iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.49984

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Detection and Classification of Cyberbullying Using CR*

Dr. M.S. Anbarasi[1], Ms. Gayathri R[2], Ms. Lydia Beryl D[3], Mr. Gowtham M[4], Mr. Naveen Kumar N[5]
[1]*Assistant Professor,* [2, 3, 4,5]*B.Tech Student, , Information Technology Puducherry TechnologicalUniversity Puducherry, India*

Abstract: ***Cyberbullying is one of the latest threats in the online world, affecting millions of people worldwide. The detection of cyberbullying became a challenging task due to the complexity involved. In this paper, we propose a novel approach for cyberbullying detection using CR* which concatenates the deep learning model's features such as Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) which helps in detecting text, audio, and emoji. Initially, we collect a dataset of cyberbullying messages from social media platforms. Then we integrate convolutional neural networks and recurrent neural networks to build our deep learning model entitled Convolutional Recurrent (CR*) to extract features from a combination of the text data, emoji, and audio. The multimodal features were then concatenated and passed through fully connected layers for classification. The proposed approach can be useful for detecting cyberbullying on various online platforms and can help prevent the spread of cyberbullying.***
Index Terms: CR*, CNN, RNN, speech recognition, cyberbullying, hybrid model, text,audio, emoji

I. INTRODUCTION

Cyberbullying is a pervasive and serious problem in today's society, particularly in online social media platforms. Cyberbullying can havesevere emotional and psychological consequences for victims, including depression, anxiety, and even suicide. Therefore, the development of effective cyberbullying detection systems is essential to prevent and mitigate the harmful effects of cyberbullying. Many researchers have recently developed various machine learning and deep learning models to detect cyberbullying. However, earlier systems have limitations and issues such as limited modality support, limited dataset size, lack of multimodal fusion, and limited audio and emoji representation. In this paper, we propose a solution for the detection of cyberbullying using text, audio, and emoji data. Our proposed system utilizes CNN and RNN model features to build our hybrid deep-learning model CR*.

This paper aims to discuss the limitations and issues of earlier cyberbullying detection systems, present our proposed system, and evaluate its performance using different metrics such as accuracy rate, precision, recall, and F1 score. The results of our proposed system will contribute to developing more effective and comprehensive cyberbullying detection systems.

II. LITERATURE SURVEY

Lee Jia Thun, Phoey Lee Teh, and Chi-Bin Cheng researched using machine learning to identify language that contains cyberbullying. The suggested method is trained to utilize features like variation in text seen in social media context and social network environment interactions. The machine may also determine through users' genderor whether they have used hate speech. To close the gaps and overcome the restrictions of current apps, this study suggests a mechanism that combines the best cyberbullying detection features. According to the study's findings, the suggested mobile application is more accurate than other ones at spotting cyberbullies.

Belal Abdullah Hezam Murshed., Jemal Abaway., et al., have proposed a hybrid deep learning model, called DEA-RNN, to detect Cyberbullying on Twitter social media networks. The Elman-type Recurrent Neural Networks (RNN) and Dolphin Echolocation Algorithm (DEA) are merged into the proposed DEA-RNN approach to enhance the Elman RNN's parameters and minimize training time. Using a dataset of10,000 tweets, they thoroughly evaluated DEA-RNN and compared its performance withthat of cutting-edge algorithms like Bi-directional long short-term memory (Bi-LSTM), RNN, SVM,Multinomial Naive Bayes (MNB), and RandomForests (RF). According to the experimentalfindings, DEA-RNN was preferred in every situation.

It performed much better than theevaluated existing methods while detecting CB on the Twitter network. In scenario 3, DEA-RNN wasmore effective and achieved an average of 90.45%accuracy, 89.52% precision, 88.98% recall, 89.25% F1-score, and 90.94% specificity.

Mithushi Raj., Kanishka Solanki., et al., have proposed a cyberbullying detection system to detect cyberbullying in text format. In this study, we put forth a deep learning system that will analyze Twitter tweets sent in real time and find any instances of cyberbullying.

Recent researchhas demonstrated that deep neural network-based methods are superior to traditional ones foridentifying texts that contain cyberbullying. Furthermore, our software can detect cyberbullying posts written in English, Hindi, and Hinglish (Multilingual data).

