Unsupervised Hybrid approaches for Cyberbullying Detection in Instagram

Abishak I., Kabilash M., Ramesh R.
B. Tech Students
Department of Computer Science and Engineering
Pondicherry Engineering College, Puducherry, India

Sheeba J.I.
Assistant Professor
Department of Computer Science and Engineering
Pondicherry Engineering College, Puducherry, India

Pradeep Devaneyan S.
Professor
Department of Mechanical Engineering
Sri Venkateshwara College of Engineering and Technology

ABSTRACT
In today's digital society, cyberbullying is serious and widespread issues affecting high number of Internet users, mostly teenager. However, increases in social media usage there is increase in the rise of cyberbullying. Cyberbullying is aggressive act carried out by some person using electronic forms of contact, repeatedly against people who cannot defend themselves. In the existing work distinguished the bullies from normal Instagram users by considering text, network and user related attributes using classifiers. Most existing cyberbullying detection methods are supervised and, thus, have mainly two key drawbacks such as labeling the data often take more time and labor and Current guidelines for labeling may not useful for future instances because of evolving social networks and different language usage. To address these limitations, this proposed work introduces for unsupervised cyberbullying detection method.
The proposed detection method will be extract linguistic attributes such as idioms, sarcasm, irony and active or passive voice. In addition, a representation learning network that learns the multi-modal session representations and a multitask learning network will simultaneously estimate bullying energy and then models the comments arriving times from Instagram data set.

General Terms
Cyberbullying Detection in Instagram

Keywords
Cyberbullying, Instagram, Multimodal, Hybrid, Unsupervised

1. INTRODUCTION
The large number of people using social platform sites has increased in the last few years. Social networking platform has become very popular in the last few years. Because social networking platform is one where we can easily share our information because of easily available, low cost etc. Nowadays in social media the people can easily interact, discuss, share by using multi modal features like multimedia text, videos pictures, and audio. Though, increase in social media usage have corresponded to increase in the occurrence of cyberbullying. Bullying, can occur at any time not limited to physical spaces (e.g., schools, colleges, sports fields or work places) can now occur anytime, anywhere [1]. Cyberbully is nothing but typically by sending messages of a threatening nature and repeated harassment or threat to an individual or group [2], has been increasing at a high rate. Previous research in the united states has found that nearly 43% of teenagers have been victims of cyberbullying. Because of this, many focused at automatically detecting cyberbullying. Though, detecting the cyberbullying in social networking is very challenging task.

In this proposed work, will focus on Instagram as the online social media platform were many are experiencing cyberbullying. Instagram is a social media platform in which users can share their images and videos to their followers either publicly or privately. In Instagram users can upload their photos or videos with relevant text, their locations, and hashtags to help other people to discover their photos. Cyberbullying in Instagram manifests itself in the form of hateful captions, hashtags and comments, humiliating images of victims with the use of fake profiles.

Even though there are many machine-learning approaches which is used in text classification tasks, there is an important drawback: they cannot combine semantic nuances of the written language. For example, considering the negation of words or sarcastic expressions with machine learning approaches is a quite challenging task. In order to come out of such difficulties, we can use deep-learning algorithms that build upon neural networks.

To build a model for classifying the abusive behavior which combines the (i) raw text, and (ii) the user, text, and network features (called metadata). At first to describe these two paths separately, which then the output is combined as a single model [3].

In addition, in this proposed work to introduce Unsupervised Cyberbullying Detection (UCD) which consists of two components such as (1) A representation learning network that learns the social media session by encoding multi-modal information such as e.g., text, time, and network. (2) A multi-task learning network that simultaneously estimates the sample bullying likelihood by using Gaussian Mixture Model (GMM) and then fits the comment inter-arrival times. Additionally, by using Hierarchical Attention Network (HAN) [4] for textual features and a Graph Auto-Encoder (GAE) [5] for the representation learning network models in social media sessions. The multitask learning network that takes the multi-modal representations such as (e.g., user, text, and social network) as input to estimate the sample bullying likelihood using a time-informed Gaussian Mixture Model (GMM). Both the components will increase their learning effectiveness. In this paper section 2 discussed the related works, section 3 is about the proposed work and section 4 described about the discussion and experimental results and finally section 5 is about the conclusion of the paper.

