A Deep Learning Approach to Detect Abusive Bengali Text

Conference Paper · June 2019

DOI: 10.1109/ICSCC.2019.8843606

5 authors, including:

Amit Kumar Das
East West University (Bangladesh)
39 PUBLICATIONS **290** CITATIONS
SEE PROFILE

Tanni Mittra
East West University (Bangladesh)
8 PUBLICATIONS **8** CITATIONS
SEE PROFILE

A Deep Learning Approach to Detect Abusive Bengali Text

Estiak Ahmed Emon[†], Shihab Rahman*, Joti Banarjee[┼] , Amit Kumar Das[‡], Tanni Mittra[§]
[†*┼‡§]Department of Computer Science and Engineering, East West University, Dhaka-1212, Bangladesh.
Email: [†]estiakemon309@gmail.com, * r.shihab95@gmail.com, [┼] banarjeejoti@gmail.com, [‡]amit.csedu@gmail.com, [§]tanni@ewubd.edu

***Abstract*— Day by day, Social media sites, online news portals and blogs commenting sections are getting saturated with abusive contents in Bangladesh. Detecting different types of abusive contents in online will not only improve these websites discussion sections but will also ensure user's safety. In this paper, several machine Learning and deep learning based algorithms e.g. Linear Support Vector Classifier (LinearSVC), Logistic Regression (Logit), Multinomial Naïve Bayes (MNB), Random Forest (RF), Artificial Neural Network (ANN), Recurrent Neural Network (RNN) with a Long Short Term Memory (LSTM) cell have been tested to detect multi-type abusive Bengali text. Besides, there has been introduced new stemming rules for Bengali language which help to achieve better performance of algorithms. Deep learning based algorithm RNN outperforms other algorithms by gaining highest accuracy 82.20%.**

Keywords—Linear Support Vector classifier; Multinomial Naïve Bayes; Long Short Term Memory; Deep Learning; Stemming etc.

I. INTRODUCTION

In this modern era, social networking sites have brought a revolutionary change in human life. Over the past decade, social network has grown in size and popularity [1]. Now people can easily communicate with each other and can share their information, feelings, and emotions using social sites in their own language and culture [19]. Currently Bangladesh has over 30 million active social network users. A survey reveals that 15 percent of active social media users increased since January 2017 [3]. In addition, Dhaka is the second highest city of active Facebook users [4]. Moreover, many of them use Twitter, Instagram, YouTube etc. [13].

As social networking sites are increasing gradually, cyber bullying is also getting more frequent [5]. People often face harassment by unknown users and strangers in social network [10]. Cybercrime and cyber bullying are rising rapidly in Bangladesh too [20]. A report by UNICEF indicates that 32 percent kids are at risk of cyber bullying in Bangladesh [7] [23]. Another study conduct by Cyber-Crime Awareness Foundation, an NGO reveals that 73.71 percent of cyber-crime victims are women [8]. Situation gets worst when several people committed suicide because of hateful messages and harassment through social network [9] [22]. Sometimes, abusive posts lead people towards real life actions and violations [17] [21].

Many researchers have focused on text classification in English and other languages to detect abusive messages, comments or images [12][6]. All over the world, significant numbers of users use Bengali language to communicate with each other [2]. Therefore, there is a good scope to detect multi type abusive Bengali text using predictive approach. Machine Learning (ML) and Deep Learning (DL) algorithms can play momentous role to detect and eliminate Bengali abusive content existing in social media.

In this paper, we introduce several types of Machine Learning (ML) and Deep Learning (DL) based algorithm, which is used to identify different types of abusive content in Bengali language. We imposed on evaluating the performance of algorithms by introducing stemming rules for Bengali language by following some Bengali grammar rules. Thus, applying these stemming rules with low dataset, higher accuracy is achieved.

The rest of the paper follows proceeding: in section II, shows the relevant papers and research. Section III, explains the methodology. Section IV, discusses about what we get from experiment. Lastly, the paper concludes with section V.

II. RELETED WORK

To detect abusive content researchers tried to show many approaches. In paper [11] author introduced a new concept that is Bag of Communities (BoC) approach. This concept can identify abusive content from a major online community. In this paper [11] Naïve Bayes (NB), LinearSVC, Logistic Regression algorithms are compared for text classification and Naïve Bayes performs best in all condition. Paper [12] used Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) with long short-term memory to classify abusive comment.