Xiao-Zhi Gao, Tapan Kumar Das, et al. have proposed Deep Convolutional NeuralNetwork(DCNN)-based automated approach to identify cyberbullying.

In addition to employing Convolutional neural network, the proposed DCNN model uses the Twitter text and GloVe embedding vector to extract the semantics from the messages. It outperformed the current models and attained the best precision, recall, and F1-score values of 0.93, 0.88, and 0.92.

III. LIMITATIONS IN THE EXISTING

A. *System*

1) *Limited Modality Support:* Some earlier systems only supported text-based cyberbullying detection, which means theywere not able to detect cyberbullying in audio or emoji data.
2) *Limited Dataset Size:* Many earlier systems were trained on small datasets, which limited their ability to generalize to new data and detect rare cyberbullying events.
3) *Lack of Multimodal Fusion:* Some earliersystems did not incorporate multimodal fusion techniques, which means they did not fully leverage the complementary nature of text, audio, and emoji data.
4) *Limited Audio Data Availability:* Cyberbullying can occur in audio formats, such as in voice messages or calls. However, cyberbullying detection foraudio formats can be challenging, which limits the effectiveness of models that rely on audio data.
5) *Limited Emoji Representation:* Emojis are a popular form of communication on social media, and can be used to convey emotions or attitudes in text messages. However, some earlier systems did not fully consider the meaning and context of emojis, which led to incorrect or biasedclassifications. Additionally, some languages have different interpretations of certain emojis, which further complicates the detection of cyberbullying in emoji data.

IV. PROPOSED WORK

Our proposed system for cyberbullying detection with text, audio, and emoji utilizes CNN and RNN model features to build our deep learning model CR* that leverages multimodal fusion techniques to improve detection accuracy.

A. *Input*

The first step will be collecting a dataset of cyberbullying messages from various social media platforms which are labelled as cyberbullying or non-cyberbullying.

B. *Process*

Initial step would be preprocessing, which involves cleaning data, and removing punctuations, stop words, single characters, and multiple spaces. If the input is in the form of audio, first it gets converted to text format. The input text is then converted into a sequence ofintegers using a tokenizer.

Then, an embeddinglayer is used to convert each integer into a dense vector of fixed size. Next, a convolutional layer is applied to capture local patterns in the text.

Apooling layer follows this to reduce thedimensionality of the output. Then, a recurrentlayer is used to capture the global context of the text. Finally, the concatenated features are passed through fully connected layers for classification. The output is a binary classification, indicatingwhether the message contains cyberbullying content or not.

C. *Output*

The hybrid model detects and classifies them as cyberbullying and not cyberbullying. The performance of the model will be evaluated usingstandard metrics such as accuracy rate, precision, recall, and F1 score. Cross-validation also hasbeen done to ensure that the model generalizes well to new data.

