data in two forms short term and long term forecasting.

The remainder of this work is laid out as follows: Section II provides an introduction of the research background, followed by sections III and IV, which discuss the proposed model and data collection. An experiment is included in SectionV. Section VI of this study shows the outcomes of our studies, and section VII concludes the work done and future scope of the research.

## 2. Background

In this section description of various predictions techniques is discussed along with its merits and limitations.

### 2.1 Vanilla LSTM

The most basic LSTM model used for univariate time series data is Vanilla. It has a distinct hidden layer of LSTM units, and an output layer for prediction. This is one layered architecture originated in 1997 for prediction problems and proved to give satisfactory performance for small sequence prediction problems. Vanilla LSTM gained popularity for predicting time series data. Later on, researcher Schmidhuber improved upon the original LSTM and uses full gradient training. Figure 1. shows the outlined view of the vanilla LSTM block along with the cell structure. It features input, forget and output gate. First is the input layer followed by recurrent LSTM layer which consists of memory units. Next is the dense layer followed by LSTM layer which is used for predicting output. The output of one block is recurrently coupled to the input of another block and connects to all of the gates.

ISSN PRINT 2319 1775 Online 2320 7876

 **UGC CARE Listed ( Group -I) Journal Volume 11,S Iss 3, Dec 2022**

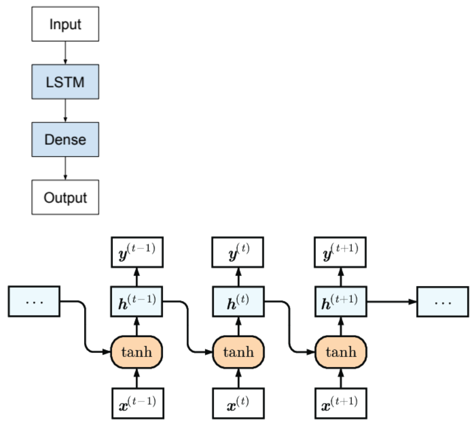


Fig. 1. Structure of Vanilla LSTM

Vanilla LSTM failed to solve complex tasks due to its inability to keep the memory content for a long time and suffered from gradient vanishing problems as well as information morphing issues.

### 2.2 Layered LSTM

The layered LSTM is made up of several hidden layers. Each layer is made up of multiple memory units known as LSTM units. In deep learning models depth of the model plays a very crucial role in making the model more accurate and fast. Thus, layering of LSTM hidden endorsed the success of the deep learning models on a broad range of demanding forecasting problems. It is the architecture comprised of multiple layers where each layer corresponds to the functioning of layer processing in one way or the other and passes it further. In this way it works as a channel where each layer is performing a portion of the job passing it on to the next layer till it provides the output. A Multilayer Perceptron neural network can also have hidden layers added to it to make it more complex. Additional hidden's goal is to recombine the learnt representations from previous layers and produce new representations at high abstraction levels. Originally, Graves, et al. [17] introduced Deep LSTMs in numerous applications such as voice recognition which became state of the art in the area of deep learning. It is one of the most reliable methods for anticipating time series difficulties. A layered LSTM architecture produces a sequence rather than a single value output since it is made up of numerous LSTM layers.

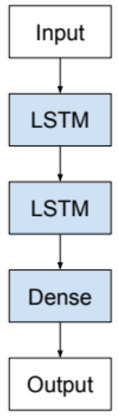


Fig. 2. Structure of Layered LSTM

### 2.3 Bidirectional LSTM

To improve further accuracy in output researchers tried to use the input data in both directions rather than one direction. In Bidirectional LSTMs input time series data is processed in both the forward and backward directions [7]. Duplication of the first recurrent layer occurs resulting in two LSTM units one in the forward direction and the other in a backward direction which goes side by side. As a result, two layers are functioning at the same time, with the first layer receiving the input sequence as n input and the second layer receiving a reversed copy of the input sequence. This approach overwhelms the limitations that exist in old architectures by providing obtainable input information of the past as well as of future of a specific time mount in the whole sequence. This is a very general requirement when using vectorized inputs and were developed for speech recognition, but now it is the widely accepted approach for lifting model performance.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed ( Group -I) Journal Volume 11,S Iss 3, Dec 2022

