AN ENHANCED GATED RECURRENT UNIT WITH AUTO-ENCODER FOR SOLVING TEXT CLASSIFICATION PROBLEMS

MUHAMMAD ZULQARNAIN

A thesis submitted in
fulfillment of the requirement for the award of the
Doctor of Philosophy

Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia

NOVEMBER 2020

In the name of Allah, Most Gracious, Most Merciful.
I praise and thank Allah.

Special thanks for my beloved father Ghulam Yasin Akhtar.

For dearest,
(Brother and Sisters)

For their love, support, enthusiasm, encouragement and motivation.

For my supervisor,
Prof. Dr. Rozaida Binti Ghazali
For her incredible help, patience, understanding, support and motivation.

For all postgraduate members, fellow friends and house mates.

This thesis is dedicated to all of you.

ACKNOWLEDGEMENT

In the name of ALLAH, witnessing that there is no God but ALLAH and Hazrat Muhammad (P.B.U.H) is HIS last prophet who made knowledge as the basis for guide the humanity towards the right path. I am not adequately capable of finding textual representation for thanking to ALMIGHTY ALLAH whose benediction bestowed me the talented teachers, provision of fruitful opportunities and enablement for undertaking and furnishing this research journey.

Firstly, my deepest gratitude for my worthy, and kind-hearted supervisors Prof. Dr. Rozaida Binti Ghazali and Dr. Yana Mazwin Mohmad Hassim for their professional and sincere guidance throughout my research work. They always guided me sincerely and honestly throughout my whole research work. Without their support, my objective was impossible to be accomplished. Their capabilities have made me able to complete my research work within prescribed timeframe.

I will never forget the educational facilities and research-oriented environment provided by the Faculty of Computer Science and Information Technology (FSKTM) at Universiti Tun Hussein Onn Malaysia (UTHM). The sincere and continuous efforts of UTHM staff and administration to provide modern and latest facilities to ensure quality education in all fields are remarkable. Their sincere efforts and approaches that have made us able for learning information technology (IT) in research oriented dimension. I also would like to thank the Ministry of Higher Education Malaysia (MOHE) and Universiti Tun Hussein Onn Malaysia for funding this research activity under the Fundamental Research Grant Scheme (FRGS/1/2017/ICT02/UTHM/02/5), vote no. 1641.

At last, it is hard fact to describe that I could not manage enough time to serve my old Parents during this long research journey but their pray are always with me unconditionally. Their sincere pray, advices, support, love and care helped me a lot in every step of my any journey in this world. My parents have always been the greatest source of inspiration and role model for me. I am adequately proud in stating that my parents are not highly educated but their role in my enablement of attaining the highest academic qualification is indescribable. I salute them as my whole academic achievements have been possible due to their selfless and restless support. . Special thanks to my beloved father that supported me morally, intellectually, spiritually or any other form which encouraged me in any way. I dedicate this research to my parents. I am also thankful to my brothers, sisters and family members for their full support and cooperation. I am also thankful to all UTHM friends for their guidance, support, encouragement and wonderful time they had with me during this long stay outside my country and never let me feeling away from my home. Thank you so much.

ABSTRACT

Classification has become an important task for categorizing documents automatically based on their respective groups. Gated Recurrent Unit (GRU) is a type of Recurrent Neural Networks (RNNs), and a deep learning algorithm that contains update gate and reset gate. It is considered as one of the most efficient text classification techniques, specifically on sequential datasets. However, GRU suffered from three major issues when it is applied for solving the text classification problems. The first drawback is the failure in data dimensionality reduction, which leads to low quality solution for the classification problems. Secondly, GRU still has difficulty in training procedure due to redundancy between update and reset gates. The reset gate creates complexity and require high processing time. Thirdly, GRU also has a problem with informative features loss in each recurrence during the training phase and high computational cost. The reason behind this failure is due to a random selection of features from datasets (or previous outputs), when applied in its standard form. Therefore, in this research, a new model namely Encoder Simplified GRU (ES-GRU) is proposed to reduce dimension of data using an Auto-Encoder (AE). Accordingly, the reset gate is replaced with an update gate in order to reduce the redundancy and complexity in the standard GRU. Finally, a Batch Normalization method is incorporated in the GRU and AE for improving the performance of the proposed ES-GRU model. The proposed model has been evaluated on seven benchmark text datasets and compared with six baselines well-known multiclass text classification approaches included standard GRU, AE, Long Short Term Memory, Convolutional Neural Network, Support Vector Machine, and Naïve Bayes. Based on various types of performance evaluation parameters, a considerable amount of improvement has been observed in the performance of the proposed model as compared to other standard classification techniques, and showed better effectiveness and efficiency of the developed model.

ABSTRAK

Pengelasan telah menjadi tugas penting untuk mengklasifikasikan dokumen secara automatik ke kategori masing-masing. *Gated recurrent unit (GRU)* adalah sejenis Rangkaian Neural Berulang (RNNs), dan algoritma pembelajaran mendalam yang mengandungi gerbang kemas kini dan gerbang penetapan semula, yang dianggap sebagai teknik klasifikasi teks yang paling efisyen, khususnya pada kumpulan data yang berjujukan. Walau bagaimanapun, GRU mempunyai tiga kelemahan utama apabila ia digunakan untuk menyelesaikan masalah klasifikasi teks pelbagai kelas. Kelemahan pertama adalah kegagalan dalam pengurangan dimensi data, yang membawa kepada penyelesaian berkualiti rendah bagi masalah klasifikasi. Kedua, GRU masih mempunyai kesukaran dalam prosedur latihan disebabkan oleh penindanan antara gerbang kemaskini dan penetapan semula. Gerbang penetapan semula membentuk kekompleksan dan menghasilkan masa pemprosesan yang tinggi. Ketiga, GRU juga mempunyai masalah dalam kehilangan ciri-ciri maklumat pada setiap pengulangan semasa fasa latihan untuk menyelesaikan masalah pengelasan pelbagai kelas. Punca disebalik kegagalan ini adalah pemilihan ciri-ciri input dari dataset (atau output sebelumnya), secara rawak, apabila ia digunakan dalam bentuk piawai. Oleh itu, di dalam kajian ini model GRU baru, iaitu *GRU Encoder Simplified (ES-GRU)* dicadangkan untuk mengurangkan dimensi dari data input berdasarkan *Auto-Encoder (AE)*. Setelah itu, gerbang penetapan semula digantikan dengan gerbang kemas kini untuk mengurangkan penindanan dan kerumitan bagi GRU piawai. Akhirnya, kaedah normalisasi berkumpulan digabungkan dalam GRU dan AE untuk meningkatkan prestasi dan ketepatan model ES-GRU yang dicadangkan. Model yang dicadangkan telah dinilai dengan tujuh kumpulan data penandaaras dan dibandingkan dengan enam pendekatan klasifikasi teks pelbagai kelas yang terkenal, termasuk GRU piawai, AE, memori jangka pendek panjang, rangkaian neural convolutional, support vector machine (SVM), dan Naïve Bayes. Berdasarkan pelbagai jenis parameter penilaian prestasi, sejumlah besar penambahbaikan telah

dilihat dalam prestasi model yang dicadangkan berbanding dengan teknik-teknik klasifikasi piawai lain yang menunjukkan keberkesanan dan kecekapan model yang dibangunkan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF SYMBOLS AND ABBREVIATIONS

GRU	-	Gated Recurrent Unit.
AE	-	Auto Encoder
RNN	-	Recurrent Neural Network
LSTM	-	Long Short Term Memory
CNN	-	Convolution Neural Network
SVM	-	Support Vector Machine
ANN	-	Artificial Neural Network
DAF	-	Decoding Activation Function
EAF	-	Encoding Activation Function
TC	-	Text Classification
DR	-	Dimensionality Reduction
BN	-	Batch Normalization
$f(x)$	-	Function of x
k_e	-	Encoded Activation Function
k_d	-	Decoded Activation Function
z_t	-	Update gate
r_t	-	Reset gate
$\widetilde{h}_t$	-	Candidate State
h_t	-	Output State
σ	-	Sigmoid activation function
Relu	-	Rectified Linear Unit
tanh	-	hyperbolic tangent Activation Function
OVO	-	One-verses-One
OVA	-	One-verses-All
RNTN	-	Recursive Neural Tensor Network
W_x	-	Input weights
U_h	-	Hidden weights with previous time step
DL	-	Deep Learning
NG	-	New Group
ML	-	Machine Learning
E	-	Reconstruction Error
t	-	Time step
NN	-	Neural Network
BPTT	-	Back Propagation Through Time

