Indonesian Journal of Computer Science
ISSN 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam No. 1, Padang, Indonesia
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

The Performance Analysis of Graph Neural Network (GNN) and Convolutional Neural Network (CNN) Algorithms for Cyberbullying Detection in Twitter Comments

Muhammad Rizki Nurfiqri[1] , Fitriyani[2]
fiqrimrn@student.telkomuniversity.ac.id, fitriyani@telkomuniversity.ac.id
[1] Informatics Department, School Of Computing, Telkom University, Bandung, Indonesia
[2] Informatics Department, School Of Computing, Telkom University, Bandung, Indonesia

Article Information

Submitted : 23 Apr 2024
Reviewed : 17 May 2024
Accepted : 15 Jun 2024

Keywords

Cyberbullying, Convolutional Neural Network (CNN), Graph Neural Network (GNN), Twitter, Performance Comparison

Abstract

Cyberbullying incidents have surged due to the expansion of social media network and advancements in internet technology, presenting a substantial challenge in online communities. Previous research utilized Support Vector Machine (SVM) techniques and obtained an accuracy rate of 71.25%. However, given the dynamic nature of cyberbullying behaviors and the necessity for more robust detection methodologies, the topic remains challenging, this study investigates the application of Convolutional Neural Network (CNN) and Graph Neural Network (GNN) techniques in detecting cyberbullying on Twitter. We chose CNN and GNN due to the capacity of neural networks to capture intricate patterns in textual and network data. The results of the experiment show that the GNN method consistently outperforms CNN in terms of f1-score, accuracy, precision, and recall. The GNN method achieves an accuracy of 80.25%, surpassing CNN 68.43%, by employing 20 epochs. Then the optimization of GNN by implementing various numbers of epochs reaches a high accuracy of 92.78 % when using 200 epochs. This validates the effectiveness of GNN in detecting cyberbullying on Twitter.

A. Introduction

In Indonesia, the internet has spread widely throughout the country, marking significant progress in information technology. Now, many social media platforms have become an integral part of daily life for people, not just conventional mass media. There are plenty of options for enhancing social interactions online, including Facebook, Twitter, Instagram, and LinkedIn. The latest data from We Are Social shows that the number of internet users in Indonesia reached 212 million in January 2023, or about 77% of the population, indicating a widening digital inclusion [1]. This increasing connectivity not only enables more people to access information but also opens up new economic opportunities, connecting businesses with larger markets, and fostering innovation in certain fields. Moreover, this digital transformation brings challenges, such as cybersecurity issues and the need for digital literacy. To ensure that this progress continues and benefits all layers of society, these challenges must be addressed [2].

On the negative side of social media development is the emergence of cyberbullying. cyberbullying can take the form of intimidation, harassment, or threats conducted through digital media such as the internet, social media, or text messages [3]. In 2023, a study by Hendry aimed to gain a broad understanding of strategies that could be used to prevent and address cyberbullying. This effort acknowledges the importance of involving different perspectives in generating effective responses [4].

Cyberbullying on Twitter is particularly challenging because it can spread information quickly. Often, these actions are concealed in subtweets, the use of specific hashtags, or closed interactions such as Direct Message groups. A study published in Emerald Insight by Bharti, Yadav, Kumar, and Yadav (2022) found that Cyber harassment is on the rise on social media, particularly against teenagers, and negatively impacts their mental health. They looked into machine learning methods and deep learning to identify cyberbullying [5]. HubSpot's Blog (2021) also discusses the increasing issue of harassment on Twitter. One example is how Twitter uses machine learning algorithms to filter search results, so content from reported or deemed offensive accounts is not prominently displayed. This indicates that offensive content can still be seen on the platform, but efforts are made to reduce it [6].

Previous research has presented various methods for detecting cyberbullying on social media. Commonly used approaches include Natural Language Processing (NLP) and Machine Learning (ML) techniques to classify text based on characteristics such as aggressiveness, use of negative words, and user behavior patterns [7]. In the application of machine learning, traditional methods such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Decision Trees have been widely applied due to their ability to process diverse data, including data from social media like Twitter [8]. These methods not only identify signs of cyberbullying but also classify the severity of cyberbullying based on features generated from social media content [9].

