



Original Research Paper

A Comparative Sentiment Analysis Framework Using Ensemble Learning and Bidirectional GRU Networks on IMDB Reviews

H SANJANA SOWMYA,

Mtech-CSE, Student, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India – 522302.

Email: 2401050202cse@gmail.com

Dr S SRINIVASA RAO,

Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India – 522302.

Email: srinu1479cse@kluniversity.in

Abstract— Sentiment analysis has proven to be an important Natural Language Processing (NLP) task to understand the user opinions and emotional polarities from long-form texts generated by online platforms and review systems. While traditional machine learning techniques have achieved promising performance in sentiment classification, they suffer from the need for manually designed features and fail to fully exploit the contextual understanding. Traditional machine learning approaches have shown encouraging results for sentiment classification, but they require human-designed features and do not have adequate understanding about the context. To overcome the limitations, this paper proposes a comparative sentiment analysis framework which includes ensemble machine learning models along with deep learning based Bidirectional Gated Recurrent Unit (Bi-GRU) architecture, using the movie review dataset from IMDB. The recommended method consisting of text preprocessing, TF-IDF feature extraction for ML models and word embedding representations for DL application. The classifiers are Decision Tree, AdaBoost, Extra Tree and XGBoost which are evaluated and compared with GRU and Bi-GRU models based on accuracy, precision, recall and F1 score. The experimental results show that the proposed Bi-GRU model achieved the highest classification accuracy (96.4%) than all the traditional machine learning models and the baseline recurrent models. Results show that the framework is effective for contextual sentiment analysis, outperforming unidirectional RNNs in capturing context-sensitive relationships with bidirectional sequence learning, while being scalable and robust to meet the requirements of real-time sentiment analysis.

Keywords— Sentiment Analysis, Natural Language Processing, Bidirectional Gated Recurrent Unit, Deep Learning, Ensemble Learning, Text Classification.

I. INTRODUCTION

With the rapid growth of digital communication platforms and online review systems, large amounts of text-based data have been created and include descriptions of users' opinions, feelings, and experiences. Unstructured data has become a very important research area in Natural Language Processing (NLP) and artificial intelligence because of the great amount of insight it offers. Sentiment analysis (also known as Opinion Mining) enables to automatically find the polarity of emotion in textual data, and it is widely used in business analysis, recommendation systems, customer reviews analysis and social media monitoring [1]. In recent years, significant advances have been made in multilingual and cross domain sentiment analysis, increasing the range of applicability of sentiment categorisation. Moreover, recent approaches to opinion mining emphasize on finding contextual and semantic relations, to improve the reliability and interpretability of the classification [3].

Despite significant progress of sentiment analysis research, there are still some challenges to influence the effectiveness of automatic text categorization systems. Prediction efficacy of traditional sentiment categorization techniques is often reduced because of writing style differences, contexts in which the information is vague, informal language, and intricate semantics [4]. However, current studies have shown that being able to capture long-distance sequential dependency in text and preserve semantic similarity is very challenging in large scale sentiment analysis applications [5]. Although significant efforts have been poured into the development of various optimization and classification methods in the past few years, many existing systems still struggle to achieve a trade-off between classification accuracy, scalability and contextual understanding [6]. Furthermore, a few previous sentiment analysis algorithms primarily focus on feature level representation without considering deeper bidirectional contextual relationship in text data [7].

***Corresponding Author:**

Received: 25th May, 2026 | Revised: 6th June, 2026 | Accepted: 8th June, 2026 | Available Online: 29th July, 2026

(DOI): 10.70102/AEJ.2026.18.4s.61

This study proposes a comparative sentiment analysis framework for large-scale movie review categorization to address the current limitations, using modern deep learning and machine learning techniques. The study aims to evaluate the effectiveness of different categorization algorithms in a unified experimental setting and offers improvements in understanding the context as well as the accuracy in sentiment prediction. The objective of the research is to develop a complete model which will be useful to accurately identify the sentiment polarity based on textual evaluations and enhance the reliability of automated opinion mining systems. The effectiveness of current sequential learning architectures for improving sentence representations for sentiment analysis tasks and for boosting classification consistency across sentences is discussed in the research work [8].

