Deep Hybrid Parallel CNN-LSTM Model for Diabetes Prediction Using Fusion of Features

Iqra Naveed[1*], Muhammad Kaleem[1]

1. Department of Electrical Engineering,
University of Management and Technology, Lahore 54770, Pakistan

Abstract-Diabetes is a persistent disease that has high prevalence and causes various complications. Using EHR to identify the individuals with the high risk of developing diabetes could improve quality of health care. This paper represents a hybrid deep model and unique engineered features to forecast diabetes. Hybrid deep model merge Long-Short Term Memory (LSTM) with Convolutional Neural Networks (CNN) and also utilizes the relative statistical data. Unlike the traditional stacked CNN-LSTM, CNN and LSTM in the proposed model extract the features in parallel and are more informative. CNN uses three filters to extract temporal features at different level using convolution technology, whereas LSTM extract the sequential relation using past visit records. These extracted features are merged with corresponding statistical data. Overall the resultant feature is the combination of multi-type (statistic and time domain) and multi-level. Moreover, the model performance is tested for two training density. Result represents the state of art performance of the proposed model for both the training samples.

Keywords- Diabetes, EHR, Hybrid Deep Model, Long-Short Term Memory (LSTM), Convolutional Neural Networks (CNN), stacked CNN-LSTM, multi-type, multi-level.

1. INTRODUCTION

Diabetes is a long term chronic disease that occurs either if the pancreas did not produce the insulin or if a body can't utilize the insulin properly, this results in increase in the sugar level in the blood [1-2]. There are three types of diabetes: Type 1 diabetes, Type 2 diabetes, and Gestational diabetes . Type 1 diabetes (insulin dependent diabetes) occurs when the pancreas fails to produce insulin [3]. Type 2 diabetes (insulin independent diabetes) occurs when the body is in-capable of utilizing the insulin [4]. According to World Health Organization (WHO) number of diabetic patients has been increased from 108 million to 422 from 1980 to 2014. There was 5% increase in death rate from diabetes during 2000 to 2016 [5]. Diabetes was the ninth major cause of death in 2019 that cause 1.5 million of death [6]. Undiagnosed diabetes could cause of amputation, death, heart stroke, kidney failure and eye disease [7]. Therefore early prediction and treatment of diabetes using electronic health record EHR is important. EHR are the electronic record that stores the medical history of millions of patients including (demographics, lab results, clinical observations, diagnosis, medical notes, and medications) [8]. Use of EHR together with artificial intelligence has made it possible to forecast the onset of disease [9-15]. Mani et al [9] implement widely used machine learning algorithms (Gaussian Naïve Bayes, Logistic Regression, CART, Random Forest and SVM) on the electronic record of Vanderbilt University Medical Center (VUMC) to predict the Type 2 diabetes. Results represents random forest (RF) has the highest AUC. Choi et al [10] extract data from Korean National Health and Nutrition Examination Survey (KNHANES) and employ two model (SVM and ANN) to screen the individuals for pre-diabetes. Authors compare their model with the logistic regression. Result shows SVM has highest AUC of 0.731. Li et al [11] gather the record of 3406 individuals from Urumqi, Xinjiang and contrast extensively used prediction models Decision Tree, SVM, Adaboost, and Bagging to predict diabetes. Results shows Adaboost has highest model performance with G-mean. Sousa et al [12], utilize 327 million record of 7000,000 individual extracted from Brazilian health plan provider. Authors implement self-attention LSTM and traditional LSTM predict the individual with high likelihood of developing diabetes. Results indicates, self-attention LSTM provide better performance with AUC of 0.83. Suvarnamukhi et al [13], implement Extreme Learning Machine (ELM) algorithm on PIMA Indian Diabetes data set and compare with CNN and ensemble model of (J48 decision tree and Naïve bayes). Results indicate, the proposed model has highest prediction accuracy. Chatterjee et al [14] extracted retinal images dataset from IDRiD (India). Data set contains total 516 images. Authors implement a novel ForeseeGAN model to predict Diabetic retinopathy. Results of this work represents proposed model obtained the highest accuracy than other machine learning models. Manoharan et al [15] extract data from National institute of diabetic and digestive and kidney disease. Data set contain the record of 768 individuals. Authors uses SMOOT to handle class imbalance in the data and implement XG boost to predict diabetes predict diabetes mellitus. Results shows XG boost provide good prediction than other widely used machine learning algorithms. In this research, a hybrid deep model and unique engineered features are proposed to forecast diabetes. Proposed model utilize two inputs to extract features from different domains of data. One input is the visit history of each patient and other input contains the corresponding statistics of each patient. Different from stack CNN-LSTM this research uses parallel CNN-LSTM structure. First, CNN extracts temporal features from the first input. Then LSTM extract sequential information from the visit history of a patient. Finally these extracted features are merged with the statistical input to forecast the diabetes. More over model performance is accessed for different training samples.