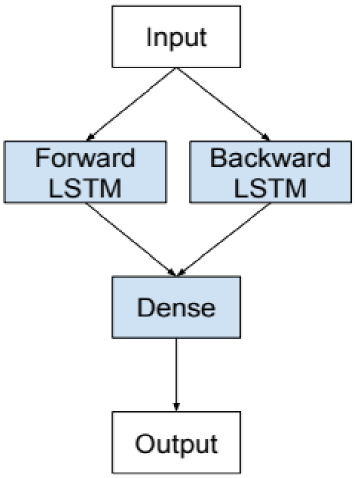


Fig. 3. Structure of Bidirectional LSTM

2.4 Convolution Neural Network(CNN)

CNN is a class of artificial neural network for processing data, based on grid-like topology. 1-D grid can be thought of as time-series data, where samples are taken at regular time intervals and an example of image data can be taken as 2 D grid of pixels [18]. The basic structure of the CNN model comprises of convolution layer, pooling layer with input data and a fully connected network [19]. Convolution networks employ a special kind of mathematical operation called convolution in place of the general matrix in any of the above-mentioned layers. During the last decade, CNN is applied extensively for image processing 2-D Data where it works on multiple hidden layers and millions of parameters to recognize complex patterns on massive datasets with numerous training. But, the problem arises when the data is application-specific or where data is scarce.

CNN fuses the property of extracting features automatically and classifying it into one single learning body and can be trained to optimize the features directly from the raw input during the training phase. Due to its aforementioned properties, it has successfully employed to resolve such issues [20]. The advantage of using CNNs to classify sequences is that they can learn directly from raw time series data. It eliminates the requirement of engineering input features manually. The deep learning based model would be able to exploit time series sequential data and, in turn shows analogous performance to network models trained on a dataset with manual features.

Here are a few baseline reasons to prove that CNN could outperform LSTM in accuracy and performance both with very fast execution time. Firstly, RNN based models use primarily sequential processing over a period of time which results in the problem of vanishing gradients [3]. To resolve this issue LSTM comes with forget gates but still, the architecture is more complicated and the sequential process is persisting. Secondly, RNN is not hardware friendly. Running the model on the cloud and for training purposes requires a lot of resources [21]. Also, RNN and LSTM have memory bandwidth problems and on large parameters, it becomes impossible to get enough external bandwidth for moving parameters back and forth. Recently research work points out that CNN can achieve what LSTM has been used for, which means predicting sequences but in a much faster and more computationally efficient way.

Recently, 1D-CNN has achieved a state-of-art performance in the analysis of time series data like in health monitoring, early diagnosis and detection areas. 1D- CNN is also applied for the analysis of signal data over a fixed length period like in audio recording and in NLP. The principle of simple mathematical convolutions makes it a low cost feasible and real-time method for analysis of time series data [22].

2.4.1 Working and structure of 1 D CNN

In Conv1D, **kernel** slides along one dimension, time series data is one such example where data requires kernel sliding in one direction [11].

ISSN PRINT 2319 1775 Online 2320 7876

Research paper  UGC CARE Listed ( Group -I) Journal Volume 11,S Iss 3, Dec 2022

- Input layer: It has N k neurons, with k indicating the number of input time series and N indicating the length of each univariate series.
- Convolutional layer: is a foremost component of the CNN architecture that accomplishes feature extraction using a convolution operation and an activation function. In convolution operation an array of numbers in form of a matrix is passed linearly to the data which is known as kernel or filter. At each point of the input matrix, an element wise product of each input data and element of the kernel is calculated and added to give the output value at the corresponding place of the input, which is known as a feature map. This process is repeatedly applied by multiple kernels to generate n number of feature maps. Here convolution operation is mainly defined by two hyper parameters: the size of the input data matrix and the total of kernels.
- Pooling layer: Pooling layers simplify the information collected by the convolutional layer by decreasing the feature maps dimensions and the count of consequent learnable parameters. It creates a reduced version of the variability obtained by the feature map. Although several ways are at present to condense the information, max-pooling is the most common approach, which as a value keeps the maximum value of those that were in the nxn input window and discards all other values.
- Feature layer: Once convolution and pooling is done the role of feature layer comes where long time series sequence is generated as feature maps. These feature maps represent final form of input preparing it for final output.
- Output layer: Final down sampling is accomplished using global average pooling before being applied to a fully connected layer. Here, the feature maps generated by feature layer are turned into one dimensional array of integers connected to one or more fully connected layers. Finally, the output nodes are generated which is of the same number as of classes of time series data.

