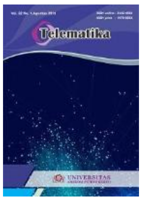
Telematika – Vol. 18, No. 2, August (2025) pp. 203-212 e-ISSN 2442-4528 | p-ISSN 1979-925X



Available online at :
http://ejournal.amikompurwokerto.ac.id/index.php/telematika/

Telematika

Accreditated SINTA “2” Kemenristek/BRIN, No. 85/M/KPT/2020

UNIVERSITAS AMIKOM PURWOKERTO

Application of Fine-Tuned Models in Sentiment Analysis of News: A Systematic Literature Review

Roni Habibi[1,*], Raul Mahya Komaran[2]

[1,2]Department of Informatics Engineering, Universitas Logistics and Business International, Bandung, Indonesia, Indonesia

ARTICLE INFO

History of the article:
Received April 1, 2025
Revised May 5, 2025
Accepted August 27, 2025

Keywords:
Fine-Tuned Models
Transformer
Sentiment Analysis
BERT
Computational Efficiency
Systematic Literature Review

Correspondence:

E-mail: roni.habibi@ulbi.ac.id

ABSTRACT

This study aims to evaluate the application of fine-tuned Transformer-based models for sentiment analysis of news texts through a Systematic Literature Review (SLR). A total of 39 peer-reviewed articles published between 2019 and 2024 were collected from the Scopus database using predefined inclusion and exclusion criteria. The review focuses on three research questions: accuracy improvement (RQ1), implementation challenges (RQ2), and computational efficiency (RQ3). The findings reveal that fine-tuned models, particularly BERT and its variants, consistently outperform traditional machine learning approaches with an average reported improvement of up to 8% in sentiment classification accuracy. However, several major challenges were identified, including high computational cost, limited availability of annotated domain-specific data, and reduced performance in multilingual or dialect-rich contexts. Various optimization strategies have been introduced in the reviewed studies, such as knowledge distillation, model compression, and hybrid approaches that integrate deep models with lexical-based techniques. Overall, this review highlights that while fine-tuned models provide notable performance gains, their scalability and deployment remain constrained by resource requirements. The findings provide a structured reference and research direction for developing more efficient, adaptive, and scalable sentiment analysis solutions for real-world applications such as news monitoring, financial analytics, and automated content moderation.

1. INTRODUCTION

The advancement of the Fourth Industrial Revolution has brought transformative changes across numerous fields, particularly in Natural Language Processing (NLP) (trimaitis, Stefanovi, Ramanauskait, & Slotkien, 2021), which has become increasingly essential in tasks such as sentiment analysis. Sentiment analysis (SA) (Rani & Kumar, 2018), or opinion mining, is the process of identifying and classifying subjective information expressed in text into categories such as positive (de Oliveira, Pisa, Lopez, de Medeiros, & Mattos, 2021), negative, or neutral. With the proliferation of online content, sentiment analysis has evolved into a crucial tool for applications in marketing (Tang, 2015), politics, customer service, and media analysis (Yadav, Kumar, & Kumar, News-based supervised sentiment analysis for prediction of futures buying behaviour).

Early approaches to sentiment analysis predominantly relied on lexicon-based methods, which utilize predefined sentiment dictionaries to detect polarity in text data (Zhao, et al., The Study on the Text Classification for Financial News Based on Partial Information, 2020). These methods are known for their interpretability and simplicity (Shen, 2020). However, they often fail to account for contextual nuances and exhibit limited generalizability across different domains (Taboada, 2011).

To overcome these limitations, statistical and machine learning (ML) models such as logistic regression, decision trees, and ensemble (Gite, et al., 2021) classifiers were introduced (G. Martín, et al., 2021). These models leverage linguistic and semantic features, including n-grams, part-of-speech (POS) tagging, and word embeddings (César, et al., A Systematic Review on Responsible Multimodal Sentiment

Analysis in Marketing Applications, 2024), leading to improved classification accuracy (Strubell, 2019). Nevertheless, these traditional models remain inadequate in capturing long-range dependencies and deeper semantic relationships, especially in complex or domain-specific texts (Kiritchenko, 2014).

