



## Deep Learning Optimization for Document Text Classification Using Gated Recurrent Unit (G R U)

<sup>1</sup>Afodia Ayuba, <sup>1</sup>Abdulsalam Ya'u Gital, <sup>2</sup>Ali Usman A., <sup>1</sup>Abubakar Umar, <sup>3</sup>Abubakar Muhammad,

<sup>1</sup>Mustapha Abdulrahman Lawal, <sup>4</sup>Abdulkadir Hamidu Alkali

<sup>1</sup>Department of Mathematical Sciences, Abubakar Tafawa Balewa University Bauchi, Nigeria

<sup>2</sup>Department of Computer Science, Federal College of Education Gombe, Nigeria

<sup>3</sup>Department of Maths and Statistics, The Federal Polytechnic Bauchi, Nigeria

<sup>4</sup>Department of Computer Engineering, University of Maiduguri, Maiduguri, Nigeria

### ABSTRACT

Text classification has become a challenge in big organizations in the area of managing large amount of data online. The classification usually applied in the tasks of Natural Language Processing. The task of organizing large documents in an organization has become quite challenging because the internal content of the files are not known. Also, manually organizing each and every file in a large collection is not practically possible as it may take hours to categorize a file based on its contents. In addition, the accuracy of classification cannot be guaranteed, due to the text being a set of data that are naturally sequential. The research work, attempts to address the issue of analysing the document efficiently using multi-level text classification approach based on GRU. The recurrent network will be used to capture long sequence data required for natural language understanding. The performance results of benchmark datasets Factory Report, Bookmarks, Reuters, EUR-Lex and RCV show that the proposed model performs better than existing algorithms in terms of the accuracy in cases of 3 datasets. These include EUR-lex, Bookmarks and RCV where the proposed model attained 74.3%, 100% and 88.44% as against the DSRM-DNN with 62.66%, 47.82 and 86.83% respectively. However, the DSRM-DNN attained classification accuracy of 91.99% for the case of Reuter's datasets demonstrating its superiority against the proposed model which attained 89.83%. Additionally, for the case of algorithm run time, the previous study ignored the computational time of the algorithms which served as index for evaluating the quality of the model. However, this study by included the computational time which attained the better run time for 15 seconds for all the bookmark datasets. On an average, we can conclude that it takes less than two (2) minutes for the proposed model to run on all the datasets. The result shows that our proposed model performs better compared to the DSRM-DNN model.

### ARTICLE INFO

Article History

Received: February, 2022

Received in revised form: March, 2022

Accepted: June, 2022

Published online: June, 2022

### KEYWORDS

Deep Learning, Recurrent Neural Network, Text Classification, Document, Optimization, Gated Recurrent Unit.

### INTRODUCTION

The amount of online document has increase greatly in recent years with so much increase in popularity of the World Wide Web and fast advancement in technology which make the text documents to become available in electronic form. Text document comprises of a large number of documents, conference materials, publications, journals, editorials, news and information etc. It is therefore of significance to develop methods for automatic classifying large collections of text

documents into predefined classes. Text categorization (TC) or Text classification is a means of organizing documents into classes by using artificial intelligence techniques. Also, text classification serves as one of the basic important tasks in text mining and means of assigning a document into predefined categories of documents automatically, based on their contents. Also, Text classification is seen as a supervised learning model that can classify text documents according to their predefined categories. The classification has search

\*Corresponding author: Afodia Ayuba. ✉ [afodia@gsu.edu.ng](mailto:afodia@gsu.edu.ng) ✉ 1Department of Mathematical Sciences, Abubakar Tafawa Balewa University Bauchi, Nigeria. © 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved



engine and is known as Web content for organizing text document properly for the efficient way of retrieving Web documents. Text classification techniques are automatically used in the area such as email filtering, medical diagnosis, indexing for document retrieval, news group filtering, documents organization of word. Text classification has become quite challenge for big organizations in managing and handling large amount of online data. Usually, can be applied extensively in the tasks of Natural Language Processing (NLP), in which files/documents can be stored effectively so that it can be retrieved whenever needs arise (Akhter, et al. 2020). Organizing large documents can be quite tasking as the content of the files are not known Also, manually organizing and searching each and every file in a large collection of documents is not practically possible which required a lot of time spent to categorize and classified a file based on its contents, in addition, the accuracy of classification cannot be guaranteed. (Elnagar, et.al, 2019)