LIST OF PUBLICATIONS

Journals:

(i) **M. Zulqarnain,** R. Ghazali, S. H. Khaleefah, and A. Rehan. (2019). “An Improved the Performance of GRU Model based on Batch Normalization for Sentence Classification,” *International Journal of Computer Science and Network Security*., vol. 19, no. 9, pp. 176–186, 2019. **(ISI)**

(ii) **M. Zulqarnain**, R. Ghazali, and Y.M.M. Hassim. (2019). “A comparative review on deep learning models for text classification.” *Indonesian Journal of Electrical Engineering and Computer Science*. Vol. 1, No. 4, pp. 886-996. **(Scopus Indexed)**

(iii) **Zulqarnain M,** Ishak SA, Ghazali R, Nawi NM, Aamir M, Mazwin Y. (2020). An improved Deep Learning Approach based on Variant Two-State GRU and Word Embedding for Sentiment Classification. *International Journal of Computer Science and Applications(IJACSA);* vol. 11, no. 1; pp. 593-603 **(Scopus and ISI)**

(iv) **M. Zulqarnain,** R. Ghazali, M. G. Ghouse, and M. F. Mushtaq, “Efficient Processing of GRU Based on Word Embedding for Text Classification,” *Int. J. Informatics Vis*., vol. 3, no. 4, pp. 377–383, 2019. **(Scopus Indexed)**

(v) **M. Zulqarnain,** R. Ghazali, Y.M.M. Hassim, and M. Rehan, “Text classification based on gated recurrent unit combines with support vector machine,” *International Journal of Electrical and Computer Engineering (IJECE).*, Vol. 10, Vo. 4, pp. 3734–3742, 2020. **(Scopus Indexed)**

CHAPTER 1

INTRODUCTION

1.1 Research Background

The rapid development of computer technologies and internet usage caused to generate huge amount of digital textual data (Wang & Qu, 2017), and to retrieve the required content from the great deal of information fast and accurately has become a common concern. Textual data are highly dimensional data, it has irrelevant and unwanted features which are difficult to manage and maintain (Sharif *et al*., 2017). In the early sixties of the 20th century, generation of excessive data was observed. A lot of online information exists in the form of texts, which is in both structured and unstructured form. The unstructured text has become fundamental problem for big organization to manage the large amount of data (Ahmed *et al*., 2016). However, machine learning helps to analyses automatically the data by identifying the patterns for making classification with minimal human intervention. In machine learning, to extract useful information and interested information from constantly increasing documents becomes a vital task. Documents can be in various formats such as word, phrase, term, pattern, concept, sentence, paragraph and text (Wang *et al*., 2018). This excessive information requires some proficient classification algorithms which can be used to assign texts into one or more classes (labels). The classification algorithms are applied on different text applications such as sentiment analysis (Do *et al*., 2019), text clustering (Yi *et al*., 2017), spam filtering (Barushka & Hajek, 2018), website classification (Wang & Qu, 2017), disease report finding (Jadhav *et al*., 2019), document summarization and text classification (Sharif *et al*., 2019).

Text classification has become an active research area over the last decade. Past studies (Kowsari et al., 2017), (Dawar, 2012) indicated that information retrieval plays an important role to improve accuracy in text classification. Textual data are highly dimensional and must be pre-processed before applying classification algorithms (Onan *et al*., 2016). Therefore, it takes much time to discover the knowledge of interest from textual data (Nam *et al*., 2014). The advent of high dimensional data has carried unprecedented challenges to machine learning researchers, making the learning task more complex and demanding computationally.

Text classification has become more active and commonly encountered decision making activity in the area of machine learning. Application of machine learning (ML) techniques for solving text classification issues is one of the basic concerns of researchers. All the classification issues are distributed into two main parts, i.e. Binary class classification and multiclass classification problems (Yeh *et al*., 2017). When the entire data is divided into two classes, is known binary classification. In contrast, a classification issue is regarded as multiclass classification issue if the dataset has more than two distinct classes. Binary classification is considered to be simpler as compared to multiclass classification problem (Don & Iacob, 2018). It is due to the fact that multiclass data has many similarities in the features set, that makes it more complicated for the classifier to distinguish them from other classes. Based on the literature studies, the well-known machine learning techniques for solving text classification issues are: Support Vector Machine (SVM) (Xu et al., 2019), Naïve Bayes (Xu, 2018) and Artificial Neural Network (ANN) (Ghiassi *et al*., 2012). Furthermore, a group of techniques known as deep learning (DL) approaches has been introduce recently for solving these complicated issues.

Deep learning algorithms are the advanced versions of existing ANNs which process some complicated problems. In existing neural networks, if there are more layers and units, there will be a higher expressional power of the network which leads to more complexity of cost functions. In order to overcome the limitations associated with traditional neural networks, deep learning algorithms have been introduced. It is an advanced approach and has been used in many applications for example transfer learning (Long *et al*., 2017), medical text classification (Hughes *et al*., 2017), computer vision (Voulodimos *et al*., 2018), natural language processing (Feature & Joseph, 2017) and many other complex applications. The reasons behind the usage of deep learning algorithms are the low cost of computing hardware, powerful processing abilities and high level of advancement in the machine learning techniques. There are three well-known DL algorithms found in the most recent research namely Convolution Neural Network (CNN), Auto Encoder (AE) and Recurrent Neural Networks (RNNs). All of these deep learning approaches have further several variants appropriate for different kinds of applications (Ahmed *et al*., 2017).

Convolution neural networks fall under the most essential deep learning algorithms based on multiple layers training approach in an efficient manner (Shone *et al*., 2018). There are mainly three layers of convolution neural network such as convolution layer, pooling layer and fully connected layer. All these layers have different roles in the general function of a neural network. Convolutions layer of CNN uses various kinds of kernels to convolve the two-dimensional data set as well as the intermediate feature map. Pooling layer works on data to compresses and makes smooth data. Max-layer selects the maximum value of the receptive field and makes data invariant to small translational changes. However, the fully connected layer converts the two dimensional feature spaces into one dimensional feature space.

Auto encoders (AEs) are a kind of artificial neural network consists of three layers such as input layer, hidden layer and output layer. These layers use the back-propagation behaviour via setting up the high-dimensional input feature set to a low-dimensional output feature set. And by doing so recover the original feature set from the output for efficient learning (Aamir *et al*., 2020). Basically, AE usually performs in two stages namely encoding and decoding. The encoding stage converts the input features to a new representation while decoding stage tries to convert this new representation back as near as possibly to its original inputs. Moreover, AE reconstructs its own inputs instead of predicting outputs from the inputs. In auto encoder, the output vector has the same dimension as the inputs. During the reduction process in auto encoder, the purpose is to minimize the reconstruction error and learned features are actually the code generated by the encoder (Ahmed *et al*., 2017). However, the novelty of the research and powerful deep RNN models is very active research area topic in deep learning community.

In recent years, RNN has been extremely used in several data mining applications to show the better performance on classification issues. RNNs are capable to capture temporal dependencies in sequence information and have shown strong semantic composition approaches for sentiment classification (Liu *et al*., 2016). The key benefit of RNNs is that they can be applied to extracts temporal sequential data with variable-length, which generates flexibilities in evaluating reviews of various lengths. RNNs have various types such as Recursive Neural Network (Cardie *et al*., 2014), Matrix Vector-RNN (Baly *et al.*, 2017), Recursive Neural Tensor Network (Socher *et al*., 2013), Long Short Term Memory (LSTMs) (Hochreiter, 1997), and Gated Recurrent Unit (GRU) (Cho *et al*., 2014). GRU is a variant of RNN family that consist of two gates such as update gate and reset gate was proposed by (Cho *et al*., 2014) and is the latest version of complex LSTM cell architecture. GRU contains three layers and fewer parameters that explained by very simple set of equations, thus need significantly less computational power. These layers are input layer, hidden layer and output layer, which are used for learning statistical features more efficiently (Xing *et al*., 2019). Text data is high dimensional data that has lot of features. A single layer cannot extract informative features from the raw data. Therefore, in recent studies on deep learning, researchers have used multiple layers to extract the most useful features from the raw data. GRU are commonly used as the similar to other types of RNN nodes, particularly when there exist some noise in the input data where usually other algorithms fail to classify the data points.

