MULTIMODAL CYBERBULLYING MEME DETECTION FROM SOCIAL MEDIA USING DEEP LEARNING APPROACH

Md. Tofael Ahmed[1, 3], Nahida Akter [1], Maqsudur Rahman[1], Abu Zafor Muhammad Touhidul Islam[2], Dipankar Das[3] and Md. Golam Rashed[3]

[1]Department of Information and Communication Technology, Comilla University, Bangladesh.
[2]Department of Electrical & Electronics Engineering, University of Rajshahi, Bangladesh.
[3]Department of Information and Communication Engineering,
University of Rajshahi, Bangladesh.

ABSTRACT

Cyberbullying includes the repeated and intentional use of digital technology to target another person with threats, harassment, or public humiliation. One of the techniques of Cyberbullying is sharing bullying memes on social media, which has increased enormously in recent years. Memes are images and texts overlapped and sometimes together they present concepts that become dubious if one of them is absent. Here, we propose a unified deep neural model for detecting bullying memes on social media. In the proposed unified architecture, VGG16-BiLSTM, consists of a VGG16 convolutional neural network for predicting the visual bullying content and a BiLSTM with one-dimensional convolution for predicting the textual bullying content. The meme is discretized by extracting the text from the image using OCR. The perceptron-based feature-level strategy for multimodal learning is used to dynamically combine the features of discrete modalities and output the final category as bullying or nonbullying type. We also create a "bullying Bengali memes dataset" for experimental evaluation. Our proposed model attained an accuracy of 87% with an F1 score of 88%. The proposed model demonstrates the capability to detect instances of Cyberbullying involving Bengali memes on various social media platforms. Consequently, it can be utilized to implement effective filtering mechanisms aimed at mitigating the prevalence of Cyberbullying.

KEYWORDS

Cyberbullying, Multimodality, Bengali Meme, Deep Learning, NLP (Natural Language Processing), Machine Vision

1. INTRODUCTION

Bullying is a harmful social problem that is spreading at a frightening rate. Bullying behaviour can be broadly divided into categories based on the following factors: type of behaviour (verbal, social, and physical), environment or platform (in person and online), mode (direct and indirect), visibility (overt and covert), and damage caused (physical and psychological), and context (location of occurrence such as home, workplace, school etc.)[1]. Cyberbullying is often covert social behaviour bullying that occurs online and causes short- and long-term psychological harmful effect for the sufferers. Online users have developed indictable and illegal ways to hurt

[1] https ://bullyingnoway.gov.au/WhatIsBullying/Pages /Types-of-bullying.aspx.

DOI: 10.5121/ijcsit.2023.15403

and humiliate people through hostile comments, memes, videos, GIFs etc. On online platforms or apps due to the increased availability of reasonable data services and social media presence Cyberbullying is very common incident in such a platform. Cyberbullying is even more harmful than face-to-face bullying because of its persistence, audience size, and speed at which damage is done. Victims of Cyberbullying have severe mental health and wellbeing concerns and overwhelming feelings. Cyberbullying can make its victims more distressed and cause low selfesteem, annoyance, frustration, sadness, social disengagement, and, in rare circumstances, the emergence of violent or suicidal tendencies [1][2].

Because of technological development, bullies can remain anonymous, difficult to find, and shielded from conflict. The victims of Cyberbullying feel as though it never ends and is intrusive. In light of this, it is of the utmost need to locate viable solutions that can detect and prevent the emotional and psychological anguish that victims are forced to endure. The prompt and accurate identification of potentially harmful posts is essential for effective prevention [3]. To proactively identify potential threats, sophisticated automatic systems are required due to the information overload on the chaotic and complex social media sites. Researchers worldwide are working to create new approaches to identify, control, and lessen the prevalence of Cyberbullying in different languages [2][4]. To effectively process, analyze, and model such sour, taunting, abusive, or unpleasant information in photos, memes, or text messages, state-of-the-art computational methods and analytical tools are necessary. Memes and other imagebased, intersexual content have become more common in social feeds in recent years [3][4].