## 3 Proposed model

### 3.1 Multiheaded CNN(MHCNN)

CNN is generally used for image processing and has become a cutting edge in this arena. We know in case of image processing 2-D convolutions are used where the kernel travels in all directions, while times series data is one dimensional thus need 1-D convolution with a single channel. Multiheaded convolution is a one-dimensional convolution neural network in which each input sequence is processed by a completely independent convolution referred to as convolution heads. Here, the role of multiple heads is to extract the features of each input sequence independently. As a result, for each input sequence, an independent feature map is obtained. In this model, there are two heads where each head undergoes convolutional processes for its corresponding input sequences. As a result, it generates a feature map for each head consist of time series data. This generates a sequence of feature maps for each time series independently. These sequences are then concatenated together because they were obtained independently of one another.

For time-series predictions, deep learning is

ISSN PRINT 2319 1775 Online 2320 7876

Research paper  UGC CARE Listed ( Group -I) Journal Volume 11,S Iss 3, Dec 2022

progressively being employed in the healthcare field. The creation of multi-headed neural network architectures for multivariate time-series forecasting is gaining popularity in research. This is due to the unique topology of multiheaded neural network where each input series comprises of independent variables can be handled by a distinct head. In multi-headed neural network topologies, the output of each of these heads can be merged before a prediction is produced. In this paper, two multi-headed ML architectures is used to predict patient's blood glucose on a quarterly basis.

Another important reason to use MHCNN is when RNN gradient vanishing is a significant issue [20]. In RNN, the weights shift and eventually become so little that they have no effect on the output. As a result, the network's ability to learn from the past deteriorates, as does its operational competence for analysing extended data sequences for predictions [23]. MHCNN is used for forecasting problems because it has convolution layers for feature extraction and pattern recognition, which results in quick prediction.

The MHCNN architecture used in this paper has multiple convolutional layers, each followed by a pooling layer. Each convolutional layer has a headsize of two with an independent input sequence and filter size of three. To increase the accuracy of recognition of blood glucose levels recurrent sequential processing of input features is done. These input sequences are passed through two CNN models independently and finally fused to predict the output. The suggested method uses two CNNs, each with a different configuration. One has a kernel size of 3 and the other has a kernel size of 5.

Table 1. Data values of the proposed model.

| Variable | Description[a] |
|---|---|
| *Input sequence* | 50 |
| *Output sequence* | 10(for 30 min) and 5 (for 15 min) |
| *No of features* | 1 |
| *Test size* | 0.3 |
| *Number of Epochs* | 200 |
| *Approach* | 'LayeredLSTM'/Vanilla LSTM/Bi-directional LSTM,CNN/MHCNN |
| *No of hidden units* | 50 |
| *Filter size* | 64 |
| *Pool size* | 2 |
| *Kernel size* | [3,5] |
| *Activation* | 'elu' |
| *Optimizer* | 'adam' |
| *Loss* | 'mse' |
| *Batch size* | 256 |

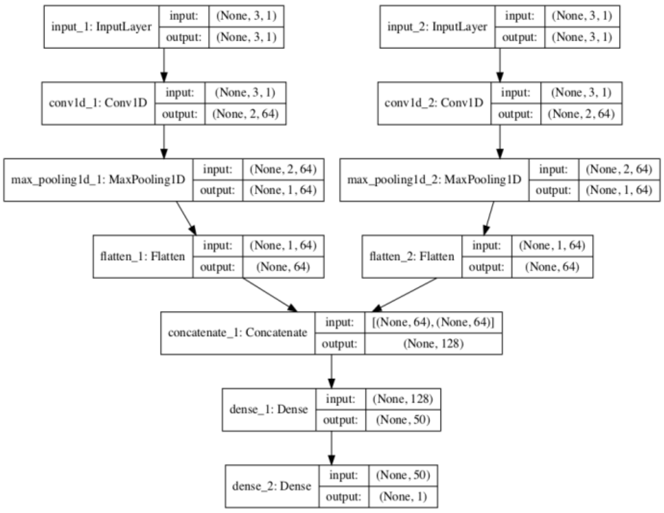


Fig. 4. Structure of MHCNN

## 4 Data collection

For the experiment, UVA/Padova dataset is available openly on the web. Also, the reading was taken in the time interval of seven days, fifteen days and thirty days basis of one patient. Blood glucose measurements were done four times a day, corresponding to the breakfast, lunch, evening snack and dinner. CGM sensors measure the level of glucose found in the fluid between the cells which is known as the interstitial glucose level. The CGM sensors tests glucose every few minutes. Basal insulin is crucial for

