Detecting Multilabel Sentiment and Emotions from Bangla YouTube Comments

Nafis Irtiza Tripto
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
nafisirtiza@cse.buet.ac.bd

Mohammed Eunus Ali
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
eunus@cse.buet.ac.bd

***Abstract*—Sentiment analysis has become a key research area in natural language processing due to its wide range of practical applications that include opinion mining, emotions extraction, trends predictions in social media, etc. Though the sentiment analysis in English language has been extensively studied in recent years, a little research has been done in the context of Bangla language, one of the most spoken languages in the world. In this paper, we present a comprehensive set of techniques to identify sentiment and extract emotions from Bangla texts. We build deep learning based models to classify a Bangla sentence with a three-class (positive, negative, neutral) and a five-class (strongly positive, positive, neutral, negative, strongly negative) sentiment label. We also build models to extract the emotion of a Bangla sentence as any one of the six basic emotions (anger, disgust, fear, joy, sadness and surprise). We evaluate the performance of our model using a new dataset of Bangla, English and Romanized Bangla comments from different types of YouTube videos. Our proposed approach shows 65.97% and 54.24% accuracy in three and five labels sentiment, respectively. We also show that the performance of our model is better for domain and language specific texts.**

Keywords*—*sentiment analysis, YouTube comments, emotion detection, Bangla language, deep learning

I. Introduction

The advent of online social networking sites such as Facebook, Twitter, and MySpace, has fueled the interest on sentiment analysis research that finds people's opinions, appraisals, evaluations, attitudes and emotions from text. The availability of huge online data and the advancement of machine learning schemes have accelerated the development of a plethora of techniques in analyzing sentiment and emotions from texts in English, French, Arabic and many other languages.

Though the sentiment analysis has been an widely studied topic in English, it is rarely studied in the context of Bangla language, which is the one of the most widely spoken and culturally rich language, with nearly 250 million of native speakers. The total number of Internet users in Bangladesh has reached 80.829 million at the end of January, 2018[1]. Moreover, Dhaka has been ranked as the third city in terms of number of active Facebook users recently[2]. Thus a large number of people express their feelings, thoughts and emotions in online platforms using Bangla text (in Romanized Bangla mostly).

Most of the works in Bangla sentiment analysis focus on identifying positive or negative texts using traditional rule based or machine learning based approach [1]–[4]. However, detecting only polarity is not enough for analyzing comments in micro-blogging or social sites as they often contain various degrees of sentiment and emotion information. For example, both *"This video is not upto the mark"* and *"What a rubbish video it is !!!"* comments are negative while detecting sentiment. However, it is evident that the second comment conveys more negative feeling as well as expresses disgust. To overcome the above limitations, we aim to build a multilabel sentiment analyzer (i.e., three and five class sentiments) and an emotion detector for Bangla and Romanized Bangla texts.

In recent period, Youtube has become one of the major social platforms and people interactively participate in the comment thread of these videos. However, sentiment analysis of short texts from these comments is challenging because of the limited amount of contextual data, the use informal language and the presence of a lot of mistakes in the text. Thus, rule and pre-defined feature based methods are not applicable for opinion mining from these texts. Recently deep learning algorithms have shown impressive performance in sentiment analysis across multiple datasets. A major benefit of these models is that they do not need to be provided with pre-defined handpicked features, rather they can learn sophisticated features from the data by themselves.

In this paper, we propose deep learning based approaches to build a multilabel sentiment and emotion analyzer. We evaluate the performance of our method in a dataset containing comments from different types YouTube videos in Bangla, English and Romanized Bangla language. Our model outperforms baseline solutions and existing methods and also

[1]https://bit.ly/2vQejwT

[2]https://bit.ly/2MyHsnq

978-1-5386-8207-4/18/$31.00 © 2018 IEEE

shows interesting characteristics in domain and language specific dataset.

The rest of the paper is organized as follows. Section II reviews the related works in opinion mining and emotion extraction. In Section III, we describe our dataset and provide an overview of proposed solution. Section IV presents the experimental results. Finally, we provide conclusions and future directions in Section V.

