International Journal of Computer Sciences and Engineering
Vol.13, Issue.6, pp.15-22, June 2025
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

Review Article

Sentiment Analysis on Social Media: A Transformer-Based Approach for Multilingual Data Insights

Sunny Suresh Nahar[1*], Prashant Prakashrao Agnihotri[2]

[1,2]School of Technology, Swami Ramanand Teerth Marathwada University, Sub-Campus, Latur, India

*Corresponding Author:

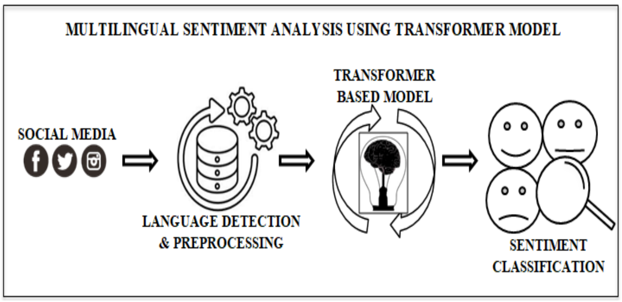
Received: 23/Apr/2025; **Accepted**: 24/May/2025; **Published**: 30/Jun/2025. **DOI**: https://doi.org/10.26438/ijcse/v13i6.1522

Abstract: This is a study conducted on the perspective of doing the sentiment analysis about the user reviews on the transformer-based social media platforms. The main focus of this model will be multi-linguistic data as the social media platforms release a tremendous amount of user-generated multilingual data. Thus, NLP techniques become the inevitable part of using them since these will help in analysis across the varied linguistic contexts where his sentiments prevail. The transformerbased models, one as Bidirectional Encoder Representations from TransformersBERT and its multi-lingual form, have been demonstrated significant improvements over traditional sentiment analysis methods. This paper includes the issues in the management of mixed-language data from social media and the proposed methodology using transformer models to conduct sentiment classification. This paper is also meant to evaluate the very potential of the transformer-based techniques in enhancing sentiment analysis in several languages and to give insight into how well such models may work in different languages while performing the same tasks.

Keywords: Multilingual Sentiment Analysis, Transformer-Based Models, Social Media Analytics, Natural Language Processing (NLP) and BERT and mBERT

Graphical Abstract

This diagram illustrates the pipeline of multilingual sentiment analysis using transformer-based models. It begins with social media data collection across platforms like Twitter, Facebook, Instagram, etc. The data undergoes language detection and preprocessing (removal of noise, special characters, emojis,etc). Then, transformer models like mBERT, BERT, etc. are fine-tuned on labeled multilingual sentiment datasets. The system classifies each post into positive, neutral, or negative sentiment categories. Finally, the output insights are visualized on a dashboard for real-time monitoring of public sentiment trends.



1. Introduction

1.1. Background and Significance

Twitter, Facebook, Insta-gram, and YouTube are probably the major social media sites and hotbeds of public expression through which the users express their opinions, emotions, and comments on selected issues. The millions of daily posts that fill the pages of these platforms have become a mine of rich resources for performing sentiment analysis. The sentiment analysis, like opinion mining, is an automatic process of computing and extracting sentiments or feelings expressed in texts [1]. The valuable insights thus gathered include public opinion, consumer behaviour, political sentiment, and social trends. It helps businesses to manage their reputation, enhance customer service social platforms such as X, Facebook, Instagram, and YouTube have been the central points of public communication where users put forward their views, emotions, and feedback on diverse topics.

The millions of daily posts on these platforms make them a rich source for sentiment analysis[1] and refine their marketing strategies. Politics can measure the public opinion of policies and leaders.

However, extracting relevant insights from social media has its unique challenges because of the informal style of communication online. Social media users frequently use slang, abbreviations, emojis, hashtags, and unconventional sentence structures which make conventional sentiment analysis models, typically designed for formal texts, rather ineffective [2]. Additionally, the short nature of the social media posts lacks context sufficient enough to analyse sentiments.

Another challenge is the multilingual aspect of social media. Users can post varying content worldwide in multiple languages; hence, sentiment analysis models are required to manage different linguistic data [3].

While most sentiment analysis researches are done in the English language, social media's multidimensional nature requires models that can handle content in various languages. This creates complications in terms of linguistic diversity, variations in contexts, and scarcity of training datasets for some languages.

Importance of Multilingual Data in Social Media Sentiment Analysis, as the nature of social media is global, it involves people from around the world with distinctive cultural backgrounds and different linguistic environments. All these elements contribute to the content. With the analysis of the sentiment in the various languages, the business can obtain a broader audience to gather feedback to improve the decision-making process at a global scale. For example, understanding different region customer sentiment will help. Understanding language differences and sentiments across those languages helps in general gathering of feedback from a larger audience, thereby enhancing decision-making on a global level [4]. For example, knowing the customer sentiment of various regions can help product and marketing strategies to be tailored to local preferences and needs.

