

Original Article

MSCT-CDMO-TFT: A Hybrid Deep Learning Framework for Sentiment Analysis and Cyberbullying Detection

D. Jagannathan¹, N. V. Balaji²

^{1,2}Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India.

¹Corresponding Author : djagannathan1@gmail.com

Received: 11 February 2026

Revised: 12 March 2026

Accepted: 15 April 2026

Published: 27 May 2026

Abstract - Cyberbullying has proliferated, particularly among youngsters, due to the expansion of digital platforms and technological advancements. Cyberbullying has emerged as a worldwide concern on social media sites, when individuals transition between networks to avoid detection. This study proposes a Cyberbullying Detection (CBD) framework that is trained using sentiment, emotional, and contextual features. The proposed CBD framework is developed using a Multi-Scale Contextual Transformer-Chaotic Dwarf Mongoose Optimizer-Temporal Fusion Transformer (MSCT-CDMO-TFT) model. The MSCT-CDMO-TFT model is evaluated with IMDB, Yelp Polarity, and Cyberbullying Classification datasets. The MSCT algorithm is applied for extracting multi-level contextual features, which enabled the model to learn local semantic patterns together with global contextual dependencies from complex textual data that contained various noise elements. The application of CDMO in feature selection led to a reduction of redundant features while maintaining essential discriminative features, which resulted in improved performance and classification accuracy. The framework also applied a TFT classifier that successfully executed temporal and contextual interaction modelling through its gated residual learning and attention mechanism capabilities. The proposed model demonstrates superior results on the cyberbullying classification dataset with 98.49% accuracy, 98.35% recall, 98.28% precision, and a 98.40% F1-score. The MSCT-CDMO-TFT model achieves better performance than all other current methods discussed in this study.

Keywords - Cyberbullying Detection, CdmO, Multi-Scale Contextual Transformer, Temporal Fusion Transformer, Sentiment Classification.

1. Introduction

The emergence of online social media platforms, the extensive availability of information and communication technologies, and the common use of smartphones and computers have resulted in users on the Internet experiencing increased freedom of speech. Moreover, social media users often possess the capacity to hide their identities, facilitating the misuse of diverse features. The prevalence of abusive language has emerged as a significant concern on social media. Offensive language denotes any mode of expression that exhibits abusive behavior intended to inflict harm on others. Numerous sorts of abusive language are evident on social media platforms, encompassing racism, sexism, hate speech, cyberbullying, and toxic comments [1].

Social media has become an essential component of social existence. Over 60% of individuals globally utilize social media. The growing interconnectivity facilitates continuous communication, exchange of data, and community development. Although social networks provide numerous

benefits, they also present significant opportunities for abusive behaviors in the online environment, such as cyber harassment, hate speech, and cyberbullying. These terms are frequently used interchangeably; however, they possess distinct limits in scope and intent [2].

Cyberbullying is characterized as the act of insulting, threatening, exposing, Or Harassing others using communication media, including Smartphones, Chats, Websites, Emails, Forums, And Communities. It aims at individuals and exploits the anonymity and wide scope of the digital landscape to cause harm, resulting in psychological trauma [3]. On the other hand, hate speech denotes cyberbullying aimed at a certain group and is defined by negative comments directed against those individuals. Thirdly, cyber harassment constitutes a comprehensive word. It is defined as a digital interaction between an aggressor and a target in the presence of an audience. Examples of such acts encompass threats, reputational harm, or harassment by email or messaging directly [4]. Each digital platform shows its



negative effects through various phenomena, which have different degrees of impact and objectives. The process of controlling these activities becomes difficult because it results in major emotional and social consequences [5].

Psychological damage causes people to experience sadness and problems with concentration and reduced self-esteem, which can develop into suicide thoughts in extreme cases. A single serious accident might adversely affect an individual's mental health [6]. Social media enables rapid information to spread, which reaches large audiences to create permanent damage through cyberbullying. The digital traces of cyberbullying continue to exist for long periods after the original content has been deleted, which results in increased negative effects for the victims. The cyberbullying problem especially affects young people, who experience it at rates higher than 55% of students who have encountered such situations. The study shows that 41% of adults have experienced cyber harassment, which demonstrates how widely this issue affects society.

Flaming	Cyber Harassment	Cyber-stalking	Denigration
Trickery	Masquerading	Exclusion	Doxing

Fig. 1 Cyberbullying types

Figure 1 demonstrates all cyberbullying types, which are defined in the following.

- **Flaming:** It describes a situation where two or more people create aggressive online dialogue through insults and taunts.
- **Cyber Harassment:** The act of sending rude and offensive messages, together with vulgar remarks, results in the intentional goal of insulting or demeaning the targeted individual.
- **Cyberstalking:** The behavior displays a pattern of sustained attacks that combine both personal threats and terrifying actions that include death threats, serious harm wishes, and non-consensual sexual threats.
- **Denigration:** It involves spreading false information, rumors, or harmful statements about a person to damage their reputation or character.
- **Masquerading:** It happens when a bully pretends to be the victim and sends disrespectful messages, which create the false belief that the victim sent those messages.
- **Trickery:** It occurs when someone tricks the victim into disclosing their private information, which the attacker uses to inflict damage or public humiliation.
- **Exclusion:** It refers to the intentional act of excluding a person from online communities through methods that

include forceful methods, blackmail, and social manipulation techniques.

- **Doxing:** It involves the public release of private or identifiable personal information, which includes documents, contact details, and family-related data through social media platforms [8].

These eight categories collectively represent the commonly recognized forms of cyberbullying in social media.

A variety of methodologies have been implemented for the identification of hate speech, encompassing traditional classifiers, neural network-based classifiers, or a hybrid of both methodologies [9]. Deep Learning (DL) is a technique utilized in Machine Learning (ML) that facilitates unsupervised learning with unlabeled data. In the domains of text classification and data mining, numerous research studies have employed DL approaches to predict and classify issues, including the detection of hate speech and the classification of opinions [10].

1.1. Problem Statement

Cyberbullying has developed into a major social and technological issue, which results in severe psychological, emotional, and social harm to its victims. The increasing presence of cyberbullying on online platforms has created multiple distinct forms of the behavior, which include harassment and impersonation, exclusion and privacy violations, yet existing studies and detection systems treat these forms in an inconsistent or ambiguous manner. The absence of a defined, organized framework for analyzing different forms of cyberbullying hinders the successful identification, prevention, and mitigation efforts. The existing issue requires researchers to create comprehensive systems that will accurately identify and describe cyberbullying with advanced DL models.

1.2. Research Novelty and Objectives

This study proposes a novel unified integration of an MSCT for hierarchical feature extraction, a CDMO algorithm exclusively for optimal feature selection, and a TFT for robust text classification. The proposed framework uses a chaos-enhanced metaheuristic feature selection stage that links deep feature extraction methods to the classification process to achieve better results than existing approaches, which depend on transformer outputs or conduct joint optimization of their classifiers. This MSCT-CDMO-TFT model enables improved generalization, balanced performance across metrics, and superior handling of both sentiment analysis and multi-class CBD tasks.

- To develop a multi-scale contextual feature extraction mechanism capable of capturing both local and global semantic information from textual data.
- To design and integrate a metaheuristic feature selection approach using CDMO to reduce feature dimensionality and redundancy.

- To employ a TFT model as an effective classifier for modeling complex contextual dependencies in text data.
- To evaluate the proposed framework on benchmark datasets, including IMDB, Yelp Polarity, and Cyberbullying Classification datasets.
- To analyze the effectiveness of each module through ablation and comparative studies.
- To compare the MSCT-CDMO-TFT model with current methodologies.
- To present and analyze the advantages and drawbacks of the MSCT-CDMO-TFT model with a conclusion and future enhancements.

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 MSCT-CDMO-TFT 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. Review of Literature

This part analyzes the current methodologies implemented for sentiment analysis and cyberbullying detection with various AI techniques. Table 1 includes a comprehensive analysis of the models analyzed, with advantages and drawbacks.

The study in [11] evaluated the effectiveness of three artificial intelligence techniques: Logistic Regression (LR), Random Forest (RF), and a Hybrid Model that combined both methods, in the field of multi-class detection of cyberbullying using a dataset obtained from Kaggle-ML repositories. The dataset was categorized into five classifications: Age, Gender, Ethnicity, Religion, and Not-Cyberbullying, subsequent to the training and validation of the models using a cyberbullying dataset. The Hybrid Model exhibited consistent performance across categories, particularly attaining better results in Not-Cyberbullying and Gender classifications.

An enhanced machine learning fuzzy logic methodology was proposed in [12] for detecting hateful speech in social media content. The methodology allowed the Fuzzy logic-based solution to manage the unpredictability and uncertainty included in the text. Fuzzy logic not only provides a methodology for addressing linguistic issues but also involves reasoning, enhancing comprehension of sentiment values within a sentence, so presenting a more dependable approach to managing linguistic challenges. Nonetheless, the research indicated that the LR-Fuzzy-Genetic Algorithm system consistently attained the most superior outcomes.

The research in [13] examined the impact of sentiment and emotional data on identifying cyberbullying from a code-

mixed tweets dataset. A multitask multimodal approach named MT-MM-Bert+VecMap, utilizing BERT and VecMap embedding techniques with emoji modality, was proposed for the detection of cyberbullying supported by emotion and sentiment analysis. The approach integrated BERT and VecMap to represent code-mixed texts efficiently. The findings indicated that the integration of BERT and Vec-Map for word vector representation was more effective in managing codemixed data than a singular representation method.

A Community-Guided Link Prediction Approach, Which Uses Sentimental Analysis (CLPSA), was developed in [14] to predict links between nodes in multilayered networks. The study showed that sentiment analysis helped people understand human emotional states better because users in the same cluster showed more willingness to connect with each other later. The Proposed Cluster Edge Betweenness method detected user clusters that showed high connection density between their members but maintained weak ties to their neighboring modules. The method established a network framework that promotes constructive connections and diminishes the frequency of negative comments.

A Cross-Platform Cyberbullying Detecting Algorithm (CPCDAlg), which was developed in [15], used multiple detection methods. The CPCD-Alg method uses DL and ML models for text analysis through BERT and network analysis through graph-based methodologies. Long Short-Term Memory (LSTM) demonstrated superiority as the most effective model for identifying cyberbullying from textual data. The analysis improved when behavioral factors and network interaction elements were included for analysis purposes. Graph-based methodologies, which include betweenness centrality, enable researchers to detect clusters that participate in such actions. The detection of cyberbullying requires researchers to include social relationships as essential elements for the work.

