Cross-Lingual Sentiment Analysis in Low-Resource Languages: A Recent Review on Tasks, Methods and Challenges

Nor Zakiah Lamin[1]*, Azwa Abdul Aziz[2]
Faculty of Computing and Multimedia, Universiti Poly-Tech Malaysia, Cheras, Kuala Lumpur, Malaysia[1]
Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UniSZA),
Tembila Campus, 22200 Besut, Terengganu, Malaysia[1, 2]
Data Science & Analytics (DASA)-Special Interest Group, Universiti Sultan Zainal Abidin (UniSZA),
21300 Kuala Nerus, Terengganu, Malaysia[2]

***Abstract*—Cross-lingual sentiment analysis (CLSA) has become increasingly important in natural language processing and machine learning, enabling the understanding of opinions across diverse linguistic communities, particularly in low-resource languages (LRLs). Despite growing attention, persistent challenges such as limited annotated data, semantic misalignment, and cultural variation in sentiment expression continue to hinder progress. This systematic literature review (SLR) examines recent developments by analyzing the tasks, methods, and challenges reported in CLSA studies focused on LRLs. Following the PRISMA 2020 framework, a comprehensive search was conducted across major databases, including Scopus, IEEE Xplore, SpringerLink, Elsevier, and Google Scholar, covering studies published between 2021 and 2025. After applying inclusion and exclusion criteria, 27 studies were selected for analysis. The findings reveal that while polarity detection remains the dominant sentiment analysis task, emerging directions such as aspect-based sentiment analysis (ABSA), emotion detection, and hate speech recognition are gaining traction. Methodologically, most studies rely on multilingual pre-trained language models (PLMs), supplemented by machine translation, transfer learning, few-shot learning, and hybrid approaches. However, key challenges remain, including the scarcity of high-quality datasets, instability of few-shot performance, difficulties in handling dialectal variation, bias in PLMs, and the lack of standardized evaluation benchmarks. This review concludes by emphasizing the need for more culturally grounded tasks, adaptive hybrid frameworks, and fairness-aware evaluation practices to build robust cross-lingual frameworks and richer linguistic resources for underrepresented languages.**

Keywords*—Cross-lingual sentiment analysis; low-resource language; natural language processing; pre-trained language models; transfer learning; few-shot learning

I. Introduction

The exponential growth of user-generated content across social media platforms, review sites, and online forums has underscored the significance of sentiment analysis in natural language processing (NLP) and machine learning. As a computational approach for identifying and categorizing opinions expressed in text, sentiment analysis plays a pivotal role in gauging public attitudes, tracking brand reputation, and informing decision-making across diverse domains.

While substantial progress has been made in sentiment analysis for resource-rich languages such as English and Chinese, extending these capabilities to a broader set of languages remains a considerable challenge. Many of the world's languages are classified as low-resource, characterized by the lack of large-scale annotated datasets, robust linguistic tools, and standardized benchmarks. This limitation hampers the development of effective sentiment analysis systems for these languages, ultimately leading to their underrepresentation in NLP applications.

Cross-lingual sentiment analysis has emerged as a promising strategy to address this disparity by leveraging knowledge from high-resource languages to improve performance in low-resource settings. Several comprehensive reviews have outlined the evolution of cross-lingual sentiment analysis and its methodological progression from translation-based models to multilingual PLMs. Techniques such as machine translation, multilingual embeddings, and transfer learning have facilitated significant advances in this area. Few-shot and zero-shot learning have recently become popular for addressing limited annotated data. However, challenges persist, particularly in managing dialectal diversity, preserving sentiment polarity across languages, and ensuring the generalizability of proposed models.

Given the growing body of research and the critical need to develop inclusive language technologies, this study conducts a systematic review of cross-lingual sentiment analysis approaches focused on low-resource languages. Specifically, it aims to examine the primary tasks addressed, the methods employed, and the key challenges reported in recent literature. By consolidating these insights, this review seeks to provide a comprehensive understanding of the current landscape and to identify avenues for future research that can advance sentiment analysis in underrepresented language contexts.

II. Basic Terminology

First, it would be useful to define the key terminology related to this research review focus:

*Corresponding author.

A. Sentiment Analysis

The increase in user-generated content on the web has transformed the online platforms into rich repositories of public opinion. People growingly use social media, review sites, and forums to share views on products, services, social issues, trends, and government policies. The surge of textual data reflecting "what people think" has become a critical source of information.

For businesses, the ability to manage in terms of identifying and analyzing these opinions is crucial for understanding customer needs, enhancing satisfaction, and informing market-driven product design. For this reason, sentiment analysis, which involves the computational study of opinions, sentiments, and emotions expressed in text, has emerged as a critical area within natural language processing and data analytics.

B. Cross-Lingual

In the context of natural language processing, cross-lingual approaches refer to methods that enable the transfer of knowledge or models across different languages. These techniques focus on leveraging data-rich languages to improve performance in languages with limited resources. Cross-lingual methods are especially valuable for tasks like sentiment analysis, where labeled data may be abundant in one language but scarce or not available in another languages. By leveraging shared linguistic representations, machine translation, or multilingual embeddings, cross-lingual models seek to generalize sentiment understanding across diverse linguistic contexts.

C. Low-Resource

Low-resource languages are languages that lack extensive computational resources, such as large, annotated datasets, linguistic tools, and benchmark datasets that are readily available for high-resource languages like English or Chinese. This shortage poses significant challenges for developing effective NLP applications, including sentiment analysis. In most cases, these languages also encompass multiple dialects and informal registers, adding further complexity. Therefore, the need for low-resource languages is critical to support more inclusive language technologies and ensure that advances in NLP will benefit a broader spectrum of linguistic communities.

