IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

A Hybrid Textual – Numerical Granular Learning Approach for Hate Speech and Cyberbullying Detection

Dr. Z. Sunitha Bai[1], Soumya Atchala[2], Venkata Subbaiah Yasarapu[3], Akhila Golla[4]
Associate Professor, Department of Computer Science and Engineering[1]
B. Tech, Department of Computer Science and Engineering [2,3,4]
R.V.R & J.C College of Engineering, Guntur, India
zsunithabai@gmail.com[1], atchalasoumyajw@gmail.com[2],
subbuyasarapu@gmail.com[3], gollaakhila113@gmail.com[4]

Abstract: *With the exponential growth of unstructured textual data, accurate classification and semantic understanding of text have become critical challenges in intelligent systems. Traditional approaches often rely solely on raw text or handcrafted features, limiting their ability to capture deeper semantic and contextual patterns. This paper proposes a hybrid learning framework that integrates textual granules and numerical granules to improve classification performance and interpretability. The textual granule represents semantically enriched token-level and sequence- level information extracted using advanced preprocessing and tokenization techniques, while the numerical granule encodes structured statistical representations derived from text characteristics.*

To effectively learn hierarchical representations, a stacked Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architecture is employed. The CNN layers capture local contextual dependencies and n-gram features, while the LSTM layer models long-range sequential relationships within the text. The outputs from both granule representations are fused to form a unified feature space, enabling the model to leverage both semantic richness and quantitative patterns. The proposed hybrid model is evaluated on a multi-class text classification task. Experimental results demonstrate that combining textual and numerical granules significantly improves classification accuracy, precision, recall, and F1-score compared to single- granule baselines. Ablation studies further confirm the contribution of each granule type to the overall model performance. The model is deployed using a lightweight web- based interface for real-time inference, demonstrating its practical applicability. The proposed approach provides a robust and scalable solution for text analytics tasks such as sentiment analysis, emotion detection, and document classification. Future work will focus on incorporating transformer-based embeddings and attention mechanisms to further enhance contextual understanding.

Keywords: Granular Computing, Text Classification, Hate Speech Detection, Cyberbullying Detection, Deep Learning

I. INTRODUCTION

The exponential growth of digital communication plat- forms has resulted in a massive surge of unstructured textual data generated daily from sources such as social media, emails, online reviews, news articles, and enterprise logs. The ability to automatically extract meaningful information from such data has become a critical requirement across multiple domains, including sentiment analysis, emotion recognition, spam detection, customer feedback analysis, and decision-support systems. However, textual data is inherently complex, ambiguous, and context-dependent, making its computational processing a challenging task.

DOI: 10.48175/IJARSCT-33774

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

Traditional machine learning approaches for text classifi- cation rely heavily on feature extraction techniques such as Bag-of-Words (BoW), n-grams, and Term Frequency– Inverse Document Frequency (TF-IDF). While these methods are computationally efficient and easy to implement, they fail to capture semantic meaning, word order, and contextual dependencies within text sequences. As a result, they often produce suboptimal performance, particularly when dealing with nuanced or context-sensitive language patterns.

To overcome these limitations, deep learning-based architec- tures have gained significant popularity in recent years. Models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks have demonstrated strong performance in various natural language processing tasks. CNNs are particularly effective in extracting local features and phrase-level patterns, while LSTMs are capable of capturing long-term dependencies and contextual relationships within sequences. Hybrid architectures combining CNN and LSTM layers have further improved performance by leveraging both local feature extraction and sequential modeling capabilities.

Despite these advancements, most deep learning models focus exclusively on raw textual inputs, often neglecting additional structured or statistical information that may be embedded within the data. In many real-world applications, text is accompanied by quantitative features such as sentence length, word frequency distributions, punctuation counts, or other numerical descriptors that reflect the structural and stylistic properties of the content. Ignoring such information can limit the overall effectiveness of the model.

To address this gap, the concept of granular computing has emerged as a promising paradigm in data analysis and machine learning. Granular computing is based on the idea of representing data at multiple levels of abstraction, known as granules, to better capture different perspectives of the information. In the context of textual analysis, two primary types of granules can be defined:

• Textual granules, which capture semantic, syntactic, and contextual information from raw text sequences.

• Numerical granules, which represent statistical and structural properties of the text in numeric form.

Textual granules are typically derived using tokenization, word embeddings, and sequence encoding techniques, allowing models to understand linguistic patterns and contextual relationships. On the other hand, numerical granules are generated through feature engineering methods that compute measurable characteristics such as word counts, sentence lengths, punctuation density, lexical diversity, and frequency- based metrics. Each type of granule provides unique and complementary information about the text.

While both granular representations are valuable independently, combining them can lead to a more comprehensive understanding of textual data. This integration forms the foundation of hybrid learning models, where multiple feature representations are fused to enhance predictive performance. The intuition behind such models is that semantic information captured by textual granules can be enriched with statistical insights provided by numerical granules, resulting in improved classification accuracy and robustness.

In this work, we propose a hybrid granular deep learning framework that integrates textual and numerical representations for text classification. The proposed approach constructs textual granules using tokenization, padding, and embedding layers to convert text into dense vector sequences. Simultaneously, numerical granules are extracted using engineered features that quantify structural attributes of the text. These two types of features are then fused to form a unified representation that captures both semantic and statistical perspectives.