Cyberbullying using a variety of content formats is reasonably widespread. The present barriers to identifying online bullying posts are social media specialization, topic reliance, and various hand-crafted elements. With end-to-end training and representation learning capabilities, deep learning approaches demonstrate their worth and produce cutting-edge results for various natural language problems [5]. Relevant works describe identifying bullying content by assessing textual, picture-based, and user data using deep learning models like CNN, RNN, and semantic image features [6][7]. However, the text-based analytics has been the focus of the most researches on online cyberaggression, harassment and toxicity detection. A few related studies have used image analysis to assess bullying content. However, visual text, such as memes that are blended with text and image, has received the most miniature exploration in the literature. An innocent text can convey a bullying sense while embedded with a specified image and vice versa. So only text or only images cannot imply the actual purpose of a meme. We have to consider both of the modalities for identifying them. Only some works have been done on memes (text embedded with images) in high resource languages, like English. Although, Bengali is one of the most widely spoken languages in the world, with 230 million speakers in Bangladesh and India and it is the 7th most commonly spoken language, spoken by about 245 million worldwide [8]. However, there are no or very little works have been done in Cyberbullying detection with mems because of limitation of available resources. Anti-social behaviour is becoming more common in Bengali, much like in other crucial languages like English.

Bengali is a very diverse and rich language. Still, it is severely under-resourced for natural language processing (NLP), primarily because it lacks the computational tools required for various NLP tasks, such as language models, labeled datasets, and effective machine learning (ML) and deep learning techniques[9][10]. Additionally, memes are frequently used content in social media. In recent years, these memes have been a significant cause of Cyberbullying. Figure 1 displays various bullying-related Bengali memes. Even though several studies have been done using Bengali textual content, we discovered only one survey on Bengali multimodal meme content. These factors encouraged us to do research work on this topic. This article proposes deep neural network architecture for predicting bullying content in Bengali memes (Bangla text combined with image).The main focus of the proposed research:

- We proposed a hybrid VGG16-BiLSTM unified deep architecture that combines a BiLSTM model for textual bullying content prediction and a VGG16 network for visual bullying content prediction.
- Use of the EasyOCR to separate the text and image and to discretize the meme's content.
- The proposed hybrid architecture processes the textual and visual components, and the early-fusion decision layer is then applied to output the final prediction.
- Creation of a *Bengali memes dataset* that contains one thousand two hundred memes from different social media platforms, including Facebook, Instagram, and Twitter, and use of this dataset to validate the performance of VGG16-BiLSTM.

Thus, for adequate decision support in identifying Cyberbullying from Bengali memes, our unifying model considers the various modality contents and analyses each modality type using deep neural learning approaches. The paper is organized as follows: The related research work is covered in Section 2, and Section 3 describes the suggested VGG16-BiLSTM model for detecting Cyberbullying in multimodal Bengali meme content. The findings are presented in Section 4, followed by the conclusion and future study in Section 5.

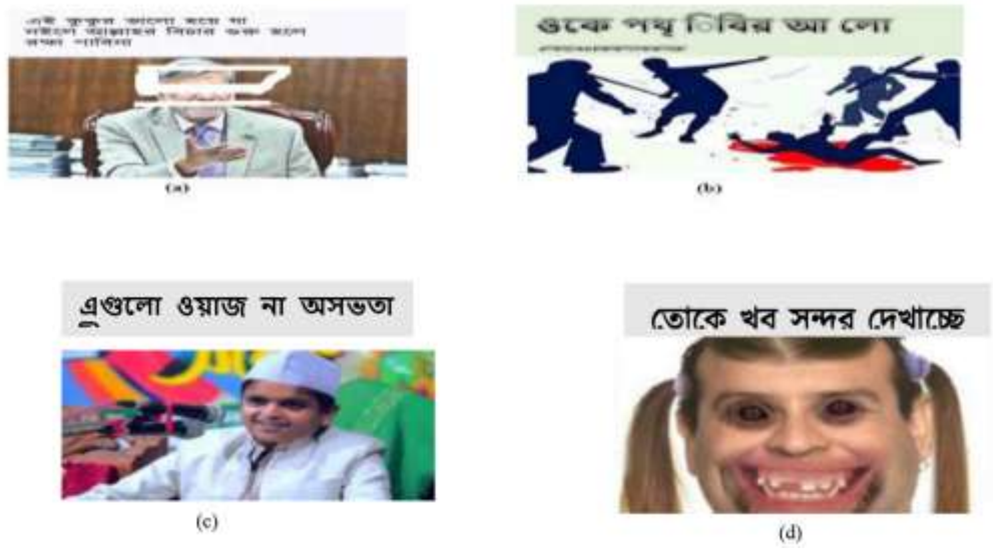


Figure 1. Some multimodal Bengali memes, where the combination of text and image creates actual meaning