A Cyberbullying Detection System-Convolutional neural Network (CBDS-ConvNet) system was presented in [16], which operated in real time to detect and stop cyberbullying through its analysis of Twitter data. The research utilized Twitter as the case study because the platform generates high volumes of data every day, while it has become recognized as a site that experiences cyberbullying incidents. The research demonstrated that CNNs possess the ability to handle complex social media conversation patterns, which dynamically evolve throughout time. The CBD-ConvNet technology enabled the immediate detection of cyberbullying activities, which created an all-encompassing system that improved online protection while advancing research efforts to reduce cyberbullying incidents.

The research in [17] investigated how word embedding technology detects multiple types of cyberbullying. The study

assessed three word embedding methods through multiple testing approaches, which included CNN, Bidirectional-LSTM, word2vec, and BERT. The dataset testing showed that BERT model fine-tuning produced the best results. The CNN-BiLSTM model demonstrated successful cyberbullying detection and classification through its combination of word embeddings and DL techniques.

The research in [18] introduced methods for detecting cyberbullying that used contextual, emotional, and sentiment features as their training data. The development of an Emotion Detection Model (EDM) used advanced Twitter datasets, which contained improved labeling for its training. The EDM system and lexicon-based methods were used to extract emotions and sentiments from cyberbullying datasets. The study results showed that fear, anger, and guilt represented the main emotions that people experienced during cybersecurity bullying incidents. The retrieved emotions were subsequently utilized as features with contextual and sentiment information for training models for detecting cyberbullying. The results demonstrated that emotional features, together with sentiment information, improved detection accuracy for CBD systems.

A stacked ensemble classification method was proposed in [19], which successfully classified tweets into three different categories: neutral and abusive language and hate speech. The approach used Terms Frequency-Inverse Documents Frequency (TF-IDF), which was derived from Twitter texts to build Support Vector Machine (SVM) classifiers that combined three main classifiers, which are RF, LR, and XGBoost. The final model results combined multiple base models into an ensemble learning system. The model performance showed improvement when the collected features were removed, while the ensemble method demonstrated better performance.

The research in [20] analyzed five common machine learning techniques, which included multinomial Naive Bayes (NB), LR, Support Vector Classifier (SVC), Decision Tree (DT), and RF. Additionally, three DL methodologies, LSTM, Bi-LSTM, and Gated Recurrent Units (GRU), successfully identified cyberbullying incidents. The method produced better results because the Focal Loss algorithm reached its highest accuracy through this approach. The majority of algorithms displayed low recall values, which demonstrated their inability to detect all relevant information.

The CNN model was implemented in [21] for processing and an Attention Mechanism-based BiLSTM with GRU (AM-BLSTM-GRU) system to predict sarcasm detection and sentiment analysis results. The system extracted auxiliary features from natural language processing, which the CNN model used to create vectors of features. The AM-BLSTM-GRU approach executed both sarcastic speech identification and sentiment analysis tasks. An Enhanced Sinogramic Reds

Deer (ESRD) approach provided an effective method for adjusting classifier parameters. The results demonstrated that the model successfully identified online harmful activities and reduced their impact.

An ensemble DL method in [22] used a tournament-selected glowworm swarm optimizer (EDL-TSGSO) for accurate detection and classification of Twitter cyberbullying incidents. The CBD system used the EDL-TSGSO approach, which employed LSTM and Adaboost classifiers for its detection and classification functions. The ELSTM-AB classifier used LSTM and Adaboost model predictions to create a combined system, which achieved better performance in classifying multiple classes. The EDL-TSGSO method demonstrated high efficiency through its experimental results on the Twitter dataset.

The research in [23] identified cyberbullying through its hybrid model, which used Valence-Aware Dictionary and Sentiment Reasoner (VADER) together with SVM that GridSearchCV improved through optimization. VADER was selected because it reliably analyzes social media texts, which show informal writing styles through its capability to detect emotional content that includes both sarcasm and multiple emotional intensities.

The complex linguistic patterns that define cyberbullying situations needed this method because it improved data classification and analysis accuracy. The HyVADSVM model accurately classified cyberbullying information through its linguistic pattern identification, which included cyberbullying characteristics.

A cyberbullying tweet classification model was developed in [24], which used Robust Optimized Bidirectional Encoders Representations From Transformer Architecture (RoBERTa) with Global Vectors (GloVe) word representation features. The transformer-driven RoBERTa model utilized Word2Vec and GloVe features to achieve the best outcomes. The Word2vec features improved model performance according to the experimental findings. The RoBERTa reached an accuracy rate of 95.9%.

The research in [25] evaluated sentiment-enhanced methods, including A Lite BERT (ALBERT), Decoding-enhanced BERT with disentangled attention (DeBERTa), Efficiently Learning an Encoder that Classifies Token Replacements Accurately (ELECTRA), HateBERT, and DeepSeek-coder-1.3b-base, that have been fine-tuned for the identification of cyberbullying on social networking sites. The research assessed how emotional expressions in user input help detect cyberbullying while determining the best models that provide accurate and efficient detection with the ability to handle large-scale operations.

Table 1. Critical analysis of the reviewed current models

Ref.	Models	Application	Advantages	Disadvantages
[11]	LR, RF, and Hybrid LR-RF ML models	Multi-class cyberbullying detection using the Kaggle dataset	The hybrid model showed stable performance across multiple classes; it was effective for structured categories.	Limited semantic understanding; relies on handcrafted features; weak performance on implicit bullying.
[12]	Fuzzy logic integrated with LR and Genetic Algorithm	Hate speech detection in social media text	Effectively handles uncertainty and linguistic vagueness; improved sentiment reasoning.	High computational complexity; rule dependency limits scalability; not suitable for large-scale real-time systems.
[13]	MT-MM-BERT + VecMap (multitask, multimodal)	Cyberbullying detection in code-mixed tweets with emoji modality	Efficient handling of code-mixed data; improved emotion and sentiment representation.	Increased model complexity; dependent on multimodal data availability.
[14]	CLPSA	Fraudulent and cyberbullying comment detection in multilayer social networks	Captures social relationships; promotes positive user interactions through clustering.	Not directly applicable to text-only platforms; high graph construction overhead.
[15]	Cross-platform DL + ML (BERT, LSTM, graph-based methods)	Cyberbullying detection across multiple social networks	Integrates behavioral, textual, and network features; strong contextual modeling.	High computational cost; complex feature integration pipeline.
[16]	CBDS-ConvNet	Real-time cyberbullying detection on Twitter	Efficient real-time performance; handles large-scale streaming data.	Limited long-range dependency modeling; weaker contextual understanding.
[17]	CNN, BiLSTM, word2vec, BERT-based models	Multiclass cyberbullying detection	Fine-tuned BERT achieved superior accuracy and effective embedding utilization.	High dependency on labeled data; computationally intensive.
[18]	Emotion Detection Model + lexicon-based sentiment features	Emotion-aware cyberbullying detection	Improved detection accuracy using emotional cues; interpretable emotional patterns.	Lexicon dependency; limited adaptability to evolving slang.
[19]	Stacked ensemble (TF-IDF + SVM, RF, LR, XGBoost)	Hate speech and abusive language detection	Ensemble improves robustness and reduces reliance on manual features.	TF-IDF lacks semantic context; it has poor performance on sarcasm.
[20]	ML (NB, LR, SVC, DT, RF) and DL (LSTM, BiLSTM, GRU) with Focal Loss	Cyberbullying detection	Focal loss improves class imbalance handling; DL models outperform ML.	Low recall for minority classes; limited contextual reasoning.
[21]	CNN + Attention-based BiLSTM-GRU with ESRD optimizer	Sentiment and sarcasm detection	Attention improves feature focus; optimized parameters enhance accuracy.	High training complexity; optimizer sensitivity.
[22]	Ensemble DL (LSTM + AdaBoost) with TSGSO optimizer	Twitter-based cyberbullying detection	Ensemble learning improves classification accuracy.	Increased computational overhead; limited interpretability.
[23]	VADER + SVM with GridSearchCV	Linguistic pattern-based cyberbullying detection	Effective for informal and sarcastic text; simple and interpretable.	Lexicon-based sentiment limits deep semantic understanding.
[24]	RoBERTa combined with GloVe/Word2Vec embeddings	Cyberbullying tweet classification	High accuracy (95.9%); strong contextual representation.	Increased model size; embedding fusion adds complexity.
[25]	Fine-tuned transformer models (ALBERT, DeBERTa, ELECTRA, HateBERT)	Emotion-aware cyberbullying detection on social platforms	ELECTRA offers strong performance with low resources; it is scalable.	Some models struggle with informal language; domain adaptation is required.

3. Materials and Methods

The proposed model uses multi-scale contextual transformer feature extraction together with chaos-driven feature selection and temporal fusion transformer classification to achieve accurate and efficient sentiment analysis and CBD across multiple text datasets. The workflow of the proposed framework is shown in Figure 2.

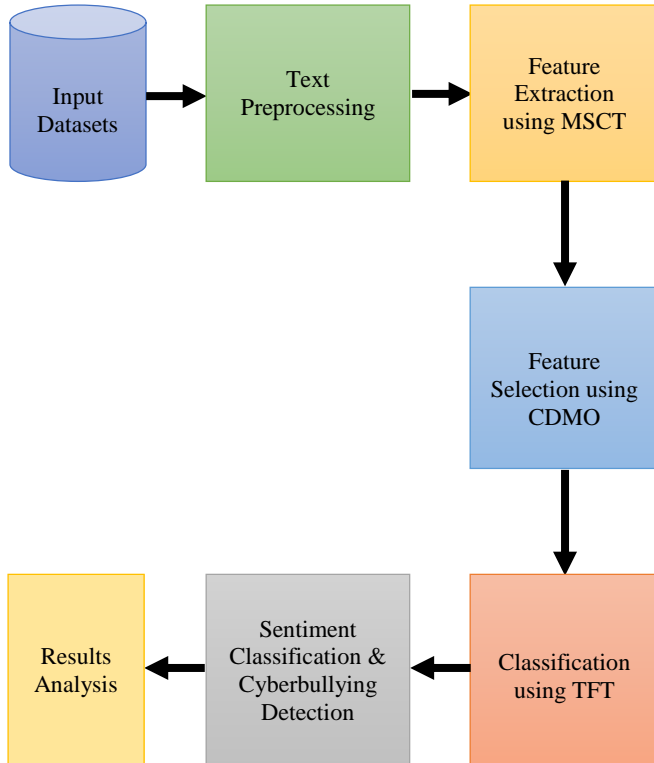


Fig. 2 Workflow of the proposed detection model

The proposed model workflow starts its process by combining three different datasets, which contain IMDB and Yelp Polarity and Cyberbullying Classification data for their respective purposes of sentiment analysis and cyberbullying detection. The raw textual data from these datasets first undergoes a complete text preprocessing process, which includes three steps: tokenization, stemming, and stop-word elimination. The process removes all irrelevant symbols together with short tokens and unnecessary linguistic elements.

