ORIGINAL ARTICLE

IJCIH
Converging Healthcare & Technology

INTERNATIONAL JOURNAL OF CONVERGENCE IN HEALTHCARE

Published by
IJCIH & Pratyaksh Medical Care LLP

www.ijcih.com

A Comparative Study of Activation Functions for Diabetes Detection Using Convolution Neural Networks (CNN)

Bhavya Sharma[1], Venu Gopal[2]

[1]Student, Department of Computer Science & Engineering, Inderprasha Engineering College, APJ Abdul Kalam Technical University, Uttar Pradesh, India, [2]Co-Founder Bootstart Biz Solutions Pvt. Ltd, Advisor-Turning Ideas Ventures

Abstract

Diabetes is a matter of concern for the health of the entire world, its diagnosis and cure are among the prime challenge for the medical fraternity, because it can be controlled but can't be cured, sooner the diagnosis the better it will be for the patient. The latest advancements in the field of Artificial intelligence (AI) found that Machine learning (ML) and Deep Learning (DL) approaches are quite effective in predicting the diabetes, and helps in controlling Blood Glucose (BG) levels.Deep learning is a budding field in predictive analysis and is often used in health care applications where the prediction of diabetes is required to be identified in an early stage. The technique Convolution Neural Networks (CNN) of one of the most promising concept, use for predictionpurpose. The performed study compares the performance of CNN model on the basis of the metrics obtained after using four activation functions viz. ReLU, eLU, tanh, and Sigmoid, over UVA/Padova dataset from UCI machine learning repository. The performance of CNN using different activation functions is measured based on various statistical measures such as recall, precision, F-score and accuracy. The experimental results show that CNN gives the maximum accuracy of 97.3% when applied with the eLU activation function.

Keywords: *CNN, pooling, activation function, deep learning, diabetes.*

Introduction

According to International Diabetes federation Atlas 2020 approximately 463 million adults are already in a grip of the disease which is expected to rise by 700 million by the end of year 2045[1]. Diabetes caused 4.2 million deaths last year and 374 million people are living at risk of developing it. Some 90% of the cases occur due to unhealthy life style habits, lack of exercise, obesity, overconsumption of alcohol etc. Rest 10% of the cases is due to genotype factors, and to maintain the levels of Blood Glucose (BG) insulin therapy is generally advised in the target range[2]. Further, to maintain the Blood Glucose (BG) levels, the insulin therapy require traditional pick and prick approachwhich is not only painful but bring mental disturbance too.

Latest advancements in the field of Artificial intelligence (AI) found that Machine learning (ML) and Deep Learning (DL) approaches are quite effective in predicting the diabetes, and helps in controlling Blood Glucose (BG) levels[3],[4]. Accurate Prediction of the BG levels makes user to take necessary precaution and timely treatment[5],[6].

Deep learning is a budding field in predictive analysis and is often used in health care applications where the prediction of diabetes is identified in an early stage.In general, the Neural Networks are the most explored tool for prediction purpose, Deep learning models like CNN can be used to train complex data and predict the output[7].

The performed study compares the performance of CNN model on the basis of the metrics obtained after using four activation functions viz. ReLU, eLU, tanh, and Sigmoid, over UVA/Padova dataset from UCI machine learnin repository

The remainder of this paper is organised as follows. *Section 2*gives an overview of related work, *Section 3* describes the architecture of the CNN and the various activation functions used in this work, *Section 4* presents methodology and the results obtained from our experiments, *Section 5* concludes the paper with discussion on the results and future work.

Literature Review: Recently various researchers explored the CNN model for prediction and classification problems, viz. Jena et al.[8] developed a CNN model to classify diabetic retinopathy by using four different kind of pooling layers with four different activation functions. The output of the model was evaluated using different evaluation parameters and distinct results were examined Hidenoriet al.[9] used CIFAR-10 dataset of RGB colour images for the 10 classes objects classification, he proposed a method of sparseness regularization for the CNN with ReLU activation function and found that the proposed method of sparseness regularization can directly dilute the effect of the internal covariate shift.

