Available online at http://scik.org
J. Math. Comput. Sci. 11 (2021), No. 4, 4301-4313
https://doi.org/10.28919/jmcs/5841
ISSN: 1927-5307

PERFORMANCE OF HAF IN ATTENTION-BILSTM FOR PREDICTING THE QUALITY OF AUTOMOBILE ANCILLARY SUPPLIERS

K. SHYAMALA, C.S. PADMASINI*

PG & Research Department of Computer Science, Dr. Ambedkar Government Arts College, Chennai, India

Abstract: Deep learning plays an important role in Machine learning and Artificial Neural Networks. Deep learning resembles the working of human brain in data processing for detecting objects and making decisions. Training and predicting data could happen through various deep learning algorithms available. One such algorithm is Bidirectional LSTM which could be used for classification and prediction of data in machine learning. Attention layer is a cognitive process in neural networks that gives attention and weights towards certain important features. The activation functions are mathematical equations used in neural networks to learn a complex problem which determines the output. Activation functions decide which feature to be activated and define the output. There are 7 types of Activation functions, they are tanh, sigmoid, relu,Leaky ReLU, Parametric ReLU and softmax. The proposed novel work uses Modified tanH with ReLU activation functions to analyze the performance of Auto Ancillary suppliers. Automobile Ancillary spare parts are one of the core industries in Indian Economy. Suppliers of Auto Ancillary Manufacturers provide the raw materials to the product Manufacturers which should be good in quality with reasonable cost and delivered on time.

Keywords: ReLU activation function; Tanh activation function; bidirectional LSTM; attention layer.

2010 AMS Subject Classification: 62M45.

---

*Corresponding author
E-mail address: padmasinics@gmail.com
Received April 10, 2021

1. INTRODUCTION

Deep learning [15] is a category of machine learning with focus on greater flexible dataset with hierarchy of concepts. Deep learning trains the data where input is given and output is extracted. These concepts were applied for supervised learning and unsupervised learning of data. To implement the Deep learning algorithm for the complex problems takes a lot of time but enable to receive desired output. The popular algorithms of Deep learning are Recurrent Neural networks, Convolutional Neural networks and Back propagation algorithms. For this research work, a type of Recurrent Neural Networks was implemented.

Neural networks are a set of neurons and Artificial Neural networks are set of artificial neurons. The nodes are connected to each other and consist of Input layer, hidden layers and output layers. The nodes and edges are connected to each other with weights. The input data travel through Input layer, hidden layer and output layer.

Bi directional LSTM is a part of Recurrent Neural Networks which feeds the input in forward direction and backward directions .The advantages of this algorithm show better performance and learn the model faster compared to LSTM. In bidirectional LSTM the forward and backward outputs are combined and added together and passed to the next layer. According to this research work, bidirectional LSTM was used for classification.

Attention layer plays a very important role in artificial neural networks to enhance the model by Aligning and translating the inputs to outputs. Attention layer decides to give importance to the features for the desired output. Activation functions in attention layer gives attention towards the features and navigates the input signals through hidden layers to the output layer for the better results. There are many types of Activation functions. Two Basic Activation functions are used for this model. Tanh function is a zero centric activation function which is easier to map negative, zero and positive values, it is very similar to sigmoid function. Rectified Linear Unit (ReLU) converges the layers very quickly for output. Modified tanh with ReLU model is an extended model of Modified tanh function for desired results. With all these existing concepts a new model with the name “Modified tanh with ReLU Activation function in Attention based Bi directional LSTM classification” was proposed.

Automobile industry is one of the core industries in Indian Economy [14]. Auto ancillary products are the backbone for Automobile vehicles. Either two-wheeler or four-wheeler needs lot of spare parts to assemble a vehicle or after sales for the vehicles to be maintained. To analyze the Auto Ancillary products on the basis of cost, quality and delivery plays a sustainable feedback from the customers and users. To be more specific quality is very important and each customer expects the maximum output from a vehicle as well as Auto Ancillary products. To maintain the quality of Auto ancillary products, the Manufacturers check the quality of raw materials from suppliers. During the production phase Auto Ancillary products were produced with at most care and with good quality products. One aspect to be considered while concentrating on quality is the raw materials from the Supplier.