2. RELATED WORK

We can exchange a plenty of information in the "Virtual Society" over the internet. We express ourselves, including our knowledge, beliefs, and opinions. Cyberbullying is undoubtedly a significant social issue, given the growing user base and reach of the internet. Researchers are mainly concerned with detecting Cyberbullying and hate speech in Bengali and other languages. Although the numerous studies have been conducted on this topic, however, the majority of which researchers are concentrated on a particular modality using only textual information. To identify hate speech, Romim et al. [11] gathered 50200 offensive Bangla comments from online social networking sites. They found the F1-score of 91% using the SVM and BiLSTM based approaches. Compared to previous pre-trained embedding's, they made the discovery that word embedding that was trained just with 1.47 million comments from social networking sites showed consistently improved modeling of hate speech detection. Ahmed et al.[12] proposed a machine learning model to detect Cyberbullying from Bangla and Romanized Bangla text using Multimodal Nave Bayes, SVM, logistic Regression, and XGBoost. They gathered comments in both Bangla and Romanized Bangla from videos of several well-known social media personalities.They found the accuracy of 76% for comments in Bangla and 84% accuracy for words in Romanized Bangla. Encoder-decoder-based ML model was suggested by Das et al.[13] to categorize user comments in Bengali from Facebook pages. They divided the total of 7425 Bengali comments into seven categories. They employed attention mechanisms, LSTMs, and GRU-based decoders and discovered that attention-based decoders had the best accuracy (77%). Ahmed et al.[14] used deep neural network (DNN) models to identify Cyberbullying from Bangla comments on social media. The comments were further divided into five categories: non-bully, sexual, threat, troll, and religious. They discovered 85% accuracy for multiclass classifiers and 87.91% for binary classifiers. Machine learning (ML) algorithms were employed by Ghosh et al. [15] to identify Cyberbullying on social media. They used TFIDF feature extraction at the N-gram level. Passive aggressive classifiers were found to have an accuracy of 78.1% compared to Random forest, SVM, Logistic Regression, and other classifiers. Karim et al. [16] suggested a Deep Hate Explainer model for Bengali language hate speech detection. They made use of several machine-learning methods and transformer-based neural architectures. They discovered that the transformer-based BERT algorithm outperforms ML and DNN with an F1 score of 88%. To categorize user Bangla comments posted on Facebook pages, Ishman and Sharmin [17] constructed a model using ML techniques and a GRU-based DNN model. They obtained 52.20% accuracy in the Random Forest method and 70.10% accuracy in the GRU-based model. Mridha et al. [18] completed another study on detecting offensive Bengali text. The approach improved the AdaBoost method by incorporating BERT and LSTM models and proposed the L-Boost model. On three different datasets, including the BERT pre-trained word-embedding vector model, they tested the L-Boost model and discovered an accuracy of 95.11%. A new dataset was created by Romim et al.[19] which contained 30000 Bengali-language comments colleccted from YouTube and Facebook comment sections. They used word embedding tools like Word2Vec, FastText, and BengFastText with DNN models. They demonstrated that SVM had the highest accuracy, reaching 87.5%.

Identifying hate speech or Cyberbullying from multimodal content like memes (text embedded with images) has become a prevalent topic in recent years. There have been a few studies on it, however most of them have focused on English multimodal posts or memes. Kumari et al. [20] proposed a model that identified the bullying comments containing texts along with images. They created a multimodal dataset with 2100 multimodal comments. For the information encoding, they employed a single-layer CNN model. It was found that single-layer CNN provides superior outcomes (recall value of 74%) for 2D representation. Another work had done on the same dataset by Kumari and Singh [21]. They employed CNN to extract features from texts and VGG16 to extract features from images. Finally, they extracted the optimum features from both text and images using a genetic algorithm. They obtained the F1-score of 78%, which was 9% higher than previously published findings on the same dataset. A Deep neural model for detecting Cyberbullying was proposed by Kumar and Sachdeva [22] across three different modalities of social media data, including textual, visual, and info graphic. They presented the CapsNet-ConvNet model for predicting bullying content. ConvNet was employed for visual content and CapsNet for text-based content. They attained 98% of AUC-ROC performance. We have found only single work on multimodal Bengali contents, model for identifying hate speech in multimodal Bengali texts and memes was proposed by Karim et al. [23]. They created a multimodal dataset for identifying hate speech and used XLM-RoBERTa, Bi-LSTM, Conv-LSTM, Monolingual Bangla BERT, and Multilingual BERT. They achieved the best result, 82% of F1-score. They stated that ResNet152 and DenseNet201 models produced F1-scores of 78% and 70%, respectively, for memes. However, they mostly paid attention on textual comments. The research goal presented in this paper is to develop a multimodal deep-learning hybrid model that can detect Cyberbullying from Bengali memes.

3. Proposed Methodology

The specifics of proposed research strategy are covered in this part, including word embedding's, multimodal learning, and the training of several neural network designs (ML/DNN/transformers).