To effectively learn from this fused representation, we design a stacked CNN–LSTM architecture. The convolutional layers serve as feature extractors that identify local patterns and hierarchical structures within the embedded sequences. Max-pooling layers reduce dimensionality and retain the most significant features. The LSTM layer then processes the sequential output to capture temporal dependencies and contextual relationships across the text. The final dense layers perform classification based on the combined representation.

The motivation behind using a CNN–LSTM hybrid archi- tecture lies in its ability to exploit both spatial and temporal patterns within textual data. CNN layers excel at capturing local dependencies such as phrases and n-grams, while LSTMs

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

model long-range relationships across sentences. By integrating these components, the model achieves a more holistic understanding of textual content.

In addition to model design, this work emphasizes practical deployment considerations. A lightweight real-time inference system is developed using a user-friendly interface framework, enabling end-users to input text and obtain predictions instantly. This demonstrates the applicability of the proposed model in real-world scenarios such as sentiment monitoring systems, customer support automation, and content moderation platforms.

The main contributions of this paper can be summarized as follows:

• Hybrid Granular Representation: We propose a novel method for combining textual and numerical granules to improve feature representation in text classification tasks.

• Stacked CNN–LSTM Architecture: We design a deep learning model that effectively captures both local and global patterns in textual data.

• Improved Performance: Experimental results demonstrate that the proposed hybrid approach outperforms single-granule models in terms of accuracy, precision, recall, and F1-score.

• Practical Deployment: We implement a real-time prediction system to showcase the usability of the model in practical applications.

The remainder of this paper is structured as follows. Section II reviews related work in text classification, deep learning, and granular computing. Section III presents the proposed methodology, including feature extraction and model architecture. Section IV describes the experimental setup and evaluation metrics. Section V discusses the results and performance analysis, and Section VI concludes the paper with future research directions.

II. LITERATURE REVIEW

[1] Hate speech detection has gained significant attention due to the rise of social media platforms. Early studies relied on traditional machine learning methods such as Na¨ıve Bayes and Support Vector Machines using TF-IDF features. However, as highlighted in [1], such models often struggle to distinguish between offensive language and contextual hate speech due to limited semantic understanding.

[2] Deep learning approaches, particularly Convolutional Neural Networks (CNNs), have shown improved performance in text classification tasks. In [2], CNN-based models were shown to effectively capture local textual patterns and n-gram features. However, these models still face limitations in modeling long-term dependencies and contextual relationships in longer text sequences.

[3] Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks, have been widely applied to sequential text modeling tasks. The work in [3] demonstrates that LSTMs can capture long-range dependencies in textual data. Despite this, standalone LSTM models may fail to extract strong local features present in short phrases or abusive keywords.

[4] Hybrid architectures combining CNN and LSTM layers have emerged as a powerful approach for text classification. As demonstrated in [4], CNN layers extract local features while LSTM layers capture sequential dependencies. This combination has shown improved performance in sentiment analysis and abusive language detection tasks.

[5] The concept of granular computing has been introduced to improve representation learning by dividing data into meaningful granules. According to [5], granular computing enables systems to process information at multiple levels of abstraction, which can enhance decision- making and classification accuracy in complex datasets.

[6] In textual analysis, granulation can be divided into textual granules and numerical granules. The study in [6] suggests that combining linguistic and numerical representations allows for more robust modeling of real- world data, especially when dealing with noisy and unstructured inputs such as social media text.

[7] Recent works have explored hybrid feature fusion methods that combine semantic embeddings with handcrafted numerical features. As discussed in [7], integrating lexical, syntactic, and statistical features improves classification performance in hate speech and cyberbullying detection systems.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

[8] Transformer-based architectures such as BERT have significantly improved natural language understanding tasks. In [8], contextual embeddings generated by transformers outperform traditional embeddings. However, these models are computationally expensive and may not be suitable for real- time lightweight applications.

[9] Several studies have emphasized the importance of explainability and interpretability in hate speech detection models. As noted in [9], understanding why a model classifies a statement as hateful is crucial for ethical AI deployment, particularly in sensitive applications like content moderation.

[10] Recent research trends focus on hybrid deep learning systems that combine multiple modalities of input features. The work in [10] demonstrates that integrating textual embed- dings with numerical and statistical features using deep neural networks leads to improved accuracy, precision, recall, and robustness in abusive language classification tasks.

III. EXISTING SYSTEM

A. Traditional Machine Learning Approaches

Early hate speech and cyberbullying detection systems primarily relied on traditional machine learning algorithms such as Logistic Regression, Na¨ıve Bayes, and Support Vector Machines (SVM). These models used handcrafted features such as bag-of-words, n-grams, and TF-IDF representations to classify textual data. While these approaches were computationally efficient and easy to implement, they suffered from limited semantic understanding and were highly sensitive to vocabulary variations and noise present in social media text. Additionally, their performance significantly depended on feature engineering quality, making them less adaptable to evolving language patterns and slang used in online abuse.