2. RELATED WORK

Automatically identifying Cyberbullying on Twitter is proposed by Nargess Tahmasbi and Elham Rastegari in December 2018. This paper introduces about to join the textual information with contextual and social characteristics and then find the important factors among them to propose a cyberbullying detection model [6]. Tracking illicit drug dealing and abuse in Instagram using multi modal analysis is proposed by Xi tong Yang and Jiebo Luo in February 2017. The main drawbacks of this method, the features show useful signals in pattern analysis, which are not detecting their significance for detecting drug dealer accounts [7]. Cyberbullying detection on social multimedia using soft computing techniques a meta-analysis is proposed by Akshi Kumar and Nitin Sachdeva in January 2019. This paper discussed the scope, feasibility and relevance of using soft computing techniques for cyberbullying detection on social media portals using textual content [8]. Approaches to Automated Detection of Cyberbullying a Survey is proposed by Semiu Salawu, Yulan He, and Joanna Lumsden in October 2017. This paper introduced the current state-of-the-art in cyberbullying detection and then provides a unique opportunity for cyberbullying research by categorizing, reviewing and identifying the current and existing work in the field. They lack of an adopted cyberbullying definition for detection purposes and a lack of large labeled cyberbullying are two key research issues facing cyberbullying detection research [9] Cyberbullying Detection on Instagram with Optimal Online Feature Selection is proposed by Mengfan Yao, Charalampos Chelmis and Daphney–Stavroula Zois in August 2018. A novel sequential approach testing formulation was proposed to address the problem of cyberbullying detection by efficiently and intelligent examining features. This method doesn't find the emotion and text features, such features has shown to being more informative for cyberbullying classification [10]. Mengfan Yao et al., proposed a approach for online cyberbullying detection in Instagram that can achieve accurate and timely detection while being scalable to the staggering rates at which content is generated. The main drawbacks of this method is experimental evaluation is limited to a single data sets that the performance of this approach should not be generalized to other platforms [11]. Detecting Cyberbullying and cyberaggression in social media is proposed by athena vakali and Gianluca stringhini. This proposed detection method could be made larger to consider not only text, user and network related features but also the linguistic attributes. To identify an Unsupervised Hybrid approaches for cyberbullying detection in Instagram instead of supervised method. In this method labeling the data is often consume more time and more labor are required and Current guidelines for labeling may not generalize to future instances. Cyberbully Images and Text Detection using convolutional Neural networks is proposed by S.V.Drishya and J.I.Sheeba in June 2019**.** This paper detected both cyberbully image and text and classify them as Harassing, Insulting, Trolling and Threatening using the popular social media Instagram. The numerical and discriminative representation, learning of textual messages is a critical issue [12]. Identifying and Classification of Cyberbully Incidents using Bystander Intervention Model is proposed by Revathy Cadiravane and J. I. Sheeba in July 2019. The focus is the direct intervention by bystanders and identifying the occurrence of cyber bullying activity in online social media platform which helps the government to yield force before many end-users enhancing a target of cyberbullying. The few bystanders ever try to reduce the conflicting effects of cyberbullying, and bystanders ever endeavor to interrupt [13]. J. I. Sheeba and B. Sri Nandhini proposed a model that focuses on detecting the cyber bullying activity in the social networking website using the Levenshtein algorithm [14]. Based on the above literature review inferred that most of the available cyberbullying detection methods are supervised and, thus, have following drawbacks: It will be needed a lot of computation time for training, the labeling the data is often take more time and labor. Also, Current guidelines for labeling may not be generalized to future instances because of many language usage and evolving social networks. Most of the existing work find only text, user and network related features but it doesn't find linguistic attributes of the posted content such as extensive use of idiomatic phrases, active or passive voice, and sarcasm or irony. To overcome above problems in this proposed work is going to identify an Unsupervised Hybrid approaches for cyberbullying detection in Instagram instead of supervised method. In addition, the proposed method also extended to consider not only text, user, and network features but also linguistic attributes such as use of idioms, sarcasm, irony and active or passive voice.

