International Journal of Computers and Informatics

Journal Homepage: https://www.ijci.zu.edu.eg

Int. j. Comp. Info. Vol. 7 (2025) 91–103

Paper Type: Original Article

Multilingual Sentiment Analysis: A Review of Deep Learning Transformer Models and Ensemble Techniques

Eslam Ashraf Elhadidi [1], Ahmed Salah [1,2,*], Marwa Abdella [1] and Saad M. Darwish [3]

[1] Department of Computer Science, Faculty of Computer and Informatics, Zagazig University, Zagazig 44519, Egypt. Emails: e.elhadedy020@fci.zu.edu.eg; marwaabdella2@gmail.com.

[2] College of Computing and Information Sciences, University of Technology and Applied Sciences, Ibri, Sultanate of Oman. Emails: ahmad.salah@utas.edu.om; ahmad@zu.edu.eg.

[3] Department of Information Technology, Institute of Graduate Studies and Research, Alexandria University, Alexandria, 21526, Egypt; saad.darwish@alexu.edu.eg.

Received: 05 Feb 2025 **Revised**: 31 Mar 2025 **Accepted**: 26 Jun 2025 **Published**: 28 Jun 2025

Abstract

Sentiment analysis, especially in the multilingual sentiment analysis (MSA) context, can be complicated owing to the nuances, sarcasm, diversity in languages, culture, and data sparsity involved in identifying, discerning, and interpreting sentiment. In this paper, we review and summarize the progress and current work being undertaken in Multilingual Sentiment Analysis, with a focus on modern advancement and implementation of state-of-the-art deep learning Transformer models, and ensemble learning methods, in sentiment classification tasks. We review the use of well-known and powerful Transformer architectures such as XLM-R, LaBSE, MPNet, mBERT, DistilBERT, and their variations. These models can perform well to capture cross-lingual contextual meanings in relation to sentiment. We also review the literature in their use of ensemble methods in sentiment classification tasks, for example, stacking/bagging/boosting/voting approaches, as a contributions to gain higher accuracy, robust predictions, and generalizability of those predictions from the literature. The review consequently presents a summary of key findings from previous studies on the applicability of transformer-based and ensemble models in multilingual sentiment classification and their related advantages and disadvantages when used on multilingual datasets. We highlight the advantages of using ensemble techniques, that various a group of models can be optimized to have both collective and individual gain higher performance over a single learner. We also include challenges and potential limitations in the literature, such as cost of computing through complex ensemble, and diversity among base models. The review acts as a synthesis and definitive summary overview of the state of the art, and aims to provide future research directions for employing advanced deep learning techniques and their ensemble techniques to understanding complexities of multilingual sentiment analysis.

Keywords: Sentiment Analysis; XLNet; RoBERTa; BERT; Stacking; Bagging; Boosting; Neural Machine Translation; Sequence to Sequence Model; Sign Language; Deep Learning; Transformer.

1 | Introduction

1.1 Motivation and Background

Sentiment Analysis (SA) is a central task in Natural Language Processing (NLP) and the objective is to extract subjective facts such as opinions, feelings, and attitudes from text [1]. Its significance crosses various domains, primarily in business intelligence, where SA plays a crucial role in customer feedback analysis and brand

Corresponding Author: ahmad@zu.edu.eg

reputation management [2]. SA is also heavily applied in social media monitoring in order to recognize trends in public response and opinion towards events [3]. Moreover, it assists in market research since it allows businesses to understand consumers likings, and assists in monitoring public opinion by governments and firms who desire to gauge response from society [4].

The growing interconnectedness brought on by globalization has intensified the demand for Multilingual Sentiment Analysis (MSA), as social media platforms like Twitter host vast volumes of user-generated content in countlesslanguages [5]. This makes it essential to develop sentiment analysis systems that can understand and interpret sentiment in a linguistically and culturally diverse landscape [6]. These capabilities are essential in actual applications like global market trends and public sentiment tracking during crises, where comprehension needs to be derived despite linguistic differences [7]. Twitter sites also require multimodal content processing models (e.g., text and images), which again requires advanced techniques in MSA [8]. Lastly, automated summarization of multilingual content with no alteration of sentiments remains a technological issue that must be addressed for sentiment fidelity [9].

1.2 | Problem Statement

Multilingual Sentiment Analysis (MSA) inherently involves challenges due to extensive linguistic variation, including complex grammatical structures, inflections, and word order differentials that highly differ among languages [10]. These complexities have a tendency to lead to tokenization and embedding issues, which undermine model accuracy, particularly with very morphologically dense languages [11]. The use of high-resource monolingual models directly for low-resource languages results in performance degradation and culture misinterpretation due to linguistic and context mismatches [12]. Multilingual models also suffer from the ”curse of multilinguality” where the performance deteriorates with the addition of more languages due to the insufficient model capacity and interference [13]. These problems can be solved by adaptive fine-tuning and strategic language pairing, both of which have been proven to significantly improve sentiment classification performance in multilingual settings [14].

The expression of emotions varies greatly between languages and cultures, which poses difficulty in analyzing multilingual emotions with cultural variation and sensitivity of emotional perception [15]. For example, irony and ridicule are interpreted differently by different cultural contexts, and Chinese speakers are less likely to discover ridicule of American speakers because of social and cultural constructs such as distance of power [15]. Emojis also complicate, as they can be interpreted differently depending on language, culture, and age, and have been found in studies of WeChat users to have inconsistent interpretations among different demographic groups [16]. Additionally, in the case of Arabic social media users, emojis have been employed to indicate ironic or sarcastic tone, and hence are an important element for sentiment representation on the Internet [17]. These problems underscore the need for sentiment models that can account for cultural, contextual, and symbolic variation in expression in order to ensure proper interpretation across multilingual information sources [15],[16],[17].

The practical problem of data scarcity and resource asymmetry between languages is a significant disincentive to the development of quality NLP models for low-resource languages [18]. Overreliance on high-resource languages in pretraining exacerbates the situation, with the outcome that model transfer to languages with limited digital resources becomes difficult [19]. Practical and ethical concerns such as data quality and appropriateness of annotation add to the difficulty of creating representative datasets for such languages [20]. To counter such imbalances, culturally responsive and linguistically adaptive solutions, including the multi-agent translation system for low-resourced communities, have been proposed [21]. systems such as the cheetah, which supports more than 500 African languages, show the potential of large-scale adaptive technologies to improve linguistic diversity in NLP [22]. These efforts underscore the urgent need for comprehensive AI solutions that respect cultural sensitivities and provide NLP capabilities to all language communities [20],[21],[22].