Sharma et al[10] also surveyed the role of activations functions in deep learning and the same was also performed by Chigozie[11] where both have enlightened the role of various activation functions in Artificial Neural Network (ANN) and their importance in Deep Learning (DL) model. Also, Gerald et al[12] studied the performance of Convolution Neural Network (CNN) model by usingdifferent activation functions. Another researcher named Harry Pratt et al.[13] used CNN architecture to develop a network to identify the intricate features involved in the classification task on the retina and provide automatic diagnosis of colour funded images using CNN without user input. Further, recognition of human activity using tri axial accelerometer data, collected from users 'smart phones, through one-dimensional (1D) Convolution Neural Network (CNN), was tested by Song-Mi Lee et al [14]Kezhi Liet al[15] studied and pointed out that artificial mechanisms of inducing insulin are going to affect patient both physically (through pain) and also surpass their moral; they proposed deep learning model for forecasting blood glucose levels.

Through the performed literature survey it is found that activation functions plays vital role towards the performance evaluation of any model, thus the performed study relates to compare the performance of various activation functions *viz. ReLU, eLU, tanh, and Sigmoid, over UVA/Padova dataset from UCI machine learnin repository,* using CNN.

Convolution Neural Network-CNN & Activation Functions: This section describes the architecture of the CNN and the various activation functions used in this work.

Convolution Neural Network-CNN: CNN is a deep feed forward multilayered neural network which is most commonly applied to feature extraction and classification problems. It consists of an input layer, hidden layers and an output layer. In a convolution neural network, the hidden layers include layers that perform convolutions. a dot product of the convolution kernel with the layer's input matrix which generates a feature map, which in turn contributes to the input of the next layer. This is followed by other layers such as pooling layers, fully connected layers, and normalization layers[16].

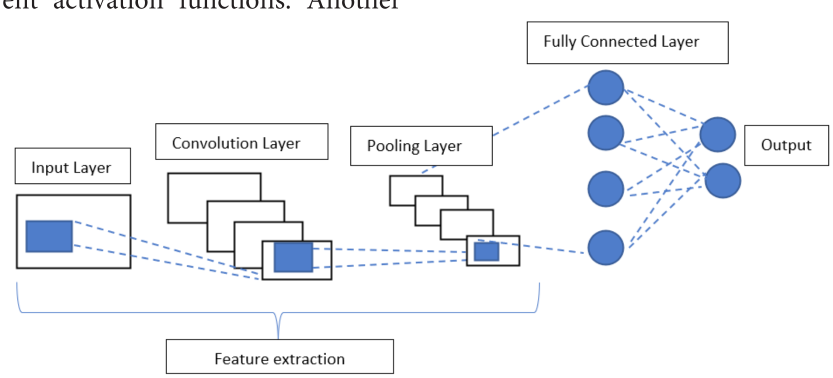


Figure 1: CNN–Architecture

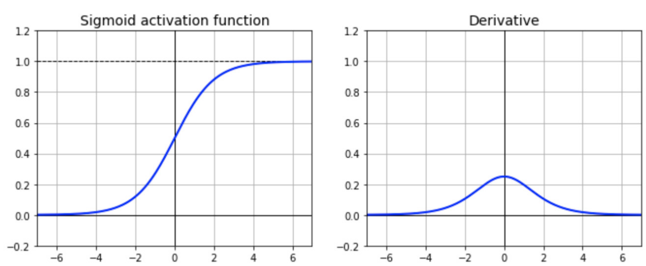
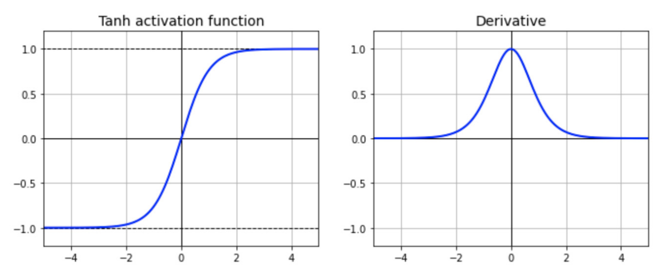
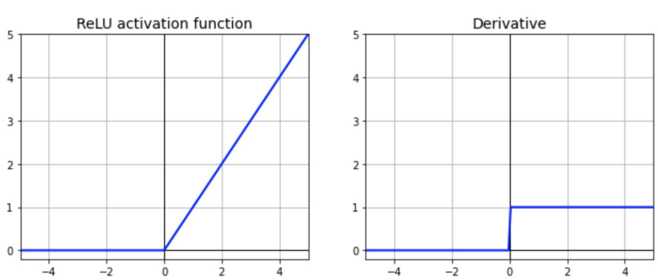
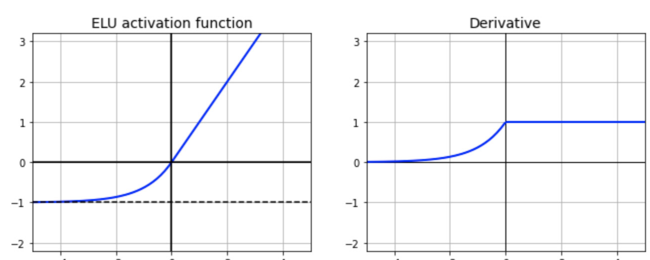
In the process of convolution learned filters are applied, which creates feature map that represent the precise position of features in the input. This process enhances the aspect of data to that extent which becomes unsuitable for further processing. Here, activation function is applied which decides which data to send further and which to reject. Further, pooling is required to reduce the spatial size of the represented feature maps[17]. It is applied on each convolved feature map independently and the best part is it can be applied multiple times in the model[19].