The emergence of deep learning models, particularly Recurrent Neural Networks (RNNs) and their variants such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), marked a turning point in sentiment analysis (Dridi, Atzeni, & Reforgiato Recupero, 2018). These models excel in sequential text modeling by retaining context through memory gates (G. Martín, et al., 2021). However, they are not without limitations (Ranathunga & Liyanage, Sentiment Analysis of Sinhala News Comments, 2021). RNN-based models suffer from vanishing gradient problems and lack parallelization capabilities, which lead to increased computational costs and longer training times (Yu & Yang, 2024).

To address these gaps, this study presents a Systematic Literature Review (SLR) of fine-tuned Transformer-based models applied to sentiment analysis of news texts. A total of 39 peer-reviewed articles published between 2019 and 2024 were identified from the Scopus database using predefined inclusion and exclusion criteria. The review is guided by three research questions: (RQ1) How much does the accuracy of fine-tuned models improve compared to traditional models in sentiment analysis of news? (RQ2) What are the major challenges in applying fine-tuned models to sentiment analysis across different types of news text and linguistic settings? and (RQ3) How does the computational efficiency of fine-tuned models compare to other machine learning models in real-world or resource-constrained scenarios? The contributions of this study are threefold: (1) synthesizing empirical evidence on the effectiveness of fine-tuned models for news sentiment analysis, (2) identifying key implementation and deployment challenges along with proposed solutions, and (3) providing a structured overview of optimization techniques and trade-offs that can guide future research and practical development of efficient, scalable sentiment analysis systems.

2. RESEARCH METHODS

This study adopts a Systematic Literature Review (SLR) approach to synthesize empirical findings related to the application of fine-tuned Transformer-based models in sentiment analysis of news texts. The review process follows the PRISMA guidelines, encompassing the formulation of research questions, systematic identification of relevant literature, screening of studies, data extraction, and qualitative synthesis of findings. The literature search was conducted exclusively using the Scopus indexing database, selected for its comprehensive coverage of peer-reviewed journals and conference proceedings in the fields of computing and data science. The search was carried out between January and March 2024, with the publication period restricted to studies published from 2019 to 2024. Relevant articles were identified using keywords associated with sentiment analysis and Transformer-based models, including terms such as "BERT," "GPT," and "RoBERTa." To ensure the quality and relevance of the reviewed studies, only peer-reviewed publications meeting predefined inclusion criteria were considered in the final analysis.

2.1. Inclusion and Exclusion Criteria

To maintain the integrity and thematic focus of this systematic review, we established a rigorous set of eligibility parameters. The inclusion phase prioritized peer-reviewed articles published in the Scopus database between 2019 and 2024, specifically those documented in English. A critical requirement for selection was the empirical implementation of fine-tuned Transformer architectures for sentiment classification, coupled with the reporting of objective performance indicators like accuracy. Any research falling outside these temporal bounds, duplicate entries, or non-peer-reviewed materials were systematically discarded. Furthermore, this review omitted studies that relied exclusively on traditional sentiment lexicons or failed to engage with Transformer-based frameworks, thereby ensuring the final synthesis reflects only the most methodologically sound and contemporary literature

2.2. Article Selection Flow

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure a transparent and systematic article selection process. The stages of article identification, screening, eligibility assessment, and final inclusion are summarized in Table 1.

Table 1. The selection procedure followed PRISMA stages:

Stage	Count
Records identified from Scopus	180
After duplicate removal	134
Excluded after title and abstract screening	53
Full-text assessed	81

Articles meeting all eligibility criteria	**39**

The systematic review process was divided into multiple phases following the PRISMA protocol:

2.3. Data Extraction Procedure

A structured data extraction process was employed to systematically capture relevant information from the selected studies. A predefined data-coding form was developed to ensure consistency across all reviewed articles. The extracted information included publication characteristics such as year of publication, country of origin, publication venue, and journal quartile. In addition, details related to the dataset type were recorded, covering various forms of news data such as headlines, full-text articles, and domain-specific news including financial and political content.

The extraction process also documented the types of models applied in each study, with particular attention given to fine-tuned Transformer-based architectures such as BERT, RoBERTa, GPT, and hybrid modeling approaches. To enable comparative analysis, evaluation metrics reported by the studies, including accuracy, precision, recall, and F1-score, were systematically collected. Furthermore, notes related to deployment considerations and computational aspects, such as training requirements and system constraints, were included to provide additional technical context.

