Sentiment and Emotion Analysis from Textual Data in Bangla Language

Abdullah Al Jamil, Dept. of CSE, Southeast University, 2016100000001@seu.edu.bd
Rifat Rahman, Dept. of CSE, Southeast University, rifat.rahman@seu.edu.bd

***Abstract*--Sentiment analysis and emotion detection play a significant role in many areas of natural language understanding because of their extensive usability in different applications like chatbot, text summarization, etc. In comparison with other languages, the studies done in the Bangla language are very little due to lack of data and proper preprocessing tools. In this work, we present a benchmark study on different models to classify both emotion and sentiment from Bangla textual data. We collect our corpus from several online social platforms and do a comparison among different traditional machine learning and neural network-based approaches. We also perform comparative analysis among various informative feature extraction techniques. We find that the sequence model-based approach with the "word embedding" feature extraction technique outperforms other approaches for emotion detection and "convolutional neural network" model provides the best accuracy for sentiment analysis.**

***Index Terms*--Sentiment analysis, Emotion detection, Machine learning, Neural network, Word Embedding, Text classification**

I. Introduction

Nowadays people share their appraisals, sentiments, evaluations, attitudes, and emotions on different kinds of social media and microblogging sites such as Facebook, Twitter, Reddit, etc. Some of these are positive, some of these are negative, some of these contain someone's anger and some of these express someone's sadness. As a result, automatic detection of sentiment and emotion has become key research in NLP and data can be collected from different social platforms. However, a large amount of the research works related to automatic emotion and sentiment identification have been done in the English language. But a little amount of research has been done in the context of the Bangla language. In this paper, we are doing a benchmark study on different traditional machine learning and neural network-based models to detect sentiment and emotion from Bangla texts.

Bangla is a widely spoken language in the world [20]. Almost 250 million people worldwide accepted Bangla as their primary language. At the end of February 2020, the number of Internet subscribers reached 99.984 million in Bangladesh [1]. Bangladeshi peoples are now involved in online activities such as - sharing viewpoints for current incidents, giving reviews of foods or products, messaging with friends, giving feedback as comments on popular micro blogging and social networking sites. Moreover, Bangladesh has attained the 10th position over the world in the sense of Facebook users' numbers [2]. So a great number of people use Bangla texts to express their emotions on social sites. People watch the news, videos, etc., and share their sentiments in the comment section by using texts. So, it is indeed very much necessary to implement an automatic model that can detect both sentiment and emotion from a given Bangla text.

Text analysis in the Bangla language is such a challenging task [21]. This is because of the lack of proper resources. There is also a scarcity of effective and built-in pre-processing tools in the Bangla language [22]. NLTK library [23] in python language provides several functionalities like stemming, lemmatizing, POS tagging, etc. in English, Spanish language. But there exists no facility in case of Bangla language. Again there are several other issues such as out-of-context data, overuse of slang words in comments and text, spelling mistakes, usage of regional and Romanized Bangla text, etc.

There have been several studies on sentiment and emotion detection in other languages. They have implemented rule-based methods [4]-[6], traditional machine learning [8], [9], [10], [11], [19], and neural network-based approaches [7], [13], [15]. They have also shown different types of feature extraction techniques. Due to the lack of corpus or labeled data set, there are few studies in the context of the Bangla language. There is no benchmark dataset through which different studies can be compared. Again most of the studies have not performed proper data pre-processing steps. Many studies have not extracted informative features from data.

In this study, we have created our corpus from different social platforms. Then we have labeled our data by deploying the corpus to the general mass of Bangladesh. Then we apply several preprocessing steps and extract various informative features. After that, we implement different models and do comparative performance analysis among all of them.

The main objective of this work is to perform comparative analysis among the performance of different traditional machine learning and neural network-based approaches. We also show the comparison among various feature extraction techniques. We aim to determine a model with high accuracy, low response time, and high scalability.

The impact of this work both in "Natural Language Processing" and "Natural Language Understanding" is huge. It has many applications. For developing a chatbot, understanding the emotions and sentiments of the users is highly needed. It has also other use-cases like text summarization, text generation, keywords extraction, etc.

