Journal of Theoretical and Applied Information Technology
15st July 2026. Vol.104. No.13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A HYBRID TEXT-GNN FRAMEWORK FOR SENTIMENT ANALYSIS AND CYBERBULLYING DETECTION IN SOCIAL MEDIA TEXT

D. JAGANNATHAN[1], N. V. BALAJI[2]
[1,2]Department of Computer Science, Karpagam Academy of Higher Education,
Coimbatore, Tamil Nadu, India.
E-mail: [1]djagannathan1@gmail.com, [2]balaji.nv@kahedu.edu.in

ABSTRACT

The fast expansion of social media platforms has resulted in a higher amount of user-created text content, which requires automatic sentiment assessment and cyberbullying identification systems to ensure secure and positive online spaces. Traditional text classification methods fail to handle informal online language because they cannot understand its complex contextual relationships and semantic dependencies. The assessment of text contains three elements that create difficulties for accurate detection: noisy text elements, different linguistic patterns, and hidden abusive language. The existing system requires a framework that combines efficiency with intelligent capabilities to extract essential text-based data and select important features for accurate classification of sentiment and cyberbullying content across multiple datasets. This paper proposes a novel hybrid sentiment classification and cyberbullying detection model. The model integrates Multi-Resolution N-gram Embedding Fusion (MR-NEF)-based feature extraction, Dynamic Binary Swordfish Movement Optimization Algorithm (DBSMOA)-based feature selection, and Graph Neural Network with Text Graph Construction (Text-GNN)-based classification. For this research, IMDb, Yelp Polarity, and the Cyberbullying Classification datasets are applied as input datasets. Initially, the input data were preprocessed using NLP techniques like tokenization, stop word removal, and lemmatization. The developed MR-NEF + DBSMOA + Text-GNN model is assessed using three datasets individually. The model achieved its best performance with results like 98.64% accuracy, 98.57% precision, 98.52% recall, and 98.50% f1-score. Compared and validated with the current models and baseline variants, the developed model outperformed all the models with improved performance.

Keywords: *Cyberbullying Detection, Sentiment Classification, MR-NEF, DBSMOA, Text-GNN.*

1. INTRODUCTION

People now use online platforms and social networks as essential tools for their daily activities because these digital technologies enable them to communicate and share information while engaging in social interactions. The digital connectivity that links people together has resulted in the emergence of multiple types of cyber abuse, which now impact millions of people throughout the world. Cyber abuse, which includes cyberbullying, hate speech, shaming, impersonation, emotional abuse, doxing, and trolling, creates serious mental health risks because it results in psychosocial disorders that include depression and anxiety [1].

Cyberbullying affects both adolescents and children and exists as a recognized form of violent behavior because it involves deliberate, repeated attacks that people direct toward their fellow associates. The number of internet users has grown exponentially, together with the rising use of social media platforms, which has created an environment that allows cyberbullying to spread among online users. The process of detecting cyberbullying remains difficult because offensive language contains hidden linguistic details and contextual information. The traditional methods require researchers to create extensive datasets through human annotation, which demands both time and financial resources [2]. Cyberbullying represents a form of violent behavior in which people use digital technology to intimidate and stalk their victims. The internet enables cyberbullying to take place at any point during the day because it allows bullies to reach their targets from any location. Cyberbullying messages, which include images and videos, can be distributed anonymously to reach a wide audience [3].

Journal of Theoretical and Applied Information Technology
15st July 2026. Vol.104. No.13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

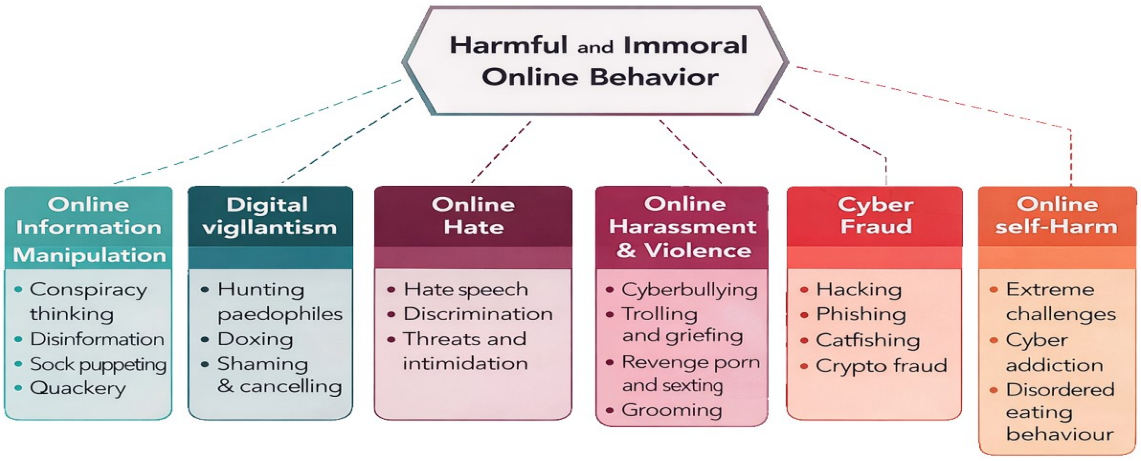

Figure 1: Online Behaviour Categories

Figure 1 demonstrates all primary categories of harmful and immoral online activities that frequently take place on digital platforms. The first level of harmful online activities is divided into six main categories, which include online information manipulation, digital vigilantism, online hate, online harassment and violence, cyber fraud, and online self-harm. Online information manipulation includes misleading practices such as conspiracy thinking, disinformation, sock puppeting, and quackery, which spread false narratives and manipulate public opinion. Digital vigilantism refers to actions such as hunting pedophiles, doxing, and public shaming, which people use to pursue online justice through exposing others. Online hate includes hate speech, together with discrimination and threats or intimidation that people direct against particular individuals or entire communities [4].

Online harassment and violence include behaviors such as cyberbullying, trolling, revenge, and grooming, which target victims through repeated abusive interactions. Cyber fraud represents malicious activities, which include hacking, phishing, and catfishing that hackers use to deceive users for financial or personal gain [5]. The category of online self-harm includes harmful digital trends, which include extreme challenges, cyber addiction, and disordered eating behaviors. The figure displays various types of online misconduct that create threats to digital safety and social well-being, which demonstrates the requirement for intelligent automated systems that can identify harmful textual content, stop cyberbullying, and abusive online communication [6].

The global problem of cyberbullying has increased due to rising social media usage. Therefore, it is critical to implement systems or models that can automatically detect and address bullying content in social media, as the consequences can result in social epidemics [7]. Researchers have created multiple predictive models that use machine learning (ML) and deep learning (DL) techniques to discover potential cyberbullies on social media platforms. Most research studies dedicated to automatic cyberbullying detection use opinion mining techniques, which analyze textual content. The identification of aggressive behavior has seen progress, but researchers still need to demonstrate its real-world effects. Social media platforms require immediate development of effective techniques that can identify and reduce cyberbullying activities among their users [8].

1.1. Problem Statement

Social media platforms and online communication platforms have expanded rapidly, resulting in increased user-generated textual content. The online space now shows widespread harmful behaviors, which include cyberbullying, hate speech, harassment, and fraudulent activities [9]. The researchers face detection challenges because these behaviors express themselves through complex linguistic patterns that contain sarcasm, implicit insults, and context-dependent expressions. Existing ML and DL approaches frequently fail to capture long-range semantic relationships, contextual dependencies, and subtle abusive patterns present in social media text. The presence of redundant and noisy textual features causes classification models to lose their ability to perform effectively. The digital communication space needs an intelligent framework that will extract essential

textual elements and choose the best features while precisely identifying harmful online content [10].

1.2. Novelty and Objectives

This research presents its novel contribution through the creation of an integrated framework that unites MR-NEF, DBSMOA, and Text-GNN to correctly identify online text sentiment and cyberbullying. The proposed MR-NEF system uses multiple n-gram resolutions to capture semantic patterns, which enable better contextual understanding than traditional methods that depend on single-level textual representations. The DBSMOA-based feature selection mechanism efficiently identifies the most informative features while eliminating redundant information, which results in better computational performance. The Text-GNN classifier creates a model that represents complex word-document relationships, which improves its ability to learn contextual dependencies and achieves better classification results across different datasets.

Research Objectives:

- To develop an intelligent framework for detecting harmful online content, including sentiment polarity and cyberbullying, from large-scale user-generated textual data.
- To design an MR-NEF mechanism that can capture contextual information from multiple n-gram levels and generate rich semantic representations of text.
- To implement the DBSMOA for effective feature selection by identifying the most informative features and eliminating redundant textual attributes.
- To employ a Text-GNN classifier that can model semantic relationships between words and documents for improved classification performance.
- To evaluate the effectiveness of the proposed MR-NEF + DBSMOA + Text-GNN framework using benchmark datasets such as IMDB, Yelp Polarity, and cyberbullying datasets.
- To compare the proposed model with existing ML and DL approaches in terms of accuracy, precision, recall, and F1-score to demonstrate its superiority.

The manuscript is structurally ordered, as Section 2 comprehensively analyzes the present models related to the study. Section 3 highlights the details of the implementation of the proposed framework using the MR-NEF + DBSMOA + Text-GNN model. Section 4 includes the experimentation findings of the developed model and a comparison with the discussed methodologies. Section 5 concludes the study with future recommendations.

2. LITERATURE REVIEW

This part analyzes the recent methodologies implemented for sentiment analysis and cyberbullying detection with various ML and DL techniques. Table 1 includes the comprehensive analysis of the models analyzed with advantages & drawbacks.

A hybrid DL model that combined Robustly Optimized Bidirectional Encoder Representation from Transformer and Bidirectional Long Short-Term Memory-based Attention framework (RoBERTa-BiLSTM) to detect cyberbullying in [11]. The system used the Optuna framework to optimize its performance. The system tested eight advanced transformer base models through an extensive cyberbullying dataset. The system achieved better F1-score results than the strong base models while it maintained essential false negative detection accuracy and required only moderate processing power.

The research in [12] created an ensemble model that combined three different classification models to detect multiple cyberbullying behavior patterns in a dataset of Twitter handles that contained cyberbullying tweets. Three multi-classification models: XGBoost, Decision Tree (DT), and Random Forest (RF) were evaluated by employing the TF-IDF feature-extraction method, and subsequently integrated these models into two ensemble strategies. Researchers investigated the most suitable machine learning methods for handling multiclass cyberbullying datasets and compared these methods to traditional machine learning approaches. The N-gram feature engineering alongside TF-IDF was utilized, attaining state-of-the-art results, especially with bigrams. The stacking classifier achieved 90.71% accuracy while the voting classifier obtained 90.44% accuracy.