3.1. Data Set

As far as our knowledge, there is no publicly available bullying Bengali meme dataset. One of the contribution in this research purpose, we created a Bengali memes dataset for bullying. To create the dataset, the bullying images are collected from the social media sites such as Facebook, Twitter, and Instagram by using specific search terms like "ugly," "fat," "animal," "cartoons of people," and "controversial political and media figures," "violation," etc. Using the same search query, and we used Google as a source for image searches. One thousand two hundred images in total are gathered from the different sources. We produced a short comment section for each image, solicited feedback from a few undergraduate students, and then completed the comments following the consensus. For making memes, we create a meme generator. We provide the images and corresponding comments as the meme generator's input and find the memes as the output. In our dataset, one column represents the images, another denotes the comments, and the last represents the memes' labels. We have a total of 1200 memes, from which 600 memes convey a bullying sense and the rest 600 memes convey a non-bullying sense. The distribution in a dataset is shown in Figure 2. And a sample of our dataset is shown in Figure 3.

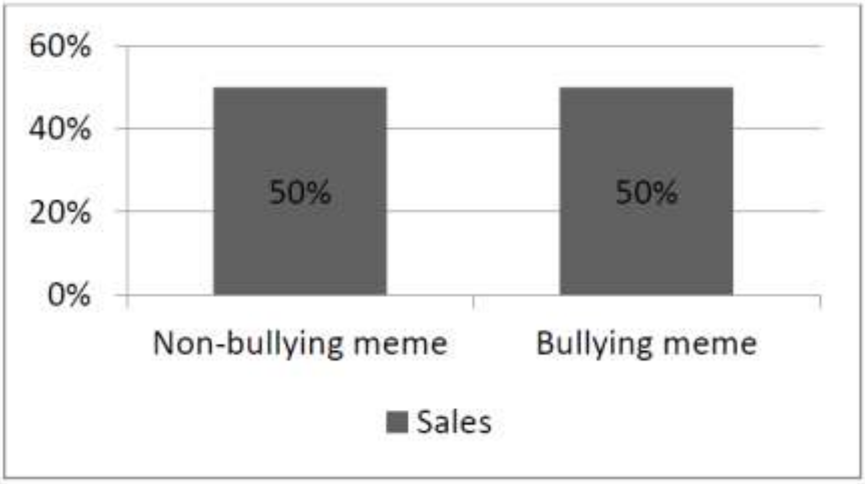
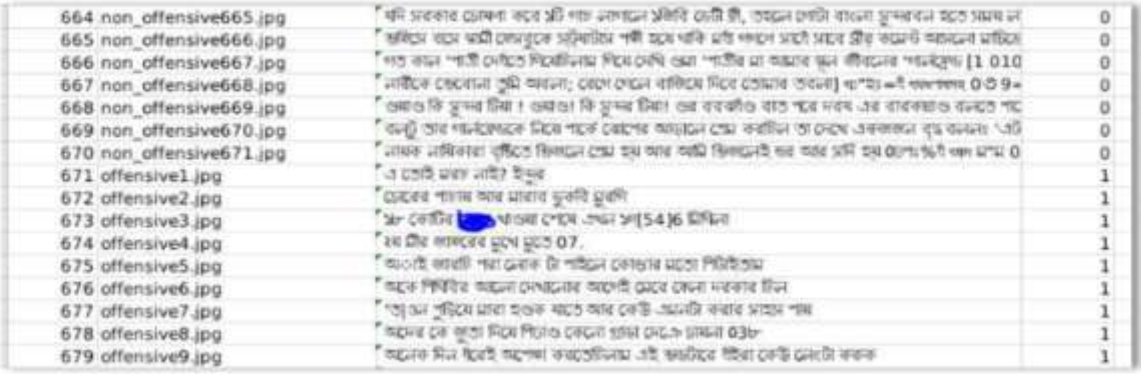


Figure 2. Bullying and Non-bullying Meme Distribution in Dataset



664	non_offensive665.jpg	[illegible]	0
665	non_offensive666.jpg	[illegible]	0
666	non_offensive667.jpg	[illegible] [1 010	0
667	non_offensive668.jpg	[illegible] 0 ৩ 9	0
668	non_offensive669.jpg	[illegible]	0
669	non_offensive670.jpg	[illegible]	0
670	non_offensive671.jpg	[illegible] 0	0
671	offensive1.jpg	[illegible]	1
672	offensive2.jpg	[illegible]	1
673	offensive3.jpg	[illegible] [redacted] [illegible] [54]6 [illegible]	1
674	offensive4.jpg	[illegible] 07.	1
675	offensive5.jpg	[illegible]	1
676	offensive6.jpg	[illegible]	1
677	offensive7.jpg	[illegible]	1
678	offensive8.jpg	[illegible] 03৮	1
679	offensive9.jpg	[illegible]	1

Figure 3. Sample Dataset

4. Proposed VGG16-BiLSTM Model

We proposed a deep neural network model which deals with the multimodal content (memes) in social media. The proposed model contains four blocks: block 1: modality discretization block, block 2: image processing block, block 3: text processing block, and block 4: classification block. The model is shown in Figure 4. The details of each block are given in the following subsections.