Suppliers of the raw material have to ensure a good quality or base for Auto Ancillary products. Raw materials play a vital role in quality sustenance and maintenance. Supplier has to produce good quality raw materials starting from a small piece to a bigger one. The dataset comprises the checking of quality at the supplier site and received the raw materials and checked at the Auto Ancillary Company also. The Supplier dataset consists of vendor test report, fitment, and process document details which give information about supplier raw materials that will be tested at their place. Raw materials follow the guidelines of the organization and tested for design, fitment and process to maintain the quality in the products. Thus the supplier dataset was taken to the novel model to classify and analyze the data. The prediction of suppliers will improve and maintain to 100% in terms of quality.

2. LITERATURE WORK

This novel model Modified tanh with ReLU Activation function in Attention based Bi directional LSTM is proposed where the study involves the Bidirectional LSTM in Deep learning used for prediction and classification. Attention layer is introduced to give attention using activation functions. The concepts for the study were available in these research papers which gave a wide knowledge about the technical factors and area where the research can be extended. With the below research papers listed, it gave an idea for the novelty of the research work.

Deep learning [4] is the recent field for machine learning techniques for data representation. Building of huge data visually with hidden layers for improving the performance leads to a better model. Alex [1] proposes the difference between unidirectional and bi directional long short term memory which was applied for speech recognition and online data for classification. Input pattern using Bi directional LSTM [2] was detected for non-repeated words and characters. The Author has proposed to extract pattern matching features with improved performances.

A Neural model was suggested by Oren [2] for NLP tasks such as word sensing, entity recognition etc. Wenhui[3]authored graph-based dependency parsing using neural network model. This model uses Bidirectional LSTM (BLSTM) to standardize contextual information. Many activation functions exists and Guifany[4] proposed improved ReLU segmentation piecewise for building a better output in convolutional neural networks. Hock Huany[5] suggested an improvised ReLU activation function for deep learning datasets which shows good performance results. Julius [6] proposed how ReLU activation functions could be used in neural networks to solve optimization problems and the loss function.

ReLU activation function was used by Hidenor [7] in the hidden layers for sparse regularization for convolutional neural network to make the inputs as zero in the training process and unwanted increase of the output can be reduced. Armenak [8] proposes a formula for integration for shallow neural network with ReLU activation functions for constructing and training the target function. Weng [9] proposed a unique structure of ReLU networks to provide two computationally efficient algorithms Fast-Lin, Fast-Lip to verify Robustness property of ReLU networks. General deep ReLU neural networks to the modulus function were proposed [10] to learn a better unique model. With these research papers referred the importance and use of ReLU activation function in variety of datasets and fields. Xiao [11] proposed to create a hidden layer for neural networks using ReLU activation function. This paved a way for creating multiple neurons with better results. Thus the ReLU activation was used for making activation function as one of the hidden layer.

Objectives as follows:

1. The Supplier dataset has varied types of data,
2. To find a solution for Multi target class labels
3. To reduce the gradient loss and increase the performance.

3. EXISTING WORK

The existing work available was tanh-activation function in Attention based BiLSTM model for prediction. The previous work was done with the basis of Bidirectional LSTM with Attention layer and modified tanh function for Export dataset. The previous work does not suit for this data as it has multi class labels. The data set had mostly categorical words which need a unique model to analyse and predict. The existing work was not suited as the dataset framework was entirely different, so there was a need for proposed work. Modified tanh function with BiLSTM showed good performance for Export dataset.

4. PROPOSED WORK

Hybrid Activation Function in Attention based BiLSTM was proposed in this research work. Bidirectional LSTM is an algorithm based on Recurrent Neural networks. This algorithm is used in machine learning concepts. For classification and prediction of the dataset, Bidirectional LSTM was used.

