IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 19 April 2024, accepted 4 May 2024, date of publication 8 May 2024, date of current version 17 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3398635

TOPICAL REVIEW

Sentiment Analysis in Low-Resource Settings: A Comprehensive Review of Approaches, Languages, and Data Sources

YUSUF ALIYU[1], ALIZA SARLAN[2], (Member, IEEE), KAMALUDDEEN USMAN DANYARO[1,3], (Member, IEEE), ABDULLAHI SANI B. A. RAHMAN[1], AND MUJAHEED ABDULLAHI[1]

[1]Department of Computer and Information Science, Universiti Teknologi PETRONAS, Seri Iskandar, Perak 32610, Malaysia
[2]Center for Foundation Studies, Universiti Teknologi PETRONAS, Seri Iskandar, Perak 32610, Malaysia
[3]Centre for Cyber-Physical Systems (C2PS), Universiti Teknologi PETRONAS, Seri Iskandar, Perak 32610, Malaysia

Corresponding author: Yusuf Aliyu (yusuf_22005103@utp.edu.my)

This work was supported by a Fundamental study of supervised machine learning techniques for autonomous defect mapping of offshore structures (Cost Centre: YUTP-FRG 015LC0-373), and in part by the Petroleum Technology Development Fund (PTDF).

ABSTRACT The field of low-resource sentiment analysis has seen significant developments in recent years. This research review SLR evaluates the approaches and data sources utilized in low-resource sentiment analysis by deep learning. The primary aim is to discover suitable approaches for future sentiment analysis in low-resource. Our studies explore various languages, models, and data sources expressing a desire to create effective approaches. Our emphasis lies in the critical evaluation of the approaches and the datasets utilized, to identify areas where further research is needed. Our analysis study adds to the existing body of literature reviews, encompassing multilingual low-resource sentiment analysis research spanning from 2018 to 2023. The findings indicate that the transfer learning approach is the most frequently used, followed by word embedding learning and machine translation systems. Additionally, the study shows that social media is the most used platform for data collection, followed by product reviews, movies, and hotels. There has been a significant surge in the adoption of pre-trained transformers, indicating a growing interest in exploring the potential of these models for low-resource languages within the natural language processing (NLP) community. This trend is largely attributed to the novel nature of these models and their feature of being non-labour intensive. However, the scarcity of annotated datasets for such languages remains a major hurdle. finally, these research findings are relevant and informative for any researcher working in the field of low-resource multilingual sentiment analysis. The study introduces a conceptual framework for performing sentiment analysis in low-resource. The study provides a valuable resource for future researchers.

INDEX TERMS Word-embedding, transfer learning, transformer, machine translation, low-resource, sentiment analysis, multilingual.

I. INTRODUCTION

Sentiment analysis or opinion mining [1], is the identification of an individual's opinions expressed in text across different languages related to various topics of interest [2]. The study of users' opinions and attitudes expressed in the written text belongs to the field of Natural Language Processing (NLP). However, the rapid expansion of the internet and consumers' active engagement in sharing, commenting, and discussing on blogs, forums, social media sites, and online shopping portals, have made sentiment analysis a particularly dynamic study subject [3]. Sentiment analysis is a multidisciplinary artificial intelligence problem [4], consisting of three analysis levels: document level, sentence level, and aspect-based level [3], [5], [6], [7], [8], [9], [10]. Moreover, the increased use of social media platforms has poised to unlock novel prospects by leveraging additional data streams to enhance and expand beyond the product review analysis [11].

The associate editor coordinating the review of this manuscript and approving it for publication was Arianna Dulizia.

Additionally, it can help businesses understand their customers' emotions and opinions by analyzing text data on social media posts and reviews. This information can be used to inform marketing and commercial strategies, enabling businesses to communicate more effectively with their customers and stay ahead of the competition [12].

The collection of several languages' text that may be categorized as positive, negative, or neutral polarity is termed a multilingual sentiment analysis [13]. Notably, sentiment analysis is becoming more significant in languages other than English. Its application ranges from business and customer care services where the users are prepared to use their local languages. However, the emergence of social media made people provide different data in different formats daily. Social media is used by many people to share information that blends several languages and cultures [14]. Therefore, relying solely on sentiment analysis conducted in English could pose a substantial risk of overlooking critical insights within written texts [15], [16]. Moreover, the growing prominence of multilingual sentiment analysis transcends language barriers and finds particular relevance in low-resource settings or languages.

Low-resource languages can be interpreted in a variety of ways, such as less researched, resource-scarce, computerized, privileged, seldom taught, or low-density [17]. Most African and Asian languages are low-resources and less explored in the NLP research [18]. Low-resource languages may be underrepresented in the global economy, as well as in social and political sectors, which might hinder economic and social progress. However, the development of technology for these languages can expand economic opportunities significantly. A language's spread and sustainability can both be aided by using NLP techniques [17].

The abundance of multilingual content online underscores the need for sentiment analysis in a variety of low-resource languages and cultures. This requirement arises from the significant influence of linguistic diversity on social analytics and social listening [19]. Drawing a connection between these observations. English remains the most extensively studied language, where much of the linguistic diversities in low-resource languages are less explored [16], [20], [21], [22], [23].

Over the years, researchers have undertaken several methods to handle issues in sentiment analysis in low-source settings, However, numerous studies have examined the need for machine learning and deep learning methods for sentiment analysis [16], [23], [24], [25], [26], [27].

Sentiment classification methods are the methods used in categorizing text into polarities. They are classified into three categories: lexicon-based, machine learning (ML), and hybrid [12], [24], [28], [29], [30], [31], [32], [33], [34], [35]. However, the lexicon-based method relies on a predefined set of patterns, often referred to as a sentiment dictionary or lexicon, where each data entry is linked with a specific sentiment orientation [17], [36]. The machine learning method leverages well-known ML algorithms to address sentiment analysis, treating it as a standard text classification issue that incorporates syntactic and linguistic attributes [21]. Hybrid techniques combine both the lexicon and the ML methods. In general, these methods can be categorized into three main groups: supervised, semi-supervised [8], and unsupervised learning [34], [36], [37], [38].

Additionally, deep learning (DL) is a subset of the machine learning methods [2], [8], based on the artificial neural network [39], [40]. The DL offer the most effective solutions for numerous challenges in image and speech recognition, as well as in the domain of the NLP [27].

A significant gap in the current literature pertains to research specifically targeting low-resource languages are deep-learning approaches. Although, several literature reviews on low-resource sentiment analysis were discovered [16], [17], [36], [41] each with distinct research emphases in machine learning methods/techniques. However, none of these studies specifically concentrate on deep learning approaches only within a low-resource context. it's essential to establish a universal set of approach criteria that can accommodate various low-resource languages.

As a result, researchers were inspired to transition their attention from the use of the ML Method to DL approaches for low-resource sentiment analysis focusing on finding an optimal approach for performing sentiment analysis tasks. Our article delves into the existing literature on sentiment analysis, specifically examining studies on low-resource languages in deep learning approaches.

The article's opening section outlines prior low-resource sentiment analysis reviews and discusses the review process in each paper. The scientific method for this review is presented in the third section. The result findings of the review are presented in the fourth section. The analysis of the findings and answers to the research questions are presented in the fifth section. The review concludes with a discussion of the key findings, and suggestions for further research. The following are the principal highlights of this research:

- To explore and report the current best approaches applied in low resource sentiment analysis with deep learning models.
- To showcase the outcomes of diverse deep learning models based on the performance metrics employed in recent studies.
- To provide future research recommendations in low-resource sentiment analysis.

Table 1 provides a list of acronyms used in this literature review paper.

II. RELATED WORK

In the study conducted by the authors, [36] focused entirely on the pre-processing techniques, and evaluation method, employed for multilingual sentiment analysis in frequently used languages. Additionally, the studies do not provide any insights into the challenges specific to low-resource settings. The studies [17] conducted a survey on past work

TABLE 1. Abbreviation list used in this review.

List of abbreviations	
1D-CNN	One-dimensional Convolutional Neural Network
1D-Conv	One-dimensional Convolution
AdopterBERT	A variant of BERT model for adoption-related tasks
araBERT	A BERT model trained for Arabic language
BiGRU	Bidirectional Gated Recurrent Unit
BiLSTM	Bidirectional Long Short-Term Memory
BERT	Bidirectional Encoder Representations from Transformers
CNN	Convolutional Neural Network
convBiLSTM	Convolutional Bidirectional Long Short-Term Memory
DL	Deep Learning
DistilBERT	Distilled version of BERT
DNN	Deep Neural Network
ELECTRA	Efficiently Learning an Encoder that Classifies Token Replacements Accurately
EEC	Equity Evaluation Corpora
FFNN	Feedforward Neural Network
GRU	Gated Recurrent Unit
HinBERT	A BERT model trained for Hindi language
LSTM	Long Short-Term Memory
MCE-CNN	Multi-Channel Ensemble Convolutional Neural Network
mBERT	Multilingual BERT
ML	Machine Learning
MT5	Multilingual Text-To-Text Transfer Transformer
MSA	Multilingual Sentiment Analysis
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
NMT	Neural Machine Translation
POS	Part-of-speech
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QAC	Quality Assessment Criteria
RNN	Recurrent Neural Network
RoBERTa	Robustly Optimized BERT Approach
RQ	Research Question
SA	Sentiment Analysis
SLR	Systematic Literature Review
SVM	Support Vector Machine
TwilBERT	A BERT model trained for Twitter data
ULMfit	Universal Language Model Fine-tuning
viBERT	A BERT model trained for Vietnamese language
XLM-R	Cross-lingual Language Model by Facebook AI

and future challenges of growth for the future in the low resource. They also recognized unanswered questions in the field. However, despite the numerous studies conducted, the specific techniques that provide optimal solutions to the challenges identified remain unclear. The author [16] conducted a comparative analysis of various techniques to determine the most effective techniques for conducting multilingual sentiment analysis across diverse languages and datasets. The study solely focused on traditional machine learning methods, overlooking the extensive research that demonstrates the superior performance of DL models in text classification. The author [24], [32], presents a comprehensive review of the latest developments in the field of Multilingual Sentiment Analysis (MSA), including the techniques employed and associated limitations. However, the review does not incorporate a comparative analysis of performance across languages and fails to identify any standout approach that excels in low-resource language settings. Another study by [41] presents a review study focusing exclusively on sentiment classification within code-mixed Indian languages. This study offers a brief overview of diverse methods and previous works in this domain but does not extensively elaborate on the distinctions and relationships among them and is limited to only Indian languages.

In contrast to the limitations observed in the related review papers conducted by different researchers, our systematic literature review (SLR) aims to address these gaps comprehensively. This study's SLR will delve into a broader spectrum of topics, including approaches, techniques, evaluation metrics, and challenges specific to low-resource settings using deep learning.

The previous reviews and reviews acknowledged the future challenges in low-resource sentiment analysis, but they fell short of providing clear insights into the optimal techniques to address these challenges. In contrast, this study aims to identify and analyze specific approaches and techniques that offer optimal solutions tailored to low-resource language settings. Moreover, while some studies have solely concentrated on traditional machine learning methods or specific language domains, This SLR will encompass a comprehensive analysis of deep learning models, across diverse languages and datasets. By doing so, This study intends to bridge the gap in the existing literature and provide a more holistic understanding of the current state and future directions of sentiment analysis in a low-resource setting.

Notably, our study is a pioneering effort to investigate and examine a specific approach with deep learning models in multilingual sentiment analysis in the context of low resources. This study aims to focus on analyzing low-resource sentiment analysis classification review approaches specific to deep learning.

III. METHODOLOGY

This section presents a set of guidelines for creating and analyzing studies on low-resource sentiment analysis. This section provides insight into the various phases of the research process when utilizing the literature review methodology. It involves identifying, assessing, and interpreting all relevant research sources about specific research questions.

A. REVIEW PROCESS

The methodology employed in this literature review adheres to the Systematic Literature Review (SLR) approach outlined by [42]. SLR is an excellent way to evaluate the range of a body of literature on a particular subject area, and it provides a clear overview of the number of literary studies that are accessible and a synthesis of their main focus [43].

B. RESEARCH PROBLEM

In light of the existing research landscape, a pertinent problem arises as there are inadequate studies that provide a comprehensive review of the most effective approaches for deep learning models in the context of low-resource sentiment analysis.

C. RESEARCH QUESTIONS

This section presents the research questions. This will help to keep the research focused on the area. The research questions and the reason to address them are as follows:

RQ1: What is the most effective approach in low-resource multilingual sentiment classification?

Addressing this RQ1 is vital for understanding the optimal approach for performing sentiment classification in low-resource settings. By exploring previous techniques and approaches utilized in this context, we can gain insights into the current state-of-the-art, as well as the advantages associated with different techniques. The findings from this investigation will help identify the most recommended techniques and features for effectively conducting sentiment classification in low-resource languages.

RQ2: What are the most common sources of data for low-resource sentiment analysis?

The purpose of question RQ2 is to determine a crucial understanding of the primary sources of data available for conducting sentiment analysis in low-resource settings. By investigating the most prevalent data sources utilized in such scenarios, we can gain insights into the types of data commonly used, their availability, and their suitability for sentiment analysis tasks.

RQ3: What are the common challenges affecting the low-resource sentiment analysis system?

this question investigation will help in identifying potential challenges and limitations associated with the available low-resource sentiment analysis system, thereby informing researchers about the scope and feasibility of conducting sentiment analysis in low-resource environments.

D. DATA SOURCES AND STRATEGY FOR SEARCHING

Selecting many literature papers plays a significant part in the general spread of information in any academic field. Therefore, The search process includes formulating the search string, doing a rigorous search, modifying the search string, and obtaining an initial list of articles from digital libraries that match the selection process. However, extensive literature was searched for pertinent papers. This is possible with the help of a suitable combination of different databases which are the most important and oldest in the research community. Repeated searches were conducted on the following databases:

TABLE 2. Search string keyword.

Search String
Sentiment analysis AND low-resource Sentiment analysis OR resource scares Sentiment analysis OR Under-resource Sentiment analysis OR Multilingual Deep learning AND Sentiment analysis classification Sentiment analysis AND low-resource OR Under-resource.

TABLE 3. Inclusion criteria.

SN	Inclusion
I1	Studies that focus on low-resource sentiment analysis
I2	The studies that use DL in low-resource sentiment classification
I3	Studies that have clear evaluation in low-resource (Under resource or Resource scares) sentiment analysis
I4	Studies based on Multilingual sentiment analysis in low-resource-based of polarity classification

- Web of Science
- Scopus
- Science Direct
- IEEE Xplore
- ACM

Several keywords were used for the search by identifying and finding phrases based on the research questions, titles, abstracts, and keywords. Particularly using the selected search Boolean in Table 2.

E. SELECTION CRITERIA FOR THE REVIEW

At this stage, the main focus revolved around the criteria for including or excluding materials. Meta-data of papers from journal papers and conference proceedings published between 2018 and 2023 were scrutinized to determine relevance.

The inclusion criteria encompassed studies meeting the inclusion criteria in Table 3 which include: I1: Studies that focus on low-resource sentiment analysis, I2: Studies that use DL only in low-resource sentiment classification, I3: Studies that have a clear evaluation in low-resource (Under resource or Resource scares) sentiment analysis, and I4: Studies based on multilingual sentiment analysis in low-resource based in polarity classification.

The exclusion criteria used to exclude articles in this study are outlined in Table 4. The following are the criteria for excluding papers deemed irrelevant: E1: Studies that do not focus on low-resource sentiment analysis. E2: The studies that use traditional ML only on multilingual low-resource sentiment classification. E3: Studies that do not have a clear evaluation in low-resource (Under resource or Resource scares) sentiment analysis. E4: Multilingual low-resource sentiment analysis in low-resource based on subjectivity classification.

TABLE 4. Exclusion criteria.

SN	Exclusion
E1	Studies that do not focus on low-resource sentiment analysis.
E2	The studies that use traditional ML only on multilingual low-resource sentiment classification.
E3	Studies articles lacking a validation or experimental result in low-resource Sentiment Analysis.
E4	Multilingual low-resource sentiment analysis in low-resource based on subjectivity classification.

F. QUALITY ASSESSMENT OF THE SELECTED ARTICLES

To ensure quality control in this SLR process, a set of quality criteria was established using a checklist that encompassed all necessary aspects and inquiries that needed to be addressed in each study. These quality criteria were usually developed as questions in the form of a checklist to determine the strength of each selected research paper, they include:

- Does the study introduce a distinctive approach in sentiment analysis for low-resource languages?
- Does the research paper address the challenges of classification tasks in low-resource language text
- Are the findings clearly stated within the paper

G. SELECTION PROCESS OF THE RESEARCH ARTICLES

The process for choosing studies that fit the requirements to answer the research questions is explained in this section. The study selection process comprises four stages: identification, screening, eligibility, and inclusion and exclusion of study criteria. During this phase of study selection, we employed the PRISMA flow diagram as a reporting guideline [44]. Figure 1 presents the process flow diagram for selecting relevant studies.

H. EXTRACTION OF INFORMATION

In the extraction phase of the related studies, we conducted a literature search across five electronic databases using predefined keywords, yielding 225 research papers. Initially, we screened the retrieved research to eliminate duplicate papers, and unrelated topics resulting in the removal of 88 papers. Subsequently, we applied inclusion and exclusion criteria to 137 articles, leading to the elimination of 62 studies. The remaining 75 articles underwent assessment using the three quality criteria, resulting in the exclusion of 19 papers. Ultimately, a total of 56 research papers (referred to as selected studies) were included in this literature review, comprising 44 papers which are 79% from journals, while 12 papers make 21% of the total from conferences. Obtaining information is necessary to address our research questions and contributions. Therefore, the generated findings were provided from the article's details discussion in the result section. Table 5 illustrates the number of selected papers in each stage.