The cleaned and preprocessed input data is fed to the MSCT feature extraction module, which extracts contextual embeddings from multiple transformer layers to create high-dimensional contextual features that represent syntactic, semantic, and sentence-level information. The CDMO algorithm achieves its primary purpose by selecting optimal features through its ability to maintain essential discriminative elements while decreasing redundancy and computational demands. The features selected are then fed into the TFT model, which successfully captures both long-term

dependencies and time-based relationships through its gated residual networks and attention systems. The framework is divided into two output layers, which handle specific tasks because they need to perform sentiment classification and fine-grained cyberbullying detection. The system achieves multi-task learning because the model can adapt to different methods of text analysis.

3.1. Details of Datasets

The proposed framework was assessed through experiments, which tested its effectiveness, robustness, and ability to generalize across multiple datasets, which included three benchmark datasets that contained sentiment analysis and cyberbullying detection tasks.

3.1.1. Internet Movie Database (IMDb) Dataset

The IMDB dataset contains textual data from 1000 movies and television shows, which includes user reviews and various metadata elements such as poster links, release year, runtime, certification, and genre. The research used the review text component to measure sentiment polarity for the assessment. The dataset contains various linguistic expressions that enable the testing of both long-range understanding and sentiment classification abilities [26].

3.1.2. Yelp Polarity Review Dataset

The Yelp Polarity Review dataset functions as a standard resource for binary sentiment classification. The dataset includes customer reviews, which have been categorized into two sentiment classes.

Negative reviews in the dataset are assigned to class 1, while positive reviews receive class 2. The dataset contains informal language and opinionated expressions with different review lengths, which make it suitable for testing the model's capability to detect sentiment cues and polarity changes [27].

3.1.3. Cyberbullying Classification Dataset

The cyberbullying classification dataset contains about 47,000 tweets, which researchers organized into six different categories. The six categories include age-based cyberbullying, ethnicity-based cyberbullying, gender-based cyberbullying, religion-based cyberbullying, other types of cyberbullying, and non-cyberbullying content. Each category contains nearly 8000 samples, which creates a balanced class distribution. The dataset demonstrates linguistic diversity together with social media platforms' typical vagueness and concealed offensive behavior, which makes it an effective benchmark for multi-class cyberbullying detection [28].

The proposed framework uses three datasets to evaluate its performance in binary sentiment classification, multi-class sentiment analysis, and fine-grained cyberbullying detection, which shows its ability to perform well in different application cases. The set of raw reviews before and after text preprocessing is illustrated in Table 2.

Table 2. Example reviews from the dataset

Before preprocessing	After preprocessing
“It seems like every other time I take my pets to Point Breeze, there is unnecessary frustration.”	Seems every other time take pet point breeze unnecessary frustration
“BED BUGS!!! Horrible place! DO NOT STAY HERE!! Stayed here for a wedding we attended.”	bed bug, horrible place, stayed here, attended a wedding
“@JoeBiden No Joe, YOU are the RACIST. They have so many video and audio clips of the RACIST shit you.”	Joe Biden is a racist, with many racist audio and video clips.
“Trump is like that jealous little school bully who cannot stand it when everybody loves the other guy.”	Trump, a jealous little school bully, stands everybody loves the other guy.
“The Hollywood movie industry is the laziest one in the entire world. It only needs a single hit to flood...”	The Hollywood movie industry is the laziest in the entire world, and it needs a single hit flood.
“I just watched this movie on its premier night out of curiosity and sheer nostalgia.”	watched movie premier night, curiosity, sheer nostalgia

3.2. Text Preprocessing

Textual data collected from online review platforms and social media sources show inherent noise because they contain redundant tokens and irrelevant symbols, and their linguistic structures remain inconsistent.

The irregularities present in the data will decrease the effectiveness of feature extraction and classification processes. The process uses a complete text preprocessing system to transform raw input text into normalized form while reducing dimensionality and maintaining essential meaning.

Let the raw textual input be represented as $R = \{r_1, r_2, \dots, r_N\}$. Where r_i denotes the i th raw review or tweet in the dataset, and N is the total number of text samples.

3.2.1. Tokenization

The process of preprocessing begins with tokenization, which breaks down each raw text instance r_i into distinct lexical units that serve as tokens. The formal expression of tokenization exists in the following equation.

$$T_i = \{w_{i1}, w_{i2}, \dots, w_{iM_i}\} \quad (1)$$

The variable w_{ij} denotes the j th token of the i th text sample while M_i represents the complete token count of r_i . The process of tokenization enables detailed language analysis because it converts complete text into separate elements.

The process of removing unneeded data includes deleting all tokens that contain special characters and punctuation marks, numeric digits, web links, and user identification tags. The process of removing three-character tokens from the system continues because these tokens offer minimal semantic value. The filtered token set is defined as follows:

$$T_i^{(f)} = \{w \in T_i | |w| \geq 3 \wedge w \notin S\} \quad (2)$$

Here, $|w|$ denotes token length and S represents the set of noisy symbols and non-linguistic elements.

3.2.2. Stop Word Removal

Stop words are high-frequency terms that contribute less to semantic interpretation. The process of stop word removal begins with common stop words, which include the terms "is," "the," "we," "I," and "an" to achieve lower dimensionality and enhanced computational efficiency.

The stop-word vocabulary can be represented by the variable W_{stop} .

$$T_i^{(s)} = T_i^{(f)} / W_{stop} \quad (3)$$

This process ensures that only semantically informative tokens are retained.

3.2.3. Stemming

Stemming functions to reduce words into their base forms, which helps to standardize different word forms. The words "acting" and "acted" transform into the base form "act". The equation of the stemming operation stands as follows:

$$T_i^{(stem)} = \{stem(w) | w \in T_i^{(s)}\} \quad (4)$$

The stemming function $stem(\cdot)$ provides a way to convert words into their base forms. This process reduces vocabulary sparsity while maintaining core semantic meaning. The text sample from the preprocessing steps produces a cleaned token sequence that represents the sample.

$$P_i = (w'_{i1}, w'_{i2}, \dots, w'_{iK_i}) \quad (5)$$

The i th text instance reaches its final preprocessed state through P_i which shows that K_i has experienced major dimensionality reduction compared to M_i [29].

The preprocessed text functions as input for the MSCT feature extraction module, which achieves efficient contextual representation learning through its ability to decrease noise and enhance semantic coherence.

3.3. Multi-Scale Contextual Transformer (MSCT)

Traditional feature extraction methods, such as bag-of-words, TF-IDF, and hash vectorization depend on two elements, which are token frequency and local contextual data, to extract text features. This dependency results in an inability to capture extended semantic relationships together with detailed contextual aspects that exist in text containing both sentiment and cyberbullying content. This study presents the MSCT-based feature extraction mechanism, which uses pretrained transformer architectures to learn hierarchical representations as a solution to the existing limitations.

The MSCT model extracts detailed text representations that show distinct characteristics because it uses the hierarchical structure of contextual information that transformer models have learned.

MSCT extracts linguistic information through its method of aggregating transformer layer representations, which enables the system to capture multiple semantic levels of language while it processes syntactic structures, phrase-level meanings, and overall contextual relationships.

The multi-scale aggregation method enables the framework to achieve complete sentiment polarity analysis, together with discovering hidden cyberbullying indicators in complex informal text. The MSCT system establishes a robust feature representation system that maintains performance throughout the entire process by preserving long-range dependencies and contextual details while avoiding information loss [30].

Let the preprocessed text sequence be expressed as $P_i = \{w_{i1}, w_{i2}, \dots, w_{iL}\}$ where L indicates the tokenized input sequence's length for the i th text sample.

3.3.1. Token Embedding and Positional Encoding

Every token w_{ij} was mapped to the representation of a dense vector using an embedding function as follows.

$$E_{ij} = Emb(w_{ij}) \quad (6)$$

Here, $E_{ij} \in \mathbb{R}^d$ Moreover, d indicates the embedding dimensions. For preserving the sequential order information, the positional encodings were added as follows.

$$Z_{ij} = E_{ij} + PE_j \quad (7)$$

Here, the variable PE_j indicates the positional embedding corresponding to the j th token position.

3.3.2. Transformer Encoding

The embedded input sequences are passed through the stack of K transformer encoding layers. The output of the k th layer is expressed as follows.

$$H^{(k)} = Transformer^{(k)}(Z) \quad (8)$$

Here, $H^{(k)} \in \mathbb{R}^{L \times d}$ includes the contextual representations at diverse semantic levels.

Every transformer layer employs the multi-head self-attention, which is computed using the following equation.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (9)$$

Here, the variables Q, K, V were the query, key, and value matrices derived from $H^{(k-1)}$, and d_k indicates the dimensionality of the key vectors.

Different transformer layers capture information at varying semantic scales:

- The first layer of a system processes basic linguistic elements through its encoding of syntactic and lexical patterns.
- The intermediate layers of a system process language through their ability to understand phrase-level meaning.
- The higher layers of a system process language by understanding complete sentence-level contextual information.

To construct the multi-scale features, the outputs from the selected layers $\{k_1, k_2, \dots, k_m\}$ were aggregated as follows.

$$F_{ms} = \sum_{j=1}^m Pool\left(H^{(k_j)}\right) \quad (10)$$

Here, $Pool(\cdot)$ indicates the mean or attention-based pooling, and summation denotes vector concatenation. The method aggregates information to create a single feature vector that maintains contextual details from different linguistic levels.

