wjert, 2024, Vol. 10 Issue 4, 221-229. Original Article ISSN 2454-695X

World Journal of Engineering Research and Technology
WJERT
www.wjert.org SJIF Impact Factor: 7.029

# CYBERGUARD: A COMPREHENSIVE CYBER BULLYING DETECTION AND SUPPORT SYSTEM

***[1]Dr. M. Hemalatha and [2]Prasanth S.**

[1]Associate Professor, PG & Research Department of Computer Science Sri Ramakrishna College of Arts & Science.

[2]UG Student, PG & Research Department of Computer Science Sri Ramakrishna College of Arts & Science.

**Article Received on 21/02/2024** **Article Revised on 11/03/2024** **Article Accepted on 01/04/2024**

***Corresponding Author**
**Dr. M. Hemalatha**
Associate Professor, PG & Research Department of Computer Science Sri Ramakrishna College of Arts & Science.

## ABSTRACT

In recent years, the rise of social media and online communication platforms has led to an increase in instances of cyber bullying, posing significant challenges to individuals' mental health and well-being. To combat this pervasive issue, we present a comprehensive Cyber bullying Detection and Reporting System with Integrated Mental Health Support. This paper utilizes techniques such as Natural Language Processing (NLP) and sentiment analysis to analyze the semantic and emotional content of user-generated text, allowing it to detect patterns indicative of cyber bullying behavior. Upon identifying potential instances of cyber bullying, the system provides users with intuitive reporting functionalities, enabling them to flag and report abusive or harmful content with ease. Furthermore, this paper goes beyond mere detection and reporting by integrating mental health support resources directly into the platform. Recognizing the detrimental impact that cyber bullying can have on individuals' mental well-being, we have incorporated links to relevant support services, help lines, and counseling resources. This holistic approach aims to provide immediate assistance and support to individuals who may be affected by cyber bullying, empowering them to seek help and access resources to address their emotional and psychological needs.

**KEYWORDS:** Cyber bulling, health care, Machine Learning, Natural Language Processing.

## INTRODUCTION

Cyber bullying is a serious problem that can have a negative impact on mental health. However, current solutions for detecting and reporting cyber bullying are often ineffective. They can be difficult to use, and they often don't provide victims with the support they need. Our project aims to develop a new Cyber bullying Detection and Reporting System with Integrated Mental Health Support. This system will use machine learning and natural language processing to identify instances of cyber bullying in online content. Once cyber bullying is detected, the system will provide victims with direct links to mental health support services. We believe that our system will be a valuable tool for victims of cyber bullying. It will make it easier for them to report cyber bullying and get the support they need. We hope that our system will help to reduce the negative impact of cyber bullying on mental health. Our goal is to empower individuals to report cyber bullying, access support resources, and foster a safer online environment.

In today's digital age, the widespread use of social media and online communication platforms has revolutionized the way people interact and connect with one another. While these platforms offer numerous benefits, they also present significant challenges, chief among them being the issue of cyber bullying. Cyber bullying, defined as the use of electronic communication to intimidate, harass, or threaten individuals, has emerged as a prevalent and pervasive problem in online spaces, posing serious risks to individuals' mental health and well-being.

Recognizing the urgent need to address this issue, our project aims to develop a comprehensive Cyber bullying Detection and Reporting System with Integrated Mental Health Support. This system is designed to leverage the capabilities of machine learning algorithms to identify and flag instances of cyber bullying in various forms of online content, ranging from text-based messages and comments to social media posts and forums. By harnessing the power of natural language processing (NLP) and sentiment analysis, the system can analyze the semantic and emotional content of user-generated text, allowing it to detect patterns indicative of cyber bullying behavior. By combining advanced technological solutions with a holistic approach to mental health support, our project seeks to create a safer and more supportive online environment where individuals can interact and express themselves freely without fear of harassment or intimidation. Through our efforts, we hope

to raise awareness, promote intervention, and empower individuals to seek help and support when faced with cyber bullying, ultimately fostering a culture of empathy, respect, and inclusivity in digital spaces.