All extracted data were tabulated and cross-validated by the research team to minimize inconsistencies and ensure the reliability of the collected information. The key variables and their corresponding examples used during the data extraction process are summarized in Table 2.

Table 2. Variables coded include:

Category	Field Examples
Model	BERT-base, BERT-large, RoBERTa-base
Dataset	Financial news, COVID-19 news, political coverage
Performance	Improvement %, accuracy, F1
Limitations	GPU needs, latency, training time

2.4. Synthesis Method

2.4.1. Formulation of Research Questions (RQs)

The review was guided by three primary research questions (RQs) to explore the effectiveness, challenges, and efficiency of Transformer-based models in sentiment analysis. These questions aimed to:

RQ1: Assess the improvement in accuracy when using fine-tuned models compared to traditional models in sentiment analysis.

RQ2: Identify the major challenges in applying fine-tuned models to sentiment analysis across different types of text.

RQ3: Compare the computational efficiency of fine-tuned models with other machine learning models.

2.4.2. Search Strategy

The search strategy involved identifying relevant keywords and visualizing them using WordCloud techniques. As illustrated in Figure 1, these visualizations helped confirm the centrality of key terms such as “BERT,” “GPT,” “Transformer,” and “sentiment analysis,” ensuring comprehensive coverage of the research domain. This Figure 1 displays a WordCloud generated from the variable extraction of the reviewed articles. The WordCloud helps visualize the most frequently used terms in the dataset, providing insight into the core concepts discussed across the selected studies. Words that appear larger indicate their higher frequency, which suggests their importance and relevance in the context of Transformer-based sentiment analysis. This visualization aids in understanding the key themes and trends within the literature.

Figure 1. WordCloud Visualization Based on Variable Extraction from Reviewed Articles

Figure 2 showcases a WordCloud derived from the metadata keywords of the articles indexed in Scopus. The keywords are drawn directly from the article abstracts and titles, representing the main topics covered in the research. The size of each word indicates its frequency in the metadata, helping to highlight the central topics and research focus within the sentiment analysis domain. This figure 2 enables a clearer identification of the major keywords and methodologies prevalent in the literature.

Figure 2. WordCloud Visualization of Metadata Keywords Extracted from Scopus

Figure 3 presents the results of keyword identification based on the articles extracted during the review process. The keywords were carefully selected to reflect the key themes and areas of study, such as "Transformer," "BERT," "GPT," and "sentiment analysis." This figure helps to map the evolution and centrality of these keywords across the literature, emphasizing the areas of Transformer-based models and their application to sentiment analysis. It provides a clearer picture of the research landscape, illustrating which terms are consistently highlighted in the body of work under review.

KEYWORD IDENTIFICATION

No	Keyword	Raw	ABS	✖	Act	View	SNA	Tag
1	analyst sentiment	41	No	*	Update	View	SNA	Tag
2	Natural Language Processing news	34	No	*	Update	View	SNA	Tag
3	machine learning sentiment news	15	No	*	Update	View	SNA	Tag
4	Systematic Sentiment	197	No	*	Update	View		Tag
5	Sentiment Detection News	32	No	*	Update	View	SNA	Tag
6	Text Classification News	68	No	*	Update	View	SNA	Tag

*Tidak bisa dilakukan lagi karena sudah mengaktifkan ekstraksi path & item

View Result

Figure 3. Keyword Identification Results from Extracted Articles

Figure 4 presents the PRISMA flow diagram, which outlines the step-by-step process used in selecting articles for the systematic literature review (SLR). Starting with 180 articles identified through the literature search, duplicates were removed, and 46 records were excluded. Further screening led to the exclusion of 53 articles that did not meet the inclusion criteria, such as the publication date range (2019–2024), peer-reviewed status, and relevance to Transformer-based sentiment analysis. Finally, 40 articles were selected for full-text review and analysis based on the type of model used, data sources, and performance evaluation metrics. The diagram illustrates the transparent and structured process, ensuring that the studies included in the review adhere to high methodological standards.