Furthermore, many files have the same titles but putting different request and required different treatments. It's only through analysing the content that can be classified correctly. The study by Chen, et al. (2018), proposed a model with high feature dimension challenges, data was sparse and the computation time is long in traditional Support Vector Machine. Similarly, the algorithms were tested on Arabic text classification, still the model experience high dimensionality problem, feature space and low precision.

Also learning of many algorithms is complex which lead to increase in parallel data in dimension (Boukil, et. al., 2018). In addition text data are sequentially in nature and there is a need to learn and use long-term dependencies to classify sequence data, because a piece of text is a sequence of words, which have dependencies between them. (Zulqarnain, et.al, 2019).

Moreover, the major limitation of study, is that the method used by deep belief neural network ignores semantic extraction and construction of the text features which degrade the speed and accuracy of the model. Therefore, as suggested by Wang, et.al, (2020), to improve the accuracy and speed, there is a need to optimize semantic extraction and construction of deep belief neural network. This suggestion worthy of future research. A number of computational intelligence algorithms

base on deep learning architectures have been employed in various areas of text classification. Deep learning was spired by artificial neural network and has an important role in the recent time. It was applied in the areas such as language processing, data processing, text classification etc.

Deep learning algorithm have various (deep) layers that enable learning and perform multiple tasks repeatedly. Deep learning machines has the ability of solving complex problems even if the dataset is diverse in Structure and unstructured. The deeper learning algorithms is fast learning, and performs better than other algorithms (Bernard Marr, 2018). On the other hand, Recurrent neural network (RNN) is commonly used in speech recognition and natural language processing. The RNN can be used for predicting a next word in a given document. The algorithm has the ability to remember important information across long stretches of time with data are stored in sequential form. Motivated by the RNN, the research aims to develop an optimized deep learning classification method via recurrent network base on GRU architecture.

## RELATED WORK

There are similar researches conducted in this context so far. Based on the review, numerous machine learning approaches have been applied to the text classification. For example, The study in Lin, et.al., (2018) proposed ARCNN is known as attention mechanism Recurrent Convolution Neural Network. The ARCNN used two dataset namely, news dataset of SCHOLAT and Fudan University document as the training dataset. The ARCNN is applied to classify the academic news. The experimental results of the ARCNN was compared with RCNN model and the results shows that ARCNN performed better.

Similarly Kilimci and Akyokuş, (2019) proposed Long Short- Term Memory (LSTM) is a trend to enhance text classification and its was used as model for word embedding with deelep learning algorithm. The Word2vec, Glove, and Fastext are the three different word embebbing models used for word representation of Turkish texts. The result shows LSTM performs better than RNN with 1% - 6% and CNN 2%-4% accuracy Improvement.

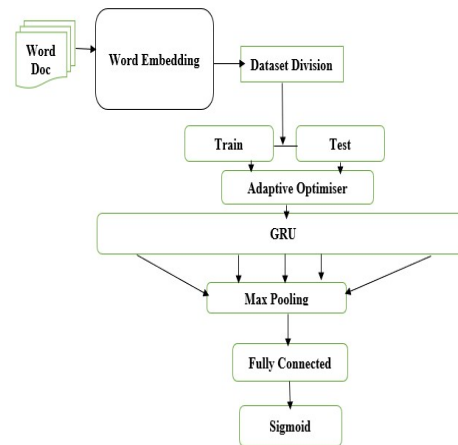
\*Corresponding author: Afodia Ayuba. ✉ [afodia@gsu.edu.ng](mailto:afodia@gsu.edu.ng) 1Department of Mathematical Sciences, Abubakar Tafawa Balewa University Bauchi, Nigeria. © 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved

More so, the study in Zulgarnain, et.al. (2019) proposed a Gated Recurrent Unit (GRU) designed to overcome the weakness in obtaining superior performance in text classification. The proposed model shows a better result on two benchmark datasets in terms of accuracy. Also, GRU obtained better results and performance in text classification than that RNN.