3. PROPOSED WORK

Unsupervised Cyberbully Detection (UCD)

Figure 1 represent the framework of the proposed system that can be explained as follows: The input data set contains multi modal input (text, image, audio and video) and metadata input (user, network and text related features) will be collected from social networks. The input of the data is sent to data preprocessing which improves quality of the input. After completion of data preprocessing the outcome of data preprocessing is sent to the neural network (Representation Learning Network and Multi-Task Learning Network) modules for detecting the cyberbully contents and finally, the cyberbully content is classified as bully, non-bully and other details.

In the proposed framework contains the following modules for identifying the cyberbully words.

3.1 Neural networks for Detecting Abuse

3.1.1 Multi modal Input

This part of the classifier only considers the multi modal input such as image, text, video, and audio. Initially, to combine all text of the single person into a single document and then convert the video and audio into text. If the text contains more words then the required length is trimmed, similarly if the text contains less words then it is padded with zeros. After preprocessed, the input text is fed to the network for learning [3].

3.1.2 Multi modal Preprocessing

In general, the data set consists of most noisy and unwanted data, so in order to make more accuracy of the input data then the process preprocessing has been applied. It includes the removing of stop words and symbols. Usually stop words are like "a", "as", "have", "is", "the", "or", etc., which consume memory space and reduce the processing time.

3.1.3 Embedding Layer

It is the first layer of the network performs an embedding layer that maps every word to the high dimensional vector. It is effective technique highly used for text classification tasks, additionally it will minimize the required training samples because to achieve a better performance.

3.1.4 RNN layer

The RNN layer is 128 units (neurons). After several experimental which conclude RNN architecture is Gate Recurrent Unit Or GRU. Additionally, in order to avoid the over fitting, here use the recurrent dropout p = 0.5 [3].

3.1.5 Metadata

The metadata network usually considers the non-sequential data. Before we feed data into neural network, will need to transform the any of the data into the numerical, either via one-hot encoding or enumeration, depending upon the particulars input.

3.1.6 Metadata Preprocessing

Before we feed data to the neural network, we need to transform any of the data into the numerical, either via one hot encoding or enumeration, depending upon the particulars input. Once this step has taken place, then each sample is represented as the features of vector of numerical.

3.1.7 Batch normalization layer

This will work best when the input of the data is zero mean, so that it will enable higher overall accuracy and faster learning. This layer takes care of data information at each level. It will reduce the over fitting and increases the stability of the neural network because usually normalization will normalize the output of the previous activation layer by subtracting the dividing by the batch standard deviation and batch mean.

3.1.8 Dense Layer

This layer uses the simple network of several, dense fully connected layers to read the metadata. In this layer bottleneck is formed. A bottleneck is a one layer which contains the few nodes when compared to previous layers. That can be used for obtaining the representation of input with the reduce dimensional [3].

3.2 Cyberbullying Detection

It contains two main components: (1) a representation learning network that uses the Hierarchical Attention Network (HAN) and then Graph Auto-Encoder (GAE) to obtain the multi-modal representations, (2) a multi-task learning network that use GMM related energy estimation task to predict the cyberbullying instances and the temporal predicting task to further reduce the session representations along with comment inter arrival times.

3.2.1 Representation Learning Network

3.2.1.1 Hierarchical Attention Network (HAN) for text

The HAN [15] is an approach used for generating the textual representation of the social media session. This is good approach in detecting the cyberbullying because it will model the two important levels of social media sessions (sequence of comments and words) and at each level it will differentiate the importance of specific words and comments and capture the long terms dependencies. The structure of hierarchical textual content will be defined as follows: the social media session which consists of sequence of comments and every comment which includes the sequence of the words. Here given the session with C as comments so that each comment i has the Li words {wit |t = 1, 2, ..., Li}, here, then use a bi directional Gated Recurrent Units (GRUs) [16] that model the both word sequence in the comment and then the comment sequence in the session.

3.2.1.2 Graph Auto-Encoder (GAE) for social networks

The representation learning network usually learns the user representation by exploring the information from the social networks in that nodes denote users with the related profile information that acts as node attributes, then the edges denote the user follower/follower information. Then, here employ GAE that will embed the user attributes as the low-dimensional vectors so that the users with related proximity in social network are very close. Since it is one of the most powerful approaches for node embedding, GAE is applied to the several challenging tasks such as the link prediction [17, 5] and then the node clustering [18]. GAE can also accurately incorporate the node features which then learn more about the interpretable representations of the user [5]. The objective of the GAE is usually encoding-decoding scheme, i.e., GAE will encode the nodes into the low dimensional vectors then it is decoded to form the original structure.