In 2020, a study at the Sardar Patel Institute of Technology by Rahul Ramesh Dalvi, Sudhanshu Baliram Chavan, and Aparna Halbe used SVM and Naïve Bayes to detect cyberbullying on Twitter. They implemented their model in real-time using the Twitter API after collecting data from various sources such as Kaggle and GitHub.

The study showed that SVM performed better with an accuracy of 71.25%, but both methods had weaknesses in interpreting tweet contexts and sentiments, as well as biases in datasets, which could affect result generalization. For detecting cyberbullying, representative data samples and accurate interpretation are crucial, according to this study and previous research [10].

The application of neural network approaches such as Graph Neural Network (GNN) and Convolutional Neural Network (CNN) has become popular for detecting cyberbullying. GNNs are particularly useful for analyzing relationships among users on social platforms because they can analyze structured data like graphs [11]. Conversely, CNNs can identify signs in text indicating cyberbullying because of their ability to process images and text [12]. Both approaches work together to detect cyberbullying. This study considers the effectiveness of GNN and CNN algorithms in detecting cyberbullying on Twitter. The choice of GNN and CNN algorithms is based on their ability to handle complex data patterns, as demonstrated by previous research [12] and [11].

B. Research Method

The general overview of the system design in this research is represented in the form of a flowchart as follows.

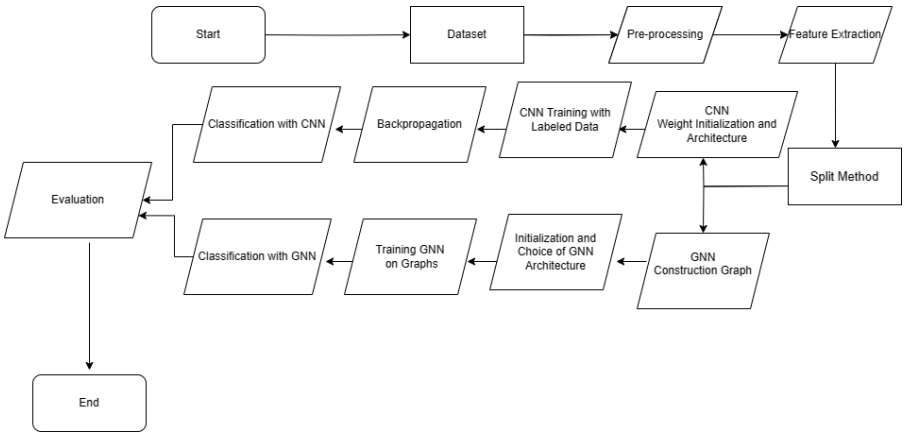

Figure 1. Flowchart Performance Comparison

From **Figure 1**. the first step in building a machine-learning model for cyberbullying detection is importing the necessary Dataset. After the dataset is processed, its features are extracted. The data is then distributed for training and testing. Labeled data is used to train CNN and GNN separately. Both are used for classifying new data after the training phase. To assess which model is more effective in detecting cyberbullying, both models are compared with metrics such as accuracy, precision, recall, and f1-score.

1. Dataset

The dataset of interest in this study is a comprehensive collection of data sourced from various platforms to facilitate the automatic detection of cyberbullying cases. Originating from several social media platforms including Kaggle, Twitter, Wikipedia Talk pages, and YouTube, this dataset includes diverse textual data. From the various platforms included in the dataset, especially data sourced from the Twitter platform used in this

research. Each entry in the dataset is carefully labeled to indicate the presence or absence of cyberbullying content, thus aiding in classification tasks. The content in this dataset includes various categories of cyberbullying such as hate speech, aggression, insults, and toxicity. The total collected data amounts to 17,803 tweets containing Bullying comments on Twitter comments, which have been collected and stored in CSV format. The CSV file likely contains columns containing information such as user IDs, comment text, and labels indicating whether the comment contains cyberbullying elements or not. Label 1 is for meaningful texts (tweets that fall within the existing categories) and label 0 is for meaningless texts (tweets not included in those categories) [13].