The aim of LSTM and GRU is to classify data based on previous time step but the working mechanism is very minor different in both. LSTM structure consists of three gates and more complex than GRU while the GRU is a latest and simple model that consists of two gates such as update gate and reset gate. Update gate decides to help the model that how much of the past information through previous time step t should be updated and pass to the future. While reset gate has the opposite functionality as compared to update gate, it applies to decide how much of the past information from the previous hidden state can be ignore in the conventional GRU model. In the natural language processing, reset gate may occur when transferring from one text to another one which is not found the semantically interrelated values. In these conditions, it is convenient to reset the stored memory in order to prevent taking a decision regarding an unrelated history. Although, GRU is one of the most effective approach has applied for solving various types of text classification problems. Many previous researchers have worked on accuracy and performance for text classification using GRU but still there are some gaps and drawbacks associated with standard GRU and there is much work needed to improve the existing GRU.

The main purpose of this research is to modify the standard GRU structure in order to reduce the complexity and improve the performance. Particularly, the main contribution of this research is three-folded: Firstly, to enhance the performance of GRU model with Auto Encoder for dimensionality reduction to solve text classification problems. This problem is the incapability of the GRU to capture the features set with some useful information. Secondly, this research evaluates to replace the reset gate with update gate in the standard GRU network design. Similarly to (Zhou *et al*., 2016), have found that to removing reset gate does not significantly impact the system performance. Thirdly**,** the integration of Batch Normalization method in the training phase of the model refers to normalize the fluctuations in the nodes values during each iteration and minimize the loss function. Finally this research replaces the hyperbolic tangent activation function (tanh) with Rectified Linear Unit (Relu) activation in candidate equation. Relu units have been demonstrated to be better performance than sigmoid non-linearities for deep learning approaches.

1.2 Problem Statement

Examining multiclass text classification issue is one of the most complex problems in machine learning. It is mainly due to the fact that there are several similarities in the feature set of all the targeted classes (Raziff *et al*., 2017). This similar behavior in different classes of the text data make it a difficult task for classifiers to distinguish between different classes. Several approaches have been proposed and methods have been developed but all of these algorithms suffer from different types of drawbacks and still there is much work needed to be done in this area to develop issue independent and efficient algorithms for solving classification issues (Nikam, 2015). In order to improve the capability of text classification algorithms, the properties of feature reduction and classification stage are targeted. In the most specific format, the motivational factors for this research are those properties of multiclass classification algorithms that directly affect the solution quality of a technique when solving these issues.

Additionally, this research work has targeted different issues associated with a text classification technique in order to enhance its performance. Although GRU is one of the most powerful approach applied for solving various types of text classification problems, but still some drawbacks are associated with this algorithm which needs proper attention to develop a technique that leads to a problem independent and high quality solution generation for solving these complicated issues. Same as other text classification models, GRU perform the text classification task in three major stages namely features extraction, feature reduction and classification. The first drawback associated with standard GRU is its failure in application independent data dimensionality reduction according to input data (Hao *et al.*, 2019). The result of this failure is the incapability of the technique to captures the finer details for possessing the useful information. Resultantly, it leads to low quality solution of the text classification issue. Secondly, based on experiment, the drawback was found to addresses the redundancy and complexity in the standard GRU structure for solving the multiclass text classification. The main reason behind this failure is the redundancy between the update gate and reset gate during the training phase. Similarly (Zhou *et al.*, 2016), found that removing reset gate does not significantly impact the system accuracy, due to the redundancy between the functionality of update and reset gates. This issue is the incapability of the GRU to the repetition of function in both reset and update gate. Finally, the third issue is concerned to RNN such as GRU when it combines with other classification techniques for solving text classification problems, the technique may still find difficulty in providing the most accurate results. The reason behind of this failure is the high computational cost and loss of informative features in the training process, when they are applied in their standard form (Davidson, 2016), (Justus *et al.*, 2019).

Therefore, three problems are the target of this research activity that are related to the standard operational steps of GRU that include: a) the failure in data dimensionality reduction that cause difficulty for the technique to capture the finer details possessing the useful information: b) redundancy and complexity between the update gate and reset gate in the standard GRU design and c) The failure of the nodes complexity and loss of informatics features in the training process. These three problems result in low quality solution of the multi class classification issues in the terms of accuracy.

1.3 Research Objectives

The main purpose of this research is to enhance and develop GRU based on AE for solving the text classification problems. This research focuses on Auto Encoder, Batch Normalization (BN) and replacing the reset gate with update gate in GRU design. In addition, this research work also seeks to find a better network architecture in order to improve the accuracy, with less execution time and computational cost.

In order to solve the aforementioned issues associated with standard GRU, following objectives have been set to achieve from the proposed work:

(i) To propose an enhanced Gated Recurrent Unit (GRU) with an Auto Encoder (E-GRU) for solving dimensionality reduction problems.
(ii) To replace reset gate with an update gate in (i) in order to reduce its complexity and redundancy, and called as ES-GRU.
(iii) To integrate batch normalization in the training phase of ES-GRU in order to boost up the training for improving the performance of ES-GRU.
(iv) To evaluate and compare the out-of-sample performance of the proposed ES-GRU with the baseline approaches for text classification

1.4 Research Scope

The proposed Gated Recurrent Unit with Auto Encoder method of this research work applied for solving multiclass text classification issues. The proposed model evaluated on seven benchmark texts datasets including: 20newsgroup, Reuters21578, Amazon reviews, AG's news, IMDB, yahoo answers and Yelp reviews. All the benchmark datasets are available online repositories (Zhang *et al*., 2018). For analyzing and testing, Softmax classifier was used as a final classification layer in the proposed model. In the comparative analysis of the proposed model, six well-known standard text classification approaches including standard GRU, standard AE, LSTM, CNN, SVM and Naïve Bayes were used.

REFERENCES

Aamir, M., Nawi, N. M., Mahdin, H. Bin, Naseem, R., & Zulqarnain, M. (2020). Auto-encoder variants for solving handwritten digits classification problem. *International Journal of Fuzzy Logic and Intelligent Systems*, *20*(1), 8–16. https://doi.org/10.5391/IJFIS.2020.20.1.8

Abreu, J., Fred, L., Macêdo, D., & Zanchettin, C. (2019). Hierarchical Attentional Hybrid Neural Networks for Document Classification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11731 LNCS*, 396–402. https://doi.org/10.1007/978-3-030-30493-5_39

Agarap, A. F. M. (2018). A Neural Network Architecture Combining Gated Recurrent Unit (GRU) and Support Vector Machine (SVM) for Intrusion Detection. In *10th International Conference on Machine Learning and Computing* (pp. 26–30).

Agnihotri, D., Verma, K., & Tripathi, P. (2017). Variable Global Feature Selection Scheme for automatic classification of text documents. *Expert Systems with Applications*, *81*, 268–281. https://doi.org/10.1016/j.eswa.2017.03.057

Ahmed, M., Paul, A. K., & Akhand, M. A. H. (2017). Stacked auto encoder training incorporating printed text data for handwritten bangla numeral recognition. *19th International Conference on Computer and Information Technology, ICCIT 2016*, 437–442. https://doi.org/10.1109/ICCITECHN.2016.7860238

Ahmed, R., Al-Khatib, W. G., & Mahmoud, S. (2016). A Survey on handwritten documents word spotting. *International Journal of Multimedia Information Retrieval*, *6*(1), 31–47. https://doi.org/10.1007/s13735-016-0110-y

Alexandrescu, A. (2018). A distributed framework for information retrieval, processing and presentation of data. *2018 22nd International Conference on System Theory, Control and Computing, ICSTCC 2018 - Proceedings*, 267–272. https://doi.org/10.1109/ICSTCC.2018.8540765

Ali, A., Shamsuddin, S. M., & Ralescu, A. L. (2015). Classification with class imbalance problem: A review. *International Journal of Advances in Soft Computing and Its Applications*, *7*(3), 176–204.