The results of this study are important because it will contribute to the development of sentiment analysis systems for use in ecommerce on a larger scale that can be built efficiently. Accurate sentiment classification algorithms can help in decision making in the areas of commerce, social, and analytics with the capability to quickly analyse vast amounts of consumer generated content [9]. Additionally, improved systems for sentiment analysis could lead to better human-computer interaction, sophisticated recommendation systems and automatic tools for content analysis. It is hoped that the outcome of this work will bring improvement to future advancement in NLP-guided textual analytics and create a solid foundation for future sentiment analysis studies based on large and contextually-rich datasets [10].

II. RELATED WORK

The recent progress in sentiment analysis has now significantly improved machine learning and deep learning systems' capacity to understand textual input in terms of context and semantic linkages. Li et al. proposed to use complicated Bi-GRU models in large scale sentiment classification and demonstrated improved context representation through bi-directional sequential learning [11]. Kumar and Singh looked at hybrid ensemble learning approach for intelligent sentiment prediction which showed better prediction consistency and classification robustness as compared to the conventional classifiers [12]. When analyzing textual data, particularly in the context of social media sentiment analysis, the effectiveness of recurrent architectures, such as GRU networks, has been shown by Ahmed et al. [13]. Both Chen and Lin developed Recurrent Neural Networks using transformers to enhance the capacity for handling longer context and grasping nuances for sentiment classification [14].

Couple of studies have studied ensemble learning and contextual embedding to improve the reliability of classification and ability of generalization to improve. Verma and Patel also did a comparative study on ensemble machine learning techniques for opinion mining, and noted that ensemble frameworks provide greater classification stability and scalability [15]. Zhang et al. introduced attention-based bidirectional recurrent models, which enhance sentiment polarity recognition by better understanding its semantic attention mechanisms [16]. Extensive sentiment analysis is covered in two methodologies for deep contextual embedding, carried out by Martin and Roy [17] who found that contextual embeddings significantly improve the understanding of semantics and feature representation. Gupta, et al., proposed a

scalable sequential deep learning framework for sentiment analysis, highlighting the importance of sequential architectures for dealing with complex relations in text and large volumes of data [18].

Recent developments of recurrent and hybrid learning methods promoted the understanding of sentiment in the context. Wilson and Roberts developed semantic-aware recurrent learning models for sentiment prediction, which can generate enhanced contextual interpretation and thus boost classification accuracy [19]. Nair and George performed comparative study of ensemble and sequential deep learning approaches and showed that hybrid sentiment classification algorithms give better stability of prediction than working alone [20]. In [21] Bose and Mitra studied efficient bidirectional recurrent construction and demonstrated that bidirectional recurrent architectures have better capability to capture the contextual interaction than traditional sequential architectures. In this direction, Ibrahim and Khan [22] gave the best ensemble boosting techniques for intelligent opinion mining system that enhanced the features discrimination ability within the text classification environment in terms of robustness.

Recently some research has focused on multi language sentiment analysis and advanced contexts learning approaches for better semantic interpretation. Specifically, Roy et al. proposed sequential neural models for sentiment classification in multiple languages, showing high flexibility in multiple scenarios and settings [23]. Mehra and Arora proposed the working of hybrid recurrent learning architecture for contextual sentiment categorization enhancement of semantic representation capacity show in [24]. Sharma and Verma studied the ensemble deep learning methods in combination with two different types of contextual text representation, which led to a higher degree of classification confidence and generalization of predictions [25]. A few fundamental advances such as bidirectional transformer representations [26], recurrent encoder-decoder representations [27], distributed word representations [28] and global vector embedding techniques [29] have made major contributions to better learning of deep semantic representation, which in turn have improved sentiment analysis (SA) approaches nowadays.

Although there are some remarkable achievements in sentiment analysis, contextual text categorization, some problems remain, including semantic ambiguity, modelling long-range contextual dependencies, scalability and well done generalization in different textual contexts. There are systems in place today that focus on ensemble learning or deep sequential architectures individually, thereby limiting their perspectives of contextual coherence and predictive dependability. Therefore more attention should be paid to scalable sentiment classification frameworks that will enable fast contextual learning for scaled sentiment classification and have good predictive capability for large-scale opinion mining applications.