Activation Functions: Activation Functions are exceptionally utilized in Convolution Neural Networks (CNN) to change an input signal into a output signal which thus is taken care of as contribution to the following layer in the stack. CNN model consists of three layers input, hidden and output layers. In neural network the sum of products of input and their corresponding weights are calculated which then finally send to an Activation function which is a node that is put at the end of or in between layers of Neural Networks[9][10]. Activation function can be either linear or non-linear depending on the function it represents. It is used to control the outputs of neural networks;thus, a proper choice of activation function improves results in neural network computing.In a linear model, output is as a linear function of the inputs.

Every node only does the Sum of weights with inputs i.e., Sum(W*x) [8]. This sum is passed to the next layer thus makes it not suitable with complexity of various parameters of data that it fed into neural network[17]. Non-linear functions are the most used activation functions as it makes easy for the model to generalize with variety of data and to differentiate between the output.

Types of activation functions: No model is complete unless the output is generated after passing through activation functions.Here are the different types of nonlinear activations are listed with their merits and demerits. Generally, the activation functions are divided on the basis of their ranges or curves which are explained below.

Table 1 : Graph of Activation Functions: sigmoid, tanh, ReLU, andeLU

$\varnothing(x) = \dfrac{1}{1+e^{-x}}$ 	f(x) = 2sigmoid(2x)–1 f(x)=tanh(x)=$\frac{2}{1+e^{-2x}} - 1$ 
Rectified Linear Unit – Re LU $f(x) = \max\,(0,x)$ $f(x) = \max\,(0,x)$ 	EXPONENTIAL LINEAR UNIT (e LU) $f(x) = \begin{cases} x & ,x \geq 0 \\ a(e^{x}-1), & x < 0 \end{cases}$ 

Methodology and Experiment

This section describes the dataset, methodology and the results obtained from performed experiments,

Table 2: Glimpse of architecture of the values set in the model.

Data Fed	Values
input_sequence	200
output_sequence	20
test_size	0.3
epochs	200
hidden_units	50
Filters	64
Poolsize	2
Kernel size	3
Activation functions	ReLu/eLu/tanh/sigmoid
Optimiser	adam
Loss	categorical_crossentropy
Batchsize	256
headsize	1

The dataset considered in this paper is generated via UVA/Padova T1D available on web. It is one of the most robust and validated framework for generating simulated cases[2]. The network architecture of our CNN model has multiple convolution layers, max-pooling layers, a fully connected layer, and a classifier layer.The following table is a glimpse of architecture of the values set in the model.

Cross entropy is used as loss function and zeropadding is applied at each layer to prevent the input volume from attenuating. In the convolution layers we vary the activation function each time the network is trained to obtain training times and accuracy reports for each activation function. Finally, SoftMax function is applied in the output layer for multiclass classification.

Experiment: For performing the experiment in the model four types of activation functions are used in the CNN model *i.e.* Sigmoid function, hyperbolic tangent function, ReLU function and e LU function. Each combination is tested in the model and the test results are analysed, based on the confusion matrix for respective activation functions the vital statistics viz. the accuracy, precision, recall, and F-score are calculated, the obtained results are summarized in Table-3 below

Table 3: Performance metrics in terms of Accuracy, Precision, Recall, F-measure

Performance Measure / Activation Functions	Accuracy $\frac{TP+TN}{N}$	Precision $\frac{TP}{TP+FP}$	Recall $\frac{TN}{TN+FN}$	F-measure $2\frac{(q.r)}{q+r}$
Elu	97.32	97.85	98.51	97.08
Relu	96.23	98.62	91.43	94.87
Sigmoid	95.81	97.52	96.79	96.39
Tanh	94.82	97.62	96.93	95.86

In all the above equations
N is total number of cases
TP is true positive i.e. correctly classified positive cases.
TN is true negative i.e. correctly classified negative cases.
FP is false positives i.e. wrongly classified negative cases.
FN is false negative i.e. wrongly classified positive cases.
q is precision and r is recall

Results and Conclusion

The Table-3 shows that ELU outperformed all other activation function with maximum accuracy of 97.35 % followed by other activation functions.