However, there are challenges associated with the analysis of multilingual data. The same sentiment can be expressed differently in different languages, depending on the sentence structure, vocabulary, and cultural context. A positive sentiment in one language may not necessarily have the same connotation in another [5]. Additionally, many social media posts are code-switched, meaning that users use a mix of languages within one post, making sentiment classification difficult. Another barrier is that the availability of sentiment-labelled data in other languages is much scarcer, with some languages lacking substantial corpora in sentiment analysis while others lack significant training data that hinders their development of models [6].

Advancements in NLP have recently been propelled by transformer models like BERT and GPT, which process entire sentences or documents at once using a self-attention mechanism [8]. Unlike traditional models like RNNs or LSTMs, transformers capture long-range dependencies and context. mBERT and XLM-R, commonly employed in multilingual sentiment analysis, are pre-trained on diverse multilingual corpora and subsequently fine-tuned for task-specific objectives [9].mBERT, for instance, is trained on 104 languages and is expected to maintain semantic meaning across them. These transformer models are highly effective for analysing content on social media, with its informal language and emojis.

1.2. Objectives and Scope of the Paper

1.2.1. Objectives:

a) To explore the integration of transformer models in various tasksaimed at sentiment detection across multilingual social media datasets.
b) To evaluate the effectiveness of mBERT and XLM-R in classifying sentiment (positive, negative, neutral) across different languages.
c) To investigate the challenges involved in processing informal social media content such as slang, emojis, and hashtags.
d) To assess the comparative effectiveness of transformer-based models and conventional sentiment analysis techniques across multilingual datasets.
e) To propose a methodology for addressing the challenges within the context of multilingual sentiment analysis on social media.

1.2.2. Scope:

Here, it provides an indepth examination of the model architectures based on transformers like mBERT and XLM-R for sentiment analysis over many languages on social media data. This paper also discusses informal languages, slang, emojis, and code-switching in multilingual content. Besides, comparative studies have also been done on modern and traditional approaches in adding value to multilingual sentiment analysis in understanding global trends, brand analyses, and public-opinion analyses.

2. Review of Multilingual Sentiment Analysis: A Transformer-Based Approach

The following section includes the review of work and analysis achieved in the past five years, highlighting advancements in multilingual sentiment analysis in social media spaces. Mostly, these studies focus on transformer-based models and how they are implemented, their advantages, and the outcomes. Each paper is vetted in the direction of underlining its contributions to improving sentiment classification across diverse linguistic environments addressing challenging situations such as informal use of language, code-switching, and complications in low-resource languages. Moreover, this review discusses the identified limitations of these studies and indicates potential avenues in future research toward filling in the existing gaps and enhancing the effectiveness of the transformer-based approach in multilingual sentiment analysis.

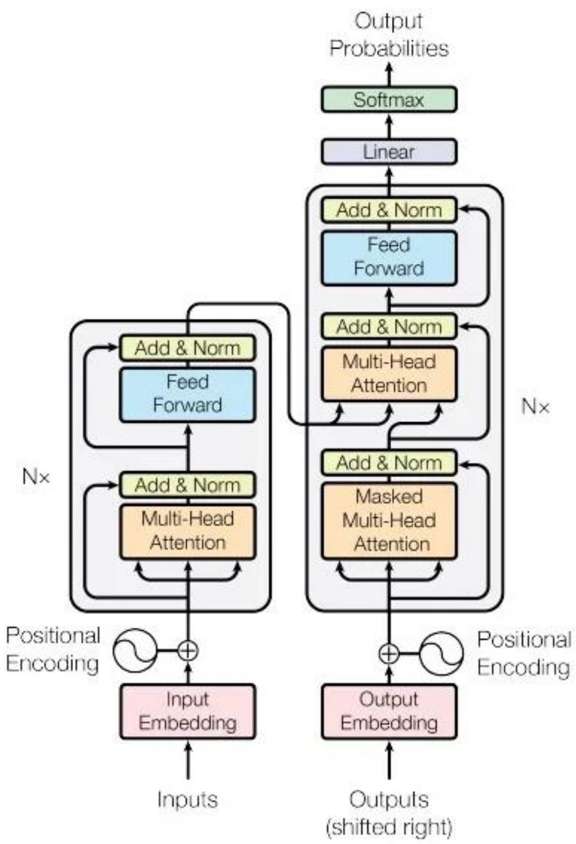


Figure 1: Transformer Based Model Architecture [7]

The sentiment analysis leverages textual data to represent user sentiments and opinions, most of which are based on social media platforms. For a long time, traditional models, such as Naïve Bayes, SVM, and RNNs, struggled with scalability and semantic understanding of informal multilingual text [10]. Considering the demand for processing user-generated multilingual content, transformer-based models became the basis for advancements in sentiment analysis.

