Detecting Bengali Spam SMS Using Recurrent Neural Network

Md. Mohsin Uddin[1], Monica Yasmin[1], M Saddam Hossain Khan[1], Md Istianatur Rahman[2], and Tabassum Islam[1]
[1] East West University, Dhaka, 1212, Bangladesh
[2] World University of Bangladesh
Email: mmuddin@ewubd.edu, {y.monicamym, mshk03.edu, mdi7357, tabassum.islam009}@gmail.com

Abstract—SMS is being spammed if the sender sends it to the targeted users to gain important personal information. If targeted users respond with personal information, it will be a great opportunity for the sender to grab their desired goal. Now, this phenomenon increases rapidly and Machine Learning (ML) is mostly used to classify this problem. In terms of Bangladesh, email spam detection is common but detecting SMS spam with the Bengali dataset is completely new as a research problem. This research is taken part to detect Bengali spam SMS using traditional Machine Learning algorithms along with Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). Then, the performances of all algorithms are compared to find the best among them. The highest testing accuracy rate is gained by both LSTM and GRU, which is 99%. To the best of our knowledge, this work is the first to apply the deep learning algorithms LSTM and GRU for detecting Bengali spam. Besides, a comparative analysis is performed with some traditional supervised ML algorithms and deep learning algorithms. Moreover, the effects of various activation functions and optimizers are also experimented on LSTM and GRU deep learning algorithms. ADAGRAD optimizer gains the best accuracy over RMSPROP, ADAMAX, ADADELTA and SGD. Finally, the best combinations of deep learning algorithms, activation functions, and optimizers are proposed based on experimental analysis.

Index Terms—SMS spam, RNN, LSTM, GRU, ADAGRAD, Machine Learning, Logistic regression, SVM, Naïve Bayes.

I. INTRODUCTION

The usage of Short message service (SMS) is increasing day-by-day. The cost of sending SMS has been dropped significantly. So, sending spam SMS to mass people is becoming easier in every mother language. Detecting spam SMS and filtering them have been acknowledged significantly in the last few years. Some major purposes of SMS spam are to spread fake and inappropriate content, advertise products and exploit targeted user's information by the means of harming [1]. Spam SMS not only harasses people but also can bring one's life in a big trouble. Spam SMS hampered privacy and manipulated people to give their personal information to unknowns.

Manuscript received September 2, 2019; revised March 6, 2020.
Corresponding author email: mmuddin@ewubd.edu
doi:10.12720/jcm.15.4.325-331

Everyone wants a safe and happy private life and to protect individuals' life privacy it is in the high necessity to stop spam SMS. In ML, spam filtering is now an interesting research area because every day new techniques arrive and researchers need to think about innovative ideas. This improves the filtering techniques and protects the people who suffer daily because of these spams. One of the major issues in detecting spam SMS is the scarce nature of dataset in English and non-English languages. Several works have been done to detect spam of emails, social tagging and so on but a few are conducted to detect the spam SMS [2] [3].

This research aims to create techniques using ML algorithms that can detect Bengali spam SMS. This paper compares the accuracy rates of ML algorithms, classifies them and explores the best algorithms that can detect SMS spam. Previous works of SMS spam detection have been done on the prepared English dataset [4] [5] but this research followed a different approach. Instead of building a model with existing Bengali datasets, a new SMS spam dataset for Bengali language is prepared.

Working with the Bengali dataset is challenging. As Bengali SMS spam detection is a completely new topic, there is no Systematic Literature Review (SLR) on this work [3]. Supervised ML algorithms including Naïve Bayes [6], Support Vector Machine (SVM) [7], Logistic Regression [8], two variants of Recurrent Neural Network LSTM [9] and GRU [10] are applied to detect spam SMS to filter them. In LSTM and GRU, different combinations of activation functions and optimizers are used to learn parameters and to adapt best performances. Activation functions such as Sigmoid, ReLU, and TanH and optimizers such as RMSPROP, ADAGRAD, ADAMAX, Stochastic Gradient Descent (SGD), and ADADELTA are used to perform a comparative analysis to find the best combinations of hidden layer activation functions as well as optimization algorithms on LSTM and GRU algorithms.