III. Systematic Literature Review

The objective of this review is to systematically examine the recent developments in cross-lingual sentiment analysis within the context of low-resource languages. This study seeks to identify the main tasks addressed, analyze the methods and techniques employed, and highlight the key challenges and research gaps reported in the literature. To achieve this, the review is guided by the following research questions:

Research Question 1 (RQ1):

What are the primary sentiment analysis tasks explored in cross-lingual studies involving low-resource languages?

Research Question 2 (RQ2):

What methods and techniques have been employed to perform cross-lingual sentiment analysis in low-resource language settings?

Research Question 3 (RQ3):

What key challenges and research gaps are identified in existing studies on cross-lingual sentiment analysis for low-resource languages?

A. Search Process

To ensure comprehensive coverage of relevant studies, a structured search was conducted across five major scientific databases, which are Scopus, Elsevier's ScienceDirect, IEEE Xplore, SpringerLink, and Google Scholar, as shown in Fig. 1. These platforms were selected for their extensive indexing of high-quality journals and conference proceedings in natural language processing, machine learning, and computational linguistics.

The search employed combinations of key terms (Table I). The review was limited to publications in English, covering the period from 2021 to 2025, to capture recent advancements in the field. In addition, Google Scholar was used to identify further studies through broader keyword searches and citation tracking, ensuring that potentially relevant grey literature and seminal works were not overlooked. Reference lists of selected articles were also manually screened to identify additional studies relevant to this review.

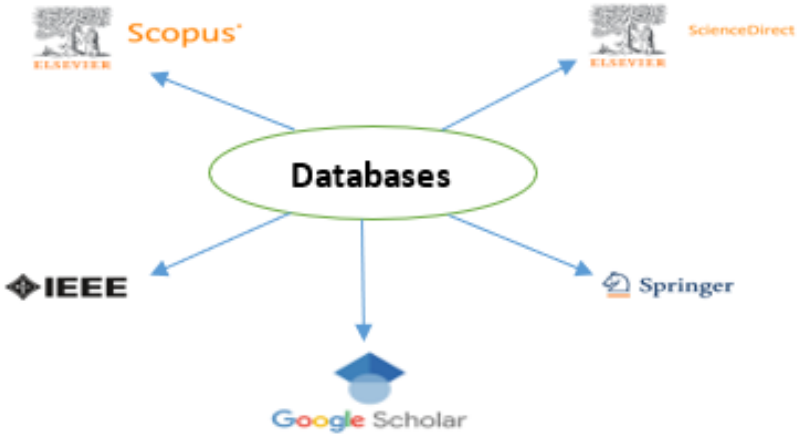

Fig. 1. Database searched.

TABLE I. TABLE OF SEARCH STRING

Database	Search Strings
Scopus	("sentiment analysis" OR "opinion mining" OR "emotion detection") AND ("cross-lingual" OR "multilingual") AND ("low-resource language" OR "under-resourced" OR "resource-scarce") AND ("few-shot learning" OR "zero-shot learning" OR "transfer learning" OR "multilingual embeddings" OR "pre-trained model") AND ("sentiment classification" OR "aspect-based sentiment analysis" OR "challenges" OR "limitations") AND PUBYEAR > 2020 AND PUBYEAR < 2026 AND PUBYEAR > 2020 AND PUBYEAR < 2026 AND (LIMIT-TO (SUBJAREA , "COMP")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (EXACTKEYWORD , "Sentiment Analysis") OR LIMIT-TO (EXACTKEYWORD , "Low Resource Languages")) AND (LIMIT-TO (SRCTYPE , "j")) Date of Access: July 2025
IEEE Xplore	("sentiment analysis" OR "opinion mining" OR "emotion detection") AND ("cross-lingual" OR "multilingual") AND ("low-resource language" OR "under-resourced" OR "resource-scarce") AND ("few-shot learning" OR "zero-shot learning" OR "transfer learning" OR "multilingual embeddings" OR "pre-trained model") AND ("sentiment classification" OR "aspect-based sentiment analysis" OR "challenges" OR "limitations") AND PUBYEAR > 2020 AND PUBYEAR < 2026 AND (LIMIT-TO (SUBJAREA , "COMP")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (EXACTKEYWORD , "Sentiment Analysis") OR LIMIT-TO (EXACTKEYWORD , "Low Resource Languages")) **Date of Access:** July 2025
SpringerLink	("sentiment analysis" OR "opinion mining" OR "emotion detection") AND ("cross-lingual" OR "multilingual") AND ("low-resource language" OR "under-resourced" OR "resource-scarce") AND ("few-shot learning" OR "zero-shot learning" OR "transfer learning" OR "multilingual embeddings" OR "pre-trained model") AND ("sentiment classification" OR "aspect-based sentiment analysis" OR "challenges" OR "limitations") Date of Access: July 2025
ScienceDirect	"cross-lingual" AND "sentiment analysis" AND ("challenges" OR "limitations" OR "gaps") AND "low-resource languages" AND "transfer learning" AND "few-shot learning" Date of Access: July 2025
Google Scholar	"cross-lingual" AND "sentiment analysis" AND ("challenges" OR "limitations" OR "gaps") AND "low-resource languages" AND "transfer learning" AND "few-shot learning" Date of Access: July 2025

B. Inclusion and Exclusion Criteria

There is a lot of literature reviews on sentiment analysis. Therefore, to ensure that the search would be manageable and focused, researchers defined some inclusion and exclusion criteria to select the papers for review as follows:

1) Inclusion criteria: Studies published during the period 2021-2025 related to cross-lingual sentiment analysis and studies on tasks, methods, and challenges related to cross-lingual sentiment analysis. The studies focused on low-resource languages instead of high-resource languages. If studies had been published in more than one journal or conference proceedings, researchers chose the most complete version for inclusion. The selection is peer-reviewed articles and written only in English.