The study in [13] proposed a method for assessing the extent of cyberbullying through the application of a DL algorithm and fuzzy logic. The research needed to detect cyberbullying comments through the analysis of Twitter data, which had been obtained from Kaggle. The researchers used Keras annotations as input data for a four-layer

LSTM network, which they used to classify the content. The researchers used fuzzy logic to evaluate the intensity of the Twitter comments. The experimental results showed that the system successfully detected cyberbullying with better detection results.

A detection model that used ML classifiers together with Natural Language Processing (NLP) methods to detect cyberbullying was developed in [14]. The research used Twitter data, which included postings and comments that researchers classified into five cyberbullying categories: religious bullying, age-related bullying, gender-based bullying, ethnic bullying, and non-cyberbullying. The model was intended to train ML classifiers subsequent to processing with NLP techniques. The researchers used five ML classifiers, which included RF, Support Vector Machine (SVM), Logistic Regression (LR), Naïve Bayes, and K-Nearest Neighbor. The RF classifier achieved the highest accuracy, which allowed it to surpass all other classification models.

A methodology for identifying cyberbullying with a DL algorithm was proposed in [15]. Three datasets from Facebook, Instagram, and Twitter were used to predict bullying incidents through LSTM analysis. The findings demonstrated the development of a successful model that detects cyberbullying better than previous detection approaches. The model achieved accuracy rates of approximately 91.26% for Facebook, 94.49% for Instagram, and 96.64% for Twitter datasets.

The research in [16] tested the effectiveness of DL and ML algorithms to recognize instances of cyberbullying and study the psychological impacts on at-risk individuals through their emotional and textual data. The researchers used three classifier algorithms, which included RF, DT, and LR; as well as DL architectures, including Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Multilayer Perceptron (MLP), and LSTM networks were employed. The researchers used TF-IDF for text vectorization, while they used TextBlob to perform sentiment analysis. The LSTM and RNN models reached their highest accuracy mark of 98 percent, which exceeded the performance of all other models. Among the conventional models, RF had superior performance, but LR demonstrated the least performance.

A cyberbullying detection model called FAEO-ECNN was designed in [17] for the identification and classification of cyberbullying on social networking platforms. The researchers used Fuzzy Adaptive Equilibrium Optimization (FAEO) clustering-based topic modeling together with Extended-CNN (ECNN) to create a system that improved cyberbullying detection accuracy. The system used unsupervised FAEO to extract hidden categories from data that researchers had already processed, and it automatically generated word clusters through its text analysis function. The cyberbullying classification system used ECNN together with the Rain Optimization (RO) method to detect cyberbullying content in written texts and online posts. The model performed successfully as a classification method because it achieved higher accuracy than other methods.

The research in [18] developed a cyberbullying detection model for Bengali and English text data using the Kaggle dataset, which was proposed. The model employed Transformer-Extra Large (XL) together with bi-directional RNN (BiGRU-BiLSTM) for its operation. The researchers used upsampling methods to solve the problem of class imbalance, whereas data augmentation techniques helped expand the size of their training dataset. The researchers used a pre-trained tokenizer to perform tokenization, which enabled them to obtain accurate semantic representations of the text. The Fusion Transformer-XL model consists of three components, which are Transformer-XL, BiLSTM, and bidirectional gated recurrent units (BiGRU). The researchers used a local interpretable model-agnostic explanation to show how the model made its decisions. The research results demonstrated that the model successfully detected cyberbullying behavior.

A cyberbullying detection system that used a heterogeneous graph-based approach was developed in [19] to analyze user interactions and textual content across different users and their session activities and comments. The method used a Heterogeneous Graph Neural Network (HGNN) to extract complex semantic relationships from a built heterogeneous graph. The model consists of three elements: a feature encoder, a heterogeneous graph encoder, and a classifier that jointly classifies social media sessions as non-bullying or bullying. The study results showed that every graph component improved the session-based cyberbullying detection system performance.

The detection framework, which combined sentiment analysis with ML methods, was introduced in [20] to enhance social media platforms for detecting cyberbullying. The framework used sentiment analysis to evaluate cyberbullying texts through their emotional and

linguistic features, which helped it identify main sentiment indicators that show harmful and safe interactions. The research identified the best text preprocessing methods, which improved text quality to support cyberbullying detection efforts. The measures provide ML models with better input data, which leads to significant improvements in their operational efficiency. The framework used several machine learning algorithms, which allowed the Extra Tree classifier to produce the best results.

The study in [21] conducted a complete research study that investigated how emotional attributes, word embeddings, and Federated Learning combine with LSTM, CNN, and Deep Neural Network (DNN), and BERT models. The research team tested different neural network designs and their associated hyperparameters to determine which combination delivered the best results. The results showed better abilities to find and handle cyberbullying incidents. The BERT model consistently outperformed all other deep learning models, which included LSTM, DNN, and CNN, in its ability to detect cyberbullying.

A hybrid feature engineering methodology that integrated word weighting with TF-IDF and the Latent Semantic Analysis (LSA) method was employed in [22]. The methodology reduced feature dimensionality and encapsulated the semantic essence of the data using SVM to distinguish between non-bullying and bullying remarks. The findings indicated that the methodology, which applied feature engineering to the LSA matrix derived from a specific dataset, achieved an accuracy rate of 97%. The model enhancement allowed for better text classification between non-bullying and bullying because LSA captured the unique features of each category.

The research presented in [23] introduced a new method that uses Neutrosophic Logic to improve cyberbullying detection through its application to Multi-Layer Perceptron (MLP) systems. The model enhanced cyberbullying detection through its solution of two primary challenges, which included the uncertainty problem and the non-specific boundary problem that exists between different types of cyberbullying. The system employs Neutrosophic Logic to resolve three classification problems, which include handling undefined cases and dealing with different levels of certainty and classification outcomes. The model results demonstrated that the system achieved better performance through the implementation of Neutrosophic Logic, which enhanced its capacity to identify different types of cyberbullying.

The study in [24] examined the utilization of CNN and Graph Neural Network (GNN) methodologies for the detection of cyberbullying on Twitter. The research selected GNN and CNN because of the ability of neural networks to discern complex patterns in network and textual data. The experimental results showed that the GNN method performed better than the CNN approach throughout all tests. The GNN optimization with the implementation of varying epoch counts achieved elevated accuracy. GNN proved successful in detecting cyberbullying on Twitter, according to this research.

A Deep Contrastive Self-Supervised Learning (DCSSL) model that incorporated a Natural Language Inference (NLI) dataset, a refined sentence encoder, and data augmentation was proposed in [25] to improve comprehension of the offensiveness and intricate semantics of cyberbullying. The DCSSL model successfully achieved two objectives because it extracted contextual links and multiple-meaning elements from cyberbullying instances, which human data annotators could not effectively identify. The model demonstrated a notable enhancement in cyberbullying datasets, which showed its usefulness for situations where human annotation processes were not possible.

Table 1: Critical Analysis of Reviewed Current Models.

Ref	Models	Application	Advantages	Limitations
[11]	RoBERTa + BiLSTM with Attention (Optuna optimized)	Cyberbullying detection	Improved F1-score; reduced false negatives; optimized hyperparameters.	Computationally intensive; depends on a pretrained transformer.
[12]	XGBoost, DT, RF with TF-IDF + Ensemble (Stacking & Voting)	Multiclass cyberbullying tweets	Strong ensemble performance; effective N-gram + TF-IDF features.	Limited deep contextual understanding; moderate accuracy (~90%).
[13]	LSTM + Fuzzy Logic	Severity assessment of cyberbullying (Twitter)	Handles uncertainty using fuzzy logic; improved detection.	Complexity in fuzzy rule design; limited scalability.

[14]	ML classifiers (RF, SVM, LR, NB, KNN) with NLP preprocessing	Multi-category cyberbullying (Twitter)	RF achieved superior accuracy; simple implementation.	Limited deep semantic modelling; feature engineering dependent.
[15]	LSTM-based DL model	Cyberbullying detection (Facebook, Instagram, Twitter)	High accuracy across platforms; robust sequential modelling.	Lacks advanced contextual transformers; moderate generalization.
[16]	ML (RF, DT, LR) + DL (RNN, CNN, MLP, LSTM) with TF-IDF & TextBlob	Cyberbullying & psychological impact analysis	LSTM/RNN achieved 98% accuracy; comparative evaluation.	TF-IDF limits semantic richness; high computational cost.
[17]	FAEO clustering + ECNN with Rain Optimization	Social media cyberbullying detection	Combines clustering and optimization; improved accuracy.	High algorithmic complexity; optimization overhead.
[18]	Transformer-XL + BiGRU + BiLSTM (Fusion Model)	Bengali cyberbullying detection	Handles class imbalance; interpretable via LIME; strong semantic capture.	Language-specific; complex architecture.
[19]	HGNN	Session-based cyberbullying detection	Captures high-order semantic interactions; graph-based insight.	Graph construction complexity; scalability concerns.
[20]	Sentiment Analysis + ML (Extra Tree best)	Social media cyberbullying detection	Improved performance via sentiment integration.	Relies heavily on preprocessing quality; limited deep contextual learning.
[21]	LSTM, CNN, DNN, BERT with Federated Learning	Cyberbullying detection	BERT outperformed others; privacy-preserving FL.	Federated setup complexity; communication overhead.
[22]	TF-IDF + LSA + SVM	Binary cyberbullying classification	97% accuracy; effective dimensionality reduction.	Limited contextual depth; class-specific bias.
[23]	Neutrosophic Logic + MLP	Fine-grained cyberbullying classification	Handles ambiguity and uncertainty effectively.	Increased model complexity; interpretability challenges.
[24]	CNN vs. GNN comparison	Twitter cyberbullying detection	GNN outperformed CNN; strong relational modelling.	High training cost; graph dependency.
[25]	DCSSL + NLI	Context-aware cyberbullying detection	Handles annotation scarcity; strong semantic representation.	Requires large pretraining resources; complex training pipeline.