BiLSTM uses the connectors and edges for transferring the data from Input to desired results. The pseudo code of the Hybrid Activation Function in Attention based BiLSTM proposed model was explained clearly in Algorithm 1. The proposed work does pre-processing first which consists of integers and characters. The numerical data would be separated and the categorical character data is separated and processed to form vectors. Class label was separated. The Cell size would be defined and passed to bidirectional LSTM. The bi directional LSTM classifies and trains the data works in forward and backward direction. Since the data was fed in both directions, training the model would be accurate .The trained output is passed to next attention layer. Customized activation with tanh and relu functions was given as input to attention layer. This converges the input to the next layer. The last layer is the dense layer which uses softmax function to form the output. The proposed model was checked with various metrics such as accuracy, precision, recall and loss functions. These metrics shows a positive note on increase in accuracy and decrease in loss. Thus the proposed model performance is better than existing model.

Algorithm1: HAF in Attention-BiLSTM

Input: Supplier data set

Output: Classification and Prediction of Supplier dataset

Step 1: Compute the vector score (v0, v1, ….., vn for all categorical texts

Step 2: Compute the binary variable for integer encoded predictor label

i.e., y(0) is represented by an array of {b(0), b(1), ……., b(n).

Step 3: Set cell dimensions for RNN and features

Step 4: To define the model, set the units for LSTM, embedding, input and output.

Step 5: Bi-LSTM layer formation to calculate the integrated sequence of inputs

- *Forward and backward LSTM Layer Concatenation based as defined below;*

  $c(t) = f(t) \oplus b(t)$

Step 6: Set the hybrid tanh (HAF) as activation function for all layers except output layer

$f = 0.5z * \tanh(\mathrm{relu}(z))$

Step 7: Compute Attention layer

- *Compute Attention Weights and context vectors and return S*

$S = \int_{t=1}^{N} a_t h_t$, *where* $a_t = \mathrm{sigmoid}(c_t * v)$ *and* $c_t = f(b + wh_t)$

w denotes weight, b denotes bias, c_t*v *denotes CV (Context vector)*

Step 8: Compute hidden dense layer using HAF

$H(t) = f(d(t-1), x(t); \theta)$, *d(t-1) signifies previous hidden state, x(t) signifies input vector,* θ *is the parameter of HAF*

Step 9: Calculate the final value using softmax AF

Step 10: $n = 1$

For $n \leq$ *epochs,*

Run the input data through all layers

Train the Network (N)

End for

Step 11: Loss calculation $Binary\ crossentropy\ loss = -\frac{1}{y}\sum_{i=1}^{y} y_i . log\overline{y_i} + (1 - y_i). log\ (1 - \overline{y_i})$

Algorithm 1: Hybrid Activation Function in Attention based Bidirectional LSTM

4.1 Pre-processing

Pre-processing of any data is the first step towards the analysis. The data was examined for the size and the parameters available. There are many pre-processing steps to be carried out. It depends upon the data set structures. There were missing values and it should be replaced by the apt values available in the dataset using fill forward method. Using this method the last valid data is updated to the missing values. The dataset consists of categorical values, numbers etc. The categorical values as words are isolated and transform those values to suitable numbers using Label encoder function. Label encoder function converts labels or words to numeric form into a machine readable form. Duplicates were removed to have unique values for analysis and convert to comma separated file for classification.

4.2 Hybrid Activation Function in Attention based BiLSTM

In the proposed model bidirectional LSTM was the base for classification and prediction. The Long short term memory algorithms internalize the concept of Artificial Neural Networks. The Artificial Neural Networks passes the inputs with connectors and edges in the layers. LSTM algorithm works on the same principle where the layers are connected through nodes. The proposed model have Input layer, Hidden Layer and output layer to extract the classification and prediction of the model. To improve the performance of the model on the basis of supervised learning, Bidirectional LSTM was introduced. The input parameters were passed to the hidden layer. BiLSTM classifies the data in both forward and backward directions. The convergence of the two outputs can be summed up, multiplied, concatenated etc. In the proposed work the forward and backward inputs were added together to form a desired output.