2) Exclusion criteria: Studies exclude informal studies (unknown conferences or journals), papers that are irrelevant to the above research questions, non-sentiment analysis in low-resource languages, and papers not written in English.

3) Screening and selection: By using the PRISMA framework, duplicates were removed, and the remaining titles and abstracts were screened to find the relevant papers. Full-text screening was conducted for shortlisted papers. 152 excluded for irrelevance. Therefore, a total of 27 papers were included in the final analysis.

4) Assessment for eligibility: At this stage, 169 full-text articles were assessed for eligibility, and 142 were excluded (not CLSA or not low-resource).

5) Data analysed and included: The 27 papers that were selected for more detailed study are summarized in Table III. This table shows the main information extracted from the selected papers, which are presented in order of the most recent year of publication.

IV. BIBLIOGRAPHY MANAGEMENT AND DOCUMENT RETRIEVAL

Researchers used Mendeley Desktop 2.135.0 to manage all the bibliographic details and citations. The studies that were identified by the above table of search process were scanned by title and abstract according to the inclusion and exclusion criteria. Then, all the papers that were identified as relevant to this research were then downloaded for data extraction and further study.

Table II provides details on the number of research studies that were discovered by the search of the databases in Fig. 1. The search process is illustrated in the form of the PRISMA framework flowchart in Fig. 2. The revised PRISMA 2020 flow diagram (see Fig. 2) details the identification, screening, eligibility, and inclusion process of the 27 selected studies. The search and selection process adhered to the PRISMA 2020 guideline for systematic reviews [28].

TABLE II. NUMBER OF RESEARCH STUDIES IDENTIFIED

Database	Number of papers found		
	Based on key terms (search strings)	Based on title	Based on abstract
Scopus	211	111	95
IEEE Xplore	14	11	11
SpringerLink	63	19	21
ScienceDirect	29	25	9
Google Scholar	406	120	34
Total	723	321	169

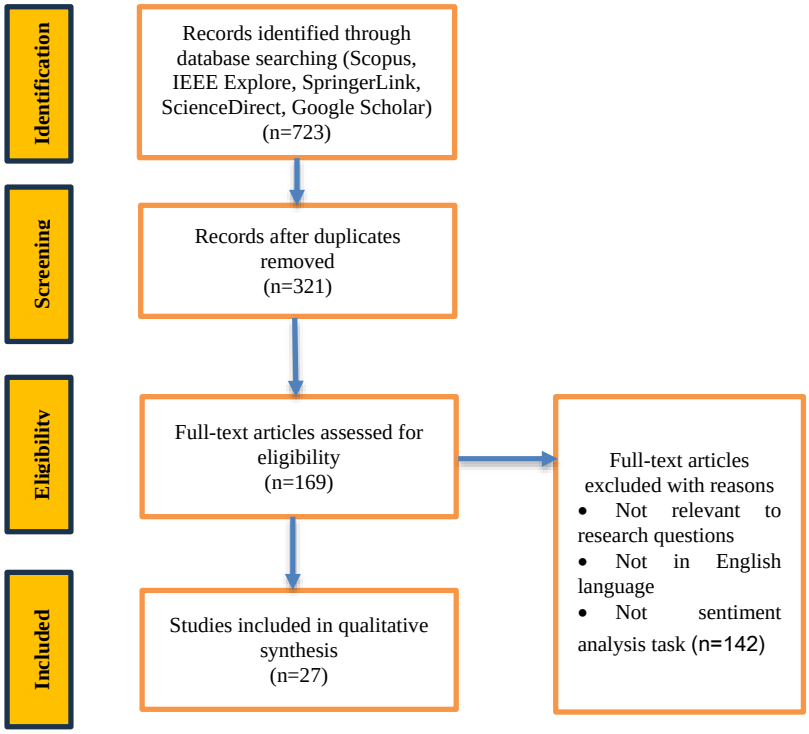

Fig. 2. Revised PRISMA 2020 flow diagram for study selection process.

V. Results and Findings

This systematic literature review analyzed 27 selected studies published between 2021 and 2025 that address sentiment analysis in cross-lingual and low-resource languages contexts. The studies encompass a range of languages, domain such as social media, app reviews, multilingual corpora, and technical approaches. The primary focus is in identifying the sentiment analysis tasks explored (RQ1), the methods and models adopted (RQ2), and the key challenges addressed (RQ3). Each study was assigned a unique identifier and categorized accordingly. Table III shows the number of studies identified.