## RELATED WORKS

Many paediatric clinics provide paper or online clinical intake forms that a patient completes prior to seeing the HCP. These clinical intake forms may include questions about medical history, family history, health behaviour, or experience. These screening forms may be useful in the case of bullying, as studies suggest it may be better to ask youth about their exposure to bullying and cyberbullying using a questionnaire rather than asking them directly.3 HCPs should review their clinical intake forms to ensure that questions related to bullying and cyberbullying are included to assess whether the patient has had experiences bullying others or being bullied by others. One common form used in the United States is the Guidelines for Adolescent Preventive Services form, which includes screening across a variety of health behaviour and experiences, including bullying.

Furthermore, it is typically recommended for HCPs to provide every adolescent patient individual time with the provider without their parents present to support discussion of topics that may be more private or stigmatising and to promote the adolescent patient's development.8 Evidence supports that many adolescents would prefer that their parents were not present when they discuss their experiences with bullying.When screening in person or following up on questionnaire answers, Lam et al.9 suggest that physicians routinely ask their patients 4 questions: 1) How often do you get bullied (or bully others)? 2) How long have you been bullied (or bullied others)? 3) Where are you bullied (or bully others)? and 4) How are you bullied (or bully others)? It is important to note during screening that most targets (and perpetrators) of cyberbullying are also bullied in traditional ways10–13; thus, screening for both types of experiences should be standard practice for HCPs. Studies support that screening can take place in a variety of clinical settings and does not need to be reserved for a well-child or acute visit related to bullying. Ranney et al.14 surveyed adolescents in an urban emergency department and found that they reported high levels of exposure to physical peer violence (46.5%), cyberbullying (46.7%), and community violence (58.9%). Results support youth's willingness to engage on these topics in urgent as well as nonfamiliar clinical settings. Thus, urgent care, school-based clinics, emergency rooms, and inpatient hospital stays are all appropriate clinical settings to address bullying. In addition to screening for

bullying experiences, HCPs should screen for health conditions known to be associated with bullying experiences. Depression and anxiety are among the most common health conditions associated with cyberbullying experiences.15–20 HCPs working with adolescents should be aware of the strong and independent association between cyberbullying and suicide17 and include screening questions for self-harm and suicide.

## PROPOSED SYSTEM

### 3.1 NAIVE BAYES CLASSIFICATION

The system used various data mining techniques within the Waikato Environment for Knowledge Analysis (WEKA) interface to predict the likelihood of survival for individuals with diabetes. In the classification process, the study focused on observing characteristics of diseases, patients, and diagnoses. Three methods were employed: the first involved the explorer interface and Naïve Bayes algorithm, known for learning statistical knowledge. The second used the Experimenter interface to conduct experiments on algorithms like Naïve Bayes. The third method utilized Knowledge Flow to assess the accuracy of Naïve Bayes on different datasets. Overall, these techniques helped analyze and predict the survivability of diabetes using different algorithms and interfaces in the WEKA tool.

### 3.2 FUZZY NEURAL NETWORK (FNN)

The performance evaluation of the Genetic Algorithm (GA)-based trained Recurrent Fuzzy Neural Network (RFNN) for heart disease diagnosis involves metrics like Root Mean Square Error (RMSE), sensitivity, specificity, precision, F-score, probability of misclassification error (PME), and accuracy. These performance metrics, quantified through equations here I will give two equations

- $\text{Sensivity} = \frac{TP}{TP+fn}$
- $\text{Sensivity} = \frac{TN}{TN+FP}$

Where the TP is True Positive, FP is False Positive and TN True Negative. That assesses the model's accuracy, ability to identify true positives and negatives, precision in positive predictions, overall balance between precision and sensitivity, error probability, and overall correctness. The goal is to achieve a highly accurate and balanced model, ensuring reliable heart disease diagnosis by correctly identifying patients with and without the condition.