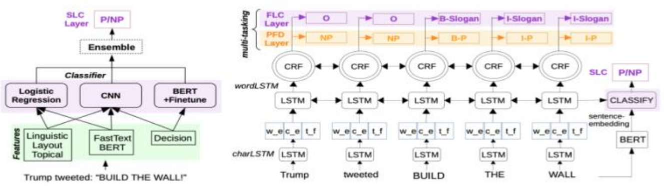


Figure 2: Overview of BERT-based Emotion Detection Models [11]

Acheampong et al. conducted a detailed review of the transformer models used for text-based emotion detection and focused on what BERT and its derivatives introduced [11]. Their work stresses the difficulties that arise in text data when one tries to capture the contextual nuances present in emotional expressions, especially when the text data is informal. They pointed out the superiority of transformer-based models over traditional recurrent neural networks, especially in handling context-dependent features of language very effectively. The authors stated that transformer architectures, BERT in particular, have enhanced emotion detection by using self-attention mechanisms [11]. Such mechanisms allow the models to extract complex contextual patterns, leading to improved sentiment classification accuracy. By training on different datasets with different emotional expressions, BERT-based models perform better than the earlier methodologies, which are stated by [11].

The above figure 2 shows how the BERT model is used in facial recognition. It uses various parts, such as input embeddings, self-attention layers, and the output layer of the binary classifier.

These reviews take on state-of-the-art results always significantly better than the traditional ones with high core metrics like accuracy and F1, thus showing how well emotion perception can be performed [11]. Moreover, the review mentions that the model generalization through training on large diverse datasets is something that can help deal with much informal language or slang that could dominate social media always. Nonetheless, limitations of all advances are tackled up in model interpretability and resource-heavy nature, which may again become barriers to broad application and availability. The authors are doing well to propose that future research explore multi-modal meaning detection methodologies that may integrate textual data with audio or visual input for further enhancement of emotion recognition [11]. This comprehensive review should certainly serve as a reverberant road map for researchers and practitioners in NLP, shedding light on foundational as well as practical grounding so far pursued with respect to transformer models in cognition tasks of emotion detection whilst urging toward untangled research into inventive Albion applications of this area, full of wedging. George Manias et al. presents a work of growing need in multilingual and domain-agnostic NLP solutions [12].

The paper contains a comparative study of multilingual BERT-based classifiers and zero-shot classification models applied to the multilingual Twitter datasets for text categorisation and sentiment analysis. These have been found essential in using multilingual datasets especially from social platforms such as X [12]. According to the research, BERT-based models outperform with the help of fine-tuning but the zero-shot approach is scalable and flexible although precision is reduced. The pre-trained multilingual embeddings provide for efficient cross-language transfer learning for better applications in multilingual NLP tasks [12]. It shows the importance of task-specific fine-tuning, particularly when there are several categories of the multilingual text categorization problem. The future work includes pre-training unknown models and improving precision with manual annotations, particularly on domain-specific tasks such as agriculture and food technology. The outcomes showcase the large area of applicability for these methods within policy making, smart cities, and telecommunications, exhibiting their significance in logical deduction from multilingual datasets to compliment various real-life contexts.

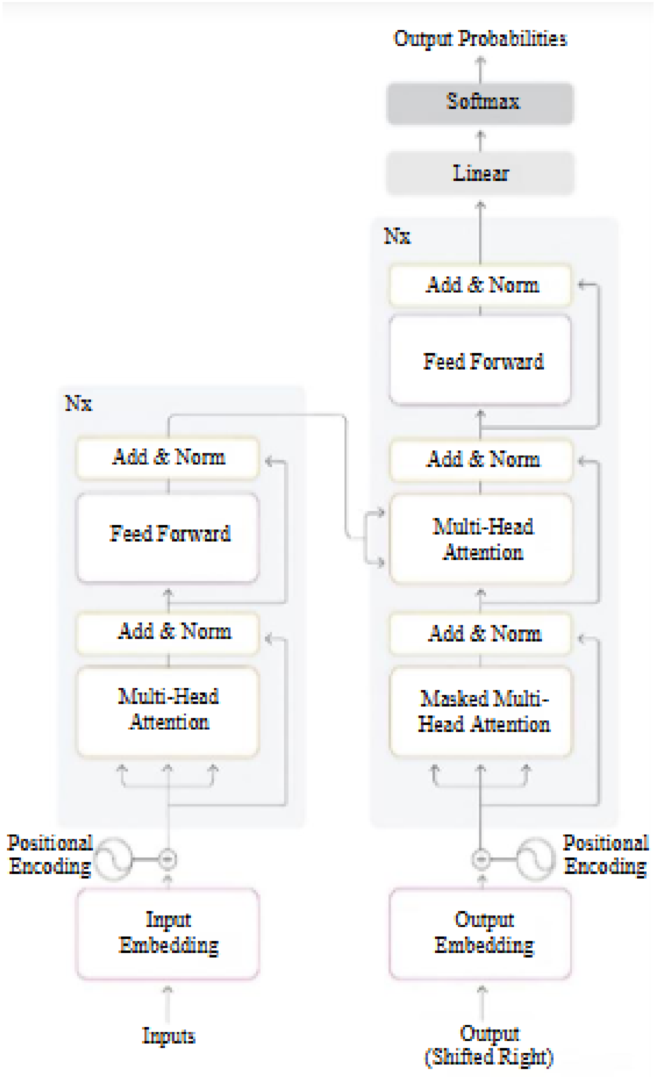


Figure 3: BERT Model [13]