TABLE III. Number of Research Studies Identified

No	Publication	Tasks / Methods	Key
1	Enhancing cross-lingual hate speech detection through contrastive and adversarial learning [2]	Tasks involved were cross-lingual hate speech detection, zero-shot and few-shot transfer learning and language transfer evaluation.	P1
		Methods used were contrastive learning where the model is trained using contrastive loss to bring semantically similar samples closer in representation space, even across different languages. It also use translation-based augmentation where text from low-resource languages is machine-translated to high resource languages such as English to utilize such annotated datasets. Multilingual Transformer Backbone utilizing XLM-R-RoBERTa as the core model and pretrained on multiple languages. It also used Dual-Encoder Setup for contrastive training and do evaluation benchmarks.	
2	Transformer-Based Abstractive Summarization of Legal Texts in Low-Resource Languages [3]	Abstractive summarization of legal documents in low-resource Indian languages (Hindi, Bengali, Telugu, Tamil, and Marathi) using cross-lingual transfer learning.	P2
		Fine-tuning of mBART and mT5 transformer models pre-trained on multilingual data; use of zero-shot and few-shot learning; translation of low-resource text into English; back-translation to maintain legal semantics.	
3	TFT-TL: Token-Level Filter Training Transfer Learning for Low-Resource Neural Machine Translation [4]	This study used Neural Machine Translation (NMT) for low-resource languages, cross-lingual transfer learning, token-level quality control in knowledge transfer and sequence prediction alignment	P3
		The method involved: ➢ Token-Level Filter Training (TFT): Filters out *unstable token predictions* from the parent model to prevent noise during transfer. ➢ Hierarchical Ranking Loss: A loss function that helps the child model better capture sequence-level dependencies and mimic the parent model's prediction order. ➢ Transfer learning framework using parameter initialization from a high-resource parent model, fine-tuning on low-resource child language and soft label supervision for consistency learning.	
4	Bridging resource gaps in cross-lingual sentiment analysis: adaptive self-	This study involved were cross-Lingual Sentiment Analysis (CLSA) that Transfers sentiment classification models trained in a high-resource source language (e.g., English) to low-resource target languages, Transfer Learning, that utilize pre-trained multilingual models to adapt sentiment knowledge from source to target	P4

	alignment with data augmentation and transfer learning [5]	languages, data augmentation to apply techniques like back-translation and synonym replacement to enhance training data diversity. Finally, performance evaluation conducted to test the model on benchmark multilingual sentiment datasets and comparing it to baseline methods. For the method, Adaptive Transfer Learning Framework is used to do fine-tuning of multilingual transformer models (e.g., mBERT, XLM-R) on labelled source language data. The augmentation techniques used back-translation to generate paraphrases and synonym replacement to create sentence variants while preserving sentiment, The training strategies applied to train the model and do evaluation metrics.	
5	A hybrid model for low-resource language text classification and comparative analysis [6]	This study tasks are on text classification for low-resource language (Malay) that focused on sentiment or review classification using a newly annotated Malay dataset, do comparative evaluation of existing tools on a low-resource dataset to assess performance gaps. There are also, Hybrid Model Development that integrates linguistic rules and transfer learning. Empirical Evaluation were performed using accuracy, F1-score, and statistical testing (paired t-test). The method involved **Tools, Data, Modelling and Evaluation**. ➢ Tools - LangDetect, spaCy, FastText, XLM-RoBERTa, and LLaMA ➢ Data - Dataset of 74,931 user reviews from apps like MyBayar, PDRM, MyJPJ, MySejahtera. A subset of 2,621 reviews was annotated manually with inter-coder reliability (Fleiss' Kappa). ➢ Modeling - Hybrid approach combining rule-based features and transfer learning. ➢ Evaluation - Accuracy: 84% and Statistical significance; $p < 0.05$ using paired t-tests on F1-scores.	P5
6	Sentiment Analysis and Emotion Detection Using Transformer Models in Multilingual Social Media Data [7]	Sentiment Classification – Classify tweets into sentiment categories (positive, neutral, negative). Emotion Detection – Identify emotions such as joy, anger, sadness, and fear from social media content. Use of pretrained transformer models: BERT, RoBERTa, and XLM-R. Application of fine-tuning on two public multilingual datasets containing tweets annotated with sentiment and emotion labels. Comparison with baseline machine learning models (e.g., SVM and LSTM).	P6
7	The Impact of Linguistic Variations on Emotion Detection: A Study of Regionally Specific Synthetic Datasets [8]	The main tasks in this paper are: Emotion Detection, to identify emotional categories (e.g., joy, sadness, anger) from social media texts, evaluation of linguistic variation to assess how regional language variations impact emotion detection performance and synthetic data generation to create regionally specific synthetic datasets for emotion classification tasks. The methods used include: ➢ Synthetic Dataset Construction: A regionally specific emotion dataset was created by generating synthetic variants of tweets from the SemEval 2018 Task 1 dataset. These variants mimic regional dialects and linguistic styles across English-speaking countries. ➢ Transformer-based Models: Multiple pre-trained transformer models (including BERT, RoBERTa, and DeBERTa) were fine-tuned and evaluated on both original and synthetic data. ➢ Cross-Dataset Evaluation: The models were tested on both the original and regionally modified datasets to observe performance changes.	P7
8	Exploring Transfer Learning in a Bidirectional Myanmar-Tedim Chin Machine Translation with the mT5 Transformer [9]	This study did data collection & corpus creation, Text Preprocessing, Model Fine-tuning (Transfer Learning), Bidirectional Machine Translation, and Model Evaluation. It creates dataset which a 10K parallel corpus from the UCSY dataset, covering diverse domains such as, Greetings, Educational stories, Communication, Leisure (e.g., travel, shopping), Resource-related texts (e.g., numbers, time). Then, preprocessing of data was segmented using Syllable-level segmentation and Word-level segmentation. Then, Model is Fine-tuned the mT5 model on both segmentation types and used BLEU score and accuracy to evaluate translation performance.	P8
9	Enhancing Low-Resource Question-Answering Performance Through Word Seeding and Customized Refinement [10]	This study to enhance Low-resource Question Answering (QA) and improve QA performance in low-resource languages (e.g., Hindi) through cross-lingual approaches. The method used Word Seeding (POS-aware noun substitution) to replace English nouns with translated/transliterated Hindi nouns to create bilingual training data, realign answer spans after word substitutions using n-gram similarity and SequenceMatcher. It applied three-stage Transfer Learning: 1. Pretrain on SQuAD 2. Intermediate MLM on Hindi corpus 3. Fine-tune on constructed bilingual QA dataset The models used mBERT and XLM-R (base & large) The evaluation are on MLQA and XQuAD datasets.	P9
10	An Empirical Evaluation of the Zero-Shot, Few-Shot, and Traditional Fine-Tuning Based Pretrained Language Models for Sentiment Analysis in Software Engineering [11]	This study evaluates sentiment classification performance on SE datasets, focusing on developer comments and discussions. It compares zero-shot, few-shot, and traditional fine-tuning methods using PLMs on sentiment classification tasks. It used pre-trained Language Models (PLMs), RoBERTa, BERT, XLNet, GPT-2, BART, T5. As for the learning settings, it applies Zero-Shot Learning, Few-Shot Learning and Traditional Fine-Tuning. For the evaluation, it used five benchmark SE sentiment datasets (e.g., Jira, Stack Overflow). Applied classification accuracy, macro-F1, and weighted-F1 as performance metrics.	P10
11	Motamot: A Dataset for Revealing the Supremacy of Large Language	This study about political sentiment analysis in Bengali language (during election period) and evaluate its performance of various Pretrained Language Models (PLMs) and Large Language Models (LLMs) in sentiment detection. It benchmark political sentiment classification using a newly created dataset.	P11