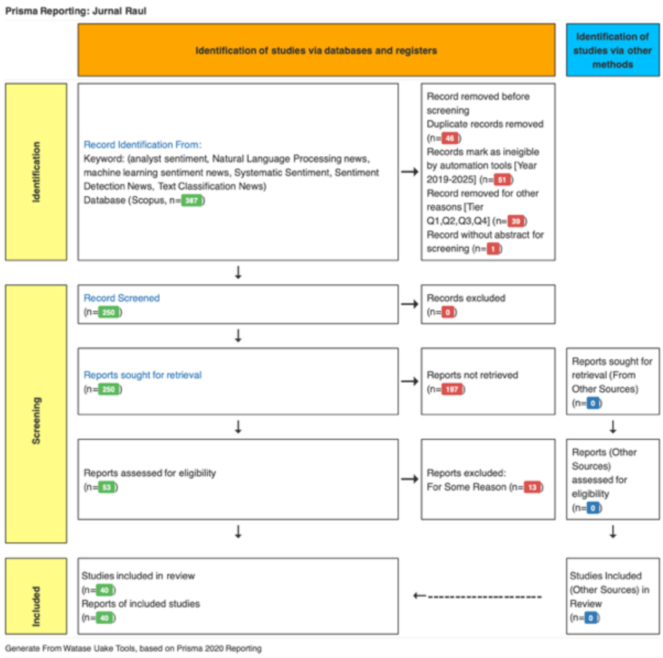


Figure 4. PRISMA Flow Diagram for Article Selection Process

2.4.3. Screening and Selection of Articles

The search strategy followed the PRISMA framework to ensure a systematic and transparent selection of relevant studies. After identifying records from the Scopus database, duplicate articles were removed, and the remaining studies were screened based on their titles and abstracts. Articles that did not meet the predefined inclusion criteria were excluded at this stage. A total of 40 articles were subsequently selected for full-text review. These articles were further categorized based on the type of model employed, distinguishing between traditional machine learning and Transformer-based approaches, as well as the nature of the data sources, including news articles and social media content. Performance metrics reported in the studies were also considered during the screening process to support comparative analysis

2.5. Data Extraction and Analysis

Data extraction and analysis were conducted using the VOSviewer and UAKE platforms to systematically examine keywords, metadata, and thematic patterns across the selected studies. Keyword and metadata extraction enabled the identification of prevailing research themes, frequently used terms, and emerging trends within the literature. To enhance interpretability, WordCloud visualizations were generated to highlight the most prominent keywords and concepts.

The extracted data were then synthesized to evaluate the performance of Transformer-based models in sentiment analysis tasks. This synthesis focused on comparing reported accuracy and evaluation metrics, identifying methodological challenges, and assessing computational efficiency and practical limitations of the models. Through this integrated analysis, the study provides a comprehensive overview of the effectiveness and research gaps related to fine-tuned Transformer-based sentiment analysis in news-related domains.

3. RESULTS AND DISCUSSION

The results of the research are presented in the form of tables, graphs, and figures, integrated with written text to discuss the findings. This section explores new insights, modifications, and the established theories in sentiment analysis using Transformer-based models. In the second phase of the screening process, the relevance and quality of each article were assessed based on the title, abstract, keywords, and supporting metadata such as publication year, citation count, and journal quartile (Q1–Q4). As a result, 39 articles were selected for in-depth analysis. Sentiment analysis is a core task in Natural Language Processing (NLP), aimed at identifying subjective information within textual data. Transformer-based models, particularly BERT and RoBERTa, have achieved state-of-the-art performance in this domain. However, their high computational demands, long training times, and significant energy consumption limit their practicality in real-time and resource-constrained environments.