Additionally Wang et al., (2020), proposed the DSRM-DNN model used to improved speed and accuracy in text categorization classification. The four datasets - RCV1-v2, Reuters- 21578, EUR-Lex, and Book marks were used. After the experiment, the results show that DSRM-DNN outperform the compared model. Recently, another study by Zulgarnain, et. al. (2021) proposed an Encoder simplified Gated Recurrent Unit (ES-GRU) for automatic categorization of document based on their respective group and pre- define classes. The model was used for accurate text classification, identified important features and removes the irrelevant, undesired features. The proposed model was evaluated on five datasets and compared with six baselines on text classifications approaches. The result shows that ES- GRU performs better than that of State- of – the- art approaches.

### METHODOLOGIES

The research proposed a deep learning framework that uses gated recurrent Unit (GRU) and max pool layers to categories text descriptions on data collected from numerous sources. The operation of the GRU is to learn dependencies between the embedding vectors, followed by max pooling operation which reduces the sequence of features vectors to a single feature vector, and the fully connected layer that maps the features to binary outputs and lastly the sigmoid operations. A sigmoid operation learns the binary cross entropy loss between the outputs and the targets. The overall structure of the proposed model is evaluated for multi-label classification base on the accuracy and speed achieved by the model. Fig. 1 depict the overall structure and define stages in the proposed model.



**Figure 1:** Structure of the Proposed Model

The proposed model as shown in Figure 1 is to increase speed of text classification and optimized semantic extraction using deep learning neural (GRU and Max Pooling) the network presents, procedure in GRU-CNN-based deep approach. After data preparation (i.e., label assignment, data pre-processing, and data set division), model optimization is then conducted. This procedure involves several rounds of model training with intermittent validations (i.e., validation is regularly conducted after model training with one batch of training data), which enables the production of an optimized GRU base Max Pool deep learning model. Then, the model is tested and Tokenize and pre-process the text data

Else return

Partition the data into training (80%) and testing (20%) partitions

Create a word encoding

(For each parameter operations)

initialize the embedding weights with random normal values.

initialize the GRU operation weights using the Glorot function

initialize the fully connect operation weights using Gaussian function

End

Define model gradient's function and specify training option

- Train using the Adam optimizer

- Train the model using a custom training loop.

(For each epoch)

\*Corresponding author: Afodia Ayuba. ✉ [afodia@gsu.edu.ng](mailto:afodia@gsu.edu.ng) 1Department of Mathematical Sciences, Abubakar Tafawa Balewa University Bauchi, Nigeria. © 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved

Convert the documents to sequences of word indices

- Convert the labels to dummy variables.
- Convert the sequences to darray objects with underlying type single and specify the dimension labels 'BCT' (batch, channel, time).
- Evaluate the model gradients and loss using dlfeval and the model Gradients function.

Clip the gradients.

- Update the network parameters using the ADAM update function.
- Validate the network using the model Predictions function
- Update the training plot.

End

Make predictions on a new set of data and calculate the numbers of true positives

Create a confusion matrix showing the true positive, false positive, and false negative counts

**Stop**

### Experimental Setup

The developed deep classifier model used MATLAB R2021a computer running operating system with 12GB RAM and Pentium® Core i7 processor. We setup the deep learning model with parameters setting tabulated below

**Table 1:** Parameter Settings

Settings	Numbers
Number Epochs	5
Mini Batch Size	256
Learning speed	01
Gradient Threshold	1
Label Threshold	0.5
Execution Environment	"auto"
Validation Frequency	Number Iterations Per Epoch
Sequence Input	1
Word Embedding Layer	1
GRU layer	1- GRU
Classification Output	Cross entropy