V. DESIGN DIAGRAM OF THE PROPOSED SYSTEM

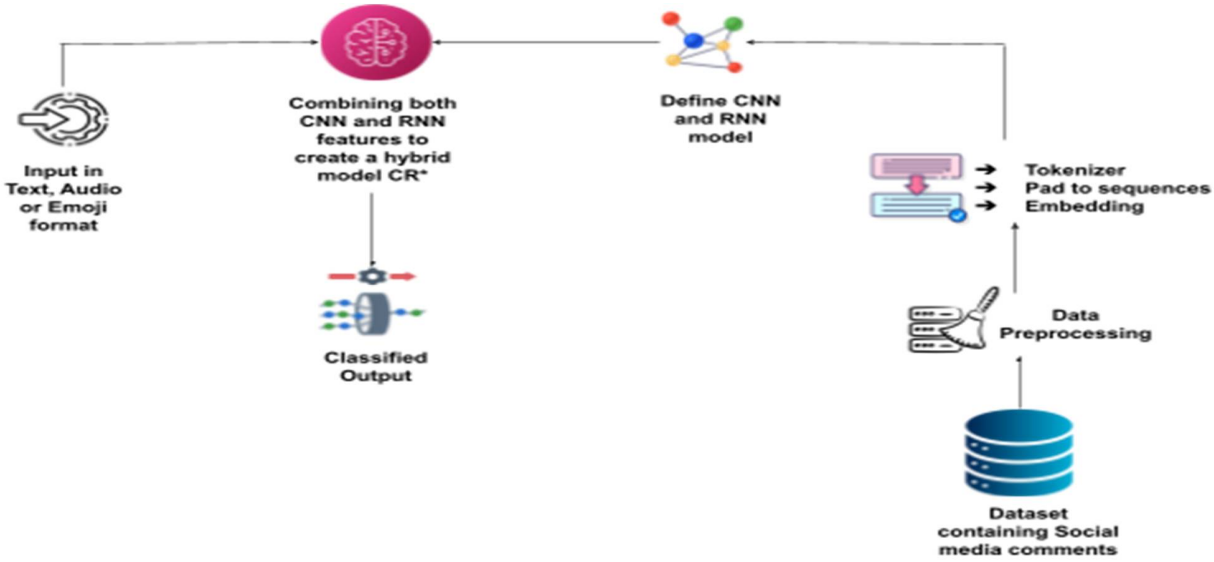


Fig 1: Design diagram of the proposed system

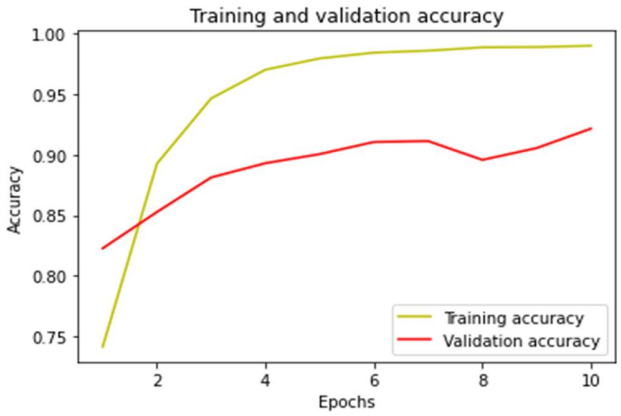


Fig 2:Training and validation accuracy

Fig 3:Training and validation loss

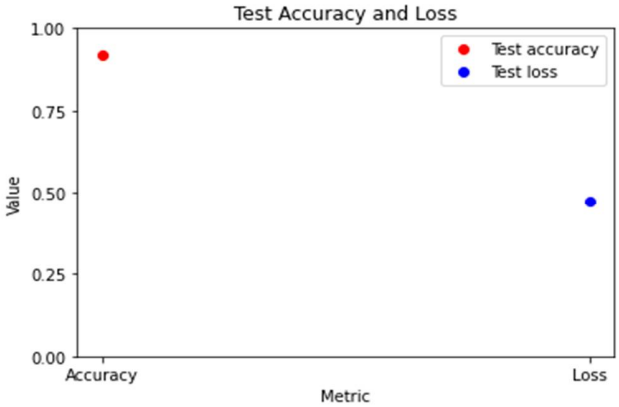


Fig 4:Testing accuracy and loss

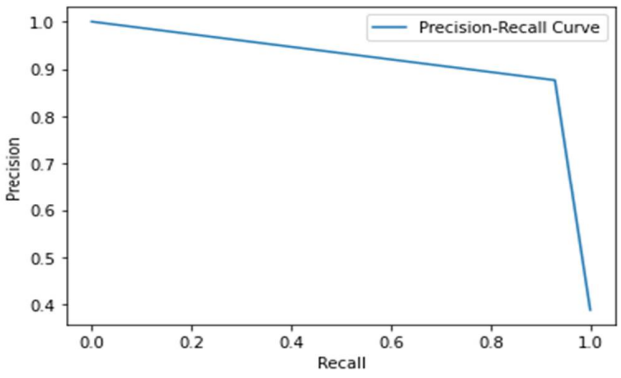


Fig 5:Precision-recall curve

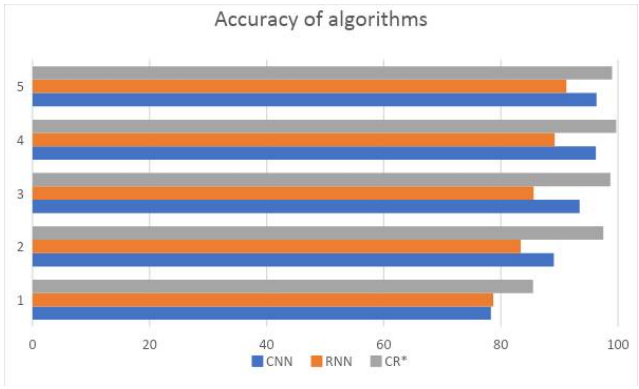


Fig 6: Comparing the accuracy of CNN, RNN and CR* with accuracy percentage in the x-axis and epochsin the y-axis

VI. RESULT ANALYSIS

A. Validating Text Input

Validating model with new / unseen data

```
[38] inp_type = input("To detect cyberbullying for Text (type 'text') for Audio (type 'audio') and for Emoji (type 'emoji') ")

To detect cyberbullying for Text (type 'text') for Audio (type 'audio') and for Emoji (type 'emoji') text

[39] if inp_type == "text":
         txt=input("Enter the input text:")
         Text_data(txt)
     elif inp_type == "audio":
         Audio_data()
     else :
         txt=input("Enter the input with emoji:")
         Emoji_data(txt)

Enter the input text:shut the fuck up you shit
1/1 [==============================] - 0s 29ms/step
Prediction score: 0.9778981
Cyberbullying
```

Fig 7:Validating text form of cyberbullying

B. Validating Audio (uploading recorded audio)

Validating model with new / unseen data

```
[31] inp_type = input("To detect cyberbullying for Text (type 'text') for Audio (type 'audio') and for Emoji (type 'emoji') ")

To detect cyberbullying for Text (type 'text') for Audio (type 'audio') and for Emoji (type 'emoji') audio

[32] if inp_type == "text":
         txt=input("Enter the input text:")
         Text_data(txt)
     elif inp_type == "audio":
         Audio_data()
     else :
         txt=input("Enter the input with emoji:")
         Emoji_data(txt)