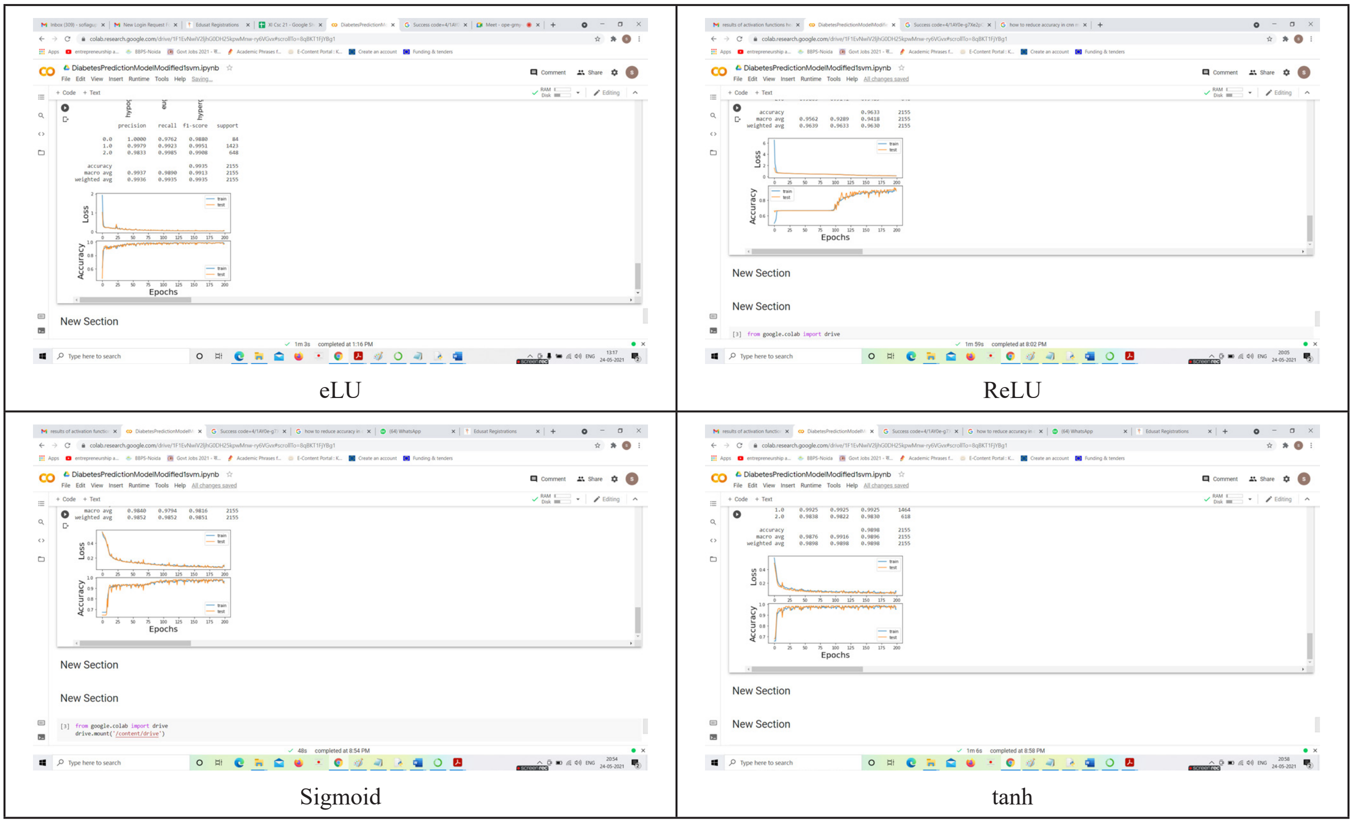


Figure 2 : Graph of Activation Functions: sigmoid, tanh, ReLU, and eLU

The Table-3 and Figure 2, given above shows accuracy and loss plot over the 200 training epochs. Results about accuracy are depicted in Table-3, however we note that as the model iterates through the dataset, value of loss decreases due to the learning ability of the model. The lower the loss values, the better the model and we have the lowest loss values on the ELU and the highest loss on the tanh.

