

Multimodal Cyberbullying Detection using Hybrid Deep Learning Algorithms

Dr. Vijayakumar V^{*1}, Dr Hari Prasad D², Adolf P³

¹ Professor & COE, Department of Computer Science, Sri Ramakrishna College of Arts and Science,
Coimbatore-641 006, Tamilnadu, India.

² Professor and Head, Department of Computer Applications, Sri Ramakrishna College of Arts and Science,
Coimbatore-641 006, Tamilnadu, India.

³ Research Assistant, Sri Ramakrishna College of Arts and Science, Coimbatore-641 006, Tamilnadu, India.

*Corresponding Author

Abstract

The usage and user of internet and social media is increasing day-by-day and consequently cyberbully vulnerabilities are also growing. Cyberbullying is an aggressive, planned behavior carried out by a group or individual. It is happening by sending, posting, sharing negative, harmful, untrue contents in online. It leads to psychiatric and emotional disorders for those affected. Hence, there is a critical requirement to develop automated methods for cyberbullying detection and prevention. Over the past few years, most existing work on cyberbullying detection has focused on text based analysis. Text and image are the important mediums in cyberbullying incident. This paper presents a hybrid deep neural model for cyberbullying detection in two different modalities of social data, namely text and image. Deep learning methods have achieved state-of-the-art results in various applications. In this paper, hybrid deep learning technique is used in multiple modalities of data to detect cyberbullying. An experiment was conducted on hybrid model (CNN and LSTM) which detects cyberbullying on text and image. The experiments are conducted on publicly available datasets and tested with telegram chat. This paper aims to direct future research on integrating video source with existing multimodal data source to prevent cyberbullying issues.

Keywords: Cyberbullying Detection, Multimodal Data, Hybrid Deep Learning Model, Image Based Bullying Detection, Text based Bully Detection.

1. INTRODUCTION

The rise of internet and social networks in the digital world has led to new definitions of social communications. It is observed that 58.11% of world populations are using the social media [38] like Telegram, Facebook, WhatsApp, Instagram, Snapchat etc. for communication and higher cyberbullying issues reported in the recent years [9][39]. Cyberbullying is a destructive, intentional act that is carried out by a group or individual. It also is growing from time to time. The effects caused by cyberbullying includes isolation, anger, mental effects of depression and anxiety, academic issues, suicidal thoughts and self-harm; behavioural effects of using drugs or alcohol, skipping regular activities [41].

The occurrence of cyberbullying can vary from multi-mode input sources such as text, images and videos. Initially, Cyberbullying was unorganized and the usage medium was

mostly text and utmost of research studies have mainly focused on analyzing of textual content, such as comments and text messages. The cyberbullying detection is still challenge due to the usage emoji, memes, text and characters in image in the communication. Now, bully events changed to organized and multi-medium data source. Social media networking websites are exclusively emphasis on photo-sharing. These trends make a shift from text to image to commit cyberbullying behaviors among victims.

The research on multimodal cyberbullying detection is more limited. The majority of research is to identify the bully event in single data source either text or images and motivation behind bullying is looked at the problem from socio-psychological and educational perspectives. There is an emergency need for detection and prevention of cyberbullying to provide the safe platform for digital users. In today's communication, most of the information sharing contents are multimodal such as text, images or videos, and also difficult to identify and detect the cyberbullying content from these multimodal data source. Technologies could significantly contribute to the development of these detection and preventative strategies. Most of the existing studies have used conventional Machine Learning (ML) models to detect cyberbullying incidents. Now, deep neural based models have shown substantial enhancement over traditional models like linear regression in detecting cyberbullying. A hybrid deep learning approach is proposed in this paper which can detect bullying image and text at the same time. T. Mahlangu et al.[11], A. Singh et al.[13] and H. Vandeboosch et al.[15] conducted notable studies in the field of detection and prevention in cyberbullying cases. Lee Jia et al.[34] developed a mobile application in order to prevent cyberbullying.

In this paper, the detailed background review of literature presented in the section 2. The proposed hybrid multimodal detection approach is discussed in the section 3. The experimental results and discussion presented in the section 4. The paper is concluded with future work in the section 5.