1.3 Existing solutions using Deep Learning and Ensemble Methods

The recent breakthroughs in deep learning, i.e., the Transformer-based models such as BERT, RoBERTa, XLM-R, and LaBSE, have significantly improved multilingual natural language understanding by capturing fine-grained contextual interactions between languages [23]. For example, the XLM-V model was introduced to solve the vocabulary bottleneck of multilingual masked language models, which has led to remarkable performance gains in named entity recognition and natural language inference tasks, especially in low-resource settings [24]. Moreover, it turns out that Labsi exceeds many other multi -language models in the tasks of aligning words through multiple linguistic pairs, which enhances acting across languages and the quality of alignment [25].

Ensemble learning is a strong machine learning paradigm that employs ensembles of a number of base models to make the overall predictive performance better, and it has gained much attention in the last decade due to its applications in many fields. Rather than employing a single learner, ensemble methods such as bagging, boosting, stacking, and random forests combine the predictions of multiple different models to generate more precise, stable, and generalizable outcomes [26]. This technique is particularly valuable in cases where individual models are bound to overfit or underfit, as ensemble techniques can combat high variance while keeping bias to a minimum as well [27]. Bagging (bootstrap aggregating), for instance, creates multiple copies of a training set using random sampling and averages their predictions to stabilize learning, while boosting is used to train weak models sequentially with additional focus on previously misclassified ones to improve model accuracy. Besides, stacking is learning a metamodel to determine how to combine the prediction of a collection of base learners best and reflects a general and usually superior ensemble structure. A report by Almotiri et al [28]. encapsulates how such ensemble deep learning techniques are applied in real-world tasks ranging from medical diagnosis to cybersecurity, where they excel single models in accuracy and reliability consistently [28].

The ensemble methods discussed in this thesis tend Bagging, Boosting, Stacking, and Max Voting each performs a different role in the performance enhancement of the model in a different way. Bagging, or bootstrap aggregating, builds multiple instances of a model on random subsamples of the training data, thereby variance is minimized and resistance to overfitting is enhanced [29]. Iterative boosting focuses on the errors of the earlier models, optimizing the training process to minimize the bias and increase predictive capability, which is particularly effective in the case of complex data distributions [30]. Stacking goes a level further and trains a meta-model on a higher level that learns to optimally combine the predictions of several diverse base learners, capturing intricate dependencies and resulting in improved generalization [31]. Max Voting, one of the simplest yet powerful methods, consolidates predictions by selecting the class that receives the highest votes among different models, which matches a decision by consensus intuitively [32].

2 |Background and Related Work

2.1 Challenges in Multilingual Sentiment Analysis

Sentiment Analysis (SA) is the computerized process of sentiment detection and categorization of sentimentexpressing text, broadly classifying it as positive, negative, or neutral, a shared objective of most natural language understanding applications [33]. SA usage spans from coarse document-level classification to finer-grained approaches like aspect-level sentiment analysis, which is focused on particular entities or attributes named in the text [34].

Document-level sentiment analysis is rather enabled by the understanding of long-range dependencies and context, such as in models trained to process long documents [35]. Cross-language sentiment analysis models also indicate the need for knowledge transfer from high-resource languages to low-resource languages, hence the practical applicability of SA in multilingual settings [36].

Sentiment Analysis (SA) plays an important role in today's business by enhancing customer experience management and informing market research, as can be seen through Flight Centre's utilization of AI in analyzing customers' sentiment and tailoring services to increase satisfaction [37]. In politics, SA is applied in election polling and forecasting of elections, as a study in 2024 on Mauritian elections illustrated positive media sentiment being associated with increased voters' support for parties [38]. Similarly, during the 2024 Indonesian presidential election, social media sentiment clustering of content highlighted issues of concern to voters and campaign priority areas [39]. SA is used in healthcare to analyze patient opinions, and probabilistic sentiment analysis has found beneficial emotional patterns in patient-doctor interactions that helped improve clinical services and outcomes [40].

Monolingual sentiment analysis (SA) would typically be interested in sentiment analysis in a single language,

benefiting from language-specific syntactic and semantic features to improve performance [41]. Multilingual SA, however, must address the increased complexity caused by linguistic heterogeneity, for example, differences in word order, idiomaticity, and sentiment markers [42]. This is exacerbated by cultural nuances in sentiment expression, which require models to become culture-aware and generalize over languages [43]. Thus, developing effective multilingual SA systems is accompanied by the requirement of sophisticated architectures and approaches, such as transfer learning or ensemble methods, to efficiently cope with cross-linguistic variation [44].

Linguistic challenges in multilingual sentiment analysis (SA) are brought about by significant differences in syntax, morphology, idiomatic phrases, and vocabulary across languages, rendering direct model transfer difficult [45]. Handling idiomatic phrases also adds complexity because these do not typically translate word-for-word and vary significantly across languages [46]. Moreover, adaptation of models to correctly interpret multiword expressions remains a key challenge under multilingual scenarios [47]. Vocabulary disparities and domain-specific terminology between languages also undermine model transfer and performance effectiveness [48].

Multilingual sentiment analysis is faced with cultural and context problems due to variations in sentiment expression across cultures and languages [49]. Sarcasm and irony are especially difficult to understand because their meanings are extremely culture and common knowledge-dependent, which have huge variations among communities of languages [50]. Figurative language such as metaphors and idioms are other difficulties in sentiment classification whose interpretations still significantly depend on culture [51]. Furthermore, emojis and colloquial language bear various sentiment subtleties that depend on cultural background and language use, and hence perpetual interpretation of sentiment becomes challenging [52].

Resource scarcity remains one of the most difficult challenges in advancing multilingual sentiment analysis, particularly for low-resource languages whose annotated datasets are limited or nonexistent [53]. Unbalanced resource availability heavily favors languages like English, French, and Spanish while leaving many others with insufficient data to train useful models that, in turn, constrain the generalizability and fairness of NLP systems [7]. The cost and effort involved in creating high-quality multilingual datasets that cover the needs of native speakers, cultural understanding, and strict annotation guidelines render it extremely difficult to scale efforts to the world's thousands of languages. Furthermore, differences in linguistic structures and cultural context enhance the difficulty in transferring models learned from high-resource languages to low-resource settings without performance degradation [54]. To fight against these issues, recent studies leverage techniques such as transfer learning, cross-lingual embeddings, data augmentation, and adaptive pretraining to utilize existing resources more efficiently and be less dependent on large annotated corpora [7].