Padmalal et al. have studied the applicability of the BERT model to sentiment analysis for social media text, focusing specifically on its capability to classify posts related to disaster [13]. The experiment made use of a balanced sampled dataset of 10,000 samples labelled with negative, neutral, and positive sentiments and computed the performance of the model concerning accuracy, F1-score, precision, and recall. BERT reached 85% accuracy,demonstrated superior performance compared to conventional approach, including Naive Bayes and Support Vector Machines (SVM). It was also shown that BERT possessed strong precision and recall across all sentiment classes by achieving an F1-score of 85.8%, where it managed linguistic nuances and colloquial expressions of social media data [13]. In this regard, the computational limitation of BERT is stressed, where comparisons are drawn to Word2Vec-LSTM and XLNet with lower accuracy but higher computational cost. Even though the BERT-Bi-LSTM model is accurate, a lot of resources are needed [13]. The model compression, optimization, interpretability through attention visualization, and adaptation for resource-constrained environments shall be future research directions. The conclusion drawn from this study is that the architecture of BERT, leveraging bidirectional processing and attention mechanisms, it offers efficiency in sentiment analysis applications, including social media tracking and analyzing brand perception. However, solving the computational challenges and extending BERT's capability to multimodal and multilingual data could further enhance its practical utility.

There is a lot of interest in code-mixed Indian language sentiment analysis for the increasing presence of multilingual social media. The linguistic diversity posed by India emphasizes the unique problem of code mixing, where it is common that users mix two different languages, as in English blending with native ones in informal writing. Social media, such as X and Facebook are treasure troves of such data, but their informal nature-includingspelling variations and grammatical inconsistencies-makes text analysis problematic [14]. Applications of ML and DL methods have demonstrated effectiveness as in handling these challenges. Studies show that Support Vector Machines (SVM) are still widely used for traditional ML methods, but LSTM and BiLSTM models are leading the DL field as they achieve more accurate results for sentiment classification.

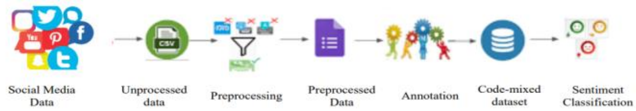


Figure 4: Sentiment Analysis Process of Code-Mixed Data [14]

The scarcity of annotated datasets and language resources continues to be a challenging issue, especially for lesser-studied language pairs like Telugu-English and Malayalam-English [14]. This review specifically emphasizes the usefulness of domain-specific frameworks in improving outcomes for sentiment classification. The results concerning handling complex code-mixed text have been remarkable through various variants of neural network architectures; especially LSTMs. Ahmad et al. provide a strong basis for potential future work toward further developing datasets, tools, and methodologies toward further advancements in multiple languages sentiment analysis.

Examining sentiment in social media content is crucial to understand public opinion, consumer behaviour, and emerging trends in different languages and geographies. Social media now generates tons of multilingual data, and transformer-based models, like BERT and GPT, offer advanced and efficient content analysis methods [15]. These models are very powerful in handling the complexities of multilingual text, giving deep contextual and sentiment understanding. Unlike traditional approaches, transformer models are very good at interpreting informal language, such as slang, emojis, and abbreviations commonly used on social media. Recent developments in these models have focused on fine-tuning them for sentiment classification across multiple languages, ensuring both precision and scalability [15]. Given such applications in the realms of marketing, public health, and even policy analysis-situations often requiring timely insight into sentiment on a global social media scale this makes them immensely valuable. Transformer-based sentiment analysis provides a potentially powerful tool in extracting actionable global social media conversational insights from it.

The literature focused on exploring sentiment detection through deep learning methodologies has undergone significant evolution featuring the replacement of more classical techniques like bag-of-words to those that are more modern and sophisticated like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). While these architectures are more adequate in general terms, they regularly struggle to capture the spatial-temporal dependencies between data sequences [16]. This gap prompted the birth of Graph Convolutional Networks (GCNs) for better representation of structured data, especially in sentiments analysis situations, where the relationship between words and their dependencies on context is decisive.