The rest of the paper is organized as follows. In Section II, we review different existing studies regarding emotion and sentiment detection both in Bangla and other languages and find their limitations. We describe our proposed methods in Section III. Section IV represents the implementation of our methodology and experimental analysis. We discuss our limitations and future research directions in Section V. Finally, we conclude this task in Section VI.

II. Related Works

Nasukawa and Yi et al. [3] introduced the term "sentiment analysis" for the first time. Most research works on sentiment analysis that appeared as early as 2000 were rule-based implementations [4]-[6]. In recent times, many works have shown impressive results by applying deep learning-based approaches in NLP-related tasks. Shirnai et al. [7] trained a model where they used a CNN-based classifier and word2vec feature extraction procedure for five-class sentiment identification by utilizing Stanford Sentiment Treebank movie review dataset and achieved 46.4% accuracy.

Nowadays the emotion analysis from texts is also a popular topic for researchers and traditional machine learning-based methods are being used. Several supervised traditional machine learning approaches had been implemented by Chaffar&Inkpen et al. [19] where they identified six basic emotions using different feature sets like BOW and N-grams. They found the Support Vector Machines (SVM) classifier as the best classifier of all ML-based classifiers for the emotion classification. Bhowmick et al. [8] applied an ensemble-based multi-label classification technique called RAKEL where they showed a novel approach for classifying news sentences into emotion classes. They obtained the F1-score of 82.1% from their model.

One biggest problem in sentiment analysis in Bangla is that there is no standard collection of data, such as - the IMDB dataset, Twitter corpus, etc. for Bangla texts. However, researchers create their datasets for training and testing models. There are very few studies in the Bangla language in which traditional ML-based methods have been applied. Chowdhury et al. [9] used the Support Vector Machine classifier and Maximum entropy-based feature extraction technique to classify the overall polarity of texts as either positive or negative. Islam et al. [10] used Naive Bayes (NB) classifier to identify sentiment from Facebook status and comments. Paul et al. [11] implemented a Multinomial Naive Bayes (MNB) classifier with mutual information as a feature extraction technique to identify Bengali texts as positive or negative from various Bengali domains texts. An approach proposed by Amin et al. [12] showed 75.5% accuracy in binary sentiment classification. They extracted the sentiment polarity score for each word from training data and applied word2vec feature extraction technique.

There also exists some neural network-based approaches regarding sentiment and emotion detection in Bangla. Hasan et al. [13] used the sequence model-based approach such as LSTM for detecting sentiments from both Bangla and Romanized Bangla texts and the F1-score of binary classification was 70%. Tripto et al. [15] developed 3 class and 5 class sentiment classifiers. They also built an emotion classifier for English, Bangla, and Romanized Bangla comments which had been collected from YouTube. They used LSTM and CNN classifiers for classification and also evaluated the results with some baseline methods.

There are also some approaches where probabilistic models have been utilized. Conditional Random Field (CRF) based classifier which is a probabilistic graphical model, has been used by Das & Bandyopadhyay et al. [14] for identifying six basic emotion tags for each word in a sentence.

As there exists no benchmark dataset, many of these studies cannot show any sorts of comparative study against others' studies. Most of the literature did not focus on the data pre-processing and data wrangling phases. Again few studies skipped the data labeling phase.

III. Methodology

In this section, we describe our proposed methodology in detail. We discuss the process of corpus creation, data pre-processing and feature extraction techniques. We also illustrate the architecture of the traditional machine learning & neural network-based models that are utilized in our study. Fig. 1 presents the workflow diagram of our methodology.