2. REVIEW OF LITERATURE

In the recent years many works were proposed related to cyberbullying detection. Our research focuses on both image and text based cyberbullying detection because only one medium can't detect the motive of the victim meritoriously in social media. The literature review presented based on the dataset used, nature of detection methods and techniques used.

Most of the authors are using the publically available datasets such as Formspring.Me, ask.fm, twitter, YouTube and Kaggle for their research. Ç. ACI et al. used multiple datasets [20]. L. Cheng et al., M. Yao et al. and Tanmayee et al. used Instagram datasets [8][12][16][17][18]. V. Banerjee et al. and H. Dani et al. detected the cyberbullying text in Twitter dataset using deep neural network techniques [7][32]. M. A. Al-Ajlan et al. [24] and N. Tahmasbi et al. [6][26] used Twitter dataset for detecting cyberbullying text. A. Bakshi et al. [19] and C. Emmerly et al. [10] used YouTube dataset, Ask.fm and crowdsourced dataset for cyberbullying detection. H. Rosa et al., conducted an in-depth analysis of different studies on automatic cyberbullying detection using formspring data [14]. A. Gangwar et al. used the NSFW dataset to detect abuse content [36]. Ibn Rafiq et al.[35] studied cyberbully detection using Vine dataset.

N. Lu et al. used Chinese weibo dataset, English tweet dataset and proposed a Char- CNNS (Character- level Convolutional Neural Network with Shortcuts) model to identify whether the text in social media contains cyberbullying based on multi lingual detection [2]. C. Van Hee et al. [21] presented automatic cyberbullying detection in social media by modelling posts written by bullies, victims, and bystanders of online bullying in English and Dutch dataset. Another major source of cyberbullying data is gaming platforms. V.

Balakrishnan et al., M. Garaigordobil et al and S. Murnion used various gaming platform data to detect cyberbullying [5][29][30].

There are few prominent works completed in cyberbully detection using machine learning and deep learning techniques. D. Mario et al.[31], P. Tata et al.[27], Talpur BA[1]., A. J. Sánchez-Medina et al.[3], V. Balakrishnan et al.[5], A. Bakshi et al., [19], S. Murnion et al. [[29] applied machine learning techniques to detect the cyberbullying. S. V. Georgakopoulos et al. used the deep neural network methods for cyberbullying detection [22]. M. A. Al-Ajlan et al. proposed a CNN-CB algorithm to exclude the need for features and yield better prediction [24]. H. Rosa et al., implemented a simple CNN, a hybrid CNN-LSTM and a mixed CNN-LSTM-DNN [14] [33] and trained via the word2vec model with Google-News, Twitter and Formspring data set. Q. H. Nguyen et al. CNN used NSFW data set for detection of pornographic content [37]. A multi modal approach is proposed by T. Pradheep et al. in cyberbully detection [23] and meme based cyberbully detection os proposed by G Campisi et al.[28]. Vijayakumar V and Hari Prasad D presented a deep learning based LSTM algorithm to detect and prevent the cyberbullying incident interfaced with chatbot [40]. The detailed summary of literature presented in the Table.1.