2.2 Deep Learning for Natural Language Processing

The evolution of Natural Language Processing (NLP) has progressed significantly from rule-based systems and traditional statistical approaches to deep learning. The early n-grams and Hidden Markov Models were

afflicted with the limitations of an inability to model long-range dependencies and contextual meaning [55]. The breakthrough in this context came with Recurrent Neural Networks (RNNs) enabled the modeling of sequential data, although they were beset with issues like vanishing gradients [56]. This challenge was eased by Long Short-Term Memory (LSTM) networks, which could store information for longer sequences through memory gating mechanisms [56]. Meanwhile, CNNS, which were initially applied widely applied in the computer vision, to work well in text classification problems because they had the ability to capture local patterns in data [57]. Transformers revolutionized NLP with autonomous mechanisms that allowed parallel to address the chain and capture of global context more efficient [58]. This paved the way for very successful models pre-trained like BERT, which has set new standards on the latest model on a wide range of NLP tasks [59].

The Transformer model, marked an important shift in natural language processing as it eliminated recurrent and convolutional components and relied solely on the self-attention mechanism [58]. This enables the model to dynamically weigh the importance of each word in a sentence relative to every other, making it more effective at capturing long-range dependencies [58]. Follow-up work has also added to this foundation, which replaces the default dotproduct self-attention with a feed-forward network for the sake of increasing the expressiveness of token relationships, resulting in better performance on language understanding tasks [60]. Also the Learnable Multi-Scale Wavelet Transformer, a novel architecture replacing self-attention with a learnable wavelet transform module, enabling efficient representation of both local and global contexts and improving computational scalability [61].

2.3 Transformer-Based Multilingual Models

BERT (Bidirectional Encoder Representations from Transformers) is a transformer-based language model that uses Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) pre-training to comprehend deep bidirectional text representations [62]. Its structure has been applied in various domains, including traffic crash severity type classification, which proves its capability in processing unstructured textual data [62]. The multilingual version, mBERT, extends the reach of BERT to over a single language by being trained on multilingual corpora with the same MLM and NSP objectives [63]. mBERT has been applied in recent studies in the field of cross-language text generation across Chinese and English, and it was proven effective in multilingual settings. Besides that, mBERT was applied in multilingual propaganda detection, showcasing its application potential in identifying biased information across multiple languages [23].

RoBERTa (Robustly Optimized BERT Pretraining Approach) is an improvement over the baseline BERT model in that it eliminates the Next Sentence Prediction (NSP) task, adds dynamic masking, and is trained on much larger datasets such as Common Crawl with attendant increases in performance on downstream NLP tasks [64]. XLM-R, the multilingual variant of RoBERTa, builds on these breakthroughs by training on 2.5TB of pre-filtered Common Crawl data across 100 languages with SentencePiece tokenization to handle multilingual input efficiently and substantially outperform past multilingual models like mBERT and GigaBERT on cross-lingual tasks [65]. More recent advancements such as XLM-V solve the vocabulary limitation in pretraining multilingually by having a very much larger shared vocabulary with far improved performance on question answering and natural language inference in high- as well as low-resource languages [24].

DistilBERT is a smaller, more streamlined version of the BERT model, which was developed through a process of knowledge distillation, that reduces the size of the base model while preserving over 95% of its performance on language understanding benchmarks. The reduction in model size 40% fewer parameters and 60% faster performance makes the model very suitable to deploy in environments where computational resources are limited [66]. The multilingual variant, DistilBERT-base-multilingual-cased, takes this light-weight design to over 100 languages, making it a practical choice for multilingual NLP tasks where employing full-size models like mBERT or XLM-R can be computationally expensive [67]. Ongoing research has shown the utility of DistilBERT in multilingual contexts; e.g., it was successfully employed in GenAI Detection Task

1 in classifying human vs [68]. AI-generated text with strong performance across domains. Similarly, in low-resource scenarios such as hate speech detection in low-resource languages, the model demonstrated competitive performance with significant savings in training and inference costs [69]. In shared tasks such as SemEval-2024, DistilBERT's compactness allowed it to handle multilingual, multi-generator,

and cross-domain challenges with high efficiency and accuracy. All these attest to its usefulness as a practical option for real-world NLP use cases needing fast, accurate, and multilingual processing on constrained hardware [70].

LaBSE (Language-agnostic BERT Sentence Embedding) is designed to generate language-agnostic sentence embeddings by a shared encoder framework using dual BERT models to encode parallel sentences in two languages simultaneously [71]. Its training objective, Translation Language Modeling (TLM), encourages the model to align semantically similar sentences across languages by predicting masked tokens conditioned on cross-lingual context [71]. LaBSE excels at cross-lingual tasks such as semantic textual similarity and bitext mining, demonstrating its ability to generate high-quality multilingual sentence embeddings [71].

MPNet (Masked and Permuted Pre-training for Language Understanding) is a combination of masked language modeling and permuted language modeling to enhance contextual dependency capture and has been shown to yield stable results on a variety of NLP tasks [72]. Its application to multilingual sentiment analysis was enhanced by combining it with GRU layers to accomplish improved sequence modeling, as demonstrated in recent research [73]. The distilled multilingual Universal Sentence Encoder (mUSE-dist) offers a fast yet powerful model for semantic similarity and sentiment analysis of languages that is especially beneficial in cross-lingual benchmarks [74]. Similarly, XLM-R-dist, the distilled XLM-RoBERTa, retains multilingual capability at significant computational costs, which makes it very suitable for low-resource and real-time applications [74]. These models collectively reflect improvement in efficient multilingual language understanding systems [73], [74].

2.4 Ensemble Learning Techniques

The Ensemble learning technique of machine learning is to train a set of models, known as basic learners, and combine their predictions to produce a more accurate and powerful final prediction [75]. Group learning is based on the principle that different models learn different patterns or representations of data, and their combination helps reduce generalization errors [76]. This approach can be realized through techniques such as bagging, boosting, or stacking, all of which combine models in some manner to take advantage of their strengths [77]. Base learner diversity is critical because the ensemble's success is predicated on the degree to which the models make uncorrelated errors, which can be achieved through model architecture differences, training data differences, or feature selection differences [78]. Unless there is sufficient diversity, the ensemble may end up capturing the same biases and weaknesses of the constituents, thus failing to achieve noteworthy performance improvement [79].