This paper has been organized in five sections which are as follows: Section I introduces the SMS spam detection problem and how important it is to resolve the problem. Section II describes the related works of solving SMS spam detection problem and the methods that researchers have applied so far, which build the idea to work in this research. Section III describes the design methodology applied in this work. Section IV represents the result part and analyzes each result with explanations and figures where needed. Section V finishes this research with the conclusion and future work, which is important for new researchers to work in this area.

II. Related Works

A comprehensive review on recent SMS spam detection challenges and filtering techniques are mention in paper [1]. The available datasets and future directions in solving the problem are also analyzed. Some researchers focused on to resolve SMS spam issues and tried to find the Systematic Literature Review (SLR) that is missing in past researches [3]. Some researchers collected and merged some datasets to build a large spam SMS dataset for English [11]. In their dataset, all the near-duplicate data were removed. Almeida et el. [5] worked on English spam SMS dataset and performed comparative analysis on a long list of traditional ML algorithms. In their work, they showed that SVM got better result than other ML techniques in terms of accuracy. Previous researches identified many problems [12], [13], which are not well highlighted and resolved in an effective way. The paper [2] worked with the experimental results. They applied two variations of RNN algorithm in detecting SMS spam. The Paper [14] represented the design and implementation of SMS spam, which is application based. According to user preferences the paper implemented their spam filtering design. Other researches tried to find the issues about E-mail spamming and their problems. Paper [12] worked with the hash algorithm and Naïve Bayesian algorithm to detect spam emails.

A comprehensive survey is performed in paper [4] where various datasets are analyzed. Their approach applied SVM and the Bayesian network to Classify spam SMS. The paper described that because of the kernel problem, higher dimensions were not used to resolve the SMS spam-filtering problem. In a review paper [4], researchers shown the existing work in this area with the previous papers and summarized them in a good way. Paper [15] extracts the features using static, temporal, and network feature extraction. Two supervised algorithm SVM and KNN were used to detect spam. They extracted the result part with performance matrices.

The volume of SMS messages increased tremendously due to mass usage of smart phones in the whole world. Previously, mobile phone user used to send SMS only in English language. Now, with the introduction of new international language support, SMS messages are being sent in various languages. Bengali language is also a widely used language for sending SMS. So a dataset preparation in Bengali language for detecting SMS spam is becoming more feasible day-by-day. Some spam detection researches were performed in Bengali dataset for detecting spam text from social networking application like Facebook [16]. A very little research works are performed in detecting Bengali SMS spam.

To the best of our knowledge, this work is the first to apply LSTM and GRU for detecting Bengali spam. Various experiments are performed with activation functions and optimizers to find the best combination in deep learning algorithms. In this paper, various activation functions such as ReLU, Sigmoid, and TanH are applied. Various experiments are run with Optimizers named: RMSPROP, ADAGRAD, ADAMAX, SGD, and ADADELTA to determine the best optimizer for Bengali SMS spam detection task.

III. Methodology

A. *Dataset and Pre-processing*

Previous spam detection experiments are performed on many existing Non-Bengali datasets but the work in this paper is performed with a completely new Bengali spam dataset. The dataset has been prepared by collecting SMS messages from mobile phones. The sample data in the Bengali SMS spam dataset is illustrated in Table I.

TABLE I: Sample Data of Bengali SMS Spam Dataset

Sample SMS Data	Class
হাজার টাকা ফ্রিতে পেতে ৫০ টাকা বিকাশ করুন।	Spam
বাংলা আমার মাতৃভাষা।	Ham

Three human annotators whose mother tongue is Bengali, annotate the dataset with Spam and Ham tags. If annotators assign two different tags for a data, then majority-voting technique is applied to determine the final tag for the data. In the prepared Bengali SMS spam dataset, the spam texts comprise of 22% and the ham texts comprise of 78%. The corpus is randomly shuffled and divided into two parts. Then, the first 80% of SMS messages are separated for training and the rest 20% of messages are used for testing. As to implement supervised ML algorithms, the dataset needs to be cleaned and pre-processed.

