ISSN: 3139-0862

A Comprehensive Review of Transfer Learning-Based Sentiment Analysis Using BERT and Transformer Models

Suvarna Chavan[1], Nilesh Alone[2]

[1]PG Student, Department of Computer Engineering
[2]Associate Professor, Department of Computer Engineering
R. H. Sapat College of Engineering, Management Studies and Research, Nashik, Maharashtra, India
[1] suvarna6808040@gmail.com, [2]nilesh.alone@ges-coengg.org

Abstract

Sentiment analysis is a task in Natural Language Processing (NLP) that detects polarity, opinions, emotions, or attitudes in text data. Traditional systems based on lexicons and hand-crafted features are relatively straightforward, but their adaptability is constrained by vocabulary, context, domain differences, and linguistic variation. Deep learning introduced CNN, LSTM, BiLSTM, GRU, and attention-based representations, while transformer-based pre-trained language models significantly improved contextual transfer. Transfer learning reduces the cost of target-domain annotation by leveraging knowledge from a source domain or a pre-trained model for a target sentiment task. This review examines transfer-learning methods for sentiment analysis, focusing on machine learning, artificial intelligence, natural language processing, Hugging Face Transformers, BERT, and the fine-tuning of pre-trained language models. The literature is organized into transfer-learning taxonomy, model development, BERT-based modifications, hybrid architectures, datasets, evaluation approaches, limitations, and emerging parameter-efficient techniques. The analysis highlights research gaps in source-domain selection, negative-transfer detection, low-resource and code-mixed evaluation, computational efficiency, explainability, continual adaptation, and reproducibility. The review suggests that promising future systems should not focus solely on increasing model size, but should identify relevant knowledge, adapt effectively, and maintain robustness across diverse domains and languages.

Keywords—Machine Learning, Artificial Intelligence, Natural Language Understanding, Opinion Mining, Transfer Learning, BERT, Bidirectional Encoder Representations from Transformers, Hugging Face Transformers, Fine-Tuning, Pre-trained Language Models, Domain Adaptation, Parameter-Efficient Fine-Tuning.

I. INTRODUCTION

Sentiment analysis, often referred to as opinion mining, seeks to computationally detect subjective information and emotions conveyed in text. It is used in product reviews, customer feedback, social platforms, opinion tracking, financial discussions, and various other applications [1]–[3]. Sentiment can be analyzed at the document, sentence, and aspect levels, representing overall polarity, sentence-level polarity, or sentiment associated with a particular entity or aspect [1], [4].

Conventional sentiment-analysis techniques include lexicon-based methods and supervised classifiers such as Naive Bayes and Support Vector Machines (SVMs). These techniques can be computationally efficient, but they often rely on manually designed lexical, n-gram, TF-IDF, or sentiment-lexicon features and may not generalize well when vocabulary, context, or domain changes [1], [3], [5].

Deep learning reduced the reliance on manual feature engineering. CNN architectures can identify local phrase patterns, whereas LSTM, BiLSTM, and GRU models capture sequential relationships; attention mechanisms can highlight informative representations [6]–[8]. Nevertheless, these models may still require considerable labeled data and task-specific training.

Transfer learning changes this approach by enabling knowledge learned from a source domain or task to support a target domain or task. Pan and Yang conceptualized transfer learning in terms of source and target domains and tasks, emphasizing the questions of what to transfer, how to transfer it, and when to transfer it [9]. Surveys of sentiment analysis have subsequently connected transfer learning with cross-domain sentiment classification, representation transfer, parameter transfer, and domain adaptation [4], [10].

The Transformer model introduced a highly parallel self-attention mechanism and enabled extensive pre-training followed by task-specific adaptation [11]. BERT subsequently introduced bidirectional pre-training for language understanding and became a major foundation for transfer-based sentiment analysis [12]. Later models such as RoBERTa, XLNet, ERNIE, and multilingual Transformer architectures extended this framework [13]–[16].