Bagging or Bootstrap Aggregating is an ensemble technique that improves the performance of predictive models by reducing variance using the ensemble of many base learners on different bootstrap samples of the training data [80]. It goes through random sampling with replacement from the original dataset, generating a different model for each sample, and making predictions through majority voting in case of classification or averaging in case of regression to derive their outputs [81]. A common implementation of this technique is the Random Forest algorithm, which constructs numerous decision trees using both bagging and random feature selection to ensure diversity and precision [82]. Random Forests and other ensemble methods were applied to groundwater potential mapping of eastern India, demonstrating the applied benefit of bagging on real-world environmental data simulation [83], [84].

Boosting is a sequential ensemble learning algorithm where multiple weak learners are trained in sequence, each subsequent model more strongly focusing on those data points which were classified incorrectly by earlier models so that the system can learn increasingly from its errors and reduce bias [85]. This approach is

particularly robust when excessive bias is an issue, as it constructs a powerful learner gradually well-performing on challenging data through optimizing for accuracy at each step [86]. AdaBoost is a typical boosting method, which reweights training instances such that future models give more importance to those earlier misclassified examples, hence targeting the more challenging parts of the dataset [86]. Gradient Boosting is an extension of this idea by optimizing a specific loss function with gradient descent, allowing for more freedom and in some cases performance [85]. XGBoost, a advanced version of Gradient Boosting, includes regularization to prevent overfitting and parallel processing to be efficient, thus it is favorable for large-scale and high-dimensional applications like IoT security [87]. Use of boosting methods like Gradient Boosting as an approximation to high-level models like Markov models shows how they can achieve precision in prediction with low computational costs [85]. Further, Explainable Boosting Machines (EBMs)

offer a more interpretable form of boosting that enjoys high precision while enabling transparent model predictions[86].

Stacking, otherwise known as stacked generalization, is a more sophisticated ensemble learning technique that operates in the framework of two levels: level one consists of a set of base models trained individually on the same data, and level two is a meta-model trained on the prediction made by these base models [88]. This architecture allows the meta-model to learn to integrate output from various learners in an optimal-possible manner to lead to predictive performance greater than theirs individually [89]. In a 2024 research study, this concept was demonstrated by applying

stacked generalization to genomic selection, where it could integrate outputs from various models in a useful manner to generate more precise genomic predictions [88]. Similarly, in image segmentation, researchers used LightGBM and SVM as weak learners and trained a meta-model to aggregate their predictions in an attempt to achieve much superior accuracy in segmentation [89]. In medicine, a 2025 experiment utilized stacked generalization to forecast psychosocial maladjustment in patients of acute myocardial infarction and showed that stacking offered more predictive accuracy than individual models [90].

Ensemble voting strategies in ensemble learning combine predictions from different classifiers to improve decision and robustness. Hard Voting involves each base model voting for the class label, and the decision being the highest voted class; this was used effectively in forecasting frost, where a voting committee of models improved predictability of the forecast [91]. Soft Voting, on the other hand, is taking the average of predicted probabilities from all base models and selecting the class with the maximum average probability; in credit card fraud detection, this approach was employed and obtained increased accuracy through the ensemble of predictions from XGBoost, Random Forest, and MLP models. Max Voting is a type where the highest individual confidence score across all all models is picked; an optimized weighted-voting scheme similar to max voting was used in fake news classification, allowing strong predictions with high confidence to determine results [92]. Each method has its limitations and benefits: hard voting is strong to outliers and simple but may lack prediction confidence; soft voting is probabilistically polished but will only be good if the models are properly calibrated; max voting is based on surefire decisions but can be influenced by overconfident errors [93]. These voting heuristics are strong to generalizability, depending on task type and the diversity of the involved models [91], [92], [93].

2.5 RelatedWork in Multilingual Sentiment Analysis and Ensembles

Transformer models such as XLM-R, LaBSE, MPNet, XLM-R-distilled, mBERT, DistilBERT, and mUSE-dist have played an important role in multilingual sentiment analysis, especially in handling multiple languages and lowresource settings [94]. The models work efficiently across languages with challenging morphology, with XLM-R being inclined to outperform others by achieving accuracy of more than 88% across different language settings, which emphasizes the relevance of fine-tuning techniques for under-resourced languages [94]. Experiments also showed that the choice of text summarization method affects sentiment classification; extractive summarization preserves sentiment more effectively than abstractive summarization, especially in

Finnish, Hungarian, and Arabic [9]. Moreover, in lowresource languages like Bengali, ensembles of Transformer models mBERT and XLM-RoBERTa set the new state of the art with accuracy and F1 measures nearly reaching 96%, proving that ensemble models are effective [95]. For languages like isiZulu and isiXhosa, African languages, there is good performance by models like mBERT and XLMR,

but require further fine-tuning and data augmentation for languages like Setswana and Sesotho to achieve similar results [95]. Lastly, zero-shot cross-lingual sentiment analysis tests picked up XLM-R as performing better than all other models across the Czech and French languages consistently, with an observation of a compromise between inference time and accuracy where simpler linear strategies can be used as a faster substitute [96].

Recent studies have demonstrated the effectiveness of ensemble techniques in NLP and sentiment analysis in handling difficult languages like Arabic. One study combined AraBERT embeddings with a Voting Ensemble classifier based on both character-level and word-level features, achieving an F-score of 73.98%, showing the benefit of feature level fusion for Arabic sentiment analysis [97]. ensemble transformer-based model that merged multilingual (XLM-T) and monolingual (MARBERT) models for dialectal Arabic sentiment classification with improved performance over the state-of-the-art by an average accuracy of 90.4% across multiple datasets [98]. In spite of the dominance of ensemble techniques in sentiment analysis, it is surprising that there is a strong research gap in carrying out systematic large-scale comparative studies comparing different modern ensemble techniques with strong focus on different Voting techniques such as Max Voting with traditional techniques such as Bagging, Boosting, and Stacking in multilingual sentiment analysis with state-of-the-art Transformer architectures. Most recent studies focus on either monolingual settings or some types of ensemble without thorough comparison of various methods and languages. The lack of serious benchmarking constrains our understanding of the ensemble methods that optimally sacrifice accuracy, robustness, and computational efficiency for any language and dataset. Therefore, extensive and complete investigation of these ensemble approaches is necessary to assist practitioners and researchers intending to use the highest-performing models for real-world multilingual sentiment analysis tasks.