II. Related Works

Sentiment analysis or opinion mining has become a major point of focus in natural language processing, with over 7,000 articles written on the subject [5]. In recent period, deep learning applications have exhibited impressive performance across different NLP tasks. Much of the work with deep learning methods have involved learning word vector representations [6] and performing Convolutional Neural Network (CNN) based feature extractor for classification [7]–[9]. Shirnai explored the performance of different deep learning architectures for semantic analysis of movie reviews and achieved 46.4% accuracy in Stanford Sentiment Treebank dataset using CNN+word2vec model for five labels sentiment classification [10].

Nowadays, besides sentiment mining, the emotional aspects in texts also attract the attention of many research areas in NLP and different researchers focus on identifying emotions. Chaffar & Inkpen adopted a supervised machine learning approach to recognize Ekman's [11] six basic emotions using different feature sets. Bhowmick et al. [12] used an ensemble based multi-label classification technique called random k-label set (RAKEL) for emotion analysis.

Recently, researchers have expressed their interest in Bengali text and there are many publications based on sentiment and emotion analysis, theme detection, topic wise opinion summarization with data resources from various Bengali corpus [13]. Different Machine learning strategies like SVM with maximum entropy [2], Naive bayes (NB) [3], Multinomial Naive Bayes (MNB) with mutual infromation as feature analysis [4] has been used to classify Bangla sentences into positive or negative in various bangla domain texts like Micrblog posts, comments from blog, translated review dataset. Amin et al [14] proposed the results of word2vec word co-occurrence score in combination with the sentiment polarity score of the words and obtained 75.5% accuracy is in two class. Hasan et al. [15] has performed sentiment analysis on Bangla and Romanized bangla text using a Long Short Term Memory (LSTM) with binary and categorical cross entropy loss and achieved 70% accuracy for two class.

All the emotion analysis in Bangla texts have been carried out by Das and Bandyopadhyay in words, phrases and sentence levels in Bangla corpora and blogs [16], [17]. They developed WordNet Affect lists in Bangla from the affect wordlists already available in English. For identifying emotions from sentence, they used a Conditional Random Field (CRF) based classifier for recognizing six basic emotion tags for different words of a sentence.

TABLE I
Language distribution of dataset

Language	3 class	5 class	Emotion	Total
Bangla	2797	1208	1006	5011
English	2389	1050	747	4189
Romanized	3724	1628	1137	6489
Total	8910	3886	2890	15689

TABLE II
Domain distribution of Dataset

Domain	3 class	5 class	Emotion
Music Video	1402	571	440
Review Video	1231	553	346
Drama Video	1188	542	352
Funny Video	1080	483	315
Report Video	535	262	140
Sports Video	737	335	232
News Video	1122	468	378
Talkshow Video	1615	672	687

III. Methodology

We provide an overview of our solution in this section. Initially, we collect a dataset from YouTube comments and annotate the data. We apply various pre-processing and word embedding techniques on texts to remove noise from data and convert it to specific input format. Then we provide the architecture of our model.

A. Dataset Creation

We extract comments from different types of video domains as discussed in Table II using YouTube API version 3.0. We manually select these videos in Bangla language based on their popularity (number of views, number of likes or dislikes) dated from 2013 to early 2018. We limit the number of comments for each video up to 50 to remove redundancy and also exclude the replies of comments since they do not possess much sentiment information. We use Google translator to detect the language of each comment as Bangla or English. The sentences which are not identified as any language are considered as Romanized Bangla. Table I and II provides a distribution over our dataset.