Table 1. Summary of Review Literature

Authors	Year	Method / Technology / Algorithm	Outcome	Challenges / Issues / Demerit / Limitation
Q. H. Nguyen, K. N. K. Nguyen, H. L. Tran, T. T. Nguyen, D. D. Phan, and D. L. Vu	2020	Mask R-CNN	Detection in Images and video	Texts and writings in image is not detected
V. Banerjee, J. Telavane, P. Gaikwad, and P. Vartak	2019	CNN	Cyberbullying detection using DNN	Single data source
L. Cheng, R. Guo, Y. Silva, D. Hall, and H. Liu	2019	Bidirectional GRU-RNN	Multi Model Cyberbullying detection with time series	Single data source
L. Cheng, J. Li, Y. N. Silva, D. L. Hall, and H. Liu	2019	Random Forest, Linear SVM, Logistic Regression	Multi Model Cyberbullying detection with time series	Less amount of data is used
L. Cheng, J. Li, Y. Silva, D. Hall, and H. Liu	2019	k*-NN	Cyberbullying Detection	Dataset is based on few keywords only
M. Yao, C. Chelmiss, and D. S. Zois	2019	CONcISE-framework	Cyberbullying detection on Instagram	Single data source
N. Tahmasbi and A. Fuchsberger	2019	NLP, Machine learning	Cyberbullying detection for helping parents	Targeted detection only
Tanmayee Patange, Jigyasa Singh, Aishwarya Thorve, Yadnyashree Somaraj Madhura Vyawahare	2019	CNN, Word2Vec, Offensiveness	Multi input cyberbullying detection	Instagram only
S. Murnion, W. J. Buchanan, A. Smales, and G. Russell	2018	Machine Learning	Sentiment analysis	Based on game data
H. Rosa, D. Matos, R. Ribeiro, L. Coheur, and J. P. Carvalho	2018	simple CNN, a hybrid CNN-LSTM and a mixed CNN-LSTM-DNN	Cyberbullying Detection	Time consumption in model building
P. Tata, S. P. Devaneyan, and J. I. Sheeba	2018	Linear SVM	detection of cyberbully	Need of new algorithms in detection

3. PROPOSED HYBRID METHOD OF CYBERBULLYING DETECTION:

The Hybrid deep learning model is developed for multimodal data based cyberbullying detection and is shown the Fig.1. The experiment data is selected from benchmark datasets and applied preprocessed techniques for improving prediction accuracy. Keras Functional API is used to build the multi-input deep neural network model. The created model is also tested with a real time input data and predictions are discussed.

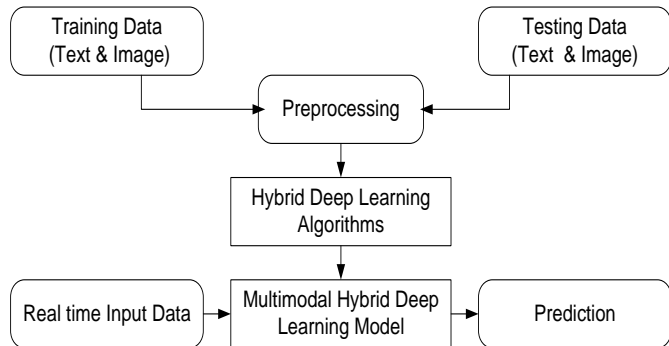


Figure 1. Proposed Hybrid Deep learning based Multi-Model Cyberbullying Detection

The detection method is discussed in the following steps:

Step 1: Data Collection and preparation:

The NSFW scrapped image dataset collected from the GitHub repository named “nswf_data_scraper” [42]. The repository contains URLs of the image scrapped from different source of internet. Text data is collected from Kaggle Toxic Comment Classification Challenge dataset. This dataset is created from comments of Wikipedia talk pages by Conversation AI team, a research initiative founded by Jigsaw and Google.

Step 2: Preprocessing:

The collected data is preprocessed to avoid possibility of wrongly tagged data. In text inputs are tokenized. Then it constructs a sequence. The image data is converted into fixed size of 90 X 90. Every image has same length and width. An image is represented by 0 – 255 values in the array. It makes smaller images by dividing it by 255.

Step 3: Build & Train the Hybrid Deep Learning Model

The proposed hybrid deep learning models which take image and text inputs to detect cyberbullying in the data. A hybrid algorithm is constructed with Convolutional Neural Network (CNN) for image and Long short-term Memory (LSTM) for text. These models are combined together using Keras Functional API. Convolutional Neural Network (CNN) is used to image based analysis which contains multiple layers of perceptrons. Each neuron in one layer is connected with all

neurons in the next layer. This network always does the convolution mathematical operation. The architecture consist an input layer, an output layer and in between there will be multiple hidden layers. After convolution operation an RELU activation function applied then followed by next convolutions like pooling layers.