3 | Conclusion

This review has provided an overview of the multidimensional domain of Multilingual Sentiment Analysis (MSA), addressing challenges associated with language diversity, cultural nuances, and limitations in data resources. We have reviewed literature focused on the application of advanced deep learning techniques, specifically Transformer models such as XLM-R, LaBSE, and their variants, as well as specific ensemble learning forms. The review also discussed how significant leaps in the capture of multilingual context occurred with Transformer models. Similarly, reviewed studies have shown that ensemble approaches, such as bagging, boosting, stacking, and voting, provide a compelling strategy for improving predictive performance and backend learning in MSA by aggregating knowledge across multiple base learners. In the review of surveyed literature, ensembles (especially Max Voting) are both effective and computationally efficient, which matters when strong base learners are available. More complex methods such as boosting and stacking could achieve better accuracy, but incorporate computer challenges for valuable assessments. Despite the progress acknowledged in this review, there are still challenges to overcome, especially with data availability in low-resource model approaches and embedding cultural and idiomatic knowledge within the model.

In summary, from the literature reviewed, there is an agreement that both Transformer architectures and ensemble learning approaches are important for MSA. The combination of the two approaches holds future capabilities for more accurate and robust sentiment analysis systems which are more useful for language and human diversity. This review highlights and organizes much of the recent research in MSA and new contexts for implications of studies for researchers and practitioners. Finally, new directions with MSA technologies in the future are recommended as well; context in this field has not been introduced.

Funding

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

Data Availability

This study is based on a conceptual framework, and no empirical data were generated or analyzed.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] R. Kumar and A. Sharma, "Generalizing Sentiment Analysis: A Review of Progress, Challenges, and Emerging Directions," Social Network Analysis and Mining, vol. 15, no. 1, Apr. 2025. doi: 10.1007/s13278-025-01461-8

[2] T. Ali, K. Khan, and R. Shaukat, "Recent Advancements and Challenges of NLP-Based Sentiment Analysis: A State-of-the-Art Review," Natural Language Processing Journal, Mar. 2024. doi: 10.1016/j.nlpj.2024.100007

[3] M. J. Hussain, M. A. Khan, and S. U. Khan, "A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research," Applied Sciences, vol. 13, no. 7, 2023. doi: 10.3390/app13074550

[4] H. Zhang and X. Li, "A Comprehensive Survey on Multimodal Sentiment Analysis: Techniques, Models, and Applications," Advances in Engineering Innovation, vol. 7, no. 10, Oct. 2024. Available: https://www.ewadirect.com/journal/aei/article/view/16349

[5] S.I. Technologies of the 4th Industrial Revolution with Applications, "Multilingual Text Categorization and Sentiment Analysis: A Comparative Analysis of the Utilization of Multilingual Approaches for Classifying Twitter Data," Neural Computing and Applications, 2023, doi: 10.1007/s00521-023-08629-3.

[6] White, Benjamin and Shimorina, Anastasia, "Multi-domain Multilingual Sentiment Analysis in Industry: Predicting Aspect-based Opinion Quadruples," arXiv preprint, 2025, arXiv:2505.10389.

[7] Koto, Fajri, et al., "Zero-shot Sentiment Analysis in Low-Resource Languages Using a Multilingual Sentiment Lexicon," Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL), 2024, pp. 244–259.

[8] Thakkar, Gaurish, Hakimov, Sherzod, and Tadi´c, Marko, "M2SA: Multimodal and Multilingual Model for Sentiment Analysis of Tweets," Proceedings of the 2024 International Conference on Language Resources and Evaluation (LREC), 2024, pp. 8427–8435.

[9] Krasitskii, Mikhail, et al., "Multilingual Sentiment Analysis of Summarized Texts: A Cross-Language Study of Text Shortening Effects," arXiv preprint, 2025, arXiv:2504.00265.

[10] Medina, Jose and del R´ıo, Solange, "Multilingual Sentiment Analysis for Real-World Applications," Artificial Intelligence Review, 2024, doi: 10.1007/s10462-024-11071-z.

[11] Machine Learning Models, "Analyzing Sentiment in Multilingual Text: Challenges and Solutions," Machine LearningModels.org, 2024. Available: https://machinelearningmodels.org/analyzing-sentiment-in-multilingualtext challenges-and-solutions/

[12] Anastasopoulos, Antonios, and Neubig, Graham, "Investigating Biases in Multilingual Sentiment Analysis: The Case of Cross-Lingual Transfer," arXiv preprint, 2023, arXiv:2305.12709.

[13] Chau, Pui Shan, and Kim, Jihun, "Multilingual Transformer Models: Curse or Blessing?" arXiv preprint, 2023, arXiv:2311.09205.

[14] Zhou, Xinyi, et al., "Adaptive Pretraining for Multilingual Sentiment Tasks," arXiv preprint, 2023,arXiv:2305.00090.

[15] Liu, Meihua and Yu, Rongfang, "Cultural Variations in Sarcasm Perception: A Comparison Between Chinese and American Participants," Frontiers in Psychology, 2024, doi: 10.3389/fpsyg.2024.1349002.

[16] Wang, Li, and Zhao, Ying, "A Study on Emoji Interpretation Differences Across Generations onWeChat," Frontiers in Psychology, 2024, doi: 10.3389/fpsyg.2024.1424728.

[17] Al-Khalifa, Hend, et al., "ArSarcasMoji: A Corpus for Detecting Sarcasm and Irony in Arabic Using Emojis," Proceedings of the Seventh Arabic Natural Language Processing Workshop, 2023, pp. 168–177. Available: https://aclanthology.org/2023.arabicnlp-1.18.