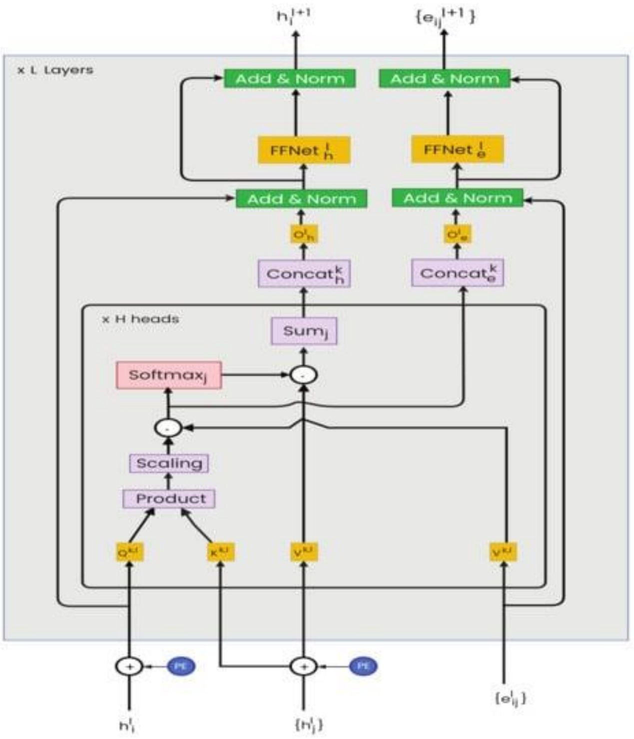


Figure 5: The ST-GCN architecture

The recent advancements in the field include Spatio-Temporal Graph Convolutional Networks (ST-GCNs), incorporating spatial and temporal information to provide significant leverage for sentiment analysis. ST-GCNs are capable of learning from the interrelations between the temporal sequence and spatial graph structure; thus, they provide great potential in capturing intricate patterns in textual data [16]. Recent surveys have shown that they outperform standard deep learning approaches in tasks such as sentiment classification, achieving maxima in enhancement of accuracy and generalizability across other datasets [16]. The architecture of ST-GCN illustrates clearly how the proposed model is described in this study-all together through a sequence of operations typically including positional encoding, feature transformation, sampling strategies, message-passing mechanisms, aggregation steps, and multi-head attention[16]. Besides, several studies indicate hyperparameter optimization, which plays an important role in improving the model's performance. It has also been observed from various studies that the learning rate and number of epochs have impacts on accuracy as well, signifying the tuning of such deep learning models must be done meticulously, especially ST-GCN, for sentiment analysis tasks.

3. Proposed Methodology

The methodology in this proposed model outlines a transformer-based approach toward sentiment analysis on multilingual data from social media. It deals with data gathering, pre-processing, model selection, fine-tuning, evaluation, and even real-time deployment for actionable insights.

Data Collection and Preprocessing: Sentiment analysis for social media will be performed by collecting data through APIs of popular social media platforms including Twitter, Facebook, and Instagram [17]. The dataset contains user posts, comments, hashtags, and other textual content. Considering the fact that data might be multilingual, a language detection algorithm such as langdetect will be used to group posts according to their languages. The text will then be preprocessed, removing URLs, special characters, and stopwords,ensuring the data is clean and analyzable.

Transformer Model Selection: Since the data set is multilingual, cross-lingual transformer models pre-trained on such tasks would be used for this purpose; for example, mBERT, XLM-R, or mT5 can be used, as they support multiple languages and understand contextual relations across languages [18]. These models are used as the base for the sentiment analysis application, leveraging their strong ability to understand languages.

Fine-Tuning the Model for Sentiment Classification: The transformer model selected will be fine-tuned on a dataset with labelled data, where it has categories: positive, negative, and neutral. This can be manually labelled or taken from existing datasets related to sentiment analysis. To further address data imbalance and ensure model performance in diverse languages, it will undergo various data augmentation methods such as back-translation and paraphrasing. A classification head will be attached to the model to predict sentiment [19].

Evaluation and Validation: The model is then tested through typical performance measurements of accuracy, precision, recall, and F1-score. Cross-lingual validation tests whether the model functions well by changing languages or dialects of language. Another approach is performing a domain-specific test using data on real-life social media so as to ensure whether the model has real-time efficiency.

Deployment and Insights: The model will be deployed in real-time environments and will analyse content on social media continuously. Trends in sentiment will be visualized through dashboards that show the shifts in opinion and key topics associated with either positive or negative sentiments [20].