3.3.3. Dimensionality Consolidation

To maintain computational efficiency, the concatenated feature vector is projected into a compact representation as follows.

$$F_c = W_p F_{ms} + b_p \quad (11)$$

Here, $W_p \in \mathbb{R}^{d' \times d_{ms}}$ and b_p were the learnable parameters and the variable d' indicates the reduced feature dimension [31].

The resulting contextual feature vector F_c serves as the high-dimensional semantic representation of the input text,

which includes all its long-range dependencies, emotional signs, and contextual relationships.

The extracted features proceed to the CDMO module, which performs optimal feature selection before the classification process.

3.4. Chaotic Dwarf Mongoose Optimizer

DMO functions as a metaheuristic method that imitates the hunting behaviors of the dwarf mongoose through its adaptive changes. The mongoose exhibits two primary compensating behavioral adaptations: a) group and prey sizes, and space usage. b). Foods Supply.

In the initialization of the population, the population of the mongoose candidates (X) was determined utilizing Equation (12). The population is created stochastically within the Lower (LB) and Upper Bounds (UB) of the specified problems.

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,d-1} & x_{1,d} \\ x_{2,1} & x_{2,2} & \dots & x_{2,d-1} & x_{2,d} \\ \vdots & \vdots & x_{i,j} & \vdots & \vdots \\ x_{n,1} & x_{n,2} & \dots & x_{n,d-1} & x_{n,d} \end{bmatrix} \quad (12)$$

Here, X represents the populations, created randomly according to Equation (12); $x_{i,j}$ denotes the position of the dimension j within the population i , n indicates the population size, and d signifies the problem dimensions.

$$x_{i,j} = VarMin + rand \times (VarMax - VarMin) \quad (13)$$

Here, $rand$ is a stochastic variable within the interval $[0, 1]$, $VarMin$ and $VarMax$ denote the lower and upper limits of an issue. The optimal solutions through iteration were the most beneficial solutions achieved so far.

The fitness of all solutions is assessed following the initialization of the population. Equation (14) computes the probability values for every population's fitness, and the alpha female (α) was selected according to the probability.

$$\alpha = \frac{fit_i}{\sum_{i=1}^n fit_i} \quad (14)$$

The n -bs corresponds to the quantity of mongoose within the alpha groups. The DMO utilizes the expression from Equation (15) to determine the prospective food location.

$$X_{i+1} = X_i + phi \times peep \quad (15)$$

Here, phi was a distributed random variable uniformly in the interval $[-1,1]$; $peep$ was the sound of the alpha female that leads the family. After every iteration, the sleeping mound was defined as per Equation (16).

$$sm_i = \frac{fit_{i+1} - fit_i}{\max\{|fit_{i+1}, fit_i|\}} \quad (16)$$

The mean value of the identified sleeping mound is specified by Equation (17).

$$\varphi = \frac{\sum_{i=1}^n sm_i}{n} \quad (17)$$

Mongoosees are recognized for their tendency to refrain from returning to prior sleeping mounds, leading scouts to seek out new ones to facilitate exploring. The scout mongoose was represented by Equation (18).

$$X_{i+1} = \begin{cases} X_i - CF \times phi \times rand \times [X_i - \vec{M}] & \text{if } \varphi_{i+1} > \varphi_i \\ X_i + CF \times phi \times rand \times [X_i - \vec{M}] & \text{elsewhere} \end{cases} \quad (18)$$

Here, CF denotes the variable that diminishes linearly with every iteration, thereby regulating the group's aggregate volatile movement. \vec{M} were the vectors that direct the mongoose's motions to its newer sleeping mounds.

Chaos is an occurrence that could demonstrate non-linear variations in future behavior once its initial state is minimally modified. Furthermore, it was characterized as the semi-random behavior produced by the non-linear deterministic frameworks. The Chaos Optimization Algorithm (COA) is a principal search algorithm that transitions parameters and variables from chaos to the solution space. It depends on identifying the global optimal for regular, stochastic, and periodically chaotic motions characteristics. Owing to its efficiency and rapid convergence, COA has been extensively utilized in numerous studies. To acquire the chaotic sets, the study utilized ten established one-dimensional maps that are commonly employed in several studies. The maps are Chebyshev, Circle, Gauss/Mouse, Iterative, Logistic, Piecewise, Sine, Singer, Sinusoidal, and Tent.

This study employs the CDMO for the feature selection process. Random numbers utilized in Equation (18) are substituted with chaotic maps to prevent returning to the same sleeping state.

$$X_{i+1} = \begin{cases} X_i - CF \times phi \times \rho \times [X_i - \vec{M}] & \text{if } \varphi_{i+1} > \varphi_i \\ X_i + CF \times phi \times \rho \times [X_i - \vec{M}] & \text{else} \end{cases} \quad (19)$$

Here, ρ was a value attained from chaotic maps. Subsequently, the study established the dimensionality of the issue, denoted as d in Equation (12), corresponding to the number of features, and assigned $VarMin$ and $VarMax$ in Equation (13) the values of 0 and 1, respectively. Every row in Equation (12) (i.e., the location of every element in X_i) is subjected to a threshold of 0.5, as the values range from 0 to 1.

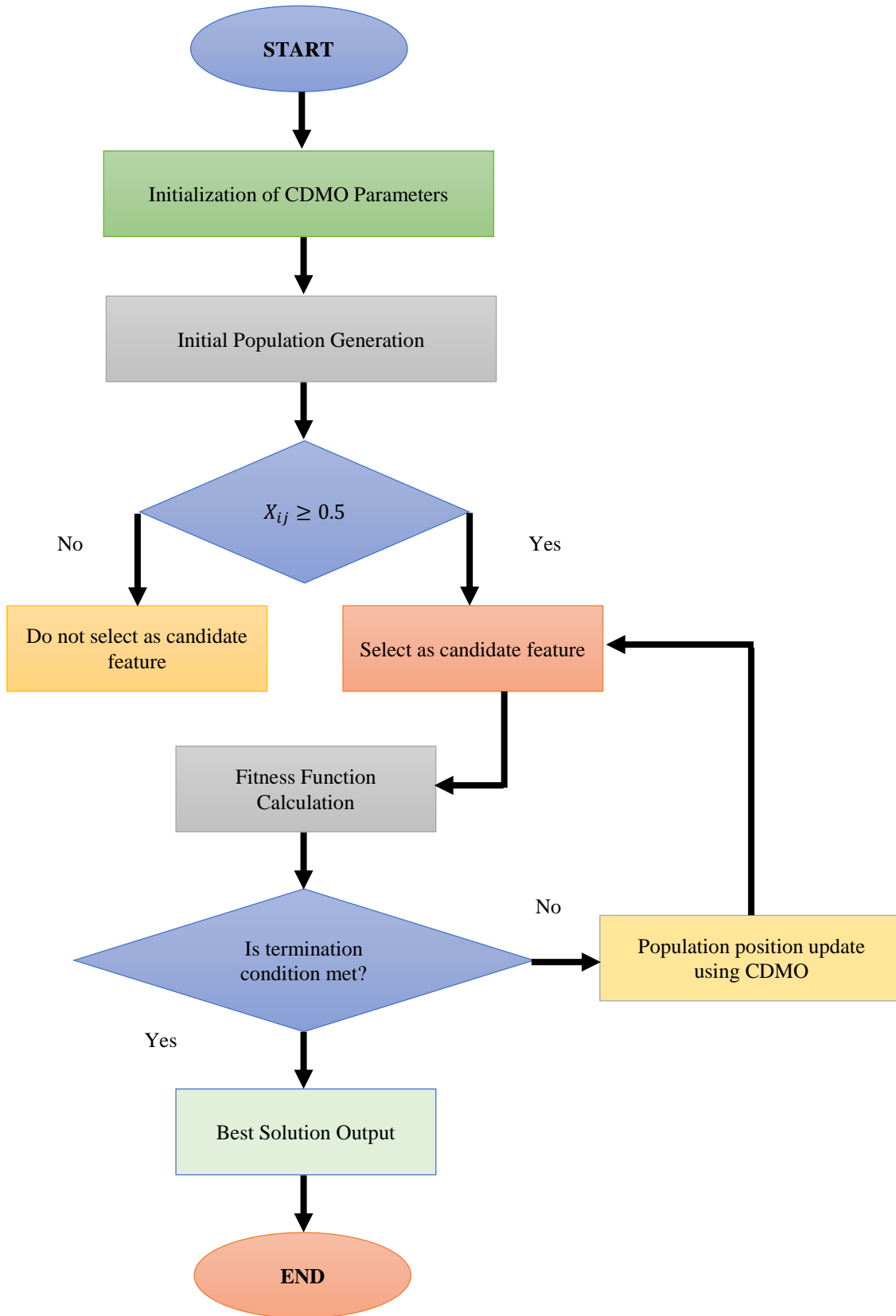


Fig. 3 Flowchart of CDMO

$$X_{i,j} = \begin{cases} 1 & x_{i,j} > 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

Subsequently, elements with positions over 0.5 are regarded as candidate features, but elements with positions below 0.5 are excluded from this solution.

The candidate features were subsequently utilized in the fitness functions, which compute the accuracy of classification of the k-NN classification method based on the employed features of the candidate.

The Fitness Function (FF) is defined as a weighted combination of classification accuracy and feature reduction rate:

$$FF = \alpha(1 - Accuracy) + \beta \frac{|X|}{d} \quad (21)$$

The equation uses control parameters α and β , which need to satisfy $\alpha + \beta = 1$, and d represents the total count of extracted features, while $|X|$ shows the number of features that have been chosen [32].

The logistic maps create chaotic sequences that direct the movement updates of the scout, alpha, and babysitter mongooses. The process of CDMO provides an optimized feature subset that the TFT model uses for its classification tasks. The integration of CDMO improves feature relevance, minimizes redundancy, and enhances classification efficiency, which makes it particularly suitable for high-dimensional contextual representations.

3.5. Temporal Fusion Transformer (TFT)

The TFT deep learning architecture uses its combined components, which include recurrent networks, attention mechanisms, and gated feature transformations, to build a unified system that enables users to understand its functionality. The system employs LSTM layers to model short-term time dependencies while using multi-head self-attention to track extended range context relationships, which enables it to extract both immediate and complete time patterns from sequential data. The system uses gated residual networks to control information flow, which enables the model to extract significant features through the processing of nonessential artifacts while keeping their chosen feature set intact. TFT works effectively as a classifier for sentiment analysis and cyberbullying detection because of its modular design, robust performance, and ability to interpret results. Thus, making it suitable for handling classification tasks that require processing contextual and sequential text data [33].