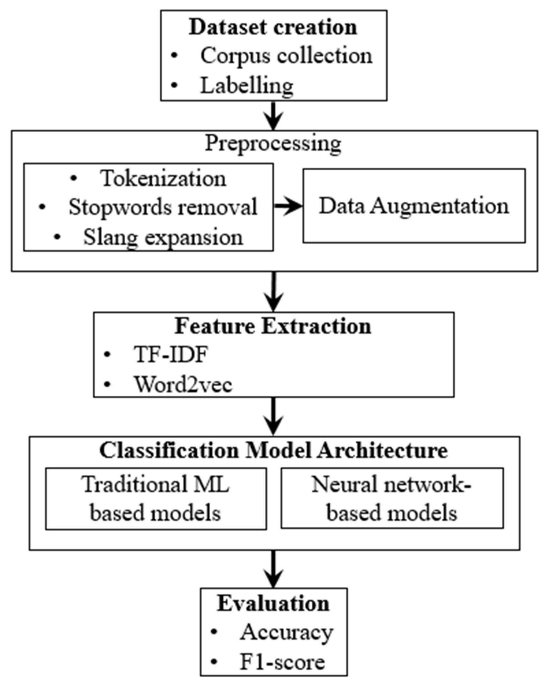


Fig. 1. Workflow diagram of our method

A. Dataset Creation

We collected Bangla textual data from different social platforms (Twitter, Youtube, Reddit, etc.) by using respective APIs. We also chose different types of popular Bengali news sites. Two types of datasets were built for our study. First one was for sentiment analysis where data were labeled with '0' for positive comments and '1' for negative comments (Fig. 2). Another one is the dataset for emotion detection where data were labeled into four distinct emotion categories: fear/surprise, sadness, anger/disgust, and joy. Here, we kept all 'anger' labels with 'disgust' and kept the 'fear' labels with 'surprise'. Some comments did not express any emotion. So, we labeled them as 'none' class (Fig. 3). Thus, we labeled all the input texts into 5 classes for emotion detection.

Total 26462 comments were included in sentiment dataset and 7732 comments into emotion dataset. TABLE I represents the distribution of our both datasets.

TABLE I: Label distribution for both sentiment analysis and emotion detection

Sentiment Analysis		Emotion Detection	
Class	Number of Comments	Class	Number of Comments
		Disgust/Anger	2714
		Joy	2118
Positive	15916	None	301
Negative	10546	Sad	1412
		Surprise/Fear	1070
Total	26462	Total	7732

These data have been labeled by the participation of general people of different ages from Bangladesh. We deployed a public domain to the people for identifying the class of the text or comments and then labeled them by majority voting. Fig. 2 and Fig. 3 showed samples from our datasets.

Input Text	Sentiment Label
<b>ভাই তোমার কথা সত্য, আমি তোমার কথার সাথে একম...	1.0
আজ হারিয়ে গেছে মানব সভ্যতা,নৈতিক ধর্ম ও চরিত্র...	0.0
আপনার কথা গুলা ১০০% সত্যি ভাইজান আপনাকে আমি সব...	1.0
আমরা মুসলিমরা আসলে বেশি আল্লাহওয়ালা দেখাতে চাচ...	0.0
আমাদের এই আজব বৈচিত্র্য যেদিন যাবে সেদিন আমরা ...	1.0

Fig. 2. Sample of sentiment dataset before preprocessing

Input Text	Emotion Label
ডিসলাইক দেয়া 32	none
ব্যাপক বিনোদন।	joy
মন খারাপ থাকলেই ঢুকে পরি আপনার চ্যানেলে।	none
Taslima Akter কন্টেন্ট ক্রিয়েটর রা এক একটা প্র...	disgust
জাতি জানতে চায়।	none

Fig. 3. Sample of emotion dataset before preprocessing

B. Preprocessing

Data preprocessing is an important phase in machine learning. Data wrangling can increase the quality of data so that informative features can be extracted and the model can perform better. Bangla comments and data often contain errors, stopwords, links, etc.We tokenized data into words and removed stopwords, links, hashtags, etc. We kept elongated words and didn't apply stemming. Elongated words often contain emotional information for multi-class classification. We replaced emoji or emoticons because emoticons inside a sentence carry the emotions of that particular sentence. As general people use slang words during the usage of social platforms, we also expanded slang words. We performed data augmentation by oversampling, single and multiple replacements of words inside a sentence with their synonyms, sentence rephrasing, etc.