This review emphasizes transfer learning for sentiment analysis rather than merely listing algorithms. It connects model architecture, transfer mechanisms, source–target relationships, data requirements, fine-tuning, evaluation, computational cost, and research gaps [4], [9], [17]–[28]. The review also considers the roles of machine learning, NLP, Hugging Face Transformers, AI, transfer learning, BERT, and fine-tuning of pre-trained language models [12], [29].

II. REVIEW SCOPE AND METHODOLOGY

The review pool consisted of sources from IEEE Xplore, Springer, ScienceDirect, Semantic Scholar, and Google Scholar. The literature was categorized according to sentiment-analysis tasks, transfer methods, representation models, target domains, language and resource levels, dataset characteristics, evaluation practices, challenges, and future research directions [1], [4], [9], [17], [18].

The review focuses on five key areas: (1) principles of sentiment analysis, (2) classification of transfer-learning approaches, (3) BERT and Transformer-based representation

ISSN: 3139-0862

learning, (4) fine-tuning for target domains and parameter-efficient adaptation, and (5) unresolved research questions. This organization maintains a clear distinction between source-domain knowledge and target-domain adaptation [4], [9], [30].

III. SENTIMENT ANALYSIS FOUNDATIONS

A. Stages and Assignments

Document-level sentiment analysis assigns a single sentiment label to an entire document, whereas sentence-level analysis categorizes individual sentences or short phrases. Aspect-based sentiment analysis associates sentiment with a particular entity, feature, or aspect [1], [4]. Fine-grained sentiment analysis can distinguish positive, negative, neutral, mixed, or specific emotional states [1].

B. Conventional Machine Learning

Lexicon-based approaches use sentiment dictionaries, polarity scores, and linguistic rules. Naive Bayes, SVM, logistic regression, and similar classifiers typically use n-grams, TF-IDF, or lexical features [1], [3], [5]. These approaches provide useful baselines with relatively low computational requirements, but their dependence on feature engineering can limit adaptability across domains [1], [5].

C. Deep Learning

CNNs are effective at extracting local features, while LSTM and BiLSTM models capture sequential relationships. GRU/BiGRU provides a less complex recurrent alternative, while hybrid CNN-RNN architectures combine local and sequential information [6]–[8], [24]. Attention mechanisms can highlight informative features; however, attention weights should not automatically be treated as reliable explanations [24], [25].

TABLE I

Approach	Representation	Main Strength	Main Limitation	Transfer Role
Lexicon	Polarity rules	Low cost	Weak contextual generalization	Limited transfer
Naive Bayes / SVM	TF-IDF, n-grams	Fast baseline	Feature engineering	Instance/feature
CNN	Local neural features	Phrase patterns	Limited transfer long context	Feature transfer
LSTM/BiLSTM/GRU	Sequential features	Context sequence	Training and tuning cost	Feature/model transfer
Attention	Salient features	Highlights useful cues	Not a complete explanation	Feature
BERT/Transformer	Contextual self-attention	Strong contextual transfer	Computational and fine-tuning cost	Parameter transfer
Hybrid Transformer	Context + local + sequential	Rich representation	Higher complexity	Hybrid transfer

a. *Development of sentiment-analysis methods compiled from the reviewed literature [1], [6]–[8], [11]–[15], [24], [25].*

IV. TRANSFER LEARNING: CONCEPT AND TAXONOMY

Transfer learning is characterized by a source domain and task together with a target domain and task. Its objective is to improve the target predictive function by leveraging knowledge from the source. A transfer-learning setting differs from conventional learning when the source and target domains or tasks are not identical. The key research questions concern what knowledge should be transferred, how it should be transferred, and when transfer is beneficial [9].

The basic taxonomy differentiates instance transfer, feature-representation transfer, parameter transfer, and relational-knowledge transfer [9]. In sentiment analysis, these mechanisms correspond to selecting relevant source examples, learning shared representations, reusing pre-trained model parameters, and transferring relationships between structured source and target settings [4], [9].

TABLE II.