As discussed earlier, the prepared dataset has two different labels, either the SMS is labeled as Spam or Ham. So, the problem is considered as a binary classification problem. The approach is to use Natural Language Processing (NLP) concepts to convert each SMS as a vector of features [17].

B. *Feature Extraction*

For building ML models, word N-Gram features are extracted. Three variations of N-Gram features have been extracted and applied for building ML models to investigate their individual importance as feature. N-Gram features named Unigram, Bigram, and Trigram are extracted and represented. Unigram is to work with one word where bigram and trigram work with two and three words, respectively. The sample tokenization patterns are illustrated in Table II. Then, all the extracted features are encoded to feature vector by using feature encoding technique.

TABLE II: SAMPLE TOKENIZATION PATTERN

N-Gram	Message	Tokenized
Uni-Gram	বাংলা আমার মাতৃভাষা।	'বাংলা', 'আমার', 'মাতৃভাষা'
Bi-Gram	বাংলা আমার মাতৃভাষা।	'বাংলা আমার', 'আমার মাতৃভাষা'
Tri-Gram	বাংলা আমার মাতৃভাষা।	'বাংলা আমার মাতৃভাষা'

C. Features to Feature Vector

After pre-processing the dataset, the features are encoded to feature vector. TF-IDF vectorization technique is used for this encoding purpose. It transforms a collection of raw text to a collection of feature vectors, which are then fed into the ML algorithms for training.

D. Baseline Algorithms

Some ML algorithms such as SVM, Naïve Bayes, and Logistic Regression are taken as baseline algorithms [5] so that our deep learning based designed models can be properly compared. Scikit-learn [18] toolkit is used to implement these baseline algorithms.

i. Support Vector Machine (SVM) [7]: SVM is used for a broader classification margin. It uses the kernel trick technique to find the optimal boundary between possible outputs. In this research, it is applied for the same validation [15]. Polynomial kernel and c=3 are used as SVM settings.

ii. Naïve Bayes [6]: It is a probabilistic classifier, which assumes that the features are independent to each other. It is used as a baseline for many ML related research. The multinomial Naïve Bayes (alpha=2.0, class_prior=None, fir_prior=True) is used in our research.

iii. Logistic Regression [8]: The logistic regression model uses sigmoid function to learn from a dataset. The default Scikit-learn function is applied for implementation.

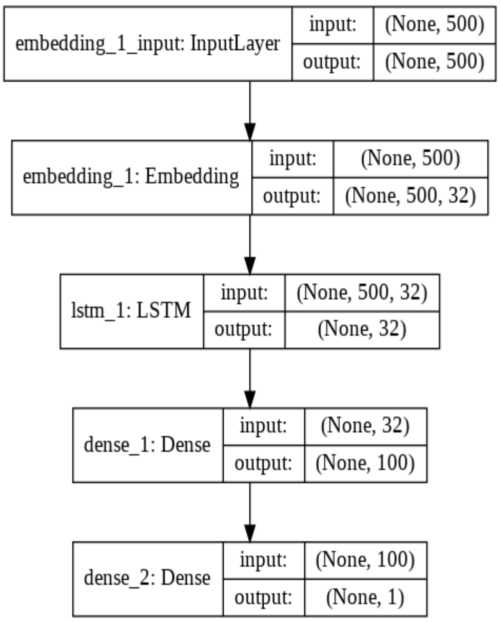

Fig. 1. Implementation of LSTM recurrent network

E. Deep Learning Algorithms

i. Recurrent Neural Network (RNN): It is a deep learning algorithm whose internal state can be used as memory to process sequential inputs [19]. In the deep learning settings, tokenization, padding, and embedding techniques are applied as pre-processing steps. Tokens are transformed into sequence. Maximum length of sequence is 500 and vocabulary size is 10000. Similar to baseline ML approaches 80:20 dataset-splitting formula is used for training and testing.

ii. Long Short-Term Memory (LSTM) [9]: RNN's most often encounter is solved by the improved version of this deep learning algorithm, called Long Short-Term Memory. RNN has a limitation called vanishing gradient problem [20] [21] whereas LSTM resolves the issue. It can trace back to several states and see what happened which ultimately results in taking efficient results. Our LSTM based design model is illustrated in Fig. 1. Like our vanilla RNN based model, similar pre-processing steps are applied here.

iii. Gated Recurrent Unit (GRU) [10]: It is a variation of RNN and at the same time, it reduces the structural complexity of LSTM by using update gate and reset gate [2]. Our GRU based model is illustrated in Fig. 2.

