Performance Comparison of Deep Learning Architectures for Cyberbullying Detection on Multi-Modal Data

Subbaraju Pericherla[1†] and E Ilavarasan[2††],
Raju.pericherla74@gmail.com Eilavarasan@pec.edu
Puducherry Technological University, Pondicherry, INDIA

Abstract
Cyberbullying drastically increased with increase of Internet and Social media networks from the last one decade. The various forms of bullying increasing with increase of smart phones . Now a days bullies are targeting the victims not only through the text , but they may also send images, videos, graphics, emojis . In this paper, we compared various types of deep learning architectures for cyberbullying detection on multi-modal data . This approach able to handle bullying detection for text and image combinational data. We mainly focused to extract image features embeddings using advance architectures such as Inception,VGG19, ResNet, Xception, MobileNet, Desnet and EfficentNet. We used RoBERTa deep learning architecture to generate word embeddings from the text data. We used various machine learning classifiers such as Logistic regression(LR), XGBoost(XG), LightGBM(GBM), Decision Tree(DT), Bernoulli Navie Bayes(BNB) and Gaussian Navie Bayes(GNB) to classify bullying and non-bullying tweets. The experiments conducted on 2100 samples of combined data of text and image. The Xception and LightGBM classifier combination performed well as compared to other combination of deep learning networks and classifiers.

Keywords:
Cyberbullying, Social Networks, Natural Language Processing, RoBERTa, EfficientNet.

1. Introduction

Social media networks are great platforms to share ideas, opinions, and express feelings to others in this digital era. Netizens use different forms of data such as text, images, audio, and videos while they share ideas or express feelings. Online harassment and abuse have been increasing with the increase of these technologies, social media, and smart devices. Online harassment, abuse, online aggressive behavior commonly referred to as cyberbullying [1]. A study has been conducted over 2000 school students of age between 12 and 18 years. The reports says that 11% of them bullied at least one time [2]. According to, United States National Crime Prevention council [3], the cyberbullying is defined as the use of digital devices to send abusive text or images to denigrate others. Adding on, it includes sending offensive gossip or spreading rumors about a target person in the social media. A recent study [4] found that 43% of the teenagers in the United States are victims of cyberbullying. Cyberbullying is more persistent by targeting the individual at a large network of people in social media which can eventually lead to several problems such as mental and psychological. Some of the victims face depression, sadness, loneliness, health degradation [5].

Cyberbullying challenges are increasing with multi-media technologies. Earlier most of the bullying activities happened through text messages. Now a days, bullies can send images, graphics, audio and videos with handy digital devices. Netizens have full freedom to send messages in social media. Short form of the messages and spelling corrections are challenges in the identification of bullying messages. Bullies use images to target the victims by sending ugly expressions, animals, or embarrassing images. Some cases they use combinational of both text and images for bullying. Most of the research works conducted with text data only for cyberbullying detection. Very few studies have been conducted with only image data and multi-modal data. In this paper, we have demonstrated the performance comparison of state-of-the-art deep learning architectures for cyberbullying detection on multimodal data.

The rest of the paper is organized as follows. Section 2 describes related works in cyberbullying detection. Section 3 outlines the workflow of cyberbullying detection. Section 4 discusses datasets and experimental results. Section 5 represents conclusion and future work.

2. Literature Review

Yong Fong et al. [6], proposed bi-directional gated recurrent unit (Bi-GRU) with self-attention mechanism. The self-attention mechanism helped to extract the weight of the words from each sentence to increase the classification accuracy. I.J Sheeba et al. [7], proposed unsupervised hybrid approach called Unsupervised Cyberbullying Detection (UCD) for cyberbullying detection for Instagram social network data. The hybrid method can extract linguistic features such as sarcasm, irony, idioms. Kiriti kumari et al. [8], proposed deep

Manuscript received November 5, 2022
Manuscript revised November 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.11.96

learning model for cyberbullying detection to classify textual comments into three categories on multi-modal data (Aggressive, moderate aggressive, non-aggressive). The bullying data consists of both text messages and images. The model was built with two deep learning architectures to handle text and image data. They applied Binary particle swarm optimization (BPSO) to select the best features. Sayantha Paul and Sriparna Saha [9] presented a unique application of Bidirectional Encoder Representations from Transformers (BERT) for cyberbullying detection. The proposed application achieves a better f1-score compared to stare of the art methods for classification task on three different datasets. Tripati et al. [10], proposed fine tune model of A Light Bidirectional Encoder Representations from Transformers (ALBERT) for cyberbullying analysis. The proposed fine-tuned model achieves f1-score 95% on real time datasets. Lu Cheng et al. [11], studied the problems of cyberbullying detection on multi-modal of social media. They proposed a novel framework called XBully to identify bullying for multi-modal data. Harsh Dani et al. [12], developed a learning framework to identify bullying messages automatically. They proposed leverage sentiment information for cyberbullying detection. Devin Soni and Vivek Singh [13] reported that audio and video features are significantly responsible for cyberbullying occurrence. They considered audio features like several words, sentiment of the spoken content, balance of voice and loudness for audio analysis. They considered visual features like number of faces, length of the visual text, scene labels for visual analysis. Vimal Balakrishnan and Hamid Arabnia[14] study deals with cyberbullying detection using psychological features including personalities, sentiment, motion . They took Big Five and Dark Traid models for psychology and used different machine learning classifiers like Navie Bayes, Random Forest, J48 to classify the tweets into four categories: normal, spammer, aggressor, and bully. Hendro Margano et al. [15], analyses bullying word on Twitter for Indonesian language. They identified “bangsat” and “anging” are the most frequently used terms for bullying in Indonesian Twitter. They used Association Rule and FP-growth algorithm to find the pattern between the words. Maral dadvar et.al.[16], considers that gender feature for identification of bullying. They used support vector machine learning algorithm as a classification of bullying and non-bullying messages. Fan Yang et al. [17], explore the challenge of identification of hate speech for multi-modal data. They present a fusion approach to integrate text and images. Kiriti kumari et al. [18], extracted features from text and images for cyberbullying detection. They used VGG16 pretrained network to extract the image features and convolution neural network to extract the features from text. They used genetic algorithms to identify the best features from combined text and image features.