Transfer Type	Mechanism	Sentiment-Analysis Interpretation	Typical Risk
Instance transfer	Select/reweight source examples	Use source reviews/posts similar to target	Wrong source examples
Feature transfer	Learn shared representation	Reduce domain/vocabulary mismatch	Overgeneralized features
Parameter transfer	Reuse pre-trained parameters	BERT/transformer fine-tuning	Negative transfer/overfit
Relational transfer	Transfer relationships	Structured or heterogeneous domains	Relationship mismatch
Hybrid transfer	Combine mechanisms	Transformer + CNN/RNN + attention	Complexity

b. *Transfer-learning taxonomy for sentiment analysis [4], [9], [10], [21]–[23].*

V. BERT AND TRANSFORMER-BASED TRANSFER

Transformers use self-attention to model relationships between tokens and produce contextual representations [11]. Unlike sequential recurrent processing, the Transformer architecture supports extensive parallel computation and has provided the foundation for large-scale pre-trained language models [11], [29].

BERT, which stands for Bidirectional Encoder Representations from Transformers, introduced bidirectional pre-training using deep Transformer representations and enabled fine-tuning for downstream NLP tasks [12]. In sentiment analysis, a BERT model can be initialized from broad pre-training and subsequently fine-tuned on a specific sentiment dataset, reducing the need to learn language representations from scratch [12], [19], [20].

RoBERTa modified the BERT pre-training strategy, whereas XLNet and ERNIE explored different pre-training objectives and approaches to knowledge integration [13]–[15]. Multilingual models improve transfer to languages with limited labeled resources, although cross-lingual and cultural differences remain challenging [16], [32].

The Hugging Face Transformers library provides access to pre-trained Transformer models, tokenizers, and task-specific components. It supports the reuse of Transformer-based NLP models across different tasks and facilitates model-adaptation workflows in research and practical applications [29].

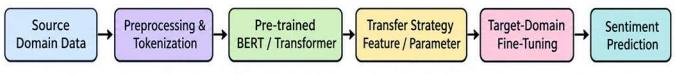


Fig. 1. Review-derived transfer-learning pipeline for sentiment analysis, synthesized from [4], [9], [12], [19], [21], [29].

ISSN: 3139-0862

VI. FINE-TUNING OF PRE-TRAINED LANGUAGE MODELS

Fine-tuning adapts a pre-trained language model to a specific sentiment task by training the model, or a selected subset of its parameters, on examples from the target domain [12], [19], [20]. A task-specific classifier is typically placed on top of the contextual representation and optimized using the target labels.

Full fine-tuning can be effective; however, it may require substantial memory and computational resources and can overfit small target datasets [19], [20]. Recent parameter-efficient fine-tuning (PEFT) methods, including prefix tuning, prompt tuning, and LoRA, reduce the number of updated parameters while retaining most of the pre-trained backbone. These methods are particularly relevant when multiple target domains share a model or when hardware resources are limited [33]–[35].

Fine-tuning performance is influenced by sequence length, batch size, learning rate, number of epochs, class distribution, and the amount of target-domain data. Research on BERT-based sentiment analysis indicates that training configurations can significantly affect classification results [19], [20].

TABLE III

Adaptation Strategy	Trainable Parameters	Data Requirement	Compute	Best Use
Feature extraction	Low	Small–moderate	Low	Fixed representation baseline
Full fine-tuning	High	Moderate	High	Strong target-domain adaptation
Layer-wise fine-tuning	Medium	Moderate	Medium–high	Controlled adaptation
Prefix / prompt tuning	Low	Low–moderate	Moderate	Efficient adaptation
LoRA / PEFT	Low	Low–moderate	Moderate	Resource-constrained transfer

c. *Comparison of strategies for adapting pre-trained models [12], [19], [20], [33]–[35].*

VII. HYBRID AND DOMAIN-SPECIFIC TRANSFER MODELS

Although Transformer models provide strong contextual representations, hybrid architectures remain useful for researchers seeking to combine complementary information. CNN layers can extract local patterns, recurrent layers can capture sequential information, and attention mechanisms can emphasize informative features [24], [25].