3.2.2 Multi-Task Learning Network

3.2.2.1 Bullying-energy estimation

A multi-task learning network will estimate the sample likelihood and then classify the samples with low likelihood as the bullying instances. The primary benefit of the energy-related models is to specify the flexibility of the energy expression [19]. So here, build a GMM-related density estimator that will form the underlying probability density function. The GMM is, usually an unsupervised learning method, which seeks to fit the multiple unimodal Gaussian distributions with multi-modal distribution these are the most normally used distributions for modelling the uni modal data. Existing works [19, 20] shown that the GMM is effective than the simple models for data with the complex structures. Then the given Multi-modal nature of social media data and the complexity, GMM will perform the density estimation tasks over the multi-modal representations [21].

3.2.2.2 Temporal dynamic fitting

The cyberbullying is usually defined as the act of repeated aggression which will develops over time [15, 22, 23]. Though, mostly the previous models consider the comment in the media session as an isolated event. So, they normally overlook the users commenting behavior of temporal dynamics. In this, we need to estimate the inter-arrival times between the sequence of comments to get the additional information from the temporal dynamics. This step will make the model to exploit the differences and commonalities across the temporal dynamic prediction and bullying-energy estimation for improving the performance of the cyberbullying detection [21].

3.3 Training the combined network

It is an approach where we need to train the full network as once and then two paths will have different convergence rates. We can induce the unpredictable interaction by using the standard back propagation on the other side of the network. To avoid these errors, we can use the interleaved training here.

In this method the data can flow through whole network, and then we can update the weight of one path. So, to train the two paths in the alternate fashionable. The result of network is more balanced and optimal and [3].

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Data set

Instagram is one of the social media where we can collect the posted captions, images, networks of followers/followings and. The data set can be downloaded from the following website.
https://gombru.github.io/2018/08/01/InstaCities1M. Here 500 posts are randomly chosen from the website.