To address these limitations, several studies have investigated lightweight machine learning models such as Support Vector Machines (SVM), which offer greater efficiency in terms of training time,

https://doi.org/ 10.35671/ telematika.v18i2.3181

inference speed, and hardware requirements. These characteristics make them well-suited for mobile applications, embedded systems, and cloud-based environments. Based on the synthesis of the selected literature, this review highlights the trade-offs between accuracy and computational efficiency across different model types. Table 1 summarizes the key findings related to model performance (RQ1), computational cost (RQ2), and real-world applicability (RQ3):

Table 1. Leading Models Identified Per Research Question

No.	Aspect	Leading Model
1	RQ1: Accuracy	B BERT, SVM
2	RQ2: Computational Cost	B BERT, Hybrid (BERT+SVM/XGBoost), Multimodal BER'
3	RQ3: Practical Use	SVM, BERT

(Wang, Wei, Tian, & Wei, 2024) reported that BERT outperformed nine traditional models in multi-class sentiment classification, achieving an 8% accuracy improvement—particularly in minority classes. However, this improvement comes with the cost of requiring GPUs with ≥12GB memory, several hours of training, and high inference latency. Yu et al. (Yu & Yang, 2024) emphasized that although Transformer models yield high accuracy, their computational inefficiency becomes a major constraint, especially in processing short or mixed-language texts. Their high energy consumption and slow processing make them less viable for large-scale deployment without robust infrastructure. Matrane et al. (Matrane, Benabbou, & Sael, 2023) noted BERT's limitations in multilingual and dialect-heavy contexts, requiring heavy preprocessing and lacking real-time responsiveness. In contrast, models like SVM and NB offer lower computational cost but slightly lower accuracy, making them suitable for real-time use. Overall, lightweight models (e.g., Naive Bayes, NRC Lexicon, ensemble methods) balance performance and efficiency, fitting well in resource-limited applications like mobile and social media analytics.

3.1. Answering Research Questions Using Meta-Analysis

Meta-analysis is a statistical approach that aggregates findings from multiple studies addressing the same research problem, enabling a more comprehensive and reliable conclusion. This review adopts a meta-analytic perspective to systematically answer the predefined research questions (RQ1–RQ3), utilizing both quantitative performance data and qualitative insights.

3.2. Data Extraction

A total of 39 primary studies were selected based on strict inclusion criteria. Numerical and categorical data were extracted using a coding framework and synthesized to compare Transformer-based models with traditional machine learning models across three core aspects: accuracy (RQ1), implementation challenges (RQ2), and computational efficiency (RQ3). This is shown in Table 2.

Table 2. Summary of Comparative Findings

Model Type	Dataset	Acc	Comp. Cost	Sources
BERT (Transformer)	news articles	3,40%	Moderate (transformer model)	(Wang, et al., 2024)
Multimodal Transformers (text, audio, visual)	Various consumer sentiment datasets (2020–2023)	Contextual only	High (multimodal)	(César, et al., 2024)
Ensemble (Bagging and Boosting), CNN, LSTM	4 benchmark datasets (e.g., Twitter, Amazon reviews)	Up to 12%	High (ensemble models)	(Tiwari, et al., 2023)
Financial domain-adapted BERT	203,886 COVID-19 articles from MarketWatch, NYTimes, Reuters (Jun–Jan 2020)	Statistical significance only	High (transformer + fine-tuning)	(Costola, et al., 2023)
Various Deep Learning Models (CNN, RNN, RecNN, DBN, etc.)	4172 tweets (UK COVID-19 related, 2021)	Up to 87.1% test accuracy for sentiment classification	High (LSTM + sentiment integration)	(Alshuwaier, et al., 2022)
LSTM, CNN, BERT, RNN (overview)	100,000+ COVID-19 news headlines (UK, India, Japan, South Korea)	90% validation accuracy	Moderate (empirical analysis)	(Ranathunga and Liyanage, 2021)
KRAKEN-SND (Semantic + Sentiment Conceptual Graph)	92 reviewed studies on student feedback from 2015–2020	LSTM 87%, LR 88% (binary); 67% (multi-class)	Low (systematic review)	(Zhao, et al., 2020)

The results obtained after extraction are sentiment analysis using the Transformer model, which is systematically evaluated based on its accuracy and efficiency, as shown in Figure 5. Among the reviewed studies, Systematic Literature Review (SLR) was the most commonly used method (46.7%), followed by case studies (16.7%) and empirical machine learning experiments (13.3%). Other approaches included NLP pipelines, regression models, and supervised or unsupervised learning techniques. This distribution highlights a strong foundation in review-based research, with a growing trend toward technical experimentation and model evaluation. It reflects the increasing emphasis on performance benchmarking and reproducibility in sentiment analysis using Transformer architectures.