The reviewed literature includes BERT-based supervised fine-tuning for sentiment classification, BERT transfer in resource-constrained settings, and hybrid models that combine pre-trained representations with CNN, BiLSTM, BiGRU, or attention mechanisms [19]–[22], [24], [28]. One fusion approach combines ERNIE, TextCNN, BiGRU, and attention mechanisms and uses parameter transfer to address data limitations [21].

Hybridization, however, increases architectural complexity and therefore requires controlled ablation studies and transparent computational reporting. An architecture that performs better on a single dataset should not be considered universally superior without validation across multiple domains [10], [21]–[25].

VIII. DATASETS AND EVALUATION

The frequently used sentiment datasets identified in the reviewed literature include IMDB, Stanford Sentiment Treebank (SST/SST-2), Yelp, the Multi-Domain Sentiment Dataset, Sentiment140, and social-media datasets [4], [17], [36]–[38]. These datasets differ in text length, annotation methods, number of classes, noise levels, and domain distributions; therefore, direct comparisons between studies should be made cautiously [4], [17].

TABLE IV

Dataset / Source	Typical Use	Transfer Challenge	Recommended Metrics
IMDB	Movie-review polarity	Long reviews / domain shift	Accuracy, Precision, Recall, F1
SST / SST-2	Sentence-level sentiment	Short contextual expressions	Accuracy, F1
Yelp	Consumer reviews	Business/domain variation	Accuracy, Macro-F1
Sentiment140	Twitter/social media	Noise, slang, imbalance	Precision, Recall, F1
Multi-Domain Reviews	Cross-domain transfer	Distribution shift	Accuracy, Macro-F1

d. *Representative sentiment datasets and evaluation considerations [4], [17], [19], [20], [36]–[38].*

Relying solely on accuracy can be misleading for imbalanced sentiment datasets. Precision, recall, and F1-score should also be reported, with macro-F1 being particularly useful when minority classes are important [4], [19], [20]. Transfer-learning studies should additionally report the source–target relationship, target-data volume, adaptation method, number of trainable parameters or computational cost, and settings required for reproducibility [9], [16], [23].

IX. COMPARATIVE ANALYSIS OF REVIEWED APPROACHES

TABLE V

Approach	Contextual Understanding	Label Requirement	Compute	Transfer Strength	Main Issue
Lexicon / SVM	Low–moderate	Moderate	Low	Low	Domain portability
CNN / RNN	Moderate	Moderate	Moderate	Moderate	Tuning/data demand
Static embeddings	Moderate	Reduced	Moderate	Moderate	Context independence
BERT fine-tuning	High	Reduced	High	High	Compute/overfit
Hybrid BERT + CNN/RNN	High	Reduced	High	High	Complexity
PEFT / LoRA	High	Low–moderate	Moderate	High	Adaptation configuration
Few-/zero-shot prompting	High	Very low	High	Potentially high	Robustness/generalization

e. *Comparative analysis of sentiment-analysis and transfer-learning approaches [1], [12], [19]–[21], [24], [28], [33]–[35], [39].*

ISSN: 3139-0862

X. ADVANTAGES OF TRANSFER LEARNING FOR SENTIMENT ANALYSIS

A. *Reduced annotation requirements*

Pre-existing knowledge can reduce the amount of labeled target-domain data required [4], [9], [19].

B. *Improved initialization*

Starting from a substantial pre-trained representation reduces the need to learn linguistic features from a random initialization [12], [29].

C. *Cross-domain application*

A single model can be adapted for product reviews, social platforms, public sentiment, or specialized domains [4], [10].

D. *Support for low-resource settings*

Multilingual and cross-lingual representations can facilitate knowledge transfer when annotations in the target language are limited [16], [32].

E. *Adaptive hybridization*

Pre-trained Transformers can be integrated with CNN, BiLSTM, BiGRU, and attention-based methods [21], [24], [28].

F. *Efficient adaptation*

PEFT approaches can reduce the number of trainable parameters and associated memory requirements [33]–[35].