Would you like to record audio using your microphone (type 'mic') or upload an existing audio file (type 'upload')? upload
Transcription: shut the fuck up you evil shit
1/1 [==============================] - 0s 374ms/step
Prediction score: 0.94527584
Cyberbullying
```

Fig 8:Validating audio form of cyberbullying through recorded audio

C. Validating Audio (using the microphone to detect audio)

Validating model with new / unseen data

```
24]: inp_type = input("To detect cyberbullying for Text (type 'text') for Audio (type 'audio') and for Emoji (type 'emoji') ")
To detect cyberbullying for Text (type 'text') for Audio (type 'audio') and for Emoji (type 'emoji') audio

33]: if inp_type == "text":
         txt=input("Enter the input text:")
         Text_data(txt)
     elif inp_type == "audio":
         Audio_data()
     else :
         txt=input("Enter the input with emoji:")
         Emoji_data(txt)
Would you like to record audio using your microphone (type 'mic') or upload an existing audio file (type 'upload')? mic
Say something!
Transcription: shut the fuck
1/1 [==============================] - 0s 21ms/step
Prediction score: 0.9876288
Cyberbullying
```

Fig 9:Validating audio form of cyberbullying through microphone

D. Validating Emoji:

Validating model with new / unseen data

```
[39] inp_type = input("To detect cyberbullying for Text (type 'text') for Audio (type 'audio') and for Emoji (type 'emoji') ")
     To detect cyberbullying for Text (type 'text') for Audio (type 'audio') and for Emoji (type 'emoji') emoji

[41] if inp_type == "text":
         txt=input("Enter the input text:")
         Text_data(txt)
     elif inp_type == "audio":
         Audio_data()
     else :
         txt=input("Enter the input with emoji:")
         Emoji_data(txt)
     Enter the input with emoji:you are a 💀
     Cyberbullying
```