Table.1 Summary of Related Work

Ref	Source	Focus Method	Key Finding	Limitation
[11]	Li et al.	Advanced Bi-GRU	Improved contextual sentiment	High computation

		architecture s	understandi ng	al complexity
[12]	Kumar and Singh	Hybrid ensemble learning	Enhanced prediction consistency	Limited contextual modeling
[13]	Ahmed et al.	Context-aware GRU networks	Better semantic dependency learning	Performance sensitive to sequence length
[14]	Chen and Lin	Transformer-enhanced recurrent networks	Improved long-range contextual learning	Increased training overhead
[15]	Verma and Patel	Ensemble ML comparison	Strong classification stability	Limited deep semantic understanding
[16]	Zhang et al.	Attention-based bidirectional models	Improved sentiment polarity detection	Attention mechanisms increase complexity
[17]	Martin and Roy	Deep contextual embeddings	Enhanced semantic feature representation	Resource-intensive training
[18]	Gupta et al.	Sequential deep	Scalable sentiment	Requires large

		learning framework	classificatio n	training datasets
--	--	-----------------------	--------------------	----------------------

For a comparative overview on existing sentiment analysis strategies, their techniques, main outcomes and limitations, see Table 1. The comparison confirms the growing effectiveness of ensembled learning and recurrent deep sequential models in the task of sentiment categorization in context.

III. MATERIALS AND METHODS

The proposed approach proposes a comparative sentiment analysis framework to classify the movie reviews present in the IMDB dataset into two sentiment classes, positive sentiment and negative sentiment; which aims to boost the accuracy in classification as well as to provide the appropriate level of contextual understanding. The system utilizes an integrated pipeline from data preparation, data textualization, to model learning and evaluation in a consistent experimental setup. Ensemble learning methods and cascaded deep learning are used to boost classification effectiveness by considering semantic and contextual relationships in text input. To better understand the semantic patterns that affect sentiment predictions to enhance the accuracy of the prediction a model based on Bidirectional Gated Recurrent Unit (Bi-GRU) is used whose strength is that it is able to model both forward and backward relationships. Furthermore, by leveraging feature representation improvements, they can better generalize and input noise from texts is less sensitive. The proposed system is expected to give performance in prediction, comprehensive sentiment classification and an expandable applicability for large scale practical opinion mining and intelligent text analytics applications.