XI. CHALLENGES AND LIMITATIONS

TABLE VI

Challenge	Why It Matters	Research Direction
Negative transfer	Irrelevant source knowledge can reduce target performance	Similarity-aware source selection
Domain shift	Vocabulary, topics and sentiment cues differ	Domain adaptation
Data scarcity	Few labeled target examples	Transfer, augmentation, few-shot learning
Class imbalance	Minority classes distort accuracy	Macro-F1 and class-aware training
Sarcasm / negation	Literal polarity may conflict with intended sentiment	Context/discourse modeling
Multilingual / code-mixed text	Language switching complicates representation	Cross-lingual transfer
High computation	Large transformers require memory and energy	PEFT, distillation, quantization
Interpretability	High confidence does not guarantee understandable evidence	Explainability and uncertainty
Continual change	Sentiment vocabulary evolves over time	Continual learning
Reproducibility	Different splits and tuning settings affect results	Standardized protocols

f. *Key challenges and research directions identified from the reviewed literature [1], [4], [9], [16], [19], [23], [32].*

XII. RESEARCH GAPS

Gap 1. Standardized source-domain selection: Numerous studies select source datasets through experimentation, yet a universally applicable and measurable method integrating semantic similarity, vocabulary overlap, label distribution, and representation distance remains insufficiently standardized [9], [10], [23].

Gap 2. Explicit detection of negative transfer: Transfer is frequently assessed using final accuracy, yet the conditions under which transferred knowledge becomes detrimental are not consistently measured in a systematic manner [9], [23].

Gap 3. Efficient BERT adaptation: Full fine-tuning can be costly for small or resource-limited implementations. A more systematic comparison of full fine-tuning, adapters, prefix tuning, and LoRA using common sentiment benchmarks is needed [33]–[35].

Gap 4. Sentiment analysis for low-resource, multilingual, and code-mixed languages: Transfer techniques can support languages with limited resources, but language-specific morphology, cultural references, Romanized text, and code mixing continue to create domain and representation challenges [16], [32].

Gap 5. Evaluation beyond benchmark accuracy: Sarcasm, negation, implicit sentiment, misspellings, adversarial text, and out-of-distribution data require more comprehensive evaluation [1], [32].

Gap 6. Transferability and explainability: Auditing high-performing Transformer systems can be challenging. Future research should assess evidence, uncertainty, bias, and domain-shift signals rather than treating attention alone as a sufficient explanation [1], [25], [32].

Gap 7. Continual and sustainable adaptation: Real-world sentiment distributions evolve over time, while large models incur costs in memory, latency, and energy. Continual learning and efficient adaptation therefore require joint evaluation [16], [32]–[35].

Gap 8. Reproducible evaluation protocols: Studies should disclose target-data size, source–target relationships, tokenizer and model versions, fine-tuning settings, random seeds, parameter counts, computational costs, and evaluation metrics to enable fair comparison of transfer-learning results [4], [9], [32].

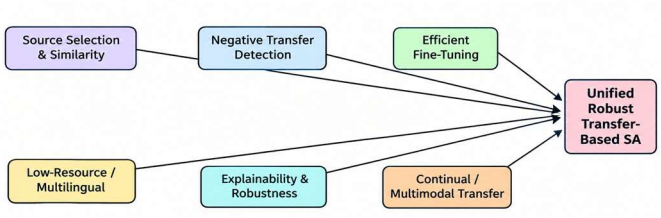


Fig. 2. Research-gap synthesis and proposed review framework derived from [4], [9], [16], [23], [32].

XIII. FUTURE DIRECTIONS

A. *Similarity-Aware Transfer*

Source selection should consider semantic similarity, lexical overlap, label distribution, and representation distance before knowledge transfer to reduce the risk of negative transfer [9], [10], [23].

B. *Parameter-Efficient Fine-Tuning*

Prefix tuning, prompt tuning, and LoRA enable pre-trained models to be adapted by updating fewer parameters, facilitating BERT-family transfer on hardware with limited resources [33]–[35].

ISSN: 3139-0862

C. *Few-Shot and Zero-Shot Sentiment Analysis*

Pre-trained language models can reduce reliance on large labeled target datasets, but evaluation should cover multiple domains and languages rather than relying on a single benchmark [32], [39].