4.1. Block 1: Modality Discretization

As we have multimodal memes, we need to extract the text from the multimodal memes so that the text and image can be passed through the respective blocks. We use EasyOCR [24] in our model to extract the text. This OCR supports more than 80 plus languages. The EasyOCR is first downloaded and installed and then imported into our environment. OCR has processed all multimodal memes. The OCR extracts the texts, and we store them in a text document. The readers are passed through the text processing block, and the images are passed through the image processing block.

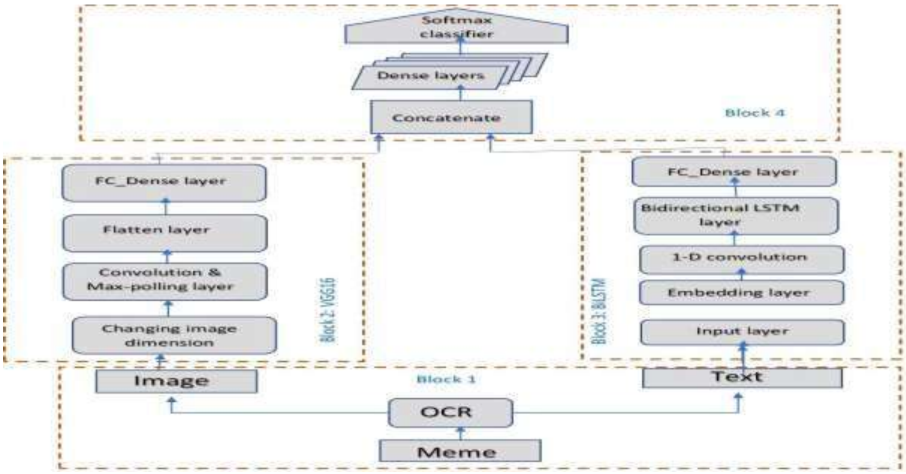


Figure 4. Proposed VGG16-BiLSTM Models

4.2. Block 2: Image Processing Block: VGG16 Network

For extracting features from images, we employ Fine-tuned VGG16 (Visual Geometry Group 16). It is a per-trained model for images and is used for extracting the features from images. There are 16 layers altogether, of which 13 are convolutional layers, and the final three are dense (or fully connected) layers. VGG16 takes an image as input whose size is (224×224× 3), and executes the convolutional operation, using (3×3) filters size in the convolution layer. We resize each image to (150×150×3) because the input images are in various sizes. As we mentioned earlier, the VGG16 model takes image sized (224×224×3); we change the input image shape dimension from (224×224×3) to (150×150×3) for our model. The image is then passed through the convolutional layers and the max-pooling layers.

Then the image is passed through flattened layer followed by two fully-connected dense layers. The size of the first fully-connected layer is 1024, and the size of the second fully-connected layer is 512. Therefore, the VGG16 network produces a feature vector of size 512. In our model, the VGG16 network is non-trainable until the two fullyconnected dense layers are fully connected. We use the ReLU activation function in every hidden layer. The ReLU activation function can be described using equation (1). For the negative value of Z ReLU activation function returns 0, and it returns Z for the positive value of Z.

eL activation function $f(z) = max(0, z)$ (1)

4.3. Block 3: Text Processing Block (Bidirectional LSTM)

For text processing, we use two-layered Bidirectional LSTM (Long Short Term Memory) named as BiLSTM. It is preceded by three one-dimensional convolution layers and followed by a single fully-connected dense layer. We keep a restriction that each text component is represented by 100 words only. The text is post-padded with zero if its length is less than 100 words. On the other hand, if the text is more than 100 words, then it is truncated to store the first 100 words only. The texts are embedded using the 300-dimensional pre-trained GloVe [24] embedding. Three one-dimensional convolution layers are used, followed by two bidirectional LSTM hidden layers. In the first convolutional layer, we use 32 filters of size 5.