2.1. Research Gap Analysis

There are still some issues that have not been fully solved regarding cyberbullying detection and sentiment analysis through ML, DL, transformer-based models, and graph-based models. The traditional ML models are primarily dependent on manually designed features and TF-IDF representation that cannot capture the semantic and contextual relationships of social media texts. Despite the advancements provided by DL and transformers in improving context awareness, most methods still lack efficiency in terms of computation, redundancy in features, and the inability to deal with implicit abuse, sarcasm, and long-range semantic dependency. Moreover, the various graph-based techniques that have been developed have concentrated only on relational learning while ignoring efficient feature optimization and multi-level semantic representation. The existing research studies lack focus on the integration of contextual embedding generation, feature optimization, and graph-based semantic learning.

To overcome these issues, this research proposes MR-NEF + DBSMOA + Text-GNN, an integrated architecture that involves the use of multi-resolution n-gram embedding fusion for obtaining contextual information, dynamic optimization-based feature selection for removing any redundant textual features, and classification using graph neural networks for capturing the complex semantic association between words and documents.

3. PROPOSED RESEARCH MODEL

The proposed system for detecting cyberbullying and classifying sentiments operates through a three-part process. This process combines high-level feature extraction, targeted feature selection, and graph-based classification methods to achieve better detection results and a deeper understanding of context. As depicted in Figure 2, the system begins its operation by using three benchmark datasets, which include IMDB, Yelp Polarity, and the Cyberbullying Classification dataset as sources of textual reviews and social media comments. The preprocessing stage of these raw textual datasets removes all noise elements, which include punctuation marks, special symbols, and irrelevant tokens, through standard Natural Language Processing (NLP) techniques. The techniques include tokenization, stop-word removal, and stemming. The process creates clean textual data that has a consistent structure for further analytical procedures.

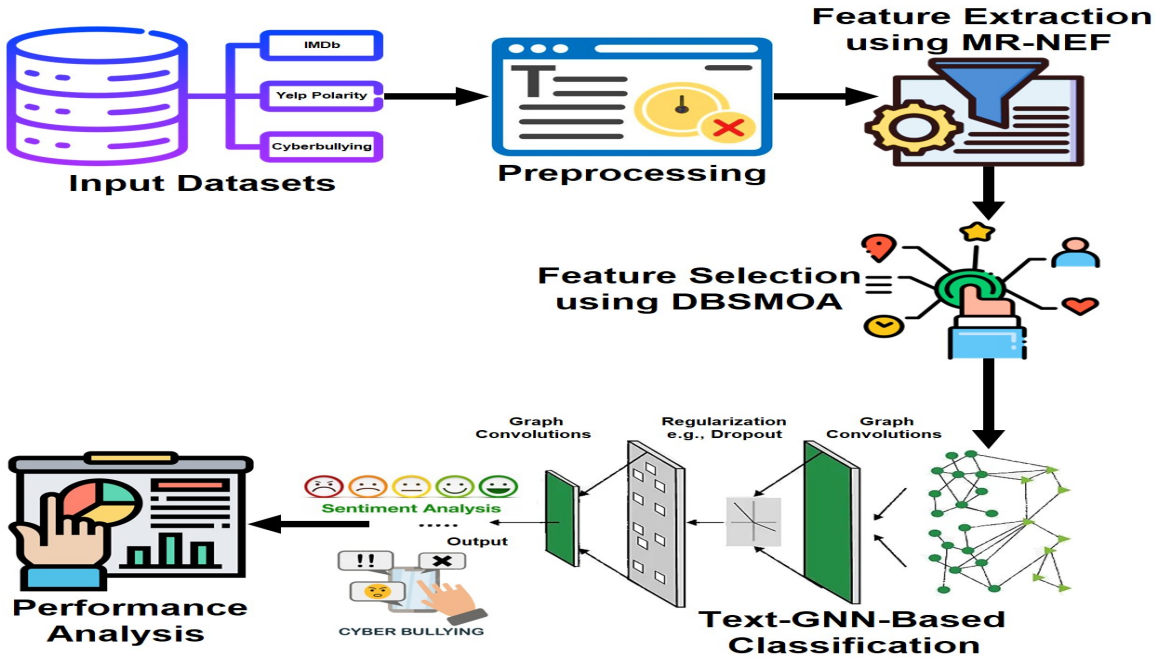

Figure 2: Workflow Pipeline of the Proposed Framework

The Multi-Resolution N-gram Embedding Fusion (MR-NEF) method performs feature extraction after the preprocessing stage by using n-gram levels to represent text through multiple n-gram levels, which include unigram, bigram, and trigram. The system creates a complete feature representation through the combination of different multidimensional embeddings, which maintains the semantic links and contextual relationships found in the textual content. The Dynamic Binary Swordfish Movement Optimization Algorithm (DBSMOA) is used for feature selection because extracted features tend to have high dimensionality, which includes both redundant and unnecessary data. The optimization algorithm uses swordfish hunting behavior to search for essential features, which results in reduced feature redundancy, better computational performance, and improved ability to distinguish different classes in the chosen features.

The optimized features create a text graph which uses words and text components as nodes, while their co-occurrence relationships form the edges of the graph. The Text-GNN model uses a graph representation to learn complex structural and semantic relationships between textual elements through its text graph construction component. The model uses graph convolution layers together with dropout regularization to propagate information across adjacent nodes, which helps it learn advanced dependencies in the textual graph. The learned graph representations proceed to the classification layer, which performs sentiment analysis and cyberbullying detection by assigning text to its correct sentiment or bullying categories. The overall system performance is then evaluated through performance analysis using standard evaluation metrics, which demonstrate how effectively the MR-NEF, DBSMOA, and Text-GNN framework identifies cyberbullying content and sentiment patterns across multiple datasets.

3.1 Dataset Description

The proposed MR-NEF-DBSMOA with the Text-GNN framework's performance and efficiency were evaluated using three widely used benchmark datasets that represent both sentiment analysis and cyberbullying detection tasks. The datasets provide multiple linguistic structures together with different text lengths and various contextual elements, which allow complete testing of the proposed model through different real-time textual cases.

Internet Movie Database (IMDb) Dataset: The IMDb dataset serves as a recognized benchmark dataset for sentiment analysis research. The dataset includes textual reviews for about 1000 movies and television shows, which have additional metadata that includes poster links, release year, runtime, certification, and genre details. This research uses only the review text to determine sentiment polarity. The reviews contain expressive opinions, narrative descriptions, and varied sentence structures, which makes the dataset suitable for evaluating the capability of the proposed framework to capture contextual semantics and long-range linguistic dependencies. The reviews contain diverse vocabulary and opinion-oriented language, which creates a challenging situation for performing sentiment classification tasks [26].

Yelp Polarity Review Dataset: This dataset serves as a standard dataset that is used to conduct binary sentiment classification experiments. The dataset includes customer reviews that were gathered from the Yelp platform, which have been divided into two sentiment categories that represent positive and negative customer assessments. The dataset assigns negative reviews to class 1, while positive reviews receive assignment to class 2. The reviews demonstrate informal language usage, subjective opinion expression, and multiple lengths of user feedback content, which simulate actual user feedback situations. The dataset features these attributes, which enable testing the proposed model's capacity to assess sentiment polarity, recognize subtle emotion signals, and accurately classify user sentiments from text reviews [27].

Cyberbullying Classification Dataset: This dataset is applied to identify harmful online activities that occur on social media networks. The dataset includes about 47000 tweets, which are collected from online sources and classified into six separate categories that include age-based cyberbullying, ethnicity-based cyberbullying, gender-based cyberbullying, religion-based cyberbullying, other forms of cyberbullying, and non-cyberbullying content. The distribution across classes maintains balance because each category includes approximately 8000 text samples. The dataset includes informal social media language, abbreviations, sarcasm, and hidden offensive expressions, which introduces significant complexity in the detection process. The dataset functions as an effective benchmark because its characteristics enable the proposed framework to demonstrate its ability to detect and classify various cyberbullying behaviors in multi-class classification scenarios [28].

The datasets applied in this research include textual reviews and social media posts that require preprocessing before being input into the proposed model. The raw textual data contains elements that include punctuation, stop words, hyperlinks, usernames, and other noisy components that will disrupt the classifier's ability to learn. This research applied standard text preprocessing steps, which included lowercasing, stop word removal, special character elimination, and token normalization to convert raw text into a structured format.

Table 2: Sample Reviews from Datasets.

Before Preprocessing	After Preprocessing
"I absolutely loved this movie! The storyline was amazing and the actors did a fantastic job."	absolutely loved movie storyline amazing actors fantastic job
"Worst service ever!!! The staff were rude and the food was cold."	worst service staff rude food cold
"@user123 Stop spreading hate, this kind of behavior is unacceptable."	stop spreading hate behavior unacceptable
"The product arrived late but the quality is pretty good overall."	product arrived late quality pretty good overall
"You people are so ignorant and offensive, learn to respect others."	people ignorant offensive learn respect others
"I watched this film last night and it was surprisingly entertaining."	watched film last night surprisingly entertaining

Table 2 shows some examples from the datasets, which include textual reviews that have not been processed and their subsequent processed state. The processed data keeps only essential tokens, which help to identify sentiment and cyberbullying while discarding unnecessary linguistic elements.

3.2 Data Preprocessing

The online platform textual data and review datasets contain unstructured, noisy information, which includes punctuation marks, hyperlinks,

emojis, user mentions, and stop words. The elements introduce redundant or irrelevant features that do not help algorithms perform at their best. The raw text required transformation through multiple preprocessing steps to create a clean, structured text system that enables sentiment classification and cyberbullying detection. Let the dataset be represented as follows.

$$D = \{(t_i, y_i)\}_{i=1}^{N} \quad (1)$$

The sentence structure of the sentence shows that t_i represents the ith raw text instance, while y_i shows the associated class label of the text, which includes positive/negative sentiments, multi-class sentiment, and cyberbullying category classification. N represents the total dataset samples. The preprocessing stage converts every raw text t_i into a normalized representation t_i' through many sequential operations.

$$t_i' = P(t_i) \quad (2)$$

The variable $P(\cdot)$ shows the preprocessing function, which includes multiple transformation steps.

Text Normalization: The initial procedure requires transforming every text character into lowercase, which will create a consistent standard for the entire text collection. The process decreases vocabulary duplication, which occurs through different case usages of words.

$$t_i^{(1)} = lower(t_i) \quad (3)$$

The lowercase form of the original text is shown by $t_i^{(1)}$.