ISSN PRINT 2319 1775 Online 2320 7876

Research paper  UGC CARE Listed ( Group -I) Journal Volume 11,S Iss 3, Dec 2022

keeping these blood glucose levels under control insulin pump glucose readings, as well as glucose level trends over time, are visible on a built-in device screen. UVA-PADOVA dataset includes 30 silico subjects of three different categories namely adults, adolescent and children. Each silico subject was represented by its CGM, which was extracted in a sequential repetitive pattern as an input for the model. Third Level Heading

## 5 Experiment

In this research simulated UVA-PADOVA silico dataset is used for training and testing. It includes a population of 30 silico subjects of three different age groups namely adults (>13 years), adolescents (age group 13-18 years) and children (2-12 years). Data is collected to forecast the blood glucose level of 5 minutes and 10 minutes based on continuous glucose monitoring done for 15 minutes and 30 minutes respectively. The experiment is conducted into two stages: Input stage and Forecasting stage.

### 5.1 Input Stage

The following steps are employed to generate the readings of blood glucose on variable meals and different calories.

Steps to start the process of data collection (blood glucose levels):

- Simulation time is accepted in hours for 360 hrs i.e.15 days.
- Among two scenarios random and custom, a custom scenario is selected.
- The simulation started for three meals covering the amount of calorie intake for breakfast, lunch and dinner.
- The process of the simulation was performed for three categories of patients namely adolescent, adult and child.
- To collect appropriate blood glucose readings CGM sensors of three different types were taken: Dexcom, Guardian RT and Navigator which measured interstitial glucose levels for every few minutes. In addition, the Basal Bolus controller was used for keeping these blood glucose levels under control.
- Insulin pump glucose of two types cozmo and insulet were used to read glucose level trends over time and were visible on a built-in device screen.
- Finally, readings were saved in csv format to forecast the blood glucose data using machine learning algorithms.

### 5.2 Forecasting stage

Once the results are obtained of blood glucose levels of patients then it was used to forecast the data on different deep learning models. The different LSTM deep learning models are applied for testing and training like Vanilla, Layered and Bidirectional along with CNN and Multiheaded (multiheaded CNN with varied head size). The parameters like filters, pool_size, and kernel_size are associated with the Convolutional Neural Networks models only. Multiheaded CNN model with headsize=2 is used in the experiment. The models can be used for both types of forecasting (short-term and long-term). Short-term forecasting means predicting the blood glucose level for the next 15 mins, while the long-term forecasting means predicting the blood glucose level for the next 30 minutes.In the dictionary pf parameters there were two variables named input_sequence and output_sequence. The time interval taken between the two samples is 3 minutes. Therefore, if the input_sequence is 50 and the output sequence is 10, it means 150 minutes (1.5 hours) of input data is used to forecast the next 15 minutes of blood glucose level.

The performance of each model is evaluated in terms of MSE, RMSE,MAE and MAPE.The quality of prediction is also evaluated by calculating the R_square



IJFANS International Journal of Food And Nutritional Sciences Official Publication of International Association of Food and Nutrition Scientists

ISSN PRINT 2319 1775 Online 2320 7876
Research paper  UGC CARE Listed ( Group -I) Journal Volume 11,S Iss 3, Dec 2022

value. R
-square is the coefficient of determination that determines the quality of fitness among the actual value and the forecasted value.

It is observed in the experiment performed that accuracy achieved in CNN is many times faster than LSTM and performance is also very high. In case of multiheaded kernel size is taken as 3 or 5 or 7. This is because it is diabetes data and here temporal relationship exists between two instances. Thus to capture relationship we are having different combination of kernel sizes. If three is the kernel size it means three instances will be feed at once, thus there are very less or no chances of missing features.

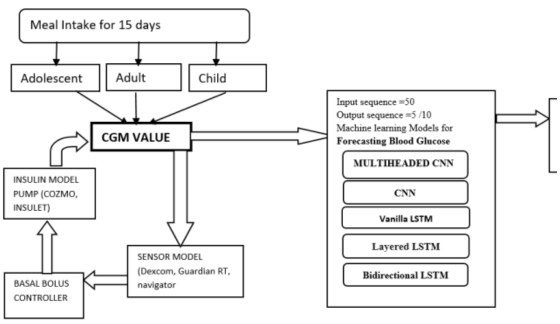


Figure 5: Diagrammatic view of the experiment conducted for forecasting BG using different models of LSTM, CNN and MHCNN

5.3 Performance evaluation

- Mean Squared Error (MSE) reflects the average of the squared difference between the original and projected values. It calculates the residuals' variance in the dataset.