C. Feature Extraction

We evaluate Tf-Idf[24] of both uni-grams and bi-grams from our corpus. We extract this feature from traditional machine learning models. Tf-Idf is the multiplication between term frequency and inverse document frequency which is highly used for keyword extraction from textual data. We also consider a simple statistical feature, for example- count vectorizer.

For neural network-based models, we use word embedding for each word. Word embedding refers to the vector representation of words. We apply the word2vec model by Mikolov et al. [25] for learning the vector representation of words from our dataset. Word2vec is an MLP neural network-based model which has 2 layers. In our literature, we take vocabulary size as 10,000. So the number of neurons in the input layer of the word2vec model is 10,000. We select the dimension size of the word vector as 100. Hence, the number of hidden nodes in the hidden layer is 100. Finally, the number of neurons in the output layer is 10,000 which is the same as the vocabulary size.

Word2vec model can be classified into two types. The first one is “Continuous Bag of words”(CBOW) where the average vector of the one-hot vectors of the context words is fed into the input layer and the one-hot vector of the center word or target word is fed into the output layer. The second one is “Skip-Gram” (SG) where the one-hot vector of the target word is fed into the input layer and the average vector of the one-hot vectors of context words is fed into the output layer.

We apply both kinds of word embedding in our neural network-based models. We also do comparative analysis among all of these feature extraction techniques.

D. Model Architecture

We have used four different deep learning methods for sentiment analysis and emotion detection. Four methods are Bidirectional Long Short Term Memory (BiLSTM), Bidirectional Gated Recurrent Unit (BiGRU), Convolutional Neural Network (CNN), and a hybrid method CNN+BiLSTM. We also applied different traditional machine learning approaches. We used all of these methods for both sentiment and emotion analysis.

1) *Deep Learning Methods:* After required preprocessing, comments are passed into a tokenizer to produce a vector of length 50. We only consider the top 10000 most frequent words from our corpus as vocabulary. So, the vocabulary size of our study is 100. We do not include the comments that are more than 50 words long for shorter comments and we pad them with zeros.

   a) *BiLSTM and BiGRU:* We have followed the same procedure that Tripto et al. [15] used for LSTM in their paper. The vectors we processed from comments are fed into an embedding layer and the weight will be word2vec embedding weights. The embedding layer produces 100-dimensional output because in the word2vec model each word's vector length is 100. The sequence of 50 words is then fed into a bidirectional LSTM layer in BiLSTM model and bidirectional GRU layer in BiGRU model. Finally, we add a dense layer with softmax activation function for emotion analysis and sigmoid for sentiment analysis.

b) *CNN:* We implemented "convolutional neural network" proposed by Zhang & Wallace [18]. After the embedding layer, a 1D convolutional layer is added. The convolution layer consists of 256 filters and the size of the kernel is 5. Then we add a global max-pooling layer. After that we add a dense layer with ReLU activation function. Finally, we add the output layer with softmax activation function.

c) *CNN+BiLSTM:* We combined CNN and BiLSTM to build this hybrid model. A BiLSTM layer has been added between 1D convolutional layer and the dense layer with ReLU activation function in CNN model.

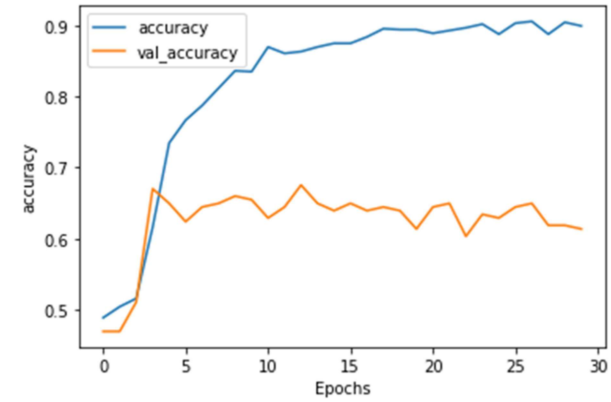
2) *Traditional Machine Learning-based Methods:* We implement several traditional machine learning approaches, for example, Support Vector Machine (SVM), Multinomial Naive Bayes (MNB), Random Forest, and Decision Tree to detect the polarity of sentiment as well as emotion. For these machine learning approaches, we extract the Tf-Idf feature. TF-IDF evaluates how relevant a word is to a text in a collection of texts. To create feature sets we also use a count vectorizer for some ML-based models in both sentiment and emotion analysis. Tfidfvectorizer considers the overall document weightage of a word while count vectorizer only counts the number of times a word appears in the document.