### 3.3 DECISION TREE

The Decision Tree Algorithm, adept a addressing both regression and classification challenges in supervised learning, holds the advantage of analyzing computational and classification data. During the training process, decision rules derived from prior data guide the model in determining the value or class of the target attribute, akin to conventional decision tree algorithms. This algorithm structures a tree with decision nodes, where each inner node corresponds to two or more leaf nodes, and the root node, containing the most significant dataset information. By subdividing the entire dataset into subsets, the algorithm creates a tree structure with leaf nodes, internal nodes, and the root node, forming a comprehensive decision tree. As the tree expands, complexity increases, enhancing the model's accuracy in decision-making based on learned rules from the training data. In the context of predicting breast tumor diseases, a comparative analysis with the Random Forest algorithm reveals lower accuracy for the Decision Tree Algorithm. Consequently, it is evident that the Random Forest algorithm outperforms, establishing itself as the superior choice for accurate disease prediction.

### 3.4 LOGISTIC REGRESSION

Logistic regression is a type of regression analysis where the dependent variable "y" is binary, typically representing two outcomes that is 0 or 1. In this context, the dependent variable is termed "Decision," indicating whether an individual is at risk of experiencing cardiac arrest and should seek medical assistance. The model aims to identify the optimal curve for a vector-oriented variable "x," incorporating diverse parameters specific to that curve. The likelihood function is employed, subsequently translated into binary values (0, 1) for actual probability estimation post the training phase using a dataset. This methodology is advantageous as it allows for the inclusion of multiple explanatory variables, which may be dichotomous, ordinal, or continuous. Consequently, logistic regression provides a quantified measure of the strength of association, adjusting for other variables in the analysis.

### 3.5 RANDOM FOREST

A study assessed machine learning models for malignant tumor detection, training Random Forest, SVM, Decision Tree, MLP, and K-NN on an initial dataset. Random Forest excelled across metrics. Using a minimal dataset, Random Forest maintained exceptional performance with 99.30% accuracy and perfect recall. This streamlined approach enhances clinical efficiency and reduces computational resources. ROC curves illustrated Random Forest's

robustness, outperforming literature models. The dataset for predicting chronic kidney disease included 1718 instances, with oversampling for balanced classes. Preprocessing involved handling outliers, imputing missing values, and feature normalization. Feature selection utilized UFS and RFECV, with Random Forest, SVM, and Decision Tree for CKD stage prediction, aiming to enhance model performance.

### 3.6 CONVOLUTIONAL NEURAL NETWORK (CNN)

The CNN architecture has two main parts: one analyzes image features through feature extraction, and the other predicts the image category. It uses five layers-CNN, max-pooling, fully interconnected, activation, and dropout layers-enhanced with about 240 filters of size 5 × 5. The input includes pre-processed FMRI, PET, and CT scan images. An 18-layered CNN achieves high accuracy by passing through multiple filters, while a 3D CNN may be less accurate than a 26-layered one in detecting Alzheimer's disease. Multi-layered CNNs effectively identify damaged neurofibrils in the brain. CNNs are widely used in related works for various applications.

## RESULTS AND DISCUSSIONS

Total 2000 data are considered here for classification. It shows the equal distribution of bully and non-bully data from the training dataset. As per reported in literature.[2,3] equal distribution of tweets gives best classification results. If the training database distribution is unequal then it may lead to wrong classification. After equal distribution, the dataset is fed to five different classifiers.