3. Workflow of multimodal based bullying detection

In this section, we have presented our selected framework [36] for cyberbullying detection as illustrated in Fig 1. In the following subsections we describe each component in detail.

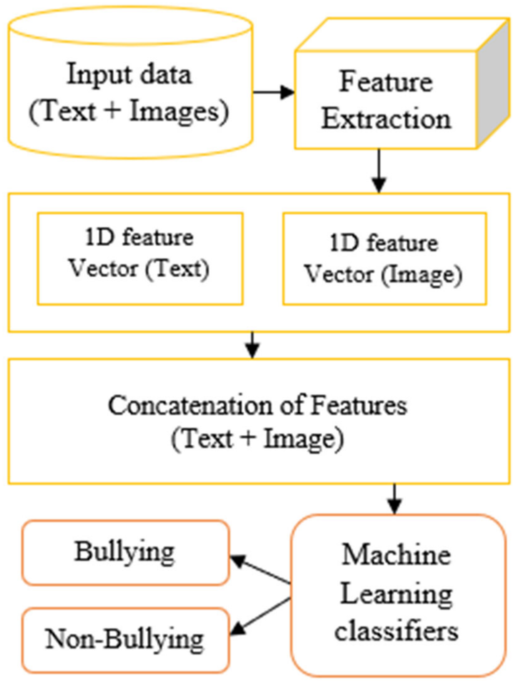

Fig 1: Workflow of the proposed Framework.

There are a number of different teaching systems that exist in different countries. The most popular system is the traditional face-to-face teaching system. In this traditional system [11,12], all the courses are taught in classes with personal contact. Classes that need practical skills benefit very much from the system where the students are allowed to attend classes and labs physically and get direct guidance from instructors.

On the other hand, distance learning (or online learning) [13,14,15,16,17,18,19,20] started as another way of teaching which offers new opportunities to people who cannot attend classes physically.

A. Data collection:

The experimental dataset is a multi-modal dataset created by Kiriti Kumari et al. [18]. It is a multi-modal dataset which is a combination of text data and image data. Totally, 2100 samples are used for all experiments. Every record in the dataset is a combination of two fields, namely image and text. Table 1 shows the summary of the dataset.

Table 1: summary of the dataset

Margin	*Text data*	*Image data*	*Combined Text+ Image*
Non-Bullying	1216	1636	1481
Bullying	884	464	619
Total	2100	2100	2100

B. Feature extraction Techniques:

Pretrained Word Embeddings:

In this section, we discussed various pre trained architectures used to extract the features from multi-modal input data. We applied from basic to advance pre-trained deep learning architectures such as Inception to Efficient Net to extract the images features and Roberta is used to extract the text features.

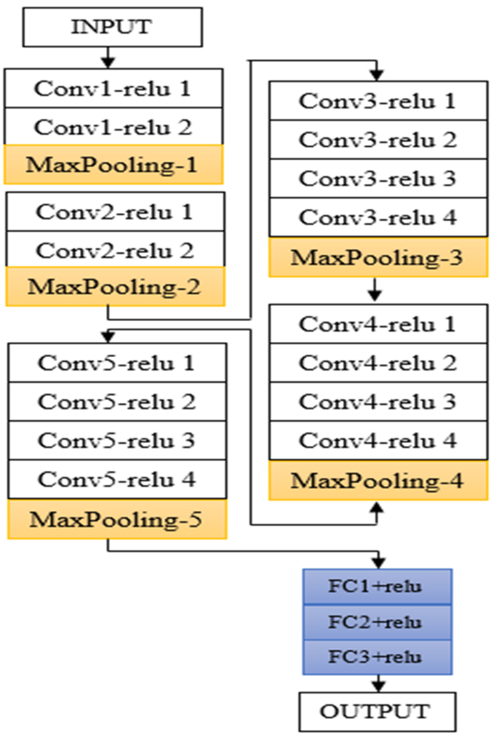

Fig 2: Architecture of VGG19 network. [19]

Visual Geometry Group (VGG):

The field of image recognition rapidly gained significance with the invention of convolution neural networks proposed by Simonian and Zisserman et al. [19]. They proposed a elegant design for CNN architecture named Visual Geometry Group. It was made with 16 layers deep compared to previous image architectures ZfNet[20] and Alex net[21]. ZfNet showed that small size filter improves the performance of the architectures in ILSVRC competition. Based on these findings, VGG replaced with large filters like 11x11 and 5x5 to 3x3. This low filter size technique reduces the computational complexity in the network training by decreasing hyperparameters. VGG network proved with results in both for localization and classification problems in images. The drawback of the VGG network is to train more than one billion parameter training, which makes the network expensive. Fig 2 shows the basic architecture of VGG.