	Models Over Transformer Models in Bengali Political Sentiment Analysis [1]	It introduced Motamot dataset: 7,058 annotated instances (positive/negative), collected from online newspaper articles in Bangladesh and used certain models: → PLMs: BanglaBERT, Bangla BERT Base, XLM-RoBERTa, mBERT, sahajBERT → LLMs: Gemini 1.5 Pro, GPT 3.5 Turbo It applied zero-shot and few-shot learning techniques and used evaluation metrics for the accuracy comparison across all models.	
12	A comparative study of cross-lingual sentiment analysis [12]	This study focuses on zero-shot Cross-lingual Sentiment Classification where it classifies sentiment in target languages (Czech, French) using models trained only in English (no labelled data in the target language). For the baseline, it develops strong monolingual models for comparison with cross-lingual approaches. It also does experimentation with embedding Techniques to study the impact of embedding normalization, vocabulary size, and in-domain vs. general embeddings for linear transformations. The method us Transformer-Based Models. Linear Transformations + CNN/LSTM to align word embeddings between source and target languages. Then, used Large Language Models (LLMs) (ChatGPT (GPT-3.5 Turbo) and LLaMA 2) in zero-shot prompting settings (no fine-tuning) to achieved competitive or superior performance with higher hardware requirements.	P12
13	Sentiment analysis on a low-resource language dataset using multimodal representation learning and cross-lingual transfer learning [13]	This study analyse sentiment from video data using text, audio, and visual modalities using Multimodal Sentiment Analysis (MSA). It performs MSA on Tamil, a low-resource language with limited annotated data. It curate and annotate a new dataset called MSAT (Multimodal Sentiment Analysis corpus in Tamil), consisting of ~300 utterances with sentiment and emotion labels for the dataset creation. To improve sentiment analysis in Tamil, it used cross-lingual transfer learning to transfer knowledge from English MSA datasets (CMU-MOSI, MOSEI, MELD). It applied SPMMAE (Shared-Private Multimodal AutoEncoder) which is a novel deep learning architecture that extracts unimodal embeddings using Self-Supervised Learning (SSL), learns both shared (common) and private (modality-specific) representations and provides a comprehensive fused representation across modalities. For the Cross-lingual Transfer Learning, the study pretrain the model on large English MSA datasets, fine-tune it on the small Tamil MSAT dataset and demonstrated 11% performance improvement on MSAT through transfer	P13
14	Exploring zero-shot and joint training cross-lingual strategies for aspect-based sentiment analysis based on contextualized multilingual language models [14]	Cross-lingual sentiment classification using multilingual product review datasets (e.g., Amazon reviews in multiple languages). ➤ The model is trained on English data and tested on other languages (zero-shot setting). Multilingual BERT (mBERT) and XLM-R pretrained models Fine-tuning on source (English) data No labeled target language data required (zero-shot cross-lingual approach)	P14
15	Data-Augmentation for Bangla-English Code-Mixed Sentiment Analysis: Enhancing Cross Linguistic Contextual Understanding [15]	Sentiment classification of code-mixed Bangla-English social media posts. ➤ Multi-class classification (positive, negative, neutral) Data augmentation techniques: back-translation, synonym replacement, transliteration Pretrained Transformer models: mBERT and IndicBERT Augmented training to enhance cross-lingual contextual understanding	P15
16	Zero-shot learning based cross-lingual sentiment analysis for Sanskrit text with insufficient labeled data [16]	Sentiment Analysis: The core task is to classify text sentiment (positive, negative, neutral). Cross-lingual Transfer: Apply models trained in one language (source) to another (target) without needing training data in the target language. Zero-shot Learning (ZSL): No labeled examples from the target language are used during training; model infers based on learned representations. Transformer-based Models: Used multilingual transformers like mBERT and XLM-R to generate language-agnostic representations. Fine-tuning on English Data: The model is trained on a labeled English sentiment dataset. Evaluation on Other Languages: Performance is evaluated on unseen target languages (e.g., Chinese, Arabic, Spanish). Cross-lingual Embedding Space: Leverage shared representation space of multilingual models to generalize across languages. Benchmarks Used: XNLI and MLDoc datasets, among others, for evaluation.	P16
17	Sentiment analysis of the Algerian social movement inception [17]	Sentiment analysis of tweets related to the Algerian Hirak social movement Comparative evaluation of models on Algerian Arabic, a low-resource language Classical ML models: Naive Bayes, SVM, Logistic Regression (LR), Decision Tree Feature extraction: Bag of Words (BoW), TF-IDF Pretrained Transformers: AraBERT, DziriBERT, XLM-R (fine-tuned) Evaluation metrics: accuracy, precision, recall, F1-score LR performed best with 68% accuracy	P17
18	How a Deep Contextualized	Cross-lingual sentiment analysis for low-resource languages Improve classification performance and provide explainability in sentiment predictions	P18