The data was partition into training and testing dataset using the cv-partition method and word encoding was created to encode object that is used in training documents as sequences of word. The parameters were defined and initializing

model by specifying the training options as depicted in Table 1. The proposed model was train with and optimized using ADAM optimizer. Finally, the false negative and false positive were calculated. Thus, the validation metrics were displayed using counts in a confusion matrix that has the true positive and false positive values. The proposed algorithm is evaluated using accuracy and speed of computation. The performance metrics of the model is expressed as:

$$\text{Accuracy} = (TP+TN) / (TP+FP+FN+TN) \quad (1)$$

### RESULT DISCUSSION

The result obtained after simulating the network on MATLAB 2021a is presented here. The results are presented in tabular and graphical forms which are analysed using standard performance evaluation metrics used for deep learning applications. The major aims of the research work are to design an improved model that can increase speed and enhance performance of the existing model (DSRM-DNN) for text classification using an optimize GRU base Max pooling for word embedding. Thus, the dataset was imported into the simulation software and fed to model. The preceding sub sections of the overall performance of the proposed model on the five selected Benchmark Datasets and finally, this subsection depicts the graphical summary for the overall experiment on the five datasets which includes (factory, bookmarks, Reuters, Uero-lex and RCV datasets) as presented in the figure below.

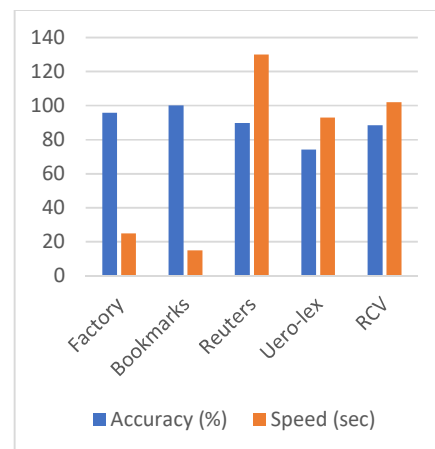


Figure 4: Performance comparison on different datasets

\*Corresponding author: Afodia Ayuba. [afodia@gsu.edu.ng](mailto:afodia@gsu.edu.ng) 1Department of Mathematical Sciences, Abubakar Tafawa Balewa University Bauchi, Nigeria. © 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved

Fig. 4 depicts the overall evaluation of proposed method on five selected benchmark datasets. The result shows that the proposed method performed better with the perfect classification accuracy of 100% for the case of bookmarks dataset within the shortest time of 15 seconds. Furthermore, Table 2 below displays the performance in terms of accuracy of the proposed algorithms against the other existing algorithms that uses the same datasets in their research.

Table 2: Performance Comparison base on accuracy for proposed model against other four different existing algorithms

The algorithms used for comparison in Table 2 includes deep neural network (DSRM-DNN), convolutional neural network base recurrent neural network (CNN-RNN), multilayer perceptron base decision tree (ML-DT) and multilayer perceptron base K nearest neighbor (ML-KNN) algorithms respectively. This result is further plotted in Fig. 5. For better explanation and discussion on how the trend continues. Figure 5 shows the performance of each compared algorithm and GRU on each dataset.

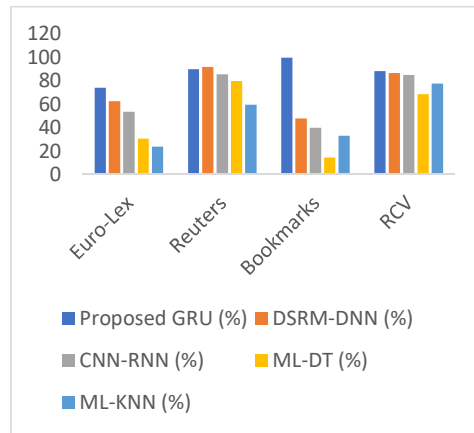


Fig. 5 Performance of each compared algorithm across the four (4) datasets

The accuracy score in percent determines the rate at which each individual model classifies the text correctly. For the case of accuracy. The higher the value, the better the detection rate and versa. Thus, from Table 2 and Fig. 5 the performance trend shows that the develop model performs better than the previous algorithms in text classification for the three (3) cases of the datasets. These include EUR-lex, Bookmarks and RCV where the proposed model attained 74.3%,