Inception:

In the year 2014, Szeged C et al. [22], proposed a novel network with deeper convolutions called Google Net. They introduced the concept of inception block in CNN, it incorporates convolutional transformations using split, transform. In this network, the convolutional layers are replaced by small blocks. These blocks are able to capture spatial information like the traditional filters of different sizes used in convolution layers (1x1,3x3, and 5x5) .

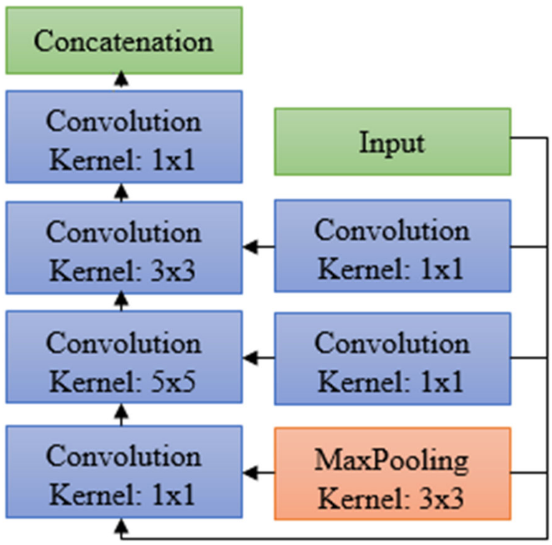

Fig 3: Basic architecture of Inception Block.[22]

The concept of split, transform and merge in Google Net helps to address the problem of images of different resolutions. Google Net decreases the computations by adding a layer of 1x1 filter. It uses sparse connections to avoid redundant information and eliminates irrelevant features. Global average pooling reduces the connection density at the last layer before connecting to the fully connected layer. These tunings in the parameters makes a rapid decrease in the number of parameters from 139 million to 5 million parameters. Fig.3 shows the architecture of the inception block.

ResNet:

He et al. [23], proposed 152 layers deep CNN, which was the winner of 2015-ILSVRC competition. ResNet is 8X deeper than previous architectures like VGG. The error rate decreased to 3.57% on the ImageNet dataset. They designed various types of ResNet architectures based on the layers from minimum to 34 to maximum 1202. The most well know architecture is ResNet50, which is 49 convolutions layers depth and finally connected a single FC layer. The innovative concept is used in ResNet by bypassing the pathway concept as shown in the Fig.4. Fig5 shows the fundamental block diagram of ResNet which consists of a conventional feed forward network plus a residual connection. The residual output is xl, mathematically represented as in Equation-1.

$$X_l = F(x_{l-1}) + x_{l-1}. \quad (1)$$

X_{l-1} ---- preceding layer output.

$F(x_{l-1})$ ----- before applying activation function.

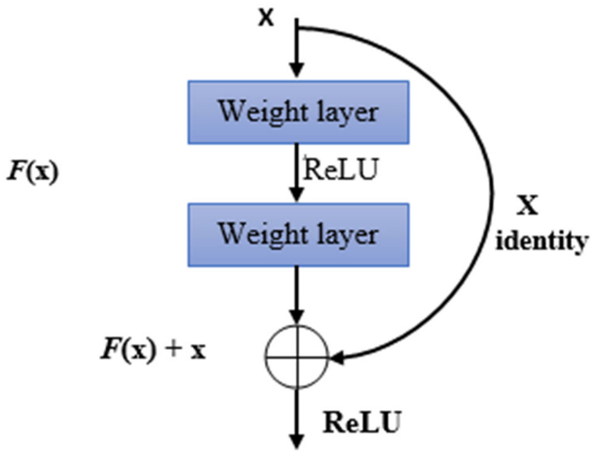

Fig 4: Residual Block in ResNet. [23]

Xception:

Francois Chollet [24] proposed a novel deep convolution neural network inspired by Inception architecture called Exception. Exception drastically reduced the complexity by exchanging a single dimension (3x3) followed by a 1x1 convolution. Decoupling and feature-map channel correlations makes the Exception network computationally efficient. Table 2 and Table 3 shows results of Exception architecture on ImageNet dataset and JST dataset.