Altınel, B., & Ganiz, M. C. (2018). Semantic text classification: A survey of past and recent advances. *Information Processing and Management*, *54*(6), 1129–1153. https://doi.org/10.1016/j.ipm.2018.08.001

Alvi, M. B., Mahoto, N. A., Alvi, M., Unar, M. A., & Akram Shaikh, M. (2018). Hybrid classification model for twitter data-A recursive preprocessing approach. *5th International Multi-Topic ICT Conference: Technologies For Future Generations, IMTIC 2018 - Proceedings*, 1–6. https://doi.org/10.1109/IMTIC.2018.8467221

Arkhipenko, A. K., Kozlov-ilya, K. I., Integral, T. J., Kirillskorniakov, S. K., Gomzin, G. A., & Turdakov, T. D. (2016). Comparison of neural network arChiteCtures for sentiment analysis of russian tweets. In *In Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference Dialogue* (pp. 50–59).

Arpit, D., Zhou, Y., Ngo, H. Q., & Govindaraju, V. (2016). Why regularized auto-encoders learn sparse representation? *33rd International Conference on Machine Learning, ICML 2016*, *1*, 211–223.

Ballester, P., & Araujo, R. M. (2016). On the performance of googlenet and alexnet applied to sketches. *30th AAAI Conference on Artificial Intelligence, AAAI 2016*, 1124–1128.

Baly, R., Hajj, H., Habash, N., Shaban, K. B., & El-Hajj, W. (2017). A sentiment treebank and morphologically enriched recursive deep models for effective sentiment analysis in Arabic. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *16*(4). https://doi.org/10.1145/3086576

Barushka, A., & Hajek, P. (2018). Spam filtering using integrated distribution-based balancing approach and regularized deep neural networks. *Applied Intelligence*, *48*(10), 3538–3556. https://doi.org/10.1007/s10489-018-1161-y

Bellec, G., Scherr, F., Hajek, E., Salaj, D., Legenstein, R., & Maass, W. (2019). Biologically inspired alternatives to backpropagation through time for learning in recurrent neural nets. *ArXiv Preprint ArXiv:1901.09049*, 1–37. Retrieved from http://arxiv.org/abs/1901.09049

Biswas, S., Chadda, E., & Ahmad, F. (2015). Sentiment Analysis with Gated

Recurrent Units. *Advances in Computer Science and Information Technology (ACSIT)*, *2*(11), 59–63.

Bsir, B., & Zrigui, M. (2018). Enhancing deep learning gender identification with gated recurrent units architecture in social text. *Computacion y Sistemas*, *22*(3), 757–766. https://doi.org/10.13053/CyS-22-3-3036

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011

Cardie, O. I. and C. (2014). Deep Recursive Neural Networks for Compositionality in Language. *In Advances in Neural Information Processing Systems*, 2096–2104.

Chen, C. W., Tseng, S. P., Kuan, T. W., & Wang, J. F. (2020). Outpatient text classification using attention-based bidirectional LSTM for robot-assisted servicing in hospital. *Information (Switzerland)*, *11*(2), 106. https://doi.org/10.3390/info11020106

Chen, M., Weinberger, K. Q., Xu, Z., & Sha, F. (2015). Marginalizing Stacked linear denoising autoencoders. *Journal of Machine Learning Research*, *16*, 3849–3875.

Chen, S., Zheng, B., & Hao, T. (2018). Capsule-based Bidirectional Gated Recurrent Unit Networks for Question Target Classification Shi. *In China Conference on Information Retrieval*, *1*(September), 67–77. https://doi.org/10.1007/978-3-030-01012-6

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., … Ma, J. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena*, *151*, 147–160. https://doi.org/10.1016/j.catena.2016.11.032

Cherabier, I., Hane, C., Oswald, M. R., & Pollefeys, M. (2016). Multi-label semantic 3D reconstruction using voxel blocks. *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016*, 601–610. https://doi.org/10.1109/3DV.2016.68

Cho, K. (2014). On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. *ArXiv*, *5*, 1–9.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-

Decoder for Statistical Machine Translation. *ArXiv*, (September), 1–15. https://doi.org/10.3115/v1/D14-1179

Chung, J. (2015). Gated Feedback Recurrent Neural Networks. *ArXiv*, *37*, 1–9.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, 1–9. https://doi.org/10.1109/IJCNN.2015.7280624

Conneau, Alexis, Holger Schwenk, Y. L. C. (2017). Very Deep Convolutional Networks for Text Classification. *ArXiv*, (Jan), 1–10.

Davidson, D. W. (2016). Modeling Missing Data in Clinical Time Series with RNNs. *Proceedings of Machine Learning for Healthcare*, *58*(4), 725–737. https://doi.org/10.2307/1936209

Dawar, M. (2012). Fast Fuzzy Feature Clustering for Text Classification. *Academy and Industry Research Collaboration Centre, Computer Science and Information Technology*, (2), 167–172. https://doi.org/10.5121/csit.2012.2317

Dey, R., & Salemt, F. M. (2017). Gate-variants of Gated Recurrent Unit (GRU) neural networks. *Midwest Symposium on Circuits and Systems*, *2017-Augus*, 1597–1600. https://doi.org/10.1109/MWSCAS.2017.8053243

Ding, D., Zhang, M., Pan, X., Yang, M., & He, X. (2017). Modeling Extreme Events in Time Series Prediction. *Computers in Industry*, (111), 1114–1122. https://doi.org/10.1145/3292500.3330896

Do, H. H., Prasad, P. W. C., Maag, A., & Alsadoon, A. (2019). Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review. *Expert Systems with Applications*, *118*, 272–299. https://doi.org/10.1016/j.eswa.2018.10.003

Don, D. R., & Iacob, I. E. (2018). Fast Multi-class Classification using Support Vector Machines. *ArXiv Preprint ArXiv*, *2*, 1–19. Retrieved from http://arxiv.org/abs/1810.09828

Dong, B., & Wang, X. (2016). Comparison deep learning method to traditional methods using for network intrusion detection. *Proceedings of 2016 8th IEEE International Conference on Communication Software and Networks, ICCSN 2016*, 581–585. https://doi.org/10.1109/ICCSN.2016.7586590

Estévez, P. A., Tesmer, M., Perez, C. A., & Zurada, J. M. (2009). Normalized mutual information feature selection. *IEEE Transactions on Neural Networks*, *20*(2), 189–201. https://doi.org/10.1109/TNN.2008.2005601

Faris, H., Aljarah, I., & Mirjalili, S. (2016). Training feedforward neural networks using multi-verse optimizer for binary classification problems. *Applied Intelligence*, *45*(2), 322–332. https://doi.org/10.1007/s10489-016-0767-1

Feature, C., K, S. G., & Joseph, S. (2017). Text Classification by Augmenting Bag of Words (BOW) Representation with Co-occurrence Feature. *IOSR Journal of Computer Engineering (IOSR-JCE)*, *16*(1), 34–38. https://doi.org/10.9790/0661-16153438

Georgakopoulos, S. V, & Plagianakos, V. P. (2017). Random resampling in the one-versus-all strategy for handling multi-class problems. *In International Conference on Engineering Applications of Neural Networks*, 111–121. https://doi.org/10.1007/978-3-319-65172-9

Ghiassi, M., Olschimke, M., Moon, B., & Arnaudo, P. (2012). Automated text classification using a dynamic artificial neural network model. *Expert Systems with Applications*, *39*(12), 10967–10976. https://doi.org/10.1016/j.eswa.2012.03.027

Giraud, S., Thérouanne, P., & Steiner, D. D. (2018). Web accessibility: Filtering redundant and irrelevant information improves website usability for blind users. *International Journal of Human Computer Studies*, *111*(October 2017), 23–35. https://doi.org/10.1016/j.ijhcs.2017.10.011

Gonzalez, G., Ash, S. Y., Vegas-Sánchez-Ferrero, G., Onieva, J. O., Rahaghi, F. N., Ross, J. C., … Washko, G. R. (2018). Disease staging and prognosis in smokers using deep learning in chest computed tomography. *American Journal of Respiratory and Critical Care Medicine*, *197*(2), 193–203. https://doi.org/10.1164/rccm.201705-0860OC

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Generative Models. *Deep Learning*, 658–729. https://doi.org/10.1109/TPAMI.2017.2766142

Graves, A., Mohamed, A., & Hinton, G. (2013). SPEECH RECOGNITION WITH DEEP RECURRENT NEURAL NETWORKS Alex Graves, Abdel-rahman Mohamed and Geoffrey Hinton Department of Computer Science, University of Toronto. *IEEEInt.Conf.Acoust.,SpeechSignal Process.*, (3), 6645–6649.

