6th International Conference on Electrical Information and Communication Technology (EICT),
07～09 December 2023, Khulna, Bangladesh

2023 6th International Conference on Electrical Information and Communication Technology (EICT) | 979-8-3503-7083-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/EICT61409.2023.10427900

Bangla Emergency Post Classification on Social Media using Transformer Based BERT Models

Alvi Ahmmed Nabil
Department of Computer Science and Engineering(CSE)
Khulna University of Engineering & Technology(KUET)
Khulna-9203, Bangladesh
Email: alvi.nabil.aan@gmail.com

Dola Das
Department of Computer Science and Engineering(CSE)
Khulna University of Engineering & Technology(KUET)
Khulna-9203, Bangladesh
Email: dola.das@cse.kuet.ac.bd

Md. Shahidul Salim
Department of Computer Science and Engineering(CSE)
Khulna University of Engineering & Technology(KUET)
Khulna-9203, Bangladesh
Email: ss@cse.kuet.ac.bd

Shamsul Arifeen
Department of Computer Science and Engineering(CSE)
Khulna University of Engineering & Technology(KUET)
Khulna-9203, Bangladesh
Email: arifeen1707008@gmail.com

H. M. Abdul Fattah
Department of Computer Science and Engineering(CSE)
Khulna University of Engineering & Technology(KUET)
Khulna-9203, Bangladesh
Email: hussainfattah@cse.kuet.ac.bd

***Abstract*—Text classification is one of the most important tasks in Natural Language Processing. As text data is growing rapidly, it needs more computational power to classify the text in a big dataset. The task is difficult for characteristic-rich languages like Bangla. Having good-quality text data significantly affects the outcome of the model that has been used to classify them. Nowadays, social media can be an important source of information. But there is a huge number of data which are of no use. As the use of social media is increasing day by day, people are posting about events around their surroundings. So, it can be an important propaganda of the media. In this study, various text classification methods were used to classify the texts in Bangla from social media, which can be categorized as emergencies that may need immediate actions from the government, local authority, or law enforcement or even may need international attention. Therefore, 5839 social media posts were collected from Facebook and Twitter, which were written in Bangla along with some mixed English words. Then, after preprocessing, various Machine Learning models, Deep Neural Network models, and Transformer based models were applied to classify them. Among these models, transformer-based XLM-RoBERTa outperformed all the other models with an F1-score of 95.25.**

***Index Terms*—Text Classification, Transformers, BERT, Emergency Post, Natural Language Processing(NLP).**

I. Introduction

It is very important to efficiently respond to various emergencies, such as natural disasters, accidents, public health events, and social security events, which have caused significant loss of life and property worldwide. Social media has emerged as a valuable tool for disaster management due to its crowdsourcing capacity, with billions of users posting about emergencies and providing valuable information for situation awareness and damage assessment. Emergency text classification in the Bangla language is very rare, even though the language is widely spoken and popular.

The main objective of this research is to create a system that can classify Bangla Emergency posts from social media. Accurate classification of emergency-related posts in Bangla can be helpful by allowing the Bengali-speaking population to report emergencies in their native language, thus facilitating more convenient and efficient assistance and resource allocation by emergency services. First, we have developed a custom dataset by collecting data from different sources, like Facebook and Twitter. Then, we applied the transformer-based model XLM-RoBERTa and other transformer-based models in order to conduct training and testing.

We created a dataset containing Bangla social media posts that can be categorized as emergency posts, manually labeled them, and then applied different classification methods.

In our research, we have discovered that transformer-based classification models work better than other methods when social media posts have a mixture of languages.

The rest of the paper is divided into the following sections: In Section II, we discuss the related works on Bangla Text Classification. Section III presents the detailed methodology of our research, which includes the embedding and classification methods. In Section IV, we describe the dataset and compare the results among various methods. In Section V, we present our conclusion of this research.

II. Related Works

A. Traditional Machine Learning Based Approaches

Omar et al. [1] used logistic regression to identify potentially harmful texts in the Bangla language. The goal of the authors was to create a machine learning model that could detect false or malicious text in Bangla, which could be useful in detecting such content on the internet. The researchers compiled a dataset of Bangla text, labelling each

978-1-6654-0906-3/21/$31.00 ©2023 IEEE

sample as either suspicious or non-suspicious. The accuracy of the model was 93.77%. Azmin et al. [2] used a Naive Bayes classifier for determining emotions in Bangla text. The authors sought to create a machine-learning model that could recognize emotions in Bangla, with potential applications in sentiment analysis, customer service, and opinion mining. This model showed 83.33% accuracy. Dhar et al. [3] tried to create a system that can automatically classify Bangla web text documents into predefined categories. The authors did this by gathering a dataset of Bangla web text documents and preprocessing using the TF-IDF-ICF scheme to remove stop words and turn them into numerical feature representations.