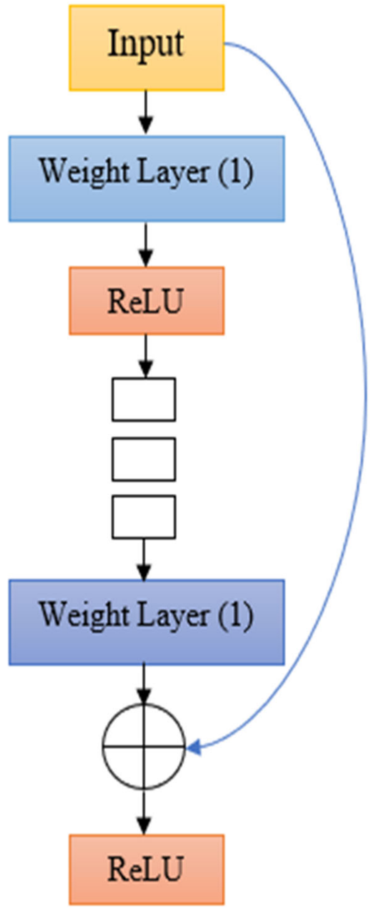

Fig 5: Basic Block diagram of ResNet.[23]

Xception:

Francois Chollet [24] proposed a novel deep convolution neural network inspired by Inception architecture called Exception. Exception drastically reduced the complexity by exchanging a single dimension (3x3) followed by a 1x1 convolution. Decoupling and feature-map channel correlations makes the Exception network computationally efficient. Table 2 and Table 3 shows results of Exception architecture on ImageNet dataset and JST dataset.

Table 2: Classification performance comparison on ImageNet dataset [24]

	Top-1 accuracy	*Top-5 accuracy*
VGG-16	0.715	0.901
ResNet-152	0.770	0.933
Inception V3	0.782	0.941
Xception	0.790	0,945

	FastEval14K MAP100
Inception V3- no FC Layers	6.36
Xception -no FC layers	6.70
Inception V3 with FC layers	6.50
Xception with FC layers	6.78

Table 3: Classification performance comparison on JFT dataset [24]

Mobile Net:

Andrew Howard et.al [25], presented an efficient model called Mobile Nets for computer vision and mobile applications. This architecture works based depth wise separable convolutions to build light weight deep neural networks. Depth wise separable convolution are made up of two layers: depth wise convolutions and pointwise convolutions. We use depth wise convolutions to apply a single filter to each input channel (input depth). Pointwise convolution, asimple1×1convolution, is then used to create a linear com-bination of the output of the depth wise layer. Mobile Net uses 3x3 depth wise separable convolutions which uses between 8 to 9 times less computation than standard convolutions. Table 4 shows Mobile Net body structure.

Type/Stride	***Filter Shape***	***Input Size***
Conv/s2	3x3x3x 32	224 x 224 x3
Conv dw/sl	3 x 3 x 32 dw	112 x 112 x 32
Conv/sl	1 x 1 x 32 x 61	112 x 112 x 32
Conv dw/s2	3 x 3 x 64 dw	112 x 112 x 64
Conv/s1	1x1 x 64 x 128	56 x 56 x 64
Cony dw/sl	3 x 3 x 128 dw	56 x 56 x 128
Conv/sl	1 x 1 x 128 x 128	56 x 56 x 128
Conv dw/s2	3 x 3 x 128 dw	56 x 56 x 128
Conv/s1	1 x 1 x 128 x 256	28 x 28 x 128
Cony dw/sl	3 x 3 x 256 dw	28 x 28 x 128
Conv/sl	1 x 1 x 256 x 512	28 x 28 x 128
Conv dw/s2	3 x 3 x 256 dw	28 x 28 x 128
Conv/s1	1 x 1 x 256 x 512	14 x 14 x 256
5 * Cony dw/sl	3 x 3 x 512 dw	14 x 14 x 512
Conv/sl	1 x 1 x 512 x 512	14 x 14 x 512
Conv dw/s2	3 x 3 x 512 dw	14 x 14 x 512
Conv/s1	1 x 1 x 512 x 1024	7 x 7 x 512
Conv dw/s2	3 x 3 x 1024 dw	7 x 7 x 1024
Conv/s1	1 x 1 x 1024x 1024	7 x 7 x 1024
Avg Pool/st	Pool 7 x 7	7 x 7 x 1024
FC/sl	1024 x 1000	1 x 1 x 1024
Softmax /SL	Classifier	1 x 1 x 1000

Table-4: Mobile Net Body structure.[25]

Dense Net:

Gao Huang et.al [26], introduced Dense convolution network (Dense Net). It uses feed forward fashion by connecting each layer to every each other layer in the network. It reduces the number of connections of n*(n+1)/2 of traditional convolution networks. Dense Net architecture reduces the vanishing grading problem and strengthens the feature propagation for the next input layers. Dense Net architecture evaluated on highly competitive benchmark datasets (ImageNet, SVHR, CIFAR-10, CIFAR-100). Fig 6 shows the basic architecture of Dense Net block.

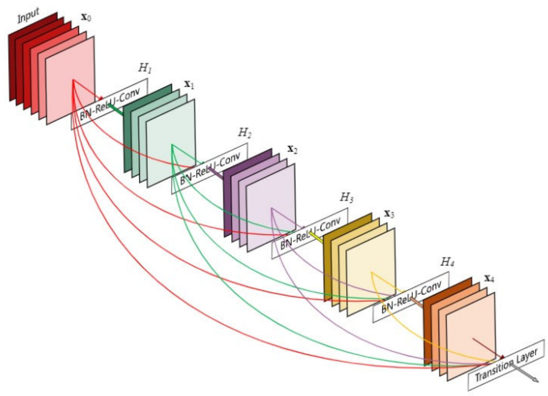

Fig 6: A 5-layer dense block [26]

EfficentNet:

Ming xing Tan and Quoc V. Le [27] proposed a new scaling method to scale up CNNs in well-structured manner. To demonstrate scaling method applied on Mobile Nets and ResNet. They use neural architecture search to design a new baseline network and scale it up to obtain family of models named as Efficient Net which performs much better efficiency and accuracy than previous convolution networks. The base EfficientNetB0 network is based on inverted residual blocks of MobileNetV2. The EfficientNet-B7 achieves 84.3% accuracy on ImageNet with less number parameters as compared to other networks as shown in Fig 7.