	Representation and Attention Mechanism Justifies Explainable Cross-Lingual Sentiment Analysis [18]	-Model architecture: → Pretrained cross-lingual transformer: XLM-RoBERTa for contextualized embeddings → Followed by LSTM (Long Short-Term Memory) for sequence modeling → Integrated attention mechanism to identify key informative words and justify polarity - Comparison with monolingual and cross-lingual baselines	
19	Prompt-based for Low-Resource Tibetan Text Classification [19]	Text classification for Tibetan language in low-resource settings. Address the lack of annotated Tibetan data for NLP tasks (including sentiment analysis and information extraction). This study used prompt-based learning to guide language models to perform classification. It leverages pre-trained language models on a large-scale unsupervised Tibetan corpus and applies few-shot learning to enable effective performance with minimal labelled data. The approach emphasizes generation-based classification, fitting the prompt-learning paradigm.	P19
20	Fake news detection in Dravidian languages using transfer learning with adaptive finetuning [20]	Fake News Detection in low-resource Dravidian languages (Telugu, Tamil, Kannada, Malayalam). This study introduced Dravidian_Fake, the first fake news dataset for Dravidian languages with 26,000 articles. It combined English ISOT dataset and Dravidian_Fake to form a multilingual corpus. Them used transfer learning with pretrained transformer models: ➢ mBERT (Multilingual BERT) ➢ XLM-RoBERTa It also applied adaptive fine-tuning strategy to better adjust models to the multilingual and low-resource scenario. The, used cross-lingual embeddings to transfer semantic knowledge from English to Dravidian languages.	P20
21	Transfer language selection for zero-shot cross-lingual abusive language detection [21]	Zero-shot Cross-Lingual Abusive Language Detection (ALD): Performing ALD in target languages without labeled training data by transferring knowledge from a source language (e.g., English). Transfer Language Selection: Identifying which source language(s) can best support abusive language detection in a given target language under a zero-shot setting. Evaluation Across Languages: Empirically assessing how the choice of source language affects ALD performance in multiple target languages (e.g., Swahili, Urdu, Arabic). This study used model architecture that utilizes the XLM-R (XLM-RoBERTa) multilingual transformer model for performing abusive language detection in a zero-shot setting. For the transfer evaluation strategy, the method involves training the model on one or more source languages and evaluating it directly on target languages without fine-tuning (zero-shot). The transferability metrics explore language similarity, geographical proximity, and cultural traits to assess and explain why certain transfer languages work better. For the experimental design, dataset and evaluation metrics are as follows: Experimental Design: Source languages: English, Arabic, Hindi, Spanish, etc. Target languages: Swahili, Urdu, and Arabic. Dataset: Uses annotated datasets (e.g., HateXplain, OLID, etc.) for source languages and tests performance on target language data. Evaluation Metrics: Primarily uses macro-F1 score and accuracy to evaluate zero-shot performance.	P21
22	Burmese Sentiment Analysis Based on Transfer Learning [22]	Sentiment Analysis of Burmese (Myanmar) texts, which is a low-resource language. Specifically, the task is to classify Burmese texts into sentiment categories (positive, negative, neutral) The study utilizes transfer learning techniques by leveraging pre-trained models such as: ➢ Multilingual BERT (mBERT) ➢ XLM-RoBERTa Fine-tuning is performed using a Burmese sentiment dataset collected and annotated manually. The model performance is compared with traditional machine learning methods like Naive Bayes, SVM, and LSTM-based deep learning models. The best results are achieved using XLM-RoBERTa, demonstrating the effectiveness of cross-lingual transfer in sentiment analysis	P22
23	Deep Persian sentiment analysis: Cross-lingual	Perform sentiment analysis on Persian, a low-resource language, Improve classification accuracy using transfer learning from English data	P23