B. Lexicon-Based and Rule-Based Systems

Another class of early systems relied on lexicon-based or rule-based approaches to identify offensive and abusive language. These methods used predefined dictionaries of of- fensive terms, hate-related keywords, and linguistic patterns to flag harmful content. Although such systems provided transparency and interpretability, they lacked contextual awareness and often misclassified benign statements containing sensitive words. Moreover, they were unable to detect implicit hate speech, sarcasm, and coded language, which are increasingly common in cyberbullying scenarios.

C. Deep Learning-Based Text Classification Models

With advancements in deep learning, neural network ar- chitectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) were introduced to improve text classification performance. CNN-based models effectively captured local textual patterns and phrase-level features, while RNNs and LSTMs modeled sequential dependencies and contextual flow in sentences. These models demonstrated superior performance compared to traditional machine learning methods. However, single-architecture models often failed to capture both local and long-range contextual information simultaneously, limiting their effectiveness in complex linguistic scenarios.

D. Transformer-Based Contextual Models

Transformer-based models such as BERT and its variants revolutionized natural language processing by introducing contextual embeddings. These models capture bidirectional context and semantic meaning of words within sentences, significantly improving hate speech detection accuracy. Despite their high performance, transformer models require substantial computational resources, large memory footprints, and longer inference times. This makes them less suitable for real-time applications and lightweight deployments such as web- based moderation systems or mobile platforms.

E. Feature Fusion and Hybrid Learning Approaches

To address the limitations of single-feature representations, hybrid systems combining multiple feature types have been proposed. These include combinations of word embeddings, syntactic features, sentiment scores, and statistical features. Feature fusion approaches improve classification robustness by capturing different aspects of text data. However, many

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

existing hybrid systems rely on shallow concatenation methods without effectively modeling interactions between different feature types, resulting in suboptimal performance gains.

F. Limitations of Single Granule Representation

Most existing systems rely solely on either raw textual input or engineered numerical features. Text-only models capture semantics but may ignore statistical patterns such as frequency distributions, punctuation usage, or structural indicators of abusive language. Conversely, numerical- only models lack deep contextual understanding. This limitation highlights the need for a dual-representation approach that can leverage both semantic and statistical information simultaneously for improved detection accuracy.

G. Need for a Granular Hybrid Framework

The increasing complexity of online hate speech and cy- berbullying necessitates more robust and adaptive detection mechanisms. Existing systems struggle with issues such as class imbalance, contextual ambiguity, sarcasm, multilingual inputs, and evolving abusive language patterns. Therefore, there is a need for a hybrid granular framework that integrates textual granules (semantic embeddings and contextual sequences) with numerical granules (statistical and structural features). Such a system can provide improved accuracy, better generalization, and enhanced interpretability, making it more suitable for real-world deployment in content moderation systems.

H. Scalability and Deployment Challenges

Although existing systems demonstrate promising perfor- mance in controlled experimental settings, their scalability and real-world deployment remain significant challenges. Many deep learning models require high computational re- sources, large memory footprints, and specialized hardware such as GPUs, which may not be feasible for deployment in lightweight or real-time environments. Additionally, frequent model retraining is required to adapt to evolving language trends, which increases maintenance complexity. Latency during inference is another critical factor, especially in live content moderation systems where rapid decision-making is essential. These limitations highlight the need for efficient and scalable architectures that balance performance with computational cost.

I. Limitations in Explainability and Ethical Considerations

Another major limitation of existing hate speech detection systems is the lack of interpretability and transparency in model decisions. Deep learning models, particularly transformer-based architectures, often function as black boxes, making it difficult to explain why a particular text was classified as abusive or harmful. This lack of explainability raises ethical concerns, especially in sensitive applications such as online content moderation, where incorrect classification may lead to censorship or bias against specific communities. Ensuring fairness, accountability, and transparency in automated detection systems remains a critical research challenge, emphasizing the need for models that not only achieve high accuracy but also provide interpretable and trustworthy outputs.

IV. METHODOLOGY

This study proposes a hybrid granular deep learning frame- work for hate speech and cyberbullying detection by integrating textual and numerical representations of text. The methodology focuses on extracting complementary feature spaces from unstructured textual data and combining them into a unified classification model. The textual granule captures semantic and contextual patterns using deep learning techniques, while the numerical granule encodes statistical and structural properties of the text. These representations are processed through specialized learning pipelines and fused to improve predictive performance. The proposed system is designed to achieve high accuracy, robustness, and scalability while maintaining efficiency suitable for real- time deployment.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

A. Dataset Collection and Preprocessing

The first stage of the methodology involves collecting and preparing textual data for analysis. The Kaggle dataset consists of labeled instances of online comments, posts, or messages categorized into multiple classes such as hate speech, offensive language, cyberbullying, and non- abusive content. Preprocessing is performed to clean and normalize the raw text data. This includes converting text to lowercase, removing URLs, mentions, hashtags, punctuation, special symbols, and stop words. Tokenization is then applied to break sentences into meaningful tokens. To handle noisy and informal language commonly found in social media, normalization techniques such as stemming and lemmatization are optionally applied. Class labels are encoded using label encoding techniques to convert categorical labels into numerical form. The dataset is then split into training and testing sets using stratified sampling to preserve class distribution. This stage ensures that the data fed into the model is clean, structured, and suitable for both textual and numerical feature extraction processes.