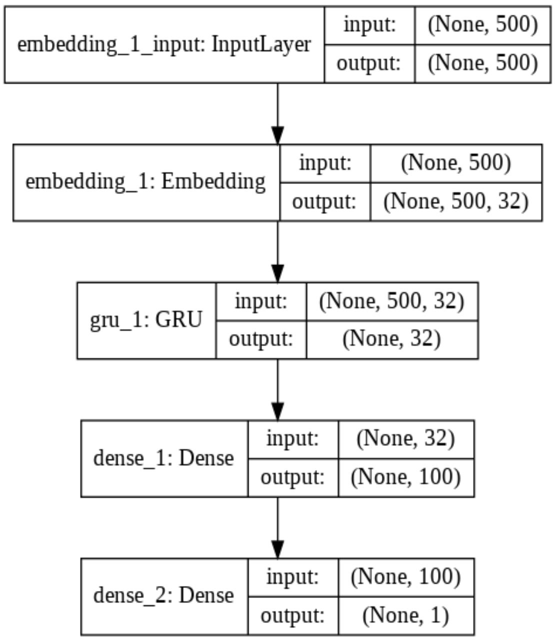

Fig. 2. Implementation of GRU recurrent network

F. Activation Function and Optimizers

Different activation functions such as Rectified Linear Unit (ReLU), Sigmoid and Tangent Hyperbolic (TanH) activation functions are used in Feature layer (dense_1 in Fig-1 and Fig-2) to determine which one gives the best accuracy. Sigmoid activation function is used in

classifying layer (dense_2 in Fig-1 and Fig-2) for all our deep learning approaches. Optimizers named RMSPROP, ADAGRAD, ADAMAX, SGD, and ADADELTA are used in the experiments. For comparison, accuracy as well as precision, recall, and f_1-score metrics are used.
Three activation functions are used in this spam filtering approach:

i. Rectified Linear Unit (ReLU): ReLU, a piecewise activation function, which is linear for +ve values and zero for –ve values. ReLU is famous for its easily training capability and better performance achievements [22].
i. Sigmoid Activation Function: This activation function introduces the nonlinearity of any model.
ii. TanH Activation Function: When logistic sigmoid function stuck and gives negative values, hyperbolic shaped tangent is used to give nearly zero output.

The following five Optimizers are used to get the better combinations for above activation functions:

i. RMSPROP: For each of the parameters, the learning rate is adapted for this optimization method [23].
ii. ADAGRAD: It is an optimizer that is adapted from the modification of SGD with per-parameter learning rate [24].
iii. ADAMAX: It is a first order gradient-based optimization algorithm and also a variant of Adam optimizer [25].
iv. ADADELTA: It is more reliable than ADAGRAD optimizer. It is robust and adapts learning rates based on the moving gradient updates [26].
v. SGD: It is an important optimization method in ML. When it is needed to store the one-element history for each parameter, SGD optimizer knows how to store this history data and accumulate new gradients in an orderly fashion [27].

IV. Result and Discussion

A. N-Gram Features

Three variations of N-Gram features have been applied individually for building baseline ML models to investigate their individual importance as feature. Individual N-Gram feature performance is compared in Table III, where accuracy is used as evaluation metric.