I. ANALYSIS OF INFORMATION

The retrieved data were combined to provide the research's answers. The narrative synthesis approach was used to answer the study questions and present the findings. However, the overview of selected studies provides insight into the distribution of published literature sourced from that particular outlet. The 56 papers matched the requirement for full-text review. Figure 2 shows the number of studies conducted per year. In 2018, there were fewer studies, while in 2021-2023, there were more studies. It demonstrates the research area's activeness.

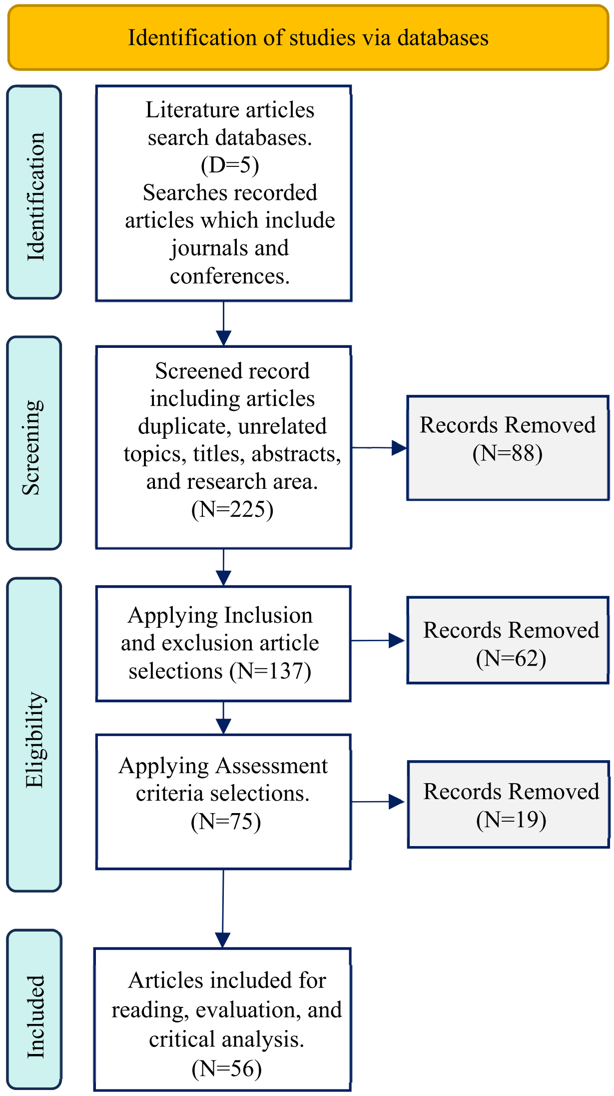

FIGURE 1. PRISMA flow diagram of the process study's selection.

IV. RESULT

This section presents the major primary study outcomes related to low-resource polarity classification text. In the subsequent section, we analyzed the chosen studies. We aim to spotlight and deliberate on the existing approach along with its characteristics, serving to inform and guide researchers in their future work. The subsection focuses on a summary of the findings of current approaches employing deep learning with low resources.

TABLE 5. The number of articles retrieved and selected study.

Search source	Number of Retrieved articles				
Data cleaning	Primary search query	Screening result	Inclusion & Exclusion Criteria		QAC
			Inc	Exc	
IEEE	29	26	18	8	12
Web Science	36	28	13	15	11
Scopus	65	37	16	21	12
ScienceDirect	31	15	8	7	5
ACM	64	31	20	11	16
Total	225	137	75	62	**56**

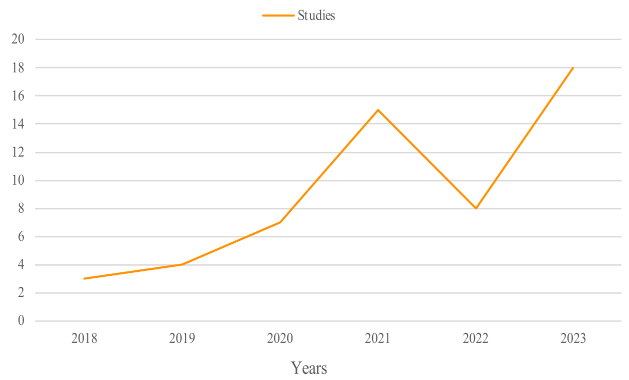

FIGURE 2. Number of studies conducted per year.

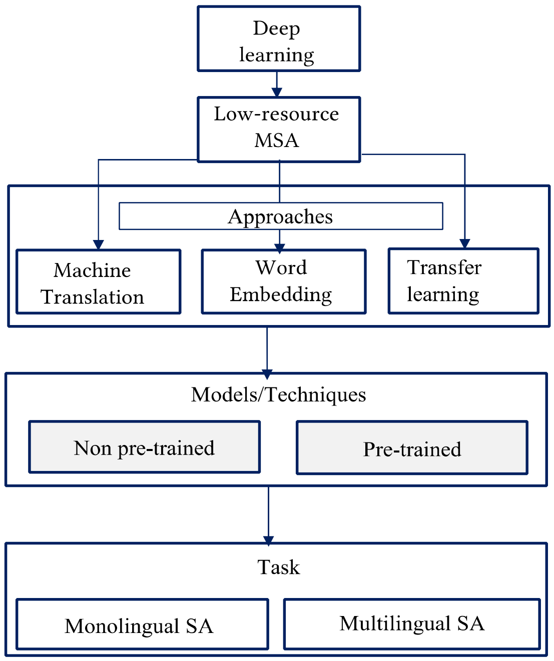

FIGURE 3. DL approaches for low-resource setting.

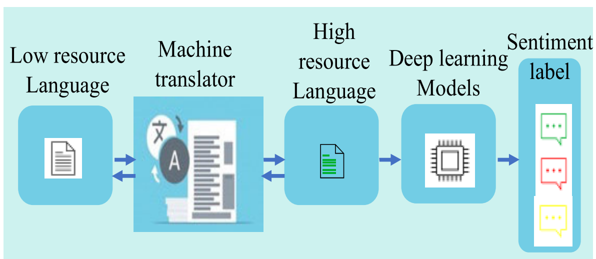

FIGURE 4. Process of machine translation for low-resource sentiment analysis.

The approaches discovered in this review are based on deep learning. Figure 3 shows the classification of approaches based on deep learning for low-resource settings. The effectiveness of each approach is discussed in detail in the following subsections.

A. DEEP LEARNING APPROACHES APPLIED FOR LOW RESOURCE SENTIMENT ANALYSIS

1) MACHINE TRANSLATION

The application of automatic translation software or translators, which are frequently employed in combination with several English knowledge bases is what is term machine translation [45]. It used the translated information it has learned by automatically interpreting the vast majority of translation instances in the corpus [25], [46]. However, for low resource languages, machine translation have been used in different research to perform sentiment analysis tasks. Figure 4 shows the use of a machine translator for the low-resource sentiment analysis process.

A. Ghafoor et al., [47] propose to find out the effect of translation from a language with abundant resources to a language with limited resources on the emotion classification task. They point out to resolve the gap that course polarity changes resulting in a decrease in sentiment performance accuracy. They used the Google machine translation API, and the English movie reviews dataset from IMDB to translate it into Urdu, Hindi, and German. Nevertheless, all models for deep learning employ random parameters that are adjusted to see how the results change. This process is repeated until the best results are obtained. Moreover, their experiment's findings indicate that English has the highest accuracy (88.37 by DNN), followed by Hindi (85.99% by Bi-LSTM), and Urdu (80.78%). To find the mistakes made by Google Translate, a subset of the data set was manually translated as part of the error analysis process in the work. Among Google-translated Urdu reviews, was found 104 terms (or groups of phrases) that were altering the mean of the review. These terms were recognized and placed into 5 groups, including confusing words and phrases, idioms, negation, sarcasm, and slang. In conclusion, based on their result they suggested that translation is not an effective method for creating a large dataset for low-resource languages.

Kumari et al. [48] argue that the MT systems frequently fail to keep the semantic meaning of sentiment information in a giving review by modulating the neuronal MT that relies on global attention (NMT) to produce translations that preserve a source phrase's non-opinionated semantic information while encoding the source phrase's underlying emotion. However, they balance both the sentiment and semantics during translation using the actor-critic. The experiment conducted in the study of [49] indicates that by selectively transliterating and translating text, the accuracy of sentiment analysis and offensive language detection can be improved, especially for the code-mixed Malayalam-English dataset. Additionally, the proposed hybrid method demonstrates that the LSTM_GRU outperformed CNN and RNN.

Sabri et al. [50] consider machine translation and BiLSTM. They use a pre-trained embedding model BERT to automatically create embedding for more than one language of tweets in a Persian-English code-mixed data set. Reference [51] highlights, the new method for bilingual sentiment analysis that uses a machine translation to compare text writing in Bengali with English translation. They used six different machine learning and deep models deployed for the task, and it achieved 58.2% performance accuracy by LSTM.

Koto and Koto, [52] experiments imply a straightforward word-for-word translation utilizing a bilingual dictionary surpasses both the transformer and LSTM models in the machine translation experiment in terms of BLEU score. Moreover, the experiment was done after manually translating an Indonesian sentiment analysis corpus into Minangkabau. Notably, one of the experiments was tested against the Minangkabau dataset and 75.91% F1-score performance with mBERT. A study by [53] explores different deep learning approaches including CNN, LSTM, FFNN, BiLSTM, and transformers for sentiment classification, employing word or sentence embedding techniques utilising Google translator. The FFNN classifier, integrating sentence transformers and the cosine similarity method, emerged as the optimal choice for both 3-class and 2-class sentiment classification tasks, demonstrating accuracies of 62.0% and 82.2%, respectively.

Certainly, machine translation still encounters numerous challenges [16], [25], [36], mostly especially in finishing assignments due to the variety and complexity of languages in different places, especially when dealing with languages with limited resources [46]. Moreover, the author [54] suggested that the use of machine translation dependent on the accuracy of the translations and the availability of comprehensive language models in the target language.

2) WORD-EMBEDDING

Word embedding is a task-specific DNN architecture that depends on word embedding vectors, which are frequently used to refine them in a supervised classification task with labelled data [55]. Thus, they are unique feature vectors in a distributional space that are influenced by all other words [56]. Furthermore, it is important for words with equivalent meanings in different languages to have similar or identical vector representations [25], [57]. Notably, word embedding has been the subject of several research recently [58], [59]. The technique's most striking feature is its exceptional ability to capture numerous lexical associations beyond simple similarities [60]. Despite this importance, there is no consensus on the appropriate degree of embedding to utilise or on how to encode a multilingual or language-independent feature space in a single model for use in the downstrea tasks [23]. Figure 5 shows the word-embedding process for low-resource sentiment analysis.

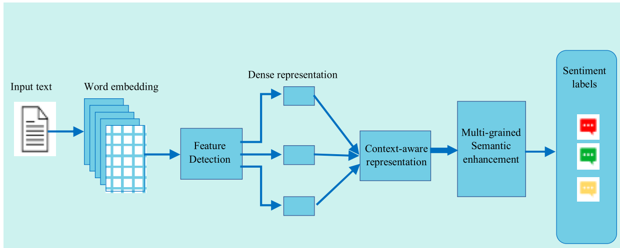

FIGURE 5. Word embedding sentiment analysis [61].

Singh et al., [62] present a dataset benchmark to identify polarity on social media datasets and find offensive sentiment phrases in the Nepali language. Moreover, the experiment was done using genism [63] and trained 300-dimension skip-gram fast-text word embeddings on both mono and multilingual datasets. However, the classification task achieved a good accuracy of 0.816% by BiLSTM models. In the study [64], the researchers conducted a comparison of the modelling outcomes achieved by deep learning models against those of other frequently utilized algorithms. Their experimental findings indicate that deep learning models surpass the baseline performance and exhibit higher accuracy, particularly evident in the case of CNNs. Reference [65] initiated research in sentiment analysis on the Bambara-French low-resources language. The author utilized convolutional neural networks (CNNs) to develop two models and four models based on long short-term memory (LSTM) and other machine learning models to experiment with a code-mixed Bambara-French social network dataset. Moreover, they use character and word representations using fixed indices to construct sentence and comment representations. One of the experiment comparisons on the models has a better accuracy of 83.3% with one-layer CNN against the other models.

Nguyen and Le Nguyen., [66] Explored N-gram BiLSTM deep learning model with word embedding approaches that represent semantic and contextual information to be consistent in predicting polarity in the YouTube dataset for the Italian language. They argued that a word may function or imply different things depending on the situation in a sentence or text. Therefore, each word is given a distinct vector by the word embedding approaches and this renders a distortion of both the word's meaning and its context. Their experiment achieved better performance accuracy. Research by [58] was carried out on an attention-based LSTM network to investigate the limitation of heavy task-dependent word embedding approaches by proposing a lexicon embedding. However, the experiment conducted on the four models verified that their customer review superposes the others with 94.10% accuracy. Also, the author [67] uses the same Hierarchical attention based on LSTM, and transformers BERT and dynamic routing-based console with BiLSTM network. They experiment with the three methods to find the best model that performed well with word2vec embedding in three-dimensional space. Interestingly, among the 14 DL models used for the study, BERT-LSTM has the highest accuracy on the cricket dataset.

Another method by [68] used a bidirectional-gated recurrent unit DL model to experiment on the word2vec Arabic version of SemEval task 5. However, The model is made up of several elements, including an embedding layer to create the vector representations of each word, and a BiGRU layer to remember long-term dependencies and extract text information simultaneously from the left and right sides. Nevertheless, the result indicated that the BiGRU transformer model performs better when compared to GRU and LSTM. Similarly, [69] discovered the problem of data sparsity which often poses a significant challenge to neural network learning most especially when dealing with languages in a low-resource setting. A DL-based approach that employs word embedding to address the challenges was suggested by the researchers. The method involves using word embeddings in combination with an LSTM model for prediction. Moreover, they experimented with two ways aspect term extraction and aspect classification. However, the result shows that it achieved an accuracy of 76.29% with multilingual embedding aspect classification.

Berrimi et al., [70] attempt to improve the performance accuracy of baseline models already developed in certain datasets for the Arabic low-resource language, They also took a step to evaluate the approach in other NLP tasks to demonstrate their approach generalization capabilities. In addition, the dataset LABR [71], HARD [72], and BRAD [73] received accuracy scores of 98.6%, 96.19%, and 95.65%, respectively. Nevertheless, an investigation was done through the use of FastText and local learnable word embeddings. However, their result shows that BiGRU typically performs better than BiLSTM, from the experiment.

Similarly, different low-resource languages suffer from the problem of class imbalance including the Bengali language. Reference [74] tries to resolve the problem by applying word embedding strategies as one of their methods in the sentiment classification task. Three DL models were used for the experiment and evaluation. However, they did not hold for the sentiment analysis dataset Because there were few words in each text [75]. Explored the word embedding approaches from FastText and Word2Vec and demonstrated how to create a sentiment analysis system for Sinhala news comments. In addition, the experiment was based on various features, machines, deep learning techniques, and settings for text cleaning and pre-processing. To select optimal classification algorithms two separate tasks were done in the deep learning models word embedding achieved 87% accuracy with binary class and 67% multiclassification by LSTM. The same language was investigated by [76] using LSTM networks with sentence states which are created using graphs of message-passing and word embeddings in fastText. It also achieves an accuracy of 87.86%.

Dahou et al., [77] experimented with the Arabic tweet dataset by inserting a variety of pre-trained word embeddings into each embedding channel's embedding block and training these channels concurrently for sentiment classification. The research introduces a novel approach to improve sentiment classification in the Arabic language by presenting a multi-channel embedding convolutional neural network (MCE-CNN). The proposed approach aims to improve Arabic sentiment classification by utilizing a multi-channel embedding convolutional neural network (MCE-CNN) that learns sentiment features from different text domains, as well as word and character n-gram levels. To enhance the performance of the model, several important procedures are carried out, including using high-quality Word Embedding Vectors (WEVs) as an initial phase for training. This approach has demonstrated promising results in improving the accuracy of Arabic sentiment classification.

Another study by [78] took a different corpus creation. They introduce a new dataset for the identification of threatening language in Urdu tweets to further the study of multilingual low-resource language sentiment analysis. However, manual annotation of data was done by an expert in binary format and evaluated using different ML and DL models with the use of pre-trained word embeddings for Urdu in fast Text [79]. The study of [80], introduces a model design, aimed at acquiring Semantic-Aspect Word Embeddings (SAWE) to discern semantics within intricate subunits of a word. The primary focus of this model is the classification of sentiment in code-mixed Telugu-English review corpora. Leveraging word embedding and SAWE as inputs, a unified deep neural network is employed, featuring a two-level architecture comprising Bidirectional Long Short-Term Memory (BiLSTM) and Gated Recurrent Unit (GRU). Remarkably, the model achieves an accuracy of 81.88%, facilitated by the implementation of the Skipgram method. Reference [81] presents a straightforward data augmentation technique applicable to existing word embedding algorithms, this method enhances cross-lingual contextual comprehension without the necessity of a parallel corpus. Experimental findings indicate that implementing this data augmentation strategy during the training of word embedding models, such as Word2Vec and FastText, proves beneficial by enabling the model to better capture cross-lingual relationships within code-mixed sentences. Another work by [82] utilizes multilingual word embeddings from FastText, comparing them with deep learning models such as CNN, LSTM, and BiLSTM. The CNN model outperformed others, achieving the highest accuracy in their evaluation. As noted by [83], hybrid approaches hold promise in mitigating sentiment errors on complex training data. The study assesses the

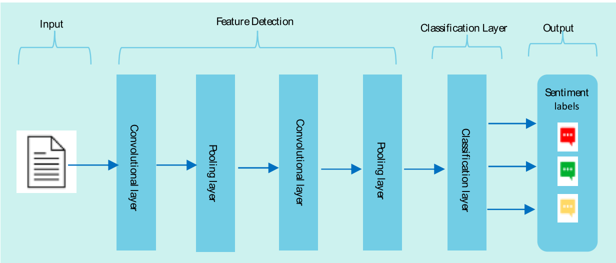

FIGURE 6. Transfer learning sentiment analysis [87].

effectiveness of hybrid deep learning models in specific Vietnamese language domains by combining CNN, LSTM, and SVM models. The achieved accuracy is 93.94%, surpassing standalone models, particularly evident with the USALUTH dataset in the case of CNN-LSTM.