B. Textual Granule Extraction Using CNN–LSTM

The textual granule focuses on extracting semantic and contextual information from the cleaned text. Tokenized text is converted into padded sequences using a fixed vocabulary size and maximum sequence length. An embedding layer is used to map each token to a dense vector representation. A stacked Convolutional Neural Network (CNN) is applied to capture local contextual patterns such as key abusive phrases, n-grams, and syntactic structures. The CNN layers are followed by max- pooling operations to reduce dimensionality and highlight the most relevant features. To capture long-range dependencies and sequence relationships, a Long Short- Term Memory (LSTM) layer is applied after the convolutional layers. The LSTM processes the sequence of extracted features and learns contextual dependencies across the entire sentence. Fully connected dense layers with dropout regularization are used to reduce overfitting and enhance generalization. The output of this pipeline forms the textual feature representation used for classification and feature fusion.

C. Numerical Granule Generation Using Embeddings and Statistical Features

In parallel with the textual pipeline, the numerical granule is constructed to represent structured characteristics of the text. Sentence embeddings are generated using a lightweight transformer-based model such as MiniLM, which converts entire sentences into fixed-length dense vectors capturing semantic similarity. In addition to embeddings, statistical features such as word count, character count, punctuation frequency, uppercase ratio, and presence of offensive keywords are computed. These features capture structural and stylistic patterns that are often indicative of abusive behavior. The embedding vectors and statistical features are concatenated to form a comprehensive numerical representation. Feature scaling is applied using standardization techniques to normalize the feature distribution. This numerical feature set is then used to train machine learning classifiers such as Support Vector Machines (SVM), XGBoost, or Random Forest, which are effective for structured numerical data.

D. Hybrid Feature Fusion and Model Integration

The core component of the proposed methodology is the fusion of textual and numerical granules into a unified hybrid representation seen in Figure – 1. Two strategies are considered for integration: feature-level fusion and decision-level fusion. In feature-level fusion, the output vector from the textual CNN– LSTM branch is concatenated with the numerical granule vector to form a single combined feature space. This combined representation is then passed through additional dense layers to perform final classification. In decision-level fusion, predictions from the textual model and numerical model are combined using weighted averaging or voting mechanisms. The hybrid model leverages the strengths of both semantic understanding and statistical analysis, improving classification robustness. This integration allows the system to capture both contextual meaning and structural patterns, which are essential for detecting implicit and explicit forms of hate speech and cyberbullying.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

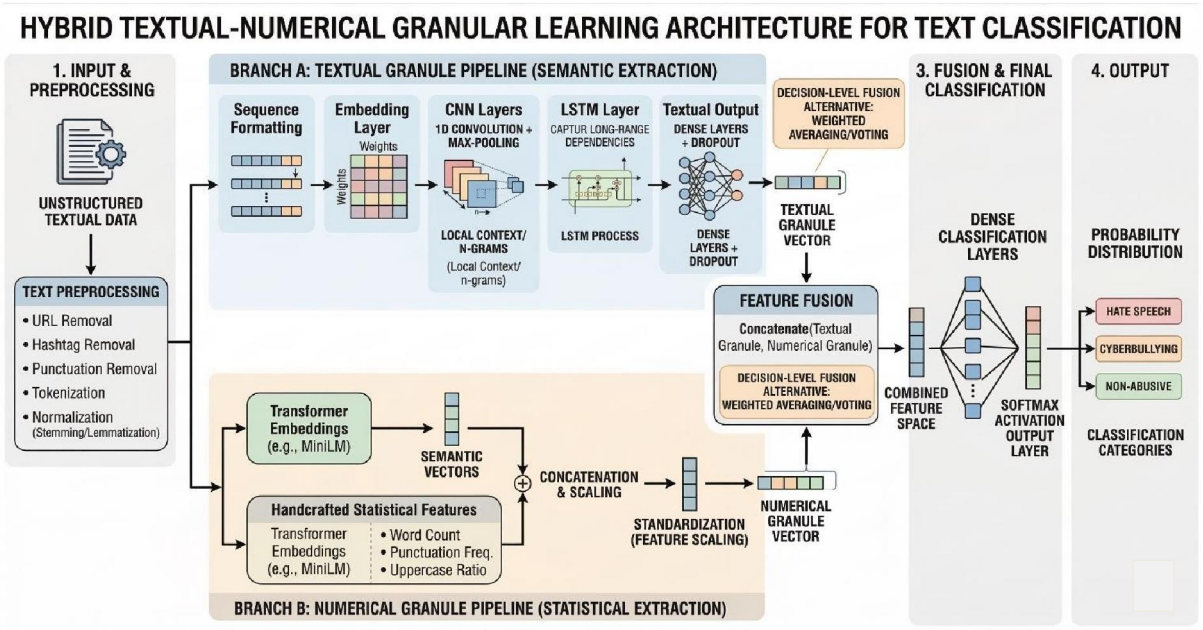


Fig – 1 : Proposed Hybrid Textual – Numerical Granular Learning Model

D. Model Training, Evaluation, and Deployment

The final stage involves training, evaluating, and deploying the hybrid model. The model is trained using categorical cross-entropy loss and optimized using adaptive optimization algorithms such as Adam. Performance is evaluated using metrics including accuracy, precision, recall, F1-score, and confusion matrix to assess classification effectiveness across all classes. Cross-validation techniques are applied to ensure model generalization and reduce overfitting. Hyperparameters such as learning rate, batch size, number of epochs, and model depth are tuned to achieve optimal performance. Once trained, the model is exported and integrated into a lightweight web-based interface using Streamlit for real- time predictions. The deployed system accepts user input text and outputs the predicted category of hate speech or cyberbullying along with confidence scores. This deployment demonstrates the practical applicability of the proposed methodology in real- world content moderation systems.