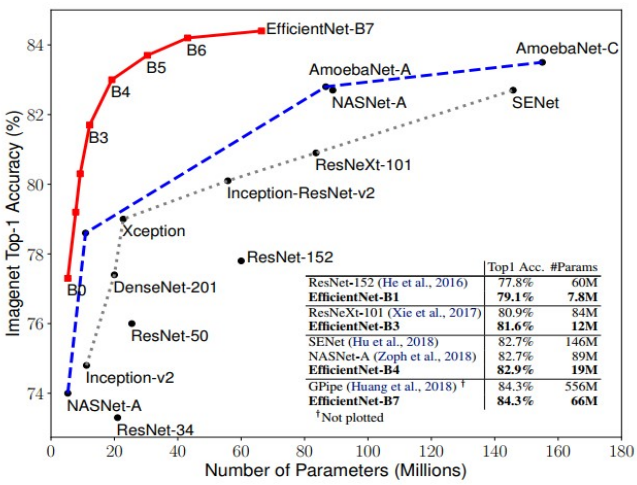

Fig 7: Model size vs ImageNet Accuracy [27].

RoBERTa:

BERT [28] (Bidirectional Encoder Representations from Transformers) architecture has bought a significant change in the Natural language processing tasks. Roberta [29] is an optimized method for BERT, produces excellent results as compared to state-of-art methods. It builds on BERT's masking strategy by modifying hyperparameters of BERT's such as next sentence pretraining objective. Roberta beats over the standard dataset like MNLI, QNLI, RTE, STS-B, and RACE as shown in table-5. We have selected Roberta architecture for computing the text features since it was demonstrated in [35] as Roberta is best choice to extract text feature embeddings for cyberbullying detection data.

Our reimplementation (with NSP loss):			
Model	***Segment-Pair***	***Sentence-Pair***	
Squad 1.1/2.0	90.4/74.7	88.7/76.2	
MNLI-m	84.0	82.9	
SST-2	92.9	82.9	
RACE	64.2	63.0	
Our reimplementation (without NSP loss):			
Full-Sentence	**Full-Sentence**	**Doc-Sentence**	
Squad 1.1/2.0	90.6/79.7	90.6/79.7	
MNLI-m	84.7	84.7	
SST-2	92.7	92.7	
RACE	65.6	65.6	
	Bertbase	Xlnet Base (K = 7)	Xlnet Base (K = 6)
Squad 1.1/2.0	88.5/76.3	-/81.3	-/81.0
MNLI-m	84.3	85.8	85.6
SST-2	92.8	92.7	93.4
RACE	64.3	66.1	66.1

Table 5 : RoBERTa performance over the benchmark datasets [29].

Machine Learning Classifier's:

In this section, we discussed a brief introduction about classifiers used for classification of bullying and non-bulling tweets. We used three different families of classifiers. 1) Regression family 2) Decision Tree family 3) Navie Bayes family.

Logistic Regression:

Logistic regression [30] is a powerful machine learning classifier for binary classification tasks. It used a logistic function called sigmoid function. The range of sigmoid functions is between 0 to 1. The equation-2 shows the sigmoid function. Fig-8 shows the logistic regression.

$$F(x) = \frac{1}{1+e^{\wedge}x} \quad (2)$$

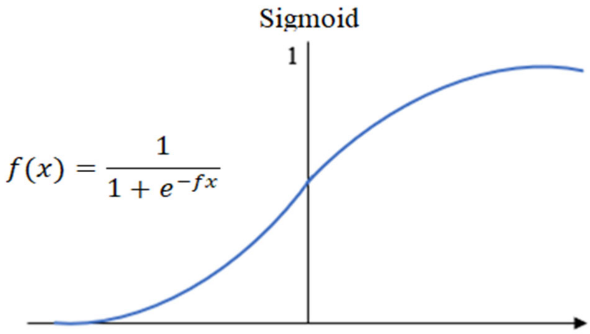

Fig 8: Logistic Regression.

Tree Classifiers (Decision Tree, XGBoost and LightGBM):
Decision tree classifier is the most well-known classifier algorithm and widely used for inductive learning method. The information entropy concept is used for labelled training data. Tianqi Chen and Carlos Guestrin[31] propose a novel sparsity-aware algorithm for sparse data weighted quantile sketch for tree learning. XGBoost (Extreme Gradient Boosting) prevents overfitting and handles missing values automatically. It computes faster due to parallel processing technique. LightGBM[32] is the best boosting algorithm for many classifications tasks. It performs better compared to CatBoost and XGBoost[33].