TABLE III: Comparison of N-Gram Feature Performance in Terms of Accuracy

N-Gram	Naïve Bayes	Logistic Regression	SVM
Uni-Gram	93%	96%	95%
Bi-Gram	90%	91%	94%
Tri-Gram	89%	89%	94%
Word N-Gram (1 to 3)	91%	95%	94%

From Table III, it can be shown that Uni-Gram performs significantly better than all other N-Grams (Bi-Gram, Tri-Gram, 1 to 3 Gram of Words) individually for all baseline algorithms. For Logistic Regression, Uni-gram feature secures 96% accuracy while for Bi-Gram and Tri-Gram, it gets 91% and 89% accuracy, respectively. It can be concluded that Uni-Gram is better as a feature than Bi-Gram and Tri-Gram for Bengali SMS spam detection. F_1-Scores of various baseline algorithms are shown in Fig. 3, where Uni-Gram is used as feature. It can be concluded that Logistic Regression is better ML algorithm for Bengali SMS spam detection than SVM and Naïve Bayes in terms of F_1-Scores.

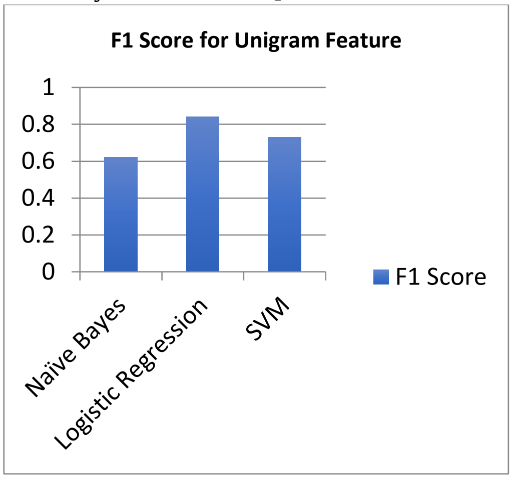

Fig. 3. F_1-Scores of baseline algorithms when Uni-Gram is used as feature

B. Baseline Algorithms VS Deep Learning Algorithms

After implementing all the basic supervised ML algorithms, it has been noticed that among all the baseline classifiers Logistic Regression gives the highest accuracy as 96%. Other supervised algorithm SVM and Naïve Bayes have accuracy rates 95% and 93%, respectively. Both deep learning algorithms LSTM and GRU get 99% accuracy. It can be concluded that the two variations of RNN algorithms (LSTM, and GRU) are the better choice than traditional baseline ML algorithms. Also, the paper provides a result of the Precision, Recall, and F_1-score [28] because accuracy can be misinterpreted sometimes. (Fig. 4)

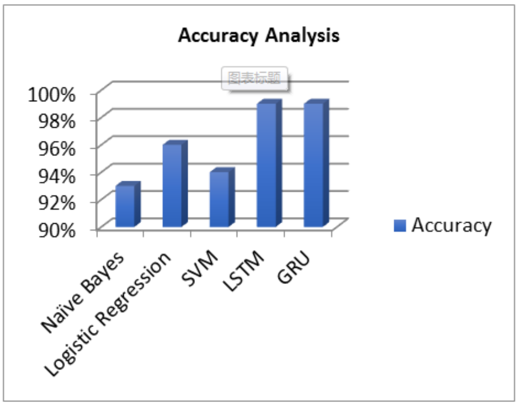

Fig. 4. Accuracy comparison on baseline algorithms VS deep learning algorithms

Precision and Recall are calculated from confusion matrix and F_1-score is the harmonic mean of precision and recall [28]. The different applied methods are compared in terms of F_1-Score and accuracy. The values of F_1-Score and accuracy metrics are shown in Table IV. It can also be observed that deep learning based methods get way better F_1-Score than all other baseline algorithms.

TABLE IV: EVALUATION OF DIFFERENT METHODS IN TERMS OF F1-SCORE AND ACCURACY

Models	F_1-Score	Accuracy
Naïve Bayes	0.62	93%
Logistic Regression	0.84	96%
SVM	0.73	95%
LSTM	**0.96**	**99%**

C. Activation Functions and Optimizers

Experiments are performed with various activation functions in feature layer and optimizers to investigate their effects on deep learning approaches. Vanilla RNN, and the variations of RNN (LSTM, and GRU) algorithms are applied in Bengali SMS spam filtering. The evaluation scores (F_1-Score and accuracy) are given in Table V, where SIGMOID activation function is used in feature layer (dense_1 in Fig-1 and Fig-2) for different deep learning algorithms.