Noise Removal: This research eliminated special characters and punctuation marks, hyperlinks, hashtags, and user mentions because these elements do not provide useful semantic information for the sentiment classification process. The operation, as given in its formal description, is given as follows:

$$t_i^{(2)} = f_{clean}(t_i^{(1)}) \quad (4)$$

Here, $f_{clean}(\cdot)$ removes every non-alphabetic character and every unnecessary token from the text.

Tokenization: The process of text cleaning leads to the creation of separate word units, which exist as individual tokens. The process of tokenization creates a token list from the original text sequence, which natural language processing models use for their operations.

$$T_i = \{w_{i1}, w_{i2}, \ldots, w_{im}\} \quad (5)$$

The tokenized text of T_i shows all its tokens while w_{ij} represents the jth word of the ith document.

Stop Word Removal: This research eliminated the common words which include "the", "is", "and", and "of" because these terms lack the essential meaning needed for their classification work. Let S indicates the predefined stop-word set. The filtered token set was given by:

$$T_i' = \{w_{ij} \in T_i | w_{ij} \notin S\} \quad (6)$$

The filtered token set contains T_i' which shows the token set after stop word removal.

Lemmatization: The text representation system underwent vocabulary reduction through lemmatization, which normalized all words into their base dictionary forms. Lemmatization converts each word into its base or dictionary form while considering its linguistic context. This step helps to preserve the semantic meaning of words while reducing different morphological variations. The lemmatization process is defined as follows:

$$w_{ij}' = f_{lemma}(w_{ij}, pos_{ij}) \quad (7)$$

The original token is represented by w_{ij} while pos_{ij} shows its assigned part-of-speech (POS) tag and w_{ij}' indicates the lemma form which results from the token [29].

Final Preprocessed Representation: The final processed text representation after all the preprocessing operations is represented as follows.

$$t_i' = \{w_{i1}', w_{i2}', \ldots, w_{ik}'\} \quad (8)$$

Here, $k \leq m$ due to the removal of noise and stop words during preprocessing. The cleaned tokens, which were generated from the cleaning process, serve as the input for both feature extraction and model training processes. The preprocessing pipeline of this system eliminates most textual data noise while enhancing machine learning and deep learning model performance through its emphasis on important linguistic elements.

3.3 MR-NEF-Based Feature Extraction

The MR-NEF method transforms preprocessed text tokens into useful numerical representations after the data cleaning process. The feature extraction method extracts semantic information through its use of multiple n-gram structure representations, which establish different contextual levels. Single-level representations cannot achieve success because they fail to detect both local word

interactions and larger context patterns. The MR-NEF framework builds its multi-resolution feature space through unigram, bigram, and trigram systems, which maintain the lexical, syntactic, and contextual meaning of input text. Let the preprocessed textual document be represented as follows:

$$t_i' = \{w_{i1}, w_{i2}, \ldots, w_{ik}\} \quad (9)$$

Here, t_i' was the ith processed document, which contains k tokens. The variable w_{ij} indicates the jth token in the sequence. The MR-NEF method creates multiple n-gram representations from the token sequence.

Unigram Representation: The unigram model considers individual words as independent features that operate separately. The system extracts essential lexical data that exists in the document. The document t_i' contains its unigram set, which exists as follows:

$$U_i = \{w_{i1}, w_{i2}, \ldots, w_{ik}\} \quad (10)$$

Every unigram token was mapped into a representation of a dense vector using an embedding function $E(\cdot)$:

$$u_{ij} = E(w_{ij}) \quad (11)$$

Here, $u_{ij} \in \mathbb{R}^d$ indicates the embedding vector of dimension d related to token wij.

Bigram Representation: The bigram model determines the relations between two following tokens, which allows the model to understand local context relationships present in the text. The bigram set was constructed as follows:

$$B_i = \{(w_{i1}, w_{i2}), (w_{i2}, w_{i3}), \ldots, (w_{i(k-1)}, w_{ik})\} \quad (12)$$

Every bigram pair was embedded into a vector representation using the embedding function:

$$b_{ij} = E(w_{ij}, w_{i(j+1)}) \quad (13)$$

Here, the variable $b_{ij} \in \mathbb{R}^d$ denotes the bigram embedding vector.

Trigram Representation: The generation of trigram features enables the system to understand advanced word linking patterns that extend beyond basic word relationships. Trigrams consider three consecutive tokens and present a broader contextual representation compared to bigram and unigram features.

$$T_i = \begin{Bmatrix} (w_{i1}, w_{i2}, w_{i3}), (w_{i2}, w_{i3}, w_{i4}), \ldots, \\ (w_{i(k-2)}, w_{i(k-1)}, w_{ik}) \end{Bmatrix} \quad (14)$$

Every trigram sequence was transformed into an embedding representation:

$$t_{ij} = E(w_{ij}, w_{i(j+1)}, w_{i(j+2)}) \quad (15)$$

Here, the variable $t_{ij} \in \mathbb{R}^d$ indicates the trigram embedding vectors.

Multi-Resolution Embedding Fusion: The extracted unigram, bigram, and trigram embeddings show semantic information through their different levels of contextual analysis. The fusion operation creates a single feature vector that combines all these representations. The aggregated embedding vectors for unigram, bigram, and trigram representations are represented by $F_i^{(u)}$, $F_i^{(b)}$, and $F_i^{(t)}$, respectively.

The process of obtaining the fused multi-resolution feature vector is represented as follows:

$$F_i = \alpha F_i^{(u)} + \beta F_i^{(b)} + \gamma F_i^{(t)} \quad (16)$$

The weighting coefficients α, β, and γ determine how much each resolution level contributes.

$$\alpha + \beta + \gamma = 1 \quad (17)$$

The embeddings can be concatenated together to maintain all contextual details that exist in the original data.

$$F_i = [F_i^{(u)} \| F_i^{(b)} \| F_i^{(t)}] \quad (18)$$

The mathematical notation ‖ represents the operation of vector concatenation.

Final Feature Representation: The resulting fused vector Fi indicates a multi-resolution semantic representation of the document, which functions as the input feature vector for the DBSMOA-based feature selection process that follows [30].

The MR-NEF method increases feature diversity through its ability to identify semantic patterns across multiple levels of detail. The basic lexical data from unigram features establishes fundamental word meaning, whereas bigram and trigram features show how words relate to each other in their specific contexts. The model achieves better performance through its ability to create distinct representations from multiple-level embeddings, which boost GNN-based classification tests for cyberbullying detection and sentiment analysis.

3.4 DBSMOA-Based Feature Selection

The MR-NEF method first extracts multi-resolution features, which produce a feature space that contains excessive, unnecessary features and irrelevant attributes. The dimensions of the feature space create two problems because they make computations more difficult while decreasing the efficiency of the classification model learning. To solve this problem, the DBSMOA method helps to find the most valuable features. The DBSMOA algorithm operates as a binary version of the Swordfish Movement Optimization (SMO) method, which models the hunting techniques used by swordfish predators. The feature selection process uses candidate solutions, which are represented as binary vectors that map each dimension to a specific feature obtained from the MR-NEF module.

Figure 3 depicts the flowchart of the DBSMOA technique used for feature selection. The process starts with the initialization of a group of swordfish agents representing potential feature subsets through their binary feature sets. The system starts by generating binary solutions, which are then assessed through a fitness function that measures the performance of each feature subset in classification tasks while maintaining a balance between feature selection and reduction. The system evaluates the fitness of all swordfish agents to determine which one has achieved the best results. The DBSMOA updates the locations of swordfish agents during each iteration, which proceeds to transform their new positions into binary feature representations. The DBSMOA evaluates the new solution's effect on the best fitness value. When the new solution shows better results, the algorithm updates the best solution, yet the algorithm advances to its next cycle. The process continues until the termination criterion, which includes reaching the maximum number of iterations, has been met. The DBSMOA provides the optimal feature subset, which serves as the basis for the upcoming Text-GNN classification phase in the proposed model [31].

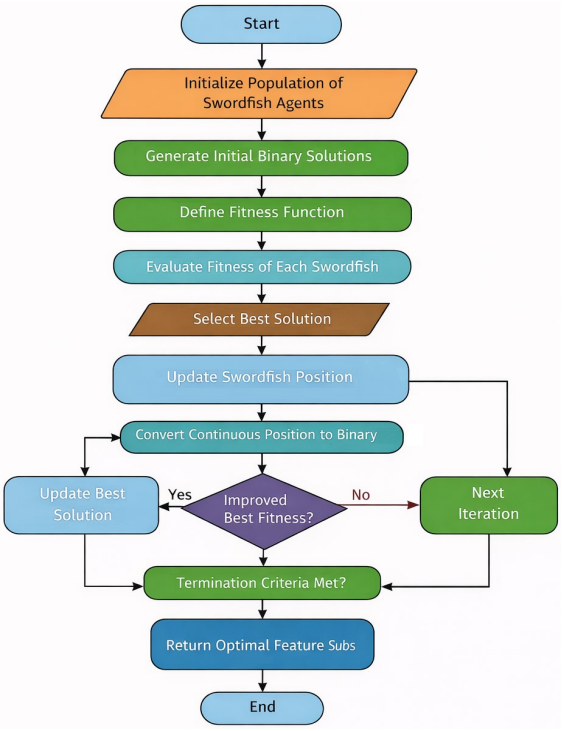

Figure 3: DBSMOA Flowchart

The DBSMOA optimization algorithm uses population-based metaheuristic techniques, which draw their inspiration from swordfish hunting behavior and their synchronized swimming patterns. The algorithm uses dynamic balancing methods to explore different areas while it seeks the best binary feature subset that will enhance classification results. The feature matrix obtained from MR-NEF will be represented by the following equation:

$$F = \{F_1, F_2, \ldots, F_N\} \in \mathbb{R}^{N \times d} \quad (19)$$

Here, N indicates the number of samples, and d denotes the total count of extracted features. The process of feature selection aims to find the best feature subset $S \subseteq d$ which keeps essential features but eliminates unnecessary features.

Binary Representation of Candidate Solutions: In DBSMOA, every candidate solution (swordfish agent) was represented as the binary vector denoting whether a feature is selected or not:

$$X_i = [x_{i1}, x_{i2}, \ldots, x_{id}] \quad (20)$$

Where,

$$x_{ij} = \begin{cases} 1, & \text{if the } j^{th} \text{ feature is selected} \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

Here, X_i represents the position of the ith swordfish agent in the binary search space.