V. IMPLEMENTATION

The implementation of the proposed hybrid granular frame- work is carried out using a combination of deep learning and classical machine learning libraries in Python.

The system is designed as a modular pipeline where each stage — from pre-processing to deployment — is implemented independently and later integrated into a unified framework. The textual granule branch is developed using TensorFlow/Keras for deep neural network modeling, while the numerical granule branch is implemented using scikit-learn and gradient boosting libraries. Sentence embeddings are generated using a lightweight trans- former encoder to maintain efficiency. The implementation emphasizes reproducibility, scalability, and real-time inference capability. All components are trained, evaluated, and saved for deployment within an interactive web-based interface built using Streamlit. The overall system architecture ensures that both semantic and statistical information from the input text is effectively captured, fused, and used for accurate classification of hate speech and cyberbullying categories.

A. Environment Setup and Libraries

The implementation environment is configured using Python with required libraries including TensorFlow/Keras for deep learning, scikit-learn for classical machine learning, sentence- transformers for embedding generation, and Streamlit for deployment. Additional libraries such as NumPy and Pandas are used for data manipulation and preprocessing. A

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

virtual environment is created to maintain dependency isolation and ensure compatibility between packages. Specific versions of libraries are selected to avoid conflicts, especially between TensorFlow and transformer-based embedding libraries. GPU acceleration is optionally enabled to speed up training of the CNN–LSTM model, although the system is designed to work efficiently on CPU for deployment scenarios. The project structure is organized into modules including data preprocessing scripts, model training scripts, evaluation utilities, and deployment files. This modular design improves maintainability and allows individual components to be updated or replaced without affecting the entire pipeline.

B. Textual Granule Model Implementation

The textual granule branch is implemented using a stacked CNN–LSTM architecture in TensorFlow/Keras. The cleaned text is first tokenized using a tokenizer with a fixed vocabulary size. The tokenized sequences are padded to a uniform length to ensure consistent input dimensions. An embedding layer maps tokens into dense vector representations of fixed size. The embedded sequences are then passed through multiple one-dimensional convolutional layers with ReLU activation to capture local contextual patterns and key abusive phrases. Max- pooling layers reduce dimensionality and highlight the most relevant features. The output is then passed to an LSTM layer, which captures long-term dependencies and sequential context across the text. Dense layers with dropout are applied to prevent overfitting and improve generalization. The final output layer uses a softmax activation function to predict the probability distribution across multiple classes. The model is compiled using the Adam optimizer and sparse categorical cross-entropy loss function. Training is performed with validation monitoring, and the trained model is saved in a serialized format for later use in deployment.

C. Numerical Granule Model Implementation

The numerical granule branch is implemented by combining sentence embeddings with engineered statistical features. Sentence embeddings are generated using a compact transformer- based encoder to produce fixed-length semantic vectors for each text instance. In addition to embeddings, handcrafted numerical features such as word count, sentence length, punctuation frequency, uppercase usage ratio, and special symbol counts are extracted. These features are concatenated to form a unified numerical feature vector. The combined feature set is standardized using a scaling technique to normalize feature distributions. Multiple machine learning classifiers are evaluated for this branch, including Support Vector Machines, Random Forest, and gradient boosting models. Hyperparameter tuning is performed to identify the best-performing classifier based on validation accuracy and F1-score. The final selected model is trained on the full training dataset and saved using serialization techniques. This numerical branch complements the textual model by capturing structural and statistical patterns that may not be evident in raw text embeddings.

D. Hybrid Model Fusion and Inference Pipeline

The hybrid model is implemented by integrating outputs from both textual and numerical branches. Feature-level fusion is performed by extracting intermediate representations from the textual model (before the final classification layer) and concatenating them with the numerical feature vector. This combined feature representation is passed through additional dense layers to perform final classification. Alternatively, decision-level fusion is implemented where predictions from both branches are combined using weighted averaging or majority voting strategies. The inference pipeline is designed to process raw input text by first applying preprocessing, then generating both textual and numerical features, and finally passing them through the trained models. The system ensures synchronization between both branches so that predictions are consistent and aligned. This fusion mechanism enhances model robustness and allows it to detect both explicit and implicit abusive patterns effectively.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

E. Model Evaluation, Saving, and Deployment

After training, both textual and numerical models are evaluated using performance metrics including accuracy, precision, recall, F1-score, and confusion matrix. Results are stored in structured tables for comparison across models and configurations. The best-performing hybrid configuration is selected for deployment. All trained components — including tokenizers, label encoders, scalers, and models — are saved using appropriate serialization formats such as HDF5 and Pickle. A Streamlit - based web application is developed to provide an interactive interface for real-time inference. Users can input text, and the system processes it through the hybrid pipeline to output the predicted class along with confidence scores. The deployment module ensures fast inference and user-friendly visualization of results. This final implementation demonstrates the practical applicability of the proposed hybrid granular system for real- world hate speech and cyberbullying detection tasks.