2. BACK GROUND AND FUNDAMENTALS OF PROPOSED MODEL

2.1 CNN

CNN contains feed-forward network. It uses various filters that convolved and pooled the input data to extract the hidden information present in raw data. CNN has sparse connections that enhance speed of convergennce, weight sharing across the network and pooling operation avoid over-fitting and reduces the dimension of data [16]. CNN can extract the information from 1D sequence, 2D images and 3D videos. In this research CNN is applied on a 1D sequence.

Convolution and pooling is the main part of CNN. Convolution is used to extract the features using various filters. Convolution operation is represented in Equation 1.

$$x_j^l = f\left(\sum_{i \in Mj} x_i^{l-1} \times k_{ij}^l + B_j^l\right) \ldots. \qquad (1)$$

Where x_j^l are the feature maps, $x_i^{l-1} \times k_{ij}^l$ is the convolutional operation between the output of x_i^{l-1} and filters k_{ij}^l , B_j^l is the bias

Next pooling layer is also an important part that is used to reduce the dimensions of raw data and also secure the invariance. So features obtained from CNN are stable especially when the raw data is noisy. CNN has three type of pooling operation: maximum pooling operation, minimum pooling operation and average pooling operation. In this research we have used maximum pooling operation which is represented in Equation 2.

$$p_j^l = max\left(q_j^{l-1}(t)\right), t \in [(j-1)w, jw] \qquad (2)$$

2.2 LSTM

Traditional feed forward networks cannot remember time series data the long term dependencies, thus unable to extract the information from the data that has long term dependencies [17]. LSTM were introduced to overcome this problem. LSTM consist of feedback network, it uses gates as memory element to remember long term sequence data as represented in Figure 1. Main part of LSTM is the top most horizontal line that acts like a conveyor belt. LSTM uses three gates to add and delete the information. Equation of LSTM gates: forget gate, input gate, output gate are represented in Equation (3) - (8).

$$\boldsymbol{f_t} = \boldsymbol{\sigma}(\boldsymbol{W_f x_t} + \boldsymbol{U_f h_{t-1}} + \boldsymbol{b_f}) \qquad (3)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \qquad (4)$$

$$\boldsymbol{\tilde{c}_t} = \boldsymbol{tanh}(\boldsymbol{W_c x_t} + \boldsymbol{U_c h_{t-1}} + \boldsymbol{b_c}) \qquad (5)$$

$$\boldsymbol{c_t} = \boldsymbol{f_t} \odot \boldsymbol{c_{t-1}} + \boldsymbol{i_t} \odot \boldsymbol{\tilde{c}_t} \qquad (6)$$

$$\boldsymbol{o_t} = \boldsymbol{\sigma}(\boldsymbol{W_o x_t} + \boldsymbol{U_o h_{t-1}} + \boldsymbol{b_o}) \qquad (7)$$

$$\boldsymbol{h_t} = \boldsymbol{o_t} \odot \boldsymbol{tanh}\,(\boldsymbol{c_t}) \qquad (8)$$

Where c_t is the cell state or memory unit that add $i_t \odot \tilde{c}_t$ or remove the $f_t \odot c_{t-1}$ the information. W_f ,W_i, W_c , W_o , b_f , b_i , b_o are the weights and bias of the gates. σ is the sigmoid activation function. x_t is the input and h_{t-1} is the value of last state.

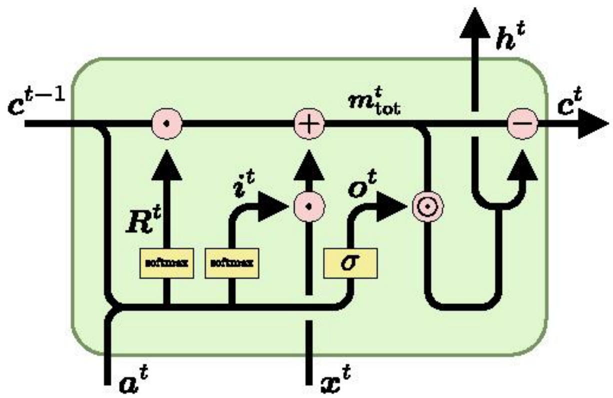

Figure 1: Structure of LSTM

2.3 STATISTICAL COMPONENTS

Statistics is used to examine the data in statistical domain. This research uses statistical information as the second input in the deep model to extract hidden patterns from raw data. Each patient has corresponding statistical components (maximum, minimum, mean, standard deviation, skewness, kurtosis) represented in Equation (9) - (14)

$$max(t) = max(x(t)) \quad (9)$$

$$min(t) = min(x(t)) \quad (10)$$

$$mean(t) = \frac{1}{M}\sum_{t=1}^{M} x(t) \quad (11)$$

$$Sd(t) = \sqrt{\frac{1}{M}\sum_{t=1}^{M}\left(x(t) - mean(t)\right)^2} \quad (12)$$

$$Skew(t) = E\left[\left(\frac{x(t)-mean(t)}{sd(t)}\right)^3\right] \quad (13)$$

$$Kurt(t) = E\left[\left(\frac{x(t)-mean(t)}{sd(t)}\right)^4\right] \quad (14)$$

3. MATERIALS AND METHODS

Framework for diabetes prediction model is divided into data collection, preprocessing and data preparation, feature engineering and feature selection, proposed model and evaluation approach. Overall schema for the proposed framework is represented in Figure 5.