B. Deep Learning Based Approaches

Huang et al. created SBEED [4] a framework for emergency event detection. The Weibo data were combined in a UTF-8 encoding format and created the emergency dataset. To classify the data, they used the Text-CNN model and calculated the text similarity using type, time, and location. Ghosh et al. [5] tried to detect depressive social media texts in the Bengali language. The authors propose a hybrid learning approach for this task, which combines two or more machine learning algorithms such as LSTM, GRU, and CNN. Alam et al. [6] developed an application of deep learning techniques for text categorization in the Bengali language. The authors propose a deep hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) network with word embedding for this task. They tried ANN, CNN, and CNN-LSTM hybrid architectures for their text classification. Rahman et al. presented [7] document classification in the Bengali language. The authors propose a deep recurrent neural network with Bidirectional Long Short-Term Memory (BiLSTM) for this task. The network architecture is designed to capture both the sequential and semantic information present in the text.

C. Transfomer-based Approaches

BanglaBERT [8] introduced a pre-trained language model using transformer architecture on Bengali text, showing superior performance over traditional machine learning models on various benchmark datasets like Named Entity Recognition, Part-of-Speech Tagging, and Sentiment Analysis in low-resource languages. DeepHateExplainer [9] introduced a deep learning system that is designed for identifying and interpreting hate speech in Bengali social media posts. It utilizes a mix of traditional machine learning and deep learning techniques, such as CNNs and LSTM networks, to effectively address the hate speech detection problem. Das et al. [10] utilized pre-trained BERT for emotion classification on a Bengali dataset and observed that the fine-tuned model outperformed SVM and MLP, demonstrating competitive performance in Bengali emotion classification. Alam et al. [11] explored the use of BERT and RoBERTa transformer-based models for Bangla text classification, comparing them against Support Vector Machines and Naive Bayes. The results demonstrate the superior performance of transformer-based models, achieving state-of-the-art results on two benchmark datasets: Bangla News articles and Bangla Sentiment Analysis.

Transformer Based Approaches are better than Traditional Machine Learning Based Approaches and Deep Learning Based Approaches. Because transformer Based Approaches hold the contextual information of the sentences.

III. Proposed Methodology

A. Model Architecture

Here, we used different variations of BERT architecture. BERT is mainly an encoder-based model that uses a subword tokenization algorithm. Fig 1 shows the basic architecture of the BERT model. The short description variations of BERT model are written below.

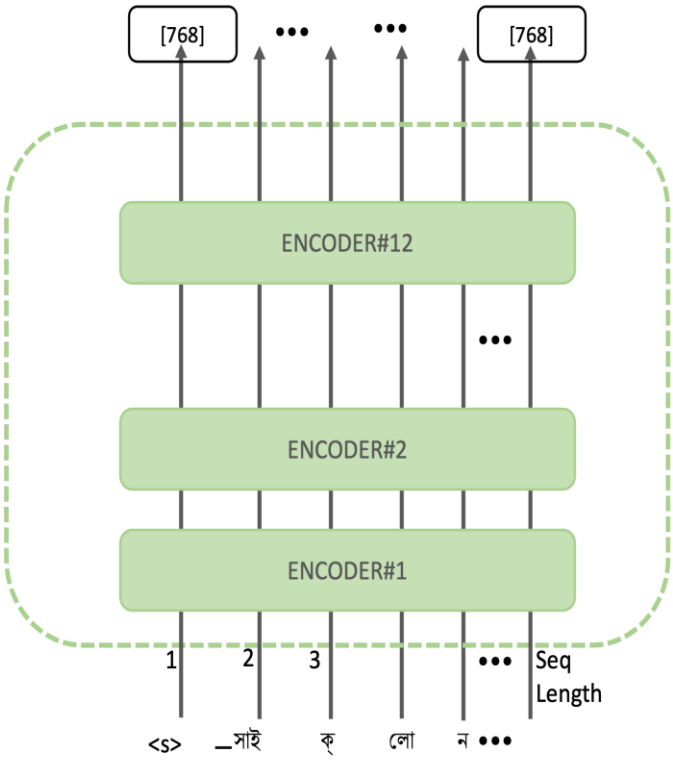

Fig. 1. Encoder architecture of BERT.