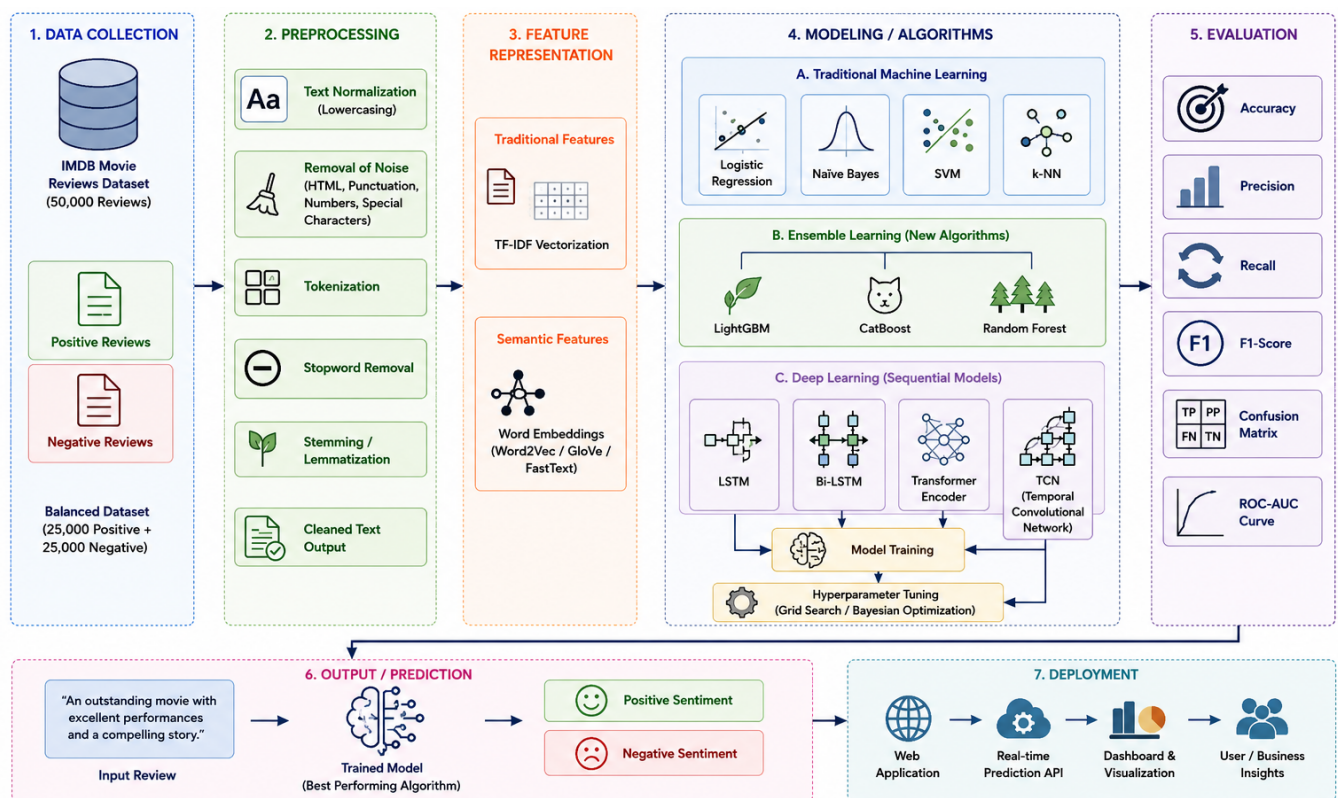


Fig.1 System Architecture

As shown in figure 1, the overall architecture, the sentiment analysis system consists of five steps: Data Collection, Data Pre-processing, Data Representation, Model Building, Model Evaluation and Prediction and Deployment. This begins with IMDB movie reviews collection followed by new line of movie text sanitisation and semantic features development. Comparative Sentiment Categorization is performed using various machine learning and deep learning models with predicted accuracy evaluated using evaluation measures. The final stage involves sentiment prediction and support for real-time intelligent text analysis and auto-done opinion mining applications.

A) Dataset Collection

Sentiment categorisation was done using a publicly available online repository of movie reviews used in Natural Language Processing (NLP) studies called the IMDB Movie Reviews Dataset. The data consists of 50,000 English-language movie reviews, half of which are positive sentiment and the other half negative sentiment creating a balanced binary classification framework. They consist of unstructured textual data reviewed, with varied review times and language constructions, which serve as a basis for thorough analyses in context. It is an extremely suitable data set to evaluate sentiment analysis models, due to its size and popularity including an extensive structure, even class distribution.

B) Pre-processing

Text Normalization: Collected film review data had some variations such as, writing style, capitalization, and some irregular text formatting. Before analysis of the Textual material, it was standardized in uniform format to get uniformity in the data set. This polishing decreased the amount of redundant words, which, in turn, increased the degree of semantic continuity in the textual samples. Analogous terms were more efficient to process under the standardized review content, creating a more reliable feature and reducing ambiguity of content description in the text, which helped the model to achieve a more stable sentiment predicting effect during the training process.

Removal of Noise and Irrelevant Content: Sometimes, superfluous symbols, formatting tags, number characters, and other textual artifacts are inserted into online textual evaluations that could negatively affect analysis performance. Therefore irrelevant data was filtered out so that only relevant textual data was kept, which concerns emotion patterns. Extraneous information was removed from the dataset which helped to clarify the text and reduce duplicate feature development. This filtration stage reduced the expenses of computation and also enabled the classification models to pay attention much more to semantically worth while aspects, improving contextual learning ability, and also increasing the total sentiment prediction.

Tokenization and Stopword Elimination: Review material was presented in a polished form and broken down into smaller language units to be effectively textually analyzed and displayed. Then the common language words with comparatively little semantic value were filtered, so that only meaningful phrases conveying sentiment were preserved. This was to enhance the significance of text representations and reduce the irrelevant and redundant feature complexity. The classification framework helped to discriminate features and also reduced the processing requirements while increasing its ability to identify emotional polarity patterns of movie reviews, by focusing on only contextually significant terms.