D. *Multilingual and Code-Switched Transfer*

Cross-lingual pre-training, language identification, normalization, and culturally sensitive adaptation are important for less-represented languages [16], [32].

E. *Continual Learning*

Deployed sentiment models need to adapt to new domains and terminology while minimizing catastrophic forgetting [16], [32].

F. *Reliable and Explainable Sentiment AI*

Future systems should provide evidence, confidence estimates, out-of-distribution indicators, and bias audits [1], [32].

G. *Efficient and Sustainable AI*

Compression, distillation, quantization, and PEFT can reduce memory, latency, and energy requirements while retaining useful pre-trained knowledge [32]–[35].

H. *Multimodal Adaptation*

Integrating text with images, audio, video, emojis, or conversational context can capture emotional information that may not be apparent from text alone [32].

CONCLUSION

Transfer learning has transformed sentiment analysis by enabling the reuse of linguistic and contextual knowledge rather than requiring each target task to be learned from scratch. The reviewed literature shows an evolution from lexicon-based and traditional machine-learning methods to CNN/RNN frameworks, attention mechanisms, Transformer-based pre-trained architectures, and hybrid transfer approaches. BERT and related Transformer models provide strong contextual representations and can be adapted to sentiment classification through fine-tuning. The Hugging Face Transformers library has also simplified access to reusable pre-trained models and NLP adaptation workflows. Hybrid models can combine contextual, local, sequential, and attention-based information, but greater architectural complexity does not inherently guarantee improved cross-domain effectiveness. Major open challenges include source-domain selection, negative transfer, domain shift, low-resource and multilingual adaptation, computational cost, interpretability, continual learning, and reproducible evaluation. Future research should therefore focus not only on larger models, but also on transfer-learning systems that determine what to transfer, when to transfer, how to adapt efficiently, and how to remain robust across domains and languages.

REFERENCES

[1] M. Wankhade, A. C. S. Rao, and C. Kulkarni, “A survey on sentiment analysis methods, applications, and challenges,” Artif. Intell. Rev., vol. 55, pp. 5731–5780, 2022.

[2] B. Liu, Sentiment Analysis and Opinion Mining. Morgan & Claypool, 2012.

[3] R. Feldman, “Techniques and applications for sentiment analysis,” Commun. ACM, vol. 56, no. 4, pp. 82–89, 2013.

[4] R. Liu, Y. Shi, C. Ji, and M. Jia, “A survey of sentiment analysis based on transfer learning,” IEEE Access, vol. 7, pp. 85401–85412, 2019.

[5] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, “Lexicon-based methods for sentiment analysis,” Comput. Linguistics, vol. 37, no. 2, pp. 267–307, 2011.

[6] Y. Kim, “Convolutional neural networks for sentence classification,” Proc. EMNLP, pp. 1746–1751, 2014.

[7] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[8] K. Cho et al., “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” Proc. EMNLP, pp. 1724–1734, 2014.

[9] S. J. Pan and Q. Yang, “A survey on transfer learning,” IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, 2010.

[10] E. Guimarães, J. Carvalho, A. Paes, and A. Plastino, “Transfer learning for Twitter sentiment analysis: Choosing an effective source dataset,” Proc. KDMiLe, pp. 161–168, 2020.

[11] A. Vaswani et al., “Attention is all you need,” Proc. NeurIPS, pp. 5998–6008, 2017.

[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” Proc. NAACL-HLT, pp. 4171–4186, 2019.

[13] Y. Liu et al., “RoBERTa: A robustly optimized BERT pretraining approach,” arXiv:1907.11692, 2019.

[14] Z. Yang et al., “XLNet: Generalized autoregressive pretraining for language understanding,” Proc. NeurIPS, 2019.

[15] Y. Sun et al., “ERNIE: Enhanced representation through knowledge integration,” arXiv:1904.09223, 2019.