The pre-processed data set is passed to the proposed model representing the embedding features, and the rnn cell size. Using Keras Backend in python, the methodology could be customized. The customized activation functions hybridized activation functions using ReLU and tanh were tested. The Rectified Linear Activation function will directly pass to the output layer if it is positive and if negative it gives zero.

Most of the neural networks use the ReLU activation function as this shows better performance and easy to train and converge fully to other layers. ReLU and tanh activation function were used as it overcomes vanishing gradient problems and learns and trains faster. The customized and hybridized activation function was passed to the attention layer. The inputs were classified using Bidirectional LSTM. The attention weights and the context vectors were calculated and converge to nodes to generate output layers.

$$\propto (x) = 0.5 * x * g(\max(0, x)) \tag{1}$$

$$\propto '(x) = 0.5 + \tanh(\max(0, x))[0.5 - \propto (x)] \tag{2}$$

$$\propto'' (x) = 1 - 2 * \propto '(x) \tanh(\max(0, x)) \tag{3}$$

The proposed work is the hybrid activation which was a combination of tanh and relu activation functions. The equation 1 shows the basic equation where x as inputs, 0.5 as a constant value with relu function i.e the values converge from 0 to 1. In the previous work a constant value 0.5 was introduced so that only half of the inputs were passed as an input, so the results showed increased performance. The extension of previous work was taken as a base for the current work. The first derivation of the equation 1 was given in equation 2. The second derivations of equation 2 were given in equation 3.From the derivation, the proposed model concludes that the input converges and narrow down to lowest terms.

The proposed work HAF in Attention- BiLSTM is represented in a diagrammatic form in Figure1. The Figure represents the sequence of proposed model. The inputs were given as X(0) , X(n) and passed to the Embedding vector. The inputs were passed to BiLSTM where the inputs were trained and learned in forward and backward directions. The concatenations of both forward and backward inputs were passed to the attention and hidden layer. The novelty of the proposed work was done in the hidden layer which was the combination of activations and passed to Drop out layer and converges to the output as y.