Population Initialization: The optimization process starts with the random creation of a population that consists of *P* swordfish agents.

$$X_i^{(0)} = rand(0,1), \quad i = 1,2,\dots,P \tag{22}$$

Here, $rand(0,1)$ creates binary values indicating the feature selection states.

Fitness Function: The quality of all the candidate feature subsets was evaluated using a fitness function that considers both classification accuracy and feature reduction. The relationship between classification accuracy and feature selection was represented by $Acc(X_i)$, which shows the classification accuracy that results from using the chosen features, and $|X_i|$ indicates the total number of chosen features. The fitness function was defined as follows:

$$Fit(X_i) = \alpha \cdot Acc(X_i) + \left(\frac{1}{-\alpha}\right)\left(-\frac{1}{\frac{|x_i|}{d}}\right) \tag{23}$$

The equation indicates the α as a weighting parameter that manages the balance between classification performance and feature reduction, while d represents the complete count of features.

Swordfish Movement Update: Each swordfish agent's movement was updated according to the best-performing solution in the population. The position update rule was expressed as follows:

$$X_i^{t+1} = X_i^t + r_1 \cdot \begin{pmatrix} X_{best}^t \\ -X_i^t \end{pmatrix} + r_2 \cdot \begin{pmatrix} X_{rand}^t \\ -X_i^t \end{pmatrix} \tag{24}$$

The current position of the *i*th agent at iteration t is represented by X_i^t, while X_{best}^t represents the global best solution and X_{rand}^t refers to a randomly selected agent, with r_1 and r_2 being random coefficients that define exploration and exploitation.

Binary Transformation: The optimization problem needs binary feature selection, so the continuous position values must be converted into binary decisions through the application of a sigmoid transfer function.

$$S(v) = \frac{1}{1+e^{-v}} \tag{25}$$

The updated binary state was attained as follows:

$$x_{ij}^{t+1} = \begin{cases} 1, & if\ S\left(v_{ij}^{t+1}\right) > rand(0,1) \\ 0, & otherwise \end{cases} \tag{26}$$

The equation indicates that v_{ij}^{t+1} represents the updated velocity component for the *j*th feature.

Optimal Feature Subset Selection: The optimization procedure updates the population through multiple iterations until either the maximum iteration count T is reached or the convergence conditions are fulfilled. The best optimum solution obtained during the search was defined as follows:

$$X^* = \arg\max_{X_i} Fitness(X_i) \tag{27}$$

Here, X^* indicates the optimal binary feature vector that chooses the most significant features from the MR-NEF feature set.

Selected Feature Representation: The reduced feature matrix was finally created as follows:

$$F^* = F \odot X^* \tag{28}$$

Here, the notation $\odot$ indicates the features' element-wise selection according to the optimal binary vector.

The DBSMOA-based feature selection approach effectively reduces feature dimensionality while retaining the most discriminative attributes. The optimization process achieves better results through DBSMOA because it creates a dynamic exploration-exploitation balance, which enhances feature subset quality, computing efficiency, and delivers stronger results for Text-GNN classification, which analyzes sentiment and detects cyberbullying.

3.5 Text-GNN Based Classification

The Text-GNN model functions as an advanced DL system that uses graph-based text representation to extract intricate semantic and contextual links present in textual data. Conventional natural language processing models, such as bag-of-words, RNNs, and CNNs, process text through linear order, which prevents them from understanding long-range dependencies and global semantic relationships between words and documents. Text-GNN builds its graph through two components, which include textual entities that contain words, phrases, and entire documents as its nodes, and each relationship includes co-occurrence, semantic similarity, and contextual association as its edges. Through this graph representation, the model uses GNN architectures to distribute information between linked nodes, which leads to the development of an advanced contextual representation.

Text-GNN achieves its purpose of processing textual data, which requires understanding sentiment, sarcasm, abusive language, and other minor linguistic patterns through its method of successively collecting data from adjacent nodes using graph convolution operations. The model uses

a graph-based learning system to study how words and documents relate to each other while maintaining the meaning that derived from their embedded representations. Text-GNN has become a strong method for multiple natural language processing tasks, which include text classification, sentiment analysis, topic modeling, and cyberbullying detection, because it can better understand complex language relationships and contextual information than traditional DL models. The Text-GNN classifier uses the optimized feature set, which results from selecting important features through the DBSMOA process. The Text-GNN model captures the complex contextual relationships between words and documents by representing textual data as a graph structure.

Graph Construction: The proposed Text-GNN model uses a heterogeneous graph to convert textual information into a graphical form because it operates differently from traditional neural networks, which process text as sequential data. Here, nodes correspond to textual units and edges represent semantic relationships between them.

The graph $G = (V, E)$ consists of:

Node set V: indicating the documents and words.

Edge set E: indicating the relations which include semantic similarity and word co-occurrence.

The analysis includes two different types of nodes:

- Word nodes represent unique terms that are extracted from the corpus.
- Document nodes represent individual text samples.

The establishment of edges occurs through the analysis of co-occurrence patterns, which emerge during the operation of a sliding context window. The system establishes an edge between two words when both words appear together in the same context window. The system creates document-word edges, which show how specific words appear throughout documents. Statistical association measures, which include Pointwise Mutual Information (PMI), serve as the basis for determining the weight of each edge.

$$PMI(i,j) = \log \frac{P(i,j)}{P(i)P(j)} \quad (29)$$

Here, $P(i,j)$ indicates the joint probability of a word pair (i,j), and $P(i)$ and $P(j)$ indicates the individual word probabilities. Positive PMI values create strong semantic connections, which enhance the connectivity of the graph structure.

Node Feature Initialization: The initial feature representation for every node was derived from the MR-NEF embeddings, which extract the textual data's multi-resolution semantic representations. The embeddings function as the input feature matrix for the system.

$$H^{(0)} = X \quad (30)$$

Here, X indicates the MR-NEF feature matrix, and $H^{(0)}$ indicates the initial node representation. The selected feature subset using the DBSMOA ensured that only the most relevant features were propagated into the graph learning stage.

Graph Convolution Operation: The Text-GNN uses graph convolution layers to learn contextual representations because these layers send information to adjacent nodes. The graph convolution operation at layer l was defined as follows.

$$H^{(l+1)} = \sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \quad (31)$$

Here, $\hat{A} = A + I$ was the adjacency matrix with self-connections, $\hat{D}$ was the degree matrix, $H^{(l)}$ was the node representation at layer l, $W^{(l)}$ was the trainable weight matrix, and σ indicates a nonlinear activation function such as ReLU. The model captures both the local and global contextual dependencies during the iterative aggregations in its neighborhoods.

Feature Propagation and Representation Learning: The graph convolution layer updates node representations during each layer by collecting data from neighboring nodes. The model extracts semantic patterns that include contextual sentiment relationships, implicit abusive language patterns, and word-document semantic connections. The network develops its ability to classify text by stacking multiple graph convolution layers to create high-level text representations.

Classification Layer: The final node representations after feature propagation undergo processing through a softmax classification layer, which predicts the sentiment and cyberbullying class labels.

$$Z = softmax\left(H^{(L)}H^{(c)}\right) \quad (32)$$

Here, $H^{(L)}$ indicates the final node embeddings after L graph convolution layers, $W^{(c)}$ was the classification weight matrix, and Z indicates the predicted probability distribution across classes.

The model training process uses the cross-entropy loss function, which is defined as follows:

$$\mathcal{L} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \quad (33)$$

Here, y_i was the true class label and $\hat{y}_i$ was the predicted probability [32].

The benefits of using Text-GNN for sentiment analysis and cyberbullying detection emerge from its three main features. The first feature of the system enables it to capture word connections through contextual information instead of relying on sequential processing. The second feature of the system enables it to understand the complete semantic relationships that exist throughout the entire document collection. The system improves document-word interaction learning through its enhanced feature interaction capabilities. The system strengthens classification performance through its ability to learn from graph-based representations. The Text-GNN classifier uses both semantic embeddings and graph-structured information to achieve better detection results in sentiment classification and cyberbullying identification through the combination of MR-NEF-based feature extraction and DBSMOA-based feature selection.

The following Algorithm 1 presents the pseudocode of the proposed MR-NEF–DBSMOA–Text-GNN framework for sentiment classification and cyberbullying detection.

Algorithm: MR-NEF-DBSMOA-Text-GNN

Input: Raw textual dataset D
Population size N
Maximum iterations Tmax
Output: Predicted sentiment / cyberbullying labels Y
Initialize
Load dataset D
Perform Text Preprocessing
Convert text to lowercase
Remove punctuation and special characters
Remove stop words
Perform tokenization
Apply stemming or lemmatization
Feature Extraction using MR-NEF
Generate multi-resolution word embeddings
Capture contextual and semantic representations
Construct feature matrix F
Initialize DBSMOA Feature Selection
Initialize swordfish population
Generate binary feature subsets
Define fitness function F(S)
Evaluate the fitness of each solution
Identify the best solution Sbest
for t = 1 to Tmax do
for each swordfish agent do
Update position using the movement equation
Convert continuous position to binary using a sigmoid transfer function
Update feature subset
Evaluate the fitness of the updated solution
If fitness improves, then
Update Sbest
end if
end for
end for
*Obtain optimal feature subset S**
Construct Text Graph
Create document nodes
Create word nodes
Establish edges using word co-occurrence
Initialize node features using selected MR-NEF features
Train Text-GNN classifier
Apply graph convolution layers
Aggregate neighborhood information
Learn node representations
Apply Softmax classification layer
Predict sentiment / cyberbullying labels Y
End

Table 3: Hyperparameter Details.

Component	Hyperparameters	Value
MR-NEF Feature Extraction	Embedding Dimension	300
	Context Window Size	5
	Dropout Rate	0.3
	Learning Rate	0.001
	Batch Size	64
DBSMOA Feature Selection	Population Size	30
	Maximum Iterations	100
	Exploration Factor	0.5
	Drift Coefficient	0.2
	Fitness Weight	0.9
	Transfer Function	Sigmoid
Text-GNN Classifier	Number of GNN Layers	2
	Hidden Layer Dimension	128
	Activation Function	ReLU
	Learning Rate	0.0005
	Dropout Rate	0.5
	Training Epochs	200
	Optimization Algorithm	Adam

This research aims to achieve optimal performance through empirical testing of hyperparameters within the MR-NEF-DBSMOA-Text-GNN framework, as shown in Table 3. The MR-NEF module uses a 300-dimensional embedding space to capture semantic information. The DBSMOA algorithm employs a population of 30 agents and 100 iterations to effectively search

for the optimal feature subset. The Text-GNN classifier uses two graph convolution layers, which operate at a hidden dimension of 128 to extract contextual information from the built text graph.