3.1 DATA COLLECTION

EHR data set has been gathered from Canadian Primary Care Sentinel Surveillance Network (CPCSSN). Dataset utilized in this research consist the 368790 record of 19,181 individuals for time interval of 1998 to 2015. Each record of a patient contains 14 attributes that includes demographics, lab results and diagnosis etc.

Out of 19,218 unique individual approximately 11054 (57.5%) are females and 8164 (42.4%) are males. Data set contains 7719 diabetic and 11499 non diabetic. Complete description of data set is represented in Figure 2 and Table 1.

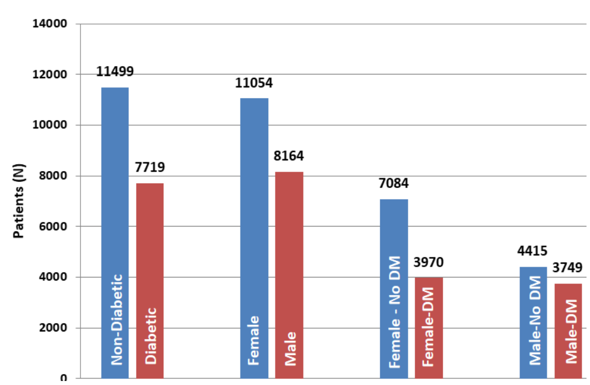

Figure 2: Distribution of data

Table 1: Characteristics of data

Demographic Information	**Findings**
Samples	368790
Patients	19218
Female patients, N (%)	11054 (57.5)
Male patients, N (%)	8164 (42.4)
Age, mean ± SD, Years	65 ±12.5
Clinical Observation	
sBP, mean ± SD, mm Hg	132.05±17.1
BMI, mean ± SD, kg/m2	30.71±9.64
Lab Values	
FBS, mean ± SD, mmol/L	6.11±1.58
A1C, mean ± SD, mmol/L	6.4±0.93
TG, mean ± SD, mmol/L	1.63±1.01
HDL, mean ± SD, mmol/L	1.35±0.4
LDL, mean ± SD, mmol/L	2.7±1.05
Cholesterol mean ± SD, mmol/L	4.79±1.23
BMI, mean ± SD, kg/m2	30.71±9.64
Clinical Diagnosis	
Hypertension, N (%)	13766(71.6)
COPD, N (%)	2138 (11.1)
OA, N (%)	7274 (37.8)
Depression, N (%)	4512 (23.4)

3.2 PREPROCESSING AND DATA PREPARATION

EHR data is noisy and contain missing values. So preprocessing is applied to clean the data [31]. All the records of diabetic individuals after the diabetes onset are removed. Individuals who have already diagnosed with diabetes before their first visit are removed from the data set. Outliers and missing values in the records are imputed by taking mean interpolation [18]. Next, the pre-processed data is normalized to remove the effect of different units in the input data [19]. To predict the patient for diabetes, visit history of a patient before the onset of diabetes is used as features vector. Whereas for no diabetic person, all the visits except the last visit is taken as the feature vector. After this, models are trained with 80% of training samples. In the course of training, models learn to distinguish the hidden patterns between normal individual and pre-diabetic individual. While remaining 20% of the samples are utilize to test the models.

3.3 FEATURE ENGINEERING AND FEATURE SELECTION

Feature engineering and feature selection extracts the important diabetic feature from already available features and select them for predictive models. This increases learning capability of the model and enhance the prediction of the model. In this research 5 unique features (pre-diabetic, TG/HDL, LDL/HDL, Total_cholesterol/HDL, obese) are engineered to extract detail characteristics of each patient. Pre diabetes is a condition when the blood sugar level is higher (5.5 mmol/L to 6.9 mmol/L) but not much higher to be called as diabetic. In most of the cases, pre-diabetic individual develop diabetes. Next, Triglycerides to high-density lipoprotein, (TG/HDL) is also an important indicator of diabetes, enhanced level of (TG/HDL) $\geq$ 1.33 increases the likelihood of diabetes. Obesity is a predecessor of insulin resistance that is link with diabetes. Individual having BMI ≥ 30 is considered as obese. Obesity causes 85% risk of developing diabetes. Further, enhanced level of total_cholesterol to high-density lipoprotein ratio and low-density lipoprotein (LDL) to high-density lipoprotein HDL ratio is also linked with diabetes. Complete description of these engineered features are represented in Table 2.