Kanclerz et al., [84] experimented with polarity classification techniques, they used two alternative neural network architectures BLSTM and CNN. The authors utilized language-agnostic embeddings and evaluated their cross-lingual approach using the PolEmo 1.0 [85] sentiment corpus test datasets translated into various languages. The study found that the BiLSTM network model, trained on language-neutral phrase embeddings, outperformed other models when evaluating single-domain datasets.

Word embedding suffers from a lot of time consumption, computer power, and large corpora of datasets to trained [57]. Again the two methods continuous bag of word (CBOW) and skip-gram word embedding method cannot handle the polysemy disambiguation [68]. However, some researchers have worked a lot to overcome this limitation by using Transfer learning.

3) TRANSFER LEARNING

Transfer learning is an approach for applying what has been learned in one area domain or language to another by using similarities in data, data distribution, models, tasks, and other factors [86]. This enables the identification of differences between source and target languages [25]. However, the transfer learning approach uses transformer models to facilitate the transfer of bidirectional contextualized representations of text, utilizing information from both the preceding and following words in all layers of the deep learning model [100]. Figure 6 presents the transfer learning process for multilingual sentiment analysis.

In an approach of transfer learning [88] explored two tasks and two benchmark corpora with large sizes in Vietnamese aspect polarity classification and aspect category detection. They use deep learning architectures to deploy supervised learning techniques for different task classifications and compare the efficacy of their manual annotation. The result achieved 74.88% f1-score with the BERT model in the multi-tasks approach in aspect polarity classification. A study by [89] explored the use of four Telugu pre-trained word embeddings to solve the problem of Telugu being a low-resource language. One of the classifiers that were used for this task is LTSM. However, they contend that their pre-trained embeddings are on par with or perhaps superior to the current multilingual pre-trained models (Transformers).

Reference [90] uses an ensemble technique in the Hindi language to create an auxiliary sentence, which transforms the ABSA issue into a task requiring sentence pairs to be classified for both mBERT-E-MV and mBERT-E-AS. However, their approach achieved a better performance result by adjusting several pre-trained BERT models and combining them to make forecasts using the suggested model. In a study by [91], a method was proposed using deep learning models, to perform sentiment analysis and detect offensive language in low-resource code-mixed data in Tamil and English. However, the results showed that the sentiment analysis task achieved only 65% accuracy on Adapter-BERT models when performed independently. The classification was carried out using various pre-trained models, including BERT, Roberta, Bi-LSTM, and Adapter-Bert. Another study by [92] identified a limitation in existing deep-learning models, noting their failure to leverage implicit language information in code-mixed text. In response, they proposed techniques involving word-level interleaving and post-sentence placement of language information to enhance the performance of BERT-based models on low-resource Code-Mixed Hindi-English datasets. The study involved a comprehensive examination of vanilla BERT-based models and their code-mixed HingBERT counterparts on corresponding benchmark datasets. The comparison included assessments of model performance with and without the incorporation of word-level language information. The findings indicate the effectiveness of the proposed language augmentation approaches across various BERT models.

Transfer learning has advantages over starting sentiment analysis from scratch, or initiating training a model which consumes a lot of time [93]. In another study [94] carried out a performance comparison of deep learning models using GRU, LSTM, and CNN with Arabic books and hotel reviews in multilingual dialects. However, their findings demonstrated that for binary and multiple-label classification, deep learning outperformed machine learning.

B. DEEP LEARNING TECHNIQUES APPLIED FOR LOW RESOURCE SENTIMENT ANALYSIS

In the subsequent section, we observed the techniques applied with one of the above approaches for sentiment analysis tasks with low resources. Based on our observation they are categorized into two pre-trained and non-pretrained. Figure 7 illustrates the taxonomy of the techniques employed.

1) NON-PRETRAINED TECHNIQUES

In a study [95] Explores all three method-based approaches of sentiment analysis for classifying Hindi tweet text. He uses machine learning, lexicon and deep learning based and proposed a domain-specific sentiment dictionary. Additionally,

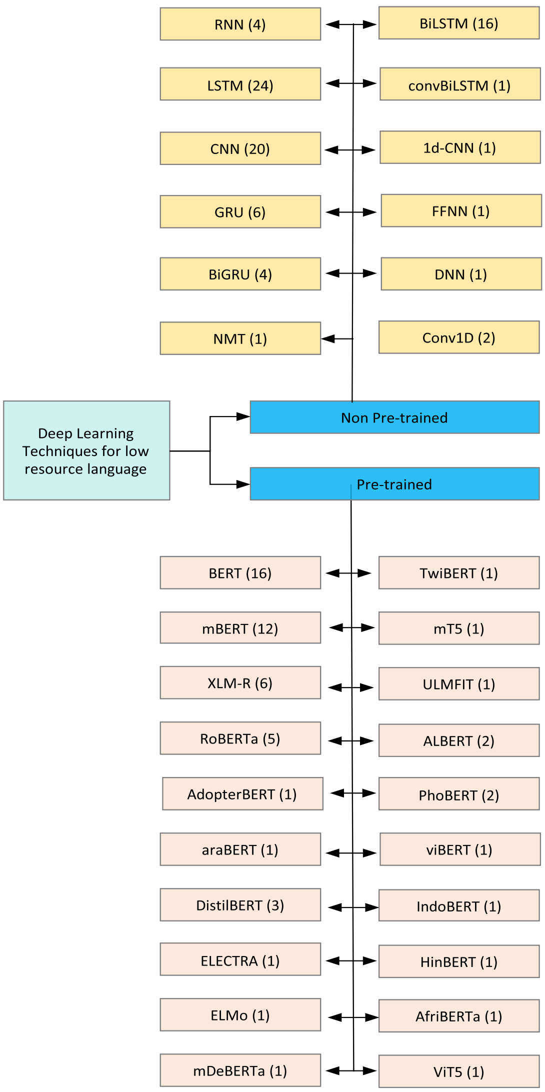

FIGURE 7. Taxonomy of the applied techniques with their number of studies in bracket.

the author proposes an integrated CNN that combines an artificial RNN and LSTM to analyse Hindi sentiment. Nevertheless, their research's goal is to identify any scalable, reliable technique that can be used with any low-resource language. However, the integrated RNN-LSTM deep learning proposed by the author outperforms the lexicon and machine learning based. Furthermore, the study points out that only tweets written in Hindi were being considered even though there are tweets in a code-mixed text which are open for further research.

Altaf et al., [35] highlighted that the majority of sentiment analysis research in the Urdu language relies on certain domains, and models are frequently tested and trained on the same dataset. for a small number of domains, to leverage the problem they develop a method model that adapts a cross-domain sentiment analysis in Urdu language. The method achieved 77% accuracy by GRU deep learning model, and it outperformed the traditional ML, RNN and LSTM models. [96] also considers the morphological feature of text written in the Punjabi language to carry out the sentiment analysis regarding the farmers' suicide in India. However, our main concern is the techniques for sentiment analysis in low-resource. Notably, the author uses a deep recurrent Neural network to solve the issue even though the data is a relatively small amount for deep-learning models.

The author [97] uses the same neural network LSTM and GRU models and achieves 87.31% accuracy. They used Bi-Indy LSTM to classify sentiment in the Arabic language with the semEval2016 [98] dataset. However, they combined the models to eliminate the recurrent neural network's usual flaws. In another study [64] was inspired to contrast machine learning with deep learning models for emotion and opinion on Algerian text. In the same way, they trace the advances in developing integrated embeddings in DL models. However, CNN with embedded embeddings produced the best results in the experiment when compared with the results obtained by traditional machine learning.

2) PRE-TRAINED (TRANSFORMERS)

The Pre-trained models or Transformers rely on transfer learning techniques, allowing researchers to focus on fine-tuning the models instead of starting the training process from scratch [25].

González et al., [99] introduced a new approach that utilizes a reply order prediction signal to improve inter-sentence coherence in Twitter conversations. Their method enhances the performance of TWilBERT, particularly in text classification tasks that require reasoning on sequences of tweets. Among the several evaluation experiments, the sentiment analysis task model performs better than the top-performing system in the competition in the TASS task version for Uruguay. The comparison study of [94] consists of a transformer model in which the result of the binary and the multi-classification on the HARD [72] dataset performed well with an accuracy of 92.7%. Similarly, [100] executed Aspect based sentiment analysis on the reviews dataset for Indonesia, They used one of the pre-trained language representation models multilingual BERT. Their experiment significantly increases performance sentiment classification with the task transformation approach.

Li et al., [101] proposed an efficient strategy that fine-tunes a pre-trained language model to perform sentiment analysis and text classification called AgglutiFiT. They fine-tuned the model using the low-noise fine-tuning dataset created by morphological analysis and stem extraction. They argued that the method is better for choosing relevant semantic and syntactic data information in Kazakh, Kyrgyz, and Uyghur low-resource languages. The Kazakh sentiment task achieved about 92.87% accuracy. Furthermore, [102] looks into a transferable BERT (TransBERT) training system that transfers knowledge from many semantic relevant supervised tasks, and also general language knowledge from large-scale unlabelled data, for a target task. The three specific tasks include sentiment analysis on IMDB and Twitter. The model is evaluated by comparing it with several baseline techniques, to assess the performance of our model accuracy.

Another work by [103] captured user sentiment from code-mixed texts, along with the low predictive nature of traditional ML learning models compared to DL counterparts, models LSTM, CNN and BiLSTM. By using code-mixed data from Hindi-English and Bengali-English, The experiment shows that the attention-based models outperformed conventional methods in terms of accuracy by 20–60%. Similarly, [104] study demonstrates the efficiency of BERT in handling the Turkish language, which has a complex morphological structure. The author experimented to compare the performance of BERT with conventional machine learning (ML) algorithms as baseline models for various natural language processing (NLP) tasks, including sentiment analysis. The results indicated that BERT outperformed the base ML algorithms, while also reducing the need for time-consuming pre-processing tasks. The findings suggest that BERT has the potential to enhance the performance of NLP tasks with minimal pre-processing efforts.

A recent research conducted by [105], introduced NaijaSenti, which is the first large-scale human-annotated dataset for sentiment analysis in low-resource languages like Igbo, Yoruba, Hausa, and Nigerian-Pidgin. The study evaluated several pre-trained models such as mBERT, XML-R, mDeBERTaV3, mDeBERTaV3, and AfriBERT to assess their effectiveness and fine-tuned their dataset using language-adaptive methods. Furthermore, the study addressed one of the major challenges, which is the under-representation of African languages in sentiment analysis.

In response to the complex nature of the Tamil language, the author [106] introduces a pioneering Multi-Stage Deep Learning Architecture (MSDLA) tailored for cross-lingual sentiment analysis in Tamil, characterized as a low-resource language. The methodology involves leveraging transfer learning from a linguistically rich source language to mitigate data constraints. The results demonstrate the superiority of the proposed model over existing techniques, particularly evident in its exceptional performance on the Tamil Movie Review dataset, where it attains an accuracy of 0.8772. Another method, introduced by a different researcher [107], is the POADL-ABSA technique, which concentrates on sentiment detection and classification. This approach involves multiple levels of operations and integrates the BiLSTM and BERT models for sentiment classification. The authors optimized the hyperparameters of the ABiLSTM model to achieve improved results in sentiment classification. Through extensive experimentation, their method, IAOADL-ABSA, demonstrated superior performance, surpassing other models with an accuracy of 98.72%.

The study by [108] employs deep learning, specifically the BERT transformer model, to conduct sentiment analysis on public opinions regarding the ongoing Russia-Ukraine war. Focusing on economic implications, non-English posts are translated to English using neural machine translation and Google Translator. The research concludes with a comparative analysis correlating global sentiments with countries' dependence on Russian oil, achieving an accuracy of 82%. Reference [109], found that sentiment analysis on code-mixed data faces challenges due to noise and this issue makes the conventional monolingual models less effective. They used mBERT, a multilingual pre-trained model, on English-Indonesian code-mixed data, achieving the highest accuracy of 76%.

There are many pre-trained language models available for Vietnamese, encompassing both monolingual and multilingual variants. However, these models differ in architecture, resulting in varied effectiveness during fine-tuning. the authors [110] find a lack of studies assessing these models on the same sentiment analysis datasets. They address this gap by introducing a fine-tuning approach to evaluate the performance of diverse pre-trained language models for Vietnamese sentiment analysis. Experimental findings highlight the superior performance of the monolingual PhoBERT and ViT5 models, setting new benchmarks on five Vietnamese sentiment analysis datasets. A similar study by [111] explores the same language and addresses a sentiment analysis challenge within the educational domain of a low-resource language, specifically examining social sentiments towards the University of Phan Thiet. The primary emphasis is on constructing a sentiment corpus for the university and investigating deep learning models including LSTM, BERT, DistilBERT, and PhoBERT. The experimental outcomes reveal that PhoBERT yields the most favourable results, achieving an F1-score of 89.68%.

In the investigation conducted by the author [112], transformer fine-tuning methods are examined for sentiment classification tasks in the Hausa language. Three pre-trained multilingual language models, namely Roberta, XLM-R, and mBERT, are utilized. The results reveal that the mBERT-base-cased model attains the highest accuracy and F1-score, both registering at 0.73%. Similar work was done by [113] on code-mixed tweets in Indonesian-Sundanese, employing the pre-trained language model, IndoBERT. The evaluation of their approach yielded a notable accuracy rate of 81%. A work conducted by [114] proposes an ensemble model to predict sentiment in code-mixed texts of Kannada and Malayalam languages. The ensemble, comprising transformer-based models, demonstrated a noteworthy weighted F1-score of 0.66 for Kannada code-mixed language. Conversely, the ensemble model within the deep learning framework excelled, achieving the highest performance with a weighted F1 score of 0.72 for the Malayalam dataset, surpassing the achievements of previous research.

The research work of [115] concentrates on sentiment analysis of code-mixed Malaysian COVID-19-related news shared on Twitter. The investigators collected a multilingual

COVID-19 Twitter dataset that includes tweets written in Malay, English, and Chinese language. Employing a data compression technique utilizing Byte-Pair Encoding (BPE), they applied two deep learning approaches CNN and mBERT models. Their findings indicate that BPE-M-BERT exhibits a slight performance edge over the CNN model, underscoring the advantageous suitability of the pre-trained M-BERT network for multilingual datasets. The investigation presented in [116] centres on the low-resource setting of Hindi-English code-mixed language, with an emphasis on improving the efficacy of various code-mixed natural language processing (NLP) tasks, including sentiment analysis. The study involves a comparative analysis of diverse Transformer Models pre-trained through unsupervised methods. The experimental outcomes reveal state-of-the-art results on relevant datasets when utilizing models based on HingBERT. These models are distinctly pre-trained on authentic code-mixed text, showcasing significant enhancements in performance on code-mixed data.

Another research by [117], delves into a deep learning-based methodology for sentiment analysis in Urdu, utilizing BERT transformers. The study introduces a novel Urdu Sentiment Analysis Dataset-23 tailored for computational linguists. Within this framework, the authors introduce USA-BERT for Urdu reviews, achieved through the fine-tuning of BERT. The paper concludes by applying the Pareto principle to evaluate USA-BERT's performance on two datasets: the state-of-the-art UCSA-21 and UDSA-23. The assessment outcomes indicate that USA-BERT outperforms existing methods, demonstrating a substantial improvement in accuracy. Table 5 Provides the Summary of the approaches, Techniques, data source and best performance metric for low resource sentiment analysis.

C. LOW RESOURCE SENTIMENT PREPROCESSING

Preprocessing are initial steps taken before feeding the data into DL models for inference learning and classification. It plays a crucial role in refining text data, particularly in platforms like blogs, Twitter, and online chats where spelling errors, slang, and multilingual terms are common [36]. However, given the complex nature of some low-resource language, diverse preprocessing measures are essential to mitigate noise, typographical discrepancies, and morphological complexity. Among the 56 primary studies examined, 10 did not document any preprocessing procedures conducted. Researchers in low-resource sentiment analysis have adopted one or more varied preprocessing stages to refine the dataset, encompassing:

1) DATA CLEANING

Data cleaning involves cleansing the text through the elimination of special characters, stop words, hashtags, diacritics, URLs, emails, user handles (@user), emoticons, emojis, punctuation, symbols, elongated letters, repeated characters, spam, sarcasm, and duplicated content.

2) STEMMING

Stemming refers to the method of reducing a word to its base form by removing affixes such as prefixes and suffixes, or by isolating the root of the word.

3) LEMMATIZATION

In contrast to Stemming, Lemmatization accurately transforms inflected words to their base form, ensuring that the resulting root word remains within the confines of the language [118]. E.g. lemmatization would map "running" to "run," while stemming might simply reduce it to "runn".

