Impact of Hidden Dense Layers in Convolutional Neural Network to enhance Performance of Classification Model

Helen Josephine V L¹, A.P.Nirmala², Vijaya Lakshmi Alluri³

¹Associate Professor, Department of Computer Applications, CMR Institute of Technology, Bengaluru, India
²Associate Professor, New Horizon College of Engineering, Bangalore
³Research Scholar, Acharya Nagarjuna University, Andhra Pradesh

helenjose.cbe@gmail.com

Abstract. Education and Health care sectors are two predominant areas where societal growth is expected through innovation and technology development. Machine Learning and Deep learning classification models have been entertained in predicting, detecting, and diagnosing major diseases in the early stage. In this research paper, we have analyzed the impact of hidden dense layers in the Convolution neural network to improve the performance of the classification model. Three different classification deep network models have been constructed, analyzed and the result was tested with a diabetes dataset. Results concluded that the more layers with deeper the network better was the classification performance. The classification model with six hidden dense layers outperforms all other less number of hidden dense layers.

Keywords: Deep Neural Network, Deep Learning, Diabetes Prediction, Classification

1. Introduction
Diabetes is a continual disease which takes place when the internal organ pancreas does not produce adequate hormone insulin or the body which is not the possible to successfully consume the segregated hormone-insulin. Uncontrolled or untreated diabetes leads to high blood sugar in the blood and every organ including the heart and vision will be function properly. The medical report states that diabetes patients' growth rate increased rapidly, from 26 million (1990) to 65 million (2016). Early detection of the disease will help the patient to regulate his food habit and undergo proper medication.

Traditional computer programs are expecting input data and algorithms to be fed to provide output. Whereas, machine learning methods are receiving input data as well as label data, which in turn will return the model or algorithm. Deep learning methods are the subset of machine learning, which have several combinations of an artificial neuron or multilayer perceptrons connected in different ways to operate several activation functions[1]. These kinds of deep learning architectures have been predominantly used to create smarter applications in all domains which are primarily based on the artificial natural network. It entertains multi-layers of nonlinear processing for feature extraction and transformation. The main idea of deep learning is exhibited in behavioral patterns in the different layers of neurons in the neocortex of the human brain[2][3].
In our present work, we constructed three different classification deep models to analyze and predict diabetes. We have achieved a high accuracy value of 98.6% for 6 dense layers with various numbers of neurons in each layer. It was compared with the other two classification deep models with 3 and 4 dense layers.

This research paper constructed three different classification deep models to predict the early detections of diabetes. Section II summarizes the existing research in this domain. Section III illustrates the Methodology to construct a deep model. Section IV discusses the datasets and experiment setup. The results were analyzed and discussed in Section V. Eventually, section VI concludes the research paper.

2. Literature Survey
Zhou et.al.,[4] utilizes MSN – Medical-Social-Network based approach to evaluate the extrapolative value of pathological factors to detects early Heart failure detection. The author makes use of EHR - Electronic Health Records which has been collected through the Heart Carer project to calculate the likeness of pathological risk factors. By using these values help to construct the weighted and un-weighted MSN – Medical-Social-Network. GD Group-division algorithm has been used to further split the MSN into two groups' namely high-risk heart failure and low-risk heart failure groups.

Dilshad et al.,[5] constructed a classification model which is based (Conv-LSTM) Convolution Long Short-Term Memory (Conv-LSTM) that was not functional yet in this observe. Researchers utilized PIDD (Pima Indians Diabetes Database) to compare the state of deep learning algorithms, Convolution-Neural-Network (CNN), Traditional LSTM (T-LSTM), and CNN-LSTM.

In one research article Author[6] entertained flattened networks which comprised of many sequences with filters (1D) to acquire similar performance like CNN. The model has been tested with the various dataset and concluded that the flattened layer is a very good alternate for the 3D filters without compromising the accuracy. The proposed model has one more advantage that one it has been trained on, does not need post-processing or physical tuning.

Numerous research has been initiated in a deep neural network with CNN[7],[8]. Liver and Brain Tumors disease has been classified by CNN and Discrete Wavelet Transform and LSTM Network, [9]. Deep CNN LSTM model for forecasting in smart cities

3. Methodology
Convolution neural networks are not only used for image classification but also all other types of dataset. In this session, will see various types of convolution.

3.1 Types of convolution
There will be different type’s convolution which has been used in the Convolution Neural network. The primary types are convolution, dilated convolution, and transpose convolution. These are often used in deep network and also in other networks which contains decoder and encoder kind of architecture [10].