Table 2: Engineered Features

Attribute	Units	Description	Feature Type
Pre-diabetic	N%	Lab Result	Binary
TG/HDL	mmol/L	Lab Result	Continuous
LDL/HDL	mmol/L	Lab Result	Continuous
Total_Cholesterol/HDL	mmol/L	Lab Results	Continuous
Obese	N%	Clinical Obseravtion	Binary

More over co-relation coefficients are used to examine the relation of the engineered features with diabetes. It estimates the strength of the features with diabetes [20]. It values ranges from 1 to -1. Correlation coefficient larger than zero represents positive relation with diabetes while correlation coefficient less than zero represents a negative relation with diabetes. Besides correlation coefficient of 0 represent no relation with diabetes. Results indicate, derived feature are strongly co-related with diabetes. According to co-relation coefficients and domain knowledge, final set of the features that are selected for the experiments are (FBS, Pre-diabetic, TG/HDL, BMI, HDL, LDL/HDL, Age_at_Exam, Sex, age, sBP. Obese, Total_Cholesterol/HDL) as represented in Figure 3.

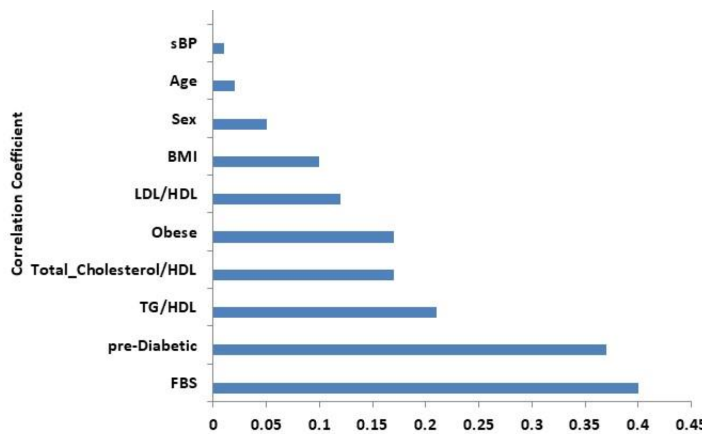

Figure 3: Feature Selection

3.4 PROPOSED MODEL

In this research a novel hybrid deep model is proposed that has two inputs. One input contains the raw data x(t) and second input contains six statistical components Statistics (x(t)) as represented in Equation (9) – (14) . Over all frame work of hybrid deep model is represent in Figure 4.

As represented in Figure 4, CNN extract the features from input data using different filters and kernels in three convolution operation CNN 1, CNN 2, CNN 3: Two convolutional layers are used to extract more hidden information from the raw data. At last "Global Max Pool" is used to reduce the dimension of extracted features. Features extracted by CNN is represented in Equation (15)

$$CNN_{Features} = CNN(x(t)) \tag{15}$$

Although CNN can extract detail hidden information from raw data, but CNN cannot extract the long term dependencies in raw data. So the proposed model employs LSTM in parallel with CNN to extract long term dependent information from the raw data. Features extracted by LSTM is represented in Equation (16)

$$LSTM_{Features} = LSTM(x(t)) \tag{16}$$

Traditional CNN-LSTM is a stacked architecture, CNN extract the features and passed to the LSTM. Unlike traditional CNN-LSTM, this paper proposed a novel parallel architecture of CNN-LSTM to extract the hidden information from raw data, and then integrate these extracted features with statistical features. Thus the features obtained are the fusion of features (time domain and statistics domain) as represented in Equation (17)

$$Fusion_{Features} = Concatenate(CNN_{Features} + LSTM_{Features} + Statistics\left(x(t)\right) \tag{17}$$

Backpropagation is utilized to update the weights [21], with respect to Mean Absolute Error (MAE) as the loss function represented in Equation (18). Adam is used as optimizer to find the convergence path.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - x_i| \tag{18}$$

4. EXPERIMENTAL VERIFICATION

To explore the predictive competency of the proposed deep model, model performance of the proposed model is compare with extensively used models (SVM, LSTM, GRU, CNN-LSTM), for two training data samples (3895, 4688). Experiments are implemented on Windows 10, 64 bit operating system having 16 GB RAM, on Intel Core i7-10750H CPU with speed of 2.60GHz.

4.1 PERFROMANCE METRICS

Multiple performance metrics (accuracy, sensitivity, specificity, precision, F1_score) are used to evaluate the efficacy of proposed the model performance of the proposed model as represented in Equation (19) –(23).

$$Accuracy = \frac{TP}{TP+TN+FP+FN} \tag{19}$$

$$\boldsymbol{Sensitivity = \frac{TP}{TP+FN}} \tag{20}$$

$$Specificity = \frac{TN}{TN+FP} \tag{21}$$

$$Precision = \frac{TP}{TP+FP} \quad (22)$$

$$F1\_score = \frac{2TP}{2TP+FP+FN} \quad (23)$$

Where TP, TN, FP, and FN are the true positive, true negative, false positive and false negative respectively.