LSTM networks are similar to RNNs with one major difference that hidden layer updates are replaced by memory cells. It consists of cells, information flow and gates. LSTM achieved by four gates. The forget gate recalls the necessary parts from the previous states based on the current event and forgets unnecessary portions. The learn gate learns from the current input and the recent events. The remember gate takes the input from the forget gate to get and learn the gate and remembers everything required for future reference of the model. The use gate maps the output from the forget gate and learns gate, considers all the factors, and provides the prediction of the current input. LSTM is used to avoid over fitting, dropping probability. The last layer is fully connected. CNN and LSTM layers are concatenated and build the model using Keras functional API. The Rectified Linear Unit (ReLU) activation function is used in models. The function returns 0 if it receives any negative input, but for any positive value x , it returns that value back. So, it can be written as $f(x)=\max(0, x)$. The Keras system can pull any layer from anywhere. This property helps to build multi input and multi output models. The functionality is used to combine LSTM and CNN.

Step 4: Training, Deployment and Testing the Model

The entire data of image and text is divided into training and testing as in 80:20 ratios. The dataset verified manually to fine tune the repeated data. The deep neural network is trained and the deployment is done using python flask server.

The deep learning model is loaded in the server. For testing process, the data is send to server from the telegram user account and cyberbullying tested the prediction accuracy. The results send back to client text and image separately tagged. The results are displayed to client as labels. The cyberbullying events are detected with real time sample data given.

4. EXPERIMENTAL RESULTS & DISCUSSION

The proposed model developed and tested using Intel i5 processor system with 8 GB RAM and 1 MBPS Internet speed. The system has in-built GPU Intel UHD Graphics, installed Windows 10 Operating system and configured the Python libraries like cv2, numpy, Tensor flow and Keras.

The model is trained with 4,590,756 parameters with separate input and output for image and text. For the text, the inputs are sent through LSTM layers and image is through CNN layers. Then, few dense layers are added on both sides. The layers are concatenated. And the results are separated from the concatenated layer and two outputs are generated based on the

image classes and text classes. For image, randomly choose 12,200 images from the dataset and trained. For functional API both the input cardinality must be equal. So, same

amount of text comments chosen from the kaggle dataset. The trained model is shown in the Fig 2.

```

Model: "model_1"
-----
Layer (type)                Output Shape          Param #    Connected to
-----
text_inp (InputLayer)       [(None, 200)]        0
embedding_1 (Embedding)     (None, 200, 128)    2560000    text_inp[0][0]
img_inp (InputLayer)        [(None, 90, 90, 3)] 0
lstm_layer (LSTM)           (None, 200, 60)     45360      embedding_1[0][0]
conv2d_2 (Conv2D)           (None, 88, 88, 32)  896        img_inp[0][0]
bidirectional_1 (Bidirectional) (None, 200, 120) 58080      lstm_layer[0][0]
max_pooling2d_2 (MaxPooling2D) (None, 44, 44, 32) 0          conv2d_2[0][0]
global_max_pooling1d_1 (GlobalM (None, 120)        0          bidirectional_1[0][0]
conv2d_3 (Conv2D)           (None, 42, 42, 16)  4624       max_pooling2d_2[0][0]
dropout_4 (Dropout)         (None, 120)         0          global_max_pooling1d_1[0][0]
max_pooling2d_3 (MaxPooling2D) (None, 21, 21, 16) 0          conv2d_3[0][0]
dense_7 (Dense)             (None, 128)         15488      dropout_4[0][0]
flatten_1 (Flatten)         (None, 7056)        0          max_pooling2d_3[0][0]
dropout_5 (Dropout)         (None, 128)         0          dense_7[0][0]
dense_5 (Dense)             (None, 256)         1806592    flatten_1[0][0]
dense_8 (Dense)             (None, 256)         33024      dropout_5[0][0]
dense_6 (Dense)             (None, 84)          21588      dense_5[0][0]
dropout_6 (Dropout)         (None, 256)         0          dense_8[0][0]
concatenate_1 (Concatenate) (None, 340)         0          dense_6[0][0]
                                dropout_6[0][0]
dense_9 (Dense)             (None, 128)         43648      concatenate_1[0][0]
dropout_7 (Dropout)         (None, 128)         0          dense_9[0][0]
img_out (Dense)             (None, 2)           682        concatenate_1[0][0]
text_out (Dense)            (None, 6)           774        dropout_7[0][0]
-----
Total params: 4,590,756
Trainable params: 4,590,756
Non-trainable params: 0
    
```