1) BERT: One of the most popular Transformer-based models is BERT, which stands for Bidirectional Encoder Representation of Transformers. BERT is used for feature extraction from a large text corpus. But these extracted features can be used for text classification. BERT generates a contextualized vector embedding for each word of the corpus as well as a special token vector that represents the overall contextualized information of that sample. BERT makes use of the encoder portion of the Transformer. Therefore, BERT can encode the syntactic and semantic information in the embedding necessary for a variety of tasks by utilizing the encoder. As said, BERT does not make use of the Transformer architecture's decoder component. Therefore, an embedding rather than a textual output is what BERT produces. This is

TABLE I
COMPARISON OF DIFFERENT VARIATIONS OF BERT MODEL

Model Name	Layers	Hidden Size	Parameters
mBERTbase	12	768	110M
mBERTlarge	24	1024	340M
BanglaBERT	12	768	110M
XLM-RoBERTabase	12	768	125M
XLM-RoBERTalarge	24	1027	355M

significant since it implies that we will need to work with the embedding in whichever application we use BERT for if the result is an embedding. For example, we can compare embeddings and get a similarity score by using methods like cosine similarity. For Bert Base, the context length is 768. This tells us that the BertBase model generates a tensor of size 768 for each word, and for Bert Large, it generates a tensor of size 1024 to hold the contextualized representation of the word in the post or sentence.

2) mBERT: To work with the Bangla Language BERT model needs to be trained on a large corpus of the Bangla language. For that reason, one of the models that we worked with is mBERT. This version of the BERT model was trained in 104 languages, whereas the vanilla BERT model was only trained in English. mBERT also has 2 varieties like BERT called mBERTbase and mBERTlarge, which has similar architecture to BERT.

3) BanglaBERT: Another variant of the BERT model is Bangla-BERT [12]. Bangla-Bert-Base is a pre-trained language model of the Bengali language using mask language modelling.

4) XLM-RoBERTa: The XLM-RoBERTa model [13] was proposed in Unsupervised Cross-lingual Representation Learning at Scale. It is based on Facebook's RoBERTa model, released in 2019. It is a large multi-lingual language model trained on 2.5TB of filtered Common Crawl data. XLM-RoBERTa is a multilingual model trained in 100 different languages. Unlike some XLM multilingual models, it does not require lang tensors to understand which language is used and should be able to determine the correct language from the input ids. Table I shows the different BERT model architecture details.

B. Proposed Framework

After collecting the data, our first task was to preprocess the data to remove all the redundant features. Preprocessing was done in 3 steps. Punctuation and Stop Word Removal, Common words removal and Stemming.

As most social media posts nowadays contain both Bangla and English words, we didn't remove any English words if the post has them in it. We also skipped this preprocessing step to test how the models work on raw data. After data collecting and optionally preprocessing, all the posts then went through tokenization, where each of the posts was represented in machine-readable vectors. Bert used the WordPiece subword-based tokenization algorithm and XLM-RoBERTa used SentencePiece tokenization. Fig 2 shows the WordPiece subword-based tokenization algorithm. Each of the words in the sequence length was converted to a vector of size 768 in this step. As the maximum length of the sentences is set to 100 for our test case, so each sentence then produces a tensor of size 100*768.

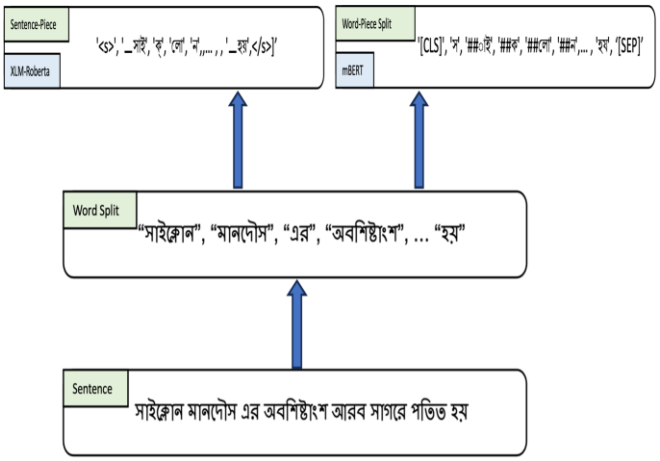

Fig. 2. WordPiece Subword-Based Tokenization

These embedding vectors are generated during the training phase that goes through self- attention layer and feed-forward neural network. After the embedding, we only take the [CLS] token's embedding for a classification task that goes through a dense layer and an activation function, which then is called "Pooler Output". The length of the "Pooler Output" tensor is 786. Finally, the "Pooler Output" goes through a linear layer that connects the tensor of length 768 to the number of classes in our case, which is 9. The one with the highest value is selected.