We annotate each sentence in database according to three or five class sentiment detection problem. For emotion anal-

TABLE III
Label distribution in dataset

3 class		5 class		Emotion	
Positive	3104	Strongly Positive	416	Anger/Disgust	823
Neutral	2805	Positive	843	Joy	762
Negative	3001	Neutral	1222	Sadness	272
		Negative	1064	Fear/Surprise	294
		Strongly Negative	341	None	739

TABLE IV
SAMPLE YOUTUBE COMMENTS

Text	Language	Domain	Classification	Label
শেষের কাহিনী টা এরকম না করলেই ভালো হতো। (*The finishing of the story could be better if it does not happen*)	Bangla	Drama Video	3 class	Negative
vi amazing video make koren oswam (*Bro, You make amazing video. Awesome.*)	Romanized	Review Video	5 class	Strongly positive
They are playing really very nice well-done girls	English	Sports Video	Emotion	Joy
ভাই এতো ভুল ইনফরমেশন দেন কেন!আজেবাজে নিউজ সব (*Bro, why do you so much wrong information! all ridiculous news.*	Bangla	News Video	Emotion	Anger/Disgust
my favorite song is it	English	Music Video	3 class	Positive
2 mohila khali hase ke? mejajta kharap hoia jay. (*Why these two women are laughing? The mood gets worse*)	Romanized	Talkshow Video	5 class	Negative

ysis, there are six major emotion categories: anger, disgust, fear, joy, sadness and surprise. Most of these emotions (except "Joy") are mainly associated with negative sentiment, with "Surprise" being the most ambiguous, as it can be associated with both positive or negative feelings. Interestingly, the number of basic human emotions can be re-categorized into just four; joy, sadness, fear/surprise, and anger/disgust [18]. Anger and disgust convey same meaning in text and it is very difficult to distinguish the difference between them even for humans. Moreover, the number of comments that express fear or surprise are relatively small and so we integrate them into a single category. However, some sentences might not any provide any emotion information and we consider them as "None" category. Therefore, there are total five labels in our emotion detection problem. The distribution of different category is shown Table III.

The annotation part has been conducted by different native Bengali speakers with various background. We have created a public domain for data annotation purpose and circulate it. Comments from YouTube often contain abusive and vulgar words, slangs and personal attack. Therefore, we ensure that all annotators are adults. We solve the conflict of multiple labels of each sentence by taking majority votes. Table IV shows a sample from our dataset.

B. Pre-processing

The comments obtained from YouTube video is noisy and often contain errors, unnecessary information and duplication. Although a lot of pre-processing steps are present in rule based sentiment analyzer, not all these steps are required our approach. We tokenize each sentence and remove stopwords from them. There are multiple resources for English and Bangla stopwords removal techniques [3,4]. For Romanized Bangla stopwords, we create our own list by manually translating each word from Bangla stopwords considering different pronunciation and spelling.

Elongated words, punctuation marks and emoticons often contain sentiment information for multi class categorization. For example, *"Greaaaaat newsss !! :D"* certainly provides more positive feeling than *"Great news"*. Therefore, we do not apply lemmatization as well as keep the elongated words, punctuation marks and hashtags also. However, we remove links, urls, user tags and mentions from comments.

[3]https://github.com/stopwords-iso/stopwords-bn
[4]https://www.nltk.org/book/ch02.html

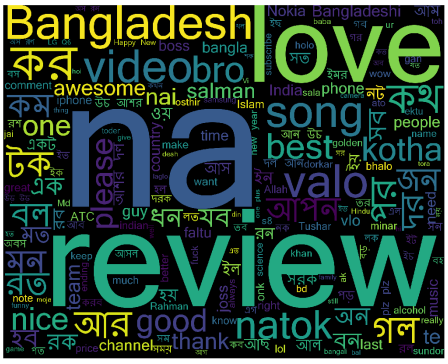
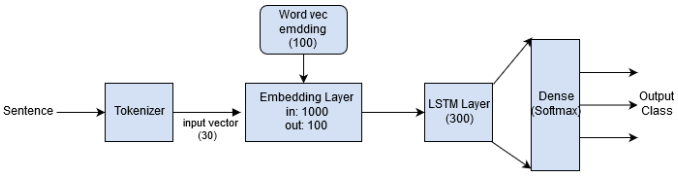