100% and 88.44% as against the DSRM-DNN with 62.66%, 47.82 and 86.83%, CNN-RNN with 53.63%, 40.11% and 85.06%, ML-DT with 30.52%, 14.80% and 68.74%, ML-KNN with 23.86%, 33.20% and 77.74%. respectively (See Fig. 5). However, the DSRM-DNN attained the highest classification accuracy of 91.99% for the case of Reuter's datasets demonstrating its superiority against the proposed model which attained 89.83%, CNN-RNN with 85.55%, ML-DT with 79.84%, ML-KNN with 59.59% respectively.

Datasets	Proposed GRU (%)	DSRM-DNN (%)	CNN-RNN (%)	ML-DT (%)	ML-KNN (%)
Euro-Lex	74.3	62.66	53.63	30.52	23.86
Reuters	89.83	91.99	85.55	79.84	59.59
Bookmarks	100	47.82	40.11	14.80	33.20
RCV	88.44	86.83	85.06	68.74	77.74

In general, we can conclude that the proposed model has achieved an improve classification performance as against the existing DSRM-DNN. In general, the proposed algorithm has an improved classification method more than the existing DSRM-DNN. The accomplishment is in agreement with the literature as GRU is one of the best algorithms for sequence-to-sequence modeling such as the current study of text classification. The GRU architecture as describes earlier allows for easy modeling of se2seq problems using the recurrent gates and long short-term memory which makes it to recall past values from the previous network state.

Furthermore, as part of our main objective to evaluate the computational cost of the proposed model, Table 3 shows the average computing time taken by the proposed model to process each of the text datasets used in the study. This result was reported after 10 independent runs on MATLAB 2021 for each dataset and the average of the time is taken and reported in order to ascertain the speed of the model.

\*Corresponding author: Afodia Ayuba. [afodia@gsu.edu.ng](mailto:afodia@gsu.edu.ng) 1Department of Mathematical Sciences, Abubakar Tafawa Balewa University Bauchi, Nigeria. © 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved



Datasets	Computing Time (s)
Euro-Lex	93
Reuters	130
Bookmarks	15
RCV	102

Table 3. Average computing time for the proposed GRU base max pool algorithms. For the case of algorithm run time, the existing study ignore the computational time of the algorithms which serve as a means to measure the quality of the models. Hence, from Table 3. This metric was not reported in the table. However, for this research, we extend on the existing study by including the computational time that served as a means for measuring the quality of the model. Thus, from Table 1. The proposed model attained better run time for 15 seconds for the case of bookmarks datasets. In general, on an average, it can be concluded that it takes less than two (2) minutes for the proposed model to run on all the datasets. Thus, this result is also in agreement with the existing literature. Base on the literature, GRU have proven to be one of the fastest variants of LSTM algorithms and recurrent neural network family in general. This is one more reason why we choose GRU as the main algorithm for this study.

## CONCLUSION

This research proposed a model gated recurrent unit (GRU) that classifies categories of text based on described text data collected from public domain. The model was evaluated base on accuracy and execution time. The experimental evaluation on MATLAB 2021a the result shows that the proposed model performs better than the compared algorithms in terms of the accuracy for the three (3) cases. These datasets include Eurolex, Bookmarks and RCV where the proposed model attained 74.3%, 100% and 88.44% as against the DSRM-DNN with 62.66%, 47.82 and 86.83% respectively. However, the DSRM-DNN attained classification accuracy of 91.99% for the case of Reuter's datasets and demonstrating its superiority against the proposed model which attained 89.83%. In general, it can be concluded that the developed model has an improve classification in terms of accuracy and speed against the existing DSRM-DNN. Additionally, for the case of algorithm run time, the existing study

ignore the computational time of the algorithms which serve as a means for measuring the quality of the model. However, the research extends on the study by including the computational time as a means for measuring the quality of the models. Thus, from the experiments, the proposed model attained better run time for 15 seconds for the case of bookmarks datasets. On an average, it can be concluded that it takes less than two (2) minutes for the proposed model to run on all the datasets. Therefore, this study has proven the suggestion of Wang, et.al, (2020) that focusing on the semantic extraction and the neural network construction optimization can be further improved, in terms of accuracy and speed of the system. However, a major limitation of this study is that the study only focuses on datasets related to only English language text. It is recommended that further work, should work on investigating datasets related to other natural language processing such as Arabic, French etc.