4.2 MODEL PERFORMANCE

Table 3 and Figure 6 represent the comparison of the proposed deep model with extensively used models (SVM, GRU, LSTM and CNN-LSTM) for two training samples. From Table 3 it is clearly visible model performance is strongly co-related with training samples. As the training samples increases, model performance increases. Large number of training samples, add positive effect to the model performance. Results represents, proposed hybrid deep model outperform the other models for both the training samples with the highest prediction accuracy of 91.04% and 89.8%. Model performance can further be improved by increasing the training samples. On other and SVM yields lowest prediction accuracy.

Detail comparison of proposed model with other models for largest training sample is examined in Table 4. Highest model performance is achieved by proposed model. Training the proposed model models with 4688 samples, provides the highest accuracy of 91.04%. Next best accuracy of 88.4% is achieved by stacked CNN-LSTM followed by LSTM and GRU with the prediction accuracy of 87.5% and 87.4%. On other hand, SVM yield lowest prediction accuracy of 86.6%.

Performance comparison in terms of sensitivity shows, the proposed model has the best sensitivity of 77.5%. SVM has second best sensitivity of 75.5% while GRU and CNN-LTM has third best sensitivity of 73.3%. Whereas, LSTM exhibits the lowest sensitivity of 68.0%

Model performance in term of specificity represents, proposed model provides the highest range of specificity of 97.0%. LSTM and GRU achieve the second highest specificity of 96.0% and 96.1% along with CNN-LSTM that yields closed performance specificity of 95.2%.

Comparison of precision indicate, propose model yields the highest precision of 92.0%. LSTM and CNN-LSTM also have satisfactory precision of 88.4% and 87.1%. Contrarily SVM and GRU has the lowest precision of 85.5% and 84.0%.

F1-score comparison exhibit, proposed model represents the best F1-score of 84.1%. SVM also perform good with the sensitivity of 81.9% while SVM and GRU have the adequate F1-score of 79.6% and 78.3%. Besides LSTM has the lowest F1-score of 76.9%.