4.2 Existing System

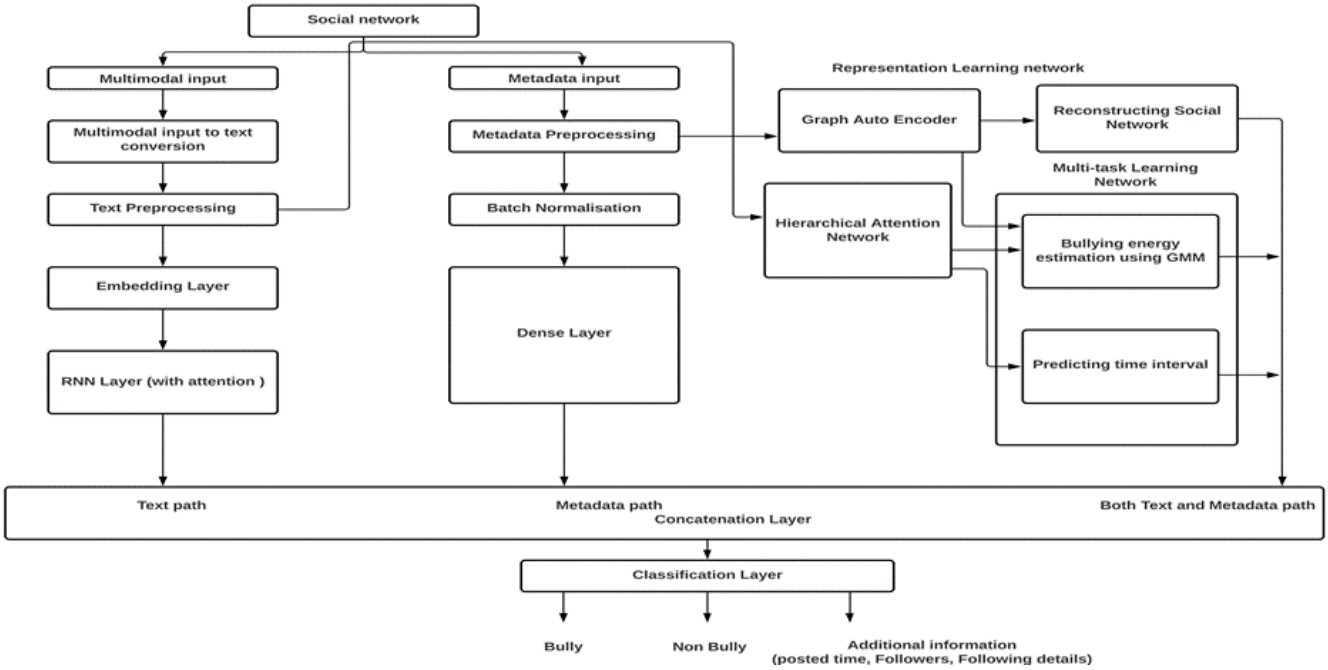

Fig 1: Proposed framework for UCD

4.2.1 Cyberbully Detection using Recurrent Neural Network (CDRNN)

This technique is mainly used for detecting abuse. The input is taken from the
https://gombru.github.io/2018/08/01/InstaCities1M.

Input is first given to the data preprocessing that will remove the irrelevant data and redundancy to improve the quality of the input. After performing the preprocessing, then the output is given to the feature extraction. With the given training data set then the preprocessed social network conservation is tested whether the bullying word is present or not. The Recurrent Neural Network is used mainly to identify the cyberbullying words is present in that conversation or not and then to display whether these words are cyberbullying words or not.

4.3 Metrics Considered for Evaluation

Then the performance of our proposed framework is measured in terms of the quality measures such as precision, recall, F-measure, classification accuracy, RMSE, sensitivity and specificity.

4.3.1 Precision

It is the fraction of the retrieved cyberbully words which are relevant to find.

$$\Pr ecision = \frac{\{A\} \cap \{B\}}{\{B\}}$$

Where A- Number of relevant cyberbully words,
B- Number. of retrieved cyberbully words

4.3.2 Recall

It is the fraction of the cyberbully words which are relevant to the query that are successfully retrieved

$$\text{Re}\,call = \frac{\{A\} \cap \{B\}}{\{A\}}$$

Where A- Number of relevant cyberbully words,
B- No. of retrieved cyberbully words.

4.3.3 F-measure

F-measure is the computation of both precision and recall. Which can be estimated by using the below formula.

$$F - Measure = 2.\frac{A.B}{A+B}$$

Where A- Precision
B - Recall

4.3.4 Classification Accuracy

Accuracy will calculate the proportion of correctly identified cyberbully words, and it can be estimated by using the following formula:

$$Classification\ Accuracy = \frac{(TP+TN)}{(TN+TP+FP+FN)}$$

In respect of cyberbully detection, the terms are evaluated in the following manner:

TP – Determined as a word being classified correctly as relating to a cyberbully category

TN–Determined the words which were non cyberbully words

FP – Determined as a cyberbully word even if it is in the non-cyberbully category

FN – Determined as a non-cyberbully word even if it is in cyberbully category.

4.3.5 Root Mean Square Error (RMSE)

It is the difference between classifying the cyberbully words predicted by a system and the cyberbully words actually observed from the input. It is estimated by using the following equation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{objects,i} - X_{items,i})^2}{n}}$$

where Xobjects is manually classified cyberbully words, and Xitems is system classified cyberbully words at time/place i. n is the number of inputs.

4.3.6 Sensitivity (also Called True Positive Rate)

Sensitivity is ability to identify a condition correctly. It will classify cyberbully words which are under the cyberbully words category in the given input. It is estimated by using the following equation:

$$Sensitivity = \frac{TP}{(TP+FN)}$$

4.3.7 Specificity (Also Called True Negative Rate)

Specificity is ability to exclude a condition correctly. It will classify cyberbully words which are not under the cyberbully words category in the given input. It is estimated by using the following equation:

$$Specificity = \frac{TN}{(TN+FP)}$$

4.4 Experimental Results and Discussion

Experiments have been repeated for randomly shuffled posts and the results are obtained for the dataset. Table 1 shows the comparison of the performance of existing and proposed techniques in terms of F-measure, Accuracy and RMSE.