In this paper we presented a brief overview of various non-linear activation functions that are used in deep learning model convolution neural network. We have also compared the performance of CNN model when different activation functions are applied for blood glucose prediction. Experiment was conducted on standard UVA/ Padova dataset of 15 days and data is classified at three levels as euglycemic, hypoglycemic, hyperglycemic. This model is evaluated with various performance metricslike accuracy, precision, recall, F-measure, and a loss function in terms of cross entropy. We considered 200 epochs for the training with longer window size. Our experimental results demonstrated that eLU achieved the highest accuracy of 97.3% among the four activation functions.

Conflict of Interest: Nil

Source of Funding: Self.

Ethical Clearance: Taken

References

1. “International Diabetes Federation,” May 2021. [Online]. Available: https://www.idf.org/aboutdiabetes/what-is-diabetes/facts-figures.html.
2. A. Facchinetti, “Continuous glucose monitoring sensors: past, present and future algorithmic challenges,” Sensors, 2016 vol. 16, no. 12, p. 2093.

3. B. J. Lee, B. Ku, J. Nam, D. D. Pham and J. Y. Kim, "Prediction of fasting plasma glucose status using anthropometric measures for diagnosing type 2 diabetes," IEEE journal of biomedical and health informatics, 2013 vol. 18, no. 2, pp. 555-561.
4. S. Oviedo, J. Vehí, R. Calm and o. Armengol, "A review of personalized blood glucose prediction strategies for T1DM patients," International journal for numerical methods in biomedical engineering, 2017 vol. 33, no. 6, p. e2833.
5. S. Zavitsanou, A. Mantalaris, M. C. Georgiadis and E. N. Pistikopoulos, "In silico closed-loop control validation studies for optimal insulin delivery in type 1 diabetes," IEEE Transactions on Biomedical Engineering, 2015 vol. 62, no. 10, pp. 2369-2378.
6. M. Vettoretti, A. Facchinetti, G. Sparacino and C. Cobelli, "Type-1 diabetes patient decision simulator for in silico testing safety and effectiveness of insulin treatments," IEEE Transactions on Biomedical Engineering, 2017 vol. 65, no. 6, pp. 1281-1290.
7. D. Vanusha and B. Amutha, "A Comparative Study on Different CNN Models Using Diabetic Retinopathy Dataset," Journal of Critical Reviews, 2020 vol. 7, no. 18.
8. M. Jena, S. P. Mishra and D. Mishra, "Empirical Analysis of Activation Functions and Pooling Layers in CNN for Classification of Diabetic Retinopathy," in International Conference on Applied Machine Learning (ICAML), 2019.
9. H. Ide and T. Kurita, "Improvement of learning for CNN with ReLU activation by sparse regularization," in International Joint Conference on Neural Networks (IJCNN), 2017.
10. S. Sharma, S. Sharma and A. Athaiya, "Activation functions in neural networks," International Journal of Engineering Applied Sciences and Technology, 2020 vol. 6, no. 12, pp. 310-316.
11. C. Nwankpa, W. Ijomah, A. Gachagan and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning..," arXiv preprint arXiv:1811.03378, 2018.
12. G. Onwujekwegn and V. Y. Yoon, "Analyzing the Impacts of Activation Functions on the Performance of Convolutional Neural Network Models," AMCIS 2020 Proceedings. 4, 2020.
13. H. Pratta, F. Coenenb, D. M. Broadbentc, S. P. Hardinga and Y. Zhenga, "Convolutional neural networks for diabetic retinopathy," Procedia computer science, 2016 vol. 90, pp. 200-205.
14. S.-M. Lee, S. M. Yoon and H. Cho, "Human activity recognition from accelerometer data using Convolutional Neural Network," in IEEE international conference on big data and smart computing (bigcomp) , 2017.
15. K. Li, J. Daniels, C. Liu, P. Herrero and P. Georgiou, "Convolutional recurrent neural networks for glucose prediction," IEEE journal of biomedical and health informatics, 2019 vol. 24, no. 2, pp. 603-613.
16. M. Kayalvizhi and D. Maheswari, "A HYBRID DEEP LEARNING ALGORITHMS FOR DIABETES MELLITUS PREDICTION USING THERMAL FOOT IMAGES," European Journal of Molecular & Clinical Medicine, 2021 vol. 7, no. 11, pp. 5176-5183.
17. C. Bircanoğlu and N. Arıca, "A comparison of activation functions in artificial neural networks," in 26th Signal Processing and Communications Applications Conference (SIU), 2018.