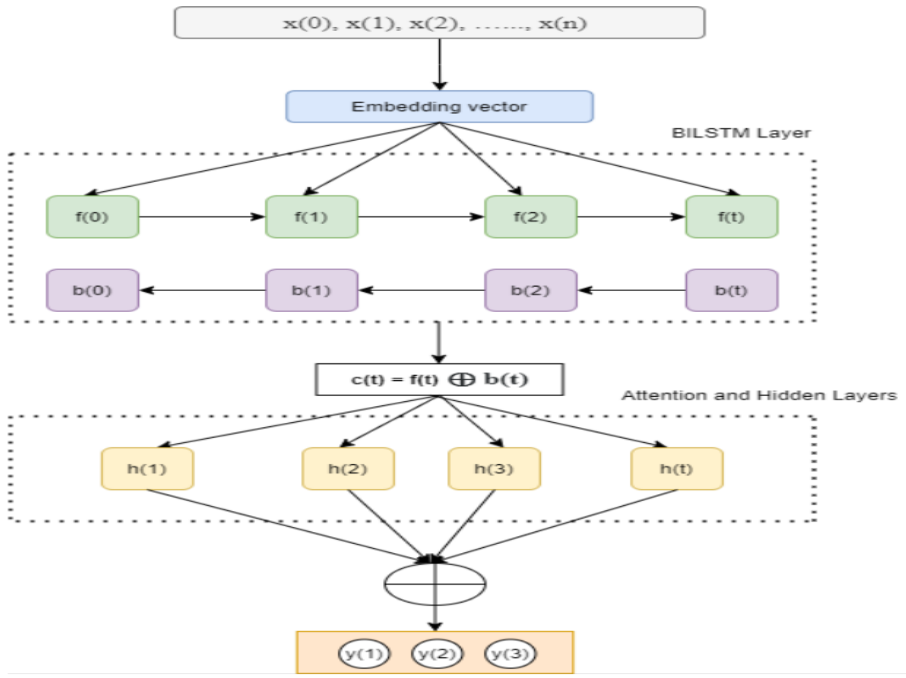

Figure 1: Diagrammatic representation of HAF in Attention-BiLSTM

4.4 Results and discussion

The proposed work shows various results to compare the existing work with the novel work. HAF in Attention based Bidirectional LSTM show better performance compared to the existing work. The comparison shows with various metrics, to highlight a few metrics which was considered for this study was accuracy, AUC, Precision, Loss, and Recall.

```
model.evaluate(x_train, y_train)

311/311 [==============================] - 5s 17ms/sample - loss: 0.1397 - tp: 275.0000 - fp: 36.0000 - tn: 586.0000 - fn: 36.0
000 - recall: 0.8842 - precision: 0.8842 - accuracy: 0.9228 - auc: 0.9883
```

Figure 2: Output for the existing work

```
model1.evaluate(x_train, y_train)

311/311 [==============================] - 3s 10ms/sample - loss: 0.1271 - tp1: 285.0000 - fp1: 26.0000 - tn1: 596.0000 - fn1:
26.0000 - recall1: 0.9164 - precision1: 0.9164 - accuracy1: 0.9443 - auc1: 0.9898
```

Figure 3: Output for the proposed work

Figure 2 and 3 represents metrics that was used to compare between tanh and HAF model. The observations [16] showed results to compare the metrics. Metrics are very important for making right decisions. The benefits of finding the better model were to help in finding accurate results with less time. Accuracy is calculated based on the ratio of correctly predicted data by total number of data available. Figure 2 shows the accuracy of 92% with existing model. Figure 3 show the accuracy of 94% with proposed model. Area under Curve (AUC) is a metric which summarizes the convergence of the curve. AUC of the model represents the measure of the values. The calculation is done using trapezoidal rule. If the AUC value is above 0.5, the model is better. If the AUC value is above 8, it is acceptable and if it is above 9 the model is represented as excellent. The higher the value of AUC shows better performance of the model.

The proposed model shows 0.9 as represented as Good result. Precision are called as positive predictive values and recall is the fraction of relevant instances that were retrieved [17].With all these selected metrics such as Accuracy, Precision , Recall and AUC the proposed work compared with existing work shows better results and performance. Table 1 represents the data comparing the existing model and the proposed model. The comparison was based on precision, recall and accuracy. The numbers shows the difference in the performance of existing and proposed work.