IV. Experimental Evaluation

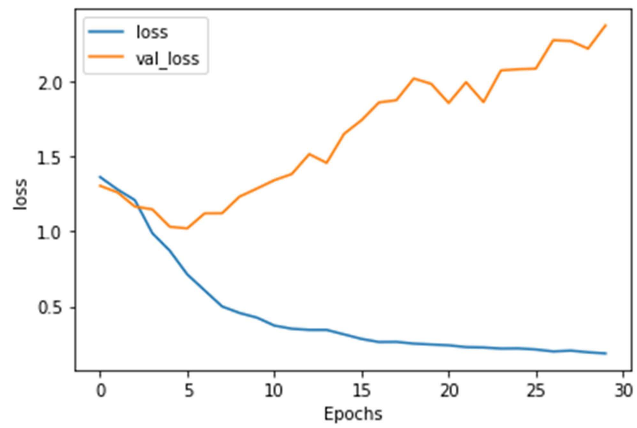
In this section, we explain the implementation details of our proposed methodologies. Then we analyze the performance of different methods. Furthermore, we do comparative analysis among all the models. We also show a comparison among different feature extraction techniques. Finally, we compare our model with other state-of-the-art works.

A. Parameterization

The total dataset has been utilized for each classification category. We used Adam optimizer and categorical cross-entropy as the loss function for multiclass emotion detection and binary cross-entropy for sentiment analysis. At first, we tried a big epoch number but the overfitting problem prevails in our approach (Fig. 4). Then we set epoch as 4 for all sentiment experiments and epoch 10 for all emotion experiments. For each experiment, we divided our datasets into 80% for training and 20% for the validation set. We present the accuracy and F1 score of the testing set as our evaluation measure.



(a) Validation and training accuracy



(b) Validation and training loss

Fig. 4. The change of validation and training curve for both accuracy and loss with the increment of epochs

TABLE II: Performance of different methods

Methods	Sentiment		Emotion	
	Accuracy	F1 score	Accuracy	F1 score
BiLSTM	0.8265	0.82043	**0.6583**	**0.618396**
BiGRU	0.8280	0.817765	0.5879397	0.5638765
CNN	**0.8318**	**0.822147**	0.5829	0.54299
BiLSTM+CNN	0.8215	0.81330	0.5879	0.5616329
NB	0.751659	0.80604	0.584158	0.531995
SVM	0.7548815	0.7975575	0.5990099	0.531048
RF	0.755645	0.802841	0.559405	0.5228075
DT	0.721967	0.774133	0.579208	0.544802

B. Result Analysis

1) *Comparative Analysis among different Models:* Table II shows the performance measurement of different methods. We can see deep learning approaches give us better results than machine learning-based approaches like SVM, NB, Random Forest, and Decision Tree models in all classification scenarios. In sentiment analysis, the CNN classifier is

performing the best of all models. But in emotion analysis classification BiLSTM is showing better performance than any other methods. NB, SVM, and Random Forest performance are almost the same. The highest achievable accuracy for sentiment class analysis is 83.18% and for emotion detection, the accuracy is 65.83%. In sentiment analysis, all deep learning models have gotten at least 5% more accuracy than ML-based solutions.

2) *Performance Comparison among Different Feature Extraction Techniques:* From Table II, we have observed that neural network-based models outperform traditional machine learning-based models. This scenario indicates that the word embedding feature is more informative than the Tf-Idf feature extraction technique.