In the second convolutional layer, we use 60 filters of size four, in the third or last convolutional layer, we use 100 filters of size 3. The subsequent two layers are bidirectional LSTM layers. Each of the bidirectional LSTM layers has 30 neurons. We kept the return sequences true for the first bidirectional LSTM layer because its output enters as the input of the second bidirectional LSTM layer. Then we apply a fully-connected dense layer. The size of the thick layer is 512. The text processing block produces a text feature vector of size 512. We use the ReLU activation function here to activate the neurons. Compared to the Sigmoid and Hyperbolic tangent, the ReLU activation function to handle successfully the vanishing gradient problem.

4.4. Block 4: Classification

In the final classification purposes, fusion between the features retrieved from textual and visual contents are taken into consideration. Multimodal fusion typically uses one of two approaches: model-free or model-level [24]. Modelfree fusion can be further divided into early fusion (feature-level) and late fusion (decision-level). Various input features are concatenated and fed into a classifier in early fusion. On the other hand, in late fusion, the predictions of multiple classifiers trained for different types of inputs are combined to give us the final result. In our model, we use early fusion. Because early fusion is carried out at the feature level, in this case, one large feature vector is created, concatenating the feature vectors from different sources and it will be used for classification.

Since vector contains a large number of features, training and classification will take longer. However, a large-size feature vector combined with appropriate learning techniques can result in significantly superior performance just one learning phase is required [25]. Therefore, the features extracted from the text and image processing blocks are concatenated to create a combined feature set of size 1024 (512 from images and 512 from texts). These concatenated features are passed through the four fully-connected dense layers. The first dense layer had 1024 neurons. After the first dense layer, we use the dropout layer to avoid the over fitting problem. The second, third and fourth dense layers have 512, 256, and 128 neurons, respectively. We use the ReLU activation function in all of our hidden dense layers. Finally, we used an output-dense layer which has two neurons and used the Softmax activation function with categorical cross-entropy (CE) as a loss function. The Softmax function activates one of the two neurons of the output layer, and it provides the final prediction result. The equation for the Softmax activation function and categorical cross-entropy (CE) is given in equations (2) and (3) respectively. Softmax function applies the usual exponential function to each element zi of the input vector z and divides the results by the total number of exponentials to normalize the values. This normalization makes ensure that the output vector's component sums equal to one.

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad (2)$$

Categorical cross-entropy $CE = -\sum_{i=1}^{N} y_i \log y_i'$, (3)

5. RESULTS AND DISCUSSIONS

We used the Scikitlearn and Keras deep learning libraries with Tensor Flow as the backend. Two separate models are utilized based on the input type, and their respective hyper parameters are tuned. Table 1 lists the model hyper parameters that were taken into consideration. Accuracy, precision, recall, and F1-Score are the parameters used to measure the model's performance. The equations for precision, recall, and F1-score are given in equations (4), (5), and (6), respectively. The proposed model achieves the performance of accuracy of 0.87, precision of 0.90, recall of 0.86, and finally we get the F1-score of 0.88. Table 2 shows the performance results of our model. Where, $\boldsymbol{TP}$: True Positives, $\boldsymbol{TN}$: True Negatives, $\boldsymbol{FP}$: False Positives, $\boldsymbol{FN}$: False Negatives.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

Table 1. Hyper parameters of the Proposed Model

Model	Hyper-parameters	Value
VGG16-BiLSTM	Image size	150×150×3
	Maximum text length	100
	Embedding dimension	300
	Number of filters	32,60,100
	Size of filters	5,4,3
	Activation function	ReLU, Softmax
	Loss Function	Categorical cross-entropy
	Optimizer	Adam
	Dorp out	0.1
	Epoch	100

Table 2. Performance Analysis Results of Proposed Model

Model	Results			
	Accuracy	Precision	Recall	F1-score
VGG16-BiLSTM	0.87	0.90	0.86	0.88

We utilized VGG16 to extract features from images, which performs incredibly better. Because an enormous amount of visual data is used to pre-train the VGG16 model. We used the weights values learn from that dataset as a starting point. Thus, even though our dataset is limited, the model can still extract the images' hidden features accurately. We updated the fully-connected dense layer of VGG16 to find a better performance. We obtained found the best result using 1024 and 512 neurons in the dense layers, respectively. Since our dataset is tiny, utilizing word vector embedding directly rarely produces good results. For this reason, we used glove2vec pre-trained word vector embedding for a bidirectional LSTM neural network. Three 1-dimensional convolutional layers with varied numbers of filters and varying filter sizes were employed before the BiLSTM layers. This gave us the noticed better results than only using bidirectional LSTM layers. We also experimented with various sizes of feature vector sizes of texts and images. The optimized result is obtained by using the feature vector size of 512 for each. After merging the features and implementing four fully-connected dense layers, the most significant 128 features were evaluated for predicting the final output. The epoch versus accuracy and loss curve for the proposed model is shown in Figure 5.