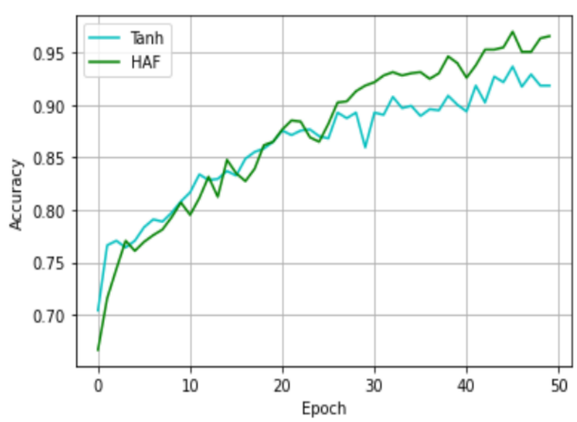

Figure 4: comparison based on Accuracy

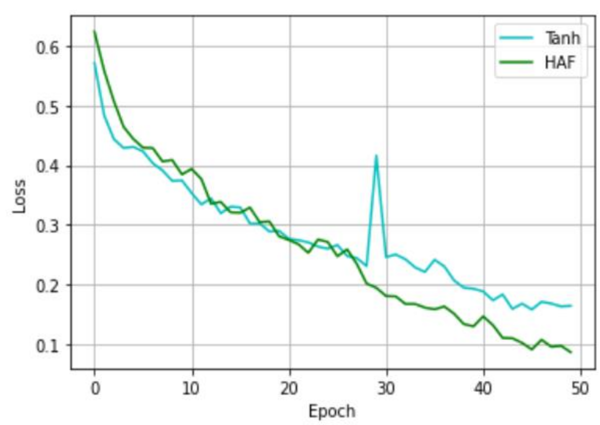

Figure 5: Comparison based on Loss function

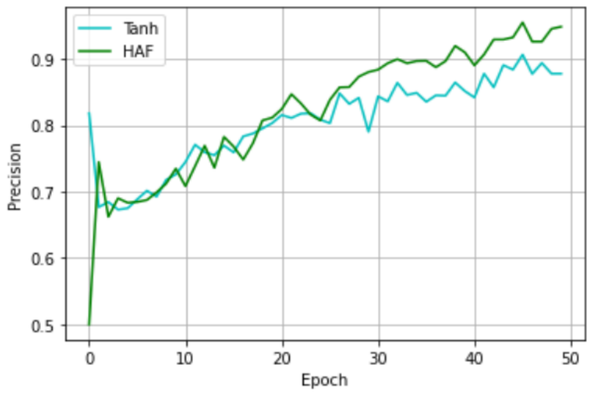

Figure 6: comparison based on precision

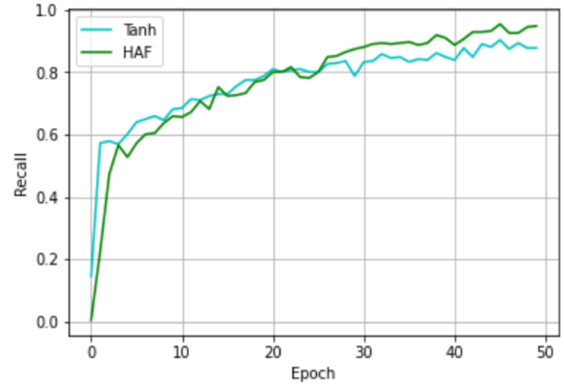

Figure 7: comparison based on Recall

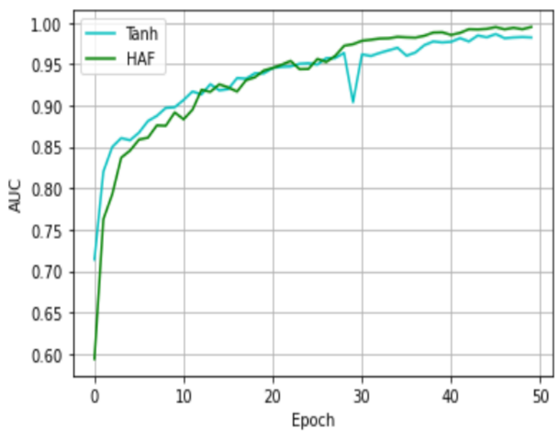

Figure 8: comparison based on AUC

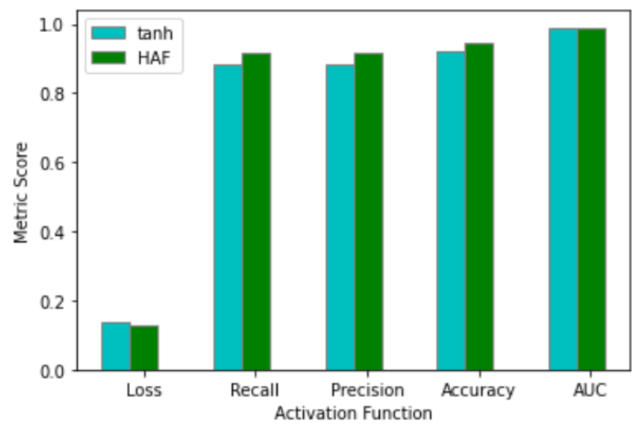

Figure 9: Overall graph represents all metrics

Table 1: Comparison of Existing and proposed model

Model	Precision	Recall	Accuracy
tanh-Attention-BiLSTM model	0.8842	0.8842	0.9228
HAF in Attention-BiLSTM	0.9164	0.9164	0.9443