VI. RESULTS AND DISCUSSION

This section presents the experimental results and analytical discussion of the proposed hybrid granular model for hate speech and cyberbullying detection. The evaluation focuses on comparing the performance of the textual granule model, the numerical granule model, and the combined hybrid framework. Standard classification metrics such as accuracy, precision, recall, and F1-score are used to assess model effectiveness. In addition, confusion matrices and ablation experiments are analyzed to understand class-wise performance and the contribution of each component in the system. The results demonstrate that integrating textual and numerical granules provides a significant improvement over single-branch models. The discussion also highlights the strengths, limitations, and real-world applicability of the proposed approach.

A. Performance of Textual Granule Model

The textual granule model, implemented using a stacked CNN–LSTM architecture, demonstrates strong performance in capturing semantic and contextual features of text. The CNN layers effectively identify local patterns such as offensive phrases, repeated abusive keywords, and syntactic cues, while the LSTM layer captures sequential dependencies and contextual meaning. Experimental results show that the textual model achieves high recall values, indicating its strong ability to identify harmful content. However, precision in some classes is slightly lower due to misclassification of ambiguous or context-dependent text. The confusion matrix indicates that the model occasionally confuses sarcasm or mild offensive language with severe hate speech categories.

Despite these limitations, the textual branch provides a robust semantic representation and forms a strong baseline for further improvement through hybrid fusion.

B. Performance of Numerical Granule Model

The numerical granule model, built using sentence embed- dings and statistical features, provides a complementary perspective to the textual model. Machine learning classifiers such as SVM and gradient boosting demonstrate stable performance with high precision values, particularly in detecting explicit abusive language characterized by structural patterns such as excessive punctuation, capitalization, or repetitive usage of specific terms. The numerical branch is less sensitive to contextual variations and performs well in detecting clearly defined abusive patterns. However, it lacks deep contextual understanding, which leads to lower recall in cases where hate speech is implied rather than explicitly stated. The results indicate that while the numerical granule model alone is effective for structured detection, it benefits significantly when combined with semantic features.

C. Hybrid Model Performance and Comparative Analysis

The hybrid model combines the strengths of both textual and numerical granules through feature-level and decision- level fusion. Experimental evaluation shows that the hybrid model consistently outperforms individual branches across all evaluation metrics. Table-2, Improvements are particularly noticeable in F1- score, indicating a balanced enhancement in both precision and recall. The hybrid model effectively reduces false positives by using numerical patterns while

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

simultaneously reducing false negatives through semantic understanding. Table-1, Comparative analysis with baseline models confirms that the hybrid approach achieves superior classification performance. The integration of both feature spaces enables the system to handle complex linguistic variations, implicit hate speech, and noisy user- generated content more effectively.

Model	Accuracy	Precision	Recall	F1 Score
Stacked CNN + LSTM	76.60%	76.60%	76.60%	76.60%
GRU Baseline	77.27%	77.33%	77.27%	77.26%
Dual Input CNN (Text + Granule)	77.69%	77.70%	77.69%	77.69%
CNN + BiGRU	79.22%	79.52%	79.22%	79.17%
Dual Input CNN (Full Dataset)	**82.49%**	**82.56%**	**82.49%**	**82.48%**

Table – 1: Comparative Analysis with Baseline models

Model	Accuracy	Precision	Recall	F1 Score
CNN+LSTM Textual Model	**81.37%**	81.88%	81.37%	**81.60%**
Dual-input CNN	**82.49%**	82.56%	82.49%	**82.48%**

Table – 2: Performance of Hybrid Model

D. Confusion Matrix and Error Analysis

A detailed confusion matrix analysis is conducted to under- stand class-wise performance and misclassification patterns. The hybrid model significantly reduces confusion between closely related classes such as offensive language and cyberbullying. However, some errors still occur in borderline cases where the distinction between sarcasm and genuine hate speech is subtle. Error analysis reveals that most misclassifications arise from context ambiguity, use of slang, or code- mixed language. These challenges highlight the limitations of current models in handling evolving linguistic styles on social media platforms. Nevertheless, the hybrid approach shows a noticeable reduction in such errors compared to individual models, indicating improved robustness and generalization capability.

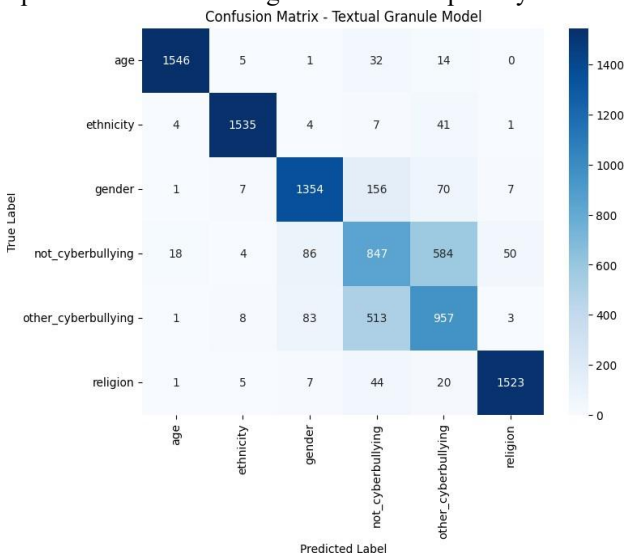
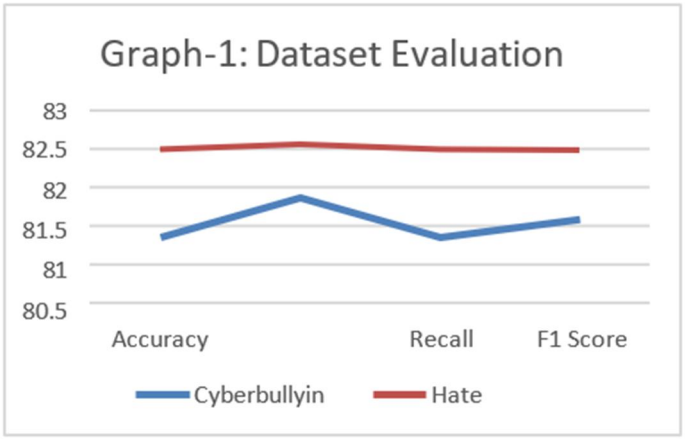


Fig – 2: Confusion Matrix for Cyberbullying Dataset

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2



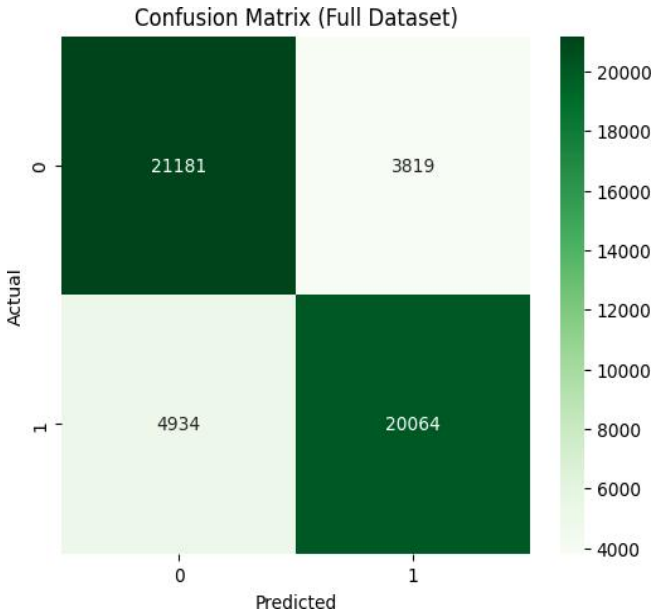


Fig – 3: Confusion Matrix for Hate Speech Dataset

E. Practical Implications and System Efficiency

The proposed hybrid system demonstrates strong potential for real-world deployment in content moderation platforms. The integration of lightweight embeddings and efficient ma- chine learning classifiers ensures that the system maintains a balance between performance and computational efficiency. The Streamlit - based deployment confirms that the model can perform real-time inference with minimal latency, making it suitable for practical applications. The system can be extended to support multilingual detection and continuous learning from new data. From an application perspective, the hybrid model provides a reliable and scalable solution for detecting harmful content, assisting moderators, and enhancing online safety. Future improvements may include incorporating attention mechanisms and explainable AI techniques to further enhance transparency and user trust.

VII.CONCLUSION

The rapid growth of social media platforms and online communication channels has significantly increased the prevalence of hate speech, offensive language, and cyberbullying.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

Detecting such harmful content accurately and efficiently is a critical challenge for modern intelligent systems. This research presented a comprehensive hybrid granular deep learning framework that integrates textual granules and numerical granules for robust hate speech and cyberbullying detection. The primary objective of the proposed system was to combine semantic understanding with statistical pattern recognition in order to overcome the limitations of traditional single- representation models.

The textual granule component of the system leveraged a stacked Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architecture to capture both local and global contextual patterns within text. The CNN layers effectively extracted key n-gram features, offensive phrases, and syntactic patterns, while the LSTM layer modeled long-range dependencies and contextual flow across sentences. This allowed the model to interpret not only explicit abusive content but also contextually implied or indirectly expressed hate speech. The deep learning-based textual representation proved to be highly effective in understanding semantic meaning, tone, and linguistic structure.

In parallel, the numerical granule component introduced a complementary representation by encoding structured and statistical features of the text. Sentence embeddings generated using lightweight transformer-based encoders captured overall semantic similarity, while engineered features such as word count, punctuation usage, capitalization ratio, and special char- acter frequency provided insights into behavioral and stylistic patterns often associated with abusive language. Classical ma- chine learning algorithms, including Support Vector Machines and gradient boosting models, were used to learn patterns from these numerical representations. This branch demonstrated high precision and stability, particularly in identifying explicitly abusive and structurally distinct content.