Another text classification problem is sentiment analysis. Paper [25] conducted sentiment analysis on Bengali and Romanized Bengali Text. Here data is categorized into positive, negative and ambiguous using deep learning and 78 percent accuracy is attained. In this paper the data set contains Romanized Bengali text. That's why, sentences bear spelling mistake and grammatically incorrect. The analysis of how other ML algorithms perform with this type of data are absent in this paper.

A few researches have been conducted on Bengali abusive text detection. Paper [14] worked with binary classification. Using 300 Facebook comments as data set the author proposed

978-1-7281-1557-3/19/$31.00 ©2019 IEEE

978-1-7281-1557-3/19/$31.00 ©2019 IEEE

an algorithm to detect abusive comment. No predictive algorithm is used in this paper. Another paper [15] worked with total 2500 Bengali data, which is collected only from popular Facebook pages. The author used Support Vector Machine (SVM), Random Forest (RF), and Multinomial Naïve Bayes (MNB) and showed a good analysis using ML algorithms. However, in both papers good preprocessing technique like stemming are absent.

The only paper [16] worked with classifier where preprocessing technique was present. In paper [16], stemming process was followed to get root form of a Bengali word. But in stemming process proper Bengali grammatical rules were not followed and it only removes suffix of a Bengali word for which getting better performance was not possible for all types of data. However, all of these issues have been considered in this paper.

III. METHODOLOGY

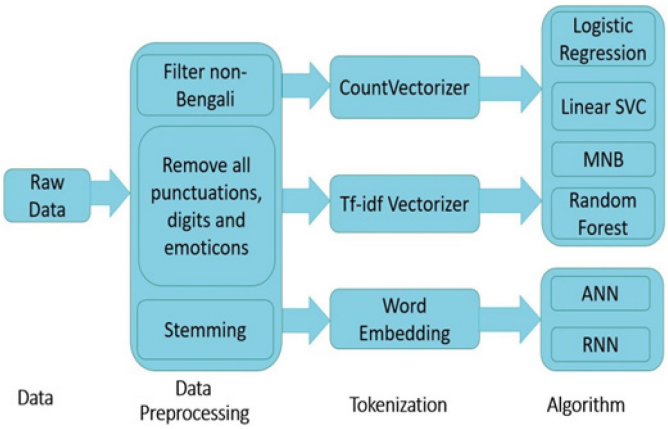

Fig. 1. Depicts the workflow of whole process. Raw data is preprocessed in 3 steps. 3 different tokenizers are used before sending to algorithms.

A. Dataset Collection

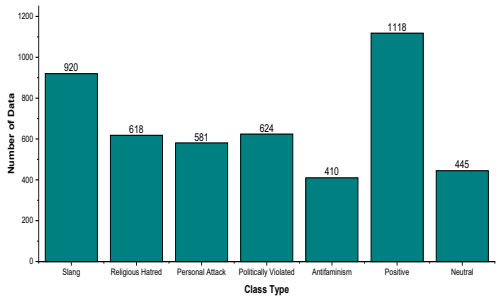

Fig. 2 shows the number of data containing each class.

For this research, data were collected from public comment sections of different social sites and online resources e.g. YouTube, Prothom Alo Online [18] and different facebook pages. In the data set, positive and neutral data are also present as two different classes. Total data set size is 4700, which were labeled in seven different classes such as slang, religious hatred, personal attack, politically violated, antifeminism, positive and neutral comment data.

B. Preprocessing

As a step of preprocessing, we collected comments which contain only Bengali language using language detector, an open source python library for language detection. Then all types of punctuations, whitespaces, emoticons and digits from dataset were removed. Each type of data was labeled manually according to their respective classes.

Stemming: In Natural Language, processing (NLP) stemming is important preprocessing technique for text analysis. Stemming approach is needed in NLP to find a word to its basic or root form. To get root form of Bengali word five rules are applied from Bengali grammar book [24].