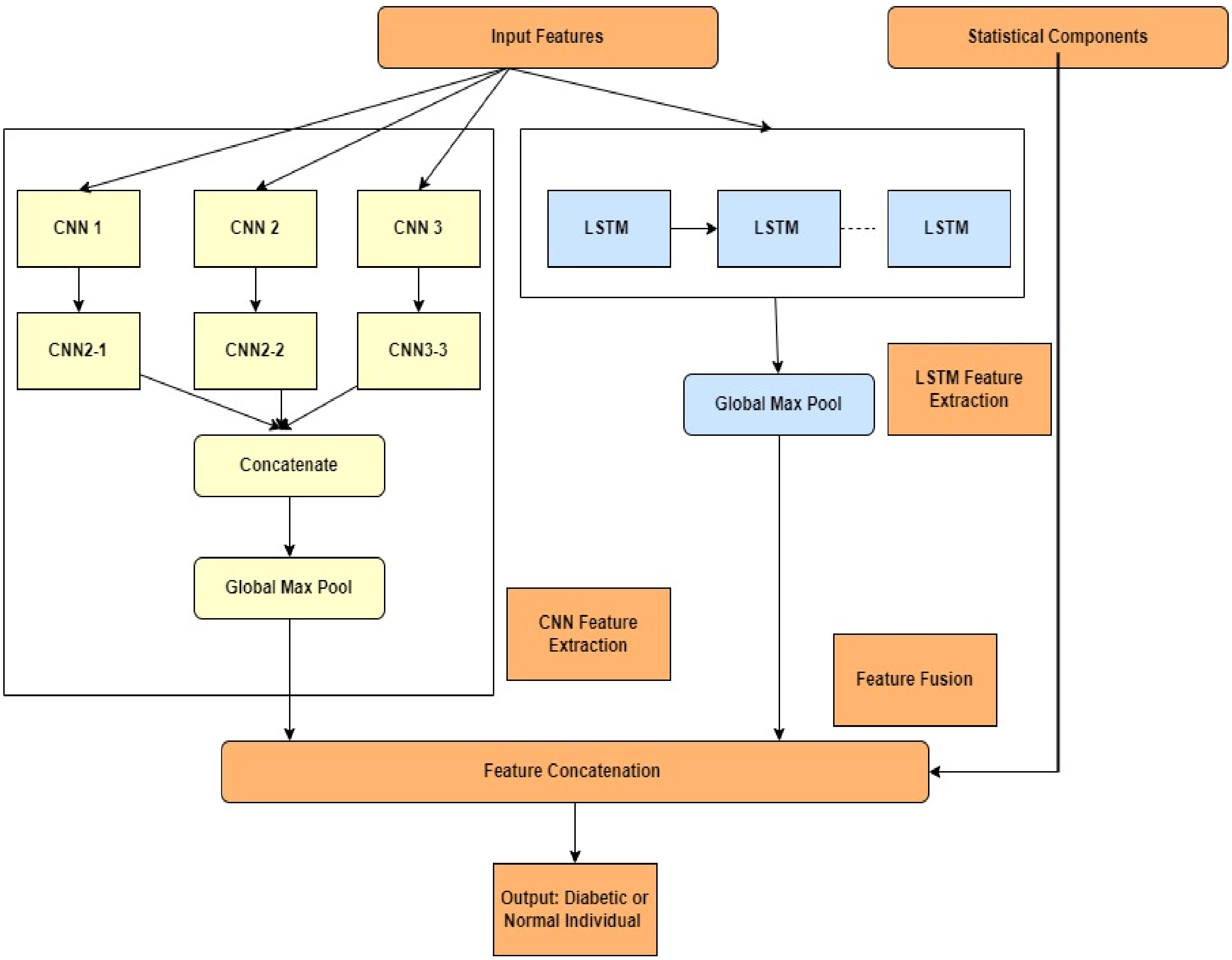

Figure 4: Architecture of proposed hybrid deep model

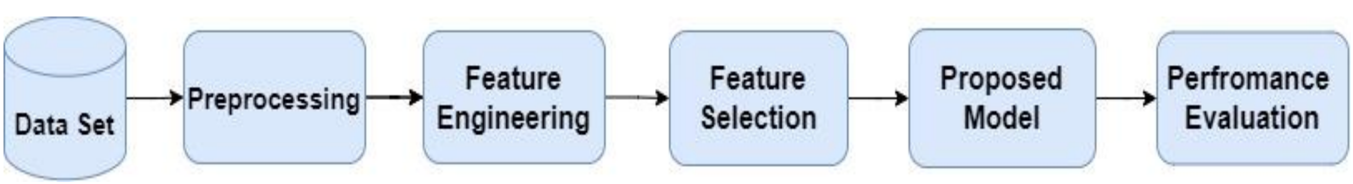

Figure 5: Framework of Proposed Model

Table 3: Accuracy comparison for different training samples

	Training data density	**Accuracy**
ADVANCE MODEL	3895	89.8
	4688	91.04
LSTM GRU	3895	86.4
	4688	87.4
	3895	86.6
	4688	87.5
CNN-LSTM	3895	87.3
	4688	88.4
SVM	3895	86.2
	4688	86.6

Table 4: Detail model comparison of proposed model

	Accuracy (%)	**Sensitivity (%)**	**Specificity (%)**	**Precision (%)**	**F1-score**
LSTM	87.4	68.0	96.0	88.4	76.9
GRU	87.5	73.3	96.1	84.0	78.3
CNN-LSTM	88.4	73.3	95.2	87.1	79.6
SVM	86.6	75.5	91.6	85.5	81.9
Enhanced Model	91.04	77.5	97.0	92.0	84.1

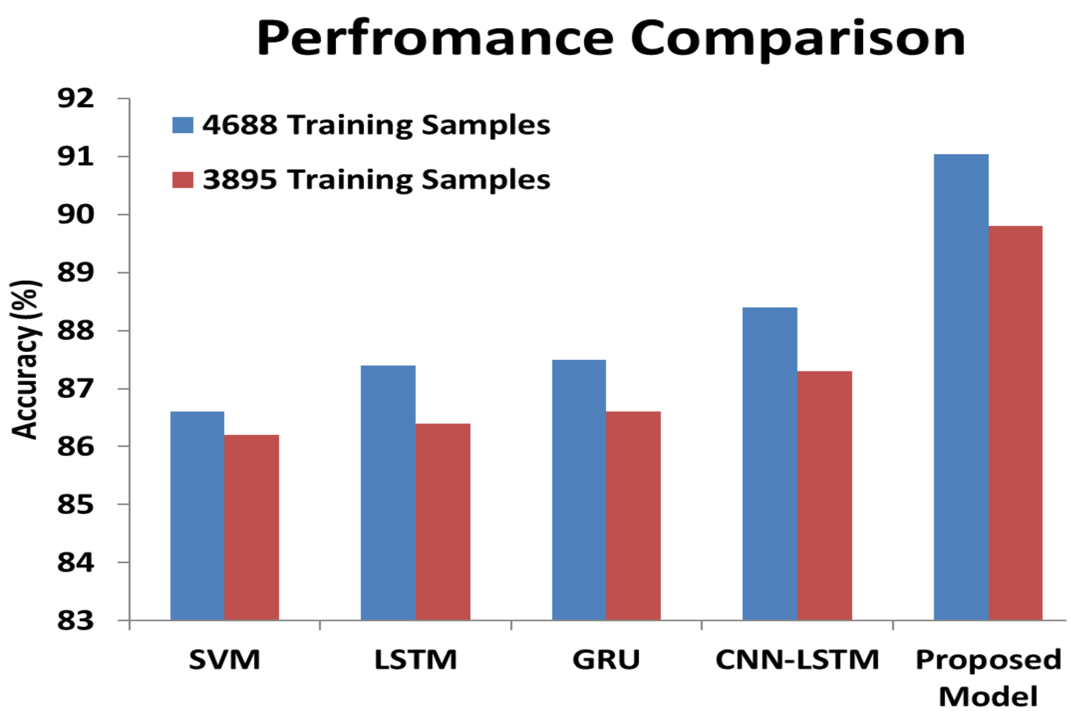

Figure 6: Accuracy comparison

5. DISCUSSION

Diabetes is an incurable disease that has various complications like heart and kidney failure, damage to eyes and blood vessels etc. It also causes many other complications like foot amputation, leg amputation, Alzheimer's, depression etc. So in order to prevent from these complications it is important to forecast the individual with high probability of getting diabetes. However by utilizing medical data (EHR) with artificial intelligence, it is possible to predict diabetes in individual. This research extracts CPCSSN EHR data to predict diabetes. Dataset utilized in this research consist the 368790 record of 19,181 individuals for time interval of 1998 to 2015.