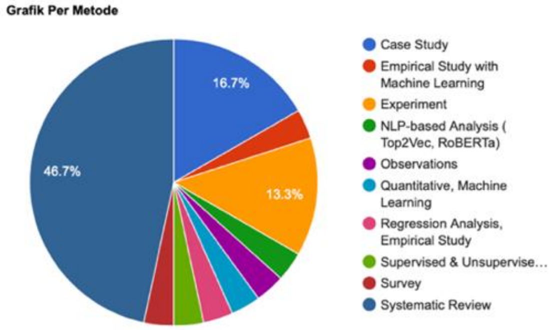


Figure 5. Research Methodologies Employed

In this review, a classification was conducted to group the articles based on their primary research focus. The largest proportion of studies fell under Sentiment Analysis, followed by Machine Learning and Natural Language Processing (NLP). Other topics included Aspect-Based Sentiment Analysis, Fake News Detection, and the application of Deep Learning Models. The classification highlights that sentiment analysis continues to be the central concern of many studies, often intersecting with subfields such as fake news classification and emotion recognition. This underscores the importance of sentiment analysis as a foundational component in broader AI-driven textual interpretation systems.

The figure 6 shows research domain proportions: Sentiment Analysis (19.1%), Machine Learning (17%), NLP (16%), and others such as classification, market prediction, and fake news detection. These domains support RQ1 (accuracy), RQ2 (deployment challenges), and RQ3 (efficiency), highlighting the diverse focus in sentiment analysis studies.

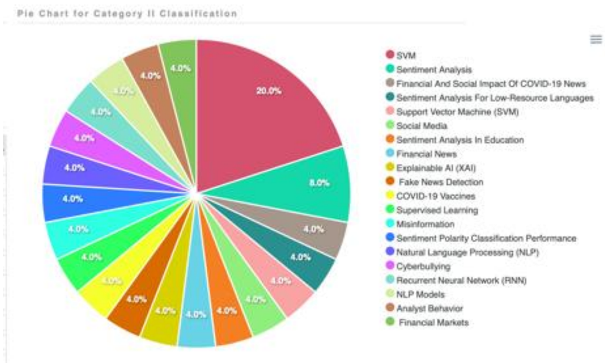


Figure 6. Classification of Research Domains in Reviewed Articles

The figure 7 shows the publication trend from 2019 to 2024 related to the use of Transformer models in sentiment analysis. It can be seen that the number of publications increased significantly in 2021 and peaked in 2022, reflecting the peak of academic interest in the application of BERT, RoBERTa, and similar models. After that, there was a slight decline, likely because the focus began to shift to efficiency and the use of lighter models.

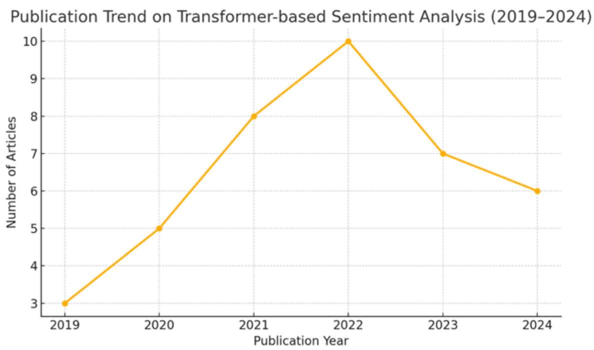


Figure 7. Publication Trend of Transformer-Based Sentiment Analysis (2019–2024)

3.3. Classification Process

This section summarizes the classification process by mapping the research questions to key findings reported in the reviewed studies. The synthesis highlights comparative performance, implementation challenges, and computational considerations of fine-tuned Transformer-based models in sentiment analysis. Table 3 presents an overview of the research questions, their corresponding findings, and the supporting studies identified in the literature.