Figure 2: Multi-Input Deep Learning Model

The performance of the proposed framework was measured in terms of the quality measures, namely Precision, Recall and F1 score. Precision is the ratio between the correct predictions and the total predictions. Recall is the ratio of the correct predictions and the total number of correct items in the set. F1 score is the weighted harmonic mean of precision and recall. Accuracy calculates the proportion of correctly identified cyberbully words. The text model attains 85% accuracy on the test data which is shown in the Fig 3. The image model attains 86% accuracy on the image data which is shown in the Fig 4.

Accuracy: 0.85

Classification Report - Text

	precision	recall	f1-score	support
tox	0.95	0.91	0.93	1968
non_tox	0.75	0.61	0.67	472
micro avg	0.91	0.85	0.88	2440
macro avg	0.85	0.76	0.80	2440
weighted avg	0.91	0.85	0.88	2440
samples avg	0.85	0.85	0.85	2440

Figure 3: Accuracy and Classification Report – Text

Accuracy: 0.86

Classification Report - Image

	precision	recall	f1-score	support
good_new1	0.84	0.88	0.86	1245
bad_new1	0.87	0.83	0.85	1195
accuracy			0.86	2440
macro avg	0.86	0.85	0.86	2440
weighted avg	0.86	0.86	0.86	2440

Figure 4: Accuracy and Classification Report – Image

The model tested with sample data in the form of image and text for cyberbullying events detection in the hybrid deep learning model. The proposed system possibly predicts four types bully events which combine the text and image. So the model is tested with sample inputs as presented in the Fig.5.

		Text	
		Non Toxic Text	Toxic Text
Image	Non Toxic \ Toxic		
	Non Toxic Image	Non Toxic Text Non Toxic Image	Toxic Text Non Toxic Image
	Toxic Image	Non Toxic Text Toxic Image	Toxic Text Toxic Image

Figure 5: Possible Cyber Bullying Events

In Fig.6 shows the output detection of non-toxic in image and text input. Toxic text and non-toxic image input event prediction is presented in Fig.7, and Fig. 8 shows non-toxic text and toxic image input. Fig. 9 shows the output toxic in both image and text inputs.