This research presents novel hybrid deep model that extract the features from different domains (time domain and statistics domain) by using only the personal history of a patient to predict diabetes as represented in Figure 4. Unlike stack CNN-LSTM, CNN and LSTM in our architecture is connected in parallel with two inputs to extract the hidden information from raw data. CNN extract detail hidden information from EHR data, but CNN cannot extract the long term dependencies. So the proposed model employs LSTM in parallel with CNN to extract long term dependent information from the raw data. Proposed integrate these extracted features with statistical features. Thus the features obtained are the fusion of features and are more informative for diabetes prediction.

More over this research also employ feature engineering and feature selection to increases learning capability of the model and to enhance the prediction of the model. Five unique features (pre-diabetic, TG/HDL, LDL/HDL, Total_cholesterol/HDL, obese) are engineered from already existent features to extract more hidden trends that separates between of prior diabetes and normal patient visit trend. Pre diabetes is a most significant diabetic features that is extracted from from fasting blood glucose level. Pre diabetes is a condition when the blood sugar level is higher (5.5 mmol/L to 6.9 mmol/L) but not much higher to be called as diabetic. In most of the cases, pre-diabetic individual develop diabetes. Next, Triglycerides to high-density lipoprotein, (TG/HDL) is also an important indicator of diabetes that is linke with insulin resistance, enhanced level of (TG/HDL) $\geq$ 1.33 increases the likelihood of diabetes. Obesity is a predecessor of insulin resistance that is link with diabetes. Individual having BMI $\geq$ 30 is considered as obese. Obesity causes 85% risk of developing diabetes. Further, enhanced level of total_cholesterol to high-density lipoprotein ratio and low-density lipoprotein (LDL) to high-density lipoprotein HDL ratio are also associated with diabetes. Further feature selection uses co-relation coefficients to investigate the co relation of features with diabetes. This represents potency of features in diabetes prediction. Results represents, derived feature are strongly co-related with diabetes. According to co-relation coefficients and domain knowledge, final set of the features that are selected for the experiments are (FBS, Pre-diabetic, TG/HDL, BMI, HDL, LDL/HDL, Age_at_Exam, Sex, sBP. Obese, Total_Cholesterol/HDL) as represented in Figure

Deep hybrid model with the above selected set of features are compared with extensively used machine and deep learning models (SVM, LSTM, GRU, CNN-LSTM) using two training data density (3895, 4688) to explore the potency of deep hybrid model and to investigate the effect of training data density on deep hybrid model. Results represents, the deep hybrid model in comparison with other deep learning models (SVM, LSTM, GRU, CNN-LSTM) continuously yields the best performance for the both set of training samples. As deep hybrid model merge Long-Short Term Memory (LSTM) with Convolutional Neural Networks (CNN) and also utilizes the relative statistical data. Thus deep hybrid model represents more hidden trend and are more informative. More over Table 3 also represents, increases number of training samples increases the model performance

This research represents some significant results, deep hybrid model can more pricesly predict the individuals with high risk of diabetes and thus can delay the progression of diabetes and can enhance the quality of health. More over this research can also minimize the cost for treating diabetic patients and can also be used for medical diagnosis.

Finally this deep hybrid model with unique engineered features has more potency to explore medical EHR data. As far as we know this research is first propose deep hybrid model that uses LSTM in parallel with CNN and also utilizes relative statistical for diabetes prediction

Limitation of this research are the absence of some diabetic predictors like parental history of diabetes and smoking history of a patient that are consider as important for diabetes prediction. More over data set also contain some missing values that effect the model performance.

6. CONCLUSION

In this research, novel hybrid deep model and unique features are presented to predict the individual with high risk of developing diabetes by utilizing personal history of a patient. Proposed model is compare with extensively used model for two training samples. Comparative analysis represents, the proposed model has provide the state of art performance for both the training samples. This shows parallel CNN-LSTM architecture has greater predictive potential than stack CNN-LSTM. In addition, results also confirm that model performance is co-related with training samples. Larger number of training samples guarantees higher model performance. In sum up advance hybrid deep model and unique engineered features are the major contribution of this research work. An additional feature work for this study includes recommendation model for the individual who are at risk of developing diabetes. Moreover proposed hybrid deep model can be implemented for different chronic disease and different data samples.