4) POS (PART-OF-SPEECH)

POS tagging is the process of labelling each word in a sentence with its corresponding part of speech, such as noun, verb, adjective etc. [36]. This tagging is crucial in sentiment analysis for low-resource languages as it helps identify the grammatical structure of sentences, enabling a more accurate interpretation of sentiment based on the context provided by different parts of speech. POS tagging enhances the success of NLP tasks [86].

5) TOKENIZATION

Tokenization is the process of breaking down a text into smaller units, typically words or sub-words, known as tokens [36]. This step is fundamental in natural language processing tasks as it forms the basis for further analysis, enabling the computer to understand and process the text effectively.

6) NORMALIZATION

Normalization, in the context of text processing, refers to the process of standardizing or transforming text data to make it more consistent and comparable across different documents or datasets [119], [120]. Normalization techniques aim to reduce variations in text data that may arise due to factors such as capitalization, long words, punctuation, or different word forms.

7) NAME ENTITY RECOGNITION (NER)

NER, or Named Entity Recognition, involves identifying and categorizing named entities within a text into predefined categories such as names of persons, organizations, locations, dates, and numerical expressions. The goal of NER is to extract meaningful information from text data.

D. LOW RESOURCE SENTIMENT FEATURE EXTRACTION

Feature extraction is a fundamental step in low-resource sentiment analysis. Feature extraction allows us to create more precise data, leading to favourable outcomes from the model [121], [122]. There are many key features employed by researchers in low resource sentiment analysis, including Partof- speech (POS) tagging, Term Frequency-Inverse Document Frequency (TF-IDF), One-hot encoding, embeddings Skipgram, Continues bag of words (CBOW) and Tokenizer features. Analysis of the literature reveals that the majority of studies utilize multiple feature types to address low-resource sentiment analysis tasks.

1) N-GRAMS
N-grams are the continuous sequences of n items. Unigrams, bigrams, and trigrams are the names given to the n-grams of sizes [16]. It is a collection of words or letters that frequently appear together within a given text [36]. It is mostly used to capture local patterns and dependencies within the data.

2) TERM-FREQUENCY INVERSE DOCUMENT FREQUENCY
TF-IDF (Term Frequency-Inverse Document Frequency) is a numerical statistic that evaluates the importance of a word in a document relative to a collection of documents [118]. It combines two components: term frequency (TF), which measures how frequently a term appears in a document, and inverse document frequency (IDF), which measures how rare a term is across all documents in the corpus.

3) ONE-HOT ENCODING
One-hot encoding is a method that converts categorical data into binary vectors, with each category represented by a single bit set to 1 and the rest set to 0.

4) CONTINUOUS BAG OF WORD (CBOW)
Continuous Bag of Words (CBOW) predicts a target word based on the context of surrounding words [123], making it useful for tasks like text classification and language modelling.

5) SKIP GRAM
Skip-gram is a neural network architecture used in natural language processing for word embedding. It predicts the context words based on a central target word, making it useful for tasks like word similarity detection.

6) EMBEDDINGS
Embeddings are dense, low-dimensional representations of words in a continuous vector space [57], obtained through unsupervised learning techniques from large text corpora. These embeddings capture semantic relationships between words, allowing similar words to have similar vector representations. They include FastText [124], Word2Vec [125] and GloVe [126].

E. LOW RESOURCE SENTIMENT EVALUATION METRIC
Evaluation metrics constitute a subset of performance assessment tools utilized to measure the efficiency of machine learning models. A system demonstrating superior predictive capabilities is assigned a higher metric score, guiding the tuning module to optimize system parameters for maximum performance.

Various metrics cater to different tasks such as classification, ranking, regression, topic modelling, clustering, confusion matrix, and logarithmic loss. However, within the scope of this study, particular emphasis is placed on metrics associated with, natural language processing. These metrics include precision, recall, F1-Score (or F-measure), and Accuracy. Metrics typically originate from the confusion matrix. It is a construct that encompasses key elements such as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) values. The metrics utilized by most of the studies in this review are:

1) ACCURACY METRIC
Evaluates how accurately pre-classified text aligns with their actual classifications. The accuracy formula is given as in (1):

$$Accuracy = (TP + TN)/(TP + FP + FN + TN) \tag{1}$$

2) RECALL METRIC
Recall, essentially, assesses the outcome by determining how many of the genuinely relevant results are retrieved. In other words, it is the count of accurately predicted positive labels out of the total positive labels. The recall formula is given as in (2):

$$Recall = TP/(TP + FN) \tag{2}$$

3) PRECISION METRIC
Precision evaluates the proportion of predicted data or samples assigned to a specific class, ensuring that the predicted samples genuinely pertain to that class. The formula is given as in equation (3):

$$Precision = TP/(TP + FP) \tag{3}$$

4) F1-SCORE METRIC
The F-score, or F-measure, represents the interplay between precision and recall, and it is inversely proportional to strike a balance between the two. It effectively mitigates the imbalance between precision and recall in the evaluation of a model's performance. The F-score, incorporating both precision and recall, provides a more nuanced and balanced evaluation. The formula is given as in equation (4):

$$F1 - score = (2 \times Precision \times Recall)/(Precision + Recall) \tag{4}$$

Table 6 provides a summary of relevant literature reviewed in this study, The models utilized, the language and reported best performance. The metrics evaluation is based on the accuracy and F1 score reported in the examined papers. Selecting the highest scores and best techniques are highlighted in bold text.

V. ANALYSIS
In this section, we conduct an analysis and assessment of the reviewed studies to address the research question posed in the Systematic Literature Review (SLR). Furthermore, we offer the study findings by proposing a standardized framework for sentiment analysis in low-resource settings.

TABLE 6. List of low-resource sentiment analysis approach, models, languages domain, and best Accuracy evaluation metric.

Ref	Approach	Model/Techniques	language	Data Domain/source	Reported best performance Metric (accuracy, F1-score)	Citation (Google Scholar)
[95]	N/A	RNN, LSTM and CNN	Hindi	Social media	**RNN-LSTM** Acc:85.01%	**28**
[35]	Word embedding	GRU, RNN and LSTM	Urdu	Sport	**GRU** Acc:77%	**6**
[96]	N/A	RNN	Punjabi	News	**RNN** Acc:90.29%	**33**
[97]	N/A	CNN and LSTM	Arabic	Hotel	**LSTM** Acc:87.31%	**50**
[64]	Word embedding	CNN, and LSTM	Algeria	Social media	**CNN** Acc:0.79%	**14**
[47]	MT	DNN, LSTM, Bi-LSTM and Conv1D	Urdu, German, and Hindi	Movies	**Bi-LSTM** Acc:85.99%	**28**
[48]	MT	NMT	Hindi	Tourism	**NMT** F1-score:91.99%	**5**
[49]	MT and Word embedding	CNN, LSTM, GRU, BiLSTM, BiGRU, XLM-R	Malayalam	Social Media	**GRU** Acc:76%	**18**
[50]	MT	BiLSTM	Persian	Social Media	**BiLSTM** Acc:66.17%	**18**
[51]	MT	LSTM	Bengali	Sport and Movies	**LSTM** F1-score: 58.2%	**38**
[52]	Word embedding and Transfer learning	mBERT, LSTM	Minangkabau	Social media	**mBERT** F1-score: 75.91%	**10**
[62]	Word embedding	BERT, BiLSTM, CNN	Nepali	Social Media	**BiLSTM** Acc:0.82%	**8**
[65]	Word-embedding	LSTM, BiLSTM, CNN	Banfara-French	Social media	**CNN** Acc:83.23%	**35**
[66]	Word embedding	BiLSTM, conBiLSTM	Italian	Social media	**conBiLSTM** Acc:62.92%	**48**
[58]	Word embedding	LSTM	Korean	Wikipedia, Product, News	**LSTM** Acc:94.10%	**131**
[68]	Word embedding and Transfer learning	GRU, BiGRU, LSTM, BiLSTM	Arabic	Hotels	**BiGRU** F1-score: 65.5%	**10**
[69]	Word embedding	LSTM	Hindi	Product	**LSTM** Acc:76.29%	**16**
[67]	Word embedding	LSTM,CNN,RNN,GRU, BERT	Bangla	Social media, Sport, Movies, Product, Restaurant, News	**BERT-LSTM** Acc:84.18%	**44**
[78]	Word embedding	1D-CNN and LSTM	Urdu	Social media	**LSTM** Acc:79.40%	**44**
[70]	Word embedding	BiLSTM and BiGRU	Arabic	Books and Hotels Reviews Sport, Politics	**BiGRU** Acc:95.90%	**7**
[89]	Transfer learning	LSTM,ELMo-Te, BERT-Te, RoBERTa-Te, ALBERT-Te, and DistilBERT-Te	Telugu	Movies, Politic and Sport	**BERT-Te** F1-score 68.72%	**14**
[74]	Word embedding	LSTM,GRU, CNN	Bengali	Social media	**CNN** F1-score 0.90%	**4**
[75]	Word embedding	SVM, LR and LSTM	Sinhala	News	**LSTM** Acc:0.67%	**12**
[76]	Word embedding	LSTM	Sinhala	News	**LSTM** Acc:87.86%.	**12**
[77]	Word embedding	CNN	Arabic	Social media and Product review	**MCE-CNN** Acc: 85.51 %	**22**

TABLE 6. *(Continued.)* List of low-resource sentiment analysis approach, models, languages domain, and best Accuracy evaluation metric.

[91]	Transfer learning	CNN,LSTM, RoBERTa. BERT, AdopterBERT	Tamil	Social media	**Adapter-Bert.** Acc:65%	**19**
[84]	Word embedding	CNN, BiLSTM	Polish	Product, Education, Hotels, Medicine	**BiLSTM** F1-score 79.91%	**40**
[88]	Transfer learning	BiLSTM, CNN, BERT	Vietnamese	Hotels, Restaurant	**BERT** F1-score 74.88%	**20**
[94]	Transfer learning	LSTM, GRU,CNN,BERT, AraBERT	Arabic	Books and Hotels Review	**araBERT** Acc:92.7%	**22**
[90]	Transfer learning	mBERT BERT	Hindi	Product, Travel (Tourism),Movies	**mBERT**-E-MV Acc:79.77%	**17**
[100]	Transfer learning	mBERT. CNN	Indonesia	Product	**mBERT** F1-score:0.9765	**17**
[99]	Transfer learning	TwilBERT, mBERT	Spanish (Uruguay)	Social media	**twilBERT** Acc: 56.21%	**54**
[101]	Transfer learning	XLM-R	Kazakh, Kyrgyz, and Uyghur	Websites	**XLM-R (AgglutiFiT)** Acc:92.89%	**11**
[102]	Transfer learning	BERT	Chinese	Movie and Social media	**BERT** Acc:90%	**3**
[103]	Transfer learning	BiLSTM, CNN, BERT	Hindi and Bengali	Social media	**BERT** Acc: 20–60%	**23**
[127]	Word embedding	CNN and BiLSTM	Hindi	Social media	**CMSA-(BiLSTM)** Acc:83.54%	**63**
[104]	Transfer learning	BERT	Turkish	Hotel and Movies	**BERT** Acc:0.99% and 0.97%	**32**
[105]	Transfer learning	mBERT, XML-R, mDeBERTaV3, and AfriBERTa	Igbo, Yoruba, Hausa, and Nig-Pidgin	Social Media	AfriBERTa-large F1-score: 78.3	**55**
[107]	Transfer learning	BERT and BiLSTM	Hindi and Arabic	Social media and Websites	**BERT(IAOADL-ABSA)** Acc:98.72%	**1**
[80]	Word embedding	BiLSTM and BiGRU	Telugu-English	Social media	**BiLSTM** Acc:81.88%	**0**
[106]	Transfer learning	mT5, XLM, mBERT, ULMFiT, BiLSTM, LSTM,MSDLA and ALBERT	Tamil	Movies	XLM **(MSDLA)** Acc:0.8772%	0
[81]	Word embedding Transfer learning	BERT,1Dconv,LSTM and mBERT	Bangla-English	Websites (google play store)	**1DconvLSTM** Acc:0.84%	**6**
[108]	MT and Transfer Learning	BERT	English and non-English	Social media	**BERT** Acc:0.82%	**5**
[109]	Transfer learning	mBERT	Indonesia-English	Social media	**mBERT** Acc:0.76%	**0**
[82]	Word embedding	CNN, LSTM, and BiLSTM	Hindi-English	Social media	**CNN** Acc:75.25%.	**4**
[53]	MT	CNN, LSTM, FFNN, BiLSTM	Amharic	Social media	**FFNN** Acc: 3-class- 62.0% Acc: 2-class-82.2%	4
[110]	Transfer learning	mBERT, PhoBERT, viBERT, ViT5, and viELECTRA	Vietnamese	Hotel, Product, Websites, and Social media	**PhoBERT** Acc:64.65	3
[111]	Transfer learning	LSTM DistilBERT, BERT-RCNN and PhoBERT	Vietnamese	Social media	**PhoBERT** Acc:89.68%	0
[83]	Word embedding	CNN, LSTM, and SVM	Vietnamese	Product and School feedback	**CNN-LSTM** Acc:93.94%	2

TABLE 6. *(Continued.)* List of low-resource sentiment analysis approach, models, languages domain, and best Accuracy evaluation metric.

[112]	Transfer learning	mBERT, RoBERTa and XLM-r	Hausa-English	Social media	**mBERT** Acc:0.73%	**0**
[92]	Transfer learning	mBERT, RoBERTa	Hindi-English	Social media	**RoBERTa** Acc:0.76%	**8**
[113]	Transfer learning	IndoBERT	Indonesian-Sundanese	Social media	**IndoBERT** Acc:0.81%	**1**
[114]	Transfer learning	BERT, DistilBERT, RoBERT and XLM-R	Kannada and Malayalam	Social media	**Ensemble** Acc:0.72%	**2**
[115]	Transfer learning	CNN and mBERT	Malay, Indian, Chinese	Social media	**mBERT** Acc:0.65%	**2**
[116]	Transfer learning	Codemixed transformer and non-codemixed	Hindi-English	Social media	**HinRoBERTa** Acc:0.73%	5
[117]	Transfer learning	BERT	Urdu	Product and Movies	**BERT** Acc:89.53%	**1**

A. (RQ1) WHAT IS THE MOST EFFECTIVE APPROACH IN LOW-RESOURCE MULTILINGUAL SENTIMENT CLASSIFICATION

To address the research question, we have observed the utilization of three broad approaches in low-resource sentiment analysis. The Machine translation approach, exemplified by [47], [48], and [53] attempts to bridge language gaps of sentiment analysis tasks in low resources but faces challenges, with translation not proving effective semantics in low-resource languages. However, word embedding, harnesses distributed representations of words but struggles with issues of polysemy and time consumption. On the other hand, the transfer learning approach, offers promise in sentiment analysis classification using pre-trained models, notably transformer architectures, as exemplified by [94], [105], [111], and [117], which excel in capturing contextual information.

Figure 8 illustrates the selected studies', approaches. The review studies of the approaches reveal that the transfer learning approach is the most prevalent in low-resource sentiment analysis, utilized in 24 studies and followed by word embedding applied in 19 studies. Machine Translation is utilized in 5 studies. Additionally, 3 studies did not employ any of these approaches, while 3 studies combined transfer learning with word embeddings. Furthermore, two studies combined machine translation with either transfer learning or word embedding, respectively. Table 7 presents the pros and cons of each of the approach studies discussed. It indicates that transfer learning has an advantage over word embedding and machine translation approaches.

The taxonomy of the techniques applied in low resource sentiment analysis with the above approaches is depicted in Figure 7. Through our investigation, we identified two primary approaches: non-pretrained techniques and pre-trained techniques. Among non-pretrained techniques, Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) were the most commonly utilized, appearing in 24 and 20 studies followed by Bidirectional LSTM (BiLSTM) in 16 studies. Gated Recurrent Unit (GRU) and Recurrent Neutral Network (RNN) were each employed in 6 and 4 studies, while Bidirectional GRU (BiGRU) were utilized in 4 respectively.

Convolutional Bidirectional LSTM (convBiLSTM), Feed Forward Neural Network (FFNN), Deep Neural Network (DNN), Neutral Machine Translation (NMT), one-dimensional CNN (1d-CNN) were used in single studies.

Numerous pre-trained or transformer techniques were identified, including Bidirectional Encoder Representations from Transformers (BERT) with the highest frequency used in 16 studies, followed by multilingual (mBERT) in 12 studies. Cross-lingual Language Model with Roberta (XLM-R) and Robustly Optimized BERT (RoBERTa) was utilized in studies 6 and 5 studies, while other variant transformers based on BERT in different languages studies were used in single or multiple studies.

Table 8 illustrates the analysis of the reviewed papers and underscores the significance of preprocessing stages in low-resource sentiment analysis tasks, with a particular emphasis on data cleaning as a present practice across the studies. However, The studies [48] and [108] were found to have used NER preprocessing step. Moreover, there is notable variation in the reporting of preprocessing steps, with certain studies such in a study like [64], [69], and [70] failing to document any preprocessing steps. This omission is often attributed to the utilization of online open-source data repositories, where the availability of already generated and preprocessed data obviates the need for explicit preprocessing steps. Studies like [51], [62], [67], and [74], highlight the critical role of stemming and lemmatization, particularly in low-resource sentiment analysis scenarios. The absence of adequate tools for stemming and lemmatization in certain

TABLE 7. Advantages and disadvantages of the approaches.