Fig 10:Validating emoji form of cyberbullying

E. Overall Accuracy of the Proposed hybrid model(CR)*

```
from sklearn.metrics import accuracy_score
y_pred = model.predict(X_test)
y_pred_binary = np.round(y_pred)

accuracy = accuracy_score(y_test, y_pred_binary)
print("Accuracy of the model: {:.2f}%".format(accuracy*100))

125/125 [==============================] - 3s 24ms/step
Accuracy of the model: 92.15%
```

Fig 11:Accuracy of the proposed hybrid model

VII. CONCLUSION

In conclusion, cyberbullying is a serious issue that affects many individuals on social media platforms, and the detection of cyberbullying is essential to prevent its harmful effects. In this paper, we presented a proposed solution forcyberbullying detection with text, audio, andemoji data using a CR* model. We discussed the limitations and issues of earlier systems, includinglimited modality support, limited dataset size, lackof multimodal fusion, and limited audio and emojirepresentation. Our proposed system overcomes these limitations and provides a more comprehensive and achieved average of 92.15% accuracy, 81.7% Precision, 93.8% Recall and
87.3% F1-score.

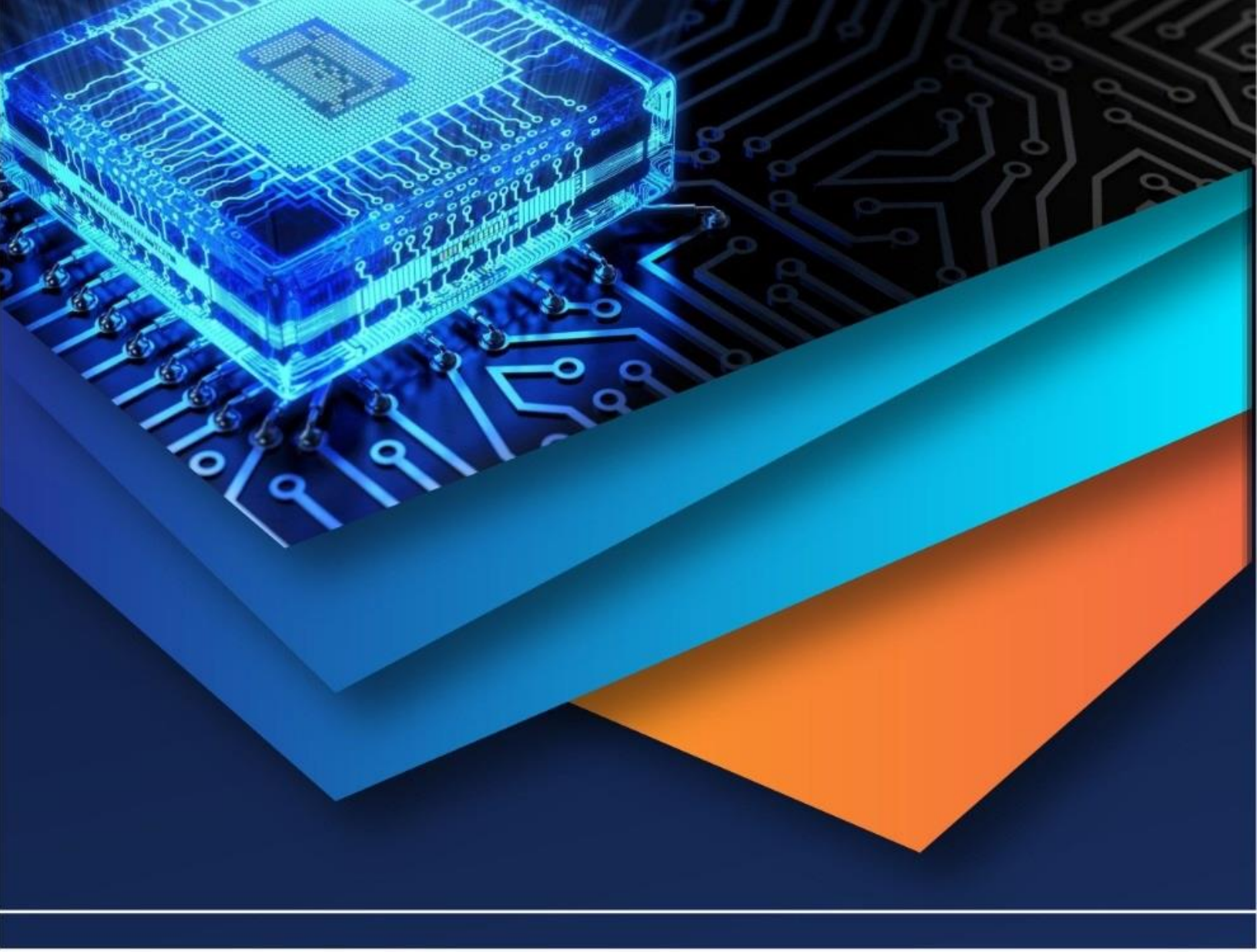
Future work can explore increasing the size and diversity of the dataset to improve the model's ability to detect rare, complex, and multilingual cyberbullying events. Overall, we believe our proposed system can make a valuable contribution to the field of cyberbullying detectionand help promote a safer and more inclusive socialmedia environment.