4. EXPERIMENTATION ANALYSIS

4.1 Experimental Setup

This section presents the experimental tests to determine how well the MR-NEF-DBSMOA-Text-GNN framework performs in text classification tests, which include sentiment analysis and cyberbullying detection. The experiments used a Python programming environment with DL libraries PyTorch and Scikit-learn to develop and test the models. The research created two subsets of the datasets, which are used for testing and training purposes with an 80:20 distribution. The experiments are performed on a system with an Intel Core i7 processor, 16 GB of RAM, and an NVIDIA GeForce RTX 3060 GPU to achieve optimal model training and model improvement results.

4.2 Performance Measures

To evaluate the performance of the MR-NEF-DBSMOA-Text-GNN framework, the model was assessed using the metrics of accuracy, precision, F1-score, and recall.

The accuracy metric calculates the total number of correctly identified cases to evaluate model efficiency, which works best when class distributions stay even across categories.

Precision assesses the relationship between correctly identified positive cases and all predicted positive cases, which holds value for cyberbullying detection because it helps reduce false positive errors.

The model's ability to find true positive cases depends on its performance in identifying actual positive instances according to the recall metric, which guarantees that all cyberbullying cases will be detected.

The F1-score is a reliable classification comparison metric. It measures classification accuracy through its dual assessment of false positive and false negative errors.

4.3 Results & Discussion

This segment presents the detailed analysis of the proposed MR-NEF-DBSMOA-Text-GNN model in evaluating three datasets and the comparison of results with the analyzed current methodologies. Additionally, an ablation study comparison was performed and discussed in detail.

Table 4: Results of MR-NEF-DBSMOA-Text-GNN model on the IMDB dataset.

Metrics	Training Data	Test Data
Accuracy	96.92	96.31
Recall	96.48	96.07
Precision	96.97	96.24
F1	96.73	96.18

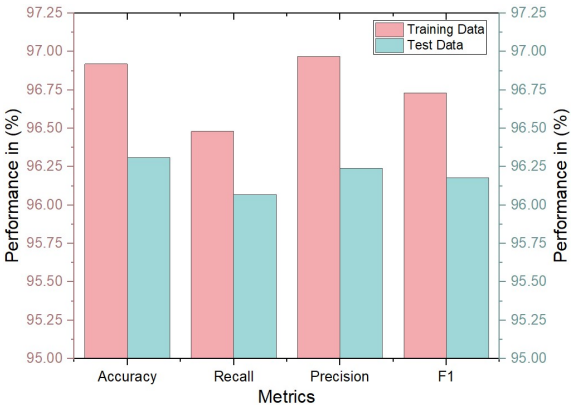

Figure 4: Graph of MR-NEF-DBSMOA-Text-GNN model's results on the IMDB dataset

Table 4 presents the results of the proposed MR-NEF-DBSMOA-Text-GNN model in performing sentiment classification on the IMDB dataset. The model achieved a 96.92% training accuracy together with 96.31% test accuracy, which demonstrates its ability to learn significant sentiment patterns from training data while still performing well on new information. The model demonstrates high accuracy in detecting positive sentiment through its recall values, which reach 96.48% for training and 96.07% for testing. The precision values of 96.97% and 96.24% show that most predicted positive sentiments are accurate because the feature extraction and selection methods used in the framework proved to be effective. The F1 scores of 96.73% and 96.18% indicates that the system achieved an equal distribution between its precision and recall processes, which leads to consistent performance results. The MR-NEF, DBSMOA, and Text-GNN integration shows strong effectiveness in enhancing sentiment classification performance because it maintains accurate and generalized predictions across all metrics for the IMDB dataset. Figure 4 depicts the graph of the proposed model's result using the IMDB dataset.

Table 5: Results of MR-NEF-DBSMOA-Text-GNN model on the Yelp dataset.

Metrics	Training Data	Test Data
Accuracy	96.98	96.52

Recall	96.69	96.21
Precision	96.93	96.44
F1	96.81	96.36

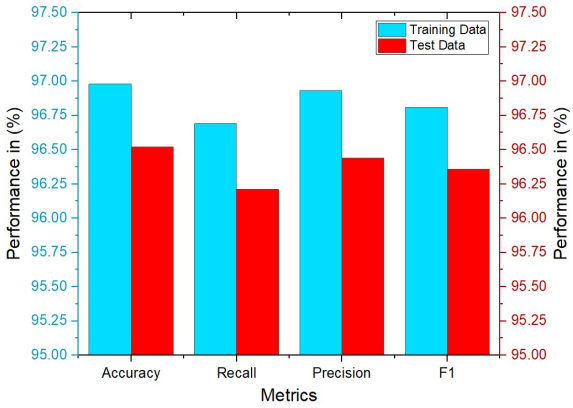

Figure 5: Graph of MR-NEF-DBSMOA-Text-GNN model's results on the Yelp dataset

Table 5 presents the results of the proposed MR-NEF-DBSMOA-Text-GNN model on the Yelp polarity dataset. The model achieved a training accuracy of 96.98% and a test accuracy of 96.52%, which shows its ability to learn sentiment patterns from training data while effectively predicting results on new test data. The recall results of 96.69% with training data and 96.21% with test data demonstrate effective positive sentiment detection ability because it correctly identifies positive sentiment cases. The precision values of 96.93% and 96.44% show that the model's predictions have high accuracy because most of its predicted positive sentiments turn out to be true. The training F1 score of 96.81% and testing F1 score of 96.36% show that the system maintains an equal balance between its precision and recall functions while the classification method remains stable. The system demonstrates complete ability to detect sentiment from customer reviews based on its stable performance across all assessment standards, which shows that the model provides better sentiment analysis results for the Yelp dataset. Figure 5 depicts the graph of the proposed model's result using the Yelp dataset.

Table 6: Results of MR-NEF-DBSMOA-Text-GNN model on the Cyberbullying Classifying dataset.

Metrics	Training Data	Test Data
Accuracy	98.92	98.63
Recall	98.71	98.55
Precision	98.86	98.47
F1	98.79	98.58

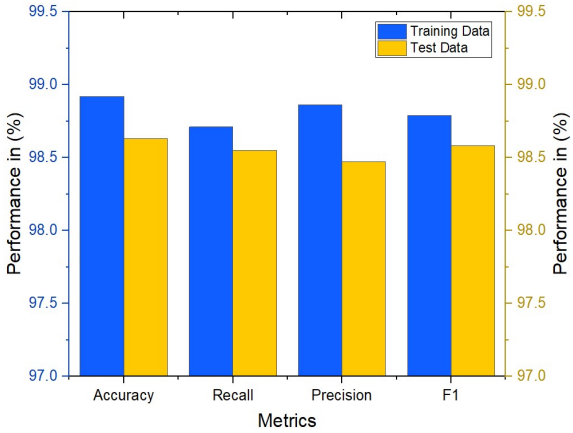

Figure 6: Graph of MR-NEF-DBSMOA-Text-GNN model's results on the Cyberbullying dataset

Table 6 presents the results, which demonstrate that the MR-NEF-DBSMOA-Text-GNN model successfully performs cyberbullying classification. The model achieved a training accuracy of 98.92% and a test accuracy of 98.63%, which demonstrates its ability to learn complex linguistic patterns used in cyberbullying while preserving its performance on new data. The model achieved 98.71% recall during training and 98.55% recall during testing, which shows that it correctly detects most cyberbullying cases while making only a few incorrect detections. The precision scores of 98.86% and 98.47% show that most cyberbullying cases, which the research identified, were accurately categorized, while the system produced only a few false positive results. The F1 scores of 98.79% (training) and 98.58% (testing) demonstrate that the model maintains a complete balance between precision and recall, which establishes the classification system as dependable and consistent. The model demonstrates its ability to detect cyberbullying content in social media text with high precision because it achieves success across all performance metrics. Figure 6 depicts the graph of the proposed model's result using the Cyberbullying dataset.

Table 7: Multiclass Results of MR-NEF-DBSMOA-Text-GNN Model on Cyberbullying Classifying Dataset.

Category	Acc	Recall	Prec	F1
Age	98.78	98.60	98.66	98.57
Religion	98.69	98.55	98.63	98.52
Gender	98.61	98.49	98.58	98.48
Ethnicity	98.57	98.46	98.53	98.45
Not Cyberbullying	98.54	98.44	98.47	98.42
Other Cyberbullying	98.65	98.58	98.55	98.56
Average	98.64	98.52	98.57	98.50

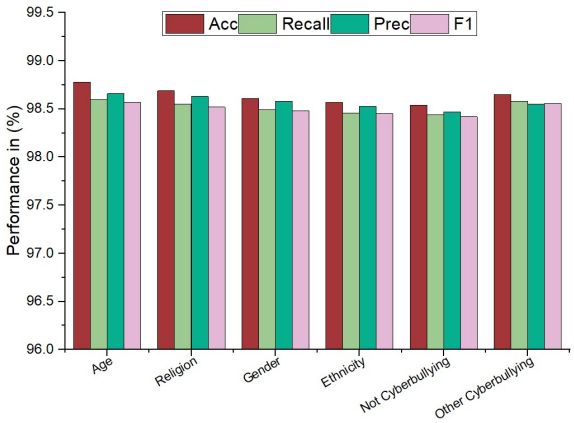

Figure 7: Graphical Chart for Comparison of Coefficient Results

Table 7 demonstrates the results of the MR-NEF-DBSMOA-Text-GNN model in multiclass classification tests against different types of cyberbullying datasets. The model demonstrates consistent performance across all classes because its accuracy ranges between 98.54% and 98.78%. The model successfully identifies all types of cyberbullying and non-cyberbullying content. The model achieved superior performance for the age-based cyberbullying class among all categories, as it reaches 98.78% accuracy, 98.60% recall, 98.66% precision, and 98.57% F1 score, which demonstrates the model's ability to identify offensive content related to age. The model demonstrates effective linguistic detection capabilities for both Religion and Gender categories, which presents a balance between their high-performance metrics. The Ethnicity and Not Cyberbullying categories maintain slightly lower yet still very high scores because the model can accurately identify harmful content while misclassifying non-harmful content. The Other Cyberbullying category shows strong detection capabilities, which demonstrate that the proposed framework can adapt to various cyberbullying detection patterns. The proposed model achieved accurate multiclass classification results with average scores of 98.64% accuracy, 98.52% recall, 98.57% precision, and 98.50% F1 score metrics, which demonstrate its dependable and balanced performance across multiple classes. The model demonstrates effective performance when detecting precise instances of cyberbullying through its fine-grained detection capabilities. Figure 7 depicts the graph of the proposed model's multiclass classification results.