Approach	Advantage	Disadvantage
Machine Translation	• Communication across diverse languages. • Leveraging high language resources.	• Variation and complexity of languages. • Highly depends on the translators' tool accuracy. • Depends on the availability of target language models. • Require rigorous preprocessing stage.
Word Embedding	• Efficient representation of word semantics. • Save time than training from scratch. • Generalize well to unseen words.	• Limited in polysemy disambiguation. • Cannot handle the syntax and semantic structure of the sentence. • Require rigorous preprocessing stage. • Require computation complexity. • Large corpora are needed for training. • Require a lot of time during the training.
Transfer learning	• Overcomes polysemy disambiguation. • Fine-tuning rather than training from scratch. • Handling the preprocessing stage. • Captures contextual information effectively. • Better performance in handling long-range dependency. • Easy to adopt different domains and language. •	• Require computational resources. • Time-consuming during the training. • Large corpora are needed for training.

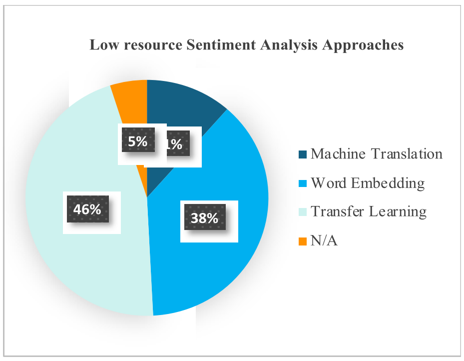

FIGURE 8. Studies' approaches distribution for low sentiment analysis.

languages within NLTK necessitates the development of tailored algorithms for these specific languages.

Tokenization and normalization emerge as near-universal practices across some of the reviewed studies like [47], [48], [115], and [116], underscoring their importance in standardizing text data for subsequent analysis. However, the advent of transfer learning and transformer techniques introduces an alternative approach, where tokenization processes are often integrated within transformer architectures. Nonetheless, challenges persist, particularly in accurately representing all languages within tokenizer frameworks due to variations in language characters and other linguistic factors.

TABLE 8. Preprocessing step.

Preprocessing step	Reference
Data cleaning	[95], [35], [96], [49], [62], [65], [58], [68], [67], [89], [74], [75], [76], [77], [88], [94], [100], [103], [104], [107], [80], [106], [81], [108], [109], [82], [53], [110], [111], [83], [92], [113], [114], [115], [116], [117]
Stemming	[51], [67], [104], [108],
Lemmatization	[62], [74], [104], [108]
Part-of-speech (POS)	[95], [67], [76], [108]
Tokenization	[47], [48], [51], [67], [74], [94], [100], [104], [80], [106], [53], [110], [83], [112], [113], [115], [116]
Normalization	[89], [77], [104], [80], [108], [83]
Name Entity Reg (NER)	[48], [108]
N/A	[64], [69], [70], [91], [84], [90], [101], [102], [127], [66]

Table 9 Shows the analysis of the feature extraction method across reviewed studies in sentiment analysis for low-resource languages. The review reveals several trends. N-grams, TFIDF, CBOW, and Skip Gram are frequently used methods, indicating their widespread usage in this

TABLE 9. Feature extraction method.

Feature	Reference
N-grams	[52], [66], [75], [78]
IF-IDF	[67], [104], [95], [35], [65], [89], [75], [115], [64], [78],
One-hot encoding	[35], [74]
CBOW	[96], , [97], [64], [68], [89], [75], [104]
Skip Gram	[80], [62], [68], [89]
Embedding	[89], [83], [69], [70], [91], [84], [127], [81], [49], [75], [76], [77], [82], [53], [117],
Contextual embedding	[90], [112], [92], [114], [116], [112], [100], [99], [101], [102], [103], [108], [91], [94], [90], [101], [64]
N/A	[47], [48], [51], [94], [88], [106]

domain. Additionally, one-hot encoding is mentioned in a few like [35] and [74], suggesting its presence in certain studies despite it various limitations. Embedding techniques, including both traditional embeddings and contextual embeddings, emerge as popular choices, with numerous studies highlighting their effectiveness. Transformer-based contextual embeddings, although not always explicitly mentioned in the utilized studies, are prevalent sentiment analysis tasks, indicating a growing trend towards leveraging transformer models for feature extraction in low-resource languages. However, several studies like [47], [48], [51], [88], [94], and [106] do not report specific feature extraction techniques, possibly due to the adoption of transformer architectures. Overall, the analysis underscores the diversity of feature extraction methods employed in sentiment analysis for low-resource languages, with a notable shift towards transformer-based approaches.

The evaluation metric of low-resources sentiment, analysis encompasses accuracy, recall, precision, and F1 score. The best-performing metrics are the only ones reported in the SLR. Notably, the accuracy metric emerges as the predominant evaluation metric, being utilized in 46 out of the 56 scrutinized studies with the only best-performing score. the F1-score, which represents the harmonic mean of precision and recall, follows a key metric, employed in 10 studies as the best-performing metric. In contrast, precision and recall are comparatively less performing score-favored evaluation criteria, in most of the implemented studies. Figure 9 illustrates the varying degrees of evaluation best-performing metric.

In terms of performance metrics, several techniques have demonstrated impressive performance. The works of [58] used LSTM in Korean sentiment using word embedding achieved an accuracy of 94.1%. The study [96] achieved a high accuracy of 90.29% for Punjabi news sentiment analysis using RNN. The author [97] employed LSTM for Arabic sentiment analysis with an accuracy of 87.31%. Additionally, Gupta et al. [95] utilizing RNN-LSTM achieved 85.01%

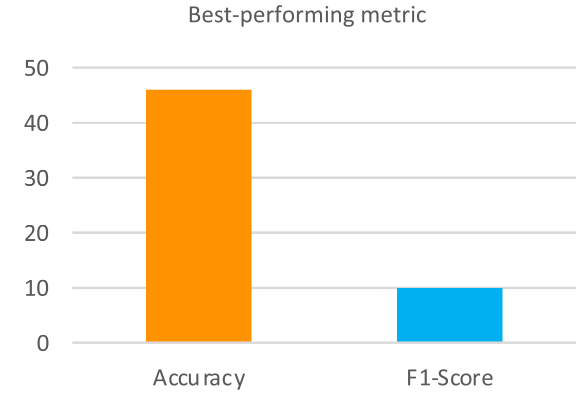

FIGURE 9. Best-performing metrics report for low resource sentiment analysis.

accuracy in Hindi social media sentiment analysis. The study of [47] used LSTM in Urdu, German, and Hindi sentiment and achieved an accuracy of 85.99% similarly, a study by Kumari et al. [48] focused on Hindi tourism sentiment using Neural Machine Translation (NMT), obtaining a remarkable accuracy of 91.99%. Moreover, The observed performance accuracy spanning beyond 80% across various studies employing word embedding approach reflects a diverse array of the approach. The transfer learning approach, specifically leveraging transformer-based models, has demonstrated exceptional accuracy, exemplified by a notable 99% and 97% accuracy in Turkish hotel and movie reviews using BERT, as highlighted in [104]. In a separate investigation by [107] focusing on Hindi and Arabic social media and websites, the utilization of the same transformer achieved an impressive accuracy of 98.72%. In another study by [100], employing the Transformer XLM-R, attained a commendable F1-score of 0.9765% using mBERT in an Indonesia product review.

Notably, various studies consistently report metric levels of 80% and above, underscoring the robust performance of transformer models across diverse languages and domains in sentiment analysis.

B. (RQ2) WHAT ARE THE MOST COMMON SOURCES OF DATA FOR LOW-RESOURCE SENTIMENT ANALYSIS?

Low-resource sentiment analysis relies on various sources of data due to the limited availability of labelled data. However, to answer the above research question, it is essential to research studies. In the outcomes of our investigation into the data utilised. We categorized the data source into two classes: single and multiple data sources. Single data source refers to instances where a paper employs only one source of data domain, such as social media, products, restaurants, or sports. In contrast, multiple source data domains involve the use of two or more data source domains in a paper, such as social media and sports, or movies and hotels.

The results, as depicted in Figure 10 specifically pertain to data source domains used for low-resource sentiment analysis. The analysis reveals the use of different data source domains, comprising 8 single data source domains, and multiple data source domains. Notably, most studies

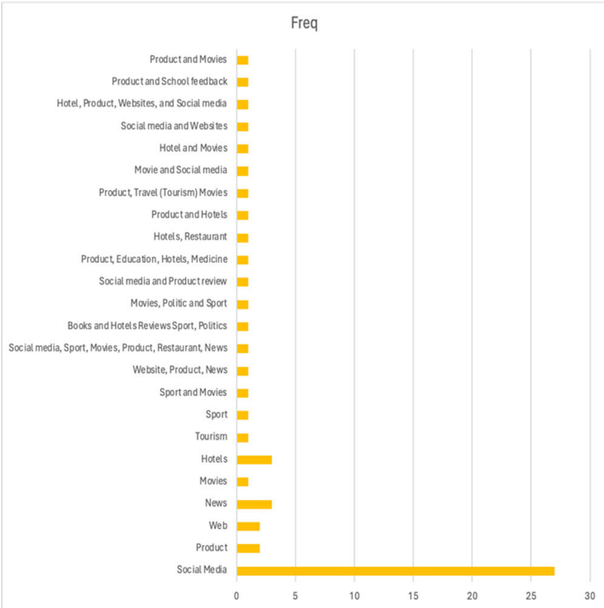

FIGURE 10. Data source/domain utilized for low resource sentiment analysis.

reveal a predominance of social media as the primary source such in studies like [62], [80], [95], [113], and [116], with 27 instances recorded. Other domains, such as product reviews [69] and [100], web content [81] and [101], news [75] and [96], movies [106], hotels [68] and [97], tourism [48], and sports [35] are also represented, although with fewer occurrences.

The utilization of multiple domains of data sources such as in [67], [70], [84], and [90] is crucial as it allows for a comprehensive evaluation of proposed models across diverse contexts and datasets, enhancing the generalizability and robustness of the findings. Different domains may exhibit distinct linguistic characteristics, sentiment expressions, and user behaviours, necessitating the assessment of sentiment analysis models across varied data sources to ensure their effectiveness across different application scenarios.

Social media data, in particular, is highly preferred and prevalent in sentiment analysis research due to data features. It offers a rich source of user-generated content, including text posts, comments, and reviews, reflecting real-time opinions of individuals [8]. This data is often readily accessible, publicly available, and abundant, facilitating large-scale data collection and analysis. Moreover, it encompasses a wide range of topics, discussions, and user demographics [86]. This makes it suitable for exploring diverse research questions and evaluating multilingual low-resource sentiment analysis models across various domains and contexts.

The study suggests a prevailing preference for social media data sources, likely driven by a desired nature of the data context with prevailing challenges. However, addressing challenges is crucial to ensure the responsible and effective use of this abundant resource.

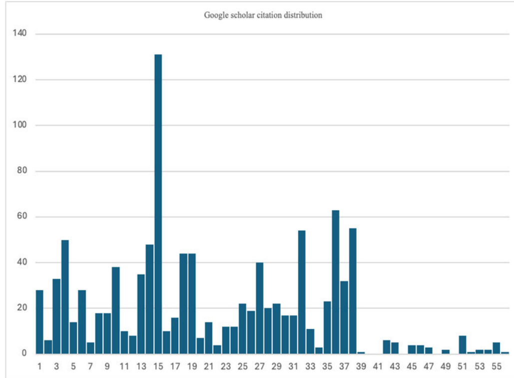

FIGURE 11. Citation score for the selected reviewed papers.

C. (RQ3) WHAT ARE THE CHALLENGES OF LOW-RESOURCE SENTIMENT ANALYSIS?

In the domain of low-resource sentiment analysis, several challenges arise, primarily from the limited availability of digital resources [89], [128]. The effectiveness of sentiment analysis is inherently linked to the accessibility of datasets, which traditionally abound for high-resource languages [48].

The scarcity of language tools, such as part-of-speech (POS) taggers, and NLTK, tailored for low-resource languages poses a significant constraint, limiting the exploration of diverse sentiment analysis methods [129]. Furthermore, the prevalent phenomenon of code-mixing or code-switching in multilingual low-resource communities and social media interactions [30], [65], [130] has witnessed a notable upswing over the past decade [131]. However, these individuals who belong to multiple online communities demonstrated a flexible ability to adjust and switch their patterns of responding to compliments based on the specific online cultural context in which they were engaged [132]. This practice involves the integration of languages with differing linguistic resources within the same textual content [30], resulting in complex linguistic features such as intra-sentential and inter-sentential code-mixing, inventive spelling variations, lexical borrowing, and phonetic typing [65]. These limitations underscore the critical necessity for comprehensive research endeavours in the arena of low-resource languages and sentiment analysis, given their escalating influence in the digital landscape.

The distribution of citation numbers among the papers analyzed reveals a skewed pattern as shown in Figure 11 with some studies having relatively low citation counts, while some papers stand out with significantly higher citation numbers. Specifically, a select few papers have garnered considerable attention, with citation numbers surpassing 40 and even reaching as high as 131.

This distribution suggests a concentration of scholarly influence among a small subset of papers, likely representing significant contributions to the field of deep learning for low-resource setting analysis. The highly cited paper [58] serves as the foundation for introducing a novel approach for augmenting embeddings sentiment lexicons that captures semantic relationships among sentiment words more effectively compared to current word embedding methods, all without requiring manually annotated sentiment corpora. Similarly, [127] used a similar approach to Incorporate linguistic features to enhance the neural network structure of the model used in their studies. the other highly cited paper [99] used an adaptation of the BERT architecture designed specifically for Spanish language usage within the context of Twitter. This indicates how the importance of augmenting embedding features can influence a model's performance for sentiment analysis, shaping the direction and discourse within the domain of deep learning approaches.

VI. DISCUSSION

The section provides a presentation of the primary research findings, offering a comprehensive overview. The section also offers a standard framework for performing sentiment analysis in low-resource for future researchers.

This SLR study of the literature on deep learning-based low-resource sentiment analysis has revealed some interesting findings.

First and foremost, across the diverse studies. The predominant trend reveals a substantial reliance on transfer learning accounting for 46% of the total studies, specifically leveraging transformer-based models. These models consistently exhibit exceptional accuracy, surpassing above 80% accuracy in various linguistic contexts and domains. Notably, the use of standalone deep learning models has shifted towards a significant adoption of pre-trained transformer models. This is largely attributed to their novel nature and feature of being non-labor-intensive. The effectiveness of transfer learning, particularly with transformers, underscores its pivotal role in addressing the challenges posed by low-resource languages. the transformers can represent and support different languages. It also can generalise between languages without being expressly trained for it [133].

Furthermore, another notable finding from this review is the prevalence of social media data as the most utilized source of data accounting for 27 papers reviewed see Figure 10. This is not surprising given the massive amount of user-generated content available on social media platforms and the potential for analyzing sentiment and opinion on a wide range of topics. However, it's important to note that social media data can present challenges, such as noise, code-mixed, use of slang and language use that may differ from other domains. In addition to social media data, product reviews, movies, and hotels were also commonly explored domains for sentiment analysis in low-resource languages, likely due to their relevance. It is interesting to note that these domains are associated with a high degree of subjectivity, making them ideal for sentiment analysis.

However, biases related to representativeness, social, and cultural or linguistic factors are pervasive in much of the available social media data [12]. Therefore, diversifying data sources by integrating traditional media and local reviews, alongside adopting mixed-methods approaches, becomes imperative to mitigate these biases. Adoption of mixed-methods diversification can involve the assessment of social biases such as gender, race, and ethnicity in the models training, as introduced by [134], through the use of Equity Evaluation Corpora (EEC). EEC comprises carefully crafted simple sentences that vary solely in their reference to different social cleavages. The discrepancies in predictions on downstream tasks among these sentences can be attributed to language models assimilating these social and cultural biases.

Additionally, while many studies emphasize accuracy as the primary criterion for selecting sentiment analysis characteristics, it's crucial to acknowledge its potential biases. Relying solely on accuracy may offer a limited perspective as it provides a singular numerical value without delineating the specific errors encountered during evaluation, as highlighted by [36].

Low-resource sentiment analysis challenge is ascribed to data limitation. Primarily due labor intensive of the annotation data [135]. However, to address this scarcity of annotated datasets, several strategies can be explored to enhance the robustness and generalizability of low-resource sentiment analysis models. Thus, data augmentation techniques offer a promising avenue by artificially increasing the size and diversity of existing annotated datasets through methods of synthetic data generation [136]. Additionally, transfer learning approaches enable leveraging knowledge from tasks with abundant data to improve performance on tasks with limited annotations. Similarly, semi-supervised learning techniques utilize both labelled and unlabeled data to train models, effectively leveraging the information present in unlabeled instances. Furthermore, domain adaptation techniques facilitate the transfer of knowledge from a source domain to a target domain with limited annotated data, enhancing the model generalization [137]. Crowdsourcing and collaboration annotation [138], can also scale annotation efforts and alleviate the burden of manual annotation, particularly for specialized domains. By exploring and implementing these diverse strategies, researchers can address the challenges posed by the scarcity of annotated datasets. They can advance the development of sentiment analysis models with improved performance and applicability across various domains.

The study merged various approaches, techniques, data acquisition, data processing data annotation, and feature detection identified through various reviewed papers of the existing literature. We propose a structured framework intended to serve as a guide for future researchers in the field. This framework comprises 10 phases:

1) Data collection.
2) Language identification system.
3) Data preprocessing.
4) Data annotation.
5) Code-switching checker.
6) Approach selection
7) Feature detection
8) Techniques selection

9) Model training and evaluation
10) System deployment

The framework involves a systematic progression. The initial stage of data collection involved sourcing data from different domains comprised of the social media [139], blog [140], product review [58], restaurant, hotels [88], movies, sports [47] etc.

The implementation of a language identification system [141], enables comprehension of the underlying data structure, facilitating the selection of a task tailored to sentiment analysis, be it monolingual, multilingual, or cross-lingual tasks.