Multi Input Cyberbully Detection

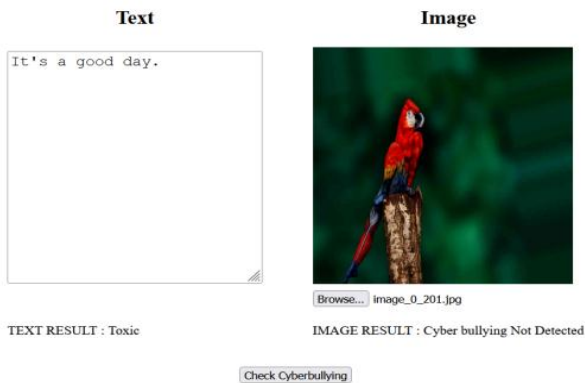


Figure 6 : Non Toxic Text and Non Toxic Image Data Test

Multi Input Cyberbully Detection

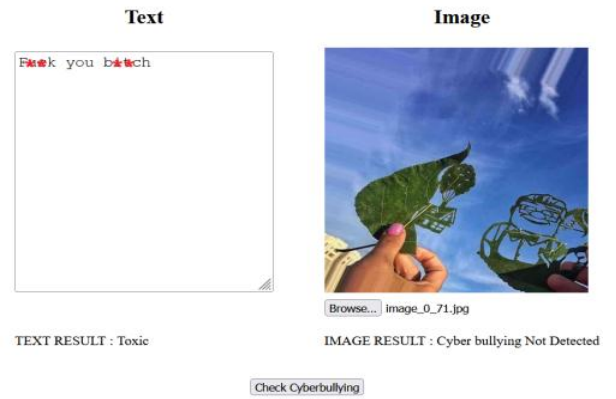


Figure 7: Toxic Text and Non Toxic Image Data Test

Multi Input Cyberbully Detection



Figure 8: Non Toxic Text and Toxic Image

Multi Input Cyberbully Detection



Figure 9: Toxic Text and Toxic Image

5. CONCLUSION AND FUTURE WORK

Due to increase usage of text, image and video based communication in social media application, cyberbullying events are growing. These events need to be detected and prevented before create harm of users. In this paper, a hybrid deep neural network model was developed which detects cyberbullying in multimodal data such as image and text. The data extracted from GitHub and Kaggle and tested with telegram real time data. The model developed with a combination of Convolutional Neural Network (CNN) and Long Short-term Memory (LSTM) using Kears Functional API. This model predicted more accurately and the image based prediction gives 86% accuracy and text based prediction gives 85% accuracy after training the model. An effective user interface system will be developed for preventing the cyberbullying event. In future, video and audio can be integrated cyberbullying detection. Images Text also is considered for detection from cyberbullying. Multi language, cross language and mix language option can be integrated.

ACKNOWLEDGMENT

This research work was funded by the ICSSR under IMPRESS Scheme. The author also thanks the ICSSR and Sri Ramakrishna College of Arts and Science who provided the support for doing the research.

REFERENCES:

- [1] Talpur BA, O'Sullivan D (2020) Cyberbullying severity detection: A machine learning approach. *PLoS ONE* 15(10): e0240924. <https://doi.org/10.1371/journal.pone.0240924>
- [2] N. Lu, G. Wu, Z. Zhang, Y. Zheng, Y. Ren, and K. K. R. Choo, "Cyberbullying detection in social media text based on character-level convolutional neural network with shortcuts," *Concurr. Comput.*, no. July 2019, pp. 1–11, 2020, doi: 10.1002/cpe.5627.
- [3] A. J. Sánchez-Medina, I. Galván-Sánchez, and M. Fernández-Monroy, "Applying artificial intelligence to explore sexual cyberbullying behaviour," *Heliyon*, vol. 6, no. 1, 2020, doi: 10.1016/j.heliyon.2020.e03218.
- [4] S. Salawu, Y. He, and J. Lumsden, "Approaches to Automated Detection of Cyberbullying: A Survey," *IEEE Trans. Affect. Comput.*, vol. 11, no. 1, pp. 3–24, 2020, doi: 10.1109/TAFFC.2017.2761757.
- [5] V. Balakrishnan, S. Khan, T. Fernandez, and H. R. Arabnia, "Cyberbullying detection on twitter using Big Five and Dark Triad features," *Pers. Individ. Dif.*, vol. 141, no. January, pp. 252–257, 2019, doi: 10.1016/j.paid.2019.01.024.