Bayes classifiers:

Navie Bayes [34] one of the most efficient inductive learning algorithms, especially for social media studies. NB classifiers works based on Bayes theorem as shown in equation 3. We used two classifiers from Navie Bayes family. 1) Bernoulli Navie Bayes and 2) Gaussian Naïve Bayes.

$$P(A/B) = \frac{P(\frac{B}{A})P(A)}{P(B)} \quad (3)$$

4 Experimental Results:

The following section presents the experimental methodology, metrics employed for evaluation and the respective outcomes. The experiments are implemented in Python using packages such as Numpy, Pandas, matplotlib, Scikit-Learn, LightGBM, and Tensorflow in Linux operating system. The methods were run on an Intel i7 8th Gen 12core CPU processor and Nvidia Max-Q 1070 32GB RAM.

Evaluation metrics

To evaluate the performance of cyberbullying detection with each of the selected image features and classifiers, the evaluation metrics such as precision, recall, f1-score are considered.

The precision is defined as the ratio of correct predictions as bullying to total number of predictions as bullying and is calculated as given in Eq. (4).

$$P = \frac{C}{C+B} \quad (4)$$

where C indicates the number of correct predictions as bullying, B indicates the number non-bullying classes that are incorrectly classified as bullying.

The recall is defined as the ratio of correct predictions as bullying to the total number of actual bullying classes and is given the Eq. (5)

$$R = \frac{C}{C+NB} \quad (5)$$

where NB refers to the total no of instances actual bullying classes wrongly predicted as non-bullying

f1-score is the weighted average of recall and precision and is computed as given in Eq. (6)

$$\text{f1-score} = 2\,\frac{P*R}{(P+R)} \quad (6)$$

We extracted text and image features from the respective pretrained architectures mentioned in section 3. Each network is involved in training with parameters from input to image to fully connected layer. Table6 shows the input image size, number of trainable parameters and non-trainable parameters and feature vector generated by each architecture. Finally, it generates 1- dimensional feature vector from the images for each architecture. Parallelly, another 1-dimentional vector is generated from RoBERTa architecture for text samples of 2100 captions for each image. We concatenated the features of 1-d image feature vector and text feature vector. The combined features supplied to machine learning classifiers to classify bullying or not.

Table 6: Parameters used in the Architectures.

Architecture	*Height*	*Width*	*Trainable parameters (in Millions)*	*Non-trainable parameters (in Millions)*	*Total parameters (in Millions)*	*Feature vector size*
Xception	*299*	*299*	*22.1M*	*.54M*	*22.6M*	*768*
Inception	*299*	*299*	*25.2M*	*.53M*	*25.7M*	*2048*
VGG19	*224*	*224*	*23.8M*	*.34M*	*24.1M*	*4096*
ResNet	*224*	*224*	*25.5M*	*.53M*	*26.3M*	*2048*
DenseNet	*224*	*224*	*20.0M*	*.22M*	*20.2M*	*1920*
MobileNet	*224*	*224*	*42.3M*	*.21M*	*42.5M*	*1024*
EfficientNet	*224*	*224*	*91.0M*	*.67M*	*91.7M*	*1408*

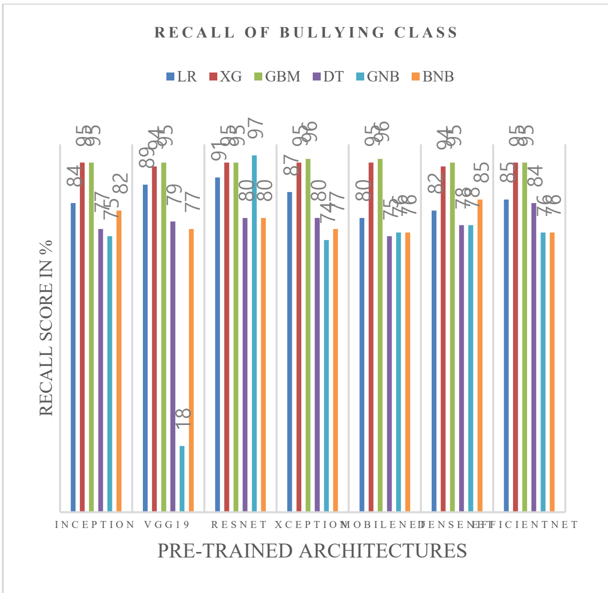

Fig. 9 a) Recall score of bullying class.

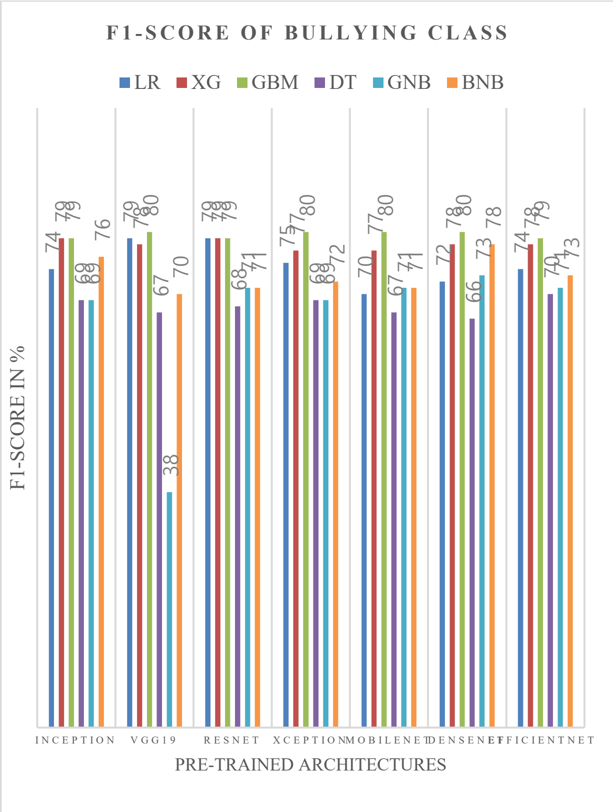

Fig. 9 b) F1-score of bullying class.