Fig 3 shows the flow diagram of our proposed model.

IV. EXPERIMENTS AND RESULTS

A. Environmental specifications

An Intel Core i7-8700K processor running at 3.7GHz, 16GB of RAM, a 512GB SSD for storage, and an NVIDIA GeForce GTX 1070ti GPU were used for the database creation and processing tasks. The model was written in Python 3.10 using a variety of libraries and tools, including PyTorch, Scipy, Numpy, Pandas, Scikit-Learn, Gensim, TensorFlow Keras, and HuggingFace pretrained models and tokenizers.

B. Experimental dataset

Data was gathered from Facebook posts, comments, and daily newspapers, as well as Twitter posts. We were able to record a variety of emergency posts and expressions thanks to this varied collection. It was difficult to collect structured data from social media for the Bangla language, in part because it

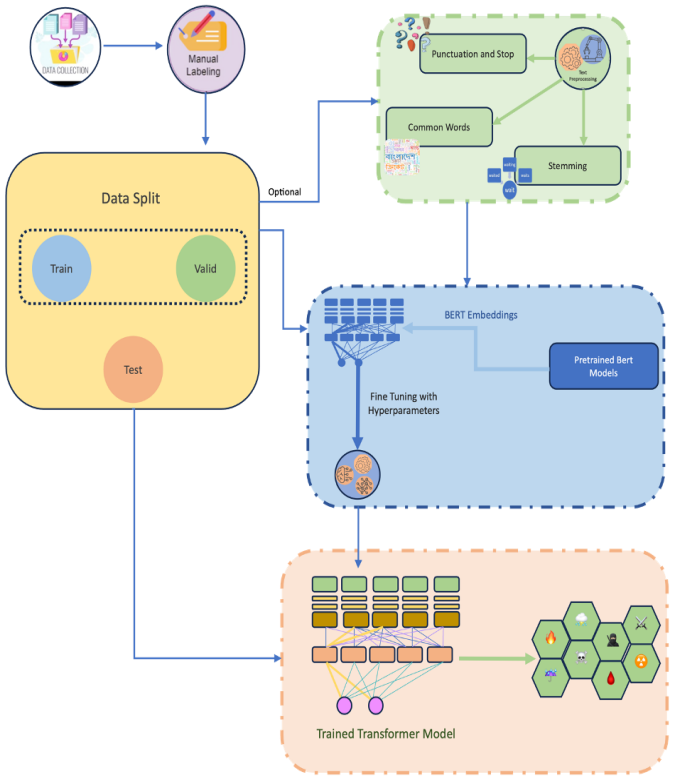

Fig. 3. Proposed Framework to classify emergency social media posts using Transformer Based BERT models.

was hard to figure out stemming rules and recognize words that should be eliminated. We manually labeled our dataset using our native Bangla proficiency to get around the problem. It is crucial to recognize that there may be disagreement over some labels, which could result in differences in the conclusions drawn from the same dataset. 3,270 training samples, 814 validation samples, and 1,752 test samples made up our dataset. The dataset was labeled with nominal categories, including accident, blood, crime, fire, natural disaster, pandemic, suicide, war, and weather. The whole data was then divided into 3 parts: Training, Validation, and testing. The training data was 56%, validation data was 14%, and test data was 30%. The dataset can be found here [14]

C. Performance evaluation metrics

The correct prediction of the model can be determined by the number of True Positive(TP) and True Negative(TN) while the wrong prediction is determined by the number of False Positive(FP) and False Negative(FN). Using these, we can calculate various evaluation metrics. For our research, we calculated the Accuracy, Precision, Recall and F1 score.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