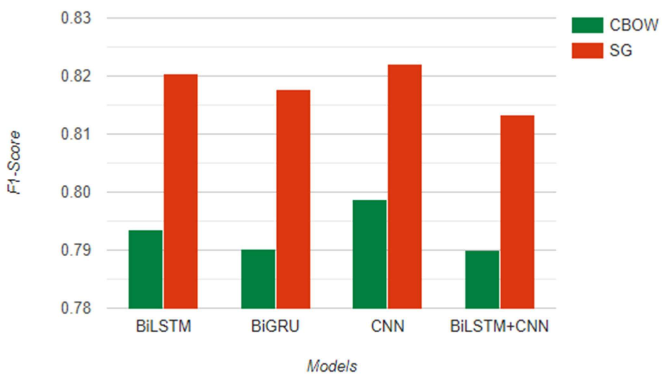


Fig. 5. Comparison between two types of word2vec model (CBOW vs SG) in the context of sentiment analysis

From Fig. 5, we may observe that in all scenarios of neural network-based approach "Skip-Gram" (SG) method is outperforming the "Continuous Bag of Words" (CBOW) method. The probable reason behind these situations is the scarcity of datasets. SG method performs well when the size of the dataset is small. On the other hand, the CBOW technique performs better when the size of the dataset is large.

3) *Comparison against State-of-the-art Works in Bangla:*

TABLE III: Comparison with other "Sentiment Analysis" studies

Sentiment Analysis		
Method	Accuracy	F1-Score
CNN (our study)	**0.8318**	**0.822147**
Islam et al. [10]	0.68	0.72
Amin et al. [12]	0.75	0.78

TABLE IV: Comparison with other "Emotion Detection" studies

Emotion Detection		
Method	Accuracy	F1-Score
Bi-LSTM (our study)	**0.6583**	**0.618396**
Tripto et al. [15]	0.592300	0.5290

From Table III and Table IV, we can see that our best-performed model outperforms other existing sentiment and emotion analysis studies. Fig. 6 depicts the confusion matrix of the sentiment analysis using CNN classifier with word2vec (SG) method.

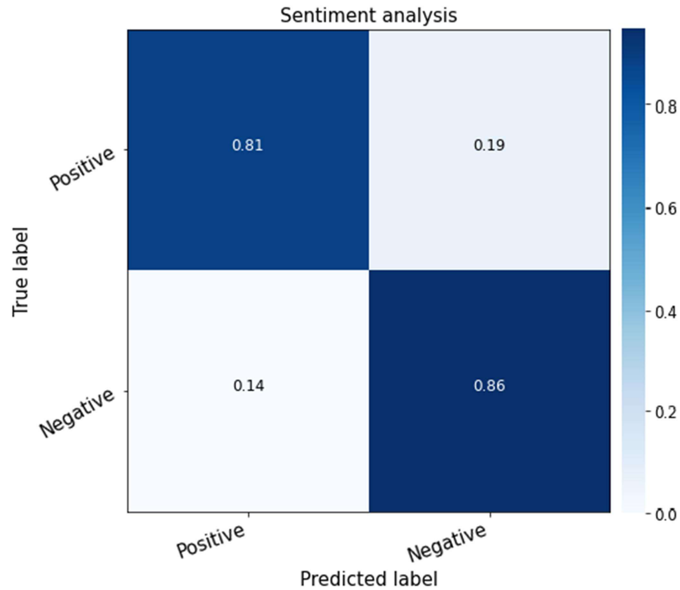


Fig. 6. Confusion matrix of the sentiment classification using CNN classifier.

V. Limitations and Future Directions

The major limitation of our work is the scarcity of datasets. Again the dataset was imbalanced for emotion detection. As a result, false-positive rates and false-negative rates are high for the emotion detection task. Furthermore, because of the scarcity of data, we cannot apply the attention-based mechanism in our study. In future, we have a plan to enrich our dataset so that it can be considered as benchmark data for both sentiment and emotion detection tasks. We also have the scheme to implement an advanced model to obtain more accuracy and F1-score.