The research uses TFT as its classification module, which operates on the optimal feature subset that the CDMO algorithm selected. The CDMO system trains the TFT system through standard gradient-based methods because it does not

perform any optimization for its architecture or operational parameters. The separation of these elements provides three benefits because it maintains model stability and reproducibility, and computational efficiency, while CDMO uses informative feature selection to improve classification results.

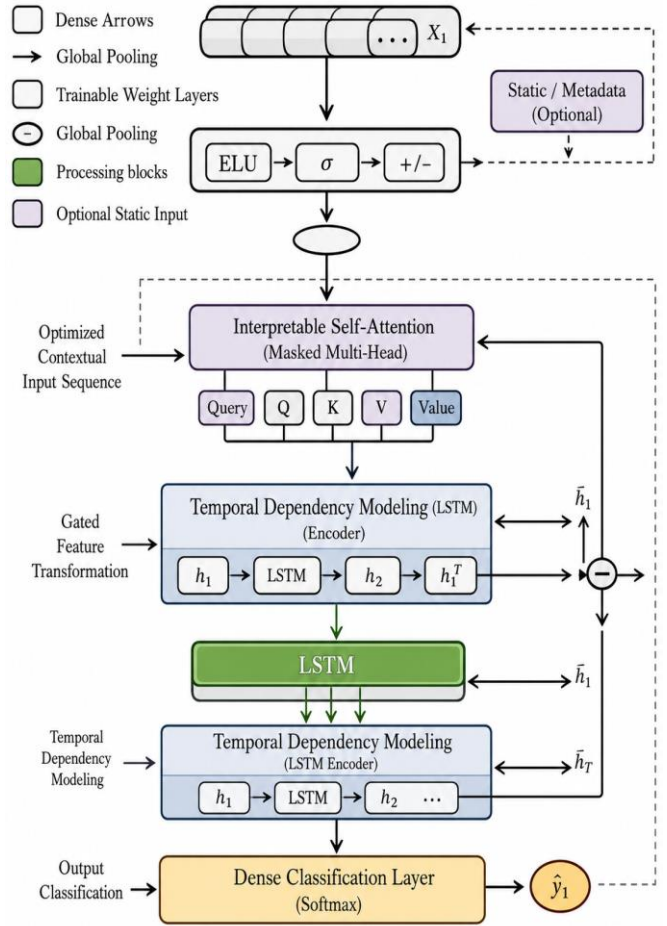


Fig. 4 TFT architecture

The TFT architecture shown in Figure 4 presents its optimized contextual features, which move sequentially toward their final classification results. The model receives its input through selected contextual features, which were chosen by CDMO. The features first undergo processing through gated feature transformation blocks, which control information distribution by using activation and gating methods to filter important features. At this point, the process allows for the combination of optional static data and metadata inputs to improve contextual understanding. The model transfers gated features to LSTM-based temporal dependency modeling layers, which will identify both short-term and long-range sequential patterns. The interpretable multi-head self-attention mechanism analyzes the temporal representations to detect which contextual elements in the sequence hold the highest importance. The attention-weighted outputs are

aggregated and forwarded to the dense classification layers with softmax activation functions, which generate the final class probabilities necessary for precise sentiment and cyberbullying classification while maintaining system interpretability and robustness.

Let the feature vector selected by CDMO for the i th input sample be denoted as follows:

$$X_i = \{x_{i1}, x_{i2}, \dots, x_{iT}\}, \quad x_{it} \in \mathbb{R}^{d_s} \quad (22)$$

The equation shows that d_s represents the reduced feature dimension which results from CDMO-based selection, and T shows the length of contextual sequences that MSCT produces.

3.5.1. Gated Feature Transformation

TFT uses Gated Residual Networks to control which features will be used and to eliminate unnecessary data.

$$GRN(x) = LayerNorm \left(x + \sigma(W_g x + b_g) \odot ELU(W_f x + b_f) \right) \quad (23)$$

The equation defines W_g, W_f and b_g, b_f as trainable parameters and $\sigma(\cdot)$ serves as the sigmoid activation function. The gating mechanism enables adaptive feature distinction between important and unimportant features while keeping CDMO feature selection intact.

3.5.2. Temporal Dependency Modelling

The gated feature sequence is processed by an LSTM-based temporal encoder:

$$h_t = LSTM(x_t, h_{t-1}) \quad (24)$$

The system uses h_t to extract both short-term and long-term relationships from the contextual feature sequence. The system enables TFT to acquire and learn time-based patterns that reveal the presence of sentiment polarity and cyberbullying activities.

3.5.3. Interpretable Self-Attention

The multi-head self-attention mechanism of TFT enables the modelling of long-distance dependencies while providing better interpretability according to the formula which follows.

$$Attention(Q, K, V) = softmax \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (25)$$

Here, the variables Q, K , and V were the LSTM hidden state projections. Attention scores are important for highlighting influential context segments that contribute most significantly to the final classification decision.

3.5.4. Output Classification

The attention-weighted temporal representation is

aggregated and fed to the dense classification layer for further processing.

$$\hat{y}_i = softmax(W_c \bar{h} + b_c) \quad (26)$$

The variables W_c and b_c were learned through the backpropagation process while \hat{y}_i shows the expected class probabilities [34]. The system supports binary classification for IMDB and Yelp datasets, and it also handles multi-class classification for cyberbullying classification tasks.

Algorithm: MSCT-CDMO-TFT-Based Sentiment and Cyberbullying Classification

```

Input: Raw textual dataset; Pretrained transformer model;
CDMO parameters (population size, iterations, threshold);
TFT hyperparameters
Output: Predicted class labels
Initialize
Text Preprocessing
For each raw text sample
    • Perform tokenization
    • Remove stop words, punctuation, digits, and special
    symbols
    • Apply stemming
    • Obtain preprocessed text
Multi-Scale Contextual Transformer Feature Extraction
(MSCT)
Encode each
Pi using the pretrained transformer
Extract contextual representations from multiple layers
Apply pooling to each selected layer output
Concatenate pooled representations to obtain a multi-scale
feature vector
Chaotic Dwarf Mongoose Optimization (CDMO) for Feature
Selection
Initialize the CDMO population randomly
While the termination condition is not met:
    • For each candidate solution
        – If Xj
        – Else, discard feature
    • Evaluate fitness using classification accuracy and feature
    penalty
    • Update population positions using chaotic mapping
Select optimal feature subset
Temporal Fusion Transformer (TFT) Classification
Input selected features into the TFT
Apply gated residual networks for feature transformation
Model temporal dependencies using LSTM layers
Apply multi-head self-attention for contextual weighting
Generate class probabilities using a softmax layer
Model Training and Evaluation
Train TFT using backpropagation and Adam optimizer
Evaluate performance using Accuracy, Precision, Recall, and
F1-score
End
    
```

The pseudocode highlights the proposed hybrid framework that starts with text data preprocessing through three steps, which include cleaning, tokenization, and lemmatization. The system performs an 80:20 data split to separate the datasets into training and testing sets. The processed text transforms into embeddings, which the system uses to evaluate multi-scale features that depend on the DAMSFE module. The FAOA algorithm optimizes these features by selecting the most informative features from the candidate feature subsets through its iterative process. The HAN-TE classifier processes the optimized features through transformer-based encoding and hierarchical attention mechanisms to produce contextual document representations used in classification. The model training process uses cross-entropy loss, while the evaluation on the test set employs standard performance metrics.

Table 3. Hyperparameter details

Module	Hyperparameter	Value
MSCT Feature Extraction	Pretrained transformer	BERT-base
	Embedding dimension	768
	Number of transformer layers	12
	Selected layers for multi-scale fusion	{4, 8, 12}
	Pooling method	Mean pooling
	Multi-scale feature dimension	2304
CDMO Feature Selection	Population size	30
	Maximum iterations	50
	Selection threshold ($X_{i,j}$)	0.5
	Chaotic map	Logistic map
	Fitness function	Accuracy + feature penalty
TFT Classifier	Input feature dimension	Selected by CDMO
	LSTM hidden units	128
	Number of LSTM layers	1
	Attention heads	4
	Dropout rate	0.2
	Activation function	ELU
Training Parameters	Optimizer	Adam
	Learning rate	0.001
	Batch size	32
	Epochs	50
	Loss function	Cross-entropy
	Validation split	80:20

The proposed pseudocode describes a sequential system that starts with the preprocessing of raw text. The preprocessing process involves two steps. The first step

requires tokenization and noise removal, stop-word elimination, and stemming to create clean input text. The preprocessed text is fed to MSCT so it can extract detailed contextual information through its multiple transformer layer feature aggregation method. The CDMO algorithm operates through feature selection, which selects optimal feature subsets that balance classification performance with feature compactness according to a fitness function. The selected features are fed into the TFT, which uses its gated feature transformation, temporal dependency modelling and attention-based weighting to create final class predictions through its softmax classifier, while the standard classification metrics assess the model's entire performance.

4. Results and Discussion

4.1. Experimental Setup

The experimental analysis tests the performance of the MSCT-CDMO-TFT framework through three publicly available datasets, which include IMDB, Yelp Polarity, and Cyberbullying Classification datasets. The study performed the experiments using Python, while the applied datasets were split to evaluate the proposed model through fair training and testing procedures. The proposed model is developed using Python 3.8 on PyTorch 2.0, and the experiments are conducted on a system that has 12 GB of operating memory. The system used for the experiments included an Intel(R) i5 CPU @3.20 GHz processor together with an NVIDIA GTX 2070 graphics processing unit.

4.2. Evaluation Metrics

To evaluate the performance of the MSCT-CDMO-TFT 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 is a reliable classification parameter. It measures classification accuracy through its dual assessment of false positive and false negative errors.