	training for low-resource languages [23]	Proposed a cross-lingual deep learning framework using English Persian data. It used cross-lingual word embeddings (static & dynamic) and trained on English Amazon reviews, tested on Persian Digikala reviews. It achieved +22% (static) and +9% (dynamic) performance improvement over monolingual methods. Required only a bilingual dictionary for embedding alignment.	
24	Zero-Shot Emotion Detection for Semi-Supervised Sentiment Analysis Using Sentence Transformers and Ensemble Learning [24]	Address sentiment analysis and emotion detection in low-resource settings. This study proposes a framework for semi-supervised sentiment classification. It enables zero-shot emotion classification without labelled data in the target domain. It utilizes Sentence Transformers (e.g., all-MiniLM-L6-v2) to encode sentences into semantic vector representations. The, implement Zero-shot Learning (ZSL) using cosine similarity between sentence embeddings and predefined emotion/sentiment labels. It uses Ensemble Learning to combine outputs from multiple classifiers (SVM, Logistic Regression, Decision Tree). This study introduces pseudo-labelling (semi-supervised learning) where model-predicted labels are iteratively added to the training set to improve classification.	P24
25	Cross-Lingual Knowledge Transferring by Structural Correspondence and Space Transfer [25]	Cross-Lingual Sentiment Analysis (CLSA) → Transferring sentiment knowledge from a label-rich source language (e.g., English) to a low-resource target language (e.g., Chinese) It used ssSCL-ST (Semi-supervised Structural Correspondence Learning with Space Transfer): ➢ Structural Correspondence Learning (SCL): Learns shared feature representations (called pivots) between languages based on co-occurrence in unlabelled data. ➢ Space Transfer: Maps feature spaces between source and target languages to align semantic spaces, allowing knowledge reuse across languages. ➢ Pivot Set Extension: Enhances the quality of shared features across domains. ➢ Semi-supervised Learning: Leverages unlabelled data in the target language to extract useful domain-specific sentiment signals	P25
26	Sentiment Analysis Using XLM-R Transformer and Zero-shot Transfer Learning on Resource-poor Indian Language [26]	Sentiment analysis for resource-poor Indian languages (Hindi); sentence-level classification of tweets and reviews. It used zero-shot cross-lingual transfer using XLM-RoBERTa; trained on English (SemEval 2017 Task 4A) and evaluated on two Hindi datasets (IITP-Movie and IITP-Product); achieves ~60.93% accuracy.	P26
27	A joint learning approach with knowledge injection for zero-shot cross-lingual hate speech detection [27]	Cross-lingual hate speech detection in low-resource languages using zero-shot learning. This study proposed joint-learning architecture using multilingual models (e.g., LASER, MUSE, Multilingual BERT) with injected external knowledge via HurtLex (a multilingual abusive word lexicon). For the testing, it tested on 6 low-resource languages using English as the source and evaluated multiple neural and machine learning models.	P27

A. Sentiment Analysis Tasks

Addressing RQ1, the corpus highlights seven task families, ranging from coarse–grained polarity detection to fine-grained, safety-critical objectives. The unevenness of this distribution carries significant methodological and practical ramifications.

1) Text classification or polarity detection: A large share of studies (P1, P2, P8, P9, P11, P13, P14, P15) focus on document or sentence level polarity. This prevalence is understandable where polarity tasks are easy to formulate and compare, and they remain useful as baseline capability checks. However, the heavy reliance on polarity implicitly rewards simplified sentiment theories, often missing stance, irony, intensifiers/negation, and code-switching phenomena that are common in low-resource language discourse.

2) Aspect-Based Sentiment Analysis (ABSA): ABSA appears in (P5, P10, P19, P22), targeting feature-level opinion in the example of service versus price. Cross-lingual ABSA in low-resource languages is data definition constrained, where aspect taxonomies are often imported from English and may not map to local discourse. An example of culturally specific service norms or product attributes. This can misrepresent label spaces and decrease transfer performance even the model's capacity is adequate. Therefore, the suggestion is to use community-informed aspect induction (unsupervised or semi-supervised discovery of aspects from in-language corpora), followed by lightweight alignment to any global schema. The evaluation can be done with partial-credit measures, example in hierarchical aspects to prevent artificial penalties on culturally grounded aspect definitions.

3) Emotion detection: Emotion detection in studies (P6, P16, P20) probes beyond polarity, for example, anger and joy. Progress is slowed by annotation scarcity and cultural variation in emotion expression (metaphor, honorifics, and emojis). Porting Western emotion inventories can mislabel affect in low-resource languages. Suggestion is to adopt culturally grounded annotation protocols and combine psychological lexicons + PLM features. It may use few-shot adjudication with bilingual experts to calibrate labels before scaling.

4) Cross-lingual transfer (XLT): Studies in P3, P4, P7, P12 and P18 adapt models from a source (typically English) to a target low-resource language. XLT often yields higher pairwise accuracy but requires per-pair tuning and incurs cost when many targets exist. As a suggestion, do incorporate target-aware adapters or vocabulary augmentation for report transfer efficiency (accuracy gain per labeled/compute unit).

5) Multilingual Sentiment Classification: MSC in studies (P17, P21, P24) trains a single model across many languages. It scales coverage but risks representation interference when typologically distant languages share parameters. It is recommended to adopt adaptive multilingual training which cluster languages by typology/script/morphology, pre-adapt per

cluster, then soft-share layers across clusters. Report per-cluster gains/losses to make interference visible.

6) Few-shot or zero-shot learning: Few or zero-shot methods studies in P23, P25 and P26 reduce dependence on labels. Gains are fragile yet sensitive to prompt wording, verbalizers, and the representativeness of very small support sets. Overclaiming of generalization from one prompt template is common. Therefore, it is suggested that prompt diversity protocols be enforced in multiple paraphrases or verbalizers and report variance across seeds or prompts. Consider its consistency of regularization or meta-learning for stability.

7) Hate speech or abusive content: One study which is P27 targets toxicity detection. This task is ethically high-stakes and linguistically nuanced, such as slurs, reclaimed terms, sarcasm, and code-switching. Importing English toxicity lexicons risks false positives and cultural harm. Therefore, the suggestion is to combine a curated community of lexicons with reclaiming flags plus contextual PLM signals and audit false positive rates on minority dialects.