Fig. 1. Wordcloud representation for YouTube Comments

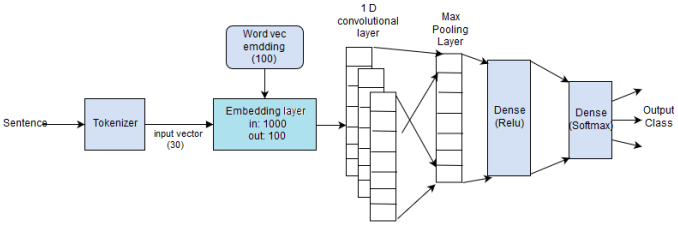
C. Word Embedding Representation

In order to implement a deep learning based model, we need to represent each word in a sentence as a vector representation. Word2vec [6] is an efficient algorithm for learning a word embedding from a text corpus that captures the syntactic meaning of words. We have implemented both Continuous Bag Of Words (CBOW) and Skip Gram (SG) model from [6]. SG approach takes an input word and attempt to estimate the probability of other words appearing close to that word. Alternatively, CBOW takes some context words as input and finds the single word that has the highest probability of fitting that context.

We create a vocabulary of size D from our text corpus. A wordcloud representation of our vocabulary is represented in Fig 1. Each sentence in the corpus is transformed into a one hot encoding vector of length D. Then we feed forward these vectors to a Neural Network consisting of one hidden layer with m nodes with linear activation function. The output layer has softmax activation function and contains D nodes where each node denotes the probability of placing the corresponding word in that sentence. By training this network, we create a $D \times m$ weight matrix connecting the D length input sentence with the m nodes hidden layer. Each row in this matrix corresponds to a word in vocabulary that



(a) LSTM Method



(b) CNN Method

Fig. 2. Architecture for sentiment and emotion classification

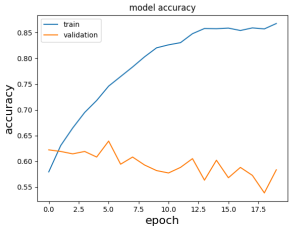
is now represented as a m length vector (We use $m = 100$ in our paper).

D. Model Architecture

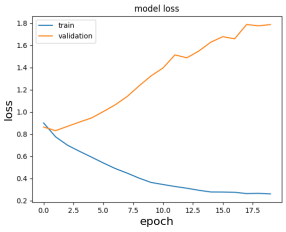
We have implemented two different approaches for opinion mining in our problem. The first model utilizes Long Short Term Memory (LSTM) with embedding layer as indicated in the work of Hasan et al. [15]. The second method employs Convolutional Neural Network (CNN) as its core layer. We have used the same model for both sentiment and emotion identification. Fig 2 shows the architecture of both model. In order to evaluate the performance of our solutions, we also implement a baseline method.

1) LSTM: After necessary prepossessing discussed in Subsection III-B, sentences are passed through a tokenizer to produce one hot encoding vector of length 30 as most of the YouTube comments are short. We only consider top 1000 most frequent words in vocabulary. We skip the sentences that are more than 30 words long and pad with zeros for shorter comments. Then these vectors are feed into an embedding layer and the weights are initialized with word2vec embedding weigths. The output dimension of embedding layer is 100 as it is the vector length of each word in word2vec model. The sequence of 30 words is then feed into a LSTM layer. Finally we add a dense layer with softmax as activation function since each sentence can belong to only one class in our scenario. The number of nodes in dense layer is equal to the number of class in specific problem.

2) CNN: The model architecture used in CNN is a variation of the approach proposed by Zhang & Wallace [9]. After the embedding layer we add 1D convolutional layer with 100 filters. Next global max pooling layer extract the maximum value from each filter and output dimension is a just one dimensional vector with length as same as the number of filters we applied. This vector is directly passed to a dense layer (Relu activation) without any filtering. Final output layer is a softmax layer with number of labels as output node.



(a) Accuracy



(b) Loss

Fig. 3. Effect of epoch number in train and validation accuracy and loss

3) Baseline Method: We utilize Support Vector Machine(SVM) and Naive Bayes (NB) as our baseline methods to detect three and five label sentiment as well as emotion. In order to generate the feature set for each sentence, we use Term Frequency Inverse Document Frequency (Tf-Idf) with n-gram tokens. We use linear kernel and l1 loss in our SVM model.