## REFERENCES

- Akhter, M. P., Jiangbin, Z., Naqvi, I. R., Abdelmajeedb, M., & Fayyaz, M. (2020). Exploring deep learning approaches for Urdu text classification in product manufacturing. *Enterprise Information Systems: 1-26*. <https://doi.org/10.1080/17517575.2020.1755455>
- Boukil, S., Biniz, M., El Adnani, F., Cherrat, Loubna, & El Moutaouakki, A.E. (2018). "Arabic text classification using deep learning technics." *International Journal of Grid and Distributed Computing 11* (9), 103-114. <http://dx.doi.org/10.14257/ijgdc.2018.11.9.09103-114>.
- Chen, X., Ouyang, C., Liu, Y., Luo, L., & Yang, X. (2018). A Hybrid Deep Learning Model for Text Classification. 2018 14th International Conference on Semantics, Knowledge and Grids (SKG). <https://doi:10.1109/skg.2018.00014>
- Du, J., Gui, L., Xu, R., & He, Y. (2017). A convolutional attention model for text classification. *National CCF conference on natural language processing and Chinese computing, Springer. 10619*,

\*Corresponding author: Afodia Ayuba. ✉ [afodia@gsu.edu.ng](mailto:afodia@gsu.edu.ng) <sup>1</sup>Department of Mathematical Sciences, Abubakar Tafawa Balewa University Bauchi, Nigeria. © 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved



- 183–195. [https://doi.org/10.1007/978-3-319-73618-1\\_16](https://doi.org/10.1007/978-3-319-73618-1_16)
- Gao, M., Li, T., & Huang, P. (2018). Text classification research based on improved Word2vec and CNN. *International Conference on Service-Oriented Computing, Springer, 11434, 126–135* [https://doi.org/10.1007/978-3-030-17642-6\\_11](https://doi.org/10.1007/978-3-030-17642-6_11)
- Guo, L., Zhang, D., Wang, L., Wang, H., and Cui, B. (2018). CRAN: a hybrid CNN-RNN attention-based model for text classification. *International Conference on Conceptual Modeling, Springer, 11157, 571–585*. [https://doi.org/10.1007/978-3-030-00847-5\\_42](https://doi.org/10.1007/978-3-030-00847-5_42)
- Hassan, A., & Mahmood, A. (2018). Convolutional Recurrent Deep Learning Model for Sentence Classification. *IEEE Access, 6, 13949–13957*. <https://doi.org/10.1109/access.2018.28148>
- Hassanzadeh, H., Nguyen, A., Karimi, S., Chu, K. (2018). "Transferability of artificial neural networks for clinical document classification across hospitals: A case study on abnormality detection radiology reports." *Journal of biomedical informatics 85,68-79*. <https://doi.org/10.1016/j.jbi.2018.07.017>
- Joo, W., Choi, K., & Kim Y. (2019). "Deep learning model for unstructured knowledge classification using structural features." *Personal and Ubiquitous Computing: 1-12*. <https://doi.org/10.1007/s00779-019-01244-x>
- Joshi, A. V. (2020). Machine learning and artificial intelligence, Springer.
- Kilimci, Z. H., & Akyokus, S. (2019). The Evaluation of Word Embedding Models and Deep Learning Algorithms for Turkish Text Classification. *2019 4th International Conference on Computer Science and Engineering (UBMK)*. <https://doi.org/10.1109/ubmk.2019.8907027>
- Lenc, L., & Král, P. (2017). Two-level neural network for multi-label document classification. *International Conference on Artificial Neural Networks, Springer, 10614, 368–375*. [https://doi.org/10.1007/978-3-319-68612-7\\_42](https://doi.org/10.1007/978-3-319-68612-7_42)
- Lin, R., Fu, C., Mao, C., Wei, J., & Li, J. (2018). Academic news text classification model based on attention mechanism and RCNN. *CCF Conference on Computer Supported Cooperative Work and Social Computing, Springer, 917, 507–516*. [https://doi.org/10.1007/978-981-13-3044-5\\_38](https://doi.org/10.1007/978-981-13-3044-5_38)
- Liu, J., Wu, Y. Chang, W., Yang, Y. (2017). Deep learning for extreme multi-label text classification. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. <http://dx.doi.org/10.1145/3077136.3080834>
- Liu, G. & Guo, J. (2019). Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neurocomputing S0925-2312(19)30106-7* <https://doi.org/10.1016/j.neucom.2019.01.078>
- Peng, H., Li, J., He, Y., Liu, Y., Bao, M., Wang, L., Soug, Y., & Yang, Q. (2018). Large-scale hierarchical text classification with recursively regularized deep graph-cnn. *Proceedings of the 2018 world wide web conference*. <https://doi.org/10.1145/3178876.3186005>
- Premchander, K., Sarma, S. S. V. N., Vaishali, K., Reddy, P. V., Anjaneyulu, M., & Nagaprasad, S. (2018). WordNet-based text categorization using convolutional neural networks. *Proceedings of International Conference on Recent Advancement on Computer and Communication, Springer*. [https://doi.org/10.1007/978-981-10-8198-9\\_25](https://doi.org/10.1007/978-981-10-8198-9_25)
- Rebala, G., A. Ravi, et al. (2019). Learning Models. *An Introduction to Machine Learning, Springer: 19-23*.
- Rebala, G., A. Ravi, et al. (2019). Machine Learning Definition and Basics. *An Introduction to Machine Learning, Springer: 1-17*.