Gultepe, E., Kamkarhaghighi, M., & Makrehchi, M. (2018). Latent Semantic Analysis Boosted Convolutional Neural Networks for Document Classification. *Proceedings - 2018 5th International Conference on Behavioral, Economic, and Socio-Cultural Computing, BESC 2018*, 93–98. https://doi.org/10.1109/BESC.2018.8697314

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. *34th International Conference on Machine Learning, ICML 2017*, *3*, 2130–2143.

Hao, Y., Sheng, Y., & Wang, J. (2019). Variant Gated Recurrent Units With Encoders to Preprocess Packets for Payload-Aware Intrusion Detection. *IEEE Access*, *7*, 49985–49998. https://doi.org/10.1109/ACCESS.2019.2910860

Hashimoto, K., Kontonatsios, G., Miwa, M., & Ananiadou, S. (2016). Topic detection using paragraph vectors to support active learning in systematic reviews. *Journal of Biomedical Informatics*, *62*, 59–65. https://doi.org/10.1016/j.jbi.2016.06.001

He, B., Guan, Y., & Dai, R. (2018). Convolutional Gated Recurrent Units for Medical Relation Classification. *Journal of Machine Learning Research*, (july), 118–127.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(9), 1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824

Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2018). NSE Stock Market Prediction Using Deep-Learning Models. *Procedia Computer Science*, *132*(Iccids), 1351–1362. https://doi.org/10.1016/j.procs.2018.05.050

Hochreiter, Sepp; (1997). Long Short Term Memory. *Neural Computation*, *9*(8), 1–32. https://doi.org/10.1144/GSL.MEM.1999.018.01.02

Hochreiter, Sepp, & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hu, B. G., & Xing, H. J. (2016). An optimization approach of deriving bounds between entropy and error from joint distribution: Case study for binary classifications. *Entropy*, *18*(2), 59. https://doi.org/10.3390/e18020059

Huang, C., Zhu, J., Liang, Y., Yang, M., Fung, G. P. C., & Luo, J. (2019). An efficient automatic multiple objectives optimization feature selection strategy for internet text classification. *International Journal of Machine Learning and Cybernetics*, *10*(5), 1151–1163. https://doi.org/10.1007/s13042-018-0793-x

Huang, L., Ding, S., Yu, S., Wang, J., & Lu, K. (2016). Chaos-enhanced Cuckoo search optimization algorithms for global optimization. *Applied Mathematical Modelling*, *40*(5–6), 3860–3875. https://doi.org/10.1016/j.apm.2015.10.052

Hughes, M., Li, I., Kotoulas, S., & Suzumura, T. (2017). Medical Text Classification using Convolutional Neural Networks. *Stud Health Technol Inform*, *235*(May), 246–250.

Ilhan, I., & Tezel, G. (2013). A genetic algorithm-support vector machine method with parameter optimization for selecting the tag SNPs. *Journal of Biomedical Informatics*, *46*(2), 328–340. https://doi.org/10.1016/j.jbi.2012.12.002

Ioffe, S., & Szegedy, C. (2015). Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift. *In Proceedings of the International Conference on Machine Learning, Lille, France*, (July), 6–11.

Islam, M. Z., Liu, J., Li, J., Liu, L., & Kang, W. (2019). A semantics aware random forest for text classification. *International Conference on Information and Knowledge Management, Proceedings*, 1061–1070. https://doi.org/10.1145/3357384.3357891

Jabreel, M., & Moreno, A. (2017). Target-dependent Sentiment Analysis of Tweets using a Bi-directional Gated Recurrent Unit. *In Proceedings Ofthe 13th International Conference on Web Information Systems AndTechnologies (WEBIST2017)*, (Webist), 80–87. https://doi.org/10.5220/0006299900800087

Jadhav, S., Kasar, R., Lade, N., Patil, M., & Kolte, S. (2019). Disease Prediction by Machine Learning from Healthcare Communities. *International Journal of Scientific Research in Science and Technology*, *5*, 8869–8869. https://doi.org/10.32628/ijsrst19633

Johnson, R., & Zhang, T. (2014). Effective Use of Word Order for Text Categorization with Convolutional Neural Networks, (2011). Retrieved from http://arxiv.org/abs/1412.1058

Justus, D., Brennan, J., Bonner, S., & McGough, A. S. (2019). Predicting the Computational Cost of Deep Learning Models. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 3873–3882. https://doi.org/10.1109/BigData.2018.8622396

Kakavand, M., Dabbagh, M., & Dehghantanha, A. (2018). Application of machine learning algorithms for android malware detection. *ACM International Conference Proceeding Series*, (November), 32–36. https://doi.org/10.1145/3293475.3293489

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A Convolutional Neural Network for Modelling Sentences. *Proceedings of the 52nd Annual Meeting of* *the Association for Computational Linguistics (Volume 1: Long Papers)*, 655–665. https://doi.org/10.3115/v1/P14-1062

Kim, D., Seo, D., Cho, S., & Kang, P. (2019). Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec. *Information Sciences*, *477*, 15–29. https://doi.org/10.1016/j.ins.2018.10.006

Knyaz, V. A., Vygolov, O., Kniaz, V. V., Vizilter, Y., Gorbatsevich, V., Luhmann, T., & Conen, N. (2018). Deep Learning of Convolutional Auto-Encoder for Image Matching and 3D Object Reconstruction in the Infrared Range. *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, *2018-Janua*, 2155–2164. https://doi.org/10.1109/ICCVW.2017.252

Konda, K., Memisevic, R., & Krueger, D. (2014). Zero-bias autoencoders and the benefits of co-adapting features. *ArXiv Preprint ArXiv:1402.3337*, (2011), 1–11. Retrieved from http://arxiv.org/abs/1402.3337

Koppel, A., Warnell, G., Stump, E., & Ribeiro, A. (2019). Parsimonious online learning with kernels via sparse projections in function space. *Journal of Machine Learning Research*, *20*(1), 1–44.

Korde, V., & Mahender, C. N. (2012). Text Classification and Classifiers: A Survey. *International Journal of Artificial Intelligence & Applications*, *3*(2), 85–99. https://doi.org/10.5121/ijaia.2012.3208

Kowsari, K., Brown, D. E., Heidarysafa, M., Meimandi, K. J., Gerber, M. S., & Barnes, L. E. (2017). HDLTex: Hierarchical Deep Learning for Text Classification. *In 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, (Octuber), 364–371. https://doi.org/10.1109/ICMLA.2017.0-134

Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. *Information (Switzerland)*, *10*(4), 1–68. https://doi.org/10.3390/info10040150

Kulkarni, A. R., Tokekar, V., & Kulkarni, P. (2012). Identifying context of text documents using Naïve Bayes classification and Apriori association rule mining. *2012 CSI 6th International Conference on Software Engineering, CONSEG 2012*, 1–4. https://doi.org/10.1109/CONSEG.2012.6349477

Kusumaningrum, R., Wiedjayanto, M. I. A., Adhy, S., & Suryono. (2017).