IV. Experimental Evaluation

In this section, we evaluate the performance of our proposed methods for three and five class sentiment analysis and emotion detection on YouTube dataset. We compare the performance of our solutions with baseline SVM and Naive Bayes(NB) approach.

A. Experimental Setup

We use Python Keras framework with Tensorflow as a background to implement all methods for training, tuning, and testing. Moreover, Gensim package has been used to implement word2vec model. Experimental evaluation was conducted on a machine with a Intel core i7 processor with 2.5GHz clock speed and 16GB RAM. The machine has also a Nvidia GTX 960M with 4GB memory and therefore Tensorflow based experiments can utilize GPU instructions.

B. Performance Evaluation and Parameterization

We have studied the efficiency and scalability of our proposed approaches by varying several parameters. We divide our dataset according to language and domain information and measure performance on each scenario. In all other cases, we use the total dataset for each classification category. Moreover, we evaluate the performance of different word2vec model by initializing embedding layer weights with them and consider both trainable and non-trainable weight matrix. We use adam optimizer and categorical cross entropy as loss function. Fig 3 shows the accuracy increases for training set as epoch number increases but loss in validation set increase also. Therefore, the problem of overfitting prevails in our approach. As a result, we set the epoch number to five in all experiments and use batch size 32.

We run the experiment ten times for each configuration and report the average performance. In each iteration, a new model is initialized with random training and testing set. For each experiment, we reserve 10% of our data for testing

TABLE V
PERFORMANCE MEASURE OF DIFFERENT APPROACHES

Method	3 class		5 class		emotion	
	Accuracy	F1 score	Accuracy	F1 score	Accuracy	F1 score
LSTM	**0.659664**	**0.63532**	**0.54242**	**0.5320**	**0.59230**	**0.5290**
CNN	0.6089	0.6052	0.521	0.52086	0.5403846	0.53465
NB	0.60791	0.5947603	.46880290	0.4802	0.5251	0.52473
SVM	0.5918542	0.589046	0.44876	0.465272	0.4926	0.4981

TABLE VI
PERFORMANCE OF DIFFERENT WORDVEC MODEL

Vectorization model	3 class		5 class		Emotion	
	Accuracy	F1 score	Acc	F1	Acc	F1
No embedding vector	0.614525	0.610142	0.5090	0.5014237	0.539230	0.51725
Continuous bag of words(cbow) (trainable)	0.622	0.62	0.51030	0.50109	0.5403846	0.53465
Continuous bag of words(cbow) (nontrainable)	0.62290	0.60425	0.50606	0.49993	0.53038	0.50461
Skip gram(SG) (trainable)	**0.6592**	**0.6550**	0.53818	0.517733	**0.59230**	0.5290
Skip gram(SG) (nontrainable)	0.639664	0.63532	**0.54242**	**0.5320**	0.5592	**0.53534**

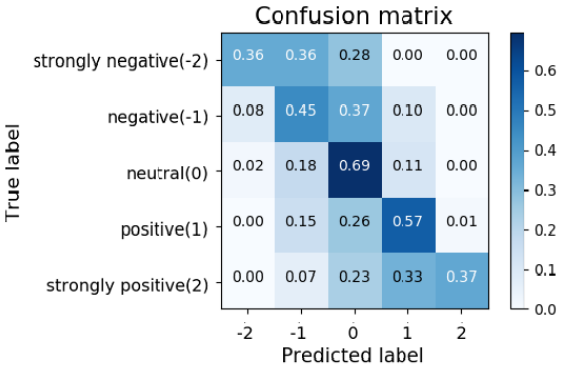


Fig. 4. Confusion matrix for 5 label classification

purpose. The rest is further divided into 80% training and 20% validation set. We present the accuracy and F1 score of testing set as our evaluation measure.