Table 1. F-measure, Accuracy, RMSE values obtained from Existing and Proposed Technique

Existing Technique (CDRNN)			Proposed Technique (UCD)		
F Measure	Accuracy	RMSE	F Measure	Accuracy	RMSE
0.853	81	0.356	0.962	88	0.257

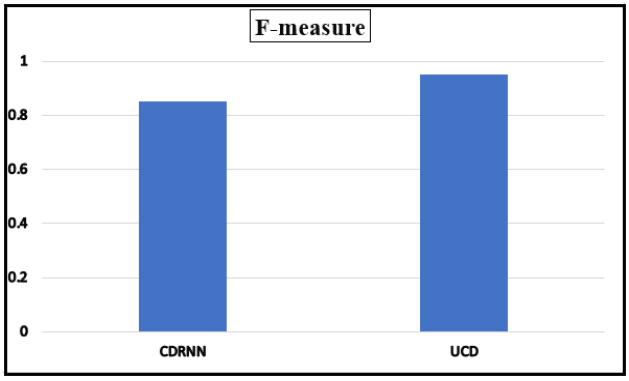

Fig 2: F-measure

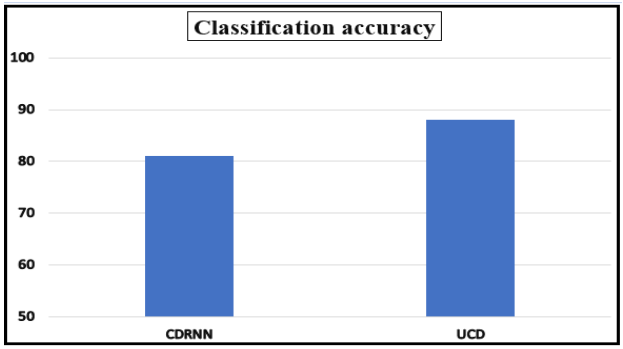

Fig 3: Classification Accuracy

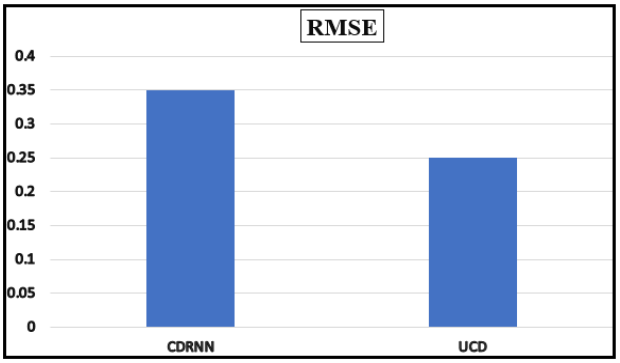

Fig 4: RMSE

Table 2 shows the comparison of the performance of existing and proposed techniques in terms of Sensitivity

Table 2. Sensitivity

	Number of Posts				
Techniques	**1-100**	**101-200**	**201-300**	**301-400**	**401-500**
CDRNN	0.66	0.619	0.64	0.66	0.62
UCD	0.72	0.715	0.73	0.698	0.70

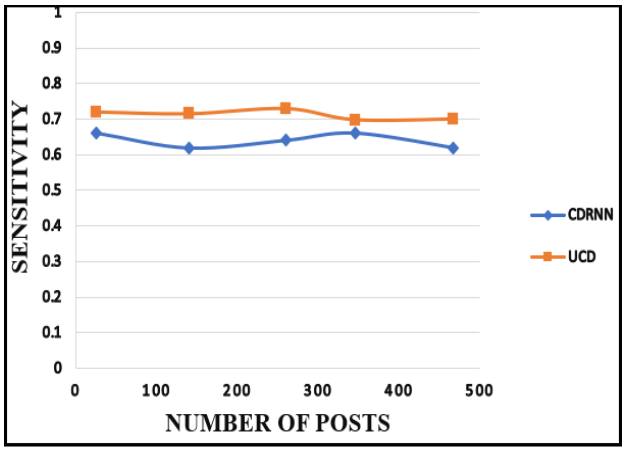

Fig 5: Performance analysis based on sensitivity values compared with various techniques