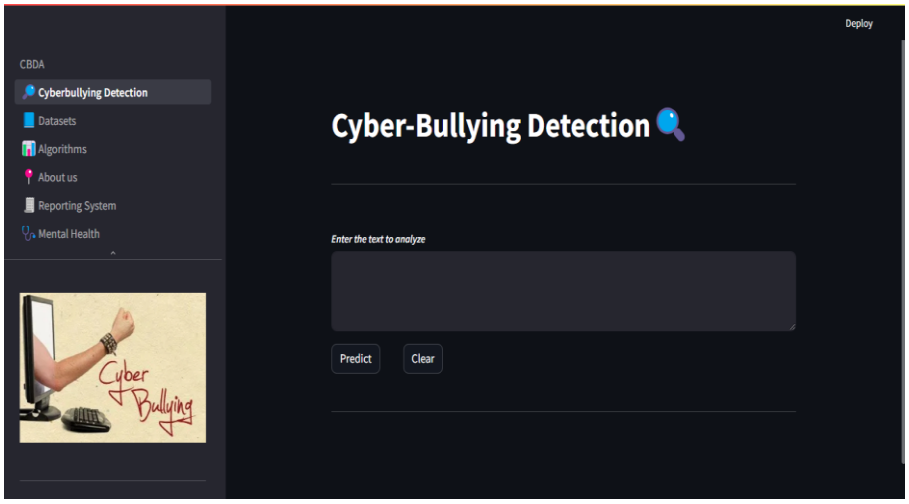


**Fig. 1: User Data.**

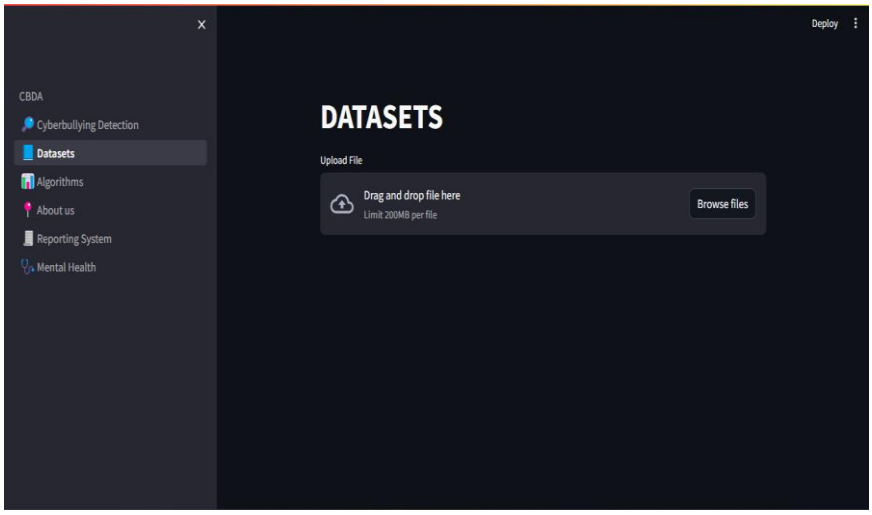


**Fig. 2: Selection of Dataset.**

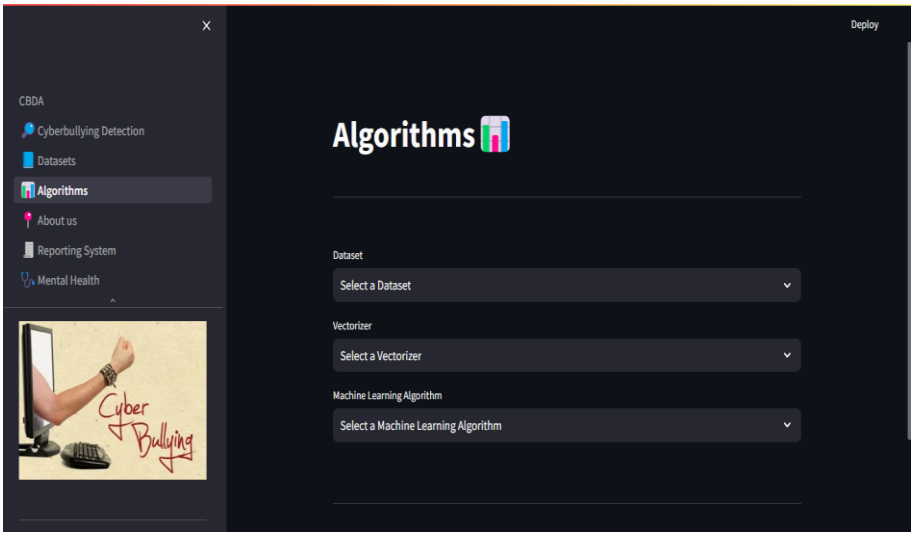


**Fig. 3: Selection of Algorithms.**


**Fig. 4: Reporting System.**

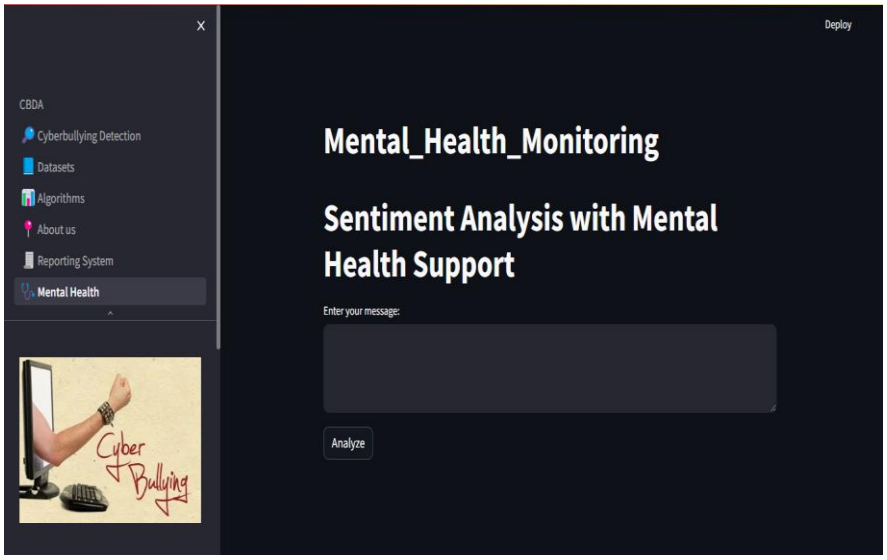


**Fig. 5: Mental Health support system.**


**Fig. 6: Classification syst.**

## CONCLUSION

The Cyber Bullying Detection and Reporting System represents a significant step forward in combating online harassment and abuse. Through the integration of Natural Language Processing (NLP) techniques and machine learning algorithms, coupled with real-time reporting functionalities, this system empowers users to identify and address instances of cyber bullying effectively. By leveraging NLP, the system can analyze text inputs from various sources, including social media platforms, messaging apps, and online forums, to detect potentially harmful content indicative of cyber bullying. The implementation of a Random Forest Classifier model further enhances the accuracy of detection, ensuring reliable results in real-world scenarios.

Moreover, the addition of a reporting system enables users to report instances of cyber bullying swiftly, prompting timely intervention and support from moderators or relevant authorities. This proactive approach not only helps prevent further harm to victims but also fosters a safer and more supportive online environment for all users. The system's seamless integration with Sup abase database technology ensures robust data management and scalability, accommodating growing user bases and evolving application requirements. Additionally, the user-friendly interface and accessibility features make it easy for individuals of all backgrounds to utilize the system effectively.