The next step is to perform Pre-processing after obtaining the dataset to be used. This process is used to address issues that arise during the data processing process, to improve data quality, and to ensure more accurate results from the process [14]. The Pre-processing process used in this research consists of five stages case folding, data cleaning, text normalization, stemming, and stopword removal. This process is depicted in the following flowchart diagram.

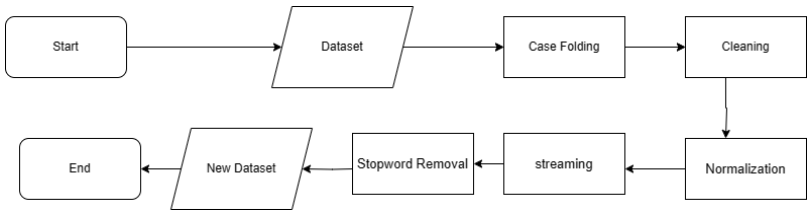

Figure 2. Flowchart Pre-processing

Figure 2 is a flow for pre-processing. after getting the next dataset case folding. Case Folding is the process of standardizing uppercase or lowercase letters, uppercase or lowercase letters are used to convert all characters to lowercase letters. Cleansing is the process of removing punctuation marks, retweet symbols, usernames, URLs, and emoticons from sentences in the dataset. Converting informal terms to formal terms is known as normalization. The alay_dictionary dictionary, which has been built in previous research, is used in this stage. This dictionary includes slang and typo words. The process of stripping or reducing words to their root or base word is known as Stemming. The PySastrawi library will be used for this research. Stopword Removal is the process of removing unimportant or irrelevant words for research purposes, thus lessening their impact on the classification process.

Feature extraction from raw data is conducted by algorithms, where features for cyberbullying detection could be a bag of words or TF-IDF. Bag of words represents text as a collection of words irrespective of sequence or context, while TF-IDF calculates word frequency in documents compared to the entire dataset. More advanced methods like Word2Vec or GloVe provide a deeper understanding of the semantic context of words used in cyberbullying.

2. CNN (Convolutional Neural Network)

Convolutional Neural Network (CNN) are essential tools for deep learning as they can solve difficult problems that cannot be addressed by traditional machine learning methods. By using layered structures, they identify important patterns and characteristics from input data such as images or text [15]. The CNN model is constructed using the TensorFlow.keras library, beginning with an embedding layer that converts tokenized input text into dense vector representations. This process facilitates the management of large input dimensions and generates richer text representations. TensorFlow.keras provides effective weight initialization by default for the embedding layer, ensuring a good initial distribution for the learning process. Below is the architecture of the model used

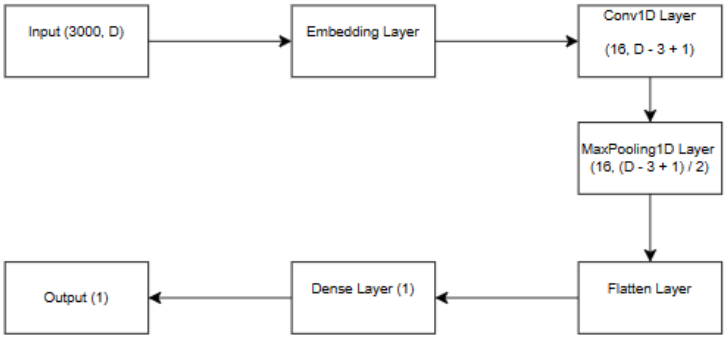

Figure 3. Flowchart CNN

From **Figure 3** show the model architecture includes

1. Embedding Layer Configured with input_dim as the maximum vocabulary size, output_dim as the embedding dimension, and input_length determined according to the maximum text length.
2. Conv1D Layer Uses 16 filters and kernel size 3 with ReLU activation for feature extraction.
3. MaxPooling1D Layer With pool size 2, reduces the output dimensions from the convolution layer and helps prevent overfitting.
4. Flatten Layer Converts the output from the previous layer into a one-dimensional vector.
5. Dense Layer (Output) One unit with sigmoid activation is used for binary classification, determining whether the input contains cyberbullying.