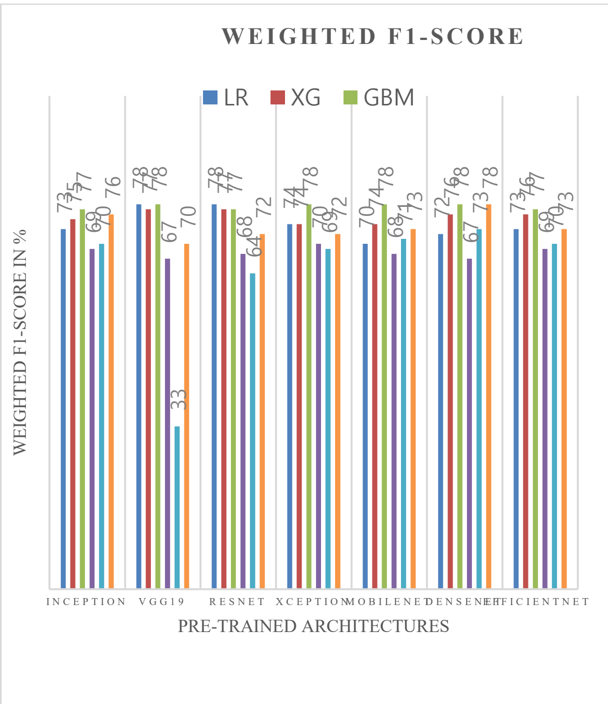

Fig. 9c) Weighted f1-score of bullying class.

ROC Curve Analysis:
ROC curve is one of the best classification metric for binary classification tasks. On x-axis, False Positive Rate(FPR) and y-axis True Positive Rate(TPR) is represented. The area covered by the curves indicates performance of the models. We also noticed that ROC(receiver operating characteristic) curve for Xception pretrained architecture. Fig 10 shows the ROC curve of six classifiers of Xception. LightGBM covered large area as compared to other classifiers, which clearly indicates, less number of mis-classifications with the combination of Xception architecture and LightGBM classifier.

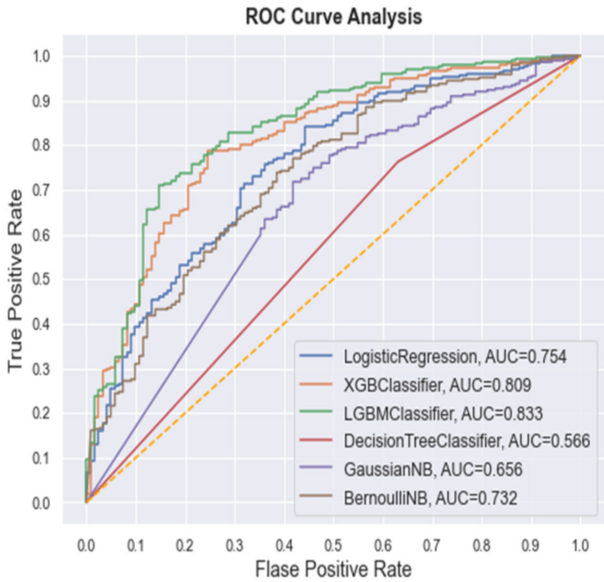

Fig. 10 ROC curve for Xception architecture.

5. Conclusion:

Data sharing in social medial networks not limited to text, people are interested to share multi-modal data. It's difficult to prevent cyberbullying content for multi-modal data. We conducted experiments with advanced deep learning architectures like Xception and Efficient Net and state of the art classifiers like LightGBM. We noticed that LightGBM and XGBoost gave better performance with weighted f1-score of 78% with Xception architecture as combination in all the cases. As part of the future work, we are exploring to work on cyberbullying detection using other multi-modal data such as audio and video.

References

[1] Peter, K.S et al, "Cyberbullying: Its nature and impact in secondary school pupils. J. Child Psychol," Psychiatry 49(4), 376–385, 2008.

[2] Van Cleemput K et al, "cyberbullying research in Flanders, Belgium and beyond: an overview of the findings," Technical report, Brussels, Belgium, 2013.

[3] US legal, available at https://defnitions.uslegal.com/c/cyber-bullying-2019

[4] Ybarra M (2010) Trends in technology-based sexual and non-sexual aggression over time and Linkages To nontechnology aggression. National Summit on Interpersonal Violence and Abuse Across the Lifespan: Forging a Shared Agenda

[5] The National Crime Prevention Cyberbullying. http://www.ncpc.org/cyberbullying. Accessed 10 March 2018.

[6] Yong Fang, Shao Shuai Yang, Bin Zhao, and Cheng Huang, "Cyberbullying Detection in Social Networks Using Bi-GRU with Self-Attention Mechanism," MDPI,2021.