TABLE V: EVALUATION SCORES OF SIGMOID ACTIVATION FUNCTION IN FEATURE LAYER

Algorithm	Optimizer	F_1-Score	Accuracy
Basic RNN	RMSPROP	0.89	97%
	SGD	0.47	87%
	ADAMAX	0.92	98%
	ADAGRAD	0.94	99%
	ADADELTA	0.93	98%
LSTM	RMSPROP	0.94	99%
	SGD	0.47	87%
	ADAMAX	0.95	99%
	ADAGRAD	0.95	99%
	ADADELTA	0.93	98%
GRU	RMSPROP	0.91	98%
	SGD	0.47	87%
	ADAMAX	0.94	99%
	ADAGRAD	**0.96**	**99%**
	ADADELTA	0.93	98%

Similarly, evaluations scores for ReLU and TanH activation functions are given in Table VI and Table VII, respectively. For all activation functions, SGD optimizer has shown very poor performances. Other four optimizers (RMSPROP, ADAMAX, ADAGRAD, ADADELTA) have shown good performances, which are evaluated in terms of F_1-Score and accuracy. For almost all the cases, ADAGRAD optimizer has shown the best performances, which are measured in terms of F_1-Score and accuracy. From Table V, VI, VII, three best combinations have been found, which are mentioned in Table VIII. The combinations are formed in such a way where activation function in feature layer and optimizer are chosen in LSTM or GRU to get the best performances. The three combinations, LSTM-TanH-ADAGRAD, LSTM-ReLU-ADAGRAD, and GRU-SIGMOID-ADAGRAD get the best performances in detecting Bengali spam SMS.

TABLE VI: EVALUATION SCORE OF RELU ACTIVATION FUNCTION IN FEATURE LAYER

Algorithm	Optimizer	F_1-Score	Accuracy
Basic RNN	RMSPROP	0.94	99%
	SGD	0.47	87%
	ADAMAX	0.93	98%
	ADAGRAD	0.95	99%
	ADADELTA	0.94	98%
LSTM	RMSPROP	0.93	98%
	SGD	0.47	87%
	ADAMAX	0.95	99%
	ADAGRAD	**0.96**	**99%**
	ADADELTA	0.93	98%
GRU	RMSPROP	0.90	98%
	SGD	0.47	87%
	ADAMAX	0.92	98%
	ADAGRAD	0.93	98%
	ADADELTA	0.95	99%

TABLE VII: EVALUATION SCORE OF TANH ACTIVATION FUNCTION IN FEATURE LAYER

Algorithm	Optimizer	F_1-Score	Accuracy
Basic RNN	RMSPROP	0.93	98%
	SGD	0.47	87%
	ADAMAX	0.90	98%
	ADAGRAD	0.86	96%
	ADADELTA	0.91	98%
LSTM	RMSPROP	0.93	98%
	SGD	0.47	87%
	ADAMAX	0.93	98%
	ADAGRAD	**0.96**	**99%**
	ADADELTA	0.93	98%
GRU	RMSPROP	0.91	98%
	SGD	0.47	87%
	ADAMAX	0.93	98%
	ADAGRAD	0.94	99%
	ADADELTA	0.93	98%

TABLE VIII: THE BEST COMBINATIONS OF ACTIVATION FUNCTIONS AND OPTIMIZERS

Combinations	Precision	Recall	F_1-Score	Accuracy
LSTM-TanH-ADAGRAD	0.93	0.99	**0.96**	**99%**
LSTM-ReLU-ADAGRAD	0.95	0.97	**0.96**	**99%**
GRU-SIGMOID-ADAGRAD	0.94	0.98	**0.96**	**99%**