- Rule 1: Article (পদাশ্রিত নির্দেশক) Inflection: Article in Bengali language is added at the end of the numeric word and demonstrative pronoun and others word. If article is removed from the Bengali word, then the Bengali word will be turned into its root form.
  Example: টাকাটা -> টাকা; Removed 'টা'(Ta)
  সারাটি -> সারা; Removed 'টি'(Ti)
  ওটি -> ও ; Removed 'টি'(Ti)

- Rule 2: Number (বচন) Inflection: By removing Singular and plural number from a word gives root form of that word.
  Example: ছাত্ররা ->ছাত্র; Removed 'রা'(Ra)
  খাতাখানা ->খাতা ;Removed 'খানা'(khana)

- Rule 3: Suffix (তদ্ধিত প্রত্যয়) Inflection: If suffix is removed from the Bengali word then the Bengali word will be turned into its root form.
  Example: চোরা ->চোর; Removed 'া'(Ra)
  জমিদারি -> জমিদার; Removed 'ি'(i)

- Rule 4: Verbal Root (ধাতুর মূল) Inflection: Another way to get root word is verbal root inflection.
  Example: খাওয়া ->খা; Removed 'ওয়া'(oa)
  খাওন -> খা ; Removed 'ওন'(on)

- Rule 5: Bibhakti(বিভক্তি) Inflection: bibhakti is attach with the root word. If number is removed from the Bengali word, then the Bengali word will be turned into its root form.
  Example: চেয়ারটিতে ->চেয়ারটি; Removed 'তে'(te)
  যাচ্ছেন -> যায় ; Repleced by 'য়'(yo)

Here root word of যাচ্ছেন is যা but for effective work purpose চ্ছেন is repleced by (য়).

TABLE I STEMMING EXAMPLE

Before stemming	After stemming
ছবিটিকে নিষিদ্ধ করা হয়েছে	ছবি নিষিদ্ধ করা হয়
কয়েক মাসের মধ্যেই সারা দেশে কার্যক্রম চালু হচ্ছে	কয়েক মাস মধ্যে সারা দেশে কার্যক্রম চালু হয়

Table I Sentences before and after stemming.

C. Parameter Tuning For Algorithms

TABLE II PARAMETER TUNING

Algorithm	Parameter for tuning	Best value
Multinomial Naïve Bayes	Alpha=[0.0001, 0.001, 0.01, 1]	Alpha = 0.01
LinearSvC	Multi-class=[crammer, ovr] Iterations=[800,1000,1300] 'Tolerance': [0.0001, 0.001, 0.01]	Multi-class=Ovr, Iterations = 800, Tolerance= 0.0001
Logistic regression	Iterations=[800,1000,1300], 'Tolerance': [0.0001, 0.001, 0.01], multi-class =*['ovr', 'auto']*	Multi-class=ovr ,Iterations=800, Tolerance= .0001
Random forest classifier	Number of trees=[100,500,1000], Maximum depth=[50,75,100]	Number of trees=1000, Maximum depth= 100
Artificial neural network	Number of layers=[2,3,4], Hidden layers=[1,2], Activation functions=[softmax,relu], Dropout =[0.1,0.2,0.3], Batch size =[64,128,256]	Number of layer =3, Hidden layer=1, Activation function=softmax, Batch size = 256, Dropout=0.1.
Recurrent Neural Network with Long Short Term Memory	Batch size=[16,32,64,128], Number of epochs=[5,7,10,15] optimizer = ['RMSprop', 'Adam', 'Nadam'], Dropout rate = [0.0, 0.1, 0.2, 0.3] recurrent dropout = [0.0, 0.1, 0.2, 0.3]	Batch_size = 16, Number of epochs=5, Optimizer=Nadam, Dropout rate=0.1, Recurrent dropout=0.2.

Table II. Parameter with values which were run using 10-fold cross validation. The best values were found for each classifier with uni-gram range.

Hyper parameters are not run directly with estimator. It is possible to increase performance of algorithm by tuning parameter. GridSearchCV is used for parameter tuning. To identify best configuration for all algorithms we use different settings. Table II. Shows all parameters and best values for algorithms. 10-fold cross validation are used to train and test the data. The best parameters are used in all subsequent process.

D. Extraction of feature and tokenizations:

We evaluate the process using three types tokenizers to analysis performance and optimize the result. For machine learning based algorithm, we use countvectorizer and tf-idf vectorizer to find best model. We use N-gram(n(1,3)) range for extracting from the text. For deep learning based approach, word embedding is used for tokenization.

IV. RESULT ANALYSIS

A. Applying machine learning algorithms using CountVectorizer

The ML algorithms are trained with CountVectorizer and the tested result is shown in Fig. 3. Multinomial Naïve Bayes gain 79.66 %, LinearSVC gain 80.93% Logistic Regression gain 77.96% and Random Forest classifier gain 73.72% accuracy.