REFRENCES

[1] R.Valdez, ,P.W. Yoon, N.Qureshi, R.F. Green, and M.J. Khoury, "Family history in public health practice: a genomic tool for disease prevention and health promotion", Annual review of public health, vol.31,pp. 69-87, January. 2010

[2] J.Samuel, S.Joya , R.M. Punnoosea , S. Josea , S.Kumar and B. Johns, "A study to evaluate prevalence of diabetes, mellitus, hypertension and dyslipidemia in patients with acute stroke, myocardial infarction or perip heral, vascular, disease", Xi'an ShiyouDaxueXuebao, vol. 65, March. 2022

[3] F. Khasanah, A. Sitasari, "diet control and regular physical activity can reduce fasting blood glucose level in Type 2 diabetes mellitus patients", Xi'an ShiyouDaxueXuebao, vol. 66 , July. 2023

[4] K. Kaul, M.J. Tarr, S.I. Ahmad, E.M. Kohner, and R. Chibber, (2013). "Introduction to diabetes mellitus", Adv Exp Med Biol, 2012

[5] H.Zhou, R.Myrzashova, and R.Zheng, "Diabetes prediction model based on an enhanced deep neural network", EURASIP Journal on Wireless Communications and Networking, vol.148, pp. 1-13, July.2020.

[6] H.Wu, S Yang, Z.Huang, J.He, and X.Wang, "Type 2 diabetes mellitus prediction model based on data mining", Informatics in Medicine Unlocked, vol.10, pp. 100-107, 2018.

[7] C.L. Morgan, C.J. Currie, N.C. Stott, M. Smithers, C.C. Butler, and J.R. Peters, "The prevalence of multiple diabetes-related complications", Diabetic medicine, vol. 17(2), pp. 146-151, 2000

[8] A. Hoerbst, and E. Ammenwerth, (2010) "Electronic health records". Methods of information in medicine, vol. 49(04), pp. 320-336, 2010.

[9] S. Mani, Y. Chen, T. Elasy, W. Clayton, and J. Denny, (2012). "Type 2 diabetes risk forecasting from EMR data using machine learning", In AMIA annual symposium proceedings, Vol. 2012, pp. 606, 2012.

[10] S.B. Choi, W.J. Kim, T.K. Yoo, S.J. Park, J.W. Chung, Y.H. Lee, and D.W. Kim, (2014). "Screening for prediabetes using machine learning models" Computational and mathematical methods in medicine, vol.2014, 2014.

[11] Y. Li, H.Li, and H.Yao, (2018). "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, vol.2018, July. 2018

[12] R.T. Sousa, L.A. Pereira, and A.S. Soares," Predicting diabetes disease evolution using financial records and recurrent neural networks", arXiv preprint arXiv:1811.09350 ,vol.2020,march 2020

[13] B.Suvarnamukhi, and M.Seshashayee, (2019). "Big data processing system for diabetes prediction using machine learning technique", IJITEE, ISSN, vol. 8, pp. 2278-3075, October. 2019

[14] S. Chattterjee, R. Mitra, S. Bari, Pranabhazra, A. Barnamsantra , A. Das, S. Pattanayak and A. Dey, "Diabettic retinopathy detection using 'ForeseeGAN': A deep learning approach, vol. 65, 2022.

[15] H.Manoharan, and D.J. Dhilipan, (2022). "Diabetes Mellitus Prediction on Class Balanced Data Set using XGBoost Algorithms", Journal of Xi'an Shiyou University, Natural Science Edition, vol. 18, pp. 480-485, January, 2022.

[16] Z. Chen, C. Li, and R.V. Sanchez, (2015). "Gearbox fault identification and classification with convolutional neural networks", Shock and Vibration, vol. 2015, October. 2015.

[17] Y. Yu, X.Si, C.Hu, and J.Zhang, " A review of recurrent neural networks: LSTM cells and network architectures", Neural computation, vol. 31, pp.1235-1270, July. 2019

[18] M. Pourahmadi, "Estimation and interpolation of missing values of a stationary time series", Journal of Time Series Analysis,vol. 10, pp. 149-169, 1989

[19] J.M. Jo, "Effectiveness of normalization pre-processing of big data to the machine learning performance", The Journal of the Korea institute of electronic communication sciences, vol. 14, pp. 547-552, 2019

[20] R.Taylor, (1990) "Interpretation of the correlation coefficient: a basic review", Journal of diagnostic medical sonography, vol. 6, pp. 35-39, 1990.

[21] A.T. Goh, (1995). "Back-propagation neural networks for modeling complex systems", Artificial intelligence in engineering, vol. 9, pp. 143-151, 1995

AUTHORS

First Author: Iqra Naveed, PhD,

Department of Electrical Engineering,

University of Management and Technology,

Lahore 54770, Pakistan,

Second Author: Muhammad Kaleem

Department of Electrical Engineering,

University of Management and Technology,

Lahore 54770, Pakistan,

Correspondence Author: Iqra Naveed