C. Result Analysis

1) Performance among different methods: Table V shows the performance measure of different methods. It is evident that both LSTM and CNN approach outperforms baseline SVM and NB approach in all classification scenario. LSTM is slightly better than CNN is most of the cases but CNN is much faster. NB performs slightly better than SVM but the difference is not statistically significant according to the work mentioned in [19]. The highest achievable accuracy for 3 and 5 class sentiment analysis is 65.97% and 54.24% respectively. For five category emotion detection, the accuracy is 59.23%. Fig 4 shows the confusion matrix for 5 class sentiment analysis. It is noticeable that the accuracy for 5 class sentiment drops due to the failure to distinguish between strongly positive and positive as well as between strongly negative and negative sentences. However, it is 10% more accurate than our baseline solution. Moreover, our method for three class sentiment scores 10% more than the method implemented in [15]. For emotion analysis, our work is the state of the art solution for Bangla language as no prior work has been conducted on identifying emotion from sentences.

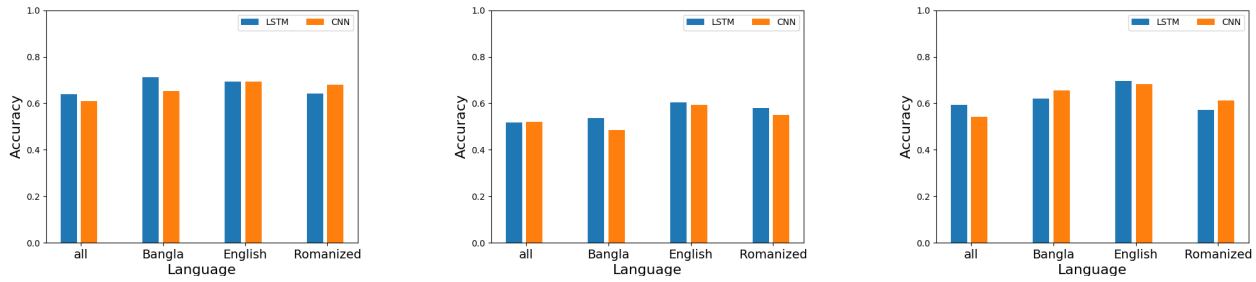
2) Wordvec performance: The performance of CBOW and SG word2vec model with both trainable and non trainable weights is listed in Table VI for LSTM method. Skip-Gram (SG) model provides highest accuracy and F1 score in all the cases. For 5 label sentiment static weights are better. However, in rest of the cases, the model performs better if we allow the weights of embedding layer trainable. CBOW also performs better than no prior initialization of weights.

3) Domain and language specific performance: Fig 5 and 6 show the performance measure of our approach in language and domain specific dataset. It is evident that accuracy increases in most of the cases for individual dataset than the combined one. Our method has higher accuracy in detecting sentiment and emotion from English texts than Romanized Bangla. The possible reason behind this scenario is that most English comments in YouTube videos are short, concise, more polarized, less error-prone and follow specific pattern. Alternatively, comments in Romanized Bangla language contain different spellings for same words, a lot of mistakes which are not easily understandable for humans also.

Interestingly, comments from review type videos score a higher accuracy in our method as they are more polarized (either appreciate the product/video or disparage it). Both of our approaches has a good accuracy in detecting sentiment and emotion from comments in music, sports, talkshow type videos. However, the accuracy decrease for news type videos as comments in this domain contain various topics and ambiguous thoughts.