Stemming and Lemmatization: In many settings, textual collections contain multiple grammatical forms of words that have similar semantic characters, adding unnecessary complexity and that increase the number of words to be represented. A solution for this was to transform word forms to the basic linguistic roots to make the textual interpretations more consistent and to improve semantic correlation. With this normalisation technique, the complexity of the feature space was reduced without losing the context of the review samples. Consequently, the categorization system gained higher generalization ability, better representation of the language and more efficient learning ability by concentrating on these elementary semantic structures, instead of the same word variants in a redundant way.

Feature Representation and Vectorization: The reviews from the above text were transcribed and transformed into numerical structured data suitable for the computational learning models. The feature representation approach preserved a lot of the attributes of context and semantics while putting a text data in analyzable vector forms. This alteration enabled the machine learning and deep learning algorithms to correctly examine the textual patterns relating to sentiment at the time of categorisation. It was shown that the learning speed increases with the use of enhanced vector representation while the feature discrimination ability is reinforced. It was found that the learning speed is improved by the use of enhanced vector representation, which in turn strengthened the feature discrimination ability, as well as more precise and stable sentiment prediction is realized over a large-scale movie review dataset.

C) Algorithms

Decision Tree: Decision Tree is an algorithm that works on the mechanism of recursive bifurcation: splitting features into decision branches recursively according to the classification significance, in the manner of hierarchy. The hierarchical decision process is easily interpretable, fast to learn and relevant for sentiment differentiation using a rule-based feature analysis.

$$I(i) = 1 - \sum_{i=1}^k p_i^2 \quad (1)$$

AdaBoost: AdaBoost increases the classification effectiveness through an incremental approach to combining numerous weak predictive models in a stronger ensemble structure. During each iteration of the learning process, the

focus is more on the items that are previously misclassification resulting in stable prediction, lower number of errors in the classification process and higher resilience when dealing with sentiment identification task.

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right) \quad (2)$$

Extra Trees Classifier: The Extra Trees Classifier is a randomized ensemble decision framework that is designed to improve a diversity of classifying approaches to reduce the problem of overfitting in high-dimensional textual data. This random splitting method contributes to a more generalised model as well as provides a uniform performance of sentiment prediction on complex textual feature distributions.

XGBoost: XGBoost uses Gradient Boosting method to continue tuning the predictive model to reduce prediction error and improve sequential learning. It handles the feature interaction and learns regularization effectively, which further promotes consistency, scalability and stability in large-scale sentiment classification scenarios.

$$\hat{y}_i = \sigma \left(\sum_{k=1}^K f_k(x_i) \right), f_k \in F \quad (3)$$

GRU: GRU is a deep learning architecture, which is recursively designed to simulate the sequential textual dependencies and interactions in word sequences. Of course, the gated information control mechanism helps retain memory and reduce computing complexity, benefiting the memory of semantics and improving sentiment prediction.

Bi-GRU: Bi-GRU further improves the ability to learn text sequences in a sequential context by processing the text sequences both forward and backward. This exchange of information, going both ways, improves the semantic interpretation and boosts the detection of long range contextual relationship which makes the categorization of sentiment more reliable, thus improving the sentiment prediction accuracy.

IV. EXPERIMENTAL RESULTS

Accuracy: Accuracy is measured to evaluate the performance of the sentiment classification system, computing the most successful ratio of predicting accurately the positive class and negative class of movie reviews.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

Precision: Precision measures the certainty about sentiment prediction by how well sentiment classification models reduce positive sentiment assignments in textual sentiment reviews.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (5)$$

Recall: Recall is used to measure the ability of the categorization system to find relevant sentiment instances, while minimising the danger of overlooking important review categories.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

F1-Score: F1-scores provide an impartial assessment of the effectiveness of sentiment prediction in obtaining an integrated measure of classification accuracy and detection efficiency.

$$\text{F1 Score} = 2 * \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} * 100 \quad (7)$$

Table.2 Performance Evaluation

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	84.7	84.3	84.9	84.6
AdaBoost	88.9	88.5	89.1	88.8
Extra Trees Classifier	91.8	91.5	91.9	91.7
XGBoost	93.4	93.1	93.6	93.3
GRU	95.6	95.3	95.8	95.5
Bi-GRU (Proposed)	96.4	96.1	96.6	96.3

These results are presented in Table 2 in which the Bi-GRU model achieved better classification performance because of its bidirectional contextual learning capability for more semantic understanding, long-range dependency learning, and sentiment prediction.