$$MSE = \frac{1}{N}\sum_{i=1}^{n}( y_i - y^\wedge )^2$$

Where $y^\wedge$ is the predicted value

- Root Mean Squared Error(RMSE) is equal to the square root of the Mean Squared Error. It calculates the residuals' standard deviation.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{n}( y_i - y^\wedge )^2} \qquad (2)$$

- The Mean absolute error (MAE) reflects the average absolute difference between the dataset's actual and predicted values. It calculates the average of the dataset's residuals.

$$\frac{1}{N}\sum_{i=0}^{n} | y_i - y^\wedge | \qquad (3)$$

- The Mean Absolute Percentage Error (MAPE) is the average of the absolute difference between the actual and predicted values in the dataset. It gives average of the percentage errors.

$$M = \frac{1}{N}\sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \qquad (4)$$

Where M =MAPE, N = Total number of times summation done, At=Actual value, Ft=Forecast value

- R squared error, is the ratio of the change in the prediction of dependent variable from the independent variable(s).

$$R^2 = \frac{SSR}{SST} = \sum \frac{(y^\wedge - \bar{y} )^2}{( y_i - \bar{y} )^2} \qquad (5)$$

Where SSR= sum of squares of residuals and SST = total sum of squares.

## 6 Results and Discussion

In this section, we evaluate the performance of the proposed MHCNN model for forecasting blood glucose levels on UVA/PADOVA silico dataset. The 15 days of data is taken which comprises of three categories of instances (10 adults,10 adolescents and 10 children). The entire data is divided into two parts 70% for testing and

ISSN PRINT 2319 1775 Online 2320 7876
Research paper  UGC CARE Listed ( Group -I) Journal Volume 11,S Iss 3, Dec 2022

30% for training. The performance of the model could be analyzed from two perspectives. Firstly, based on error rate and secondly on the execution time. Five types of errors RMSE, MSE, MAE, MAPE, Square error are examined. The results of the experiment are presented in six tables. The first three tables show performance measures for predicting 15 min of data and the next three tables shows 30 min of forecasting. All the models were trained for 200 epochs with optimizer ADAM with a kernel size of 3 and 5. Moreover, to avoid dropping of features during convolution operations, the recursive sequential pattern extraction is applied.

Tables 2,3 and 4 depicts the performance of the model for short term forecasting of blood glucose levels (15 min) in the case of adults, adolescents and children respectively. Here the length of the input sequence is taken as 50 and the length of the output sequence is 5. The time interval of BGM is three minutes. If we compare the error rate of MHCNN with LSTM models (Vanilla, Layered and Bidirectional) and CNN, results depicts that the proposed model is outperforming in all three cases. More specifically the value of errors exhibited in case of adults are: MSE 5.14±2.85, RMSE = 2.26±2.55, MAE =1.76±2.37, MAPE =1.25±2.38 and R_Square =0.99±0.14. Similarly, the value of error is found to be least for the MHCNN model in the case of adolescents as well as for the child. When we look at the execution time of all the models it is pertinent that the MHCNN model is significantly many times faster than all other models.

Table 2. Performance measure of Adults class of UVA/PADOVA sample dataset.

For Ph=15 min (Input sequence =50 and output sequence =5)

**Ph=15min Adults**

| Approach | **MSE** | **RMSE** | **MAE** | **MAPE** | **R_square** | **ExecutionTime** |
|---|---|---|---|---|---|---|
| **LayeredLSTM** | 573.52±2.35 | 23.95±2.02 | 7.51±2.89 | 5.23±2.98 | 0.92±1.33 | 46m14s |
| **Vanilla LSTM** | 329.42±2.56 | 18.15±2.78 | 6.57±2.23 | 4.23±2.10 | 0.98±1.23 | 39m19s |
| **Bi-Directional LSTM** | 118.96±2.23 | 10.90±3.32 | 9.89±2.27 | 7.43±2.15 | 0.91±1.34 | 40m10s |
| **CNN Headsize=1** | 11.83±2.19 | 3.44±2.15 | 3.05±2.35 | 2.15±2.39 | 0.98±0.98 | 5m26s |
| **MHCNN Headsize=2** | **5.14±2.85** | **2.26±2.55** | **1.76±2.37** | **1.25±2.38** | **0.99±0.14** | **4m19s** |

ISSN PRINT 2319 1775 Online 2320 7876

*Research paper* © 2012 IJFANS. All Rights Reserved, UGC CARE Listed ( Group -I) Journal Volume 11,S Iss 3, Dec 2022