4. Results and Discussion

The reviewed studies collectively highlight the effectiveness of transformer-based and hybrid deep learning models in sentiment classification tasks across varied datasets. LSTM models still show strength in structured language contexts, as evidenced by Qixuan [21], who reported over 98% accuracy in classifying Weibo sentiments. BERT-based approaches generally outperformed traditional models, with Xie [22] and Dhanalakshmi et al. [23] confirming superior results for BERT over RNN, CNN, and LSTM on general and COVID-19-related Twitter data, respectively. Wang [24] further validated this trend, showing a clear performance margin favoring BERT (87.4%) over LSTM (83.2%). Hybrid architectures integrating BERT with BiLSTM, such as those by Smitha et al. [26] and Nkhata et al. [32], achieved accuracy upwards of 92%, with strong results even on complex datasets like SST-5. Talaat [25] demonstrated the benefits of enhancing BERT with BiGRU/LSTM, yielding 91.37% accuracy on Twitter sentiment data. Meanwhile, Pookduang et al. [28] and Semary et al. [31] reported impressive outcomes using RoBERTa, exceeding 96% accuracy, while Sinha et al. [27] achieved 94.3% in financial sentiment using finBERT.

Table 1. A Performance Comparison of Sentiment Analysis Techniques in Prior Studies

Sr. N.	Author(s)	Method	Accuracy / F1
1	[3]Christian et al.	BERT+RoBERTa (Model Averaging)	68%
2	[21]Qixuan	LSTM	98.31% accuracy, 98.28% F1
3	[22]Y. Xie	BERT, RNN, CNN	BERT: 92.27%
4	[23]Dhanalakshmi et al.	BERT, LSTM	BERT: 78–92%; LSTM: 71–81% across categories
5	[24]Wang	BERT, LSTM	87.4% (BERT), 83.2% (LSTM)
6	[25]Talaat	Hybrid BERT	91.37%
7	[13] Padmalal et al.	BERT + Bi-LSTM + Dilated CNN	85.30%
8	[26]Smitha et al.	BERT + BiLSTM	92.64% accuracy, F1=91.46%
9	[32]Nkhata et al.	BERT + BiLSTM	97.67% (IMDb), 59.48% (SST-5)
10	[28]Pookduang et al.	RoBERTa	96.3% accuracy, 98.1% F1
11	[29]D. Nayyar et al.	RoBERTa + RNN	84.60%
12	[30]Rahman et al.	RoBERTa + BiLSTM	92.36% (IMDb), 82.25% (Sent140)
13	[31]Semary et al.	RoBERTa + CNN + LSTM	96.3% (IMDb), 94.2% (Twitter)
14	[27]Sinha et al.	RoBERTa, finBERT	94.30%
15	[15]Miah et al. (2024)	Transformer + LLM (Multimodal)	F1 >90%

From a critical standpoint, the shift toward transformer-based hybrids suggests growing recognition of the importance of combining contextual awareness with sequential modeling capabilities. While models like RoBERTa-BiLSTM [30] and RoBERTa-CNN-LSTM [31] deliver top-tier accuracy across IMDb, Twitter, and Sentiment140 datasets, they also introduce higher computational demands. This trade-off between performance and efficiency remains a recurring theme, especially for domain-tuned architectures such as Nayyar et al.'s RoBERTa-RNN [29]. Another consistent observation is the limitation in cross-domain generalizability. Models trained on domain-specific datasets, like those targeting Amazon [28] or movie reviews [32], risk performance degradation when applied to broader sentiment contexts. Despite these challenges, the empirical evidence across all studies confirms that hybrid architectures—particularly those blending RoBERTa or BERT with RNN-based layers—consistently outperform conventional models, reaffirming their position as state-of-the-art for sentiment classification.

5. Conclusion & Scope

5.1. Conclusion

This study delves into the evolving field of sentiment analysis on social media by leveraging transformer-based models to extract meaningful insights from multilingual datasets. With the exponential growth of user-generated content across platforms like Twitter, Facebook, and Instagram—often expressed in diverse languages and informal styles—traditional sentiment analysis techniques have shown limitations in accuracy and scalability. In contrast, transformer architectures such as BERT, mBERT, and XLM-R (XLM-RoBERTa) have demonstrated significant advantages due to their ability to capture deep contextual relationships and cross-lingual patterns.

Through a comprehensive review of recent literature and methodologies, this paper has highlighted the benefits and challenges associated with using transformer models for sentiment classification. While the strengths of these models include improved semantic understanding and robustness to noisy, code-mixed inputs, challenges such as computational cost, model interpretability, and availability of annotated multilingual data persist. The methodology proposed in this paper advocates for a fine-tuned transformer approach that addresses these issues by efficiently processing multilingual social media content, ultimately classifying sentiments into positive, negative, or neutral categories with higher reliability.

What sets this approach apart is its capacity to generalize across languages without needing separate models for each, making it a scalable and cost-effective solution. By capturing nuances in local dialects, emojis, and slang, the transformer-based system outperforms conventional machine learning models that struggle with such informal variability. Additionally, the application of techniques like data augmentation, cross-lingual transfer learning, and real-time sentiment tracking adds further value for industries and researchers aiming to understand public opinion at a global scale.

In essence, this research contributes to the growing body of work that supports the use of advanced NLP models for sentiment analysis in complex linguistic environments.