Table 3 shows the comparison of the performance of existing and proposed techniques in terms of Specificity

Table 3. Specificity

	Number of Posts				
Techniques	**1-100**	**101-200**	**201-300**	**301-400**	**401-500**
CDRNN	0.69	0.62	0.65	0.66	0.644
UCD	0.72	0.71	0.739	0.76	0.67

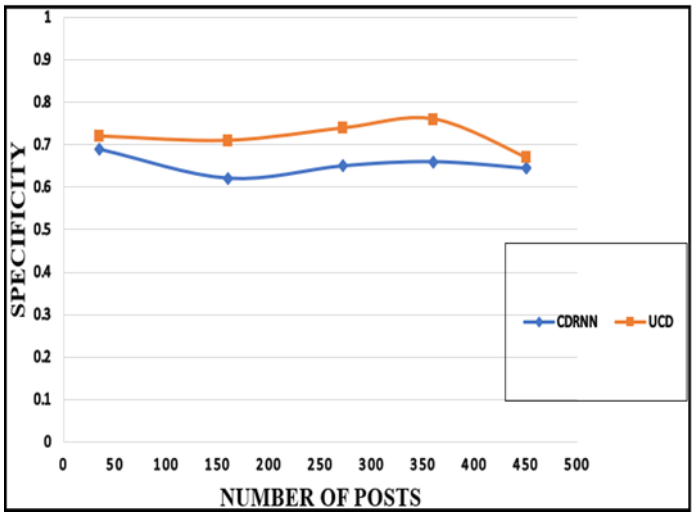

Fig 6: Performance analysis based on specificity values compared with various techniques

The performance is evaluated with the posts collected from the Instagram data set by randomly shuffled 500 posts, results are compared with existing system and proposed system. Figure 2 shows the F-measure values obtained from different techniques. CDRNN denotes the cyberbully detection using Recurrent Neural Network, and UCD denotes Unsupervised cyberbully detection. Figure 2 shown that the values of F-measure obtained using UCD is much better than the values obtained using CDRNN. Thus, the result shows the proposed system perform better than the existing system. Figure 3 shows that the graph of classification accuracy obtained from different techniques. Thus, the proposed system shows more accuracy than the existing system. Figure 4 shows the RMSE graph where system using the UCD has shown less error value than the system using the CDRNN technique. Specificity and sensitivity values are calculated and then the results obtained from the given data set are shown in Figures 5 and 6. It is observed from Figures 5 and 6 thus UCD as achieved the good specificity and sensitivity measure when compared with the existing technique CDRNN.

5. ACKNOWLEDGMENTS

Our thanks to the author J. I. Sheeba who has done the Monitoring and Guidance part and S. Pradeep Devaneyan who has done for validation part.

6. CONCLUSION

The existing system consider only the text, user and network related features for detecting cyberbullying detection. But in the proposed work will find the linguistic attributes such as, use of idioms, sarcasm, irony and active or passive voice along with existing system features. Additionally, proposed system will support the unsupervised cyberbullying detection from the Instagram. This proposed framework shown better performance, while our aim is to control the users from the cyberbully in Instagram who are becoming the victims.

7. REFERENCES

[1] P. K. Smith, J. Mahdavi, M. Carvalho, and N. Tippett, "An investigation into cyberbullying, its forms, awareness and impact, and the relationship between age and gender in cyberbullying,", Research Brief No. RBX03- 06. London: DfES, 2006.

[2] Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & Tippett, N. (2008). "Cyberbullying: its nature and impact in secondary school pupils. Journal of Child Psychology and Psychiatry, 49(4), 376–385.

[3] Athena vakali and Gianluca stringhini 2019, "Detecting cyberbullying and cyberaggression in Social media", ACM transaction on web vol. 13, no.13 Article 17.

[4] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy, "Hierarchical attention networks for document classification", Proceedings of NAACL-HLT 2016, pages 1480–1489.

[5] Thomas N Kipf and Max Welling. "Variational graph auto-encoders". arXiv preprint arXiv:1611.07308 (2016).

[6] Nargess Tahmasbi and Elham Rastegari, "A Socio-Contextual Approach in Automated Detection of Public Cyberbullying on Twitter", ACM Transactions on Social Computing, Volume: 1, No. 4, Article 15, 22 pages, December 2018.