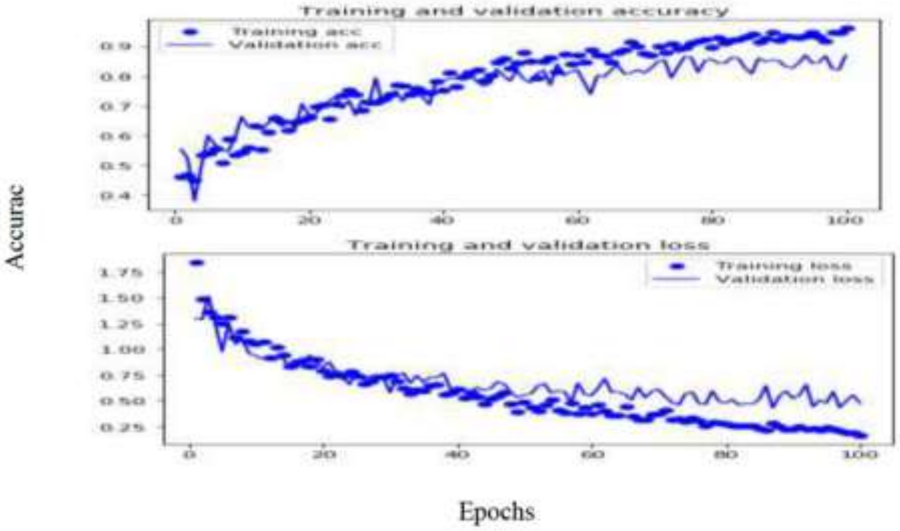


Figure 5. Epochs vs. Accuracy and Loss of the Model

6. Conclusion and Future Work

The internet and social media have opened up new modes of communication, oppression, and empowerment. Cyberbullying is a problematic and it is getting worse due to the heightened mental health hazards associated with the use of social media. Multimodality is a widespread form of Cyberbullying. Thus, building a predictive modeling is crucial to identifying such kind of bullying activities. The proposed research develop a hybrid deep neural network model to detect Cyberbullying from Bengali memes (text with image), and we found a satisfactory performance of the model.

In the future, we increase our dataset and try to include more dimensions. The transformer-based techniques BERT, Bangla-BERT, ELECTRA, XLNet, RoBERTa, and Distil BERT have recently seen use in the NLP field. We have a plan to use these methods and will analyze their performance. For images, we will use other CNN methods and analyze the performance. We will also emphasize working with other modalities, such as audio and video.

References

[1] M. A. Campbell, "Cyber Bullying An Old Problem in a New Guise?," *Aust. J. Guid. Couns.*, vol. 15, no. 1, pp. 68–76, 2005, doi: 10.1375/ajgc.15.1.68.

[2] .Pawar and . . aje, "Multilingual cyberbullying detection system," *IEEE Int. Conf. Electro Inf. Technol.*, vol. 2019-May, pp. 040–044, 2019, doi: 10.1109/EIT.2019.8833846.

[3] Velioglu and J. ose, "Detecting Hate peech in Memes sing Multimodal Deep Learning Approaches Prizewinning solution to Hateful Memes Challenge," Dec. 2020, [Online]. A*vailable: http://arxiv.org/abs/2012.12975.*

[4] S. Pramanick, . harma, D. Dimitrov, M. . Akhtar, P. Nakov, and T. Chakraborty, "MOMENTA A Multimodal Framework for Detecting Harmful Memes and Their Targets," ep. 2021, [Online]. Available *http://arxiv.org/abs/2109.05184.*

[5] T. Young, D. Hazarika, . Poria, and E. Cambria, " ecent trends in deep learning based natural language processing [eview Article]," *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, 2018, doi: 10.1109/MCI.2018.2840738.

[6] A. Solanki, S. Kumar, and A. Nayyar, *Emerging Trends and Applications of Machine Learning*, vol. i. 2020.

[7] M. Dadvar and K. Eckert, “Cyberbullying detection in social networks using deep learning based models,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12393 LNCS, pp. 245–255, 2020, doi: 10.1007/978-3-030-59065-9_20.

[8] F. Alam *et al.*, “A eview of Bangla Natural Language Processing Tasks and the tility of Transformer Models,” 2021, [Online]. Available: *http://arxiv.org/abs/2107.03844.*

[9] O. Sen *et al.*, “Bangla natural language processing A comprehensive analysis of classical, machine learning, and deep learning-based methods,” *IEEE Access*, vol. 10, pp. 38999–39044, 2022, doi: 10.1109/ACCESS.2022.3165563.