VIII. ACKNOWLEDGEMENT

We are deeply indebted to Dr.M.S.Anbarasi, Assistant Professor, Department of Information Technology, Puducherry Technological University,Puducherry, India.

REFERENCES

[1] Lee Jia Thun; Phoey Lee Teh; Chi-Bin Cheng;(2022). CyberAid: Are your children safe fromcyberbullying? Journal of King Saud University - Computer and Information Sciences, Elsevier, –. doi:10.1016/j.jksuci.2021.03.001
[2] B. A. H. Murshed, J. Abawajy, S. Mallappa, M. A. N. Saif and H. D. E. Al-Ariki, "DEA-RNN: A Hybrid Deep Learning Approach for Cyberbullying Detection in Twitter Social Media Platform," in IEEE Access, vol. 10, pp. 25857-25871, 2022, doi:10.1109/ACCESS.2022.3153675.
[3] P. K. Roy, A. K. Tripathy, T. K. Das, and
[4] X. -Z. Gao, "A Framework for Hate Speech Detection Using Deep Convolutional Neural Network," in IEEE Access, vol. 8, pp. 204951-204962, 2021, doi: 10.1109/ACCESS.2020.3037073.
[5] Raj, M., Singh, S., Solanki, K. et al. An Application to Detect Cyberbullying Using Machine Learning and Deep Learning Techniques. SN COMPUT. SCI. 3, 401 (2022),Springer. https://doi.org/10.1007/s42979-022-01308-5
[6] Bellmore, A. J. Calvin, J.-M. Xu, and X. Zhu, ‘‘The five W’s of ‘bullying on Twitter: Who, what, why, where, and when,’’ Comput. Hum.
[7] Behav., vol. 44, pp. 305–314, Mar. 2022
[8] Dadvar, M., Jong, D. F., Ordelman, R., & Trieschnigg, D. (2018). Improved cyberbullying detection using gender information. In Proceedings of the Twelfth Dutch-Belgian Information Retrieval (DIR 2022). The University of Ghent.
[9] M. A. Al-Garadi, K. D. Varathan, and S.
[10] Ravana, ‘‘Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network,’’ Comput. Hum. Behav., vol. 63, pp. 433–443, Oct. 2022.
[11] R. Zhao, A. Zhou, and K. Mao, ‘‘Automatic detection of cyberbullying on social networks based on bullying features,’’ in Proc. 17th Int. Conf. Distrib. Comput. Netw., 2022, Art. no. 43.
[12] Y. Chen, Y. Zhou, S. Zhu, and H. Xu, ‘‘Detecting offensive language in social media to protect adolescent online safety,’’ in Proc. Int. Conf. Privacy, Security., Risk Trust (PASSAT), Sep. 2022, pp. 71–80.
[13] Nahar, V., Li, X., Pang, C., & Zhang, Y. (2021). Cyberbullying detection based on text-stream classification. In The 11th Australasian Data Mining Conference (AusDM 2022).
[14] Dinakar, K., Reichart, R., and Lieberman, H., “Modelling the detection of textual cyberbullying,” Social Mobile Web Workshop at International Conference on Weblog and Social Media, Barcelona, Spain, July 17-21,2021.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY