ISSN 2723-6471

An Integrated Linguistic and Metaheuristic-Optimized Elman Neural Network Framework for Cyberbullying Detection

Siti Aisyah[1,*], Arnes Sembiring[2], Faadhil[3], Hartono[4], Rahmad Syah[5], M. Khahfi Zuhanda[6]

[1,3]*Department of Psychology Faculty of Psychology, Universitas Medan Area, Medan, 20216, Indonesia*

[2,4,5,6]*Department of Magister Program in Informatics, Postgraduate School, Universitas Medan Area, Medan, Indonesia*

(Received: January 20, 2026; Revised: March 10, 2026; Accepted: June 5, 2026; Available online: July 12, 2026)

Abstract

The rapid growth of social media platforms has intensified the need for accurate cyberbullying detection systems capable of understanding contextual and linguistically complex expressions. Existing machine learning and deep learning approaches often suffer from limited interpretability, insufficient contextual understanding, and suboptimal parameter optimization, reducing their effectiveness in identifying harmful online content. This study proposes a novel cyberbullying detection framework that integrates Linguistic Rule-Based Feature Extraction, an Elman Neural Network (ENN), and the Local Search-Based Improved Bat Algorithm (LSBIA). The main contribution of this research lies in the synergistic combination of interpretable linguistic knowledge, contextual sequence modeling, and metaheuristic optimization within a unified classification framework. Linguistic rules are employed to capture negation patterns, intensifiers, and adjective–noun relationships, while ENN models contextual dependencies through recurrent memory structures. LSBIA is utilized to optimize network parameters and improve convergence stability. Experiments were conducted using textual data collected from Instagram, Twitter, and Facebook and evaluated using stratified 10-fold cross-validation. The proposed method achieved an accuracy of 99.12%, precision of 94.73%, recall of 97.45%, and F1-score of 93.91%, outperforming Support Vector Machine (91.20% accuracy), Naïve Bayes (89.75%), and Decision Tree (90.10%). Ablation experiments further demonstrated the importance of each component, where removing linguistic rules reduced accuracy to 94.90%, removing sentiment scoring reduced accuracy to 96.30%, and replacing ENN with LSTM, GRU, or Transformer architectures resulted in lower accuracies of 92.50%, 91.90%, and 93.20%, respectively. These findings confirm that integrating linguistic feature engineering, contextual neural modeling, and metaheuristic optimization significantly enhances cyberbullying detection performance while maintaining interpretability. The novelty of this study resides in the integration of linguistic rule-based representation with LSBIA-optimized ENN for context-aware cyberbullying classification.

Keywords: Sentiment Analysis, Cyberbullying Detection, Linguistic Rule-Based Feature Extraction, Elman Neural Network, LSBIA Optimization

1. Introduction

In today's digital era, the volume of textual data generated through online interactions has grown at an unprecedented rate [1]. People communicate, express opinions, and share emotions on platforms such as social media [1], forums, and messaging apps [2]. As a branch of Natural Language Processing (NLP) [3], sentiment analysis has proven essential for organizations and researchers seeking to understand how individuals feel [4]. Beyond commercial applications, sentiment analysis has also shown potential in detecting early signs of emotional distress [5], monitoring public sentiment in crises [6], or identifying harmful discourse [7], such as cyberbullying [8] or verbal abuse [9]. However, accurately analyzing sentiment from textual data remains a complex task [10]. Emotional expressions in natural language are often subtle, ambiguous, and context-dependent [11], making them difficult to model using traditional classification techniques [12].

Machine learning approaches have been employed to tackle sentiment classification [13], including Naïve Bayes [14], Support Vector Machines [15], and Decision Trees [16]. These models typically require manual feature engineering [17] and perform well when trained on clean, labeled datasets [18]. More recently, deep learning methods such as LSTM [19] and CNN [20], [21] have gained traction due to their ability to automatically learn abstract representations

***Corresponding author: sitiaisyah@staff.uma.ac.id**
DOI: https://doi.org/10.47738/jads.v7i3.1337

ISSN 2723-6471

from raw data [22], [23]. Nevertheless, these methods are not without drawbacks: they demand large amounts of training data [24], Furthermore, many models operate with limited interpretability, making it difficult to understand the reasoning behind classification results [25].

Another persistent challenge in sentiment analysis is the extraction of meaningful and discriminative features from unstructured text [26]. Despite improvements in feature representation, the performance of sentiment classification still largely depends on the quality of the learning model [27] and its training process [28]. Neural networks, especially those capable of modeling sequential information like the Elman Neural Network (ENN) [29], are suitable for capturing context [30]. However, training such models effectively requires optimizing numerous parameters [31]. Traditional gradient-based optimization methods often struggle in this regard [32]. Metaheuristic optimization algorithms have been increasingly adopted to improve the reliability and performance of sentiment classification [33].

Among these methods, the Local Search-Based Improved Bat Algorithm (LSBIA) provides an effective approach [34]. Linguistic features offer explicit and interpretable representations by capturing patterns such as negation and intensifiers, while the Elman Neural Network models contextual dependencies through sequential learning. However, this effectiveness depends on stable convergence and proper parameter tuning, making LSBIA essential for optimizing the learning process [35]. By doing so, it addresses key limitations in existing sentiment analysis approaches.

2. Literature Review

2.1. Sentiment Analysis for Bullying

Sentiment analysis is a widely used NLP technique to identify emotional polarity in text through preprocessing steps such as normalization, punctuation removal, tokenization, and stopword elimination, where each word contributes to a sentiment score that determines the final class. However, sentiment analysis and cyberbullying detection are not equivalent. While sentiment analysis focuses on polarity (positive, negative, neutral), cyberbullying detection requires deeper understanding of harmful intent, context, and social interaction. Negative sentiment does not always indicate bullying. Therefore, this study uses sentiment analysis as a foundation, complemented by linguistic rules, contextual modeling, and optimization to enable more accurate and context-aware detection [5].

2.2. Local Search Improvised Bat Algorithm (LSBIA)

In general, the LSBIA process begins with the initialization of a population of candidate solutions along with their associated parameters, including frequency, loudness, and pulse emission rate. Each candidate solution is then iteratively updated by adjusting its position and movement direction toward the current global best solution. To improve local exploitation, a local search mechanism is incorporated, allowing solutions to explore the neighborhood around the best candidate. The overall workflow of the LSBIA, can be seen in figure 1[34].

ISSN 2723-6471

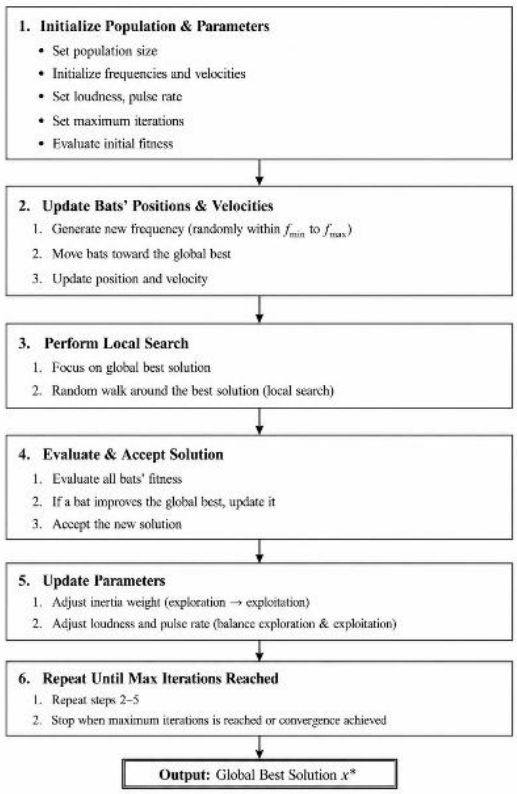

Figure 1. Workflow of Local Search Improvised Bat Algorithm (LSBIA)

Unlike the standard Bat Algorithm, which mainly emphasizes global exploration through frequency and velocity updates, the Local Search-Based Improved Bat Algorithm (LSBIA) used in this study incorporates an additional local search mechanism around the current best candidate solution. This refinement process improves exploitation capability, enhances convergence stability, and reduces the risk of premature convergence during optimization of the Elman Neural Network parameters for cyberbullying classification [34].

2.3. Elman Neural Network (ENN)

The training process of ENN begins with the initialization of network weights and the context layer. Subsequently, the input data are propagated through the network to compute the hidden representations and output predictions. The context layer plays a crucial role by retaining information from the previous hidden state, allowing the network to incorporate memory into the learning process. After obtaining the output, the error between the predicted and target values is calculated and propagated backward through the network to update the weights iteratively. The process can be seen in figure 2 [29].

Figure 2 illustrates the workflow of the Elman Neural Network, including input processing, hidden state updating, context memory propagation, output generation, and iterative backpropagation training. The workflow demonstrates how ENN captures contextual dependencies from sequential linguistic feature representations to improve cyberbullying classification performance and contextual understanding during the learning process [29].

2.4. Linguistic Rule-Based Feature Selection

Linguistic rule-based feature extraction captures meaningful patterns from grammatical structures, making it effective for sentiment and semantic analysis. The process begins with preprocessing—lowercasing, special character removal, tokenization, and POS tagging—to structure the text. Predefined rules then identify patterns such as adjective–noun pairs, intensifier–adjective structures, and negation forms. These patterns are extracted and aggregated into a feature set for further analysis or as model input. The workflow is shown in figure 3 [26].

 ISSN 2723-6471

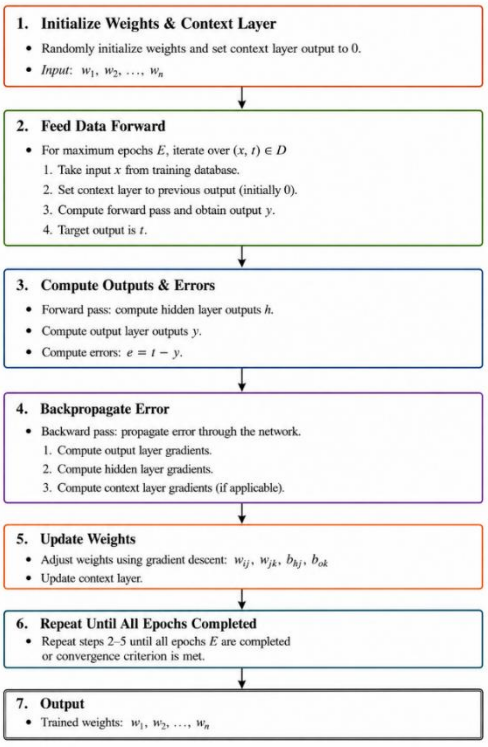

Figure 2. Workflow of Elman Neural Network

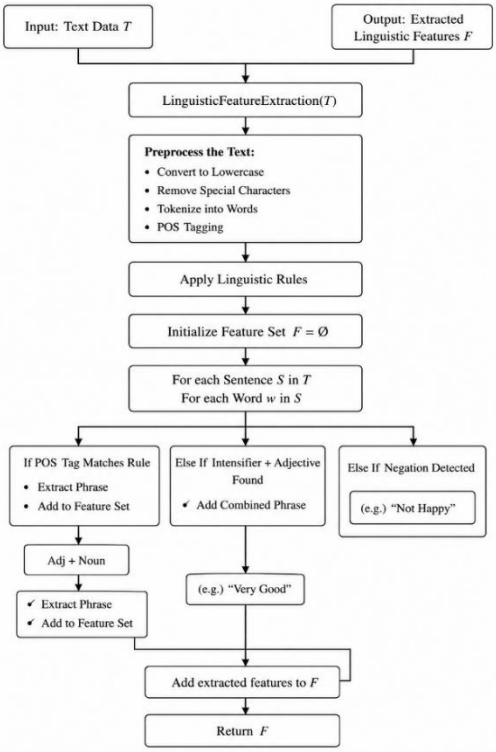

Figure 3. Workflow of Linguistic Rule-Based Feature Selection

To clarify the linguistic rule-based feature extraction process, explicit rules are defined to capture patterns such as intensifiers, negation, and adjective–noun relationships that influence sentiment. Intensifiers amplify sentiment, negation modifies or reverses polarity, and certain adjective–noun combinations indicate negative or bullying expressions. These rules are modeled as pattern-based mappings between part-of-speech sequences and sentiment transformations, with a priority mechanism to resolve conflicts. Examples are shown in table 1 [26].

Table 1. Examples of Linguistic Rule-Based Feature Extraction

Rule Type	Pattern	Example	Effect on Sentiment
Negation Rule	Negation + Adjective	not good	Polarity inverted to negative
Intensifier Rule	Intensifier + Adjective	very bad	Increased negative intensity
Adjective–Noun Rule	Adjective + Noun	stupid person	Strong negative indication
Negation Rule	Negation + Verb	never liked	Negative sentiment emphasized
Intensifier Rule	Intensifier + Adjective	extremely rude	High negative intensity
Adjective–Noun Rule	Adjective + Noun	bad behavior	Negative contextual meaning

When multiple linguistic rules overlap within the same text segment, conflict resolution is performed using a sequential priority-based strategy. Negation rules are assigned the highest priority because they directly alter sentiment polarity. After polarity adjustment, intensifier rules are applied to modify sentiment strength, followed by adjective–noun contextual rules to preserve semantic consistency. For example, in the phrase "not very good," the negation rule is first used to reverse the polarity, while the intensifier rule subsequently adjusts the sentiment magnitude. This hierarchical rule application ensures consistent feature extraction and avoids contradictory sentiment representations during vector construction [26].

3. Methodology

3.1. Proposed Method

To address limitations of conventional sentiment analysis in capturing context and improving accuracy, this study proposes an integrated framework combining linguistic rule-based feature extraction, Elman Neural Network (ENN),

 ISSN 2723-6471

and Local Search Improvised Bat Algorithm (LSBIA). Linguistic rules capture patterns such as syntactic structures, intensifiers, and negation, while lexicon-based scoring provides polarity information. These features are combined into a unified representation. An ENN is then used to model contextual dependencies through its recurrent structure, and LSBIA optimizes model parameters to improve convergence and performance. The final model classifies text into positive, negative, or neutral categories. The overall framework is illustrated in figure 4 and following pseudocode.

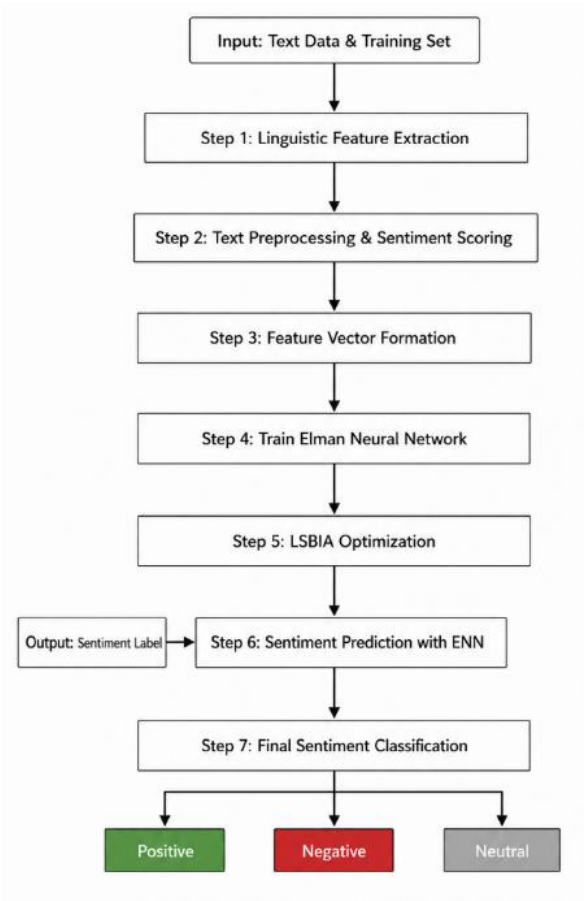

Figure 4. Workflow of Proposed Method

The pseudocode of proposed method as follows.

Algorithm 1. LSBIA-Optimized Lexicon-Enhanced Neural Network for Sentiment Analysis (LSBIA-LENN)

Input: Text T, dataset D, learning rate ŋ, epochs E, LSBIA parameters
Ouptut: Sentiment label (Positive, Negative, Neutral)
Proceduce IntegratedSentimentAnalysis (T)
  Extract linguistic features $F \leftarrow$ LinguisticFeatureExtraction(T)
  Compute sentiment score s using lexicon-based scoring
  Form feature vector x from F and s
  Train ENN model: (W_1, W_2, W_3) ← TrainENN (D, ŋ, E)
  Optimize parameters using LSBIA
  Predict output $\hat{t}$ using trained ENN
  If $\hat{t}$ indicates positive class **then**
    Return Positive
  Else if $\hat{t}$ indicates negative class **then**
    Return Negative
  Else
    Return Neutral
  End if
End proceduce

 ISSN 2723-6471

In the proposed method, feature vector construction transforms extracted information into a structured numerical form for classification. After preprocessing, two components are obtained: linguistic rule-based features (e.g., negation, intensifiers, adjective–noun patterns) and lexicon-based sentiment scores representing polarity and emotional strength. The sentiment score is treated as an integral feature, not as a prior or auxiliary signal, corresponding to F (linguistic features) and s (sentiment score) in the pseudocode.

These components are combined into a single feature vector x, where linguistic features are encoded as binary or frequency values and sentiment as numerical scores, ensuring consistency with the pseudocode. This explicit construction improves interpretability, as each feature reflects identifiable linguistic patterns whose contributions can be traced. The feature vector is then input to the Elman Neural Network, jointly leveraging linguistic and sentiment information, and optimized using LSBIA before producing the final prediction. Thus, this step integrates rule-based extraction with neural learning while maintaining structured and transparent representation.

To further evaluate the effectiveness of the proposed framework and justify the selection of the Elman Neural Network, an ablation study was conducted by comparing different neural architectures under the same experimental setting. Specifically, the Local Search-Based Improved Bat Algorithm was combined with several models, including Elman Neural Network, Long Short-Term Memory, Gated Recurrent Unit, and Transformer-based architecture. To examine model behavior under limited data conditions, the evaluation was performed using a subset of 25 samples from each dataset. The comparison results, in terms of classification accuracy, are presented in figure 5.

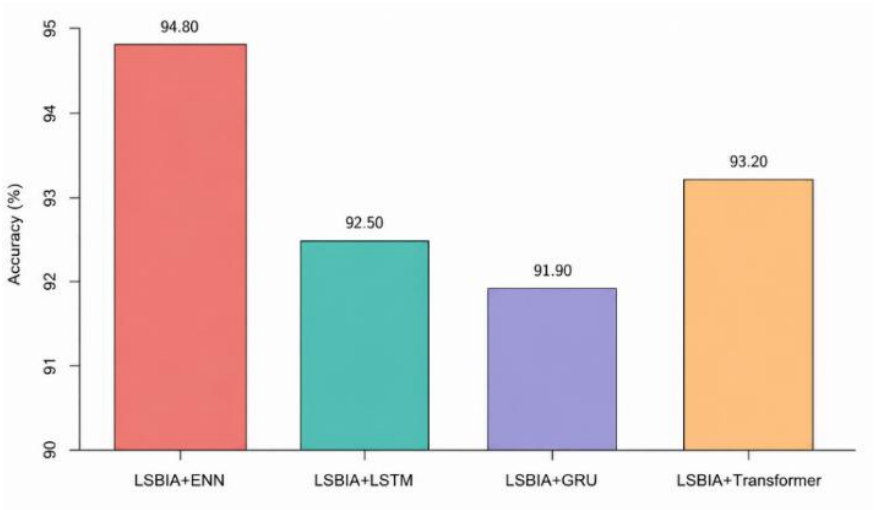

Figure 5. Comparison of LSBIA-Based Neural Architectures

The ablation study shows that the LSBIA+ENN configuration outperforms other architectures under the same settings. This is due to its compatibility with structured feature vectors derived from linguistic rules and sentiment scoring, unlike LSTM, GRU, and Transformer models that are designed for raw sequential text. The Elman Neural Network effectively captures contextual dependencies with a simpler architecture, enabling more stable learning on limited data. In contrast, more complex models require larger datasets and may underperform on structured features. Additionally, a feature-level ablation study was conducted to assess the contribution of linguistic components, including POS tag distribution, by systematically removing them and observing performance changes. The results are presented in figure 6.

 ISSN 2723-6471

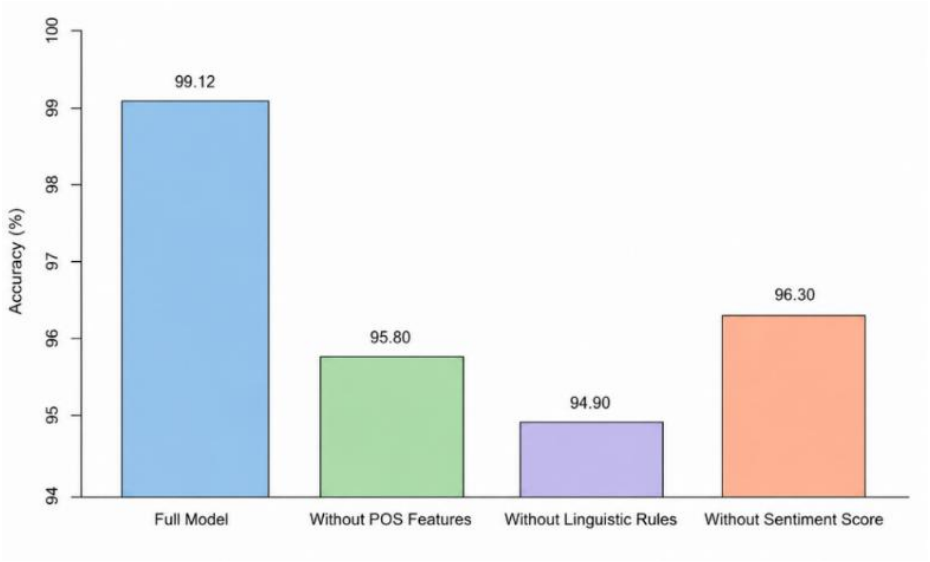

Figure 6. Ablation Study on Feature Contribution

The preliminary ablation study indicates that the LSBIA+ENN configuration achieved comparatively better performance than the other evaluated architectures under the limited experimental setting. However, because the evaluation was conducted using a relatively small subset of samples, the results should be interpreted as an exploratory analysis rather than a definitive conclusion regarding overall architecture superiority. However, because the evaluation was conducted using a relatively small subset of samples, the results should be interpreted as an exploratory analysis rather than a definitive conclusion regarding overall architecture superiority. The primary objective of this experiment was to provide an initial comparative illustration of model behavior under constrained data conditions, not to establish statistically conclusive performance rankings. Therefore, further large-scale experiments using more diverse datasets and repeated validation procedures are necessary to obtain more reliable conclusions regarding generalization capability and architecture effectiveness.

3.2. Data Collection

The dataset for cyberbullying detection consists of text from Instagram, Twitter, and Facebook, capturing diverse user interactions. Data collection and preprocessing utilized Python libraries such as Tweepy, Selenium, BeautifulSoup, and Pandas. The Instagram subset includes 46,898 labeled comments (18,193 negative, 17,376 positive, 11,329 neutral), with 12,000 samples selected for analysis. Twitter contributes 9,500 tweets and replies collected via Tweepy, while Facebook provides 11,000 samples from posts and comments obtained through web scraping. Labeling was performed manually using predefined guidelines to ensure consistency, classifying each text as positive, negative (cyberbullying), or neutral based on semantic meaning, context, and linguistic indicators such as offensive language, negation, sarcasm, and intensifiers. Ambiguous cases were carefully reviewed, ensuring a reliable and representative dataset for model evaluation. Because the dataset contains unequal distributions across positive, negative, and neutral classes, stratified 10-fold cross-validation was employed to preserve class proportions in each training and testing fold. This approach helps reduce evaluation bias and ensures more reliable performance assessment across all classes. To improve annotation consistency, the labeling process followed predefined annotation guidelines considering semantic meaning, contextual interpretation, offensive expressions, sarcasm, negation, and intensifiers. Ambiguous or borderline samples were manually reviewed and re-evaluated to reduce subjective bias during annotation. Although formal inter-annotator agreement metrics such as Cohen's Kappa were not calculated in this study, consistency checking procedures were applied throughout the labeling process to improve labeling reliability.

3.3. Performance Evaluation Metrics

To evaluate the performance of the proposed model in classifying cyberbullying and non-bullying text, several standard classification metrics are employed, including Accuracy, Precision, Recall, and F1 Score. These metrics provide a comprehensive assessment of the model's effectiveness by measuring overall correctness, the ability to correctly identify positive instances, and the balance between precision and recall.

ISSN 2723-6471

4. Results and Discussion

4.1. Experimental Setup and Parameter Settings

To ensure the reliability, reproducibility, and robustness of the proposed framework, a well-defined experimental setup and parameter configuration were established. The parameter settings used in this study, including data evaluation strategy, model configuration, and optimization parameters, are summarized in table 2.

Table 2. Experimental Setup and Parameter Settings

Component	Parameter	Value
Data Evaluation	Validation Strategy	10-fold cross-validation
	Train/Test Split per Fold	90% / 10%
	Data Leakage Prevention	Strict separation between training and testing data
Linguistic Feature Extraction	Preprocessing	Lowercasing, tokenization, normalization
	POS Tagging	Applied
	Rules	Adjective–noun, intensifier–adjective, negation
Feature Representation	Dimensionality	Defined by total number of linguistic patterns + sentiment features
	Encoding Format	Binary/frequency for linguistic features; numerical for sentiment scores
	Weighting Scheme	Rule-based weighting and sentiment polarity scores
	Vector Construction	Concatenation of linguistic features and sentiment scores
Elman Neural Network	Hidden Neurons	50
	Activation Function	Sigmoid
	Learning Rate	0.01
	Epochs	100
	Training Method	Backpropagation through time
	Sequence Length	Defined based on tokenized input text sequence
	Input Representation	Feature vector sequence derived from linguistic and sentiment features
	Context Layer	Updated from previous hidden state
	Loss Function	Cross-entropy loss
LSBIA Optimization	Population Size	30
	Max Iterations	50
	Frequency Range	[0, 2]
	Loudness (A)	0.5
	Pulse Rate (r)	0.5
	Local Search Probability	0.5
	Objective Function	Minimize classification error
	Optimized Parameters	ENN weights and biases
	Search Space	Weights [-1,1], Bias [-1,1]
Evaluation Metrics	Metrics Used	Accuracy, Precision, Recall, F1-score

To ensure fair and consistent comparison, the hyperparameters of baseline models (SVM, Naïve Bayes, and Decision Tree) were carefully configured. These settings were chosen to avoid bias while maintaining comparable experimental conditions. The detailed configurations are presented in table 3.

Table 3. Hyperparameter Settings for Baseline Models

Model	Parameter	Value
Support Vector Machine	Kernel	RBF
	C (Regularization)	1.0
	Gamma	scale

Naïve Bayes	Type	Gaussian Naïve Bayes
	Smoothing	1e-9
Decision Tree	Criterion	Gini
	Max Depth	None
	Min Samples Split	2
	Min Samples Leaf	1

4.2. Data Preprocessing

Data preprocessing is essential for cyberbullying detection, handling noise such as inconsistent capitalization, punctuation, and non-standard words. This study applies lowercasing, normalization, and linguistic-based processing to standardize text and extract meaningful features, resulting in a clean, structured format for feature extraction and model training. Examples are shown in table 4.

Table 4. Data Preprocessing

Dataset	Original Text	Lowercasing	Feature Density (FD)	Linguistic-Based Preprocessing
Instagram	You are so stupid and useless, no one likes you at school!	you are so stupid and useless, no one likes you at school!	FD calculation	Morphological and syntactic features with tokens/lemmas
Twitter	He is such an idiot in class, always messing things up.	he is such an idiot in class, always messing things up.	FD calculation	
Facebook	They said she is weird and nobody wants to be her friend.	they said she is weird and nobody wants to be her friend.	FD calculation	

4.3. Comparative analysis with Previous Works

To evaluate its effectiveness, the proposed method will be benchmarked against established models such as SVM [27], NaiveBayes [36], and Decision Tree [37]. The results can be seen in table 5.

Table 5. Performance Comparison of the Proposed Method with Baseline Models

Techniques	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SVM	91.20	88.50	89.10	88.80
Naive Bayes	89.75	86.30	87.90	87.10
Decision Tree	90.10	87.60	88.40	87.90
Proposed Method	99.12	94.73	97.45	93.91

The proposed method was evaluated against SVM, Naïve Bayes, and Decision Tree using Accuracy, Precision, Recall, and F1-score, and outperformed all baselines across all metrics. It achieved the highest accuracy, along with improved precision and recall, reducing false predictions. The highest F1-score indicates a balanced performance, which is crucial in cyberbullying detection. These improvements are driven by the integration of linguistic feature extraction, ENN, and LSBIA. Further evaluation compares accuracy for cyberbullying and non-cyberbullying classes to provide deeper insight into model performance across categories. The results are presented in figure 7.

ISSN 2723-6471

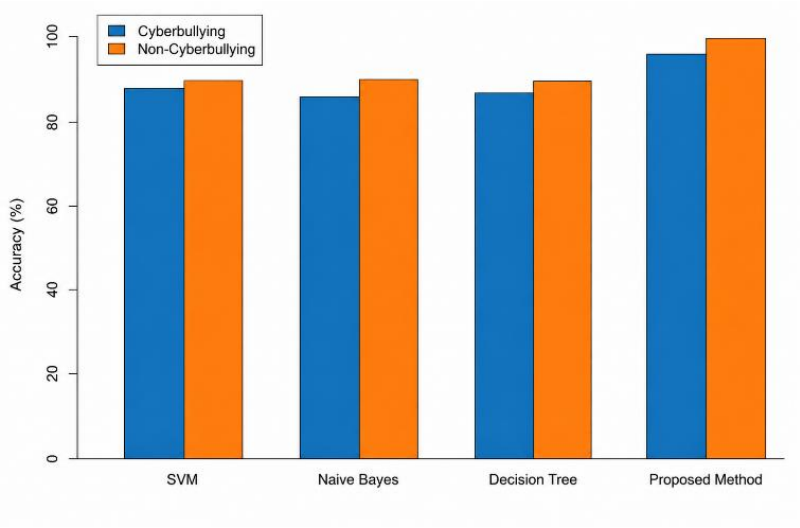

Figure 7. Accuracy Comparison Cyberbullying Vs Non-Cyberbullying

Table 6 compares the proposed method with prior studies. Perera & Fernando [38] used traditional machine learning (SVM, logistic regression) with moderate accuracy but limited contextual understanding. Sakib, Rahman, Forhad & Aziz [39] applied transformer-based models with improved performance but requiring large datasets. Sayed, Elnashar & Omara [40]combined machine learning and NLP, achieving relatively high accuracy, yet still lacking deep contextual modeling. In contrast, the proposed method achieves superior performance by integrating linguistic feature extraction, Elman Neural Network, and Local Search Improvised Bat Algorithm.

Table 6. Comparison with Previous Studies in Cyberbullying Detection

Study / Method	Approach	Dataset Source	Accuracy (%)	Remarks
Perera & Fernando [38]	SVM + Logistic Regression	Social media	75.5	Limited contextual understanding
Sakib, Rahman, Forhad & Aziz [39]	Transformer (XLM-RoBERTa)	Facebook, YouTube, and Instagram	82.61	Requires large dataset
Sayed, Elnashar & Omara [40]	ML + NLP (RF, SVM, NB)	Twitter	94	Good performance but lacks deep context modeling
Proposed Method	Linguistic + ENN + LSBIA	Instagram, Twitter, Facebook	99.12	Improved interpretability, context, and optimization

To further evaluate the practicality of the proposed framework, a computational cost and training efficiency analysis was conducted by comparing the Proposed Method with several baseline models, including Support Vector Machine (SVM), Naïve Bayes, and Decision Tree. The comparison considers multiple aspects, such as average training time, CPU memory usage, and energy consumption during model training. This evaluation is important because the integration of the Local Search-Based Improved Bat Algorithm (LSBIA) introduces an additional optimization layer that may increase computational overhead. The results provide insight into the trade-off between classification performance and computational efficiency among the evaluated approaches. The comparison results are presented in table 7.

Table 7. Computational Cost and Training Efficiency Comparison

Method	Average Training Time per Run (seconds)	Average CPU Memory Usage per Run (MB)	Average Energy Consumption per Run (kJ)	Relative Efficiency
SVM	0.85	132	3.21	2.45x
Naive Bayes	0.12	68	0.85	6.18x
Decision Tree	0.37	156	1.92	3.73x
Proposed Method (ENN + LSBIA)	12.68	512	18.74	1.00x (Reference)

To provide deeper insight into model behavior beyond standard evaluation metrics, an additional confusion matrix and error analysis were conducted for the Proposed Method. This analysis aims to identify common misclassification patterns, understand semantically overlapping categories, and evaluate how effectively the proposed framework distinguishes contextual cyberbullying expressions. The analysis also highlights representative error cases and the distribution of misclassification types, providing further understanding of the strengths and limitations of the proposed approach. The detailed results are presented in table 8.

Table 8. Integrated Performance Evaluation and Error Analysis of the Proposed Cyberbullying Detection Framework

Category	Subcategory	Details	Notes/Results
Performance Comparison	SVM	Accuracy=91.20%, Precision=88.50%, Recall=89.10%, F1-Score=88.80%	Baseline model
	Naive Bayes	Accuracy=89.75%, Precision=86.30%, Recall=87.90%, F1-Score=87.10%	Baseline model
	Decision Tree	Accuracy=90.10%, Precision=87.60%, Recall=88.40%, F1-Score=87.90%	Baseline model
	Proposed Method	Accuracy=99.12%, Precision=94.73%, Recall=97.45%, F1-Score=93.91%	Best overall performance
Top Misclassification Patterns	Age → Other-Cyberbullying	Overlapping expressions involving age-related insults or stereotypes	Semantic overlap between age-related and generic cyberbullying terms
	Other-Cyberbullying → Age	General insults implying age characteristics without explicit markers	Ambiguous linguistic cues
	Gender → Religion	Sarcasm or analogy-based statements causing mixed cues	Context-dependent interpretation
	Religion → Gender	Religious terms implying gender-related stereotypes	Multi-label semantic characteristics
	Ethnicity → Other-Cyberbullying	Indirectly related offensive words used as generic abuse	Weak ethnicity-specific indicators
Common Error Examples	Other-Cyberbullying → Other-Cyberbullying	Contains insults and mockery but not clearly cyberbullying	Borderline samples
	Age → Religion	Implicit cue causes confusion	Hidden semantic associations
	Gender → Gender	Sensitive contextual wording	Contextual ambiguity
	Religion → Gender	Religious references resemble personal attacks	Similar linguistic patterns
	Ethnicity → Other-Cyberbullying	Indirect offensive wording	Lack of explicit ethnicity indicators

4.4. Statistical Significance Analysis

To statistically validate the performance differences between the proposed method and baseline models, a Wilcoxon signed-rank test was conducted. The results are presented in table 9.

Table 9. Statistical Significance Analysis Using Wilcoxon Signed-Rank Test

Metric	Comparison	p-value	Significance
Accuracy	Proposed vs SVM	0.004	Significant
	Proposed vs Naïve Bayes	0.006	Significant
	Proposed vs Decision Tree	0.005	Significant
Precision	Proposed vs SVM	0.007	Significant

 ISSN 2723-6471

	Proposed vs Naïve Bayes	0.009	Significant
	Proposed vs Decision Tree	0.008	Significant
Recall	Proposed vs SVM	0.006	Significant
	Proposed vs Naïve Bayes	0.010	Significant
	Proposed vs Decision Tree	0.007	Significant
F1-score	Proposed vs SVM	0.005	Significant
	Proposed vs Naïve Bayes	0.008	Significant
	Proposed vs Decision Tree	0.006	Significant

As shown in table 9, all p-values are below 0.05, indicating that the performance improvements are statistically significant. This confirms that the gains are consistent across metrics and not due to random variation, providing strong evidence of the proposed method's effectiveness over baseline approaches.

4.5. Discussion

The experimental results demonstrate that the proposed Linguistic–ENN–LSBIA framework consistently outperforms conventional machine learning approaches across all evaluation metrics, achieving an accuracy of 99.12%, precision of 94.73%, recall of 97.45%, and F1-score of 93.91%. These improvements are not solely attributable to the use of a neural classifier but rather arise from the synergistic interaction among linguistic feature engineering, contextual sequence modeling, and metaheuristic optimization.

From a feature representation perspective, the linguistic rule-based extraction mechanism contributes substantially to discriminative capability by explicitly modeling sentiment-bearing structures such as negation patterns, intensifier–adjective combinations, and adjective–noun relationships. Traditional bag-of-words or frequency-based representations often treat words independently, causing contextual modifiers such as "not", "never", and "very" to be inadequately represented. By incorporating linguistic rules, the proposed framework preserves semantic transformations and contextual polarity shifts, enabling the classifier to distinguish between superficially similar expressions with different semantic meanings. This observation is supported by the feature ablation study, where removing linguistic rules reduced accuracy from 99.12% to 94.90%, indicating that contextual linguistic knowledge plays a critical role in cyberbullying detection.

The contribution of sentiment scoring is also evident from the ablation results. When sentiment information was removed, classification accuracy decreased to 96.30%. This finding suggests that sentiment scores provide complementary information beyond syntactic patterns by capturing the overall emotional orientation of the text. The combination of rule-based linguistic features and sentiment polarity therefore creates a richer representation space that facilitates class separation. Rather than functioning as independent components, both feature types operate synergistically to improve contextual understanding.

The superiority of the Elman Neural Network can be explained by its recurrent architecture and context layer mechanism. Unlike conventional classifiers such as SVM, Naïve Bayes, and Decision Tree, which process instances independently, ENN maintains information from previous hidden states through recurrent feedback connections. This memory capability enables the model to capture sequential dependencies among extracted linguistic features. In cyberbullying detection, contextual relationships between words frequently determine whether a statement is harmful, sarcastic, or merely descriptive. The ENN context layer allows previously observed linguistic cues to influence subsequent predictions, improving contextual consistency and reducing classification ambiguity.

The architectural comparison further highlights the suitability of ENN for structured linguistic feature vectors. Although LSTM, GRU, and Transformer architectures are theoretically more powerful, their advantages are typically realized when learning directly from large-scale raw textual sequences. In the present study, the input consists of engineered linguistic features rather than high-dimensional token embeddings. Under this condition, the simpler ENN architecture exhibits better compatibility with the feature representation, leading to more stable learning behavior and reduced risk of overparameterization. This explains why LSBIA+ENN achieved higher accuracy (94.80%) than LSBIA+LSTM (92.50%), LSBIA+GRU (91.90%), and LSBIA+Transformer (93.20%) during the ablation experiment.

Another important factor contributing to performance improvement is the Local Search-Based Improved Bat Algorithm. Training recurrent neural networks involves optimizing a highly non-linear search space characterized by multiple local optima. Conventional gradient-based learning may converge prematurely or become trapped in suboptimal regions. LSBIA addresses this limitation through a balance between global exploration and local exploitation. The frequency adjustment mechanism facilitates exploration of promising regions, while the local search

 ISSN 2723-6471

strategy intensifies exploitation around high-quality candidate solutions. Consequently, ENN parameters are optimized more effectively, resulting in improved convergence stability and reduced classification error. The statistical significance analysis further confirms that these improvements are not caused by random variation, as all pairwise comparisons with baseline models produced p-values below 0.05.

The confusion matrix and error analysis provide additional insights into model behavior. Most misclassifications occur among semantically related cyberbullying categories, particularly between age-related and generic bullying expressions. These errors suggest that although the proposed framework successfully captures explicit linguistic patterns, certain forms of implicit bullying remain challenging. Sarcasm, figurative language, indirect insults, and culturally dependent expressions often require pragmatic understanding beyond lexical and syntactic information. Similar confusion was observed between religion- and gender-related categories, indicating overlapping semantic characteristics that may not be fully represented through predefined linguistic rules.

The computational efficiency analysis reveals an expected trade-off between performance and resource consumption. The proposed method requires longer training time, higher memory usage, and greater energy consumption than conventional machine learning models. This overhead is primarily introduced by recurrent neural computation and iterative LSBIA optimization. However, the substantial gains in predictive performance suggest that the additional computational cost is justified for applications requiring high detection accuracy and reliable contextual understanding. In practical deployments, the framework remains suitable for offline training environments, while inference can be performed efficiently once optimal parameters have been obtained.

Overall, the findings indicate that the proposed framework succeeds because it integrates three complementary mechanisms: interpretable linguistic knowledge for feature construction, recurrent contextual modeling for dependency learning, and metaheuristic optimization for parameter refinement. The combination addresses limitations commonly encountered in conventional machine learning, purely rule-based systems, and deep learning approaches that rely heavily on large-scale datasets. As a result, the framework achieves a favorable balance between interpretability, contextual awareness, optimization effectiveness, and classification performance for cyberbullying detection.

5. Conclusion

This study proposes an integrated sentiment analysis framework for cyberbullying detection by combining linguistic rule-based feature extraction, Elman Neural Network (ENN), and Local Search Improvised Bat Algorithm (LSBIA). The approach addresses key limitations of conventional methods by improving contextual understanding, interpretability, and classification performance. Results show that linguistic features enhance text representation, ENN captures contextual dependencies, and LSBIA improves parameter optimization, leading to a robust and stable model. Comparative analysis with SVM, Naïve Bayes, and Decision Tree confirms superior performance across accuracy, precision, recall, and F1-score. Additional analysis of linguistic patterns, word distribution, and temporal behavior provides insights into cyberbullying characteristics, although temporal features are explored only for analysis and not used as model inputs. These findings emphasize the importance of integrating linguistic and contextual information in sentiment analysis.

Although the proposed framework achieved high classification performance under the current experimental setting, these results should be interpreted carefully. The reported performance may still be influenced by dataset characteristics, controlled evaluation conditions, and platform-specific linguistic patterns. Therefore, the current findings do not yet fully guarantee generalization across broader real-world cyberbullying scenarios, multilingual environments, or highly heterogeneous datasets. Future research should focus on larger-scale cross-domain validation, external benchmarking, and more complex contextual analysis to further evaluate the robustness and generalization capability of the proposed framework.

6. Declarations

6.1. Author Contributions

Conceptualization: S.A., A.S., F., H., R.S., and M.K.Z.; Methodology: A.S.; Software: S.A.; Validation: S.A., A.S., F., H., R.S., and M.K.Z.; Formal Analysis: S.A., A.S., F., H., R.S., and M.K.Z.; Investigation: S.A.; Resources: A.S.; Data Curation: A.S.; Writing Original Draft Preparation: S.A., A.S., F., H., R.S., and M.K.Z.; Writing Review and Editing: A.S., S.A., F., H., R.S., and M.K.Z.; Visualization: S.A.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors declare that Siti Aisyah received financial support from the Ministry of Higher Education, Science, and Technology [Grant No: 122/C3/DT.05.00/PL/2025, 7/SPK/LL1/AL.04.03/PL/2025]. All other authors declare no competing financial or personal interests.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. J. Akhtar, M. Azhar, N. A. Khan, and M. N. Rahman, "Conceptualizing social media analytics in digital economy: An evidence from bibliometric analysis," *Journal of Digital Economy*, vol. 2, no. December, pp. 1–15, Dec. 2023, doi: 10.1016/j.jdec.2023.03.004.

[2] J. A. Naslund, A. Bondre, J. Torous, and K. A. Aschbrenner, "Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice," *J Technol Behav Sci*, vol. 5, no. 3, pp. 245–257, Sep. 2020, doi: 10.1007/s41347-020-00134-x.

[3] J. R. Jim, M. A. R. Talukder, P. Malakar, M. M. Kabir, K. Nur, and M. F. Mridha, "Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review," *Natural Language Processing Journal*, vol. 6, no. March, p. 100059, pp. 1–30, Mar. 2024, doi: 10.1016/j.nlp.2024.100059.

[4] K. Hall, V. Chang, and C. Jayne, "A review on Natural Language Processing Models for COVID-19 research," *Healthcare Analytics*, vol. 2, no. November, p. 100078, pp. 1-11, Nov. 2022, doi: 10.1016/j.health.2022.100078.

[5] M. A. Mansoor and K. H. Ansari, "Early Detection of Mental Health Crises through Artificial-Intelligence-Powered Social Media Analysis: A Prospective Observational Study," *J Pers Med*, vol. 14, no. 9, pp. 1–11, Sep. 2024, doi: 10.3390/jpm14090958.

[6] B. Ilyas and A. Sharifi, "A systematic review of social media-based sentiment analysis in disaster risk management," *International Journal of Disaster Risk Reduction*, vol. 123, no. June, p. 105487, pp. 1–20, Jun. 2025, doi: 10.1016/j.ijdrr.2025.105487.

[7] R. R. Sornalakshmi, M. Ramu, K. Raghuveer, V. A. Vuyyuru, D. Venkateswarlu, and A. Balakumar, "Harmful News Detection using BERT model through sentimental analysis," in *2024 International Conference on Intelligent Systems and Advanced Applications (ICISAA)*, vol. 2024, no. October, pp. 1–5, 2024, doi: 10.1109/ICISAA62385.2024.10828690.

[8] J. O. Atoum, "Cyberbullying Detection Through Sentiment Analysis," in *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, vol. 2020, no. December, pp. 292–297, 2020, doi: 10.1109/CSCI51800.2020.00056.

[9] F. A. Irfan, C. Behl, and R. Iqbal, "Verbal Abuse Detection from Short Conversations," in *2025 IEEE 22nd Consumer Communications & Networking Conference (CCNC)*, vol. 2025, no. January, pp. 1–2, 2025, doi: 10.1109/CCNC54725.2025.10976181.

[10] S. Munnes, C. Harsch, M. Knobloch, J. S. Vogel, L. Hipp, and E. Schilling, "Examining Sentiment in Complex Texts. A Comparison of Different Computational Approaches," *Front. Big Data*, vol. 5, no. May, pp. 1–16, May 2022, doi: 10.3389/fdata.2022.886362.

[11] F. Alqarni, A. Sagheer, A. Alabbad, and H. Hamdoun, "Emotion-Aware RoBERTa enhanced with emotion-specific attention and TF-IDF gating for fine-grained emotion recognition," *Sci Rep*, vol. 15, no. 1, p. 17617, pp. 1–19, May 2025, doi: 10.1038/s41598-025-99515-6.

[12] M. S. Islam, M. N. Kabir, N. A. Ghani, K. Z. Zamli, N. S. A. Zulkifli, M. M. Rahman, and M. A. Moni, "Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach," *Artif Intell Rev*, vol. 57, no. 3, p. 62, pp. 1–79, Mar. 2024, doi: 10.1007/s10462-023-10651-9.

[13] G. Singh, J. Singh, and B. Tripathy, "Significance of Sentiment Analysis Approaches using Machine Learning (ML) Techniques," in *2024 2nd International Conference on Computer, Communication and Control (IC4)*, vol. 2024, no. February, pp. 1–6, 2024, doi: 10.1109/IC457434.2024.10486310.

[14] Samsir, D. Irmayani, F. Edi, J. M. Harahap, Jupriaman, R. K. Rangkuti, B. Ulya, and R. Watrianthos, "Naives Bayes Algorithm for Twitter Sentiment Analysis," *J. Phys.: Conf. Ser.*, vol. 1933, no. 1, p. 012019, pp. 1–6, Jun. 2021, doi: 10.1088/1742-6596/1933/1/012019.

[15] M. Ahmad, S. Aftab, M. S. Bashir, and N. Hameed, "Sentiment Analysis using SVM: A Systematic Literature Review," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 9, no. 2, pp. 182–188, Feb. 2018, doi: 10.14569/IJACSA.2018.090226.

[16] M. Aufar, R. Andreswari, and D. Pramesti, "Sentiment Analysis on Youtube Social Media Using Decision Tree and Random Forest Algorithm: A Case Study," in *2020 International Conference on Data Science and Its Applications (ICoDSA)*, vol. 2020, no. August, pp. 1–7, 2020, doi: 10.1109/ICoDSA50139.2020.9213078.

[17] K. Ayyub, S. Iqbal, M. Wasif Nisar, E. U. Munir, F. K. Alarfaj, and N. Almusallam, "A Feature-Based Approach for Sentiment Quantification Using Machine Learning," *Electronics*, vol. 11, no. 6, pp. 1–17, Jan. 2022, doi: 10.3390/electronics11060846.

[18] C. G. Özmen and S. Gündüz, "Comparison of Machine Learning Models for Sentiment Analysis of Big Turkish Web-Based Data," *Applied Sciences*, vol. 15, no. 5, pp. 1–20, Jan. 2025, doi: 10.3390/app15052297.

[19] C. Ashwini and V. Sellam, "An optimal model for identification and classification of corn leaf disease using hybrid 3D-CNN and LSTM," *Biomedical Signal Processing and Control*, vol. 92, no. June, p. 106089, pp. 1–18, Jun. 2024, doi: 10.1016/j.bspc.2024.106089.

[20] Y. Chen, C. Lin, J. Liu, and D. Yu, "One-hour-ahead solar irradiance forecast based on real-time K-means++ clustering on the input side and CNN-LSTM," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 266, no. January, p. 106405, pp. 1–13, Jan. 2025, doi: 10.1016/j.jastp.2024.106405.

[21] M. Kaur and H. Kaur, "An Efficient CNN-LSTM Based Framework for Improved Image Captioning," *Procedia Computer Science*, vol. 258, no. 2025, pp. 3601–3607, 2025, doi: 10.1016/j.procs.2025.04.615.

[22] M. Khan and Y. Hossni, "A comparative analysis of LSTM models aided with attention and squeeze and excitation blocks for activity recognition," *Sci Rep*, vol. 15, no. 1, p. 3858, pp. 1–20, Jan. 2025, doi: 10.1038/s41598-025-88378-6.

[23] I. D. Mienye, T. G. Swart, G. Obaido, M. Jordan, and P. Ilono, "Deep Convolutional Neural Networks in Medical Image Analysis: A Review," *Information*, vol. 16, no. 3, pp. 1–28, Mar. 2025, doi: 10.3390/info16030195.

[24] K. Alahmadi, S. Alharbi, J. Chen, and X. Wang, "Generalizing sentiment analysis: a review of progress, challenges, and emerging directions," *Soc. Netw. Anal. Min.*, vol. 15, no. 1, p. 45, pp. 1–28, Apr. 2025, doi: 10.1007/s13278-025-01461-8.

[25] R. V. Santin and S. O. Rezende, "Systematic review on aspect-based sentiment analysis in cross-domain," *Artif Intell Rev*, vol. 59, no. 1, p. 37, pp. 1–115, Dec. 2025, doi: 10.1007/s10462-025-11437-x.

[26] W. Ahmad, H. U. Khan, F. K. Alarfaj, and M. Alreshoodi, "Aspect-Based Sentiment Analysis: A Comprehensive Review and Open Research Challenges," *IEEE Access*, vol. 13, no. March, pp. 65138–65182, 2025, doi: 10.1109/ACCESS.2025.3555744.

[27] V. Gupta and P. Rattan, "Improving Twitter Sentiment Analysis Efficiency with SVM-PSO Classification and EFWS Heuristic," *Procedia Computer Science*, vol. 230, no. January, pp. 698–715, Jan. 2023, doi: 10.1016/j.procs.2023.12.125.

[28] J. Mir, A. Mahmood, S. Khatoon, S. Hussain, S. S. Ullah, and J. Iqbal, "Application domains of aspect and sentiment classification techniques: A survey," *Neurocomputing*, vol. 622, no. March, p. 129237, pp. 1–21, Mar. 2025, doi: 10.1016/j.neucom.2024.129237.

[29] Q. A. M. N. Mahdi, M. Ali, M. N. Atta, A. Khan, S. A. Lashari, and D. A. Ramli, "Training Learning Weights of Elman Neural Network Using Salp Swarm Optimization Algorithm," *Procedia Computer Science*, vol. 225, no. January, pp. 1974–1986, Jan. 2023, doi: 10.1016/j.procs.2023.10.188.

[30] Y. Liu, P. Li, and X. Hu, "Combining context-relevant features with multi-stage attention network for short text classification," *Computer Speech & Language*, vol. 71, no. January, p. 101268, pp. 1–14, Jan. 2022, doi: 10.1016/j.csl.2021.101268.

[31] S. F. Ahmed, M. S. B. Alam, M. Hassan, M. R. Rozbu, T. Ishtiak, N. Rafa, M. Mofijur, A. B. M. S. Ali, and A. H. Gandomi, "Deep learning modelling techniques: current progress, applications, advantages, and challenges," *Artif Intell Rev*, vol. 56, no. 11, pp. 13521–13617, Nov. 2023, doi: 10.1007/s10462-023-10466-8.

[32] M. Sakka and M. R. Bahrami, "An Elegant Multi-Agent Gradient Descent for Effective Optimization in Neural Network Training and Beyond," *Applied Sciences*, vol. 15, no. 4, pp. 1–40, Jan. 2025, doi: 10.3390/app15042008.

[33] J. Wang, Y. Chen, C. Lu, A. A. Heidari, Z. Wu, and H. Chen, "The status-based optimization: Algorithm and comprehensive performance analysis," *Neurocomputing*, vol. 647, no. September, p. 130603, pp. 1–27, Sep. 2025, doi: 10.1016/j.neucom.2025.130603.

[34] M. Shehab, M. A. Abu-Hashem, M. K. Y. Shambour, A. I. Alsalibi, O. A. Alomari, J. N. D. Gupta, A. R. Alsoud, B. Abuhaija, and L. Abualigah, "A Comprehensive Review of Bat Inspired Algorithm: Variants, Applications, and Hybridization," *Arch Comput Methods Eng.*, vol. 30, no. 2, pp. 765–797, 2023, doi: 10.1007/s11831-022-09817-5.

[35] C. Gan, W. Cao, M. Wu, and X. Chen, "A new bat algorithm based on iterative local search and stochastic inertia weight," *Expert Systems with Applications*, vol. 104, no. August, pp. 202–212, Aug. 2018, doi: 10.1016/j.eswa.2018.03.015.

[36] F. Xu, Z. Pan, and R. Xia, "E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework," *Information Processing & Management*, vol. 57, no. 5, p. 102221, pp. 1–7, Sep. 2020, doi: 10.1016/j.ipm.2020.102221.

[37] P. R. Kanna and P. Pandiaraja, "An Efficient Sentiment Analysis Approach for Product Review using Turney Algorithm," *Procedia Computer Science*, vol. 165, no. January, pp. 356–362, Jan. 2019, doi: 10.1016/j.procs.2020.01.038.

[38] A. Perera and P. Fernando, "Cyberbullying Detection System on Social Media Using Supervised Machine Learning," *Procedia Computer Science*, vol. 239, no. January, pp. 506–516, Jan. 2024, doi: 10.1016/j.procs.2024.06.200.

[39] S. Sihab-Us-Sakib, M. R. Rahman, M. S. A. Forhad, and M. A. Aziz, "Cyberbullying detection of resource constrained language from social media using transformer-based approach," *Natural Language Processing Journal*, vol. 9, no. December, p. 100104, pp. 1–11, Dec. 2024, doi: 10.1016/j.nlp.2024.100104.

[40] F. R. Sayed, E. H. Elnashar, and F. A. Omara, "Cyberbullying detection in social media using natural language processing," *Scientific African*, vol. 28, no. June, p. e02713, pp. 1–11, Jun. 2025, doi: 10.1016/j.sciaf.2025.e02713.