Table 8: Comparison of Results with Current Models.

Models	Acc	Recall	Prec	F1
RoBERTa + BiLSTM [11]	94.80	95.30	96.40	95.80
Stacking Classifier [12]	90.71	90.60	90.85	90.63
LSTM + Fuzzy Logic [13]	93.67	93.62	93.65	93.64
RF [14]	94.20	94.00	94.00	94.00
LSTM [15]	96.64	96.55	96.80	96.67
RNN [16]	98.00	96.00	98.00	97.00
FAEO + ECNN [17]	92.91	92.28	92.53	92.40
Fusion Model [18]	98.37	98.22	98.47	98.36
Extra Tree [20]	95.38	NA	NA	95.00
FedBERT [21]	92.15	89.15	90.25	88.65
LSA + SVM [22]	97.00	94.00	95.00	97.00
MLP [23]	95.00	97.00	98.00	98.00
GNN [24]	92.78	80.15	96.51	87.57
DCSSL + NLI [25]	93.49	94.75	90.46	92.31
Proposed Model	98.64	98.52	98.57	98.50

The comparison results presented in Table 8 demonstrate the effectiveness of the proposed MR-NEF-DBSMOA-Text-GNN model when evaluated against several current models for cyberbullying detection. Traditional models such as Stacking Classifier, RF, and FAEO+ECNN show comparatively lower performance, with accuracy values generally below 95%, indicating limitations in capturing complex contextual patterns in textual data. The performance of LSTM, RNN, and GNN models shows improvement through their execution, yet still displays moderate discrepancies between their recall and precision results. The combination of deep contextual embeddings with neural architectures leads to better results in the RoBERTa + BiLSTM and Fusion Model hybrid approaches, which achieved accuracy rates of 94.80% and 98.37%, respectively. The proposed model achieves its best results through balanced performance, which delivers 98.64% accuracy, 98.52% recall, 98.57% precision, and 98.50% F1 score while surpassing most existing methods. The proposed model improves performance because it accurately models the relationships between different textual elements. The results verify that the proposed framework delivers better accuracy and stronger performance for cyberbullying

detection than current leading models. Figure 8 depicts the graph of the proposed model's results compared with current models.

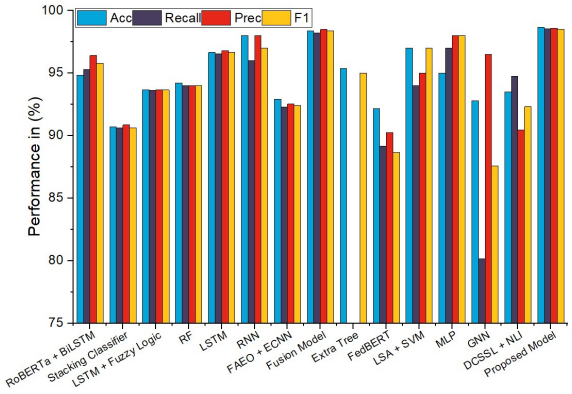

Figure 8: Graph of Results Comparison with Current Results

Table 9: Comparison of Results with Current Models.

Models	Acc	Recall	Prec	F1
Text-GNN (Baseline)	96.82	96.44	96.71	96.57
MR-NEF + Text-GNN	97.48	97.12	97.35	97.23
MR-NEF + DBSMOA + BiLSTM	97.96	97.65	97.88	97.76
MR-NEF + DBSMOA + CNN	98.21	98.03	98.16	98.09
Proposed Model	98.64	98.52	98.57	98.50

Table 9 provides ablation study results that demonstrate how each variant of the proposed framework contributes to its overall function. The baseline Text-GNN model achieved fair performance. However, the integration of MR-NEF-based feature extraction improves the representation of textual information by capturing multi-resolution contextual patterns, which results in noticeable performance gains. The integration of DBSMOA-based feature selection eliminates redundant features that have low informational value, resulting in improved classification performance and system reliability. The tests with BiLSTM and CNN classifiers demonstrate limited performance enhancement, but they still fall short of the graph-based learning method. The complete proposed architecture achieves the highest performance across all metrics, confirming that the combined effect of multi-resolution embeddings, optimized feature selection, and graph-based semantic learning significantly strengthens the model's capability for accurate sentiment and cyberbullying detection. Figure 9 depicts the graph of the ablation study results comparison performance.

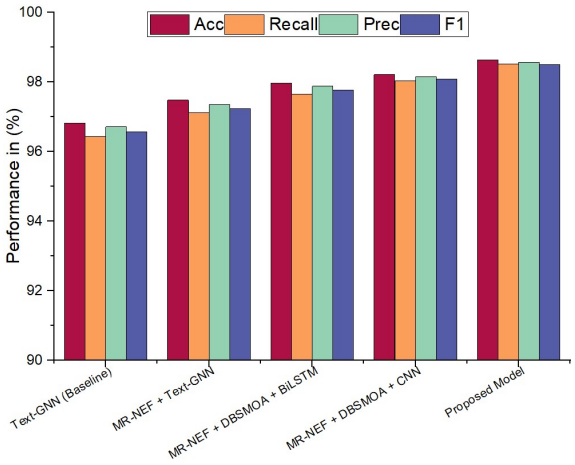

Figure 9: Graph of Ablation Study Results Comparison

The MR-NEF-DBSMOA-Text-GNN framework, which we proposed, provides multiple benefits for both text classification and cyberbullying detection tasks. The MR-NEF mechanism enabled the model to capture semantic data through its n-gram system, which expands its capability to analyze both nearby and wide-ranging linguistic patterns. The DBSMOA method enabled effective selection of essential features while it removes unnecessary attributes, which leads to reduced processing needs and better results in identification tasks. The Text-GNN classifier used graph structures to model word-document relationships, which enabled the system to identify long-range dependencies and semantic connections that traditional sequential models fail to detect. The integrated framework achieved high classification accuracy while maintaining strong performance across different datasets.

The proposed model includes a few constraints that limit its effectiveness. The Text-GNN framework faces increased memory demands and processing requirements when it needs to build graph structures for very large text collections. The model performance is determined through the proper parameter adjustment, which needs to be tested during the DBSMOA optimization procedure. The method requires testing on specific benchmark datasets, while its ability to function in noisy multilingual real-world social media environments needs more testing to prove its effectiveness.

4.4 Critical Discussion and Limitations

Even though the MR-NEF-DBSMOA-Text-GNN framework outperformed in several benchmarking datasets, several limitations should be taken into account carefully. The new framework involves multi-resolution embedding, optimization-based feature selection, and GNN learning methods that increase the accuracy of classification tasks and contextual comprehension. Nevertheless, the implementation of these approaches complicates the overall algorithm and requires more computational power. Building text graphs at scale and applying DBSMOA optimizations require significant computing memory and time consumption.

A further limitation of the proposed approach is that it is mostly evaluated on benchmark datasets with a rather structured class distribution. In a real-time environment, however, there may be various languages used in social media texts, such as slang terms, emojis, code mixing, and incomplete sentences. These linguistic elements can potentially impact the ability of the proposed model to generalize. Moreover, the construction of the graph relies heavily on the co-occurrence of words and contextual windows, and inappropriate values of graph parameters might have an adverse effect on classification results.

This model additionally needs fine-tuning of hyperparameters related to embedding size, graph convolution layer, optimization coefficients, and feature selection. These hyperparameters can greatly affect the model's performance and its ability to converge. In addition, while the proposed model shows excellent results in terms of classification accuracy, the problem of explainability persists because of the nature of graph neural networks as black-box models. This interprets predictions made by these models as hard, particularly when dealing with issues like cyberbullying.

However, despite these limitations, the research model is very effective in terms of modeling semantic relationships and identifying harmful text patterns accurately. This research has provided a solid basis for future work in developing techniques that include lightweight graph models, multilingual support, integration of transformers into graph learning, explainability in AI, and real-time applications for online monitoring systems.

5. CONCLUSION

This research proposed a new hybrid approach, MR-NEF-DBSMOA-Text-GNN, that was used for sentiment analysis and cyberbullying detection on social media text. The main novelty in this work was the use of a combination of multi-resolution n-gram embedding fusion, dynamic optimization-based feature selection, and graph neural networks-based contextual learning in one intelligent framework. In contrast to the traditional ML/DL techniques that focus on using sequential textual features, the proposed approach was capable of capturing semantic dependencies, context relations, and latent abusive language patterns through graph-based representation learning. The MR-NEF scheme helped improve contextual semantics extraction by integrating unigrams, bigrams, and trigrams, whereas the DBSMOA technique facilitated dimension reduction and improved classification. Moreover, the Text-GNN classifier effectively modeled complex document-word interactions and long-distance semantic relationships to classify sentiment and cyberbullying accurately. Experiments performed using the IMDB, Yelp Polarity, and Cyberbullying datasets showed that the proposed approach obtained better and balanced results, achieving an accuracy of 98.64%, precision of 98.57%, recall of 98.52%, and F1-score of 98.50%, surpassing all other baseline and state-of-the-art approaches. This research will influence the development of intelligent automated systems to monitor harmful content, moderation on social networks, increase online safety, and prevent cyberbullying. In general, the proposed approach makes an important contribution to the development of efficient text analysis and cyberbullying detection techniques.

This research makes significant contributions to the existing knowledge on sentiment analysis and cyberbullying by presenting an integrated approach that integrates the concepts of multi-resolution semantic embeddings, feature selection through optimization, and graph-based context learning into one framework. Unlike previous studies, which consider each concept independently, the suggested MR-NEF-DBSMOA-Text-GNN approach considers all three aspects to improve context understanding, feature representation, and robust classification. The research further shows that by using graph-based semantic modeling in conjunction with optimization-based features, it is possible to effectively detect complex and implicit abusive language structures on social media texts.

In future, this work will focus on extending the proposed framework to handle large-scale and multilingual social media datasets to improve its generalization capability. The combination of

contextual transformer-based embeddings together with dynamic graph construction methods will improve semantic representation capabilities. The implementation of explainable AI systems will help users understand how cyberbullying detection systems make their decisions. Additionally, the research will investigate two areas, which include methods for real-time system deployment and techniques to optimize the processing of streaming data in actual online monitoring environments.