4.3. Discussion of Results

This segment presents the detailed analysis of the proposed MSCT-CDMO-TFT 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 MSCT-CDMO-TFT model on IMDB dataset

Metrics	Training Data	Test Data
Accuracy	96.84	96.12
Recall	96.27	95.68
Precision	96.91	96.03
F1	96.58	95.89

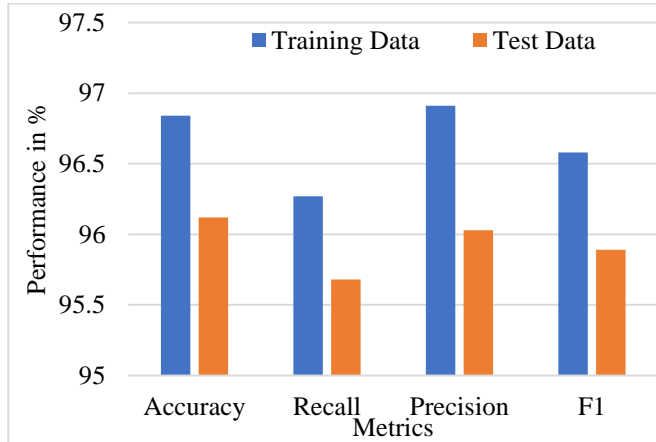


Fig. 5 Graph of MSCT-CDMO-TFT model's results on IMDB dataset

The proposed MSCT-CDMO-TFT model achieved high classification results on the IMDB dataset according to Table 4, which shows its performance in both training and testing stages. The model demonstrates strong generalization capabilities because it achieves an accuracy of 96.84% in training and an accuracy of 96.12% in testing. The small performance drop between these two metrics demonstrates that the model maintains effective control overfitting. The model achieves its highest performance through precision, which measures 96.91% for training and 96.03% for testing. This measurement shows how well the model detects positive sentiment while reducing incorrect positive results. Note that the training recall is 96.27% and the testing recall is 95.68%. This indicates that the model successfully captures most of the sentiment-related patterns relevant to the task, avoiding missing important examples. The F1-score, which combines precision and recall, shows strong performance with 96.58% accuracy during training and 95.89% accuracy during testing, which demonstrates the system's ability to classify different cases. The results show that the combination of multi-scale contextual feature extraction with optimized feature selection and temporal modelling achieves effective results. This demonstrates that the proposed framework correctly performs sentiment classification on large textual datasets. Figure 5 depicts the graph of the proposed model's result using the IMDB dataset.

Table 5. Results of the MSCT-CDMO-TFT model on the Yelp dataset

Metrics	Training Data	Test Data
Accuracy	96.93	96.41
Recall	96.58	96.07
Precision	96.87	96.29
F1	96.72	96.18

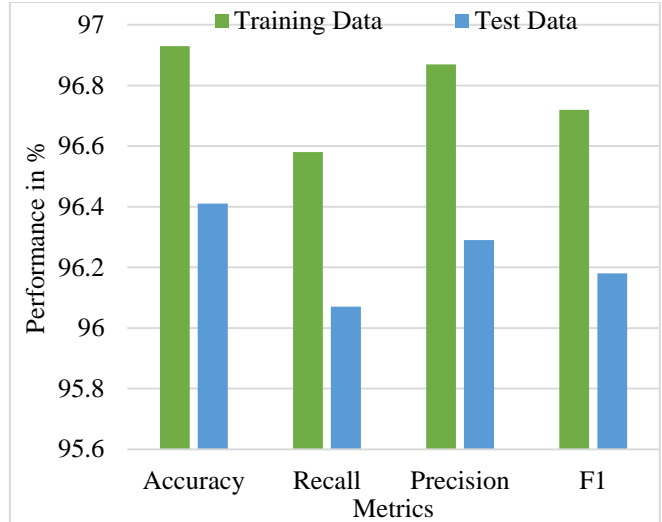


Fig. 6 Graph of MSCT-CDMO-TFT model's results on the Yelp dataset

The performance of the MSCT-CDMO-TFT model on the Yelp Polarity dataset shows its ability to maintain high performance during both training and testing stages, according to Table 5. The training accuracy of 96.93% and testing accuracy of 96.41% demonstrate that the system learned sentiment patterns effectively while maintaining performance ability, which shows the system's strength and ability to perform in multiple scenarios. The model demonstrates its capacity to accurately identify positive and negative reviews because it achieved training precision of 96.87% and testing precision of 96.29%, which demonstrates its ability to control false positive results that are vital for sentiment analysis applications. The model demonstrates its ability to identify the most relevant sentiment cases through its testing results, which achieved a 96.07% score, and its training results, which achieved a 96.58% score. Balanced F1-score is 96.72% in training, and 96.18% is more stable, even at 96.18% for testing, where the precision-recall trade-off is not close when compared to one another. The results demonstrate that multi-scale contextual feature extraction combined with optimized feature selection successfully processes large-scale polarity-based reviews while the model achieves high-accuracy results in sentiment classification. Figure 6 depicts the graph of the proposed model's result using the Yelp dataset.

Table 6. Results of the MSCT-CDMO-TFT model on the cyberbullying classifying dataset

Metrics	Training Data	Test Data
Accuracy	98.86	98.51
Recall	98.64	98.47
Precision	98.78	98.39
F1	98.71	98.44

In Table 6, the proposed MSCT-CDMO-TFT model shows its performance results on the Cyberbullying Classification dataset between its training and testing phases, with exceptionally high and consistent classification results.

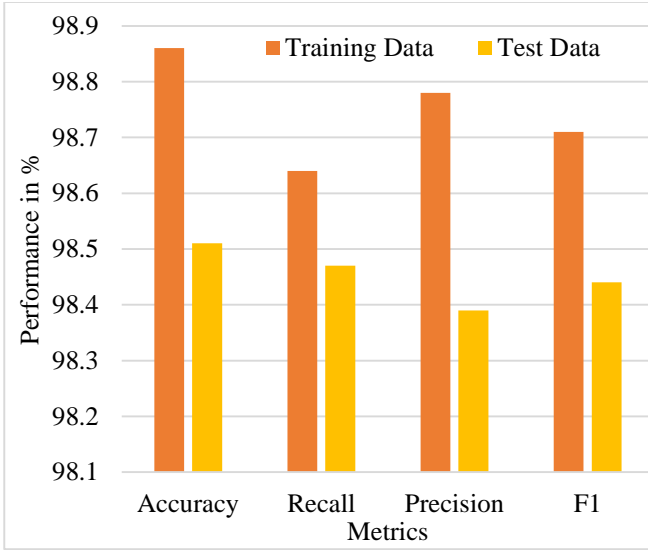


Fig. 7 Graph of MSCT-CDMO-TFT model’s results on the cyberbullying classifying dataset

The model demonstrates strong learning capacity through its training achievement of 98.86% accuracy and its test

performance, which reached 98.51% accuracy. The system maintained its performance with only slight reductions because it faced the difficult task of detecting multiple types of cyberbullying. The precision values of 98.78% for training and 98.39% for testing highlight the model’s effectiveness in accurately identifying cyberbullying-related instances while minimizing false positive classifications, which is critical for preventing incorrect flagging of benign content. The model demonstrates success in detecting actual cyberbullying instances, according to its recall scores of 98.64% for training and 98.47% for testing, which also guarantees that no harmful content will be missed. The F1-score, which combines precision and recall, maintains a high value of 98.71% during training and 98.44% during testing, which demonstrates strong classification performance across all tested cases. The results demonstrate that the combination of multi-scale contextual feature extraction with chaos-driven feature selection and temporal modelling produces an effective method for cyberbullying detection, which the proposed framework proves to work on large social media datasets. Figure 7 depicts the graph of the proposed model’s result using the cyberbullying classifying dataset.

Table 7. Multiclass results of the MSCT-CDMO-TFT model on the cyberbullying classifying dataset

Category	Accuracy	Recall	Precision	F1
Age	98.62	98.41	98.35	98.48
Religion	98.44	98.32	98.27	98.39
Gender	98.53	98.36	98.31	98.43
Ethnicity	98.38	98.29	98.21	98.33
Not Cyberbullying	98.67	98.49	98.44	98.55
Other Cyberbullying	98.31	98.23	98.12	98.22
Average	98.49	98.35	98.28	98.40

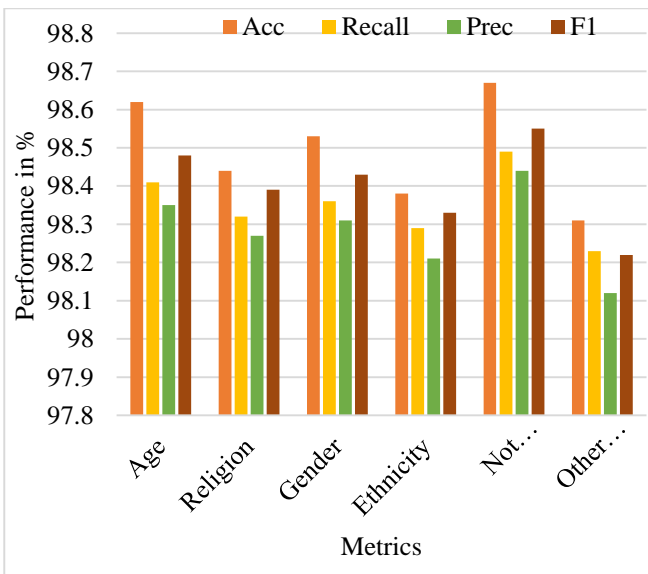


Fig. 8 Graph of MSCT-CDMO-TFT model’s multiclass results

Table 7 shows that the MSCT-CDMO-TFT model achieves its best performance on the Cyberbullying

Classification dataset through its ability to accurately classify all six designated categories. The model achieves its highest performance level for the Not Cyberbullying category because it reaches an accuracy of 98.67% and an F1-score of 98.55%, which shows its ability to identify safe content while minimizing incorrect alerts. The model demonstrates high accuracy and balanced precision-recall performance for Age and Gender classes because it successfully detects specific linguistic features that link to these cyberbullying types. The performance of Ethnicity and Other Cyberbullying shows a slight decrease from its previous level, yet maintains a strong performance because these two categories display more concealed contextual patterns and shared language features. Constant and well-balanced progress is being obtained by the Religion class in dealing with the representation of sensitive and subtle expressions. The framework demonstrates consistent and reliable performance through its average scores of 98.49% accuracy, 98.35% recall, 98.28% precision, and 98.40% F1-score across different cyberbullying types. The system demonstrates effective performance through its multi-scale contextual feature extraction and chaos-driven feature selection and temporal modelling techniques, which solve

complex multiclass cyberbullying detection challenges. Figure 8 depicts the graph of the proposed model's multiclass results using the cyberbullying classifying dataset.