**REFERENCES**

1. Hinduja, S., & Patchin, J. W. Cyberbullying: An update and synthesis of the research, 2018.
2. Ptaszynski, M., & Dybala, P. Cyberbullying detection using machine learning techniques, 2019.
3. Fortuna, P., Nunes, S., & Rodrigues, F. Cyberbullying detection in social networks, 2018.
4. Chatzakou, D., Kourtellis, N., Blackburn, J., De Cristofaro, E., Stringhini, G., & Vakali, A. Mean birds: Detecting aggression and bullying on Twitter, 2017.
5. Sánchez, D., Díaz, D., Velasco, F., & Herrera, F. Genetic feature selection for improving classification in large datasets: A case of study with learning from online comments, 2012.
6. Boyd, D. It's complicated: The social lives of networked teens, 2014.
7. Williams, K. R., & Guerra, N. G. Prevalence and predictors of internet bullying, 2007.
8. Patchin, J. W., & Hinduja, S. Cyberbullying and self-esteem, 2015.
9. Spears, B., Taddeo, C. M., Daly, A. L., Stretton, A., & Karklins, L. T. Cyberbullying, help-seeking and mental health in young Australians: Implications for public health, 2015.
10. Kowalski, R. M., Limber, S. P., & Agatston, P. W. Cyberbullying: Bullying in the digital age, 2012.
11. A. Kumar, R. S. Umurzoqovich, N. D. Duong, P. Kanani, A. Kuppusamy, M. Praneesh, and M. N. Hieu, ‘‘An intrusion identification and prevention for cloud computing: From the perspective of deep learning,’’ Optik, 2022; 270: 170044.