After that is the Model CNN stage-trained using the labeled dataset, consisting of text comments and cyberbullying labels, through 20 epochs with a batch size of 32, using binary_crossentropy as the loss function and adam optimizer. During the training phase, the backpropagation technique is used to calculate the gradient of the loss function concerning all weights in the model and update those weights to minimize the loss, allowing the model to learn from mistakes iteratively and improve classification accuracy. After the model is trained, it is used to classify new comments as cyberbullying or

not, and model evaluation is performed using metrics such as accuracy, precision, recall, and F1-score to assess its performance in detecting cyberbullying in the context of social media.

3. GNN (Graph Neural Network)

Machine learning method Graph Neural Network (GNN) is used to process data in graph structures. GNN has become more efficient and effective since its introduction. GNN has become important for various purposes, such as predicting protein interactions and creating recommendation systems [16]. The GNN model is built using the igraph library for node and edge creation and the PyTorch library for GCN model implementation. First, nodes and edges are constructed using functions provided by the igraph library, allowing for graph representation of data. Next, this graph representation is used as input for the GCN model built using the PyTorch library. The GCN model consists of multiple layers, each involving the process of aggregating information from the neighbors of each node in the graph. This process allows the model to gain a better understanding of the relationships between nodes in the graph and produce stronger feature representations for tasks related to graphs. The weights of the GCN model are initialized effectively, ensuring that the learning process starts from a good weight distribution. With this architecture, the GNN model can effectively handle structured data such as graphs and produce quality results in various graph analysis tasks. Below is the model architecture used

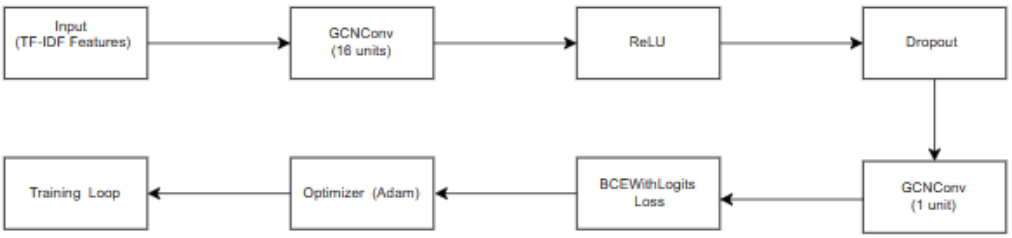

Figure 4. Flowchart GNN

Figure 4 The model architecture includes

1. Input Data TF-IDF features generated from Twitter comment text. TF-IDF is used as a numerical representation of text to inform the model about the importance of specific words in the context of the dataset as a whole.
2. GCNConv Layer The model uses two GCNConv layers which are the core of the Graph Convolutional Network (GCN) for convolution operations on graphs.
3. ReLU Activation Function After each GCNConv layer, the ReLU (Rectified Linear Unit) activation function is applied. This function adds non-linearity to the model, allowing it to learn more complex relationships in the data.
4. First Layer With 16 units, tasked with extracting and learning lower-level graph features.

5. Second Layer With 1 unit, aims to aggregate information learned by the first layer and prepare it for classification.
6. Dropout To avoid overfitting, the dropout technique is applied after the convolution operation. Dropout works by randomly removing some units from the layer during training, which helps make the model more robust to unseen data.
7. Binary Cross Entropy (BCE) Loss Function Used to calculate the difference between predictions and actual labels. In this context, BCE is suitable for binary classification tasks like cyberbullying detection.
8. Adam Optimizer is used to adjust network weights based on the gradient of the loss function. Adam is a popular optimizer because of its ability to combine the advantages of two other algorithms, AdaGrad and RMSProp, making it effective for various types of tasks.
9. Training Loop Involves iteration through the dataset, where at each iteration, the model performs a forward pass to calculate predictions, computes loss, performs backpropagation to adjust weights, and uses the optimizer to update model parameters.