The results comparison in Table 8 demonstrates that the proposed MSCT-CDMO-TFT model outperforms various machine learning methods, fuzzy systems, deep learning techniques, and transformer-based methods. The early hybrid and fuzzy models, which include Hybrid LR-RF and LR-Fuzzy-GA and NB-Fuzzy-GA, achieve moderate accuracy results, which range from 90% to 93%, but their performance suffers because their recall rates remain low, which shows their inability to detect intricate contextual patterns. The accuracy of MT-BERT shows low performance because researchers face difficulties when they try to use generic pretrained transformers for their specific tasks without any specialized training. The traditional LSTM system provides equal performance to the proposed model but falls short because it cannot effectively handle multiple levels of contextual information. The evaluation results demonstrate

that all four models, which include CBDS-ConvNet, EDM, Stacked Ensemble, and CNN-BiLSTM-GRU, fail to achieve their targets because they either lose recall ability or precision ability. The advanced classifiers Focal Loss EDL-TSGSO and HyVAD-SVM show strong performance results, although HyVAD-SVM approaches the proposed model, their testing results show unbalanced performance across different evaluation metrics. The transformer-based models RoBERTa and ELECTRA achieve competitive results, but they fail to use all the specific features needed for each task. The proposed model demonstrates superior accuracy through its balanced testing results, which include 98.49% accuracy, 98.35% recall, 98.28% precision, and a 98.40% F1-score. The model achieves better performance than all other tested methods because it uses multi-scale contextual feature extraction and CDMO-based optimal feature selection and TFT-powered temporal classification. Figure 9 depicts the graph of the proposed model's multiclass results compared with current models.

Table 8. Results comparison

Models	Accuracy	Recall	Precision	F1
Hybrid LR-RF [11]	92.91	85.79	93.63	89.54
LR-Fuzzy-GA [12]	93.52	87.70	88.94	93.46
NB-Fuzzy-GA [12]	90.30	77.94	85.19	89.75
MT-BERT [13]	82.05	NA	NA	82.27
LSTM [15]	98.00	98.00	98.00	98.00
CBDS-ConvNet [16]	77.65	63.86	56.26	60.20
EDM [18]	96.00	97.00	94.00	96.00
Stacked Ensemble [19]	96.00	96.00	96.00	96.00
Focal Loss [20]	97.00	78.88	92.72	85.54
CNN-BiLSTM-GRU [21]	96.70	95.40	95.10	95.30
EDL-TSGSO [22]	95.49	95.49	NA	95.36
HyVADSVM [23]	98.15	98.15	98.19	97.37
RoBERTa [24]	95.90	95.00	96.00	96.00
ELECTRA [25]	88.35	88.35	88.35	88.35
Proposed Model	98.49	98.35	98.28	98.40

The ablation study results in Table 9 demonstrate that all framework components combine to multiply their individual effects. The MSCT-CDMO-TFT model achieves its highest performance with an accuracy of 98.49% and an F1-score of 98.40% while demonstrating the success of combining multi-scale contextual feature extraction with chaos-driven feature selection and temporal fusion classification methods.

The removal of CDMO from MSCT-TFT leads to a significant metric decline, which demonstrates that feature selection is essential for removing unnecessary and unimportant features while achieving better discrimination results. The research shows that using a single-scale transformer instead of a multi-scale transformer will decrease performance because multiple semantic level contextual data

should be used for optimal performance. The use of LSTM instead of the TFT classifier causes more performance loss, which shows that gated residual learning together with attention mechanisms provides better modelling for complex dependency patterns.

The lowest results are obtained using traditional TF-IDF with SVM, emphasizing the limitations of shallow feature representations compared to deep contextual embeddings. The ablation analysis demonstrates that all modules together with their individual contributions create performance improvements, which need to be used together as the essential method for achieving high performance in cyberbullying and sentiment classification. Figure 10 depicts the graph of the ablation study results comparison.

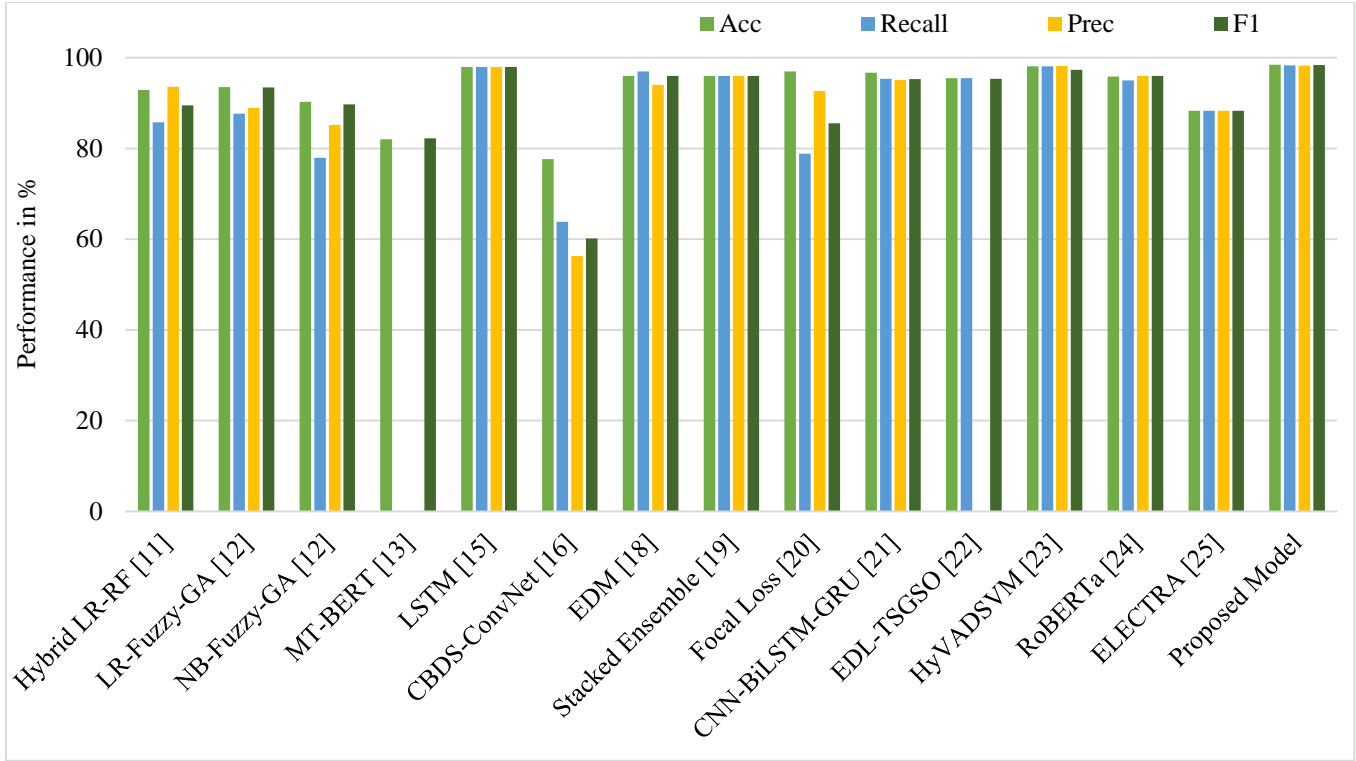


Fig. 9 Graph of results comparison

Table 9. Ablation study results comparison

Model Variant	Accuracy	Recall	Precision	F1
Traditional TF-IDF + SVM	95.96	95.81	95.74	95.77
MSCT + LSTM (without CDMO and TFT)	96.84	96.67	96.59	96.68
MSCT + CDMO + LSTM (without TFT)	97.31	97.18	97.09	97.19
Single-scale Transformer + CDMO + TFT	97.64	97.49	97.42	97.52
MSCT + TFT (without CDMO)	97.92	97.78	97.71	97.81
MSCT + CDMO + TFT (Proposed)	98.49	98.35	98.28	98.40

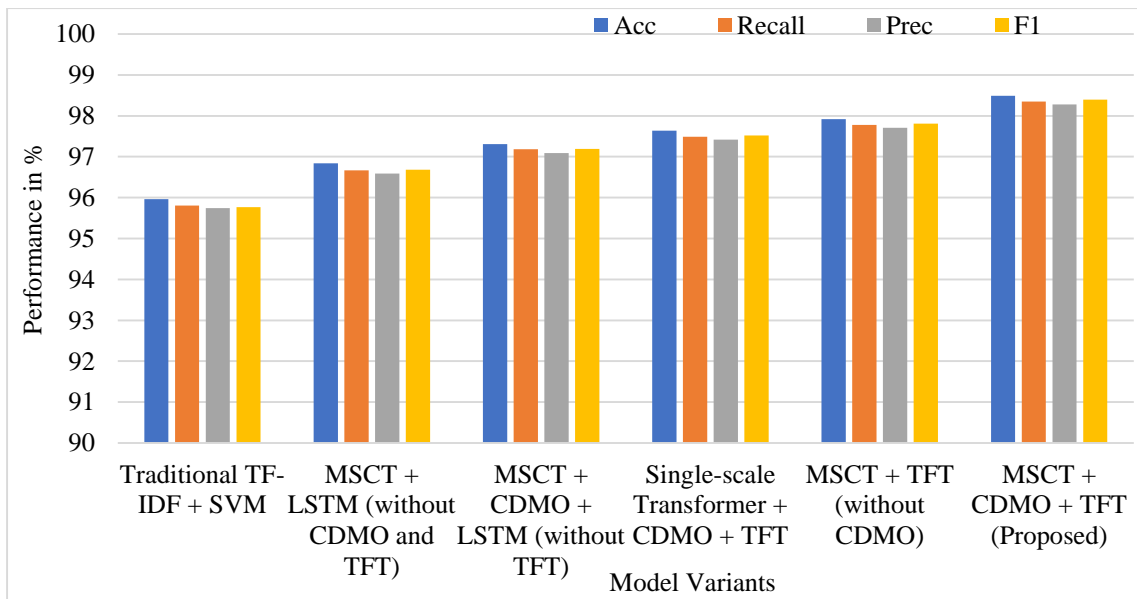


Fig. 10 Graph of ablation study results comparison

4.4. Advantages & Limitations

The MSCT achieves its purpose through the aggregation of different transformer layer representations, which enables the system to capture linguistic data across various semantic levels. The system effectively models hidden emotional states together with intricate patterns of cyberbullying behavior. The CDMO algorithm selects the most discriminative feature subset, which reduces dimensionality through feature elimination while maintaining essential classification data. The TFT employs gated residual learning together with its temporal modelling and attention mechanisms to achieve accuracy in processing long-range dependencies and contextual interactions that exceed the capabilities of standard classifiers. The system demonstrates its ability to handle both binary and multi-class text classification tasks because it maintains high accuracy, precision, recall, and F1-score results throughout three different datasets, which include IMDB, Yelp, and cyberbullying datasets. The proposed model maintains a strong balance between precision and recall because it minimizes both false positive and false negative errors, which function as a significant requirement in cyberbullying detection systems. The framework components function independently, which enables users to replace or enhance the extraction and selection of features and classification modules without needing to redesign the entire system.