Table 3. Research Questions, Key Findings, and Supporting Studies

Research Question	Explanation	Supporting Studies
RQ1: How much does the accuracy of the fine-tuned model improve compared to traditional models in sentiment analysis?	Fine-tuned models such as BERT show a significant improvement in sentiment classification accuracy compared to traditional machine learning models like SVM and Naive Bayes. For instance, BERT demonstrated an 8% accuracy gain over nine traditional models in a multi-class classification task.	(Wang, Wei, Tian, & Wei, 2024)
RQ2: What are the major challenges in applying fine-tuned models to sentiment analysis on different types of texts?	A key challenge lies in handling multilingual and dialectal variations, particularly in low-resource languages such as Arabic dialects. These variations often lead to inconsistent results and demand intensive preprocessing efforts.	(Yu & Yang, 2024)
RQ3: How does the computational efficiency of fine-tuned models compare to other machine learning models in sentiment analysis tasks?	Fine-tuned models typically require extensive computational resources, including high-memory GPUs (≥12GB), prolonged training times (several hours to days), and elevated energy consumption. In contrast, traditional models are more efficient and suitable for resource-constrained environments.	(Matrane, Benabbou, & Sael, 2023)

Traditional models like SVM and Naïve Bayes are efficient in resource-limited settings, requiring minimal hardware and offering fast inference. Lexicon-based models (e.g., NRC Lexicon) are low-cost but less effective for nuanced tasks. Ensemble learning balances robustness and accuracy and is widely used in real-time systems. Hardware choices also impact efficiency, with TPUs consuming more energy than GPUs or CPUs. Although Transformers provide superior accuracy, they demand high resources. Recent studies explore transfer learning, XAI, data augmentation, hybrid models, federated learning, and multimodal analysis to improve efficiency and adaptability in complex domains.

4. CONCLUSION

This systematic literature review analyzed 39 peer-reviewed articles (2019–2024) to evaluate the performance, implementation challenges, and computational efficiency of Transformer-based models—particularly BERT—in sentiment analysis, compared to traditional models such as SVM, Naive Bayes (NB), and Random Forest (RF).

For RQ1 on accuracy, Transformer-based models consistently outperformed traditional methods, with BERT achieving an average improvement of around 8% due to its bidirectional contextual encoding and large-scale pretraining. Although SVM showed lower accuracy, it remains relevant because of its simplicity and computational efficiency. Regarding RQ2, key implementation challenges include linguistic diversity, limited data availability for low-resource languages, and poor model interpretability, which complicate deployment in multilingual and real-world settings. For RQ3 on computational efficiency, Transformer models were found to be resource-intensive, requiring substantial GPU memory, long training times, and high energy consumption, making them less suitable for real-time or embedded applications. In contrast, traditional models such as SVM and Naïve Bayes offer faster inference and lower hardware requirements.

Despite its contributions, this review is limited by its reliance on Scopus-indexed, English-language sources, which may exclude relevant studies. Moreover, the rapid evolution of NLP models could render some findings outdated. Future research should explore domain-adaptive pretraining, model compression (e.g., distillation, quantization, pruning, LoRA), and the integration of Explainable AI (XAI) to enhance interpretability. Hybrid models combining lexicon-based and deep learning methods also hold promise, especially for low-resource and edge environments. In summary, while Transformer-based models deliver superior accuracy and adaptability, their broader application requires focused improvements in efficiency, transparency, and scalability for real-world deployment.

REFERENCES

Alshuwaier, F., Areshey, A., and Poon, J. (2022). Applications and Enhancement of Document-Based Sentiment Analysis in Deep learning Methods Systematic Literature Review. *Intelligent Systems with Applications*.

Bengio, Y. S. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 157-166.

César, I., Pereira, I., Rodrigues, F., Miguéis, V. L., Nicola, S., Madureira, A., . . . Oliveira, D. A. (2024). A Systematic Review on Responsible Multimodal Sentiment Analysis in Marketing Applications. *IEEE Access*.

César, I., Pereira, I., Rodrigues, F., Miguéis, V. L., Nicola, S., Madureira, A., . . . Oliveira, D. A. (2024). A Systematic Review on Responsible Multimodal Sentiment Analysis in Marketing Applications. *IEEE Access*.

Chung, J. G. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint*, arXiv:1412.355.

Costola, M., Hinz, O., Nofer, M., and Pelizzon, L. (2023). Machine learning sentiment analysis, COVID-19 news and stock market reactions. *Research in International Business and Finance*.

de Oliveira, N. R., Pisa, P. S., Lopez, M. A., de Medeiros, D. S., and Mattos, D. M. (2021). Identifying Fake News on Social Networks Based on Natural Language Processing Trends and Challenges. *Information*.