Next, the stage for GNN is the construction stage, where data is processed in the form of a graph, with nodes representing entities and edges representing relationships between those entities, for example, in the case of cyberbullying, nodes represent text comments and edges represent '1' labels for comments detected as cyberbullying and '0' for those that do not. This process uses functions from the Igraph library. Next, selecting the appropriate Graph Neural Network (GNN) such as a Graph Convolutional Network (GCN) becomes crucial in the initialization stage, as GCN has been proven effective in handling structured data in graphs by modeling local relationships between nodes. The model training process involves continuous iterations, where the model updates its node representations by considering information from its neighbors to predict whether a comment includes cyberbullying or not. Backpropagation methods and the Adam optimizer are used to adjust the model weights based on loss calculations. Once the model is trained, it is applied to classify new comments to determine whether they contain cyberbullying or not. Model evaluation is performed using metrics such as accuracy, precision, recall, and F1-score to assess its performance in detecting cyberbullying in the context of social media.

4. Evaluation

The statistical methods used to thoroughly evaluate Graph Neural Network (GNN) and Convolutional Neural Network (CNN) in detecting cyberbullying. An analysis of variance (ANOVA) will be performed to determine if there is a statistically significant difference in performance to determine which model performs best under various dataset conditions [17]. The accuracy, precision, recall, and F1-score of both models will be analyzed with the same number of epochs between GNN and CNN themselves. Thus, a strong recommendation on the best cyberbullying detection algorithm can be provided based on empirical evidence. Various strong statistical techniques will be used to provide a comprehensive analysis of the

performance of GNN and CNN algorithms. First, the f1-score, accuracy, precision, and recall will be calculated using the following formulas

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

$$Precision = \frac{TP}{(TN + FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

$$F1 - Score = \frac{2TP}{2(TP + FP + FN)} \quad (4)$$

Additionally, the optimization experiments of GNN with epochs parameter variations (20, 50, 75, 100, 150, and 200) will be evaluated. Identical statistical techniques to those used in comparing GNN and CNN will be adopted. Analysis of variance (ANOVA) will be used to evaluate whether there is a significant difference in performance between various epochs parameters. Additionally, accuracy, precision, recall, and F1-score calculations will be performed for each epochs parameter using the same formulas used in comparing GNN and CNN. Those evaluation has a strong focus on statistical analysis, and model performance evaluation, which will provide valuable insights for cyberbullying detection on Twitter.

C. Result and Discussion

The result is present experimental results testing Convolutional Neural Network (CNN) and Graph Neural Network (GNN) with pre-determined parameters. An analysis of their performance in classification tasks.

1. CNN vs GNN Experiment Results

The following table shows the average results of three times running experiments of testing CNN and GNN by using 20 Epochs parameters.

Table 1. CNN VS GNN Experiment Result

CNN AND GNN		
Epochs 20		
INFORMATION	**CNN**	**GNN**
ACCURACY	68,43%	80.25 %
PRECISION	50,00%	88.03 %
RECALL	34,21%	43.71 %
F1-SCORE	40.63 %	58.41 %
Time	13.890 Second	1,13 Second

From these experimental results on **Table 1** , several analyses can be obtained

1. Overall Performance GNN shows superior performance compared to CNN in terms of accuracy, precision, recall, and f1-score, indicating GNN's ability to classify data.

2. Accuracy GNN has a higher accuracy rate than CNN in each experiment iteration, indicating GNN's ability to recognize data patterns better.
3. Precision and Recall GNN also shows higher precision and recall values compared to CNN. This indicates that GNN is more effective in identifying and classifying data samples.
4. Computational Time GNN has an advantage in terms of more efficient computational time. GNN requires much less time to train the model and perform inference compared to CNN.