Subsequently, a series of pre-processing tasks are undertaken to refine the acquired unstructured nature of the data [142]. These tasks encompass various stages, including data cleaning, stemming, lemmatization, part-of-speech tagging, normalization, named entity recognition (NER), and tokenization. However, not all stages may be applied to every data type or task at hand. The importance of each stage varies depending on the nature of the task at hand [143], and the available data. Additionally, stemming and lemmatization are crucial steps in the preprocessing stage, particularly in low-resource languages where dedicated tools like NLTK may not be readily available. Hence, researchers should explore the creation of language-specific tools or algorithms to address these needs effectively. Moreover, using autoencoder (AE) models has demonstrated effectiveness in overcoming the obstacles related to text data processing in sentiment analysis [144].

Data annotation, a pivotal step in language identification systems, follows, necessitating the establishment of appropriate labels for the dataset. Annotation methods may range from manual [145], semi-automated [146] and automated approaches [147]. with shared tasks and crowdsourcing emerging as common strategies for manual annotation [121]. Although manual annotation is labor labour-intensive and time-consuming task [146]. The automatic data annotation leverages auto-encoders [148]. The annotation based on polarity classification is in two types thus: binary or multiple class annotation. To advance to subsequent step, a rigorous evaluation of the annotated data is needed by experts in the specific language to ensure its validity, coupled with an assessment of quality criteria to uphold dataset annotation integrity.

The code-switching checker aims to detect and identify occurrences of code-switching in the text data. By discerning language switches, the sentiment analysis framework can adjust its processing accordingly through language identification. This adaptation enables the framework to determine data augmentation strategies for improved system performance.

Approach selection: the framework offers the flexibility to select an approach, such as machine translation, word embedding, or transfer learning. However, given the aforementioned benefits associated with each approach, the preference is strongly inclined towards employing the transfer learning approach.

Feature extraction is subsequently conducted to convert textual data into numerical formats, utilizing methodologies like N-grams, TF-IDF, and embeddings such as GloVe, FastText, or Word2Vec. This process is particularly relevant when opting for either the word embedding approach or machine translation. However, the need for such steps is mitigated by transfer learning. The features are extracted using transformer contextual embeddings and a transformer tokenizer [149].

The subsequent phase is the classification models: pre-trained (transformers) and non-pretrained models. This involves training, validation, and testing sets sampled from the dataset. The training set is employed to uncover underlying patterns, the validation set is utilised for model refinement and hyperparameter tuning. The testing set is employed to evaluate model performance through metrics such as accuracy, precision, recall, and F-score at the final stage.

The model derived from the classification system modelling phase serves as the foundation for sentiment analysis in low-resource languages. Figure 12 provides a visual depiction of the comprehensive proposed framework.

VII. STUDY'S IMPLICATION

This SLR offers significant implications for future researchers in low-resource sentiment analysis. Prior studies have often covered a range of approaches and techniques for sentiment analysis in low-resource settings. They often treat them as a single method. However, none have extensively identified the most effective techniques specifically tailored for low-resource sentiment analysis.

The study introduces a conceptual framework for developing low-resource sentiment analysis applications. This framework is proposed to be valuable for researchers, and practical application in low-resource studies. It outlines general steps that can serve as practical guidelines for emerging researchers. Key components of the framework include language identification, task selection, preprocessing, data annotation, code-switching checker, approach selection, feature detection techniques, model selection and training, evaluation, and deployment. The importance of each step is emphasized, with effective implementation of the framework in classification modelling to derive optimal sentiment solutions.

VIII. FUTURE DIRECTION

This section highlights key gap considerations specific to the low-resource languages. The extracted main points serve as foundational insights for guiding future research:

Data Limitation for Low-Resource Sentiment Analysis: Future researchers are encouraged to actively participate in generating datasets for low-resource languages. Additionally, providing standardized benchmark datasets for low-resource languages can enhance result consistency and facilitate cross-study comparisons.

Expanding Approaches With Reduced Dependency on Labelled Data: Data annotation is a tedious task in sentiment

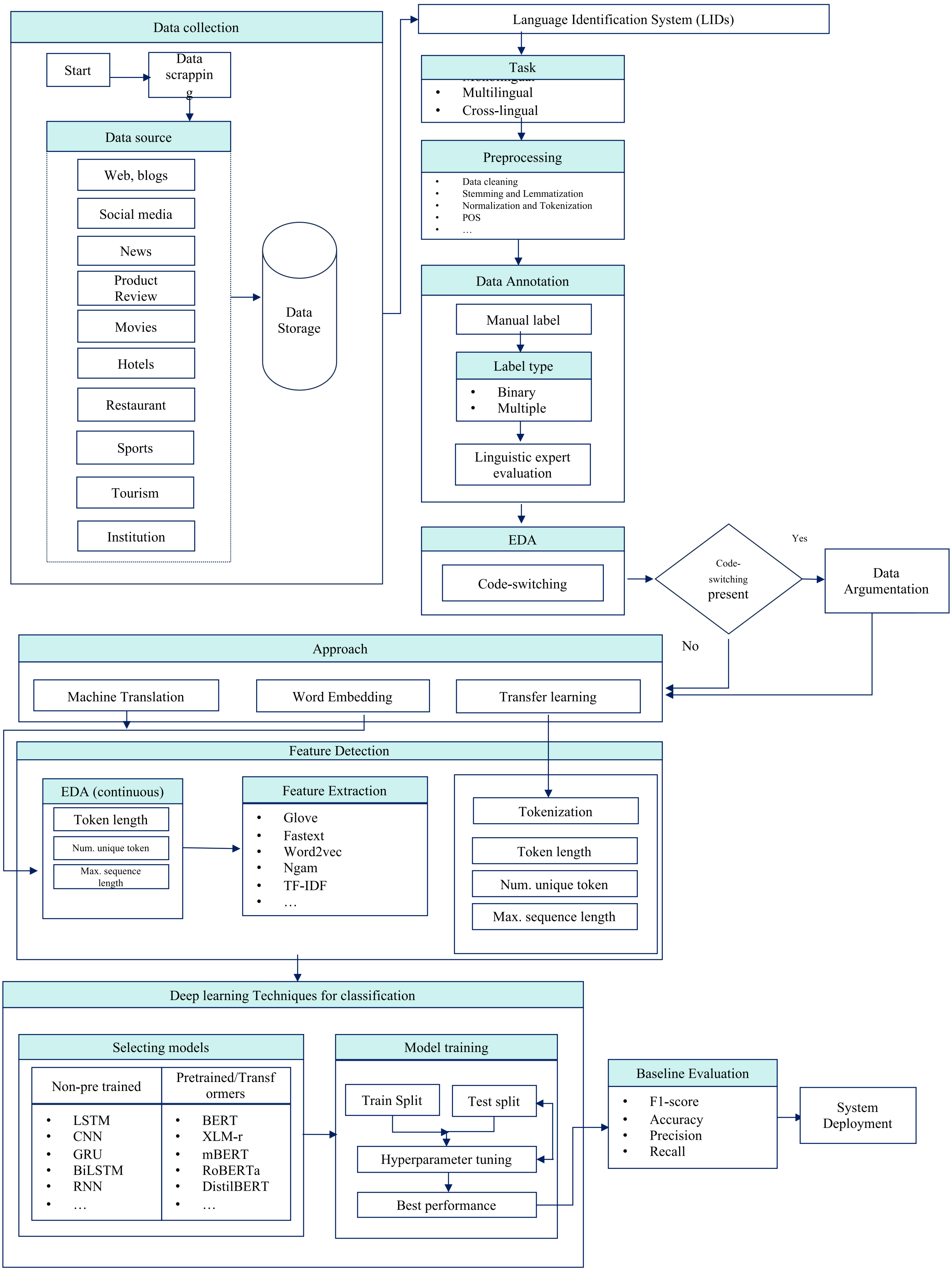

FIGURE 12. Framework of the low-resource sentiment analysis.

analysis. Thereby, given such limitations associated with obtaining standard data annotation for low-resource language. It is recommended to prioritize methods such as unsupervised learning through a transfer learning approach (zero-shot learning). The transfer learning approach through transformer models alleviates the need for extensive data with labelled sentiment.

Data Argumentation and Domain Knowledge: Future research is recommended to delve into data augmentation techniques and leverage domain knowledge to address the challenges inherent in code-mixed datasets, particularly those sourced from social media and product reviews. These datasets often exhibit characteristics such as noise, data imbalance, text variation, grammatical errors, spelling inaccuracies, and informal writing styles. Researchers are encouraged to explore a diverse range of data sources beyond social media, including news articles, emails, and online reviews in various languages. By doing so, they can mitigate the challenges associated with sentiment analysis in low-resource settings.

IX. CONCLUSION

In conclusion, the insights gained from this SLR of the literature on deep learning-based low-resource multilingual sentiment analysis provide valuable guidance for researchers in this field. By understanding the most common sources and data domains utilized and the most popular deep learning models i.e., transformers, researchers can design better experiments and develop more accurate and reliable sentiment analysis models for low-resource languages.

Moreover, It highlights the opportunities that exist in this rapidly evolving field. Transfer learning, followed by, word embeddings, are the most commonly used approaches used, with social media data being the most commonly used platform for sentiment analysis in low-resource languages. The use of deep learning models employing transformer models is on the rise, indicating the growing interest in investigating pre-trained models for low-resource languages.

The prevalence of social media data, product reviews, movies, and hotels as common data source domains for sentiment analysis suggests that these areas may provide a fruitful avenue for further research. Furthermore, future research in low-resource multilingual sentiment analysis should focus on developing new approaches by utilizing transformers that can address the challenges posed by limited labelled data. This includes the development of methods for domain adaptation and data augmentation. Finally, a proposed framework is provided which can serve as a guide for sentiment analysis in low-resource. In conclusion, low-resource multilingual sentiment analysis is a challenging yet promising area of research. The availability of various sources and data domains, as well as the use of deep learning models employing transformer models, offers a promising way forward. The findings from this review have important implications for researchers working in this field, and we hope that this discussion will contribute to the development of new and effective approaches for low-resource sentiment analysis.

ACKNOWLEDGMENT

The authors extend their gratitude to ''Universiti Teknologi PETRONAS (UTP). They acknowledge their support, participation, and knowledge contributions, which played a pivotal role in facilitating the dissemination of their research findings, active involvement in emerging developments, and the establishment of valuable professional connections in our field. They highly value the investment made in their academic and practical development.

Informed Consent: Since there were no human or animal subjects in the study, informed consent was not required.

Conflict of Interest: The authors confirm that they do not have any competing interests or conflicts of interest.

REFERENCES

[1] Y. Chen and C.-F. Wang, ''Introduction,'' in *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, B. Liu, Ed. Cambridge, U.K.: Cambridge Univ. Press, 2015, pp. 1–15.

[2] M. Araújo, A. Pereira, and F. Benevenuto, ''A comparative study of machine translation for multilingual sentence-level sentiment analysis,'' *Inf. Sci.*, vol. 512, pp. 1078–1102, Feb. 2020, doi: 10.1016/j.ins.2019.10.031.

[3] V. Ganganwar and R. Rajalakshmi, ''Implicit aspect extraction for sentiment analysis: A survey of recent approaches,'' *Proc. Comput. Sci.*, vol. 165, pp. 485–491, Jan. 2019, doi: 10.1016/j.procs.2020.01.010.

[4] N. N. Yusof, A. Mohamed, and S. Rahman, ''A review of contextual information for context-based approach in sentiment analysis,'' *Int. J. Mach. Learn. Comput.*, vol. 8, pp. 399–403, Jan. 2018, doi: 10.18178/ijmlc.2018.8.4.719.

[5] T. A. Rana and Y.-N. Cheah, ''Aspect extraction in sentiment analysis: Comparative analysis and survey,'' *Artif. Intell. Rev.*, vol. 46, no. 4, pp. 459–483, Dec. 2016, doi: 10.1007/s10462-016-9472-z.

[6] A. S. Shafie, N. M. Sharef, M. A. A. Murad, and A. Azman, ''Aspect extraction performance with POS tag pattern of dependency relation in aspect-based sentiment analysis,'' in *Proc. 4th Int. Conf. Inf. Retr. Knowl. Manage. (CAMP)*, Mar. 2018, pp. 1–6, doi: 10.1109/INFRKM.2018.8464692.

[7] M. Alassaf and A. M. Qamar, ''Aspect-based sentiment analysis of Arabic tweets in the education sector using a hybrid feature selection method,'' in *Proc. 14th Int. Conf. Innov. Inf. Technol. (IIT)*, Nov. 2020, pp. 178–185, doi: 10.1109/IIT50501.2020.9299026.

[8] A. Ligthart, C. Catal, and B. Tekinerdogan, ''Systematic reviews in sentiment analysis: A tertiary study,'' *Artif. Intell. Rev.*, vol. 54, no. 7, pp. 4997–5053, Oct. 2021, doi: 10.1007/s10462-021-09973-3.

[9] M. Tubishat, N. Idris, and M. A. M. Abushariah, ''Implicit aspect extraction in sentiment analysis: Review, taxonomy, oppportunities, and open challenges,'' *Inf. Process. Manage.*, vol. 54, no. 4, pp. 545–563, Jul. 2018, doi: 10.1016/j.ipm.2018.03.008.

[10] A. Nazir, Y. Rao, L. Wu, and L. Sun, ''Issues and challenges of aspect-based sentiment analysis: A comprehensive survey,'' *IEEE Trans. Affect. Comput.*, vol. 13, no. 2, pp. 845–863, Apr. 2022, doi: 10.1109/TAFFC.2020.2970399.

[11] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.-F. Chang, and M. Pantic, ''A survey of multimodal sentiment analysis,'' *Image Vis. Comput.*, vol. 65, pp. 3–14, Sep. 2017, doi: 10.1016/j.imavis.2017.08.003.

[12] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P.-M. Cuenca-Jiménez, ''A review on sentiment analysis from social media platforms,'' *Expert Syst. Appl.*, vol. 223, Aug. 2023, Art. no. 119862, doi: 10.1016/j.eswa.2023.119862.

[13] M. Sangeetha and K. Nimala, ''Exploration of sentiment analysis techniques on a multilingual dataset dealing with tamil-english reviews,'' in *Proc. Int. Conf. Adv. Comput., Commun. Appl. Informat. (ACCAI)*, Jan. 2022, pp. 1–8, doi: 10.1109/ACCAI53970.2022.9752612.

[14] A. Chatterjee, V. Guptha, P. Chopra, and A. Das, ''Minority positive sampling for switching points—An anecdote for the code-mixing language modeling,'' in *Proc. LREC*, 2020, pp. 6228–6236.

[15] A. I. Abubakar, A. Roko, A. M. Bui, and I. Saidu, ''An enhanced feature acquisition for sentiment analysis of English and hausa tweets,'' *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 9, pp. 102–110, 2021, doi: 10.14569/ijacsa.2021.0120913.

[16] K. Dashtipour, S. Poria, A. Hussain, E. Cambria, A. Y. A. Hawalah, A. Gelbukh, and Q. Zhou, ''Multilingual sentiment analysis: State of the art and independent comparison of techniques,'' *Cognit. Comput.*, vol. 8, no. 4, pp. 757–771, Aug. 2016, doi: 10.1007/s12559-016-9415-7.

[17] A. Magueresse, V. Carles, and E. Heetderks, "Low-resource languages: A review of past work and future challenges," 2020, *arXiv:2006.07264*.

[18] F. Laumann. (2022). *Low-Resource Language: What Does It Mean*. Accessed: 2022. [Online]. Available: https://medium.com/neuralspace/low-resource-language-what-does-it-mean-d067ec85dea5

[19] *Multilingual Sentiment Analysis: How to Do It Right*. Accessed: Oct. 7, 2023. [Online]. Available: https://www.crisoltranslations.com/our-blog/

[20] J. Z. Maitama, N. Idris, A. Abdi, L. Shuib, and R. Fauzi, "A systematic review on implicit and explicit aspect extraction in sentiment analysis," *IEEE Access*, vol. 8, pp. 194166–194191, 2020, doi: 10.1109/ACCESS.2020.3031217.

[21] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014, doi: 10.1016/j.asej.2014.04.011.

[22] M. A. Hedderich, L. Lange, H. Adel, J. Strötgen, and D. Klakow, "A survey on recent approaches for natural language processing in low-resource scenarios," 2020, *arXiv:2010.12309*.

[23] M. M. Agüero-Torales, J. I. A. Salas, and A. G. López-Herrera, "Deep learning and multilingual sentiment analysis on social media data: An overview," *Appl. Soft Comput.*, vol. 107, Aug. 2021, Art. no. 107373, doi: 10.1016/j.asoc.2021.107373.

[24] K. R. Mabokela, T. Celik, and M. Raborife, "Multilingual sentiment analysis for under-resourced languages: A systematic review of the landscape," *IEEE Access*, vol. 11, pp. 15996–16020, 2023, doi: 10.1109/ACCESS.2022.3224136.

[25] Y. Xu, H. Cao, W. Du, and W. Wang, "A survey of cross-lingual sentiment analysis: Methodologies, models and evaluations," *Data Sci. Eng.*, vol. 7, no. 3, pp. 279–299, Sep. 2022, doi: 10.1007/s41019-022-00187-3.

[26] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *WIREs Data Mining Knowl. Discovery*, vol. 8, no. 4, p. e1253, Jul. 2018, doi: 10.1002/widm.1253.

[27] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," *Electronics*, vol. 9, no. 3, p. 483, Mar. 2020. [Online]. Available: https://www.mdpi.com/2079-9292/9/3/483

[28] A. Kumar and G. Garg, "Systematic literature review on context-based sentiment analysis in social multimedia," *Multimedia Tools Appl.*, vol. 79, nos. 21–22, pp. 15349–15380, Jun. 2020, doi: 10.1007/s11042-019-7346-5.