TABLE II
SAMPLE DATA OF OUR DATASET

Posts	Class
১২ জনের করোনা শনাক্ত, ১১ জনই ঢাকার	pandemic
বিশ্বনাথের তেলিকোনা গ্রামে গভীর রাতে ভয়াবহ অগ্নিকাণ্ড	fire
ইউক্রেন জুড়ে রাশিয়ার ক্ষেপণাস্ত্র ও ড্রোন হামলায় নিহত ১১	war
O+ রক্তদাতারা এগিয়ে আসুন রোগীর সমস্যা: ক্যান্সার (রক্তশুন্যতা)	blood

$$Recall = \frac{TP}{TP + TN} \tag{3}$$

$$Accuracy = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

D. Results on Emergency Post Using Transformer-Based Models

Deep learning requires millions of data points, whereas machine learning uses thousands. Generally, machine learning algorithms exhibit good performance on fragile datasets. Large volumes of data are needed for Deep Learning to comprehend and outperform conventional machine learning algorithms. Table III shows this statement is true because our traditional machine-learning algorithms perform better than other deep-learning algorithms. The idea of pre-training on huge corpora and fine-tuning on task-specific data was first presented by BERT. With lesser amounts of labeled data, models can be fine-tuned for specific tasks after learning general language representations through pre-training. Long-range dependencies must be captured in many sequence-based tasks. The inability of conventional recurrent neural networks (RNNs) to capture long-term dependencies may be caused by disappearing or ballooning gradient issues. By using self-attention processes, transformers like BERT avoid this problem by being able to take into account the relationships between distant words in a sequence. The ability of BERT to capture word semantic associations has improved. Words are taught to be represented in a dense vector space where words that are similar are near to one another. Transformer models outperform other machine learning and deep learning models primarily for those reasons. However, we also observe that XLM-RoBERTa outperforms the other Bert models; this was shown in their masking challenge. In this sense, each sentence in BERT is essentially masked in ten different ways during the single masking operation that takes place during data preparation. As a result, the model will only encounter those ten different forms of each sentence throughout training. In Roberta, however, masking takes place during training. As a result, masking is completed for each sentence when it is included in a minibatch; as a result, unlike in BERT, there is no limit to the number of possible variations of each sentence's mask. Batch sizes of 8, 16, and 32 and learning rates of 0.00001, 0.0003, and 0.001 were the only parameter options available to us

TABLE III
COMPARISON OF EVALUATION METRICS BETWEEN DIFFERENT MODELS.

Model Name	Pr(%)	Re(%)	F1 score(%)
Logistic Regression	86.31	85.02	84.04
Multinomial Naive Bayes	74.17	66.07	60.76
K-Nearest Neighbors	83.80	82.83	82.01
LSTM	20.24	44.99	27.92
BiLSTM CNN	62.42	64.65	60.37
mBERT	93.77	93.66	93.59
BanglaBERT	94.73	94.67	94.63
XLM-RoBERTa	95.22	95.38	95.25

as Transformer model takes a lot of computing power. The mBERT, BanglaBERT, and XLM-RoBERTa models were the main subjects of our study. Following meticulous parameter adjustment, we discovered that the optimal testing accuracy for the mBERT model was 94.37% when using a batch size of 16 and a learning rate of 0.00003. For BanglaBERT, this accuracy was 95.20%.We kept the batch size and learning rate from our earlier studies for our XLM-RoBERTa model research. The best results were obtained with a batch size of 16 and a learning rate of 0.00001, yielding an exceptional testing accuracy of 95.38%. Fig 4 training accuracy, validation accuracy and Fig 5 shows the confusion matrix.

Table III shows the comparative analysis of different evaluation metrics for the different transformer models.

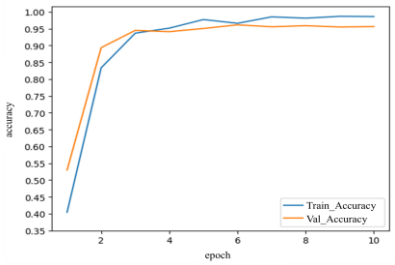

Fig. 4. Training and validation accuracy graph of XLM-RoBERTa.

V. CONCLUSION

We endeavored to develop an automated solution for detecting emergencies in Bangladesh, encompassing events such as earthquakes and natural disasters. To achieve this goal, we proposed the implementation of a transformer-based system, specifically XLM-RoBERTa. Our experiments entailed utilizing accidental emergency posts gathered from various