.

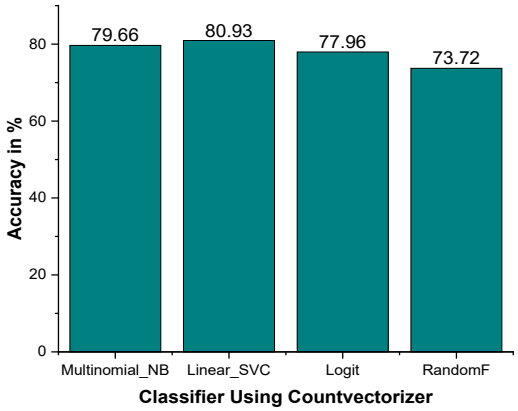

Fig. 3. Shows accuracy of different algorithms using CountVectorizer.

From Fig. 3, Linear SVC with CountVectorizer achieves highest accuracy than other classifiers. Besides, the accuracy of Multinomial Naïve Bayes is also noticeable which is very close to the Linear SVC. Other two algorithms result are poor.

B. Applying machine learning algorithms using tf-idf Vectorizer

The ML algorithms are trained with tf-idf Vectorizer. The tested results are shown in Fig. 4. Multinomial naïve Bayes gains 77.11%, LinearSVC gain 80.29% Logistic Regression gains 75.42% and Random Forest classifier gains 74.15% accuracy.

TABLE III PRECISION, RECALL, F1_SCORE, SUPPORT USING COUNTVECTORIZER

Class_name	Precision	Recall	F1_score	Support
Slang	0.84	0.93	0.88	80
Religious hatred	0.86	0.87	0.86	69
Personal attack	0.72	0.74	0.73	57
Politically violated	0.78	0.82	0.80	56
Antifeminism	0.79	0.67	0.72	45
Positive	0.86	0.88	0.87	117
Neutral	0.69	0.56	0.62	48

Table III Chart of all classes Precision, Recall and F-1 Score of Linear SVC classifier using Countvectorizer.

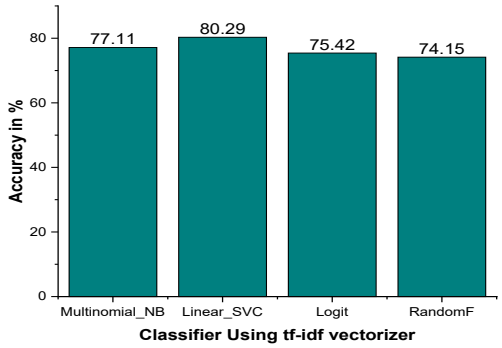

Fig. 4. Accuracy of different algorithm using TF-IDF Vectorizer.

From Fig. 4. We see that, Linear SVC with tf-idf Vectorizer shows better accuracy than other classifiers. Although, from Fig. 3 We see that, LinearSVC with countVectorizer shows slightly better accuracy.

TABLE IV PRECISION, RECALL, F1_SCORE, SUPPORT USING TF-IDF VECTORIZER

Class_name	Precision	Recall	F1_score	Support
Slang	0.82	0.86	0.84	99
Religious hatred	0.74	0.78	0.76	51
Personal attack	0.82	0.73	0.77	51
Politically violated	0.82	0.92	0.87	66
Antifeminism	0.73	0.59	0.65	41
Positive	0.87	0.90	0.88	115
Neutral	0.67	0.59	0.63	49

Table IV. Precision, Recall and f-1 Score of Linear SVC classifier using tf-idf Vectorizer.

C. Applying Deep learning algorithm

Deep learning based algorithms are trained with text to sequence which is keras provided tokenizer. From Fig. 5. We see that ANN achieves 34.65% accuracy where RNN with LSTM cells achieves 82.20 % accuracy. RNN shows better accuracy than artificial neural network. The gap of performance between these two algorithms is noticeable.