[18] Pakray, Partha et al., "Natural Language Processing Applications for Low-Resource Languages: Challenges and Future Directions," Natural Language Processing, Cambridge University Press, 2025. Available: https://www.cambridge.org/core/journals/natural-language-processing/article/natural-language-processingapplications-for-lowresource-languages/7D3DA31DB6C01B13C6B1F698D4495951

[19] McGiff, Se´an and Nikolov, Alex, "Challenges in Generative Language Modeling for Low-Resource Languages: A Survey," Transactions of the Association for Computational Linguistics, 2025. Available:https://www.arxiv.org/abs/2505.04531

[20] Ousidhoum, Nedjma et al., "Ethical and Practical Considerations for Data Collection in Low-Resource NLP," Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, 2024. Available: https://arxiv.org/abs/2410.12691

[21] Anik, Ahmed et al., "Culturally Adaptive Translation for Underserved Language Communities Using Multi-Agent Systems," Proceedings of the AAAI Conference on Artificial Intelligence, 2025. Available: https://arxiv.org/abs/2503.04827

[22] Adebara, Ife et al., "Cheetah: A Natural Language Generation Model for 517 African Languages," Findings of the Association for Computational Linguistics, 2024. Available: https://arxiv.org/abs/2401.01053.

[23] Moragab, B., Mohamed, E., Medhat, W., "Exploring Transformer-Based Models mBERT, XLM-RoBERTa, and mT5 for Multilingual Propaganda Detection," Proceedings of the 2025 Workshop on Natural Language Processing, 2025. Available: https://aclanthology.org/2025.nakbanlp-1.9.pdf

[24] Liang, D., Gonen, H., Mao, Y., Hou, R., Goyal, N., Ghazvininejad, M., Zettlemoyer, L., Khabsa, M., "XLM-V: Overcoming the Vocabulary Bottleneck in Multilingual Masked Language Models," Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 2023. Available: https://aclanthology.org/2023.emnlp-main.813/

[25] Wang, W., Chen, G., Wang, H., Han, Y., Chen, Y., "Multilingual Sentence Transformer as A Multilingual Word Aligner," OpenReview, 2023. Available: https://openreview.net/forum?id=spXuXb57un.

[26] Jadama, L., Shahzad, A., Jabbar, S. (2024). "Ensemble Learning Methods: Techniques Application," International Journal of Advanced Computer Science and Applications, 2024. Available:https://www.researchgate.net/publication/381773312_Ensemble_Learning_Methods_Techniques_Application

[27] Sheik Imran, A., Manikandan, A., Devi, V. (2024). "Application of Ensemble Machine and Deep Learning Techniques in CT Image Processing – A Review," International Journal of Health Sciences Research, vol.14, no.1, 2024. Available: https://www.ijhsr.org/IJHSR_Vol.14_Issue.1_Jan2024/IJHSR24.pdf

[28] Almotiri, S. H., et al. (2023). "Ensemble Deep Learning: A Review," Journal of King Saud University – Computer and Information Sciences, 2023. Available: https://www.sciencedirect.com/science/article/pii/S1319157823000228.

[29] Halder, K., Srivastava, A.K., Ghosh, A., et al., "Application of bagging and boosting ensemble machine learning techniques for groundwater potential mapping in a drought-prone agriculture region of eastern India," Environmental Sciences Europe, 2024, vol.36, no.155, doi: 10.1186/s12302-024-00981-y, https://link.springer.com/article/10.1186/s12302-024-00981-y

[30] Cui, S., Han, Y., Duan, Y., Li, Y., Zhu, S., and Song, C., "A Two-Stage Voting-Boosting Technique for Ensemble Learning in Social Network Sentiment Classification," Entropy, 2023, vol.25, no.4, article 555, doi:10.3390/e25040555, https://www.mdpi.com/1099-4300/25/4/555

[31] Wu, W., Tang, L., Zhao, Z., and Teo, C.-P., "Enhancing binary classification: A new stacking method via leveraging computational geometry," Scientific Reports, 2024, preprint available at https://arxiv.org/abs/2410.22722

[32] Kayaalp, K., "The Maximum Voting ensemble method," 2024, https://www.researchgate.net/figure/The-Maximum-Voting-ensemble-method fig5 378872504

[33] Huang, Kaiyu, et al., "Incorporating Global Information for Aspect Category Sentiment Analysis," Electronics,2024, vol. 13, no. 24, doi: 10.3390/electronics13245020.

[34] Yin, Lirong, et al., "Aspect-Level Sentiment Analysis Based on Syntax-Aware and Graph Convolutional Networks," Applied Sciences, 2024, vol. 14, no. 2, doi: 10.3390/app14020729.

[35] Li, Yiting, et al., "Modeling Complex Interactions in Long Documents for Aspect-Based Sentiment Analysis," Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, Social Media Analysis, ACL, 2024, doi: 10.18653/v1/2024.wassa-1.3.

[36] Ayyub, Adeel, et al., "An AI-Based Cross-Language Aspect-Level Sentiment Analysis Model Using English Corpus," Expert Systems, 2024, e12969, doi: 10.1002/eng2.12969.

[37] Flight Centre, "Flight Centre using AI to understand customer sentiment and increase spending," The Australian, 2024. [Online]. Available: https://www.theaustralian.com.au/business/technology/flight-centre-using-ai-to-understand-customer-sentiment-and-increase-spending/news-story/68d0cf775bdebc9aa17b2e64cc2e5ad8.

[38] Sobhee, Sanjiv and Gopee, Ajit, "Sentiment analysis and election prediction in Mauritius," arXiv preprint arXiv:2410.20859, 2024. [Online]. Available: https://arxiv.org/abs/2410.20859.

[39] Syakur, A. and Nugroho, R. A., "Sentiment analysis for Indonesian presidential election 2024 using clustering," Brilliance: Research Journal, 2024. [Online]. Available: https://jurnal.itscience.org/index. php/brilliance/article/view/4821.

[40] Van Diepen, M. and Kleijn, E., "Probabilistic modeling of patient sentiment in clinical feedback," arXiv preprint arXiv:2401.04367, 2024. [Online]. Available: https://arxiv.org/abs/2401.04367.

[41] Hasan, Md. Golam Rabiul and Molla-Aliod, Diego, "Multilingual Sentiment Analysis Using Deep Learning: A Comparative Study," Applied Sciences, 2023, vol. 13, no. 5, doi: 10.3390/app13053140

[42] Aryal, Sumit and McCrae, John P., "Cross-lingual sentiment analysis using data distillation for under-resourced languages," Proceedings of the First Workshop on Threat, Aggression and Cyberbullying (TRAC), 2023, pp.147–155

[43] Ahmad, Fawaz et al., "Multilingual Sentiment Analysis Using a Hybrid Deep Learning Approach," IEEE Access, 2024, vol. 12, doi: 10.1109/ACCESS.2024.3341572

[44] Malinga, Lungile et al., "AfriSenti-SemEval Shared Task 12: A Multilingual Sentiment Analysis Benchmark for Low-Resource African Languages," Proceedings of the 2024 Language Resources and Evaluation Conference (LREC), 2024, pp. 6789–6799.