VI. Conclusion

In this study, we do a benchmark study on both traditional machine learning and neural network-based methods to detect both sentiment and emotion from Bangla text. We perform comparative analysis among different feature extraction techniques and models. We find that the Convolutional neural network with word2vec (SG) approach is giving the best accuracy for sentiment classification and Bi-LSTM with word2vec (SG) is performing the best in terms of emotion detection. We also find that neural network-based methods outperform traditional machine learning-based approaches. This study will contribute highly in different applications of both "Natural Language Processing" and "Natural Language Understanding".

References

[1] BTRC, 2020, Internet Subscribers in Bangladesh February, accessed 1 February 2021, http://www.btrc.gov.bd/content/internet-subscribers-bangladesh-february-2020

[2] Statista, 2021, Facebook users by country 2020, accessed 1 February 2021, https://www.statista.com/statistics/268136/top-15-countries-based-on-number-of-facebook-users/

[3] Nasukawa, T. and J. Yi. Sentiment analysis: Capturing favorability using natural language processing. in Proceedings of the 2nd international conference on Knowledge capture. 2003. ACM.

[4] Pang, B., L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. in Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. 2002. Association for Computational Linguistics.

[5] Das, S. and M. Chen. Yahoo! for Amazon: Extracting market sentiment from stock message boards. in Proceedings of the Asia Pacific finance association annual conference (APFA). 2001. Bangkok, Thailand.

[6] Wiebe, J. Learning subjective adjectives from corpora. in AAAI/IAAI. 2000

[7] H. Shirani-Mehr, “Applications of deep learning to sentiment analysis of movie reviews,” in Technical Report. Stanford University, 2014.

[8] 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.

[9] 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.

[10] M. S. Islam, M. A. Islam, M. A. Hossain, and J. J. Dey, “Supervised approach of sentimentality extractionfrom bengalifacebook status,” in Computer and Information Technology (ICCIT), 2016 19th International Conference on. IEEE, 2016, pp. 383–387.

[11] 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.

[12] 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.

[13] A. Hassan, M. R. Amin, N. Mohammed, and A. Azad, “Sentiment analysis on bangla and romanizedbangla text (brbt) using deep recurrent models,” arXiv preprint arXiv:1610.00369, 2016.

[14] D. Das and S. Bandyopadhyay, “Developing bengaliwordnet affect for analyzing emotion,” in International Conference on the Computer Processing of Oriental Languages, 2010, pp. 35– 40.

[15] N. IrtizaTripto and M. Eunus Ali, "Detecting Multilabel Sentiment and Emotions from Bangla YouTube Comments," 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, 2018, pp. 1-6, doi: 10.1109/ICBSLP.2018.8554875.

[16] 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.

[17] Cho, Kyunghyun; van Merrienboer, Bart; Gulcehre, Caglar; Bahdanau, Dzmitry; Bougares, Fethi; Schwenk, Holger; Bengio, Yoshua (2014). "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation". arXiv:1406.1078

[18] 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.

[19] Chaffar, Soumaya, and Diana Inkpen. "Using a heterogeneous dataset for emotion analysis in text." In *Canadian conference on artificial intelligence*, pp. 62-67. Springer, Berlin, Heidelberg, 2011.

[20] Sharif, S. M. A., Nabeel Mohammed, SifatMomen, and NafeesMansoor. "Classification of bangla compound characters using a hog-cnn hybrid model." In *Proceedings of the International Conference on Computing and Communication Systems*, pp. 403-411. Springer, Singapore, 2018.

[21] Chowdhury, Shaika, and Wasifa Chowdhury. "Performing sentiment analysis in Bangla microblog posts." In *2014 International Conference on Informatics, Electronics & Vision (ICIEV)*, pp. 1-6. IEEE, 2014.

[22] Karim, Mohammad A., ed. *Technical challenges and design issues in bangla language processing*. IGI Global, 2013.

[23] Loper, Edward, and Steven Bird. "Nltk: The natural language toolkit." *arXiv preprint cs/0205028* (2002).

[24] Gerard Salton and Christopher Buckley. 1988. Term-weighting approaches in automatic text retrieval. Information processing & management 24, 5 (1988), 513–523.

[25] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013).