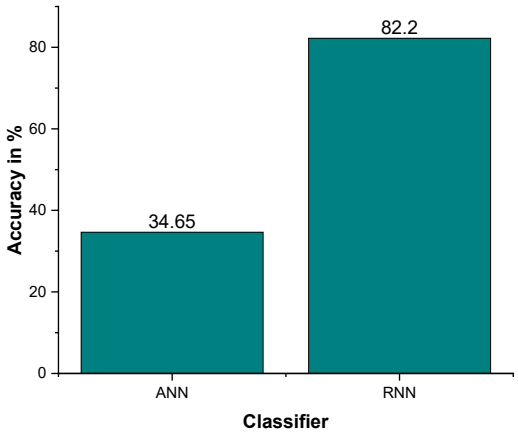

Fig. 5. Accuracy of different deep learning algorithms.

TABLE V PRECISION, RECALL, F1_SCORE, SUPPORT USING RNN

Class_name	Precision	Recall	F1_score	Support
Slang	0.89	0.92	0.91	89
Religious hatred	0.83	0.84	0.84	70
Personal_attack	0.80	0.80	0.80	56
Politically Violated	0.78	0.89	0.83	65
Antifeminism	0.91	0.67	0.77	30
Positive	0.88	0.81	0.84	112
Neutral	0.62	0.66	0.64	50

Table V. The chart of all classes Precision, Recall and F-1 Score of RNN.

D. Comparison between RNN, Linear SVC tf-idf Vectorizer and Linear SVC CountVectorizer

From Fig. 6, we see that RNN with LSTM cell outperforms all other machine learning algorithms by achieving accuracy 82.20%.

Here RNN achieves highest accuracy. RNN gives best performance with high precision (0.83), high recall (0.82) and high f1-score (0.82) while linear SVC with CountVectorizer achieved precision 0.81, recall 0.81, f1-score 0.81 and linear SVC with tf-idf Vectorizer achieved 0.80 in precision, recall and f1-score.

For our multiclass classification problem 90% training and 10 % testing brings highest accuracy. In RNN model with 3 hidden layers is used. Nesterov-accelerated Adaptive Moment

Estimation(NADAM) optimizer is the combination of Adam, RMSProp and momentum, a modified algorithm of Adam. As our classification is multiclass, so categorical cross entropy with Nadam optimizers achieved highest accuracy.

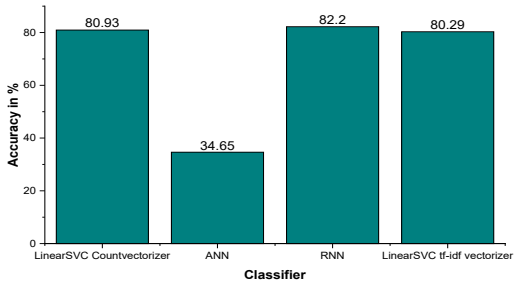

Fig. 6. Shows that RNN outperforms all other algorithms with accuracy achieved 82.20%.

Dropout for node was kept 0.1 and recurrent dropout was kept 0.2 to prevent over fitting the model. Tensorflow library is adopted for RNN model. Tensorflow, an open source python library for deep learning based algorithm.

V. CONCLUSION

This experiment helps to analyze the performance of all algorithms by following Bengali grammar rule. In this paper, among machine learning and deep learning algorithms, RNN with LSTM cell performs best to detect Bengali abusive text. In future, this experiment will be extended by applying other deep learning algorithms such as Deep Neural Network (DNN), Convolutional Neural Network (CNN) with Bengali spelling correcting process for detecting abusive Bengali text.

REFERENCES

[1] Global social media research summary 2019. [Online]. Available: https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/.[Accessed On: 10- Feb- 2019].

[2] J. Islam, M. Mubassira, M. R. Islam and A. K. Das, "A Speech Recognition System for Bengali Language using Recurrent Neural Network," 2019 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019.

[3] 30 Bangladesh's Digital Marketing and Social Media Marketing Stats and Facts. [Online]. Available: https://www.soravjain.com/digital-marketing-and-social-media-marketing-stats-and-facts-of-bangladesh/.[Accessed On: 10- Feb- 2019].

[4] http://digiology.xyz/demographics-facebook-population-bangladesh-april-2018/.[Accessed On: 15- Feb- 2019].

[5] "A Majority of Teens Have Experienced Some Form of Cyberbullying", Pew Research Center: Internet, Science & Tech, 2019. [Online]. Available: http://www.pewinternet.org/2018/09/27/a-majority-of-teens-have-experienced-some-form-of-cyberbullying/.[Accessed On: 10- Feb- 2019].