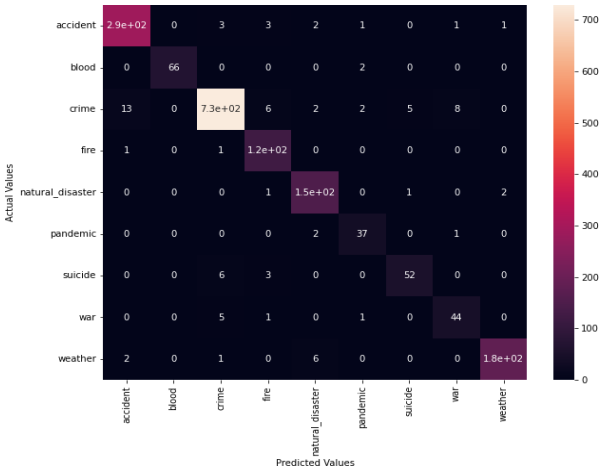

Fig. 5. Multiclass confusion matrix tested with XLM-RoBERTa.

sources such as daily newspapers and Facebook. To validate the effectiveness of our model, we employed different DNN transformer-based models, as well as traditional machine learning approaches, on our newly constructed emergency posts dataset. Notably, our results indicated that XLM-RoBERTa, leveraging self-attention and feed-forward neural network mechanisms to capture contextual dependencies within sentences, outperformed the DNN models. However, Our system categorizes some sentences as a natural disaster, which is actually an accident. These illustrations highlight the difficulties presented by ambiguous and varied linguistic patterns in emergency posts. Looking ahead, we envision further enhancements to our classifier's efficiency. This involves refining our pre-processing and feature engineering steps to enhance the quality of the input data. Additionally, we plan to explore the integration of multi-task learning, enabling the model to tackle multiple related classification tasks, including classifying the type of emergency and assessing the level of urgency.

REFERENCES

[1] Sharif Omar, Mohammed Moshiul Hoque, "Automatic detection of suspicious Bangla text using logistic regression.," Intelligent Computing and Optimization: Proceedings of the 2nd International Conference on Intelligent Computing and Optimization 2019 (ICO 2019), pp. 581-590, 2020.

[2] Azmin, Sara, Kingshuk Dhar., "Emotion detection from bangla text corpus using naive bayes classifier," th International Conference on Electrical Information and Communication Technology (EICT), pp. 1-5, 2019.

[3] Dhar, Ankita, Niladri Sekhar Dash, Kaushik Roy, "Categorization of Bangla web text documents based on TF-IDF-ICF text analysis scheme.," ocial Transformation–Digital Way: 52nd Annual Convention of the Computer Society of India, CSI 2017, Kolkata, India, vol. 52, pp. 477-484, 2018.

[4] Huang, Lida, Gang Liu, Tao Chen, Hongyong Yuan, Panpan Shi, Yujia Miao, "Similarity-based emergency event detection in social media," Journal of Safety Science and Resilience, vol. 2, no. 1, pp. 11-19, 2021.

[5] Ghosh Tapotosh, M. Shamim Kaiser, "Bangla Depressive Social Media Text Detection Using Hybrid Deep Learning Approach," Third International Conference on Trends in Computational and Cognitive Engineering: TCCE 2021, pp. 111-120.

[6] Alam S, Haque MA, Rahman A, ”Bengali Text Categorization Based on Deep Hybrid CNN–LSTM Network with Word Embedding,” International Conference on Innovations in Science, Engineering and Technology (ICISET) 2022, pp. 577-582, 2022.
[7] Rahman S, Chakraborty P, ”Bangla document classification using deep recurrent neural network with BiLSTM,” International Conference on Machine Intelligence and Data Science Applications: MIDAS 2020, pp. 507-519, 2021.
[8] Bhattacharjee A, Hasan T, Samin K, Rahman MS, Iqbal A, Shahriyar R. Banglabert, ”Combating embedding barrier for low-resource language understanding. arXiv preprint arXiv:2101.00204.,” 2021.
[9] BKarim MR, Dey SK, Islam T, Sarker S, Menon MH, Hossain K, Hossain MA, Decker S., ”Deephateexplainer: Explainable hate speech detection in under-resourced Bengali language.,” In2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA) 2021, pp. 1-10, 2021.
[10] Das A, Sharif O, Hoque MM, Sarker IH., ”Emotion classification in a resource-constrained language using the transformer-based approach,” 2021.
[11] Alam T, Khan A, Alam F., ”Bangla text classification using transformers,” 2020.
[12] S. Sarker, ‘BanglaBERT: Bengali Mask Language Model for Bengali Language Understanding’. 2020.
[13] Conneau, Alexis, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzma´n, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. ”Unsupervised cross-lingual representation learning at scale.” arXiv preprint arXiv:1911.02116 (2019).
[14] https://github.com/AlviNabil/Bangla-Emergency-Post-From-Social-Media