[10] M. . Karim, B. . Chakravarthi, J. P. McCrae, and M. Cochez, “Classification Benchmarks for nder-resourced Bengali Language based on Multichannel Convolutional-L TM Network,” Apr. 2020, [Online]. Available *http://arxiv.org/abs/2004.07807.*

[11] N. omim, M. Ahmed, M. . Islam, A. en harma, H. Talukder, and M. . Amin, “BD-SHS: A Benchmark Dataset for Learning to Detect Online Bangla Hate peech in Different ocial Contexts,” *Proc. Lang. Resour. Eval. Conf.*, pp. 5153– 5162, 2022, [Online]. Available: https://aclanthology.org/2022.lrec-1.552.

[12] M. T. Ahmed, M. ahman, . Nur, A. Z. M. T. Islam, and D. Das, “Natural language processing and machine learning based cyberbullying detection for Bangla and omanized Bangla texts,” *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 20, no. 1, pp. 89–97, 2022, doi: 10.12928/TELKOMNIKA.v20i1.18630.

[13] A. K. Das, A. Al Asif, A. Paul, and M. N. Hossain, “Bangla hate speech detection on social media using attention-based recurrent neural network,” *J. Intell. Syst.*, vol. 30, no. 1, pp. 578–591, 2021, doi: 10.1515/jisys-2020-0060.

[14] M. F. Ahmed, Z. Mahmud, Z. T. Biash, A. A. N. yen, and ..., “Cyberbullying detection using deep neural network from social media comments in bangla language,” *arXiv Prepr. arXs* 2021, [Online]. Available: *https://arxiv.org/abs/2106.04506%0Ahttps://arxiv.org/pdf/2106.04506.*

[15] R.Ghosh, B. T. tudent, . Nowal, and G. Manju, “IJE T-Social Media Cyberbullying Detection using Machine Learning in Bengali Language ocial Media Cyberbullying Detection using Machine Learning in Bengali Language,” *IJERT J. Int. J. Eng. Res. Technol.*, 2021, [Online]. Available: www.ijert.org.

[16] M. R. Karim *et al.*, DeepHateExplainer: Explainable Hate Speech Detection in Under-resourced Bengali Language**,** vol. 1, no. 1. *Association for Computing Machinery*, 2021. doi: 10.1109/DSAA53316.2021.9564230.

[17] A. M. Ishmam and . harmin, “Hateful speech detection in public facebook pages for the bengali language,” *Proc. - 18th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2019*, no. December, pp. 555–560, 2019, doi: 10.1109/ICMLA.2019.00104.

[18] M. F. Mridha, M. A. H. Wadud, M. A. Hamid, M. M. Monowar, M. Abdullah-Al-Wadud, and A. Alamri, “L-Boost: Identifying Offensive Texts from Social Media Post in Bengali,” *IEEE Access*, vol. 9, pp. 164681–164699, 2021, doi: 10.1109/ACCESS.2021.3134154.

[19] N. omim, M. Ahmed, H. Talukder, and M. aiful Islam, “Hate peech Detection in the Bengali Language A Dataset and Its Baseline Evaluation,” no. March 2022, pp. 457–468, 2021, doi: 10.1007/978-981-16-0586-4_37.

[20] K. Kumari, J. P. ingh, Y. K. Dwivedi, and N. P. ana, “Towards Cyberbullying-free social media in smart cities: a unified multi-modal approach,” *Soft Comput.*, vol. 24, no. 15, pp. 11059–11070, Aug. 2020, doi: 10.1007/s00500-01904550-x.

[21] K. Kumari and J. P. ingh, “Identification of cyberbullying on multi-modal social media posts using genetic algorithm,” *Trans. Emerg. Telecommun. Technol.*, vol. 32, no. 2, Feb. 2021, doi: 10.1002/ett.3907.

[22] A. Kumar and N. achdeva, “Multimodal cyberbullying detection using capsule network with dynamic routing and deep convolutional neural network,” 2021. doi *10.1007/s00530-020-00747-5.*

[23] M. . Karim, . K. Dey, T. Islam, and B. . Chakravarthi, “Multimodal Hate Speech Detection from Bengali Memes and Texts,” Apr. 2022, [Online]. Available *http://arxiv.org/abs/2204.10196.*

[24] K. melyakov, A. Chupryna, D. Darahan, and . Midina, “Effectiveness of modern text recognition solutions and tools for common data sources,” *CEUR Workshop Proc.*, vol. 2870, pp. 154–165, 2021.

[25] M. Ebersbach, . Herms, and M. Eibl, “Fusion methods for ICD10 code classification of death certificates in multilingual corpora,” *CEUR Workshop Proc.*, vol. 1866, no. September, 2017.