The experimental results show that GNN significantly outperforms CNN in terms of accuracy, precision, recall, and f1-score. GNN's ability to consider the complex context and relationships between comments in the form of a graph provides an advantage in detecting cyberbullying. Meanwhile, CNN's limitations in handling structured data and its focus on local feature extraction lead to relatively lower performance in this context. These findings have significant practical implications, including the potential use of GNN in improving cyberbullying detection algorithms on social media platforms. Additionally, insights gained from this comparison can assist algorithm developers in selecting the appropriate approach for similar cyberbullying detection problems.

2. Optimization GNN Experiment Results

The following table shows the average results of three times running experiments of testing GNN itself with different epochs parameters

Table 2. Optimization GNN Experiment Results

OPTIMIZATION GNN EXPERIMENT RESULTS				
EPOCHS	**ACCURACY**	**PRECISION**	**RECALL**	**F1-SCORE**
20	80.25 %	88.03 %	43.71 %	58.39 %
50	85.90 %	88.85 %	63.55 %	74.09 %
75	87.74 %	91.19 %	67.93 %	77.86 %
100	89.04 %	92.95 %	70.83 %	80.40 %
150	91.04 %	95.17 %	75.60 %	84.26 %
200	92.78 %	96.51 %	80.15 %	87.57 %

From the conducted experiments in **Table 2**, several significant analyses can be revealed

1. Overall Performance The experiments show that the higher the number of epochs used, the higher the performance of the GNN model in terms of accuracy, precision, recall, and f1-score. These results illustrate the superior capability of GNN in classifying data better with increasing training iterations.
2. Accuracy Consistently, the GNN model demonstrates higher accuracy rates than the previous iterations in each trial. This indicates that increasing training iterations provides the model with better capabilities in recognizing data patterns and producing more accurate predictions.

3. Precision and Recall GNN also shows an increase in precision and recall values with the increasing number of epochs. This indicates the effectiveness of GNN in identifying and classifying data samples better, especially in minimizing potential prediction errors.
4. Computational Time Although GNN produces better performance in terms of accuracy and prediction quality, the execution time required to train the GNN model also increases with the increasing number of epochs. Therefore, there is a need to consider the trade-off between model performance and computational time when selecting the optimal number of epochs for training.
5. Thus, although GNN demonstrates superior performance in terms of accuracy, precision, recall, and f1-score, it is important to consider computational time aspects when evaluating the model's performance in different application contexts.

D. Conclusion

The result of this research is that Graph Neural Network (GNN) significantly outperforms Convolutional Neural Network (CNN) in detecting cyberbullying in Twitter comments. Statistical analysis confirms the superiority of GNN in accuracy, precision, recall, and F1-score, with Anova test results showing a significant difference ($p < 0.05$) between the two in classifying cyberbullying data. GNN consistently excels in all tested aspects, indicating significant performance differences.

The performance evaluation of experiment shows that GNN performs better than CNN, with the average experimental results showing GNN achieving higher values in all metrics. In the experiment result **Table 1** by using 20 epochs parameters, GNN has an accuracy of 80.25%, while CNN only reaches 68.43%. Similarly, with precision, recall, and F1-score, GNN consistently excels. This indicates the ability of GNN to consider the context and complex relationships between comments in graphical structures. However, it can also be observed that CNN is weaker compared to SVM, with an accuracy of 71.25%. However, compared to GNN, SVM is also weaker [10] .

Increasing the number of epochs in GNN leads to improved performance, as experimental results indicate increased accuracy, precision, recall, and F1-score with escalating training iterations. In the experiment result **Table 2** with 20 epochs, the accuracy stands at 80.25%, whereas with 200 epochs, it escalates to 92.78%. However, the escalation in computational time poses a consideration in selecting the optimal number of epochs. This study underscores the superiority of GNN in detecting cyberbullying compared to CNN, as well as the significance of adjusting training parameters such as the number of epochs to attain optimal performance in varying contexts.