CONCLUSION

HAF in Attention-BiLSTM was proposed to analyze the performance of Suppliers to Auto Ancillary Manufacturers. Suppliers provide Raw materials to the product Manufacturers. Raw material has to maintain standard procedures for the quality and delivery aspects. If there was a quality compromise, then the products manufactured would not give the quality expected by the customers. This is like a chain reaction for the quality maintenance and delivery logistics. The supplier raw materials would be a one of deciding factor for maintaining quality standards. The proposed model applies the hybrid of Activation functions such as tanh and ReLU to analyze the performance of the model. This model showed better performance comparing the existing model. The Proposed Model showed 94% of the Accuracy and increase in other metrics such as Precision, Recall and Area under curve. These metrics showed the increase in the performance compared with existing work.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] A. Graves, J. Schmidhuber, Framewise phoneme classification with bidirectional LSTM networks, in: Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005., IEEE, Montreal, Que., Canada, 2005: pp. 2047–2052.

[2] O. Melamud, J. Goldberger, I. Dagan, Context2Vec: Learning generic context embedding with bidirectional LSTM. In: Proceedings of the Conference on Computational Natural Language Learning. Berlin, 2016. 51–61.

[3] W. Wang, B. Chang, Graph-based Dependency Parsing with Bidirectional LSTM, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Berlin, Germany, 2016: pp. 2306–2315.

[4] G. Lin, W. Shen, Research on convolutional neural network based on improved Relu piecewise activation function, Procedia Computer Sci. 131 (2018), 977–984.

[5] H.H. Chieng, N. Wahid, O. Pauline, S.R.K. Perla, Flatten-T Swish: a thresholded ReLU-Swish-like activation function for deep learning, Int. J. Adv. Intell. Inform. 4 (2018), 76-86.

[6] J. Berner, D. Elbrächter, P. Grohs, How degenerate is the parametrization of neural networks with the ReLU activation function?, ArXiv:1905.09803 [Cs, Math, Stat]. (2019).

[7] H. Ide, T. Kurita, Improvement of learning for CNN with ReLU activation by sparse regularization, in: 2017 International Joint Conference on Neural Networks (IJCNN), IEEE, Anchorage, AK, USA, 2017: pp. 2684–2691.

[8] A. Dereventsov, A. Petrosyan, C. Webster, Neural network integral representations with the ReLU activation function, in: Proceedings of the First Mathematical and Scientific Machine Learning Conference, Proc. Mach. Learn. Res. 107, PMLR, 2020, pp. 128–143.

[9] L. Weng, H. Zhang, H. Chen, et al. Towards Fast Computation of Certified Robustness for ReLU Networks, Inderjit Dhillon, in: Proceedings of the 35th International Conference on Machine Learning, PMLR 80 (2018), 5276-5285.

[10] D. Yarotsky, Optimal approximation of continuous functions by very deep ReLU networks, in: Proceedings of the 31st Conference On Learning Theory, PMLR 75 (2018), 639-649.

[11] X. Zhang, Y. Yu, L. Wang, Q. Gu, Learning one-hidden-layer ReLU networks via gradient descent, in: Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics, PMLR 89 (2019), 1524-1534.

[12] https://s3.us-east-2.amazonaws.com/hkg-website-assets/static/pages/files/DeepLearning.pdf

[13] https://synapse.koreamed.org/upload/SynapseData/PDFData/1088HIR/hir-22-351.pdf

[14] A. Gautam, Strategizing Indian automobile industry for gaining a sustainable competitive advantage for firms operating in global marketplace, Manage. Insight, 3(1) (2020), 55-69.

[15] https://www.analyticssteps.com/blogs/deep-learning-overview-practical-examples-popular-algorithms

[16] S. Deepak, P.M. Ameer, Brain tumor classification using deep CNN features via transfer learning, Computers Biol. Med. 111 (2019), 103345.

[17] https://en.wikipedia.org/wiki/Precision_and_recall