5.2. Scope:

Future developments in this area could focus on enhancing the efficiency and scalability of transformer-based models, particularly for deployment in low-resource or real-time applications. Model compression techniques and hardware optimization can make these solutions more accessible. Additionally, improving model interpretability—through attention visualization and explainable AI—would increase trust and transparency, especially in critical domains such as public health and governance.

Another important direction is the inclusion of multimodal data sources, such as image and audio-based sentiment cues, to enrich text-based analysis and offer deeper emotional understanding. There is also a pressing need to develop larger, annotated multilingual datasets, especially for underrepresented languages and dialects. Finally, incorporating privacy-preserving machine learning techniques would allow sentiment analysis tools to function ethically in compliance with data protection regulations.

Together, these future enhancements will not only strengthen the proposed approach but also expand its impact across a wide array of sectors including marketing, policy-making, disaster response, and international relations.

Author's statements:

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request. For experimental replication, public datasets from platforms like Twitter or Kaggle were also referenced, in accordance with respective usage policies.

Conflict of Interest

The authors declare that they have no competing interests in relation to the publication of this manuscript.

Funding Source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author's Contributions

Sunny Suresh Nahar: Conceptualization, Methodology, Literature Review, Draft Writing. **Dr. Prashant P. Agnihotri**: Supervision, Technical Review, critical revision and Approval.

Acknowledgment

The authors express sincere gratitude to the School of Technology, Swami Ramanand Teerth Marathwada University Sub-Campus, Latur, for providing the research environment. Special thanks to peer reviewers and editorial members for their constructive feedback.

References

[1] Bing Liu, "Sentiment Analysis and Opinion Mining", Springer Nature, India, pp.**1-250**, **2022**.

[2] Israa Khalaf Salman Al-Tameemi, et al., "A Comprehensive Review of Visual–Textual Sentiment Analysis from Social Media Networks", *Journal of Computational Social Science,* Vol.**7,** Issue.**3,** pp.**2767-2838**, **2024**.

[3] Christian H., Suhartono D., Chowanda A., Zamli K. Z., "Text Based Personality Prediction from Multiple Social Media Data Sources Using Pre-Trained Language Model and Model Averaging", *Journal of Big Data,* **Vol.8, Issue.1, pp.1-12**, 2021.

[4] Neeraj Anand Sharma, ABM Shawkat Ali, Muhammad Ashad Kabir, "A Review of Sentiment Analysis: Tasks, Applications, and Deep Learning Techniques", *International Journal of Data Science and Analytics*, Vol.**6,** Issue.**2,** pp.**1-38**, **2024**.

[5] Yongliang Wang, Ali Derakhshan, Ziwen Pan, "Positioning an Agenda on a Loving Pedagogy in Second Language Acquisition: Conceptualization, Practice, and Research", *Frontiers in Psychology*, Vol.**13,** pp.**1-12**, **2022**.

[6] Arghya Bhattacharya, "Breaking Language Barriers: A Study On Advancing Aspect-Based Sentiment Analysis for Low Resource Languages", International Institute of Information Technology, India, pp.**1-100**, **2023**.

[7] Sehgal, Sats. "Sentiment Analysis with Transformer Models with a Twist of Psychology." *Medium*, Towards Data Science, **2021**.

[8] Ansar, Wazib, Saptarsi Goswami, and Amlan Chakrabarti. "A Survey on Transformers in NLP with Focus on Efficiency." *arXiv preprint* arXiv:2406.16893, **2024**.

[9] Pavel Přibáň, Jakub Šmíd, Josef Steinberger, Adam Mištera, "A Comparative Study of Cross-Lingual Sentiment Analysis", *Expert Systems with Applications*, Vol.**247, 2024**.

[10] MA Bashar, Tahayna, "A Coherent Knowledge-Driven Deep Learning Model for Idiomatic-Aware Sentiment Analysis of Unstructured Text Using BERT Transformer", Universiti Tunku Abdul Rahman, Malaysia, pp.**1-85**, **2023**.

[11] Francisca Adoma Acheampong, Henry Nunoo-Mensah, Wenyu Chen, "Transformer Models for Text-Based Emotion Detection: A Review of BERT-Based Approaches", *Artificial Intelligence Review*, Vol.**54,** Issue.**8,** pp.**5789-5829**, **2021**.

[12] George Manias, et al., "Multilingual Text Categorization and Sentiment Analysis: A Comparative Analysis of the Utilization of Multilingual Approaches for Classifying Twitter Data", *Neural Computing and Applications*, Vol.**35,** Issue.**29,** pp.**21415-21431**, **2023**.

[13] Padmalal, et al., "Enhancing Sentiment Analysis in Social Media Texts Using Transformer-Based NLP Models", *International Journal of Computer Science and Applications*, Vol**.15,** Issue**.1,** pp**.22-30**, **2024**.