[6] R. A. Tuhin, B. K. Paul, F. Nawrine, M. Akter and A. K. Das, " An Automated System of Sentiment Analysis from Bangla Text using Supervised Learning Techniques," 2019 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019.

[7] Cyber Safety in Bangladesh: 32pc children bullied online.[Online]. Available: https://www.thedailystar.net/country/safer-internet-day-2019-prevent-bullying-of-children-online-in-bangladesh-unicef-1697785. [Accessed On: 10- Feb- 2019].

[8] M. Akter, F. T. Zohra and A. K. Das, "Q-MAC: QoS and mobility aware optimal resource allocation for dynamic application offloading in mobile cloud computing," 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox's Bazar, 2017, pp. 803-808.

[9] Cyber violence against women: the case of Bangladesh.[Online]. Available: https://www.genderit.org/articles/cyber-violence-against-women-case-bangladesh. [Accessed On: 03- Mar- 2019].

[10] A. K. Das, A. Ashrafi and M. Ahmmad, "Joint Cognition of Both Human and Machine for Predicting Criminal Punishment in Judicial System," 2019 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019.

[11] E. Chandrasekharan, M. Samory, A. Srinivasan, E. Gilbert , "The Bag of Communities: Identifying Abusive Behavior Online with Preexisting Internet Data", CHI 2017, May 6–11, 2017, Denver, CO, USA.

[12] T. Chu, K. Jue, M. Wang," Comment Abuse Classification with Deep Learning".

[13] T. Adhikary, A. K. Das, M. A. Razzaque, A. Almogren, M. Alrubaian, and M. M. Hassan, "Quality of Service Aware Reliable Task Scheduling in Vehicular Cloud Computing," Mobile Networks and Applications, Volume 21, Issue 3, pp 482-493, June 2016.

[14] M. G. Hussain, T. A. Mahmud, W. Akthar, "An Approach to Detect Abusive Bangla Text", International Conference on Innovation in Engineering and Technology (ICIET) 27-29 December, 2018.

[15] S. C. Eshan, M. S. Hasan," An application of Machine Learning to Detect Abusive Bengali Text".

[16] A. M. Ishmam, J. Arman," Automated Hate Speech Detection for Bengali Language in Social Media using Natural Language Processing and Machine Learning Approach".

[17] M. A. A. Mamun, J. A. Puspo and A. K. Das, "An intelligent smartphone based approach using IoT for ensuring safe driving," 2017 International Conference on Electrical Engineering and Computer Science (ICECOS), Palembang, 2017, pp. 217-223.

[18] Prothom Alo [Online]. Available: https://blog.mukto-mona.com. [Accessed On: 10- Feb- 2019].

[19] T. Adhikary, A. K. Das, M. A. Razzaque, M. Alrubaian, M. M. Hassan, and A. Alamri, "Quality of service aware cloud resource provisioning for social multimedia services and applications," Multimedia Tools and Applications, Volume 76, Issue 12, pp 14485-14509, June 2017.

[20] Women increasingly falling prey to cyberbullying.[Online]. Available: http://m.theindependentbd.com/post/171850. [Accessed On: 11- Jan- 2019].

[21] Hindus again attacked in Bangladesh on false rumours of defaming Muhammad in facebook.", *Struggle for Hindu Existence*, 2019. [Online]. Available: https://hinduexistence.org/2014/05/07/hindus-again-attacked-in-bangladesh-on-false-rumours-of-defaming-muhammad-in-facebook/. [Accessed On: 18- Feb- 2019].

[22] A. K. Das, T. Adhikary, M. A. Razzaque, M. Alrubaian, M. M. Hassan, Z. Uddin, and B. Song, "Big media healthcare data processing in cloud: a collaborative resource management perspective," Cluster Computing, Volume 20, Issue 2, pp 1599-1614, June 2017.

[23] A. Tashnim, S. Nowshin, F. Akter and A. K. Das, "Interactive interface design for learning numeracy and calculation for children with autism," 2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE), Phuket, 2017, pp. 1-6.

[24] M. Chowdhury, M. H. Chowdhury,"NCTB Bangla Grammer for Class 9-10".

[25] A. Hassan, M. R. Amin, A. K. A. Azada, N. Mohammed, "Sentiment Analysis on Bangla and Romanized Bangla Text (BRBT) using Deep Recurrent models".