- [6] N. Tahmasbi and A. Fuchsberger, "ChatterShield – A multi-platform cyberbullying detection system for parents," *25th Am. Conf. Inf. Syst. AMCIS 2019*, pp. 1–5, 2019.
- [7] V. Banerjee, J. Telavane, P. Gaikwad, and P. Vartak, "Detection of Cyberbullying Using Deep Neural Network," *2019 5th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2019*, pp. 604–607, 2019, doi: 10.1109/ICACCS.2019.8728378.
- [8] L. Cheng, R. Guo, Y. Silva, D. Hall, and H. Liu, "Hierarchical attention networks for cyberbullying detection on the instagram social network," *SIAM Int. Conf. Data Mining, SDM 2019*, pp. 235–243, 2019, doi: 10.1137/1.9781611975673.27.
- [9] R. Sprugnoli, S. Menini, S. Tonelli, F. Oncini, and E. Piras, "Creating a WhatsApp Dataset to Study Pre-teen Cyberbullying," pp. 51–59, 2019, doi: 10.18653/v1/w18-5107.
- [10] Emmery, C., Verhoeven, B., De Pauw, G. et al. Current limitations in cyberbullying detection: On evaluation criteria, reproducibility, and data scarcity. *Lang Resources & Evaluation* (2020). <https://doi.org/10.1007/s10579-020-09509-1>
- [11] T. Mahlangu, C. Tu, and P. Owolawi, "A review of automated detection methods for cyberbullying," *2018 Int. Conf. Intell. Innov. Comput. Appl. ICONIC 2018*, pp. 1–5, 2019, doi: 10.1109/ICONIC.2018.8601278.
- [12] L. Cheng, J. Li, Y. Silva, D. Hall, and H. Liu, "Pi-Bully: Personalized cyberbullying detection with peer influence," *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2019-Augus, pp. 5829–5835, 2019, doi: 10.24963/ijcai.2019/808.
- [13] A. Singh and M. Kaur, "Content-based cybercrime detection: A concise review," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 8, pp. 1193–1207, 2019.
- [14] H. Rosa et al., "Automatic cyberbullying detection: A systematic review," *Comput. Human Behav.*, vol. 93, no. October 2018, pp. 333–345, 2019, doi: 10.1016/j.chb.2018.12.021.
- [15] H. Vandebosch, *Cyberbullying Prevention, Detection and Intervention*. Springer International Publishing, 2019.
- [16] M. Yao, C. Chelmiss, and D. S. Zois, "Cyberbullying ends here: Towards robust detection of cyberbullying in social media," *Web Conf. 2019 - Proc. World Wide Web Conf. WWW 2019*, pp. 3427–3433, 2019, doi: 10.1145/3308558.3313462.
- [17] Y. S. M. V. Tanmayee Patange, Jigyasa Singh, Aishwarya Thorve, "DETECTION OF CYBERHECTORING ON INSTAGRAM," pp. 5–8, 2019.
- [18] L. Cheng, J. Li, Y. N. Silva, D. L. Hall, and H. Liu, "XBully: Cyberbullying Detection within a Multi-Modal Context," pp. 339–347, 2019, doi: 10.1145/3289600.3291037.
- [19] A. Bakshi and A. K. Patel, *Empirical Analysis of Supervised Machine Learning Techniques for*