[14] Gazi Imtiyaz Ahmad, et al., "Machine Learning Techniques for Sentiment Analysis of Code-Mixed and Switched Indian Social Media Text Corpus: A Comprehensive Review", *International Journal of Advanced Computer Science and Applications*, Vol**.13,** Issue**.2,** pp**.150-160**, **2022**.

[15] Miah, Md Saef Ullah, et al. "A Multimodal Approach to Cross-Lingual Sentiment Analysis with Ensemble of Transformer and LLM." *Scientific Reports*, **2024**.

[16] Barakat AlBadani, et al., "Transformer-Based Graph Convolutional Network for Sentiment Analysis", *Applied Sciences*, Vol**.12,** Issue**.3,** pp**.1-12**, **2022**.

[17] Chiara Zucco, et al., "Sentiment Analysis for Mining Texts and Social Networks Data: Methods and Tools", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol**.10,** Issue**.1,** pp**.1-15**, **2020**.

[18] Jean, Guillaume. "Cross-Lingual Transfer Learning for Low-Resource NLP Tasks: Leveraging Multilingual Pretrained Models." **2023**.

[19] K. Sangeetha, D. Prabha, "Sentiment Analysis of Student Feedback Using Multi-Head Attention Fusion Model of Word and Context Embedding for LSTM", *Journal of Ambient Intelligence and Humanized Computing*, Vol**.12,** Issue**.3,** pp**.4117-4126**, **2021**.

[20] A. Shaji George, T. Baskar, "Leveraging Big Data and Sentiment Analysis for Actionable Insights: A Review of Data Mining Approaches for Social Media", *Partners Universal International Innovation Journal*, Vol**.2,** Issue**.4,** pp**.39-59**, **2024**.

[21] Y. Qixuan, "Three-Class Text Sentiment Analysis Based on LSTM", In the Proceedings of the *2024 International Conference on Artificial Intelligence and Data Science*, China, pp**.112-117**, **2024**.

[22] Y. Xie, "Sentiment Analysis of Xiaohongshu Texts Based on the RoBERTa Model", *Applied Computing and Engineering*, Vol**.11,** Issue**.2,** pp**.75-80**, **2024**.

[23] Dhanalakshmi, et al., "COVID-19 Twitter Data Analysis using LSTM and BERT Techniques", *International Journal of Engineering Trends and Technology*, Vol**.72,** No**.1,** pp**.45-52**, **2024**.

[24] X. Wang, "Analysis of Neural Networks in Text Sentiment", *Applied Computing and Engineering*, Vol**.76,** pp**.55-60**, **2024**.

[25] A. S. Talaat, "Sentiment Analysis Classification System Using Hybrid BERT Models", *Journal of Big Data*, Vol**.10,** Issue**.1,** pp**.12-19**, **2023**.

[26] K. Smitha, et al., "Deep Learning Approach for Sentiment-Based Rating Prediction", *International Journal of Cloud Computing and Database Management*, Vol**.5,** Issue**.2,** pp**.100-108**, **2024**.

[27] A. Sinha et al., "SEntFiN 1.0: Financial Sentiment Analysis Using finBERT and RoBERTa," *arXiv preprint*, arXiv:2305.12257, 2023.

[28] N. Pookduang, et al., "Sentiment Analysis of Amazon Reviews Using RoBERTa", *Engineering, Technology & Applied Science Research*, Vol**.15,** pp**.50-58**, **2025**.

[29] D. Nayyar, N. Chandel, R. Kaur, "Enhanced RoBERTa-RNN Model for Sentiment Analysis", *Multimedia Tools and Applications*, Vol**.83,** pp**.18765–18784**, **2024**.

[30] H. A. Semary, et al., "Sentiment Analysis Using RoBERTa-CNN-LSTM", *Frontiers in Human Neuroscience*, Vol**.17,** pp**.1-10**, **2023**.

[31] A. Nkhata, et al., "Enhanced BERT-BiLSTM Model for Improved Sentiment Classification", In the Proceedings of the *2025 International Conference on Artificial Intelligence Systems*, South Africa, pp**.89-94**, **2025**.

AUTHORS PROFILE

Sunny Nahar earned his B.Sc IT and MCA, from University of Mumbai. He has published few research papers in reputed international journals and conferences including IEEE. Research interests in areas like Cloud Computing, Data Mining, and Machine Learning. He has teaching experience.

Dr. Prashant P. Agnihotri has completed M.Sc. in computer Science and qualified UGC-SET in Computer Science and Applications. He has awarded Ph.D. degree in Computer Science. He is working as Associate Professor in the School of Technology, S.R.T.M. University Nanded, Sub-Campus, Latur. He has published more than 15 research articles in reputed journal. His areas of interests are Pattern Recognition (Speech), Signal Processing, Social Media and Core Computer Science subjects.