However, the model has certain limitations. Using transformer-based feature extraction together with TFT classification requires more computing power and memory space, which creates difficulties for immediate use on devices that have limited resources. The effectiveness of MSCT depends on pretrained transformer representations, which determine its performance through the selection process and language model quality assessment. The CDMO method helps to enhance feature compactness, but its metaheuristic-based feature selection method requires extra computational resources throughout the training process. The combined transformer and TFT architecture operates as a deep black-box model because its attention mechanisms offer partial interpretability but prevent specific feature attribution from being accurately determined. The hyperparameter settings, which include population size for CDMO and attention heads for TFT, need to be adjusted when testing their performance on datasets that contain different linguistic features. The existing system functions only with English texts, and its effectiveness needs to be modified for use with low-resource and multilingual datasets.

References

- [1] Mohammed Hussein Obaida, Saleh Mesbah Elkaffas, and Shawkat Kamal Guirguis, "Deep Learning Algorithm for Cyber-Bullying Detections in Social Media Platform," *IEEE Access*, vol. 12, pp. 76901-76908, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] Jana Lekscha, and Milad Mirbabaie, "How to Analyze Cyberbullying on Social Media Platform: A Systematic Literatures Review in Information System," *i-com*, vol. 24, no. 2, pp. 385-405, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

5. Conclusion

This study proposed a sentiment analysis and cyberbullying detection framework using the MSCT-CDMO-TFT model. The model combined MSCT-based feature extraction with CDMO-based feature selection and a TFT-based classifier. The model was evaluated with IMDB, Yelp Polarity, and Cyberbullying Classification Datasets. The MSCT module succeeded in extracting multi-level contextual features, which enabled the model to learn local semantic patterns together with global contextual dependencies from complex textual data that contained various noise elements. The application of CDMO in feature selection led to a reduction of redundant features while maintaining essential discriminative features, which resulted in improved performance and classification accuracy. The framework received its additional strength from the TFT classifier that successfully executed temporal and contextual interaction modelling through its gated residual learning and attention mechanism capabilities. The proposed model demonstrates superior accuracy through its balanced testing results, which include 98.49% accuracy, 98.35% recall, 98.28% precision, and a 98.40% F1-score. The model achieves better performance than all other tested methods because it uses multi-scale contextual feature extraction and CDMO-based optimal feature selection and TFT-powered temporal classification.

In the future, the study intends to develop the proposed MSCT-CDMO-TFT framework, which currently works with social media platforms by adding support for multilingual and low-resource languages. The model can achieve better performance during real-time deployment through the development of lightweight transformer variants, which decrease its computational requirements. The model's interpretability will improve through the addition of explainable AI techniques, which will help test the system's performance using larger and different types of data sets.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

Not applicable.

- [3] Vidhushavarshini Suresh Kumar et al., "A Comprehensive Review of TriBERT-X: A Concatenated Transformers Approach for Explainable Cyberbullying Detections on Twitter," *International Conferences on Intelligent System and Digital Transformations (ICISD)*, pp. 77-89, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [4] Daniyar Sultan et al., "A Review of Machine Learning Techniques in Cyberbullying Detection," *Computers, Material & Continua*, vol. 74, no. 3, pp. 5625-5640, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [5] Bedour Alrashidi et al., "A Review on Abusive Contents Automatic Detections: Approaches, Challenge and Opportunities," *PeerJ Computer Sciences*, vol. 8, pp. 1-32, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] Malliga Subramanian et al., "A Survey on Hate Speech Detection and Sentiment Analysis using Machine Learning and Deep Learning Models," *Alexandria Engineering Journal*, vol. 80, pp. 110-121, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] Swapnil Sanjaykumar Mane, Suman Kundu, and Rajesh Sharma, "A Survey on Online Aggressions: Contents Detections and Behavioural Analysis on Social Medias," *ACM Computing Survey*, vol. 57, no. 7, pp. 1-36, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [8] Biodoumoye George Bokolo, and Qingzhong Liu, "Artificial Intelligences in Social Media Forensic: A Comprehensive Survey and Analysis," *Electronics*, vol. 13, no. 9, pp. 1-24, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [9] Hooayda Allwaibed et al., "Cyberbullying Detections Approaches for Arabic Text: A Systematic Literatures Review," *Frontiers in Artificial Intelligences*, vol. 8, pp. 1-13, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] Golam Md Mohiuddin, Md Shohel Sayeed, and Ong Lee Yeng, "Deep Learning Model for Culturally Aware Cyberbullying Detections in Muslims Societies: A Systematic Review," *Discover Artificial Intelligence*, vol. 5, pp. 1-38, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [11] Anthony Obogo Otiko et al., "A Hybrid Machine Learnings Approach for Multi-Class Cyberbullying Detections," *SSRN*, pp. 1-10, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [12] Lida Ketsbaia et al., "A Multi-Stages Machine Learnings and Fuzzy Approach to Cyber-Hates Detections," *IEEE Access*, vol. 11, pp. 56046-56065, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [13] Krishanu Maity, Abhishek Kumar, and Sriparna Saha, "A Multitask Multimodal Framework for Sentiment and Emotion-Aided Cyberbullying Detection," *IEEE Internet Computing*, vol. 26, no. 4, pp. 68-78, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [14] Snigdha Luthra, Rakesh Kumar, and Meenu Gupta, "A Novel Hybrid Model Based on Sentiment Analysis Using Community Guided Link Prediction for Connecting People With Similar Interest," *IEEE Access*, vol. 12, pp. 139435-139445, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [15] Atika Gupta, Priya Matta, and Bhasker Pant, "An Integrated Users Matching Framework for Cross-Platforms Cyberbullying Detections," *Discover Computing*, vol. 28, pp. 1-24, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] O. Ayodeji et al., "CBDS-ConvNet: A Cyber-Bullying Detection Model using Convolutional Neural Network," *Communications on Applied Electronics*, vol. 7, no. 40, pp. 11-21, 2025. [[Google Scholar](#)] [[Publisher Link](#)]
- [17] Azhi Faraj, and Semih Utku, "Comparative Analysis of Words Embedding for Multiclass Cyberbullying Detections," *UHD Journal of Sciences and Technology*, vol. 8, no. 1, pp. 55-63, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [18] Mohammed Al-Hashedi et al., "Cyberbullying Detection Based on Emotion," *IEEE Access*, vol. 11, pp. 53907-53918, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [19] Muhammad Mubeen et al., "Cyberbullying-Related Automated Hate Speech Detection on Social Media Platforms Using Stack Ensemble Classification Method," *International Journal of Computational Intelligences System*, vol. 18, pp. 1-24, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [20] Amr Mohamed El Koshiry et al., "Detecting Cyberbullying using Deep Learnings Technique: A Pre-Trained Glove and Focal Loss Techniques," *PeerJ Computer Science*, vol. 10, pp. 1-33, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [21] Raghu Dhupati et al., "Enhancing Sarcasm Detection in Sentiment Analysis for Cyberspace Safety Using Advanced Deep Learning Techniques," *Scientific Report*, vol. 15, pp. 1-19, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [22] Ravuri Daniel et al., "Ensemble Learning with Tournament Selected Glowworm Swarm Optimization Algorithm for Cyberbullying Detection on Social Media," *IEEE Access*, vol. 11, pp. 123392-123400, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [23] Siti Ernawati, Frieyadi Frieyadi, and Eka Rini Yulia, "HyVADSVM: Hybrid VADER-SVM and GridSearchCV Optimization for Enhancing Cyberbullying Detection," *Journal of Robotic and Controls*, vol. 6, no. 1, pp. 101-113, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [24] Arwa A. Jamjoom et al., "RoBERTaNET: Enhanced RoBERTa Transformer Based Model for Cyberbullying Detection with GloVe Features," *IEEE Access*, vol. 12, pp. 58950-58959, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [25] Adamu Philipo et al., "Sentiment-Enhanced Cyberbullying Detection Models on Social Media Platforms," *ACM Transactions on the Web*, vol. 20, no. 1, pp. 1-26, 2026. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [26] Zeeshan Shaikat et al., "Sentiment Analysis on IMDB Using Lexicon and Neural Networks," *SN Applied Science*, vol. 2, pp. 1-10, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [27] Mujiono Sadikin, and Abi Fauzan, "Evaluation of Machine Learning Approach for Sentiment Analysis using Yelp Dataset," *European Journal of Electrical Engineering and Computer Sciences*, vol. 7, no. 6, pp. 58-64, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [28] Meruert Yerekeshva et al., "A Hybrid BERT-CNN with Multihead Self-Attention for Automated Cyberbullying Detection," *Engineering, Technology & Applied Sciences Research*, vol. 16, no. 1, pp. 30716-30724, 2026. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [29] Naga Surekha Jonnala et al., "Leveraging Hybrid Models for Accurate Sentiments Analysis of Twitter Data," *Scientific Report*, vol. 15, pp. 1-9, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [30] Ziyang He, and Ziyi Wang, "Multi-Scale Contextual Transformer for MRI Super-Resolution," *2024 5th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*, Wenzhou, China, pp. 517-520, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [31] Cheng da et al., "MCATD: Multi-Scale Contextual Attention Transformer Diffusion for Unsupervised Low-Light Image Enhancement," *IEEE Access*, vol. 13, pp. 97348-97363, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [32] Mohammed Abdelrazek, Mohamed Abd Elaziz, and A. H. El-Baz, "CDMO: Chaotic Dwarf Mongoose Optimization Algorithm for Feature Selection," *Scientific Reports*, vol. 14, pp. 1-18, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [33] Bryan Lim et al., "Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748-1764, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [34] Iyan E. Mulia et al., "Multi-Horizon Prediction of Tropical Cyclone Intensity and its Interpretability with Temporal Fusion Transformer," *Scientific Reports*, vol. 15, pp. 1-16, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]