A) Naïve Convolution

Naïve Convolution is having three important parameters, Kernel size – height and width, Stride – height and width, Padding. The dimension of the output feature map is governed by this parameter. The output width and output height are calculated by equation No (1) and equation No. (2). Kernal width will be subtracted from input width and it will be added with 2 times padding ad this value will be divided with stride width and the entire result will be added with one.
\[ O_w = \frac{I_w - K_w + 2p}{S_w} + 1 \]
\[ O_h = \frac{I_h - K_h + 2p}{S_h} + 1 \]  

B) Dilated convolution

Dilated is also known as Atrous convolution. In this type of convolution, the kernel is filled with zero to see a larger receptive field. It has an additional parameter called "Dilation Rate". If the dilation rate is set 2 for any 3x3 convolution then it would produce the same result as naive 5x5 convolution with all other parameters.

C) Transposed Convolution

The idea behind the transposed convolution is to assist in increasing the output feature map size. It helped in encoder-decoder networks to increase the spatial dimension of the feature map. In transposed convolution rebuilt the original optical evolution and executes a convolution. The input image will be padded before the convolution operation.

3.2 Classification deep model

Single or several convolution layers that regularly come with sub-sampling stage consist of Convolution neural networks. These layers are then connected with several fully-connected layers. The complete architecture will be similar to standard multilayer neural networks. The image data set will be primarily used on CNN algorithms. But CNN algorithms will also be used in the numerical dataset to bring out high accuracy and performance.

Keras is one of the open-source libraries which has python interface for artificial neural network, especially for the Tensorflow open source library. It provides four important pre-build layers namely core layers, convolution layers, pooling layers, and recurrent layers. By using these layers, we would create any complex neural network easily. It contains sequential and functional models. The word sequential means linear combinations. Sequential model is easy, smallest as well as can signify nearly all available neural networks. The dense layer is the general and often used layer that contains a deeply connected neural network layer. It can be used as it means a hidden layer that is associated to each and every one node in the next layer. Any dense layer will perform the below operation

\[
\text{Output} = \text{activation}(\text{dot(input, kernel)} + \text{bias})
\]  

The sequential model contains accurately one input tensor and one output tensor in each layer, which is appropriate for a simple stack of layers. When a sequential model has been built, it acts like a layer that contains an input and output attribute. In the sequential model, these attribute to create a model which excerpts the output of all intermediate layers. A Sequential model is not suitable when the model has several inputs and several outputs and when there is an expectation for non-linear topology.

Hidden dense layers place a very vital role in the neural network. In this research work, we have constructed three sequential models. Each has its dense layers and the various number of neurons in each layer. The second sequential model network architecture is presented in Figure. 1 which contains input layers. The dataset contains numerical data that has to be normalized in the pre-processing step. Seven attributes were normalized and concatenated. It is followed by several dense layers. The second classification deep model contains 4 dense layers, each layer contains 32, 24, 12, 8 neurons respectively. In the third classification, deep models are comprised of 6 dense layers and each layer has a different number of neurons. The first dense layer has 64 neurons, second dense layer contains 64 neurons, third dense layer contains 32 neurons, followed by fourth layer which has 24
neurons. The fifth and sixth layer has 12 and 8 respectively. Figure 2 depicts the architecture diagram.

Figure 1. Classification deep model with 4 layers
4. Dataset & Experiment Setup
The performance of the Deep-Convolution Neural Network model was estimated through a famous diabetes database called Pima Indian Diabetes Database (PIDD) and it was downloaded from the UC Irvine Machine Learning Repository. This dataset includes 768 instances. Dataset has 9 important attributes to identify diabetic disease. Table 1 shows the attributes of the dataset.

| Attribute No with observation | Description                              |
|------------------------------|------------------------------------------|
| 1-NTP                        | Number of times pregnant                 |
| 2-PGC                        | Plasma glucose concentration a 2 hours in an oral |
| 3-DBP                        | Diastolic blood pressure (mm Hg)         |
| 4-TSFT                       | Triceps skin fold thickness (mm)         |
| 5-H2SI                       | Hour serum insulin (mu U/ml)             |
| 6-BMI                        | Body mass index (weight in kg/(height in m)^2) |
| 7-DPF                        | Diabetes pedigree function               |
| 8-AGE                        | Patient's Age                            |
| 9-Lable                      | Class 0 or Class 1 - Diagnosis of diabetes with |

5. Experiment Result and Analysis
To speed up feed-forward execution, we have increased hidden layers and engaged a flattened convolution neural network. Table No. 2 shows the overall performance of three different classifications deep models with the various number of hidden layers.