Cyberbullying Detection, vol. 2, no. January. Springer Singapore, 2019.

- [20] Ç. ACI, E. ÇÜRÜK, and E. S. EŞSİZ, "Automatic Detection of Cyberbullying in Formspring.Me, Myspace and Youtube Social Networks," *Turkish J. Eng.*, vol. 3, no. 4, pp. 168–178, 2019, doi: 10.31127/tuje.554417.
- [21] C. Van Hee *et al.*, "Automatic detection of cyberbullying in social media text," *PLoS One*, vol. 13, no. 10, pp. 1–22, 2018, doi: 10.1371/journal.pone.0203794.
- [22] S. V. Georgakopoulos, A. G. Vrahatis, S. K. Tasoulis, and V. P. Plagianakos, "Convolutional neural networks for toxic comment classification," *ACM Int. Conf. Proceeding Ser.*, no. April, 2018, doi: 10.1145/3200947.3208069.
- [23] T. Pradheep, J. . Sheeba, T. Yogeshwaran, and S. Pradeep Devaneyan, "Automatic Multi Model Cyber Bullying Detection from Social Networks," *SSRN Electron. J.*, pp. 248–254, 2018, doi: 10.2139/ssrn.3123710.
- [24] M. A. Al-Ajlan and M. Ykhlef, "Deep learning algorithm for cyberbullying detection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 9, pp. 199–205, 2018, doi: 10.14569/ijacsa.2018.090927.
- [25] M. Dadvar and K. Eckert, "Cyberbullying Detection in Social Networks Using Deep Learning Based Models; A Reproducibility Study," pp. 1–13, 2018, [Online]. Available: <http://arxiv.org/abs/1812.08046>.
- [26] N. Tahmasbi and E. Rastegari, "A Socio-contextual Approach in Automated Detection of Cyberbullying," *Proc. 51st Hawaii Int. Conf. Syst. Sci.*, vol. 9, pp. 2151–2160, 2018, doi: 10.24251/hicss.2018.269.
- [27] P. Tata, S. P. Devaneyan, and J. I. Sheeba, "Improved cyberbully detection techniques using multiple correlation coefficient from forum corpus," *Int. J. Auton. Comput.*, vol. 3, no. 2, p. 152, 2018, doi: 10.1504/ijac.2018.10018810.
- [28] G. Campisi, S. Corchs, E. Fersini, F. Gasparini, and M. Mantovani, "Automatic detection of sexist content in memes," p. 3247, 2018.
- [29] S. Murnion, W. J. Buchanan, A. Smales, and G. Russell, "Machine learning and semantic analysis of in-game chat for cyberbullying," *Comput. Secur.*, vol. 76, pp. 197–213, 2018, doi: 10.1016/j.cose.2018.02.016.
- [30] M. Garaigordobil and V. Martínez-Valderrey, "Technological resources to prevent cyberbullying during adolescence: The Cyberprogram 2.0 program and the cooperative Cybereduca 2.0 Videogame," *Front. Psychol.*, vol. 9, no. MAY, pp. 1–12, 2018, doi: 10.3389/fpsyg.2018.00745.
- [31] D. Mario and P. G. P. Madhuri, "A Study to Learn Robust and Discriminative Representation to Tackle Cyber bullying Detection www.ijsear.com * corresponding Author," vol. 5, no. 6, pp. 643–646, 2017.
- [32] H. Dani, J. Li, and H. Liu, "Sentiment Informed Cyberbullying Detection in Social Media," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10534 LNAI, pp. 52–67, 2017, doi: 10.1007/978-3-319-71249-9_4.
- [33] H. Rosa, D. Matos, R. Ribeiro, L. Coheur and J. P. Carvalho, "A "Deeper" Look at Detecting Cyberbullying in Social Networks," 2018 International Joint Conference on Neural Networks (IJCNN), 2018, pp. 1-8, doi: 10.1109/IJCNN.2018.8489211.
- [34] Lee Jia Thun, Phoey Lee Teh, Chi-Bin Cheng, CyberAid: Are your children safe from cyberbullying?, *Journal of King Saud University - Computer and Information Sciences*, 2021, <https://doi.org/10.1016/j.jksuci.2021.03.001>. (<https://www.sciencedirect.com/science/article/pii/S1319157821000537>)
- [35] Ibn Rafiq, R.; Hosseinmardi, H.; Han, R.; Lv, Q.; Mishra, S.; Mattson, S.A. Careful What You Share in Six Seconds.ss2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015—ASONAM '15; Association for Computing Machinery (ACM): New York, NY, USA, 2015; pp. 617–622.
- [36] A. Gangwar, E. Fidalgo, E. Alegre, and V. González-Castro, "Pornography and child sexual abuse detection in image and video: A comparative evaluation," *IET Semin. Dig.*, vol. 2017, no. 5, pp. 37–42, 2017, doi: 10.1049/ic.2017.0046.
- [37] Q. H. Nguyen, K. N. K. Nguyen, H. L. Tran, T. T. Nguyen, D. D. Phan, and D. L. Vu, "Multi-level detector for pornographic content using CNN models," *Proc. - 2020 RIVF Int. Conf. Comput. Commun. Technol. RIVF 2020*, 2020, doi: 10.1109/RIVF48685.2020.9140734.
- [38]. <https://news.uga.edu/social-media-addiction-linked-to-cyberbullying/>
- [39]. Ojasvi Jain, Muskan Gupta, Sidh Satam, Siba Panda, "Has the COVID-19 pandemic affected the susceptibility to cyberbullying in India?," *Computers in Human Behavior Reports*, Volume 2, 2020, <https://doi.org/10.1016/j.chbr.2020.100029>.
- [40]. Vijayakumar V and Hari Prasad D, "Intelligent Chatbot Development for Text based Cyberbullying Prevention," *International Journal of New Innovations in Engineering and Technology*, Volume 17, Issue 1, pp. 73-81, June 2021.
- [41]. <https://timesofindia.indiatimes.com/readersblog/riyable/effect-of-cyberbullying-5884/>
- [42]. https://github.com/alex000kim/nsfw_data_scraper