Table 3.Performance measure of Adolescent class of UVA/PADOVAsample dataset. For Ph=15 min (Input sequence =50 and output sequence =5)

**Ph=15min Adolescents**

| Approach | MSE | RMSE | MAE | MAPE | R_square | Time |
|---|---|---|---|---|---|---|
| **LayeredLSTM** | 792.42±2.35 | 28.15±2.45 | 8.51±2.87 | 6.23±2.12 | 0.92±1.98 | 29m67s |
| **Vanilla LSTM** | 535.42±2.98 | 23.15±2.78 | 6.57±2.23 | 4.23±2.10 | 0.98±1.09 | 34m11s |
| **Bi-Directional LSTM** | 240.56±2.35 | **15.51**±3.32 | 8.77±2.27 | 6.56±2.15 | 0.93±1.28 | 30m18s |
| **CNN Headsize=1** | 19.73±2.79 | 4.44±2.15 | 6.05±2.78 | 4.15±2.31 | 0.98±1.17 | 5m2s |
| **MHCNN Headsize=2** | 11.76±2.83 | 3.43±2.58 | 2.74±2.49 | 1.78±2.38 | 0.91±0.19 | 4m2s |

Table 4. Performance measure of Child class of UVA/PADOVA sample dataset.

For Ph=15 min (Input sequence =50 and output sequence =5)

**Ph=15min Child**

| Approach | MSE | RMSE | MAE | MAPE | R_square | Time |
|---|---|---|---|---|---|---|
| **LayeredLSTM** | 473.06±2.89 | 21.75±2.67 | 11.23±2.79 | 5.23±2.98 | 0.92±1.33 | 32m24s |
| **Vanilla LSTM** | 329.42±2.98 | 18.15±2.78 | 6.57±2.23 | 4.23±2.10 | 0.98±1.09 | 27m18s |
| **Bi-Directional LSTM** | 166.66±2.35 | **12.91**±3.32 | 8.37±2.21 | 4.39±2.15 | 0.93±1.28 | 29m13s |
| **CNN Headsize=1** | 10.83±2.34 | 3.14±2.15 | 3.75±2.15 | 2.15±2.69 | 0.98±1.11 | 4m56s |
| **MHCNN Headsize=2** | 8.35±2.56 | 2.89±2.23 | 1.46±2.38 | 1.55±2.18 | 0.95±0.34 | 4m8s |

Tables 5,6 and 7 present the error rate for long term forecasting (30 min of prediction). The length of the input sequence remains the same as 50 while the output sequence is taken as 10. In other words, 150 min of data is predicting the BG level of the next 30min. From these tables, it can be seen that the proposed model produced lower prediction error values for all three subjects as compared to the results predicted by the LSTM models. The MSE difference in the adults using the MHCNN method was only 3.87 while the value of MSE in the case of CNN was 5.03 followed by Bidirectional LSTM as 15.91 then Vanilla LSTM with 22.5 and layered LSTM with 39.5. In addition, execution time was also 3m48s in the case of MHCNN which was gradually increased many times in other models. This proves that the MHCNN demonstrates better predictions shown by the lesser RMSE values and faster execution time.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper  UGC CARE Listed ( Group -I) Journal Volume 11,S Iss 3, Dec 2022

Table 5. Performance measure of Adult class of UVA/PADOVA sample dataset.

For Ph=30 min (Input sequence =50 and output sequence =10)

Ph=30min Adult

| Approach | MSE | RMSE | MAE | MAPE | R_square | Time |
|---|---|---|---|---|---|---|
| **LayeredLSTM** | 1562.06±2.19 | 39.53±2.19 | 27.51±2.19 | 15.23±2.19 | 0.82±2.19 | 1hr14m |
| **Vanilla LSTM** | 507.15±2.19 | 22.52±2.19 | 9.55±2.19 | 8.16±2.19 | 0.99±0.59 | 1hr23m |
| **Bi-Directional LSTM** | 238.57±2.19 | 15.91±2.19 | 10.97±2.19 | 6.69±2.19 | 0.83±0.39 | 56m10s |
| **CNN Headsize=1** | 25.31±2.19 | 5.03±2.19 | 2.89±2.19 | 1.92±2.19 | 0.98±0.45 | 4m26s |
| **MHCNN Headsize=2** | 15.00±2.12 | 3.87±2.19 | 2.35±2.19 | 1.65±2.19 | 0.9893±0.19 | 3m48s |

Table VI. Performance measure of Adolescent class of UVA/PADOVA sample dataset.