V. CONCLUSION AND FURURE WORK

The automatic classification of Bengali SMS into spam or ham is a challenging task as Bengali dataset on spam SMS is very rare. In this paper, we present our prepared Bengali spam SMS dataset. The goal of this research is to classify a Bengali SMS in terms of spam or ham. Some supervised ML algorithms along with Basic RNN, LSTM and GRU are applied as separate runs to design spam detection models. An elaborate comparative analysis is performed with some baseline supervised ML algorithms along with the deep neural network algorithms LSTM and GRU as well. We found that Deep learning algorithms LSTM and GRU perform better than baseline ML algorithms in terms of F_1-Score and accuracy. Both LSTM and GRU secured 99% accuracy, which is the highest accuracy among all accuracies achieved by ML techniques. Activation functions and Optimizers are compared to extract the best performances. The ADAGRAD optimizer gets the best F1-Score and accuracy for all deep learning based Bengali spam detection techniques. In future, Transfer learning [29] and Deep Belief network (DBN) [30] algorithms can be applied for spam detection. Various feature-engineering techniques can also be experimented. Various NLP related dependency relations and semantic approaches could be applied to classify the SMS. Due to short length nature of SMS, the detection of SMS spam is challenging. At the same time, users can send SMS using mixture of various languages. The SMS spam detection task is also challenging for Multilanguage coded SMS. In future, we can also work on Multilanguage coded SMS spam dataset. We can apply various ML techniques to classify them properly.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Md. Mohsin Uddin generated research idea, conducted the research and reviewed the whole work while Monica Yasmin implemented the ideas. Monica Yeasmin and Md. Mohsin Uddin wrote the manuscript. M Saddam Hossain Khan, Md Istianatur Rahman, and Tabassum Islam helped to generate some ideas and critically reviewed the manuscript. All authors discussed the research results and approved the final manuscript version.

References

[1] M. A. Shafi'I, M. S. A. Latiff, H. Chiroma, *et al.*, "A review on mobile SMS spam filtering techniques," *IEEE Access*, vol. 5, pp. 15650-15666, 2017.

[2] P. Poomka, W. Pongsena, N. Kerdprasop, and K. Kerdprasop, "SMS spam detection based on long short-term memory and gated recurrent unit," *International Journal of Future Computer and Communication* vol. 8, no.1, pp. 11-15, 2019.

[3] L. N. Lota and B. M. M. Hossain, "A systematic literature review on SMS spam detection techniques," *International Journal of Information Technology and Computer Science*, vol. 9, no. 7, pp. 42-50, 2017.

[4] Q. Xu, E. W. Xiang, and Q. Yang, "SMS spam detection using non-content features," *IEEE Intelligent Systems*, vol. 27, no. 6, pp. 44-51, 2012.

[5] T. Almeida, J. M. G. Hidalgo, and T. P. Silva, "Towards sms spam filtering: Results under a new dataset," *International Journal of Information Security Science*, vol. 2, no. 1, pp. 1-18, 2013.

[6] K. P. Murphy, "Naive bayes classifiers," University of British Columbia, 2006.

[7] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, 2001.

[8] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic Regression*, New York: Springer-Verlag, 2002.

[9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735-1780, 1997.

[10] I. Ahmed, D. Guan, and T. C. Chung, "Sms classification based on navïe bayes classifier and apriori algorithm frequent itemset," *International Journal of Machine Learning and Computing*, vol. 4, no. 2, p. 183, 2014.

[11] T. A. Almeida, J. M. G. Hidalgo, and A. Yamakami, "Contributions to the study of SMS spam filtering: new collection and results," in *Proc. 11th ACM Symposium on Document Engineering*, September 2011, pp. 259-262.

[12] S. Bhatia, "A study on spam detection methods for safe SMS communication," *International Journal of Engineering & Technology*, vol. 7, pp. 790–792, 2018.

[13] A. Narayan and P. Saxena, "The curse of 140 characters: Evaluating the efficacy of sms spam detection on android," in *Proc. Third ACM workshop on Security and Privacy in Smartphones & Mobile Devices*, 2013, pp. 33–42.

[14] K. Yadav, S. K. Saha, P. Kumaraguru, and R. Kumra, "Take control of your smses: Designing an usable spam sms filtering system," in *Proc. IEEE 13th International Conference on Mobile Data Management*, 2012, pp. 352–355.