[7] Xi tong Yang and Jiebo Luo, "Tracking illicit drug dealing and abuse in Instagram using multi modal analysis", ACM Transactions on Intelligent Systems and Technology, Vol. 8, No. 4, Article 58, 15 pages, February 2017.

[8] Akshi Kumar and Nitin Sachdeva, "Cyberbullying detection on social multimedia using soft computing techniques: a meta-analysis", Springer on Multimedia Tools and Applications, pages:23973–24010, 23 January 2019.

[9] Semiu Salawu, Yulan He, and Joanna Lumsden, "Approaches to Automated Detection of Cyberbullying: A Survey", IEEE Transactions on Affective Computing, Vol. 11, Issue: 1, pages: 3-24, 10 October 2017.

[10] Mengfan Yao, Charalampos Chelmis and Daphney–Stavroula Zois, "Cyberbullying Detection on Instagram with Optimal Online Feature Selection", IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 28-31 august 2018.

[11] Mengfan Yao, Charalampos Chelmis, and Daphney–Stavroula Zois. 2019. "Cyberbullying Ends Here: Towards Robust Detection of Cyberbullying in Social Media". In Proceedings of the 2019 World Wide Web Conference (WWW'19), Pages 3427-3433 13–17 May 2019.

[12] S.V.Drishya and J.I.Sheeba" Cyberbully Images and Text Detection using convolutional Neural networks" CiiT International Journal of Fuzzy Systems, Vol 11, No 2, April - June 2019

[13] Revathy Cadiravane and J.I.Sheeba "Identification and Classification of Cyberbully Incidents using Bystander Intervention Model" IJRTE, ISSN: 2277-3878, Volume-8 Issue-2S4, July 2019

[14] B.Sri Nandhini, J.I.Sheeba,: Cyberbullying Detection and Classification Using Information Retrieval Algorithm. In: ICARCSET '15, ACM, pp.1-5 (2015)

[15] Lu Cheng, Ruocheng Guo, Yasin Silva, Deborah Hall, and Huan Liu. 2019. Hierarchical Attention Networks for Cyberbullying Detection on the Instagram Social Network. In SDM, pages:235-243.

[16] [Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate". Published as a conference paper at ICLR 2015, pages: 1-15.

[17] Aditya Grover, Aaron Zweig, and Stefano Ermon. "Graphite: Iterative generative modeling of graphs". Proceedings of the 36th International Conference on Machine Learning, PMLR 97:2434-2444, 2019.

[18] Guillaume Salha, Romain Hennequin, Viet Anh Tran, and Michalis Vazirgiannis. "A degeneracy framework for scalable graph auto encoders". arXiv preprint arXiv:1902.08813 (2019).

[19] Shuangfei Zhai, Yu Cheng, Weining Lu, and Zhongfei Zhang. "Deep structured energy related models for anomaly detection". Proceedings of the 33rd International Conference on Machine Learning, PMLR 48:1100-1109, 2016.

[20] Bo Zong, Qi Song, Martin Renqiang Min, Wei Cheng, Cristian Lumezanu, Daeki Cho, and Haifeng Chen. 2018. "Deep auto encoding gaussian mixture model for unsupervised anomaly detection". In ICLR, pages:1-19.

[21] Lu Cheng, Kai Shu, Siqi Wu, Yasin N. Silva, Deborah L. Hall, Huan Liu. 2020. "Unsupervised Cyberbullying Detection via Time-Informed Gaussian Mixture Model". In Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20), October 19–23, 2020, Virtual Event, Ireland. ACM, New York, NY, USA, 10 pages.

[22] Dinakar, K., Jones, B., Havasi, C., Lieberman, H., & Picard, R. (2012). Common Sense Reasoning for Detection, Prevention, and Mitigation of Cyberbullying. ACM Transactions on Interactive Intelligent Systems, 2(3), 1–30.

[23] Devin Soni and Vivek Singh. 2018. "Time Reveals All Wounds: Modeling Temporal Characteristics of Cyberbullying". Proceedings of the International AAAI Conference on Web and Social Media, 12(1), pages: 648-687.