Classification of Indonesian news articles based on Latent Dirichlet Allocation. *Proceedings of 2016 International Conference on Data and Software Engineering, ICoDSE 2016*, 1–5. https://doi.org/10.1109/ICODSE.2016.7936106

Lafferty, J., MeCallum, A., & Pereira, F. (2017). Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. *Majalah Ilmiah Pengkajian Industri*, *11*(1), 1–84. https://doi.org/10.29122/mipi.v11i1.2792

Lai, S., Xu, L., Liu, K., & Zhao, J. (2018). Recurrent Convolutional Neural Networks for Text Classification. *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2267–2273.

Lamiya, K., & Mohan, A. (2018). Classification of Short Text Using Various Preprocessing Techniques: An Empirical Evaluation. *ArXiv*, *3*(January), 161–168. https://doi.org/10.1007/978-981-10-8633-5

Le, T. T. H., Kim, J., & Kim, H. (2017). Classification performance using gated recurrent unit Recurrent Neural Network on energy disaggregation. *Proceedings - International Conference on Machine Learning and Cybernetics*, *1*(November 2017), 105–110. https://doi.org/10.1109/ICMLC.2016.7860885

Lei Geng, Haiyue Wang, Z. X. (2019). Fully Convolutional Network With Gated Recurrent Unit for Hatching Egg Activity Classification. *IEEE Access*, *7*, 92378–92387.

Liu, H., & Cocea, M. (2017). Semi-random partitioning of data into training and test sets in granular computing context. *Granular Computing*, *2*(4), 357–386. https://doi.org/10.1007/s41066-017-0049-2

Liu, Lei. (2016). Hierarchical learning for large multi-class network classification. *Proceedings - International Conference on Pattern Recognition*, *0*, 2307–2312. https://doi.org/10.1109/ICPR.2016.7899980

Liu, Liping, Wang, N., Chen, Z., & Guo, L. (2018). A Novel Spectrum Scheduling Scheme with Ant Colony Optimization Algorithm. *Algorithms*, *11*(2), 16. https://doi.org/10.3390/a11020016

Liu, M., Zhang, D., Chen, S., & Xue, H. (2016). Joint binary classifier learning for ECOC-based multi-class classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(11), 2335–2341. https://doi.org/10.1109/TPAMI.2015.2430325

Liu, P., Qiu, X., & Huang, X. (2016). Recurrent Neural Network for Text Classification with Multi-Task Learning. *Proceedings of the 25th International Joint Conference on Artificial Intelligence IJCAI-16*, to appear. Retrieved from http://arxiv.org/abs/1605.05101

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). Neurocomputing A survey of deep neural network architectures and their applications ☆. *Neurocomputing*, *234*(December 2016), 11–26. https://doi.org/10.1016/j.neucom.2016.12.038

Loh, W. Y., Eltinge, J., Cho, M. J., & Li, Y. (2019). Classification and regression trees and forests for incomplete data from sample surveys. *Statistica Sinica*, *29*(1), 431–453. https://doi.org/10.5705/ss.202017.0225

Long, M., Zhu, H., Wang, J., & Jordan, M. I. (2017). Deep transfer learning with joint adaptation networks. *34th International Conference on Machine Learning, ICML 2017*, *5*, 3470–3479.

Lu, G., Gan, J., Yin, J., Luo, Z., Li, B., & Zhao, X. (2020). Multi-task learning using a hybrid representation for text classification. *Neural Computing and Applications*, *32*(11), 6467–6480. https://doi.org/10.1007/s00521-018-3934-y

Lukasik, M., & Cohn, T. (2016). Convolution kernels for discriminative learning from streaming text. *30th AAAI Conference on Artificial Intelligence, AAAI 2016*, 2757–2763.

Luo, L. xia. (2019). Network text sentiment analysis method combining LDA text representation and GRU-CNN. *Personal and Ubiquitous Computing*, *23*(3–4), 405–412. https://doi.org/10.1007/s00779-018-1183-9

Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). Learning Word Vectors for Sentiment Analysis. *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics, Hum. Lang. Technol.*, *1*, 142–150.

Malhotra, P., Vig, L., Agarwal, P., & Shroff, G. (2017). TimeNet: Pre-trained deep recurrent neural network for time series classification.

Manevitz, L. ., & Yousef, M. (2011). One-Class SVMs for Document Classification. *Contraception Fertilite Sexualite*, *7*(3), 215–222.

Marini, F., & Walczak, B. (2015). Particle swarm optimization (PSO). A tutorial. *Chemometrics and Intelligent Laboratory Systems*, *149*, 153–165. https://doi.org/10.1016/j.chemolab.2015.08.020

Metwally, A. A., Yu, P. S., Reiman, D., Dai, Y., Finn, P. W., & Perkins, D. L.

(2019). Utilizing longitudinal microbiome taxonomic profiles to predict food allergy via long short-term memory networks. *PLoS Computational Biology*, *15*(2), 1–16. https://doi.org/10.1371/journal.pcbi.1006693

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2020). Deep Learning Based Text Classification: A Comprehensive Review. *ArXiv Preprint ArXiv:2004.03705*, *1*(1), 1–42. Retrieved from http://arxiv.org/abs/2004.03705

Moazenzadeh, R., Mohammadi, B., Shamshirband, S., & Chau, K. W. (2018). Coupling a firefly algorithm with support vector regression to predict evaporation in northern iran. *Engineering Applications of Computational Fluid Mechanics*, *12*(1), 584–597. https://doi.org/10.1080/19942060.2018.1482476

Naik, V. A., & Desai, A. A. (2018). Online Handwritten Gujarati Numeral Recognition Using Support Vector Machine. *International Journal of Computer Sciences and Engineering*, *6*(9), 416–421. https://doi.org/10.26438/ijcse/v6i9.416421

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, *2*(1), 1–21. https://doi.org/10.1186/s40537-014-0007-7

Nam, J., Kim, J., Loza Mencía, E., Gurevych, I., & Fürnkranz, J. (2014). Large-scale multi-label text classification - Revisiting neural networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8725 LNAI*(PART 2), 437–452. https://doi.org/10.1007/978-3-662-44851-9_28

Nayak, J., Naik, B., & Behera, H. S. (2015). A Comprehensive Survey on Support Vector Machine in Data Mining Tasks: Applications & Challenges. *International Journal of Database Theory and Application*, *8*(1), 169–186. https://doi.org/10.14257/ijdta.2015.8.1.18

Nguyen, D., Nguyen, C., Duong-Ba, T., Nguyen, H., Nguyen, A., & Tran, T. (2017). Joint network coding and machine learning for error-prone wireless broadcast. In *2017 IEEE 7th Annual Computing and Communication Workshop and Conference, CCWC 2017*. https://doi.org/10.1109/CCWC.2017.7868415

Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., López García, Á., Heredia, I., … Hluchý, L. (2019). Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. *Artificial Intelligence Review*, *52*(1), 77–124. https://doi.org/10.1007/s10462-018-09679-z

Nikam, S. S. (2015). A Comparative Study of Classification Techniques in Data Mining Algorithms. *International Journal of Modern Trends in Engineering & Research*, *4*(7), 58–63. https://doi.org/10.21884/ijmter.2017.4211.vxayk

Nilashi, M., Ibrahim, O., & Bagherifard, K. (2018). A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques. *Expert Systems with Applications*, *92*, 507–520. https://doi.org/10.1016/j.eswa.2017.09.058

Olson, D. L., & Wu, D. (2016). Predictive Data Mining Models. *Machine Learning and Soft Computing*, 105. https://doi.org/10.1007/978-981-10-2543-3

Onan, A., Korukoğlu, S., & Bulut, H. (2016). Ensemble of keyword extraction methods and classifiers in text classification. *Expert Systems with Applications*, *57*, 232–247. https://doi.org/10.1016/j.eswa.2016.03.045

Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., … Iyengar, S. S. (2018). A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys*, *51*(5), 23–51. https://doi.org/10.1145/3234150

Qdwxudo, H. Z., Surfhvvlqj, O., Hduqlqj, H. H. S., Phfkdqlvp, W., & Yhfwru, R. U. G. (2018). A Short Text Sementic Classification Method for Power Grid Service Based on Attention Gated Recurrent Unit Neural Network. *The 2018 5th International Conference on Systems and Informatics (ICSAI 2018)*, 1105–1110.

Rajendran, G., Chitturi, B., & Poornachandran, P. (2018). ScienceDirect ScienceDirect Stance-In-Depth Stance-In-Depth Approach Deep Neural Approach to to Stance Stance Classification. *Procedia Computer Science*, *132*(Iccids), 1646–1653. https://doi.org/10.1016/j.procs.2018.05.132

Rana, R., Epps, J., Jurdak, R., Li, X., Goecke, R., Brereton, M., & Soar, J. (2016). Gated Recurrent Unit (GRU) for Emotion Classification from Noisy Speech. *ArXiv Preprint ArXiv:1612.07778*, 1–9.

Ravanelli, M., Brakel, P., Omologo, M., & Bengio, Y. (2017). Improving speech recognition by revising gated recurrent units. *ArXiv*, 2–6.

Raziff, A. R., Sulaiman, M. N., Mustapha, N., Perumal, T., & Mohd Pozi, M. S. (2017). Multiclass classification method in handheld based smartphone gait identification. *Journal of Telecommunication, Electronic and Computer*

Engineering, *9*(2–12), 59–65.

Salman, H., Grover, J., & Shankar, T. (2018). Supervised Learning in Multilayer Spiking Neural Networks. *Neural Computation*, *1541*, 1514–1541. https://doi.org/10.1162/NECO

Samarawickrama, A. J. P., & Fernando, T. G. I. (2018). A recurrent neural network approach in predicting daily stock prices an application to the Sri Lankan stock market. *2017 IEEE International Conference on Industrial and Information Systems, ICIIS 2017 - Proceedings*, *2018-Janua*, 1–6. https://doi.org/10.1109/ICIINFS.2017.8300345

Saqib, S. M., Kundi, F. M., & Ahmad, S. (2018). Unsupervised Learning Method for Sorting Positive and Negative Reviews Using LSI (Latent Semantic Indexing) with Automatic Generated Queries. *IJCSNS International Journal of Computer Science and Network Security*, *18*(1), 56–62.

Sayed, S., Nassef, M., Badr, A., & Farag, I. (2019). A Nested Genetic Algorithm for feature selection in high-dimensional cancer Microarray datasets. *Expert Systems with Applications*, *121*, 233–243. https://doi.org/10.1016/j.eswa.2018.12.022

Sen, A., Islam, M. M., Murase, K., & Yao, X. (2016). Binarization with Boosting and Oversampling for Multiclass Classification. *IEEE Transactions on Cybernetics*, *46*(5), 1078–1091. https://doi.org/10.1109/TCYB.2015.2423295

Septiawan, W. M., & Endah, S. N. (2019). Suitable Recurrent Neural Network for Air Quality Prediction with Backpropagation Through Time. *2018 2nd International Conference on Informatics and Computational Sciences, ICICoS 2018*, 196–201. https://doi.org/10.1109/ICICOS.2018.8621720

Sharif, W., Samsudin, N. A., Deris, M. M., & Aamir, M. (2017). Improved relative discriminative criterion feature ranking technique for text classification. *International Journal of Artificial Intelligence*, *15*(2), 61–78.

Sharif, W., Tun, U., Onn, H., Tri, I., & Yanto, R. (2019). An Optimised Support Vector Machine with Ringed Seal Search Algorithm for Efficient Text Classification. *Journal of Engineering Science and Technology*, *14*(3), 1601–1613.

Sharma, N., & Singh, M. (2017). Modifying Naive Bayes classifier for multinomial text classification. *2016 International Conference on Recent Advances and Innovations in Engineering, ICRAIE 2016*, 1–7.

https://doi.org/10.1109/ICRAIE.2016.7939519

Shen, D., Sun, J. T., Li, H., Yang, Q., & Chen, Z. (2007). Document summarization using conditional random fields. *IJCAI International Joint Conference on Artificial Intelligence*, 2862–2867.

Shen, G., Tan, Q., Zhang, H., Zeng, P., & Xu, J. (2018). ScienceDirect ScienceDirect Deep Learning with Gated Recurrent Unit Networks for Financial Deep Learning with Gated Recurrent Unit Networks for Financial Sequence Predictions Sequence Predictions. *Procedia Computer Science*, *131*, 895–903. https://doi.org/10.1016/j.procs.2018.04.298

Shickel, B., Tighe, P. J., & Bihorac, A. (2018). Deep EHR : A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. *IEEE Journal of Biomedical and Health Informatics*, *22*(5), 1589–1604. https://doi.org/10.1109/JBHI.2017.2767063

Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A Deep Learning Approach to Network Intrusion Detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *2*(1), 41–50. https://doi.org/10.1109/tetci.2017.2772792

Siddiqui, M. A. (2016). An empirical evaluation of text classification and feature selection methods. *Artificial Intelligence Research*, *5*(2), 70–81. https://doi.org/10.5430/air.v5n2p70

Singh, A., Thakur, N., & Sharma, A. (2016). A review of supervised machine learning algorithms. *Proceedings of the 10th INDIACom; 2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom 2016*, 1310–1315.

Sivakumar, B., & Srilatha, K. (2016). A novel method to segment blood vessels and optic disc in the fundus retinal images. *Research Journal of Pharmaceutical, Biological and Chemical Sciences*, *7*(3), 365–373. https://doi.org/10.15680/IJIRCCE.2016.

Socher, R., Huval, B., Manning, C. D., & Ng, A. Y. (2012). Semantic Compositionality through Recursive Matrix-Vector Spaces. *In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, (July), 1201–1211.

Socher, R., Perelygin, A., & Wu, J. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of the …*, (October), 1631–1642. https://doi.org/10.1371/journal.pone.0073791

Srividhya, V., & Anitha, R. (2010). Evaluating Preprocessing Techniques in Text Categor ization. *International Journal of Computer Science and Application*, *47*(April), 49–51.

Sun, Y., Zhang, M., Sun, Z., & Tan, T. (2018). Demographic Analysis from Biometric Data: Achievements, Challenges, and New Frontiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(2), 332–351. https://doi.org/10.1109/TPAMI.2017.2669035

Tang, B., He, H., Baggenstoss, P. M., & Kay, S. (2016). A Bayesian Classification Approach Using Class-Specific Features for Text Categorization. *IEEE Transactions on Knowledge and Data Engineering*, *28*(6), 1602–1606. https://doi.org/10.1109/TKDE.2016.2522427

Tang, D., Qin, B., & Liu, T. (2015). Document Modeling with Gated Recurrent Neural Network for Sentiment Classification. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, (September), 1422–1432. https://doi.org/10.18653/v1/D15-1167

Tian, Z., Rong, W., Shi, L., Liu, J., & Xiong, Z. (2018). *Attention Aware Bidirectional Gated Recurrent Unit Based Framework for Sentiment Analysis* (Vol. 3). Springer International Publishing. https://doi.org/10.1007/978-3-319-99365-2

Vairaprakash Gurusamy Madurai, S. K. M. (2014). Preprocessing Techniques for Text Mining. *Conference ICLR*, (October), 1–7.

Varma, M. K. S. (2016). Pixel-based Classification Using Support Vector Machine Classifier. *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, 51–55. https://doi.org/10.1109/IACC.2016.20

Velliangiri, S., Alagumuthukrishnan, S., & Thankumar Joseph, S. I. (2019). A Review of Dimensionality Reduction Techniques for Efficient Computation. *Procedia Computer Science*, *165*, 104–111. https://doi.org/10.1016/j.procs.2020.01.079

Vijayarani, S., Ilamathi, J., & Nithya, M. (2015). Preprocessing Techniques for Text Mining - An Overview. *International Journal of Computer Science & Communication Networks*, *5*(1), 7–16. Retrieved from http://www.ijcscn.com/Documents/Volumes/vol5issue1/ijcscn2015050102.pdf

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep

Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, *2018*, 1–13. https://doi.org/10.1155/2018/7068349

Wang, B., Wang, L., Wei, Q., & Liu, L. (2018). TextZoo, a New Benchmark for Reconsidering Text Classification. *ArXiv Preprint ArXiv:1802.03656*, 1–2. Retrieved from http://arxiv.org/abs/1802.03656

Wang, D., & Mao, K. (2018). Multimodal object classification using bidirectional gated recurrent unit networks. *Proceedings - 2018 IEEE 3rd International Conference on Data Science in Cyberspace, DSC 2018*, 685–690. https://doi.org/10.1109/DSC.2018.00109

Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, *119*, 3–11. https://doi.org/10.1016/j.patrec.2018.02.010

Wang, Y., Yao, H., & Zhao, S. (2016). Auto-encoder based dimensionality reduction. *Neurocomputing*, *184*(November), 232–242. https://doi.org/10.1016/j.neucom.2015.08.104

Wang, Z., & Qu, Z. (2017). Research on Web Text Classification Algorithm Based on Improved CNN and SVM. *IEEE*, 1958–1961.

Wu, Q., Teney, D., Wang, P., Shen, C., Dick, A., & van den Hengel, A. (2017). Visual question answering: A survey of methods and datasets. *Computer Vision and Image Understanding*, *163*, 21–40. https://doi.org/10.1016/j.cviu.2017.05.001

Wu, W., Liao, W., Miao, J., & Du, G. (2019). Using gated recurrent unit network to forecast short-term load considering impact of electricity price. *Energy Procedia*, *158*, 3369–3374. https://doi.org/10.1016/j.egypro.2019.01.950

Xiao, Y., & Cho, K. (2016). Efficient Character-level Document Classification by Combining Convolution and Recurrent Layers. *ArXiv*, *1602*(00367). Retrieved from http://arxiv.org/abs/1602.00367

Xing, Y., & Xiao, C. (2019). A GRU Model for Aspect Level Sentiment Analysis. *Journal of Physics: Conference Series*, *1302*, 032042. https://doi.org/10.1088/1742-6596/1302/3/032042

Xu, Jiacheng, Chen, D., Qiu, X., & Huang, X. (2016). Cached Long Short-Term Memory Neural Networks for Document-Level Sentiment Classification. *ArXiv*, 1610–1620.

Xu, Jie, Xu, C., Zou, B., Tang, Y. Y., Peng, J., & You, X. (2019). New Incremental

Learning Algorithm with Support Vector Machines. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *49*(11), 2230–2241. https://doi.org/10.1109/TSMC.2018.2791511

Xu, S. (2018). Bayesian Naïve Bayes classifiers to text classification. *Journal of Information Science*, *44*(1), 48–59. https://doi.org/10.1177/0165551516677946

Xue, B., Zhang, M., Browne, W. N., & Yao, X. (2016). A Survey on Evolutionary Computation Approaches to Feature Selection. *IEEE Transactions on Evolutionary Computation*, *20*(4), 606–626. https://doi.org/10.1109/TEVC.2015.2504420

Yan, Z., Yang, K., Wang, Z., Yang, B., Kaizuka, T., & Nakano, K. (2019). Time to lane change and completion prediction based on Gated Recurrent Unit Network. *IEEE Intelligent Vehicles Symposium, Proceedings*, *2019-June*(June), 102–107. https://doi.org/10.1109/IVS.2019.8813838

YangZhenYu, J. S. (2017). A Study on Text Classification Based on Stacked. *2017 First International Conference on Electronics Instrumentation & Information Systems*, 1–6.

Yeh, C. K., Wu, W. C., Ko, W. J., & Wang, Y. C. F. (2017). Learning deep latent spaces for multi-label classification. *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, 2838–2844.

Yenter, A., & Verma, A. (2017). Deep CNN-LSTM with combined kernels from multiple branches for IMDb review sentiment analysis. *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2017*, *2018-Janua*, 540–546. https://doi.org/10.1109/UEMCON.2017.8249013

Yi, J., Zhang, Y., Zhao, X., & Wan, J. (2017). A Novel Text Clustering Approach Using Deep-Learning Vocabulary Network. *Mathematical Problems in Engineering*, *2017*. https://doi.org/10.1155/2017/8310934

Yin, W., Kann, K., Yu, M., & Schütze, H. (2017). Comparative Study of CNN and RNN for Natural Language Processing. https://doi.org/10.14569/IJACSA.2017.080657

Yogatama, D., Dyer, C., Ling, W., & Blunsom, P. (2017). Generative and Discriminative Text Classification with Recurrent Neural Networks. *ArXiv*, (May), 1–9.

Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing [Review Article]. *IEEE Computational Intelligence Magazine*, *13*(3), 55–75. https://doi.org/10.1109/MCI.2018.2840738

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8689 LNCS*(PART 1), 818–833. https://doi.org/10.1007/978-3-319-10590-1_53

Zhai, S., & Zhang, Z. (2016). Semisupervised autoencoder for sentiment analysis. *30th AAAI Conference on Artificial Intelligence, AAAI 2016*, *13902*, 1394–1400.

Zhang, C., Wang, H., Liu, Y., Wu, D., Liao, Y., & Wang, B. (2008). Automatic keyword extraction from documents using conditional random fields. *Journal of Computational Information Systems*, *4*(3), 1169–1180.

Zhang, J., Liu, F., Xu, W., & Yu, H. (2019). Feature fusion text classification model combining CNN and BiGRU with multi-attention mechanism. *Future Internet*, *11*(11), 237. https://doi.org/10.3390/fi11110237

Zhang, L., Zhou, Y., Duan, X., & Chen, R. (2018). A Hierarchical multi-input and output Bi-GRU Model for Sentiment Analysis on Customer Reviews. *IOP Conference Series: Materials Science and Engineering*, *322*(6). https://doi.org/10.1088/1757-899X/322/6/062007

Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information Fusion*, *42*(November 2017), 146–157. https://doi.org/10.1016/j.inffus.2017.10.006

Zhang, Q., Yang, L. T., Chen, Z., Li, P., & Bu, F. (2019). An Adaptive Dropout Deep Computation Model for Industrial IoT Big Data Learning with Crowdsourcing to Cloud Computing. *IEEE Transactions on Industrial Informatics*, *15*(4), 2330–2337. https://doi.org/10.1109/TII.2018.2791424

Zhang, X., & LeCun, Y. (2018). Text detection in Arabic news video based on SWT operator and convolutional auto-encoders. *In 2016 12th IAPR Workshop on Document Analysis Systems (DAS)*, 13–18. Retrieved from http://arxiv.org/abs/1802.01817

Zhang, Y., Lu, W., Ou, W., Zhang, G., Zhang, X., Cheng, J., & Zhang, W. (2019). Chinese medical question answer selection via hybrid models based on CNN and GRU. *Multimedia Tools and Applications*, *79*(21–22), 14751–14776.

https://doi.org/10.1007/s11042-019-7240-1

Zhang, Z. L., Luo, X. G., García, S., Tang, J. F., & Herrera, F. (2017). Exploring the effectiveness of dynamic ensemble selection in the one-versus-one scheme. *Knowledge-Based Systems*, *125*, 53–63. https://doi.org/10.1016/j.knosys.2017.03.026

Zhao, Z., Li, H., Zhao, R., & Wang, X. (2016). Crossing-line crowd counting with two-phase deep neural networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9912 LNCS*, 712–726. https://doi.org/10.1007/978-3-319-46484-8_43

Zhou, G. B., Wu, J., Zhang, C. L., & Zhou, Z. H. (2016). Minimal gated unit for recurrent neural networks. *International Journal of Automation and Computing*, *13*(3), 226–234. https://doi.org/10.1007/s11633-016-1006-2

VITA

The author of this thesis is Muhammad Zulqarnain. He was born in January 03, 1993, in Bahawalpur, Pakistan. He completed his Bachelor's degree in department of Computer Science and Information Technology from Government Sadiq Egerton College Bahawalpur in 2011 and graduated in Computer Science from The Islamia University Bahawalpur in 2014. He completed his MS degree in Computer Science from National College Business Administration & Economics in 2016. During his MS study, he served as lecturer of computer science in City College Bahawalpur. To find his passion for research and improve the knowledge in Machine leaning and Deep learning, he has completed his PhD research at Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM), Johor, Malaysia. Additionally, he has published several journals and conference proceedings. His research interests are in the area of text classification, sentiment analysis, artificial intelligence, machine learning and deep learning.