| Total Number of dense layers | No of the neurons in each layer | Accuracy | Precision | Recall | F1 Score |
|------------------------------|---------------------------------|----------|-----------|--------|----------|
| 3                            | 12,12,8                         | 0.832    | 0.762     | 0.754  | 0.758    |
| 4                            | 32,24,12,8                      | 0.895    | 0.819     | 0.896  | 0.856    |
| 6                            | 64,64,32,24,12,8                | 0.986    | 0.964     | 0.996  | 0.986    |

Figure 3 and Figure 4 depicts the loss and accuracy graph for training and validation for the second and third classification deep model. Batch size has been set as 32 and the total number of epochs is 500 for all the classification deep models.
The first classification deep model contains 3 hidden layers with 12, 12, and 8 neurons in each layer respectively. Classification performance is measured with four important parameters namely Accuracy, Precision, Recall, and F1 Score. The result shows that accuracy 83%, precision 76%, recall 75%, and F1 Score 75%. The second classification deep model comprised of 4 hidden layers with 32, 24, 12, and 8 neurons. Performance has been increased 7%, precision by 8%, recall by 15%, and F1 Score by 10%. The final test was conducted with 6 hidden layers classification deep model. Each layer has 64, 64, 32, 24, 12, and 8 neurons respectively. The confusion matrix presented for 4 dense and 6 dense layers with various numbers of neurons in each layer in Figure 4 and Figure 5 respectively. When there is an increase in the number of layers and number of neurons in each layer, CNN outperforms compared to the other classification algorithms. The final experiment outcomes produce 99% accuracy, 96% precision, 100% Recall, and 98% F1 Score. Classification Accuracy has been increased from 83% to 99% from 3 dense layers to 6 dense layers. If we further increase the hidden layer the performance will be saturated since it reaches the highest learning point.
6. Conclusion

In this research paper, we have constructed three different classification deep models to predict type 2 diabetes. The models were analyzed based on the classification performance metrics: accuracy, precision, recall, and F1-score. The results concluded that the deeper the model, which has more dense layers with more neurons in each hidden layer, outputs the less the number of hidden dense layers. The more hidden layers with the deeper the network, better were the classification performance.

References

[1] Krizhevsky, A., Sutskever, I., and Hinton, G. E.. 2012, Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, pp. 1097–1105.

[2] Montufar GF, Pascanu R, Cho K, Bengio Y., 2014, On the number of linear regions of deep neural networks. In: Advances in Neural Information Processing Systems. Red Hook, NY: Curran Associates, Inc. pp. 2924–32.

[3] Mahrishi M., Morwal S. (2020) “Index Point Detection and Semantic Indexing of Videos—A Comparative Review”. In: Pant M., Kumar Sharma T., Arya R., Sahana B., Zolfagharinia H. (eds) Soft Computing: Theories and Applications. Advances in Intelligent Systems and Computing, vol 1154. Springer, Singapore.

[4] C. Zou, A. Li, A. Hou, Z. Zhang, Z. Zhang, F. Wang, Modeling methodology for early warning of chronic heart failure based on real medical big data. Expert Syst. Appl., Published online. https://doi.org/10.1016/j.eswa.2020.113361

[5] Rahman, Md & Saha, Indrajit & Islam, Dilshad & Mukti, Rokeya Jahan. (2020). A Deep Learning Approach based on Convolutional LSTM for Detecting Diabetes. Computational Biology and Chemistry. 88. 107329. 10.1016/j.compbiolchem.2020.107329.

[6] Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, J. Dean, 2019, A guide to deep learning in healthcare. Nat. Med. 25(1), 24–29

[7] Zhou, H., Myrzashova, R. & Zheng, R., 2020, Diabetes prediction model based on an enhanced deep neural network. J Wireless Com Network , 148. https://doi.org/10.1186/s13638-020-01765-7.
[8] Kutlu H, Avcı E., 2019, A Novel Method for Classifying Liver and Brain Tumors Using Convolutional Neural Networks, Discrete Wavelet Transform and Long Short-Term Memory Networks. Sensors (Basel, Switzerland). Apr;19(9). DOI: 10.3390/s19091992.

[9] Huang CJ, Kuo PH. 2018, A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities. Sensors (Basel) Jul 10;18(7):2220. doi: 10.3390/s18072220. PMID: 29996546; PMCID: PMC6069282.

[10] Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks, 2018, An overview and application in radiology. Insights Imaging Vol.9, pp.611–629. https://doi.org/10.1007/s13244-018-0639-9.