[29] Q. A. Xu, V. Chang, and C. Jayne, "A systematic review of social media-based sentiment analysis: Emerging trends and challenges," *Decis. Anal. J.*, vol. 3, Jun. 2022, Art. no. 100073, doi: 10.1016/j.dajour.2022.100073.

[30] N. H. Mahadzir, "Sentiment analysis of code-mixed text: A review," *Turkish J. Comput. Math. Educ.* , vol. 12, no. 3, pp. 2469–2478, Apr. 2021.

[31] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif. Intell. Rev.*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.

[32] S. Sagnika, A. Pattanaik, B. Shankar Prasad Mishra, and S. K. Meher, "A review on multi-lingual sentiment analysis by machine learning methods," *J. Eng. Sci. Technol. Rev.*, vol. 13, no. 2, pp. 154–166, Apr. 2020, doi: 10.25103/jestr.132.19.

[33] G. I. Ahmad, J. Singla, and N. Nikita, "Review on sentiment analysis of Indian languages with a special focus on code mixed Indian languages," in *Proc. Int. Conf. Autom., Comput. Technol. Manage. (ICACTM)*, Apr. 2019, pp. 352–356, doi: 10.1109/ICACTM.2019.8776796.

[34] D. S. Kulkarni and S. S. Rodd, "Sentiment analysis in Hindi—A survey on the state-of-the-art techniques," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 21, no. 1, pp. 1–46, Jan. 2022, doi: 10.1145/3469722.

[35] A. Altaf, M. W. Anwar, M. H. Jamal, S. Hassan, U. I. Bajwa, G. S. Choi, and I. Ashraf, "Deep learning based cross domain sentiment classification for Urdu language," *IEEE Access*, vol. 10, pp. 102135–102147, 2022, doi: 10.1109/ACCESS.2022.3208164.

[36] N. A. S. Abdullah and N. I. A. Rusli, "Multilingual sentiment analysis: A systematic literature review," *Pertanika J. Sci. Technol.*, vol. 29, no. 1, pp. 445–470, 2021.

[37] N. A. D. Suhaimi and H. Abas, "A systematic literature review on supervised machine learning algorithms," *Perintis Ejournal*, vol. 10, no. 1, pp. 1–24, 2020.

[38] A. Alsayat, "Improving sentiment analysis for social media applications using an ensemble deep learning language model," *Arabian J. Sci. Eng.*, vol. 47, no. 2, pp. 2499–2511, Feb. 2022, doi: 10.1007/s13369-021-06227-w.

[39] K. Cortis and B. Davis, "Over a decade of social opinion mining: A systematic review," *Artif. Intell. Rev.*, vol. 54, no. 7, pp. 4873–4965, Oct. 2021, doi: 10.1007/s10462-021-10030-2.

[40] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowl.-Based Syst.*, vol. 226, Aug. 2021, Art. no. 107134, doi: 10.1016/j.knosys.2021.107134.

[41] G. I. Ahmad, J. Singla, A. Ali, A. A. Reshi, and A. A. Salameh, "Machine learning techniques for sentiment analysis of code-mixed and switched Indian social media text corpus—A comprehensive review," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 2, pp. 455–467, 2022.

[42] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Keele Univ., Keele, U.K., Univ. Durham, Durham, U.K., EBSE Tech. Rep. EBSE-2007-01, 2007, vol. 2, pp. 1–57.

[43] Z. Munn, M. D. J. Peters, C. Stern, C. Tufanaru, A. McArthur, and E. Aromataris, "Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach," *BMC Med. Res. Methodol.*, vol. 18, no. 1, p. 143, Dec. 2018, doi: 10.1186/s12874-018-0611-x.

[44] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *Int. J. Surg.*, vol. 88, Jan. 2020, Art. no. 105906.

[45] S. L. Lo, E. Cambria, R. Chiong, and D. Cornforth, "Multilingual sentiment analysis: From formal to informal and scarce resource languages," *Artif. Intell. Rev.*, vol. 48, no. 4, pp. 499–527, Dec. 2017, doi: 10.1007/s10462-016-9508-4.

[46] D. Liu, N. Ma, F. Yang, and X. Yang, "A survey of low resource neural machine translation," in *Proc. 4th Int. Conf. Mech., Control Comput. Eng. (ICMCCE)*, Oct. 2019, pp. 1–11, doi: 10.1109/ICMCCE48743.2019.00017.

[47] A. Ghafoor, A. S. Imran, S. M. Daudpota, Z. Kastrati, R. Batra, and M. A. Wani, "The impact of translating resource-rich datasets to low-resource languages through multi-lingual text processing," *IEEE Access*, vol. 9, pp. 124478–124490, 2021, doi: 10.1109/ACCESS.2021.3110285.

[48] D. Kumari, A. Ekbal, R. Haque, P. Bhattacharyya, and A. Way, "Reinforced NMT for sentiment and content preservation in low-resource scenario," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 20, no. 4, pp. 1–27, Jul. 2021, doi: 10.1145/3450970.

[49] S. Thara and P. Poornachandran, "Social media text analytics of Malayalam–English code-mixed using deep learning," *J. Big Data*, vol. 9, no. 1, p. 45, Dec. 2022, doi: 10.1186/s40537-022-00594-3.

[50] N. Sabri, A. Edalat, and B. Bahrak, "Sentiment analysis of Persian–English code-mixed texts," in *Proc. 26th Int. Comput. Conf., Comput. Soc. Iran (CSICC)*, Mar. 2021, pp. 1–4, doi: 10.1109/CSICC52343.2021.9420605.

[51] S. Sazzed and S. Jayarathna, "A sentiment classification in Bengali and machine translated English corpus," in *Proc. IEEE 20th Int. Conf. Inf. Reuse Integr. Data Sci. (IRI)*, Jul. 2019, pp. 107–114, doi: 10.1109/IRI.2019.00029.

[52] F. Koto and I. Koto, "Towards computational linguistics in Minangkabau language: Studies on sentiment analysis and machine translation," in *Proc. Pacific Asia Conf. Lang., Inf. Comput.*, 2020, pp. 138–148.

[53] S. G. Tesfagergish, R. Damasevicius, and J. Kapočiūtė-Dzikienė, "Deep learning-based sentiment classification in amharic using multi-lingual datasets," *Comput. Sci. Inf. Syst.*, vol. 20, no. 4, pp. 1459–1481, 2023.

[54] T. Isbister, F. Carlsson, and M. Sahlgren, "Should we stop training more monolingual models, and simply use machine translation instead?" 2021, *arXiv:2104.10441*.

[55] J. Noh and R. Kavuluru, "Improved biomedical word embeddings in the transformer era," *J. Biomed. Informat.*, vol. 120, Aug. 2021, Art. no. 103867, doi: 10.1016/j.jbi.2021.103867.

[56] R. A. Salama, A. Youssef, and A. Fahmy, "Morphological word embedding for Arabic," *Proc. Comput. Sci.*, vol. 142, pp. 83–93, Jan. 2018, doi: 10.1016/j.procs.2018.10.463.

[57] M. Toshevska, F. Stojanovska, and J. Kalajdjieski, "Comparative analysis of word embeddings for capturing word similarities," 2020, *arXiv:2005.03812*.

[58] M. Song, H. Park, and K.-S. Shin, "Attention-based long short-term memory network using sentiment lexicon embedding for aspect-level sentiment analysis in Korean," *Inf. Process. Manage.*, vol. 56, no. 3, pp. 637–653, May 2019, doi: 10.1016/j.ipm.2018.12.005.

[59] H. Nankani, H. Dutta, H. Shrivastava, P. V. N. S. R. Krishna, D. Mahata, and R. R. Shah, "Multilingual sentiment analysis," in *Deep Learning-Based Approaches for Sentiment Analysis*, B. Agarwal, R. Nayak, N. Mittal, and S. Patnaik, Eds. Singapore: Springer, 2020, pp. 193–236.

[60] S. Jameel, Z. Bouraoui, and S. Schockaert, "Unsupervised learning of distributional relation vectors," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, 2018, pp. 23–33, doi: 10.18653/v1/p18-1003.

[61] L. Wei, Z. Wang, J. Xu, Y. Shi, Q. Wang, L. Shi, Y. Tao, and Y. Gao, "A lightweight sentiment analysis framework for a micro-intelligent terminal," *Sensors*, vol. 23, no. 2, p. 741, Jan. 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/2/741

[62] O. M. Singh, S. Timilsina, B. K. Bal, and A. Joshi, "Aspect based abusive sentiment detection in Nepali social media texts," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Dec. 2020, pp. 301–308, doi: 10.1109/ASONAM49781.2020.9381292.

[63] R. Řehůřek and P. Sojka, "Software framework for topic modelling with large corpora," in *Proc. LREC Workshop New Challenges NLP Frameworks*. Brno, Czech Republic: Masaryk Univ., 2010, pp. 45–50, doi: 10.13140/2.1.2393.1847.

[64] L. Moudjari and K. Akli-Astouati, "An experimental study on sentiment classification of Algerian dialect texts," *Proc. Comput. Sci.*, vol. 176, pp. 1151–1159, Jan. 2020, doi: 10.1016/j.procs.2020.09.111.

[65] A. Konate and R. Du, "Sentiment analysis of code-mixed Bambara–French social media text using deep learning techniques," *Wuhan Univ. J. Natural Sci.*, vol. 23, no. 3, pp. 237–243, Jun. 2018, doi: 10.1007/s11859-018-1316-z.

[66] H. T. Nguyen and M. Le Nguyen, "Multilingual opinion mining on Youtube—A convolutional N-gram BiLSTM word embedding," *Inf. Process. Manage.*, vol. 54, no. 3, pp. 451–462, May 2018, doi: 10.1016/j.ipm.2018.02.001.

[67] N. R. Bhowmik, M. Arifuzzaman, and M. R. H. Mondal, "Sentiment analysis on Bangla text using extended lexicon dictionary and deep learning algorithms," *Array*, vol. 13, Mar. 2022, Art. no. 100123, doi: 10.1016/j.array.2021.100123.

[68] R. Bensoltane, "Comparing word embedding models for Arabic aspect category detection using a deep learning-based approach," in *Proc. E3S Web Conf.*, vol. 297, 2021, p. 1072, doi: 10.1051/e3sconf/202129701072.

[69] M. S. Akhtar, P. Sawant, S. Sen, A. Ekbal, and P. Bhattacharyya, "Improving word embedding coverage in less-resourced languages through multi-linguality and cross-linguality: A case study with aspect-based sentiment analysis," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 18, no. 2, pp. 1–22, Jun. 2019, doi: 10.1145/3273931.

[70] M. Berrimi, M. Oussalah, A. Moussaoui, and M. Saidi, "Attention mechanism architecture for Arabic sentiment analysis," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 22, no. 4, pp. 1–26, Apr. 2023, doi: 10.1145/3578265.

[71] M. Aly and A. Atiya, "LABR: A large scale Arabic book reviews dataset," in *Proc. 51st Annu. Meeting Assoc. Comput. Linguistics*, 2013, pp. 494–498.

[72] A. Elnagar, Y. Khalifa, and A. Einea, "Hotel Arabic—Reviews dataset construction for sentiment analysis applications," in *Intelligent Natural Language Processing: Trends and Applications* (Studies in Computational Intelligence), vol. 740, K. Shaalan, A. Hassanien, and F. Tolba, Eds. Cham, Switzerland: Springer, 2018, pp. 35–52, doi: 10.1007/978-3-319-67056-0_3.

[73] A. Elnagar and O. Einea, "BRAD 1.0: Book reviews in Arabic dataset," in *Proc. IEEE/ACS 13th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Nov. 2016, pp. 1–8.

[74] M. Rafi-Ur-Rashid, M. Mahbub, and M. A. Adnan, "Breaking the curse of class imbalance: Bangla text classification," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 21, no. 5, pp. 1–21, Sep. 2022, doi: 10.1145/3511601.

[75] S. Ranathunga and I. U. Liyanage, "Sentiment analysis of Sinhala news comments," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 20, no. 4, pp. 1–23, Jul. 2021, doi: 10.1145/3445035.

[76] P. Demotte, L. Senevirathne, B. Karunanayake, U. Munasinghe, and S. Ranathunga, "Sentiment analysis of sinhala news comments using sentence-state LSTM networks," in *Proc. Moratuwa Eng. Res. Conf. (MERCon)*, Jul. 2020, pp. 283–288, doi: 10.1109/MERCon50084.2020.9185327.

[77] A. Dahou, S. Xiong, J. Zhou, and M. A. Elaziz, "Multi-channel embedding convolutional neural network model for Arabic sentiment classification," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 18, no. 4, pp. 1–23, Dec. 2019, doi: 10.1145/3314941.

[78] M. Amjad, N. Ashraf, A. Zhila, G. Sidorov, A. Zubiaga, and A. Gelbukh, "Threatening language detection and target identification in Urdu tweets," *IEEE Access*, vol. 9, pp. 128302–128313, 2021, doi: 10.1109/ACCESS.2021.3112500.

[79] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics*, 2016, pp. 427–431.

[80] U. R. Rayala, K. Seshadri, and N. B. Sristy, "Sentiment analysis of code-mixed telugu-english data leveraging syllable and word embeddings," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 22, no. 10, pp. 1–30, Oct. 2023, doi: 10.1145/3620670.

[81] M. Tareq, Md. F. Islam, S. Deb, S. Rahman, and A. A. Mahmud, "Data-augmentation for Bangla–English code-mixed sentiment analysis: Enhancing cross linguistic contextual understanding," *IEEE Access*, vol. 11, pp. 51657–51671, 2023, doi: 10.1109/ACCESS.2023.3277787.

[82] S. Das and T. Singh, "Sentiment recognition of hinglish code mixed data using deep learning models based approach," in *Proc. 13th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence)*, Jan. 2023, pp. 265–269, doi: 10.1109/Confluence56041.2023.10048879.

[83] C. N. Dang, M. N. Moreno-García, F. De la Prieta, K. V. Nguyen, and V. M. Ngo, "Sentiment analysis for Vietnamese–based hybrid deep learning models," in *Hybrid Artificial Intelligent Systems*. Cham, Switzerland: Springer, 2023, pp. 293–303.

[84] K. Kanclerz, P. Miłkowski, and J. Kocoń, "Cross-lingual deep neural transfer learning in sentiment analysis," *Proc. Comput. Sci.*, vol. 176, pp. 128–137, Jan. 2020, doi: 10.1016/j.procs.2020.08.014.

[85] J. Kocoń, M. Zaśko-Zielińska, and P. Miłkowski, "Multi-level analysis and recognition of the text sentiment on the example of consumer opinions," in *Proc. Int. Conf. Recent Adv. Natural Lang. Process.*, Oct. 2019, pp. 559–567.

[86] 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, doi: 10.1109/ACCESS.2019.2925059.

[87] M. T. Jones. (2023). *Transfer Learning for Deep Learning*. Accessed: Mar. 10, 2023. [Online]. Available: https://developer.ibm.com/articles/transfer-learning-for-deep-learning/

[88] D. Van Thin, N. L.-T. Nguyen, T. M. Truong, L. S. Le, and D. T. Vo, "Two new large corpora for Vietnamese aspect-based sentiment analysis at sentence level," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 20, no. 4, pp. 1–22, Jul. 2021, doi: 10.1145/3446678.

[89] M. Marreddy, S. R. Oota, L. S. Vakada, V. C. Chinni, and R. Mamidi, "Am I a resource-poor language? Data sets, embeddings, models and analysis for four different NLP tasks in Telugu language," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 22, no. 1, pp. 1–34, Jan. 2023, doi: 10.1145/3531535.

[90] A. Pathak, S. Kumar, P. Roy, and B.-G. Kim, "Aspect-based sentiment analysis in Hindi language by ensembling pre-trained mBERT models," *Electronics*, vol. 10, no. 21, p. 2641, Oct. 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/21/2641

[91] K. Shanmugavadivel, V. E. Sathishkumar, S. Raja, T. B. Lingaiah, S. Neelakandan, and M. Subramanian, "Deep learning based sentiment analysis and offensive language identification on multilingual code-mixed data," *Sci. Rep.*, vol. 12, no. 1, p. 21557, Dec. 2022, doi: 10.1038/s41598-022-26092-3.

[92] G. Takawane, A. Phaltankar, V. Patwardhan, A. Patil, R. Joshi, and M. S. Takalikar, "Language augmentation approach for code-mixed text classification," *Natural Lang. Process. J.*, vol. 5, Dec. 2023, Art. no. 100042, doi: 10.1016/j.nlp.2023.100042.

[93] J. Tao and X. Fang, "Toward multi-label sentiment analysis: A transfer learning based approach," *J. Big Data*, vol. 7, no. 1, p. 1, Dec. 2020, doi: 10.1186/s40537-019-0278-0.

[94] A. B. Nassif, A. M. Darya, and A. Elnagar, "Empirical evaluation of shallow and deep learning classifiers for Arabic sentiment analysis," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 21, no. 1, pp. 1–25, Jan. 2022, doi: 10.1145/3466171.

[95] V. Gupta, N. Jain, S. Shubham, A. Madan, A. Chaudhary, and Q. Xin, "Toward integrated CNN-based sentiment analysis of tweets for scarce-resource language—Hindi," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 20, no. 5, pp. 1–23, Sep. 2021, doi: 10.1145/3450447.

[96] J. Singh, G. Singh, R. Singh, and P. Singh, "Morphological evaluation and sentiment analysis of Punjabi text using deep learning classification," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 33, no. 5, pp. 508–517, Jun. 2021, doi: 10.1016/j.jksuci.2018.04.003.