\*Corresponding author: Afodia Ayuba. ✉ [afodia@gsu.edu.ng](mailto:afodia@gsu.edu.ng) 1Department of Mathematical Sciences, Abubakar Tafawa Balewa University Bauchi, Nigeria. © 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved



- Semberecki, P. & Maciejewski, P. (2017). Deep learning methods for subject text classification of articles. *2017 Federated Conference on Computer Science and Information Systems (FedCSIS), IEEE*, 11, 357–360.  
<https://doi.org/10.15439/2017F414>
- Wang, L., Chen, Q., Sun, Q., & Jiba, Y. (2018). Research on Short Text Classification Method Based on Convolution Neural Network. *International Conference on Intelligent and Interactive Systems and Applications, Springer*, 885, 397–403.  
[https://doi.org/10.1007/978-3-030-02804-6\\_53](https://doi.org/10.1007/978-3-030-02804-6_53)
- Wang, T., Liu, L., Liu, N., Zhang, H., Zhang, L., Feng, S. (2020). A multi-label text classification method via dynamic semantic representation model and deep neural network. *Applied Intelligence* 50(8): 2339-2351.  
<https://doi.org/10.1007/s10489-020-01680-w>
- Wu, H., Li, D., & Cheng, M. (2019). Chinese Text Classification Based on Character-Level CNN and SVM. *School of Computer and Communication Engineering*, 986, 227–238. [https://doi.org/10.1007/978-981-13-6473-0\\_20](https://doi.org/10.1007/978-981-13-6473-0_20)
- Zheng, J., Cai, F., Chen, W., Feng, C., Chen, H. (2018). Hierarchical neural representation for document classification. *Cognitive Computation* 11(2)317-327.  
<https://doi.org/10.1007/s12559-018-9621-6>
- Zhong, B., Xing, X., Love, P., Wang, X., & Luo, H. (2019). "Convolutional neural network: Deep learning-based classification of building quality problems." *Advanced Engineering Informatics* 40, 46-57.  
<https://doi.org/10.1016/j.aei.2019.02.009>
- Zulgarnain, M., Ghazali, R., Ghouse, M.G., Mushtaq, M. F. (2019). Efficient Processing of GRU Based on Word Embedding for Text Classification. *International Journal On Informatics Visualization*, 3(4), 377-383.