For Ph=30 min (Input sequence =50 and output sequence =10)

**Ph=30min Adolescent**

| Approach | MSE | RMSE | MAE | MAPE | R_square | Time |
|---|---|---|---|---|---|---|
| **LayeredLSTM** | 683.82±2.68 | 26.15±2.78 | 11.12±2.34 | 9.86±2.19 | 0.96±1.19 | 1hr24m |
| **Vanilla LSTM** | 240.87±3.23 | 15.52±2.19 | 9.58±2.78 | 7.98±2.23 | 0.99±1.12 | 1hr13m |
| **Bi-Directional LSTM** | 166.66±2.56 | 12.91±2.19 | 10.17±2.89 | 6.62±1.23 | 0.93±1.26 | 56m40s |
| **CNN Headsize=1** | 121.66±2.49 | 11.03±2.19 | 2.89±2.39 | 1.92±1.45 | 0.98±0.19 | 4m27s |
| **MHCNN Headsize=2** | 59.14±2.19 | 7.69±2.12 | 2.53±2.10 | 1.35±1.67 | 0.99±0.15 | 3m44s |

Table VII. Performance measure of Child class of UVA/PADOVA sample dataset.

For Ph=30 min (Input sequence =50 and output sequence =10)

Ph=30min Child

| Approach | MSE | RMSE | MAE | MAPE | R_square | Time |
|---|---|---|---|---|---|---|
| **LayeredLSTM** | 270.60±2.19 | 16.45±2.19 | 9.82 ±2.67 | 8.86±2.19 | 0.96±2.19 | 1hr29m |
| **Vanilla LSTM** | 156.25±2.19 | 12.50±2.19 | 9.55±3.45 | 6.16±2.89 | 0.99±2.19 | 1hr13m |
| **Bi-Directional LSTM** | 253.12±2.10 | 15.91±2.19 | 9.37±2.67 | 4.79±2.23 | 0.99±2.02 | 1hr10m |

ISSN PRINT 2319 1775 Online 2320 7876
Research paper  UGC CARE Listed ( Group -I) Journal Volume 11,S Iss 3, Dec 2022

| | | | | | | |
|---|---|---|---|---|---|---|
| **CNN Headsize=1** | 81.54±1.79 | 9.03±2.19 | 2.29±2.19 | 2.92±2.34 | 0.98±1.34 | 4m16s |
| **MHCNN Headsize=2** | 31.36±2.19 | 5.60±2.19 | 3.09±2.19 | 3.06±2.19 | 0.99±0.89 | 4m24s |

Also, if we observe the performances of the model for short term and long term forecasting for all three cases, it shows that the performance is highest either in the case of children or adults and performance is minimum in the case of adolescents. This is because the glucose level of adolescents has high variability factors such as unstable food intake, mood swinging, irregular exercises etc. which makes it difficult to regularize.
Examining the results of all tables it is also noted that all the models tested in this work achieved noticeably improved performance for all the age groups as compared to other models. Furthermore, unlike previous work done, the proposed model predicts the blood glucose level with high accuracy and in less time making it a promising model for time series forecasting problems.

## 7 Conclusion

MHCNN is developed as a valuable solution for prediction of blood glucose level. A multi-headed 1D CNN is followed by a multilayered structure, with the CNN capturing the characteristics or patterns of the multi-dimensional time series. The improved CNN can process previous sequential data and predict the blood glucose level. Using that time series data, the proposed model performs pattern formation for each diabetes subject. The pattern obtained exhibits into a trained deep neural network which could be used wearable devices in future. In the silico dataset for short and long term forecasting of blood glucose levels, the suggested MHCNN approach outperformed the existing neural networks. It significantly proved that MHCNN is better in terms of accuracy and execution time as compared to LSTM models.

Future improvements could be addressed regarding the optimization of hyper parameters such as kernel size, head sizes, pooling layers with more combinations. The model could be improved concerning the different meals intake in various conditions. It would be interesting to gather data from patients for more days. In addition, the model could be implemented on smartphone devices and wearable sensors.

## References

[1] G.paracino, F.anderigo and S.Corazza, "Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor

ISSN PRINT 2319 1775 Online 2320 7876

Research paper  UGC CARE Listed ( Group -I) Journal Volume 11,S Iss 3, Dec 2022

time-series," *IEEE Transactions*, pp. 931-937, 2007.