[15] E. G. Dadaa, J. S. Bassia, H. Chiromab, S. M. Abdulhamidc, A. O. Adetunmbid, and O. E. Ajibuwae, "Machine learning for email spam filtering: Review, approaches and open research problems," 2019.

[16] T. Islam, S. Latif, and N. Ahmed, "Using social networks to detect malicious bangla text content," in *Proc. 1st International Conference on Advances in Science, Engineering and Robotics Technology*, May 2019, pp. 1-4.

[17] F. Olsson, "A literature survey of active machine learning in the context of natural language processing," *Swedish Institute of Computer Science*, 2009.

[18] Pedregosa, Fabian, *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.

[19] T. Mikolov, M. Karafiát, L. Burget, J. Černocký, and S. Khudanpur, "Recurrent neural network based language model," in *Proc. Eleventh Annual Conference of the International Speech Communication Association*, 2010

[20] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proc. International Conference on Machine Learning*, February 2013, pp. 1310-1318.

[21] K. Cho, B. V. Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical

machine translation," arXiv preprint arXiv:1406.1078, 2014.
[22] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 807-814.
[23] T. Tijmen and H. Geoffrey, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural Networks for Machine Learning*, 2012.
[24] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, no. 7, pp. 2121-2159, 2011.
[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014
[26] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," arXiv preprint arXiv: 1212.5701, 2012.
[27] T. Zhang, "Solving large scale linear prediction problems using stochastic gradient descent algorithms," in *Proc. Twenty-First International Conference on Machine Learning*, July 2004, p. 116.
[28] D. L. Olson and D. Delen, *Advanced Data Mining Techniques*, Springer Science & Business Media, 2008.
[29] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, IGI Global, 2010, pp. 242-264.
[30] G. E. Hinton, "Deep belief networks," *Scholarpedia*, vol. 4, no. 5, p. 5947, 2009.

Md. Mohsin Uddin is currently a Lecturer of the Department of Computer Science and Engineering at East West University, Dhaka. He joined East West University in April 2018. He has obtained B. Sc. Engg. degree in Computer Science and Engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka. He has obtained his MS degree in Computer Science from University of Lethbridge, Alberta, Canada. He worked as a software engineer in various multinational software companies in Canada and Bangladesh. He has 6+ years of professional experience in Big data, NLP, Machine learning systems design and software development. He has several research publications in renowned international conferences. His research areas are NLP, Machine learning, and Big data.

Monica Yasmin is currently a student of the Department of Computer Science and Engineering at East West University, Dhaka. She admitted to East West University in January 2016. She was born in Pabna, Bangladesh, in 1997. She obtained her Secondary School Certificate (SSC) and Higher Secondary Certificate (HSC) both from Ispahani Girls' School and College, Dhaka, Bangladesh, in 2013 and 2015 respectively. She is an undergraduate student and her research interests include Machine Learning, algorithms, deep neural network and recurrent neural network.

M Saddam Hossain Khan is currently working as a Senior Lecturer in CSE department at East West University, Bangladesh. Previously, he also served as a Lecturer in the same institution from 2017 to 2019. He started his career as a lecturer in Atish Dipankar University of Science & Technology, Bangladesh in 2011 under Faculty of Science & Technology and served there for almost 2 years. He received his Master's degree in Information Technology from University of Stuttgart, Germany in 2016 and Bachelor's degree in Electrical and Electronic Engineering from East West University, Bangladesh in 2011. His research interests include Analog Circuit Design, Optoelectronics, Internet of Things (IoT), Machine Learning and Artificial Intelligence.

Md Istianatur Rahman received B.Sc. in Electrical and Electronic Engineering from East West University in 2011. He is the Lecturer in World University of Bangladesh Since 2015. He has supervised over 10 B.Sc. students. His research interest includes Artificial Neural Network modeling and optimization, smart renewable system modelling and optimization.

Tabassum Islam is pursuing B.Sc. in Computer Science & Engineering from East West University, a leading post graduate institute that deals with state-of-the-art technology in the field of Computer Science & Engineering in Bangladesh. Besides study, she is working as an undergraduate teaching assistant in the same field in EWU. Her research interests are app development, machine learning and networking.