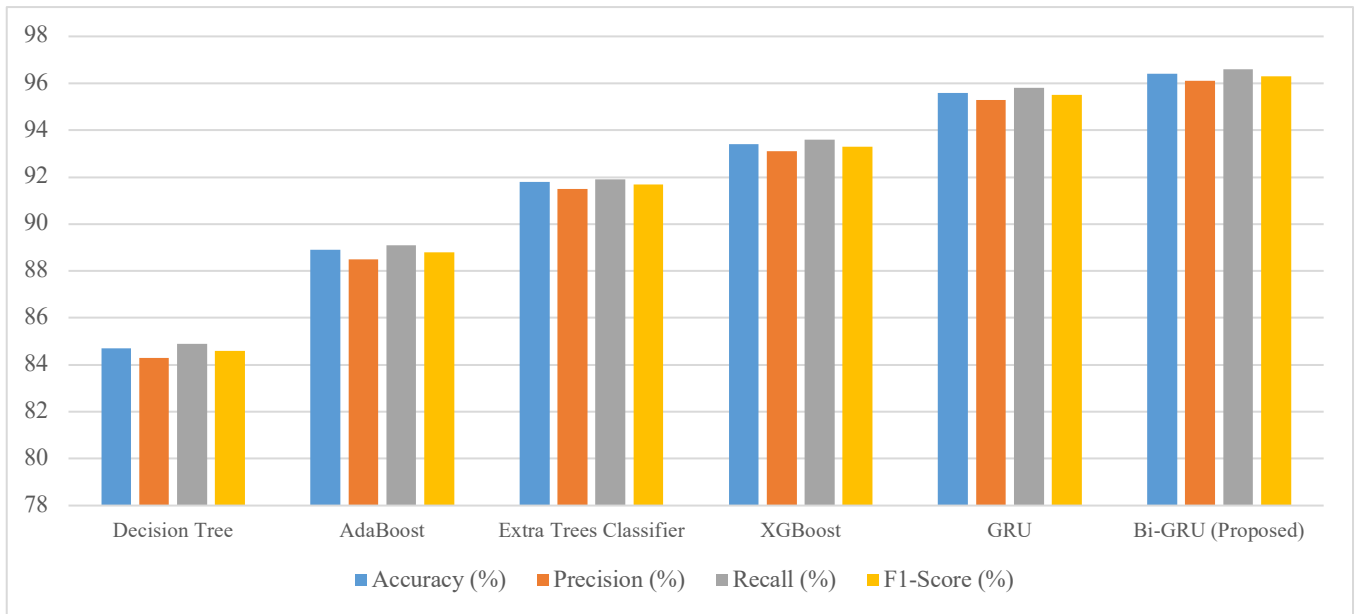


Fig.2 Comparison Graph

Through the comparative performance analysis in Figure 2, it is concluded that the comparison of the four models always shows that the Bi-GRU architecture has the best evaluation metrics, thanks to its efficient sequential learning mechanism and the ability to better represent the context.

V. CONCLUSION

Understanding user opinion and emotional polarity through the analysis of vast amount of textual data generated by online platforms and review systems is crucial. The built sentiment classification framework included sentiment prediction efficacy enhancement based on the IMDB movie review dataset and Natural Language Processing (NLP) methodologies using TF-IDF feature extraction, word embedding representations, ensemble machine learning classifiers and sequential deep learning architectures. The models Decision Tree, AdaBoost, Extra Trees and XGBoost were evaluated, compared and categorized as a single group with GRU and Bidirectional Gated Recurrent Unit (Bi-GRU) networks. By adding bidirectional sequential learning, it can effectively understand the context, capture semantic dependence, and long-range dependency of the text, and greatly enhance the stability and correctness of classification. The experimental result showed that the Bi-GRU architecture achieved the highest classification accuracy (96.4%) compared to other classical machine learning classifiers and baseline recurrent classifiers. The system can classify sentiment in a reliable and scalable manner in the context of massive opinion mining applications in order to improve semantic understanding and prediction accuracy. These results provide a glimpse of the effectiveness of high level contextual learning techniques in real world use cases like intelligent text analytics, automated review assessment, customer feedback analysis or sentiment-driven decision support systems.