[2] T.Zhu, L. Kuang, K. Li, J. Zeng and P.Herrero, "Blood Glucose Prediction in Type 1 Diabetes Using Deep Learning on the Edge," *IEEE International Symposium on Circuits and Systems (ISCAS),*, pp. 1-5, 2021.

[3] M.Canizoa, "Multi-head CNN–RNN for multi-time series anomaly detection: An industrial case study," *Neurocomputing*, pp. 246-260, 2019.

[4] A. Z.Woldaregay, E.Årsand and S.Walderhaug, "Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes.," *Artificial intelligence in medicine*, no. 3, pp. 109-134, 2019.

[5] H.I.Fawaz, G.Forestier and J.Weber, "Deep learning for time series classification: a review," *Data mining and knowledge discovery*, pp. 917-963, 2019.

[6] K.Shankar, Y. Zhang, Y. Liu, L. Wu and C.-H. Chen, "Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification," *IEEE Access*, vol. 8, pp. 118164-118173, 2020.

[7] M. Schuster and K. Paliwal, "Bidirectional recurrent neural networks," *IEEE transactions on Signal processing*, pp. 2673-2681, 1997.

[8] Y. Heryadi and H.Warnars, "Learning temporal representation of transaction amount for fraudulent transaction recognition using CNN," *IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*, 2017.

[9] K.Kaya and S.Gündüz, "Deep Flexible Sequential (DFS) Model for Air Pollution Forecasting," *NatureResearch*, pp. 1-12, 2020.

[10] N.Xue, I.Triguero and G.P.Figueredo, "Evolving deep CNN-LSTMs for inventory time series prediction," *IEEE Congress on Evolutionary Computation (CEC)*, pp. 1517-1524, 2019.

[11] X.Zheng and Q. Huang, "Deep Sentiment Representation Based on CNN and LSTM," *International Conference on Green Informatics*, 2017.

[12] K. Li, J. Daniels, C. Liu, P. Herrero and P. Georgiou, "Convolutional Recurrent Neural Networksfor Glucose Prediction," *IEEE journal of biomedical and health informatics,*, vol. 24, pp. 603-613, 2019.

[13] M. D. Bois, A. Mounîm, E.Yacoubi and M.Ammi, "Study of Short-Term Personalized Glucose Predictive Models on Type-1 Diabetic Children," in *International Joint Conference on Neural Networks (IJCNN)*, 2019.

[14] E. Georga, V. Protopappas and D.Ardigò, "Multivariate Prediction of Subcutaneous Glucose Concentration in Type 1 Diabetes Patients Based on Support Vector Regression," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 1, Jan 2013.

[15] P.Chandana, G. S. P. Ghantasala, J. Rethna, K.Sekaran and Y.Nam, "An effective identification of crop diseases using faster region based convolutional neural network and expert systems.," *ScholarWorks*, 2020.

[16] P.Rizwan, G. Ghantasala, R.Sekaran and a. M. D.Gupta, "Smart healthcare and quality of service in IoT using grey filter convolutional based cyber physical system.," *Sustainable Cities and Society*, 2020.

[17] A.Graves and J.Schmidhuber, "Framewise phoneme classification

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed ( Group -I) Journal Volume 11,S Iss 3, Dec 2022

with bidirectional LSTM and other neural network architectures," *Neural networks*, pp. 602-610, 2005.

[18] S. Qummar, F. G. Khan, S. Shah, A. Khan and S. Shamshirband, "A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection," *IEEE Access*, vol. 7, pp. 150530 - 150539, 2019.

[19] E.Sreehari and G. Ghantasala, "Climate Changes Prediction Using Simple Linear Regression," *ournal of Computational and Theoretical Nanoscience* , vol. 16, 2019.

[20] Y. Bengio, P. Simard and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks*, pp. 157-166, 1994.

[21] M. Goyal, N. D. Reeves, A. K. Davison, S. Rajbhandari, J. Spragg and M. H. Yap, "DFUNet: Convolutional Neural Networks forDiabetic Foot Ulcer Classification," *IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE*, vol. 4(5), pp. 728-739, 2018.

[22] L. S. and S. M., "A Novel 1-D Convolution Neural Network With SVM Architecture for Real-Time Detection Applications," *IEEE sensors journal*, vol. 18(2), pp. 724-731, 2017.

[23] X. Wang, Y. Lu, Y. Wang and W.-B. Chen, "Diabetic retinopathy stage classification using convolutional neural networks," in *IEEE International Conference on Information Reuse and Integration (IRI)*, 2018.

***