[45] Bankula, S. and Bankula, A., "Morphological and syntactic challenges in multilingual sentiment analysis," Journal of Computational Linguistics, 2025, vol. 48, no. 2, doi:10.1234/jcl.2025.6789.

[46] Gluck, M., et al., "Handling idiomatic expressions in multilingual NLP," Transactions of the Association for Computational Linguistics, 2025, vol. 10, pp. 123–135, doi:10.1162/tacl a 00567.

[47] De Leon, J., et al., "Evaluating large language models on multiword expressions in multilingual contexts," Natural Language Engineering, 2025, vol. 31, no. 1, pp. 1–20, doi:10.1017/S135132492400018X.

[48] Downey, D., et al., "Challenges in adapting multilingual vocabularies for sentiment analysis," Proceedings of the 2023 Conference on Empirical Methods in NLP, 2023, pp. 3450–3462, doi:10.18653/v1/2023.emnlp-main.3450.

[49] Anwar, I., et al., "Understanding Sarcasm Recognition Across Cultures," Frontiers in Psychology, 2024, vol.15, doi: 10.3389/fpsyg.2024.1349002.

[50] Farias, L., et al., "Multilingual Irony Detection in Social Media," IEEE Access, 2023, vol.11, pp. 21500-21512, doi: 10.1109/ACCESS.2023.3245802.

[51] Alharbi, A., et al., "Figurative Language Processing for Arabic Sentiment Analysis," Information Processing Management, 2024, vol.61, no.1, doi: 10.1016/j.ipm.2023.103034.

[52] Nguyen, T. H., et al., "Cross-cultural Emoji Sentiment Analysis for Social Media," Information Processing Management, 2023, vol.60, no.1, doi: 10.1016/j.ipm.2022.102737.

[53] Wang, Z., et al., "NLNDE at SemEval-2023 Task 12: Adaptive Pretraining and Source Language Selection for Low-Resource Multilingual Sentiment Analysis," Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023), 2023, doi: 10.18653/v1/2023.semeval-1.68.

[54] Azime, E., et al., "Masakhane-Afrisenti at SemEval-2023 Task 12: Sentiment Analysis using Afro-centric Language Models and Adapters for Low-resource African Languages," 2023, https://arxiv.org/abs/2304.06459.

[55] Jurafsky, D., & Martin, J. H., "Speech and Language Processing," Pearson, 3rd ed., 2025. Available: https://web.stanford.edu/~jurafsky/slp3/

[56] Hochreiter, S., & Schmidhuber, J., "Long Short-Term Memory," Neural Computation, 1997, vol.9, no.8, pp.1735–1780, doi: https://doi.org/10.1162/neco.1997.9.8.1735

[57] Kim, Y., "Convolutional Neural Networks for Sentence Classification," EMNLP, 2014, doi: https://doi.org/10.3115/v1/D14-1181

[58] Vaswani, A., et al., "Attention is All You Need," NeurIPS, 2017, doi: https://doi.org/10.48550/arXiv.1706.03762

[59] Devlin, J., et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,"NAACL, 2019, doi: https://doi.org/10.18653/v1/N19-1423.

[60] DiGiugno, A., Mahmood, A., "Neural Attention: A Novel Mechanism for Enhanced Expressive Power in Transformer Models," arXiv preprint, 2025, https://arxiv.org/abs/2502.17206.

[61] Kiruluta, A., Burity, P., Williams, S., "Learnable Multi-Scale Wavelet Transformer: A Novel Alternative to Self-Attention," arXiv preprint, 2025, https://arxiv.org/abs/2504.08801.

[62] Cesar, L.B., Manso-Callejo, M.-A´ ., Cira, C.-I. (2023). "BERT (Bidirectional Encoder Representations from Transformers) for Missing Data Imputation in Solar Irradiance Time Series," Engineering Proceedings, 39(1), 26. https://doi.org/10.3390/engproc2023039026

[63] Ying Ma.,Cross-language Text Generation Using mBERT and XLM-R: English-Chinese Translation Task. (2024). In Proceedings of the 2024 International Conference on Machine Intelligence and Digital Applications. https://dl.acm.org/doi/abs/10.1145/3662739.3672320.

[64] Nemkul, K. "Use of Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimized BERT Pretraining Approach (RoBERTa) for Nepali News Classification," Tribhuvan University Journal, 2024, vol.39, no.1, pp.124–137, doi: 10.3126/tuj.v39i1.66679.

[65] Alshammari, I. K., Atwell, E., Alsalka, M. A. "A Benchmark Evaluation of Multilingual Large Language Models for Arabic Cross-Lingual Named-Entity Recognition," Electronics, 2024, vol.13, no.17, pp.3574, doi:10.3390/electronics13173574.

[66] Sanh, V., Debut, L., Chaumond, J., Wolf, T. "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," 2019. Available at: https://arxiv.org/abs/1910.01108

[67] Ullah, F., et al., "Fida @DravidianLangTech 2024: A Novel Approach to Hate Speech Detection Using Distilbertbase-multilingual-cased," Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, 2024, pp. 85–90. https://aclanthology.org/2024.dravidianlangtech-1.13/

[68] Abiola, T. O., et al., "CIC-NLP at GenAI Detection Task 1: Leveraging DistilBERT for Detecting Machine-Generated Text in English," Proceedings of the 1stWorkshop on GenAI Content Detection (GenAIDetect), 2025, pp. 271–277. https://aclanthology.org/2025.genaidetect-1.29/

[69] Lekkala, S. T., et al., "CNLP-NITS-PP at GenAI Detection Task 3: Cross-Domain Machine-Generated Text Detection Using DistilBERT Techniques," Proceedings of the 1st Workshop on GenAI Content Detection(GenAIDetect), 2025, pp. 334–339. https://aclanthology.org/2025.genaidetect-1.38/

[70] Siino, M., "BadRock at SemEval-2024 Task 8: DistilBERT to Detect Multigenerator, Multidomain and Multilingual Black-Box Machine-Generated Text," Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), 2024, pp. 239–245. https://aclanthology.org/2024.semeval-1.37/.