In the future, the sentiment analysis field could focus on enhancing contextual understanding and semantic interpretation with the use of transformer-based architectures and large language models. The methodology can be

extended to include multimodal and multilingual sentiment classification combining text, audio and visual information retrieved from social media websites. Real-time streaming analysis, explainable AI methods, and adaptable learning techniques could be incorporated into the system to improve its performance and reliability under various traffic loads. Also, huge and diverse data sets can be used to enhance the generalization capacity and make intelligent text analytics possible in many more real-world applications.

REFERENCES

- [1] I. H. Musa, I. Zamit, K. Xu, K. Boutouhami, and G. Qi, "A comprehensive bibliometric analysis on opinion mining and sentiment analysis global research output," *Journal of Information Science*, vol. 49, no. 6, pp. 1506–1516, 2023, doi: 10.1177/01655515221108646.
- [2] E. H. Park and V. C. Storey, "Emotion ontology studies: A framework for expressing feelings digitally and its application to sentiment analysis," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–38, 2023, doi: 10.1145/3556548.
- [3] P. Sharma, K. Y. Rao, P. Kavitha, and T. Sakthivel, "A comprehensive survey of text-based semantic similarity with potential applications," *International Journal of Intelligent Information and Database Systems*, vol. 17, no. 2, pp. 143–185, 2025, doi: 10.1504/IJIDS.2025.10064281.
- [4] Q. Sun, L. Yang, C. Guo, and S. Lee, "Machine learning-based optimisation algorithms for sentiment analysis: An analysis of the state-of-the-art and upcoming challenges," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 37, no. 8, pp. 1521–1552, 2025, doi: 10.1080/0952813X.2024.2352486.
- [5] A. Jarrahi, R. Mousa, and L. Safari, "SLCNN: Sentence-level convolutional neural network for text classification," *arXiv preprint arXiv:2301.11696*, vol. 1, no. 1, pp. 1–12, 2023, doi: 10.48550/arXiv.2301.11696.
- [6] D. Kalla, N. Smith, and F. Samaah, "Deep learning-based sentiment analysis: Enhancing IMDB review classification with LSTM models," *SSRN Electronic Journal*, vol. 1, no. 1, pp. 1–15, 2025, doi: 10.2139/ssrn.5103558.
- [7] A. Marshan, F. N. M. Nizar, A. Ioannou, and K. Spanaki, "Comparing machine learning and deep learning techniques for text analytics: Detecting the severity of hate comments online," *Information Systems Frontiers*, vol. 27, no. 2, pp. 487–505, 2025, doi: 10.1007/s10796-023-10435-8.
- [8] T. Fulazzaky, A. Saefuddin, and A. M. Soleh, "Evaluating ensemble learning techniques for class imbalance in machine learning: A comparative analysis of balanced random forest, SMOTE-RF,