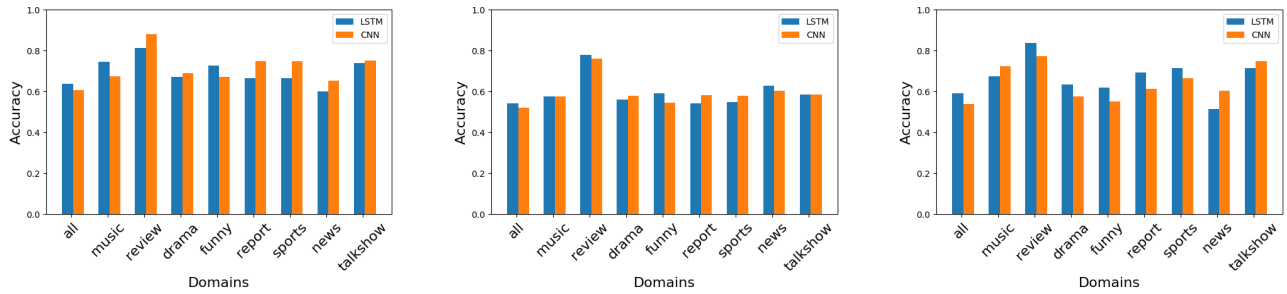
V. CONCLUSION

In this paper, we have developed deep learning based models to detect multilabel sentiment and emotion for Bangla and Romanized Bangla sentences. We have collected a dataset of comments from various YouTube videos and evaluate the performance of our models. Our model for multilabel



(a) 3 class sentiment (b) 5 class sentiment (c) Emotion

Fig. 5. Performance on specific languages



(a) 3 class sentiment (b) 5 class sentiment (c) Emotion

Fig. 6. Performance on specific domains

sentiment achieved at least 10% more accuracy than baseline solutions and existing approaches. We have also observed that the performance of our approach increases in domain or language specific texts. In future, we aim to include multiple aspects and topic information in sentiment and emotion detection.

References

[1] A. Das and S. Bandyopadhyay, "Sentiwordnet for bangla," *Knowledge Sharing Event-4: Task*, vol. 2, 2010.

[2] S. Chowdhury and W. Chowdhury, "Performing sentiment analysis in bangla microblog posts," in *Informatics, Electronics & Vision (ICIEV), 2014 International Conference on*. IEEE, 2014, pp. 1–6.

[3] M. S. Islam, M. A. Islam, M. A. Hossain, and J. J. Dey, "Supervised approach of sentimentality extraction from bengali facebook status," in *Computer and Information Technology (ICCIT), 2016 19th International Conference on*. IEEE, 2016, pp. 383–387.

[4] A. K. Paul and P. C. Shill, "Sentiment mining from bangla data using mutual information," in *Electrical, Computer & Telecommunication Engineering (ICECTE), International Conference on*. IEEE, 2016, pp. 1–4.

[5] R. Feldman, "Techniques and applications for sentiment analysis," *Communications of the ACM*, vol. 56, no. 4, pp. 82–89, 2013.

[6] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[7] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[8] C. dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 69–78.

[9] Y. Zhang and B. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," *arXiv preprint arXiv:1510.03820*, 2015.

[10] H. Shirani-Mehr, "Applications of deep learning to sentiment analysis of movie reviews," in *Technical Report*. Stanford University, 2014.

[11] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[12] P. K. Bhowmick, "Reader perspective emotion analysis in text through ensemble based multi-label classification framework," *Computer and Information Science*, vol. 2, no. 4, p. 64, 2009.

[13] V. K. Singh, "Sentiment analysis research on bengali language texts," *International Journal of Advanced Scientific Research Development (IJASRD)*, vol. 02, pp. 122–127, 2015.

[14] M. Al-Amin, M. S. Islam, and S. D. Uzzal, "Sentiment analysis of bengali comments with word2vec and sentiment information of words," in *Electrical, Computer and Communication Engineering (ECCE), International Conference on*. IEEE, 2017, pp. 186–190.

[15] A. Hassan, M. R. Amin, N. Mohammed, and A. Azad, "Sentiment analysis on bangla and romanized bangla text (brbt) using deep recurrent models," *arXiv preprint arXiv:1610.00369*, 2016.

[16] D. Das and S. Bandyopadhyay, "Developing bengali wordnet affect for analyzing emotion," in *International Conference on the Computer Processing of Oriental Languages*, 2010, pp. 35–40.

[17] D. Das, S. Roy, and S. Bandyopadhyay, "Emotion tracking on blogs-a case study for bengali," in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, 2012, pp. 447–456.

[18] R. E. Jack, O. G. Garrod, and P. G. Schyns, "Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time," *Current biology*, vol. 24, no. 2, pp. 187–192, 2014.

[19] T. Pranckevičius and V. Marcinkevičius, "Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification," *Baltic Journal of Modern Computing*, vol. 5, no. 2, p. 221, 2017.