Dridi, A., Atzeni, M., and Reforgiato Recupero, D. (2018). FineNews fine-grained semantic sentiment analysis on financial microblogs and news. *International Journal of Machine Learning and Cybernetics*.

G. Martín, A., Fernández-Isabel, A., González-Fernández, C., Lancho, C., Cuesta, M., and Martín de Diego, I. (2021). Suspicious news detection through semantic and sentiment measures. *Engineering Applications of Artificial Intelligence*.

Gbaje, E. S., Agwu, C., Odigie, I. O., and Dauda Yani, S. (2023). Curtailing fake news creation and dissemination in Nigeria Twitter social network and sentiment analysis approaches. *Journal of Information Science*.

Gite, S., Khatavkar, H., Kotecha, K., Srivastava, S., Maheshwari, P., and Pandey, N. (2021). Explainable stock prices prediction from financial news articles using sentiment analysis. *PeerJ Computer Science*.

Hochreiter, S. and. (1997). Long short-term memory. *Neural Computation*, 1735-1780.

Kiritchenko, S. Z. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research, 50*, 723-762.

Lan, Z. e. (2020). ALBERT: A lite BERT for self-supervised learning of language representations. *ICLR*.

Ligthart, A., Catal, C., and Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis a tertiary study. *Artificial Intelligence Review*.

Matrane, Y., Benabbou, F., and Sael, N. (2023). A systematic literature review of Arabic dialect sentiment analysis. *Journal of King Saud University - Computer and Information Sciences* .

Powers, D. M. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. *Journal of Machine Learning Technologies, 2*(1).

Radford, A. e. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8), 9.

Ranathunga, S., and Liyanage, I. U. (2021). Sentiment Analysis of Sinhala News Comments. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

Ranathunga, S., and Liyanage, I. U. (2021). Sentiment Analysis of Sinhala News Comments ACM Transactions on Asian and Low-Resource . *Language Information Processing*.

Rani, S., and Kumar, P. (2018). A journey of Indian languages over sentiment analysis a systematic review. *Artificial Intelligence Review*.

Seifollahi, S., and Shajari, M. (2018). Word sense disambiguation application in sentiment analysis of news headlines an applied approach to FOREX market prediction. *journal of Intelligent Information Systems*.

Shen, W. e. (2020). BERT-based dynamic multi-scale graph neural network for multi-hop reasoning question answering. *EMNLP*, 7041-7051.

Strubell, E. e. (2019). Energy and policy considerations for deep learning in NLP. *ACL*, 3645-3650.

Taboada, M. and. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 267-307.

Tang, D. Q. (2015). Document modeling with gated recurrent neural networks for sentiment classification. *EMNLP*, 1422-1432.

Tiwari, D., Nagpal, B., Bhati, B. S., Mishra, A., and Kumar, M. (2023). A systematic review of social network sentiment analysis with comparative study of ensemble-based techniques. *Artificial Intelligence Review*.

trimaitis, R., Stefanovi, P., Ramanauskait, S., and Slotkien, A. (2021). Financial Context News Sentiment Analysis for the Lithuanian Language. *Applied Sciences*.

Wang, J., Wei, J., Tian, F., and Wei, Y. (2024). A comparative study of machine learning models for sentiment analysis of transboundary rivers news media articles. *Soft Computing* .

Yadav, R., Kumar, A. V., and Kumar, A. (News-based supervised sentiment analysis for prediction of futures buying behaviour). News-based supervised sentiment analysis for prediction of futures buying behaviour. *IIMB Management Review*.

Yu, L., and Yang, L. (2024). News media in crisis a sentiment and emotion analysis of US news articles on unemployment in the COVID-19 pandemic. *Humanities and Social Sciences Communications*.

Zhao, W., Zhang, G., Yuan, G., Liu, J., Shan, H., and Zhang, S. (2020). The Study on the Text Classification for Financial News Based on Partial Information. *IEEE Access*.

Zhao, W., Zhang, G., Yuan, G., Liu, J., Shan, H., and Zhang, S. (2020). The Study on the Text Classification for Financial News Based on Partial Information. *IEEE Access*.