[16] A. Conneau et al., “Unsupervised cross-lingual representation learning at scale,” Proc. ACL, pp. 8440–8451, 2020.

[17] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” J. Big Data, vol. 3, p. 9, 2016.

[18] F. Zhuang et al., “A comprehensive survey on transfer learning,” Proc. IEEE, vol. 109, no. 1, pp. 43–76, 2021.

[19] N. J. P. Kowsher et al., “Transfer learning for sentiment analysis using BERT based supervised fine-tuning,” Sensors, vol. 22, no. 11, p. 4157, 2022.

[20] A. Areshey and H. Mathkour, “Transfer learning for sentiment classification using BERT model,” Sensors, vol. 23, no. 11, p. 5232, 2023.

[21] Z. Zhao, W. Liu, and K. Wang, “Research on sentiment analysis method of opinion mining based on multi-model fusion transfer learning,” J. Big Data, vol. 10, p. 155, 2023.

[22] K. Jahanbin and M. A. Z. Chahooki, “A hybrid deep implicit neural model for sentiment analysis via transfer learning,” IEEE Access, vol. 12, 2024.

[23] Z. Wang, Z. Dai, B. Póczos, and J. Carbonell, “Characterizing and avoiding negative transfer,” Proc. IEEE/CVF CVPR, pp. 11293–11302, 2019.

[24] M. Wankhade, C. S. R. Annavarapu, and A. Abraham, “CBMAFM: CNN-BiLSTM multi-attention fusion mechanism for sentiment classification,” Multimedia Tools Appl., vol. 83, 2024.

[25] J. Mutinda, W. Mwangi, and G. Okeyo, “Sentiment analysis of text reviews using lexicon-enhanced BERT embedding with convolutional neural network,” Appl. Sci., vol. 13, no. 3, p. 1445, 2023.

[26] P. Atandoh et al., “Integrated deep learning paradigm for document-based sentiment analysis,” J. King Saud Univ. Comput. Inf. Sci., vol. 35, 2023.

[27] A. Iqbal et al., “Sentiment analysis of consumer reviews using deep learning,” Sustainability, vol. 14, no. 17, p. 10844, 2022.

[28] J. Zhang et al., “Text sentiment classification based on BERT embedding and sliced multi-head self-attention Bi-GRU,” Sensors, vol. 23, no. 3, p. 1481, 2023.

[29] T. Wolf et al., “Transformers: State-of-the-art natural language processing,” Proc. EMNLP Demonstrations, pp. 38–45, 2020.

[30] S. Ruder, “Transfer learning in natural language processing,” Ph.D. dissertation, National University of Ireland Galway, 2019.

[31] R. Minaee et al., “Deep learning-based text classification: A comprehensive review,” ACM Comput. Surv., vol. 54, no. 3, 2021.

[32] A. Areshey, “Exploring transformer models for sentiment classification: A comparison of BERT, RoBERTa, ALBERT, DistilBERT, and XLNet,” Expert Systems, 2024.

ISSN: 3139-0862

[33] X. L. Li and P. Liang, “Prefix-tuning: Optimizing continuous prompts for generation,” Proc. ACL-IJCNLP, pp. 4582–4597, 2021.

[34] E. J. Hu et al., “LoRA: Low-rank adaptation of large language models,” Proc. ICLR, 2022.

[35] B. Lester, R. Al-Rfou, and N. Constant, “The power of scale for parameter-efficient prompt tuning,” Proc. EMNLP, pp. 3045–3059, 2021.

[36] A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision,” Stanford University, 2009.

[37] A. Maas et al., “Learning word vectors for sentiment analysis,” Proc. ACL, pp. 142–150, 2011.

[38] R. Socher et al., “Recursive deep models for semantic compositionality over a sentiment treebank,” Proc. EMNLP, pp. 1631–1642, 2013.

[39] T. Brown et al., “Language models are few-shot learners,” Proc. NeurIPS, pp. 1877–1901, 2020.

[40] A. de Arriba, M. Oriol, and X. Franch, “Applying transfer learning to sentiment analysis in social media,” Proc. IEEE REW, 2021.