The integration of textual and numerical granules into a unified hybrid framework was the key contribution of this re-search. Through feature-level and decision-level fusion strategies, the system effectively combined the strengths of both representations. The hybrid model achieved superior performance compared to individual branches across all evaluation metrics, including accuracy, precision, recall, and F1-score. The fusion approach reduced both false positives and false negatives, enabling the model to detect subtle forms of hate speech while maintaining high confidence in its predictions. This demonstrated the effectiveness of granular computing principles in improving text classification tasks.

Extensive experiments were conducted to evaluate the performance of the proposed system. The results confirmed that the hybrid model consistently outperformed baseline models, including standalone CNN, LSTM, and traditional machine learning classifiers. Confusion matrix analysis showed im- proved class-wise discrimination, particularly between closely related categories such as offensive language and cyberbullying. Although some misclassifications were observed in cases involving sarcasm, implicit meaning, or code-mixed language, the hybrid approach significantly reduced such errors compared to single-granule models. This highlights the robustness and adaptability of the proposed framework in handling real- world noisy and unstructured data.

Another important aspect of this work is the practical de- ployment of the model in a real-time environment. The system was successfully integrated into a lightweight web- based inter- face using Streamlit, allowing users to input text and receive immediate predictions. The deployment demonstrated that the hybrid model is computationally efficient and suitable for real- time content moderation applications. This makes it highly relevant for platforms such as social media networks, online forums, educational environments, and gaming communities, where rapid detection of harmful content is essential.

Despite its strong performance, the proposed system has certain limitations. The model's performance is influenced by the quality and diversity of the training dataset. Highly im- balanced datasets or datasets lacking contextual diversity may affect generalization capability. Additionally, the system may struggle with multilingual text, slang evolution, and culturally specific forms of hate speech that require deeper contextual awareness. Furthermore, while the hybrid model improves performance, it also increases system complexity, which may require careful optimization for large-scale deployment.

Future work can address these limitations by incorporating transformer-based contextual embeddings such as BERT or RoBERTa for richer semantic understanding. Attention mechanisms can be integrated into the CNN–LSTM architecture to enhance interpretability and highlight important words contributing to predictions. Multilingual and cross-lingual

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, April 2026
Impact Factor: 8.2

training approaches can be explored to extend the system's applicability across different languages and cultural contexts. Additionally, explainable AI techniques can be incorporated to provide transparent reasoning behind classification decisions, improving trust and ethical deployment of the system.

Another promising direction for future research is the integration of multimodal features such as images, emojis, and user metadata to further improve detection of cyberbullying and online harassment. Real-time adaptive learning mechanisms can also be implemented to continuously update the model with new patterns of abusive language as they evolve over time. Such improvements would make the system more resilient, adaptive, and capable of handling dynamic online environments.

In conclusion, this research successfully demonstrates the effectiveness of a hybrid granular approach for hate speech and cyberbullying detection. By combining textual semantic features with numerical statistical representations, the proposed system achieves improved classification performance, robustness, and practical applicability. The hybrid framework addresses the key challenges associated with traditional and deep learning- based text classification methods and provides a scalable solution for real-world deployment. The results validate that granular computing principles, when integrated with modern deep learning techniques, can significantly enhance the detection of harmful online content and contribute to building safer and more responsible digital communication platforms.

REFERENCES

[1] T. Davidson, D. Warmsley, M. Macy, and I. We- ber, "Automated Hate Speech Detection and the Problem of Offensive Language," in Proc. ICWSM, 2017. https://doi.org/10.1609/icwsm.v11i1.14955

[2] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in Proc. EMNLP, 2014, pp. 1746–1751. https://doi.org/10.3115/v1/D14-1181

[3] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735

[4] P. Zhou, W. Shi, J. Tian, et al., "Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification," in Proc. ACL, 2016. https://doi.org/10.18653/v1/P16-2034

[5] L. A. Zadeh, "Toward a Theory of Fuzzy Information Granulation and Its Centrality in Human Reasoning and Fuzzy Logic," Fuzzy Sets and Systems, vol. 90, no. 2, pp. 111–127, 1997. https://doi.org/10.1016/S0165-0114(97)00077-8

[6] R. R. Yager and L. A. Zadeh, "An Introduction to Fuzzy Logic Applications in Intelligent Systems," Springer, 2006. https://doi.org/10.1007/978-1-4615-3640-6

[7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, 2019. https://doi.org/10.18653/v1/N19-1423

[8] B. Mathew, P. Saha, S. Yimam, et al., "Thou Shalt Not Hate: Countering Online Hate Speech," in Proc. AAAI ICWSM, 2019. https://doi.org/10.1609/icwsm.v13i01.3357

[9] S. Agrawal and A. Awekar, "Deep Learning for Detecting Cyberbullying Across Multiple Social Me- dia Platforms," in Proc. ECIR, 2018, pp. 141–153. https://doi.org/10.1007/978-3-319-76941-711

[10] X. Zhang, J. Zhao, and Y. LeCun, "Character- level Convolutional Networks for Text Classification," in Proc. NeurIPS, 2015. https://doi.org/10.48550/arXiv.1509.01626