[97] S. Al-Dabet, S. Tedmori, and M. Al-Smadi, "Enhancing Arabic aspect-based sentiment analysis using deep learning models," *Comput. Speech Lang.*, vol. 69, Sep. 2021, Art. no. 101224, doi: 10.1016/j.csl.2021.101224.

[98] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, I. Androutsopoulos, and N. Bel, "SemEval-2016 task 5: Aspect based sentiment analysis," in *Proc. Int. Workshop Semantic Eval.*, 2016, pp. 19–30.

[99] J. Á. González, L.-F. Hurtado, and F. Pla, "TWilBert: Pre-trained deep bidirectional transformers for Spanish Twitter," *Neurocomputing*, vol. 426, pp. 58–69, Feb. 2021, doi: 10.1016/j.neucom.2020.09.078.

[100] A. N. Azhar and M. L. Khodra, "Fine-tuning pretrained multilingual BERT model for Indonesian aspect-based sentiment analysis," in *Proc. 7th Int. Conf. Advance Inform., Concepts, Theory Appl. (ICAICTA)*, Sep. 2020, pp. 1–6.

[101] Z. Li, X. Li, J. Sheng, and W. Slamu, "AgglutiFiT: Efficient low-resource agglutinative language model fine-tuning," *IEEE Access*, vol. 8, pp. 148489–148499, 2020, doi: 10.1109/ACCESS.2020.3015854.

[102] Z. Li, X. Ding, and T. Liu, "TransBERT: A three-stage pre-training technology for story-ending prediction," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 20, no. 1, pp. 1–20, Jan. 2021, doi: 10.1145/3427669.

[103] A. Jamatia, S. D. Swamy, B. Gambäck, A. Das, and S. Debbarma, "Deep learning based sentiment analysis in a code-mixed English–Hindi and English–Bengali social media corpus," *Int. J. Artif. Intell. Tools*, vol. 29, no. 5, Aug. 2020, Art. no. 2050014, doi: 10.1142/s0218213020500141.

[104] A. Özçift, K. Akarsu, F. Yumuk, and C. Söylemez, "Advancing natural language processing (NLP) applications of morphologically rich languages with bidirectional encoder representations from transformers (BERT): An empirical case study for Turkish," *Automatika*, vol. 62, no. 2, pp. 226–238, Apr. 2021, doi: 10.1080/00051144.2021.1922150.

[105] S. H. Muhammad, D. I. Adelani, S. Ruder, I. S. Ahmad, I. Abdulmumin, B. S. Bello, M. Choudhury, C. C. Emezue, S. S. Abdullahi, A. Aremu, A. Jeorge, and P. Brazdil, "NaijaSenti: A Nigerian Twitter sentiment corpus for multilingual sentiment analysis," 2022, *arXiv:2201.08277*.

[106] J. Prakash and A. A. S. Vijay, "Cross-lingual sentiment analysis of Tamil language using a multi-stage deep learning architecture," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 22, no. 12, pp. 1–28, Dec. 2023, doi: 10.1145/3631391.

[107] H. A. Rasool, F. Abedi, A. G. Ismaeel, A. H. Abbas, R. Khalid, A. Alkhayyat, M. M. Jaber, and A. Garg, "Pelican optimization algorithm with deep learning for aspect based sentiment analysis on Asian low resource languages," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 22, no. 7, Jul. 2023, doi: 10.1145/3608949.

[108] P. Dominic, N. Purushothaman, A. S. A. Kumar, A. Prabagaran, and J. A. Blessy, "Multilingual sentiment analysis using deep-learning architectures," in *Proc. 5th Int. Conf. Smart Syst. Inventive Technol. (ICSSIT)*, Jan. 2023, pp. 1077–1083, doi: 10.1109/ICSSIT55814.2023.10060993.

[109] H. R. Utomo and A. Romadhony, "Sentiment analysis on Indonesia–English code-mixed data," in *Proc. IEEE 8th Int. Conf. Converg. Technol. (I2CT)*, Apr. 2023, pp. 1–6, doi: 10.1109/I2CT57861.2023.10126234.

[110] D. Van Thin, D. N. Hao, and N. L.-T. Nguyen, "Vietnamese sentiment analysis: An overview and comparative study of fine-tuning pretrained language models," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 22, no. 6, pp. 1–27, Jun. 2023, doi: 10.1145/3589131.

[111] Q.-T. Vo, P. Tran, and T. Tran, "Sentiment analysis for a low-resource language: A study on a Vietnamese University," *ICIC Exp. Lett.*, vol. 14, pp. 1115–1124, Jan. 2023, doi: 10.24507/icicelb.14.11.1115.

[112] A. Yusuf, A. Sarlan, K. U. Danyaro, and A. S. B. A. Rahman, "Fine-tuning multilingual transformers for Hausa–English sentiment analysis," in *Proc. 13th Int. Conf. Inf. Technol. Asia (CITA)*, Aug. 2023, pp. 13–18, doi: 10.1109/cita58204.2023.10262742.

[113] H. Najiha and A. Romadhony, "Sentiment analysis on indonesian-sundanese code-mixed data," in *Proc. IEEE 8th Int. Conf. Converg. Technol. (I2CT)*, Apr. 2023, pp. 1–7, doi: 10.1109/I2CT57861.2023.10126254.

[114] P. K. Roy, "Deep ensemble network for sentiment analysis in bi-lingual low-resource languages," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 23, no. 1, pp. 1–16, Jan. 2024, doi: 10.1145/3600229.

[115] J. T. H. Kong, F. H. Juwono, I. Y. Ngu, I. G. D. Nugraha, Y. Maraden, and W. K. Wong, "A mixed Malay–English language COVID-19 Twitter dataset: A sentiment analysis," *Big Data Cognit. Comput.*, vol. 7, no. 2, p. 61, Mar. 2023. [Online]. Available: https://www.mdpi.com/2504-2289/7/2/61

[116] A. Patil, V. Patwardhan, A. Phaltankar, G. Takawane, and R. Joshi, "Comparative study of pre-trained BERT models for code-mixed hindi-english data," in *Proc. IEEE 8th Int. Conf. Converg. Technol. (I2CT)*, Apr. 2023, pp. 1–7, doi: 10.1109/I2CT57861.2023.10126273.

[117] M. R. Ashraf, Y. Jana, Q. Umer, M. A. Jaffar, S. Chung, and W. Y. Ramay, "BERT-based sentiment analysis for low-resourced languages: A case study of Urdu language," *IEEE Access*, vol. 11, pp. 110245–110259, 2023, doi: 10.1109/ACCESS.2023.3322101.

[118] P. K. Jain, R. Pamula, and G. Srivastava, "A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews," *Comput. Sci. Rev.*, vol. 41, Aug. 2021, Art. no. 100413, doi: 10.1016/j.cosrev.2021.100413.

[119] K. S. S. Varma, P. Sathineni, and R. Mamidi, "Sentiment analysis in code-mixed Telugu–English text with unsupervised data normalization," in *Proc. Conf. Recent Adv. Natural Lang. Process.-Deep Learn. Natural Lang. Process. Methods Appl.*, 2021, pp. 753–760.

[120] A. M. Barik, R. Mahendra, and M. Adriani, "Normalization of indonesian-english code-mixed Twitter data," in *Proc. 5th Workshop Noisy User-generated Text (W-NUT)*, 2019, pp. 417–424.

[121] A. F. Hidayatullah, A. Qazi, D. T. C. Lai, and R. A. Apong, "A systematic review on language identification of code-mixed text: Techniques, data availability, challenges, and framework development," *IEEE Access*, vol. 10, pp. 122812–122831, 2022, doi: 10.1109/ACCESS.2022.3223703.

[122] F. Djatmiko, R. Ferdiana, and M. Faris, "A review of sentiment analysis for non-English language," in *Proc. Int. Conf. Artif. Intell. Inf. Technol. (ICAIIT)*, Mar. 2019, pp. 448–451, doi: 10.1109/ICAIIT.2019.8834552.

[123] A. Fesseha, S. Xiong, E. D. Emiru, M. Diallo, and A. Dahou, "Text classification based on convolutional neural networks and word embedding for low-resource languages: Tigrinya," *Information*, vol. 12, no. 2, p. 52, Jan. 2021. [Online]. Available: https://www.mdpi.com/2078-2489/12/2/52

[124] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," 2016, *arXiv:1607.01759*.

[125] M. A. Fauzi, "Word2Vec model for sentiment analysis of product reviews in Indonesian language," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 1, p. 525, Feb. 2019.

[126] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.

[127] Y. K. Lal, V. Kumar, M. Dhar, M. Shrivastava, and P. Koehn, "De-mixing sentiment from code-mixed text," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics, Student Res. Workshop*, 2019, pp. 371–377, doi: 10.18653/v1/p19-2052.

[128] A. Ekbal, P. Bhattacharyya, T. Saha, A. Kumar, and S. Srivastava, "HindiMD: A multi-domain corpora for low-resource sentiment analysis," in *Proc. 13th Lang. Resour. Eval. Conf.*, 2022, pp. 7061–7070.

[129] L. S. Meetei, T. D. Singh, S. K. Borgohain, and S. Bandyopadhyay, "Low resource language specific pre-processing and features for sentiment analysis task," *Lang. Resour. Eval.*, vol. 55, no. 4, pp. 947–969, Dec. 2021, doi: 10.1007/s10579-021-09541-9.

[130] R. Srinivasan and C. N. Subalalitha, "Sentimental analysis from imbalanced code-mixed data using machine learning approaches," *Distrib. Parallel Databases*, vol. 41, no. 1, pp. 37–52, Mar. 2021, doi: 10.1007/s10619-021-07331-4.

[131] A. F. Hidayatullah, R. A. Apong, D. T. C. Lai, and A. Qazi, "Corpus creation and language identification for code-mixed Indonesian–Javanese–English tweets," *PeerJ Comput. Sci.*, vol. 9, p. e1312, Jun. 2023, doi: 10.7717/peerj-cs.1312.

[132] Z. Rasekh Eslami and S. Zohoor, "Second language (L2) pragmatics and computer assisted language learning (CALL)," *Technol. Assist. Lang. Educ.*, vol. 1, no. 3, pp. 1–17, 2023.

[133] T. Pires, E. Schlinger, and D. Garrette, "How multilingual is multilingual BERT?" in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 4996–5001.

[134] A. Câmara, N. Taneja, T. Azad, E. Allaway, and R. Zemel, "Mapping the multilingual margins: Intersectional biases of sentiment analysis systems in English, Spanish, and Arabic," 2022, *arXiv:2204.03558*.

[135] V. Batanović, M. Cvetanović, and B. Nikolić, "A versatile framework for resource-limited sentiment articulation, annotation, and analysis of short texts," *PLoS ONE*, vol. 15, no. 11, Nov. 2020, Art. no. e0242050.

[136] B. Samanta, N. Ganguly, and S. Chakrabarti, "Improved sentiment detection via label transfer from monolingual to synthetic code-switched text," 2019, *arXiv:1906.05725*.

[137] S. E. Taher and M. Shamsfard, "Adversarial weakly supervised domain adaptation for few shot sentiment analysis," in *Proc. 7th Int. Conf. Web Res. (ICWR)*, May 2021, pp. 119–124, doi: 10.1109/ICWR51868.2021.9443023.

[138] M. Neves and J. Ševa, "An extensive review of tools for manual annotation of documents," *Briefings Bioinf.*, vol. 22, no. 1, pp. 146–163, Jan. 2021, doi: 10.1093/bib/bbz130.

[139] Q. Hou, M. Han, and Z. Cai, "Survey on data analysis in social media: A practical application aspect," *Big Data Mining Anal.*, vol. 3, no. 4, pp. 259–279, Dec. 2020, doi: 10.26599/BDMA.2020.9020006.

[140] M. N. Hussain, A. Obadimu, K. K. Bandeli, M. Nooman, S. Al-Khateeb, and N. Agarwal, "A framework for blog data collection: Challenges and opportunities," in *Proc. IARIA Int. Symp. Designing, Validating, Using Datasets*, 2017, pp. 35–40.

[141] O. Ojo, A. Gelbukh, H. Calvo, A. Feldman, O. Adebanji, and J. Armenta-Segura, "Language identification at the word level in code-mixed texts using character sequence and word embedding," in *Proc. 19th Int. Conf. Natural Lang. Process. (ICON)*, 2022, pp. 1–6.

[142] S. Pradha, M. N. Halgamuge, and N. Tran Quoc Vinh, "Effective text data preprocessing technique for sentiment analysis in social media data," in *Proc. 11th Int. Conf. Knowl. Syst. Eng. (KSE)*, Oct. 2019, pp. 1–8, doi: 10.1109/KSE.2019.8919368.

[143] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data," *Frontiers Energy Res.*, vol. 9, Mar. 2021, Art. no. 652801.

[144] K. Berahmand, F. Daneshfar, E. S. Salehi, Y. Li, and Y. Xu, "Autoencoders and their applications in machine learning: A survey," *Artif. Intell. Rev.*, vol. 57, no. 2, p. 28, Feb. 2024.

[145] M. D'Orazio, E. Di Giuseppe, and G. Bernardini, "Automatic detection of maintenance requests: Comparison of human manual annotation and sentiment analysis techniques," *Autom. Construct.*, vol. 134, Feb. 2022, Art. no. 104068, doi: 10.1016/j.autcon.2021.104068.

[146] R. Alahmary and H. Al-Dossari, "A semiautomatic annotation approach for sentiment analysis," *J. Inf. Sci.*, vol. 49, no. 2, pp. 398–410, Apr. 2023, doi: 10.1177/01655515211006594.

[147] M. A. Qureshi, M. Asif, M. F. Hassan, G. Mustafa, M. K. Ehsan, A. Ali, and U. Sajid, "A novel auto-annotation technique for aspect level sentiment analysis," *Comput., Mater. Continua*, vol. 70, no. 3, pp. 4987–5004, 2022.

[148] C. Wu, F. Wu, S. Wu, Z. Yuan, J. Liu, and Y. Huang, "Semi-supervised dimensional sentiment analysis with variational autoencoder," *Knowl.-Based Syst.*, vol. 165, pp. 30–39, Feb. 2019, doi: 10.1016/j.knosys.2018.11.018.

[149] S. Tabinda Kokab, S. Asghar, and S. Naz, "Transformer-based deep learning models for the sentiment analysis of social media data," *Array*, vol. 14, Jul. 2022, Art. no. 100157.

YUSUF ALIYU received the B.Tech. degree in computer science from Abubakar Tafawa Balewa University, Nigeria, in 2018. He is currently pursuing the M.Sc. degree in information technology research with Universiti Teknologi PETRONAS, focusing on deep learning and NLP. He is skilled in data science, machine learning, deep learning, computer networks, and web development. He holds certifications in deep learning, NLP, usable security, and advanced web programming. He is an Academic Contributor with papers in sentiment analysis areas, dedicated to advancing knowledge in computer science and data science.

ALIZA SARLAN (Member, IEEE) received the bachelor's degree in IT from Universiti Utara Malaysia, in 1996, the master's degree in IT from Queensland University, Australia, in 2002, and the Ph.D. degree in IT from Universiti Teknologi PETRONAS (UTP), Malaysia, in 2015.

From 2019 to 2023, she was the Chair of the Computer and Information Sciences Department, UTP, and currently holds the position of Chair of the Centre for Foundation Studies (CFS-UTP). She is also a Researcher with the Center of Data Science (CeRDaS), UTP, focusing on solving complex challenges in the upstream oil and gas industry using computer science. Additionally, she conducts professional training in data analytics for the public and industry. Her research interests include human behavior and technology adoption, information system analysis and modeling, software reliability, and software testing. She has made significant contributions to the fields of human factors in software development and technology's impact on people's lives. Since 2018, she has been a member of Malaysia Board of Technologists (MBOT).

KAMALUDDEEN USMAN DANYARO (Member, IEEE) received the bachelor's degree in mathematics from Bayero University, Nigeria, the master's degree in business information technology from Northumbria University, U.K., and the Ph.D. degree from Universiti Teknologi PETRONAS (UTP), Malaysia.

He is currently a Lecturer with UTP. He is also a Researcher with the Center for Research in Data Science, UTP. Previously, he was a Postdoctoral Researcher and a Data Engineer with the Offshore Engineering Centre, UTP. With over six years of industry experience and more than seven years of research and academic experience, he is recognized as a MikroTik Certified Network Associate (MTCNA) and a MikroTik Certified Routing Engineer (MTCRE). He is affiliated with professional organizations, such as IEEE, ACM, and AIS. He has an extensive publication record with over 50 peer-reviewed journals and conference papers and serves as a reviewer for numerous conferences and journals. His work primarily revolves around semantic web and data science. He has also secured multiple academic and industrial research grants. He actively promotes collaborations between industry and academia, particularly in the oil and gas sector. His research interests include data science, computer networking, and blockchain.

ABDULLAHI SANI B. A. RAHMAN received the degree in informatique (industrial systems) from the University of La Rochelle, France, in 1995, and the master's degree in computer science (distributed computing) from Universiti Putra Malaysia. Currently, he is a Lecturer with Universiti Teknologi PETRONAS, Malaysia, where he is also a member of the Institute of Autonomous Systems. He is also a Registered Professional Technologist. His research interests include cybersecurity, data analytics, and machine learning.

MUJAHEED ABDULLAHI received the B.Sc. degree in information technology from Infrastructure University Kuala Lumpur, Malaysia, in 2018, and the M.Sc. degree in information technology from Universiti Teknologi PETRONAS, Malaysia, in 2022, where he is currently pursuing the Ph.D. degree in information technology. He was a Graduate Research Assistant. His research interests include machine learning, data analytics, cybersecurity, and the Internet of Things.