E. References

[1] "The Changing World Of Digital in 2023," We Are Social . Accessed: May 24, 2024. [Online]. Available: https://wearesocial.com/id/blog/2023/01/the-changing-world-of-digital-in-2023-2/

[2] Monavia Ayu Rizaty, "Pengguna Internet di Indonesia Sentuh 212 Juta pada 2023," DataIndonesia. Accessed: Nov. 22, 2023. [Online]. Available: https://dataindonesia.id/internet/detail/pengguna-internet-di-indonesia-sentuh-212-juta-pada-2023

[3] E. Merdan, "The Ugly Face of the Digital World," 2022, pp. 489–505. doi: 10.4018/978-1-7998-9187-1.ch021.

[4] B. P. Hendry, L. ann M. Hellsten, L. J. McIntyre, and B. R. R. Smith, "Recommendations for cyberbullying prevention and intervention: A Western Canadian perspective from key stakeholders," *Front Psychol*, vol. 14, 2023, doi: 10.3389/fpsyg.2023.1067484.

[5] Shubham Bharti, Arun Kumar Yadav, Mohit Kumar, and Divakar Yadav, "Cyberbullying detection from tweets using deep learning," cyberbullying *detection from tweets using deep learning*, Jul. 2021, Accessed: Nov. 22, 2023. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/K-01-2021-0061/full/html

[6] Sophia Bernazzani, "How Twitter Is Fighting Harassment & cyberbullying ," Hubspor. Accessed: Nov. 23, 2023. [Online]. Available: https://blog.hubspot.com/marketing/twitter-harassment-cyberbullying

[7] J.-M. Xu, K.-S. Jun, X. Zhu, and A. Bellmore, "Learning from Bullying Traces in Social Media." Accessed: Apr. 28, 2024. [Online]. Available: https://aclanthology.org/N12-1084

[8] D. Sultan *et al.*, "A Review of Machine Learning Techniques in cyberbullying Detection," *Computers, Materials and Continua*, vol. 74, no. 3. Tech Science Press, pp. 2625–2640, 2023. doi: 10.32604/cmc.2023.033682.

[9] B. A. Talpur and D. O'Sullivan, "Cyberbullying severity detection: A machine learning approach," *PLoS One*, vol. 15, no. 10 October, Oct. 2020, doi: 10.1371/journal.pone.0240924.

[10] Vaigai College of Engineering and Institute of Electrical and Electronics Engineers, *Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020) : 13-15 May, 2020*.

[11] A. Bouliche and A. Rezoug, "Detection of cyberbullying in Arabic social media using dynamic graph neural network ⋆," 2022. [Online]. Available: http://ceur-ws.org

[12] A. Alhlou and A. Alam, "Bullying Tweets Detection using CNN-Attention," *International Journal on Cybernetics & Informatics*, vol. 12, no. 1, pp. 1–14, Jan. 2023, doi: 10.5121/ijci.2023.120106.

[13] SAURABH SHAHANE, "Cyberbullying Dataset," Kaggle.com. Accessed: Apr. 15, 2024. [Online]. Available: https://www.kaggle.com/datasets/saurabhshahane/cyberbullying-dataset

[14] Febiana Anistya and Erwin Budi Setiawan, "Hate Speech Detection on Twitter in Indonesia with Feature Expansion Using GloVe," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 6, pp. 1044–1051, Dec. 2021, doi: 10.29207/resti.v5i6.3521.

[15] J. M. Vaz and S. Balaji, "Convolutional neural networks (CNNs): concepts and applications in pharmacogenomics," *Mol Divers*, vol. 25, no. 3, pp. 1569–1584, Aug. 2021, doi: 10.1007/s11030-021-10225-3.

[16] X. Ma *et al.*, "A Comprehensive Survey on Graph Anomaly Detection with Deep Learning," *IEEE Trans Knowl Data Eng*, 2021, doi: 10.1109/TKDE.2021.3118815.

[17] Engineering Statistics Handbook, "One-way ANOVA overview." Accessed: Dec. 25, 2023. [Online]. Available: https://itl.nist.gov/div898/handbook/prc/section4/prc431.htm