REFERENCES:

[1] J.A. Diaz-Garrcia and J.P. Carvallho, A Literature Review of Textual Cyber Abuse Detections Using Cutting-Edge Natural Languages Processing Technique: Languages Model and Large Languages Model, Wiley Interdisciplinary Review: Data Mining and Knowledge Discovery, Vol. 15, No. 3, 2025, e70029. https://doi.org/10.1002/widm.70029

[2] S. Wang, A.S. Shibghatulah, T.J. Iqbal, and K.H. Koey, A review of multimodal-based emotions recognition technique for cyberbullying detections in online social media platform, Neural Computing and Application, Vol. 36, No. 35, 2024, pp. 21923-21956. https://doi.org/10.1007/s00521-024-10371-3

[3] M.T. Hasan, M.A.E. Hosain, M.S.H. Muktha, A. Aktar, M. Ahmad, and S. Islam, A review on deep-learnings-based cyberbullying detections, Future Internet, Vol. 15, No. 5, 2023, 179. https://doi.org/10.3390/fi15050179

[4] A. Sundaram, H. Subramanian, S.H. Ab Hammid, and A.M. Norr, A systematic literatures review on social media slang analytic in contemporary discourses, IEEE Access, Vol. 11, 2023, pp. 132457-471. https://doi.org/10.1109/ACCESS.2023.3334278

[5] V.M. Ngo and P.Q. Dao, A Systematic Review of Sentiment Analysis System Applied to Textual Data, In International Conferences on Multi-disciplinary Trend in Artificial Intelligences, Singapore: Springer Nature Singapore, 2025, pp. 239-51. https://doi.org/10.1007/978-981-95-4963-4_20

[6] L.M. Al-Harigy, H.A. Al-Naim, N. Moradpor, and Z. Tan, Building toward automated cyberbullying detections: A comparative analysis, Computational Intelligences and Neurosciences, Vol. 2022, No. 1, 2022, 4794227. https://doi.org/10.1155/2022/4794227

[7] A.G. Philipo, D.S. Sarwat, J. Ding, M. Danesmand, and H. Ning, Cyberbullying detections: Exploring dataset, technologies, and approaches on social media platform, ACM Computing Survey, Vol. 58, No. 7, 2026, pp. 1-35. https://doi.org/10.1145/3785654

[8] J.S. Malik, H. Qio, G. Pang, and A. van den Henggel, Deep learning for hate speech detections: a comparative study, International Journal of Data Sciences and Analytics, Vol. 20, No. 4, 2025, pp. 3053-68. https://doi.org/10.1007/s41060-024-00650-6

[9] E.A. Elhadidi, A. Sallah, M. Abdela, and S.M. Darawish, Multilingual Sentiments Analysis: A Review of Deep Learning Transformers Model and Ensembled Technique, International Journal of Computer and Informatics (Zagazig University), Vol. 7, 2025, pp. 91-103. https://www.ijci.zu.edu.eg/index.php/ijci/article/view/107

[10] A. Shah, S. Varshney, and M. Meharotra, Threats on online social networks platform: classifications, detections, and preventions technique, Multimedia Tools and Application, Vol. 84, No. 16, 2025, pp. 17083-115. https://doi.org/10.1007/s11042-024-19724-5

[11] M.A. Mahdi, S.M. Fatti, M.G. Ragab, M.A.G. Hazaber, S. Ahamed, S.A. Saad, and M. Al-Shallabi, A Novel Hybrid Attentions-Based RoBERTa-BiLSTM Model for Cyberbullying Detections, Mathematical and Computational Application, Vol. 30, No. 4, 2025, 91. https://doi.org/10.3390/mca30040091

[12] A.F. Alqahtani and M. Ilyas, An ensemble-based multi-classifications machine learning classifier approach to detect multiple classes of cyberbullying, Machine Learning and Knowledge Extractions, Vol. 6, No. 1, 2024, pp. 156-70. https://doi.org/10.3390/make6010009

[13] M.H. Obaid, S.K. Guirrguis, and S.M. Elkafas, Cyberbullying detections and severity determinations models, IEEE Access, Vol. 11, 2023, pp. 97391-399. https://doi.org/10.1109/ACCESS.2023.3313113

[14] F.R. Sayed, E.H. Elnasar, and F.A. Omarra, Cyberbullying detections in social media using natural languages processing, Scientific African, Vol. 28, 2025, e02713. https://doi.org/10.1016/j.sciaf.2025.e02713

[15] M.H. Obaida, S.M. Elkafas, and S.K. Guirrguis, Deep learning algorithm for cyber-bullying detections in social media platform, IEEE Access, Vol. 12, 2024, pp. 76901-908.

https://doi.org/10.1109/ACCESS.2024.3406595

[16] A.M. Fashakh, M. Çevik, Ş.K. Aydoğan, and A.A. Ibrahem, Detections cyberbullying using AI and sentiment analysis to examine psychological impact on vulnerable group, Egyptian Informatics Journal, Vol. 32, 2025, 100856. https://doi.org/10.1016/j.eij.2025.100856

[17] B.A.H. Murshed, Suresh, J. Abawajy, M.A.N. Saif, H.M. Abdulwaha, and F.A. Ghanem, FAEO-ECNN: cyberbullying detections in social media platform using topics modelling and deep learning, Multimedia Tools and Application, Vol. 82, No. 30, 2023, pp. 46611-650. https://doi.org/10.1007/s11042-023-15372-3

[18] M.M. Hossain, M.S. Hosain, M.S. Hosain, M.F. Mrida, M. Safran, S. Alfarhod, and D. Che, Fusing Transformer-XL with bi-directional recurrent network for cyberbullying detections, PeerJ Computer Science, Vol. 11, 2025, e2940. https://doi.org/10.7717/peerj-cs.2940

[19] M. Buyankhishig, T. Shwe, I. Mendonça, and M. Aritsuggi, Heterogeneous graph neural networks framework for sessions-based cyberbullying detections, IEEE Access, Vol. 13, 2025, pp. 110926-940. https://doi.org/10.1109/ACCESS.2025.3583332

[20] M.F. Almufareh, N.Z. Jhanihi, M. Humayun, G.N. Alwakkid, D. Javeed, and S.N. Almuayqil, Integrating sentiments analysis with machine learnings for cyberbullying detections on social media, IEEE Access, Vol. 13, 2025, pp. 78348-359. https://doi.org/10.1109/ACCESS.2025.3558843

[21] N.A. Samee, U. Khan, S. Khan, M.M. Jamjom, M. Sharrif, and D.H. Kim, Safeguarding online space: a powerful fusion of federated learning, word embedding, and emotional feature for cyberbullying detections, IEEE Access, Vol. 11, 2023, pp. 124524-541. https://doi.org/10.1109/ACCESS.2023.3329347

[22] W. Wulandari, H. Makmmur, D.F. Suriantto, A.A.N. Rissal, N.A.E. Budiarti, S.G. Zain, and A. Waheid, Semantic features engineering with LSA-SVM for cyberbullying comments classifications on Instagram, Informatica, Vol. 49, No. 15, 2025, pp. 165–78. https://doi.org/10.31449/inf.v49i15.6992

[23] Y.M. Ibrahim, R. Esameldin, and S.M. Saad, Social media forensic: An adaptive cyberbullying-related hates speech detections approach based on neural network with uncertainty, IEEE Access, Vol. 12, 2024, pp. 59474-484. https://doi.org/10.1109/ACCESS.2024.3393295

[24] M.R. Nurfiqri, The performance analysis of graph neural networks (GNN) and convolutional neural networks (CNN) algorithm for cyberbullying detections in Twitter comment, The Indonesian Journal of Computer Sciences, Vol. 13, No. 3, 2024, pp. 3656-67. https://doi.org/10.33022/ijcs.v13i3.3940

[25] L.M. Al-Harigy, H.A. Al-Naim, N. Moradpor, and Z. Tan, Towards a cyberbullying detections approach: fine-tuned contrastive self-supervised learnings for data augmentations, International Journal of Data Sciences and Analytics, Vol. 19, No. 3, 2025, pp. 469-90. https://doi.org/10.1007/s41060-024-00607-9

[26] K. Karunarathna and R. Rupasinga, Learning to use normalization technique for preprocessing and classifications of texts document, International Journal of Multidisciplinary Studies, Vol. 9, No. 2, 2022, pp. 69-81.

[27] Z. Shaukat, A.A. Zulfiqar, C. Xio, M. Azeem, and T. Mahmood, Sentiments analysis on IMDB using lexicons and neural network, SN Applied Science, Vol. 2, No. 2, 2020, 148. https://doi.org/10.1007/s42452-019-1926-x

[28] M. Sadikin and A. Fazan, Evaluations of machine learning approach for sentiments analysis using Yelp datasets, European Journal of Electrical Engineering and Computer Sciences, Vol. 7, No. 6, 2023, pp. 58-64. https://doi.org/10.24018/ejece.2023.7.6.583.

[29] A. Palagati, S.K. Bala, S.A.J. Babullo, L. Raja, K.K. Nataraja, and R. Kalimuthu, Comparative Analysis of Machine Learning Algorithm and Dataset for Detecting Cyberbullying on Social Media Platform, In 2024 International Conferences on Computing and Intelligent Reality Technologies (ICCIRT), 2024, pp. 391-96. https://doi.org/10.1109/ICCIRT59484.2024.10922033

[30] S. Chen, B. Lang, Y. Cheng, and C. Xie, Detections of algorithmically generated malicious domains name with features fusion

of meaningful words segmentations and N-Gram sequence, Applied Science, Vol. 13, No. 7, 2023, 4406. https://doi.org/10.3390/app13074406

[31] F.H. Rizk, K.S. Gabber, M.M. Eid, D.S. Khafaga, A.A. Alhusan, and E.S.M. El-kenawey, Dynamic binary swordfish movements optimizations algorithms for features selections, Journal of Big Data, Vol. 13, No. 8, 2026. https://doi.org/10.1186/s40537-025-01328-x

[32] S.D. Cui, K. Duan, W. Ma, and H. Shinou, CMGN: Text GNN and RWKV MLP-mixers combined with cross-features fusions for fake news detections, Neurocomputing, Vol. 633, 2025, 129811. https://doi.org/10.1016/j.neucom.2025.129811