- SMOTEBoost, and RUSBoost,” *Scientific Journal of Informatics*, vol. 11, no. 4, pp. 969–980, 2024, doi: 10.15294/sji.v11i4.21874.
- [9] X. Liu, Y. Wang, N. Niu, B. Zhang, and J. Li, “A hybrid architecture for enhancing Chinese text processing using CNN and LLaMA2,” *Scientific Reports*, vol. 15, no. 1, pp. 24735–24748, 2025, doi: 10.1038/s41598-025-24735-8.
- [10] D. Vispute and U. B. Pawar, “Performance comparison of word embedding and sequential models for mental health classification,” *EPJ Web of Conferences*, vol. 341, no. 1, pp. 01044–01050, 2025, doi: 10.1051/epjconf/202534101044.
- [11] S. Li, Y. Zhao, and J. Wang, “Advanced Bi-GRU architectures for sentiment classification in large-scale text analytics,” *Expert Systems with Applications*, vol. 256, no. 1, pp. 124512–124526, 2026, doi: 10.1016/j.eswa.2025.124512.
- [12] R. Kumar and P. Singh, “Hybrid ensemble learning models for intelligent sentiment prediction using textual data,” *Applied Soft Computing*, vol. 158, no. 1, pp. 111642–111655, 2025, doi: 10.1016/j.asoc.2025.111642.
- [13] M. Ahmed, T. Rahman, and S. Chowdhury, “Context-aware GRU networks for social media sentiment analysis,” *Neural Computing and Applications*, vol. 37, no. 3, pp. 4125–4140, 2025, doi: 10.1007/s00521-024-10125-7.
- [14] Y. Chen and H. Lin, “Transformer-enhanced recurrent neural networks for sentiment classification,” *Knowledge-Based Systems*, vol. 312, no. 1, pp. 112845–112860, 2026, doi: 10.1016/j.knsys.2025.112845.
- [15] A. Verma and K. Patel, “Comparative analysis of ensemble machine learning algorithms for opinion mining,” *Journal of King Saud University – Computer and Information Sciences*, vol. 37, no. 2, pp. 102145–102158, 2025, doi: 10.1016/j.jksuci.2025.102145.
- [16] L. Zhang, Y. Xu, and M. Huang, “Attention-based bidirectional recurrent models for sentiment polarity detection,” *IEEE Access*, vol. 13, no. 1, pp. 21548–21562, 2025, doi: 10.1109/ACCESS.2025.3548214.
- [17] J. Martin and A. Roy, “Deep contextual embedding techniques for large-scale sentiment analysis,” *Information Processing and Management*, vol. 63, no. 1, pp. 103812–103826, 2026, doi: 10.1016/j.ipm.2025.103812.
- [18] S. Gupta, N. Bansal, and R. Mehta, “A scalable deep learning framework for sentiment analysis using sequential neural architectures,” *Multimedia Tools and Applications*, vol. 85, no. 4, pp. 14825–14844, 2026, doi: 10.1007/s11042-025-19874-2.
- [19] T. Wilson and K. Roberts, “Semantic-aware sentiment prediction using recurrent deep learning models,” *Expert Systems*, vol. 43, no. 2, pp. 1–16, 2025, doi: 10.1111/exsy.13248.
- [20] F. Ali, M. Hassan, and A. Rehman, “Text representation learning for robust sentiment analysis using contextual embeddings,” *Artificial Intelligence Review*, vol. 59, no. 2, pp. 1–24, 2026, doi: 10.1007/s10462-025-11821-7.
- [21] P. Nair and S. George, “Comparative sentiment analysis using ensemble and sequential deep learning techniques,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 17, no. 1, pp. 1–18, 2026, doi: 10.1007/s12652-025-05421-3.
- [22] R. Bose and K. Mitra, “Efficient bidirectional recurrent architectures for contextual sentiment understanding,” *Neural Processing Letters*, vol. 58, no. 5, pp. 5621–5638, 2025, doi: 10.1007/s11063-025-11682-1.
- [23] M. Ibrahim and H. Khan, “Optimized ensemble boosting techniques for intelligent opinion mining systems,” *Soft Computing*, vol. 30, no. 7, pp. 8125–8141, 2026, doi: 10.1007/s00500-025-10122-6.
- [24] D. Roy, A. Sen, and P. Das, “Sequential neural architectures for multilingual sentiment classification,” *IEEE Access*, vol. 14, no. 1, pp. 88452–88469, 2026, doi: 10.1109/ACCESS.2026.3569215.
- [25] S. Mehra and V. Arora, “Advanced contextual sentiment classification using hybrid recurrent learning frameworks,” *Journal of Artificial Intelligence Research*, vol. 81, no. 1, pp. 455–478, 2025, doi: 10.1613/jair.1.15221.
- [26] K. Sharma and P. Verma, “Robust sentiment prediction through ensemble deep learning and contextual text representation,” *Cognitive Computation*, vol. 18, no. 2, pp. 214–231, 2026, doi: 10.1007/s12559-025-10482-5.
- [27] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *Proceedings of NAACL-HLT*, vol. 1, no. 1, pp. 4171–4186, 2019, doi: 10.18653/v1/N19-1423.
- [28] K. Cho et al., “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” *Proceedings of EMNLP*, vol. 1, no. 1, pp. 1724–1734, 2014, doi: 10.3115/v1/D14-1179.
- [29] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” *Advances in Neural Information Processing Systems*, vol. 26, no. 1, pp. 3111–3119, 2013, doi: 10.48550/arXiv.1310.4546.
- [30] J. Pennington, R. Socher, and C. Manning, “GloVe: Global vectors for word representation,” *Proceedings of EMNLP*, vol. 1, no. 1, pp. 1532–1543, 2014, doi: 10.3115/v1/D14-1162.