[7] I.J Sheeba et al, "Unsupervised Hybrid approaches for Cyberbullying Detection in Instagram," International journal of computer applications, vol.174,2021.

[8] Kiriti kumari et al, "Multi-modal aggression identification using Convolutional Neural Network and Binary Particle Swarm Optimization," Future Generation Computer Systems, Elsevier, vol. 118, may-2021.

[9] Samantha Paul and Sriparna Saha, "CyberBERT: BERT for cyberbullying identification," Multimedia systems, Springer,2020.

[10] Jatin Karthik Tripathy et al, "ALBERT-based fine-tuning model for cyberbullying analysis," Multimedia systems, Springer,2020.

[11] Cheng, Lu et al. "XBully: Cyberbullying Detection within a Multi-Modal Context," Proceeding of the 12th ACM International Conference on Web Search and Data Mining,2019.

[12] Dani H., Li J., Liu H. "Sentiment Informed Cyberbullying Detection in Social Media," In Machine Learning and Knowledge Discovery in Databases. Lecture Notes in Computer Science, vol 10534. Springer, 2017.

[13] Devin Soni and Vivek Singh, "See No Evil, Hear No Evil: Audio-Visual-Textual Cyberbullying Detection", Proceedings of the ACM on Human-Computer Interaction, Vol. 2, 2018.

[14] Vimala Balakrishnan, Shahzaib Khan, Hamid R. Arabnia, "Improving Cyberbullying Detection using Twitter Users," Psychological Features and Machine Learning, Computers & Security, Elsevier-2019.

[15] Henderson Margono et.al, "Mining Indonesian Cyber Bullying Patterns in Social Networks,", Proceedings of the Thirty-Seventh Australasian Computer Science Conference - Volume 147, January 2014

[16] Dadvar M., Trieschnigg D., Ordelman R., de Jong F, "Improving Cyberbullying Detection with User Context.", Advances in Information Retrieval. ECIR 2013. Lecture Notes in Computer Science, vol 7814. Springer, Berlin, Heidelberg.

[17] Yang, Fan, et al., "Exploring Deep Multimodal Fusion of Text and Photo for Hate Speech Classification.", Proceedings of the Third Workshop on Abusive Language Online, Association for Computational Linguistics,2019.

[18] Kiriti Kumari and Jyothi Prakash Singh, "Identification of cyberbullying on multi-modal social

[19] Media posts using genetic algorithm," Wiley,2020.

[20] Simonyan, Karen and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014.

[21] Antioquia, Arren Matthew C. et al. "ZipNet: ZFNet-level Accuracy with 48× Fewer Parameters." 2018 IEEE Visual Communications and Image Processing (VCIP) (2018): 1-4.

[22] Krizhevsky A, Sutskever I, Hinton GE. "Imagenet classification with deep convolutional neural Networks," Commun ACM. 2017;60(6):84–90.

[23] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. "Going deeper with convolutions," In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 1–9.

[24] He, Kaiming et al. "Deep Residual Learning for Image Recognition." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 770-778.

[25] Chollet, François, "Xception: Deep Learning with Depthwise Separable Convolutions." 2017

IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017): 1800-1807.
[26] Andrew Howard et.al, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.",2017.
[27] Huang, Gao, et al, "Densely Connected Convolutional Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017): 2261-2269.
[28] Tan, Mingxing and Quoc V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," 2019.
[29] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," pp. 4171-4186, Oct-2018. DOI:10.18653/v1/N19-1423
[30] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov, " Roberta: A robustly optimized BERT pretraining approach", 2019.
[31] LeCessie S, Van Houwelingen JC, "Ridge estimators in logistic regression". J R Stat Soc Ser C (Appl Stat). 1992;41(1):191–201.
[32] Chen, Tianqi and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2016).
[33] Ke, Guolin, et al. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." NIPS (2017).
[34] Smith, P.K.; Mahdavi, J.; Carvalho, M.; Fisher, S.; Russell, S.; Tippett, N. "Cyberbullying: Its nature and impact in secondary school pupils," J. Child Psychol. Psychiatry 2008, 49, 376–385.
[35] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofe P, Weiss R, Dubourg V, et al. "Scikit-learn: machine learning in python," J Mach Learn Res. 2011;12:2825–30

Percherla Subbaraju received the B.tech. and M.Tech. degrees, from Andhra Univ. in 2006 and 2011, respectively. After working as a assistant professor (from 2010), in the Dept. of Information Technology, SRKREC Bhimavaram., and His research interest includes Text Processing, Machine and Deep Learning. He is a member of AICTE, ISTE, CSI, IEEE in India.

E. Ilavarasan received the post graduate degree MTech in Computer Science and Engineering from Pondicherry University, Puducherry, India, in 1997 and the PhD in Computer Science and Engineering from the same University, in 2008. He is currently working as a Professor in the Department of Computer Science and Engineering at Pondicherry Engineering College, Puducherry, India. He is instrumental in setting up the Microprocessor laboratory and the Embedded System Laboratory funded by MODROBS-AICTE. Presently he is supervising five research scholars. His research interests include parallel and distributed systems, operating systems security, web services computing and embedded systems. He has organized National and International conferences with faculty members working in the Pondicherry Engineering College. He has published more than 50 research papers in the international journals and conferences. He had more than 25 years of experience in teaching.