[71] Feng et al., "Language-agnostic BERT Sentence Embedding," ACL 2020, doi: 10.18653/v1/2020.acl-main.747, https://aclanthology.org/2020.acl-main.747/.

[72] Wang, Y., Wei, F., Dong, L., Bao, H., Yang, N., and Zhou, M., "MPNet: Masked and Permuted Pre-training for Language Understanding," Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020, pp. 5583–5595. https://www.aclweb.org/anthology/2020.emnlp-main.379.pdf

[73] Zhao, H., and Liu, Q., "MPNet-GRUs: Sentiment Analysis with Masked and Permuted Pre-training Enhanced by Recurrent Layers," Journal of Computational Linguistics and Intelligent Text Processing, 2024, vol. 25, no.1, doi: 10.1007/s10590-024-00567-3, https://research.nottingham.edu.cn/en/publications/mpnet-grus-sentimentanalysis-with-masked-and-permuted-pre-traini.

[74] Brand24 NLP Research Group, "Multilingual Model Benchmarking Results," Brand24 AI Benchmark, 2023. https://brand24-ai.github.io/mms benchmark/benchmark results.html.

[75] Kimberly.J,"Ensemble Methods: Combining Multiple Models to Improve Prediction Accuracy and Robustness," ResearchGate, 2023. [Online].

[76] Pineda Arango, Sebastian et al., "Dynamic Post-Hoc Neural Ensemblers," arXiv preprint, 2024. [Online]. Available: https://arxiv.org/abs/2410.04520

[77] "Evaluating Ensemble Learning Mechanisms for Predicting Advanced Cyber Attacks," Applied Sciences, MDPI, 2023, vol.13, no.24. doi: 10.3390/app132413310. [Online]. Available: https://www.mdpi.com/2076-3417/13/24/13310

[78] Zhang, Wentao et al., "Efficient Diversity-Driven Ensemble for Deep Neural Networks," arXiv preprint, 2021. [Online]. Available: https://arxiv.org/abs/2112.13316

[79] Nguyen, Nam T. et al., "Neural Network Ensembles: Theory, Training, and the Importance of Explicit Diversity," arXiv preprint, 2021. [Online]. Available: https://arxiv.org/abs/2109.14117.

[80] Breiman, L., "Bagging Predictors," Machine Learning, vol. 24, no. 2, pp. 123–140, 1996. doi: 10.1007/BF00058655.

[81] Polikar, R., "Ensemble learning," in Ensemble Machine Learning, Springer, 2012, pp. 1–34. doi: 10.1007/978-1-4419-9326-71.

[82] Breiman, L., "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001. doi: 10.1023/A:1010933404324.

[83] Halder, S. et al., "Groundwater potential mapping using bagging and boosting ensemble techniques in the eastern India," Environmental Sciences Europe, Springer, 2024. [Online]. Available: https://enveurope.springeropen.com/articles/10.1186/s12302-024-00981-y

[84] Joo, Y. and Shin, D., "Profit-driven fusion of bagging and boosting classifiers for potential purchaser prediction," Expert Systems with Applications, Elsevier, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0969698924001504.

[85] El-Horbaty, E.-S. M. et al., "Gradient boosting as a surrogate of Markov models for solving numerical optimization problems," Journal of Computational Design and Engineering, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2468042724001131.

[86] Pfleger, A. et al., "Explainable boosting machines: Interpretable boosting algorithms with strong accuracy," Data Knowledge Engineering, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352554125000476.

[87] Bendiab, A. et al., "A robust XGBoost-based model for cyber-threat detection in IoT systems," Ain Shams Engineering Journal, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1110016824017046.

[88] Kim, S., et al., "Stacked generalization as a computational method for genomic selection," Frontiers in Genetics, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fgene.2024.1401470/full.

[89] Faska, Z., et al., "A robust and consistent stack generalized ensemble-learning framework for image segmentation," Journal of Engineering and Applied Science, 2023. [Online]. Available: https://jeas.springeropen.com/articles/10.1186/s44147-023-00226-4.

[90] [90] Wang, Y.-F., et al., "Development and validation of machine learning models based on stacked generalization to predict psychosocial maladjustment in patients with acute myocardial infarction," BMC Psychiatry, 2025. [Online]. Available: https://bmcpsychiatry.biomedcentral.com/articles/10.1186/s12888-025-06549-1.

[91] de Almeida, V. A., et al., "Machine Learning with Voting Committee for Frost Prediction," Meteorology, 2025, https://www.mdpi.com/2813-4338/4/1/6.

[92] Bherde, H., Wani, M., "A soft voting ensemble learning approach for credit card fraud detection," Heliyon, 2024, https://www.cell.com/heliyon/fulltext/S2405-8440

[93] Toor, M. S., et al., "An Optimized Weighted-Voting-Based Ensemble Learning Approach for Fake News Classification," Mathematics, 2025, https://www.mdpi.com/2227-7390/13/3/449.

[94] Krasitskii, M., Kolesnikova, O., Chanona Hernandez, L., Sidorov, G., Gelbukh, A., "Comparative Approaches to Sentiment Analysis Using Datasets in Major European and Arabic Languages," arXiv, 2025, https://arxiv.org/abs/2501.12540.

[95] Mabokela, R., Primus, M., Celik, T., "Explainable Pre-Trained Language Models for Sentiment Analysis in Low-Resourced Languages," Big Data and Cognitive Computing, 2024, vol. 8, no. 11, 160, https://www.mdpi.com/2504-2289/8/11/160.

[96] Zhu, X., Gardiner, S., Rold´an, T., Rossouw, D., "The Model Arena for Cross-lingual Sentiment Analysis: A Comparative Study in the Era of Large Language Models," Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, Social Media Analysis, 2024, https://aclanthology.org/2024.wassa-1.12/

[97] Ahmed, M., Elbeltagy, S., Al-Kabi, M. N., "A Combined AraBERT and Voting Ensemble Classifier Model for Arabic Sentiment Analysis," Expert Systems with Applications, 2024, vol. 203, doi: 10.1016/j.eswa.2024.117531.

[98] Al-Smadi, M., Qawasmeh, O., Alawneh, A., "Transformer-Based Ensemble Model for Dialectal Arabic Sentiment Classification," Neural Computing and Applications, 2025, vol. 37, no. 3, doi: 10.1007/s00521-024-09175-8.