In summary, the evidence for RQ1 indicates that the field is task diversifying, but cross-lingual sentiment analysis for low-resource languages remains over-indexed on polarity. To avoid distorted progress signals, future work should elevate fine-grained tasks with culturally valid label spaces and robustness of the first evaluation.

B. Methods and Techniques in Cross-Lingual Sentiment Analysis

Addressing RQ2, the findings reveal that methods extend beyond PLM-centric transfer to include translation-based pipelines, contrastive or self-supervised learning, and hybrid strategies that combine neural architectures with linguistic resources, where each offering different trade-offs in scalability and robustness. The strongest studies compose these elements rather than betting on a single paradigm.

1) Pre-Trained Multilingual Language Models (PLMs): PLMs such as mBERT, XLM-R, and MT5 that are used in (P1, P3, P4, P7, P12, P18, P23, P25) serve as the most common baseline for cross-lingual sentiment analysis. These models are trained on massive multilingual corpora and can transfer knowledge across languages without requiring paralled data, making them highly appealing for low-resource settings. They provide quick performance gains by leveraging shared representations between high-resource languages and low-resource languages. However, PLMs also show significant limitations where their performance does not equally benefit high-resource languages that dominate pretraining data. As the benefits degrade sharply for low-resource languages, specifically in those with rich morphology, code switching, or non-Latin scripts. This imbalance raises concerns about fairness and inclusivity. Similar conclusions were drawn in recent works highlighting the critical role of pre-trained language models and their adaptability across linguistic boundaries [29], [30]. To address these limitations, researchers propose bias-aware fine-tuning, which introduces adapter layers for parameter-efficient specialization per language cluster and applying morphology-aware tokenization to reduce subword fragmentation and preserve sentiment-bearing morphemes. These strategies could make PLMs more equitable and effective for underrepresented languages.

2) Few or zero-shot learning: These learning methods in studies (P23, P25) aim to minimize dependence on any annotated datasets by transferring sentiment knowledge across languages with little or no labelled data in the target language. In few-shot settings, the model is provided with only a few labelled examples, while zero-shot learning transfers directly without any labelled target data, that is often relying on prompts or shared representations. These approaches offer label efficiency, making it highly attractive for low-resource languages where annotation is scarce. However, the findings show that performance is highly variable, depending on prompt wording, random seeds, or the representativeness of the small labelled set. This instability makes few-shot and zero-shot methods unreliable for consistent deployment in sensitive domains. To address this, researchers recommend reporting confidence intervals across multiple prompt templates rather than relying on single run results that employ agreement-based exemplar selection to reduce topic bias in few-shot settings and incorporating consistency losses across paraphrases to stabilize output. These improvements would strengthen the reliability of low label transfer methods in cross-lingual sentiment analysis.

3) Translation-based pipelines: Translation-based pipelines, employed in studies (P2, P9, P15), provide a practical workaround for low-resource languages by leveraging resources from high-resource languages. The process typically involves translating text written in a low-resource language (LRL) into a high-resource language (HRL) such as English, and then applying a sentiment classifier trained on the HRL (e.g., an English sentiment model). This approach is attractive because it allows researchers to bypass the lack of annotated datasets in LRLs by "borrowing" established tools and datasets from HRLs. However, this method is prone to translation noise, where idioms, sarcasm, negations, or culturally specific expressions are mistranslated, leading to distorted sentiment predictions. Translation systems may also encode systematic bias against informal registers, dialectal variants, or slang common in social media, further reducing reliability. To improve robustness, researchers suggest training with noise-augmented machine translation (MT) outputs, such as back-translations or synthetic error injection, adding quality-estimation filtering to remove unreliable translations, and calibrating classification thresholds by domain to account for translation-induced variability. While practical, translation pipelines should therefore be treated as interim solutions rather than definitive methods for sustainable CLSA in low-resource settings.

4) Contrastive or self-supervised alignment: Studies such as P18 and P27 adopt contrastive learning and self-supervised approaches to improve cross-lingual alignment without relying on parallel corpora, which are often unavailable in low-resource languages. These methods train models to bring semantically

similar pairs closer together and push dissimilar ones apart in the embedding space, thereby creating shared cross-lingual representations. While this reduces dependence on expensive parallel datasets, the effectiveness of these methods depends heavily on the availability of large volumes of unlabeled in-language text, which is typically scarce in LRLs. To mitigate this limitation, researchers suggest combining contrastive loss with weak lexicon anchors, such as small bilingual lexicons or cognate lists, and leveraging distant supervision signals from social media, such as emojis or reaction labels, to reduce data demands and improve robustness.

5) Hybrid approaches (rules + neural) robustness and interpretability: Hybrid approaches, as explored in P14 and P20, integrate rule-based features with pre-trained language models to improve robustness in sentiment analysis for low-resource languages. These systems are particularly effective in handling linguistic phenomena such as negation, intensifiers, and morphological variations (e.g., prefixes and affixes) that often confuse purely neural models. Although hybrids are sometimes dismissed as outdated, they remain valuable in LRL contexts, where PLMs frequently struggle with dialectal drift and morphological complexity. The main drawbacks are the engineering effort required to develop and maintain language-specific rules and the limited portability across languages. Suggested design patterns include using rules as weak supervision signals to generate high-precision training data, adopting noise-aware objectives to handle imperfect labels, applying constrained decoding to avoid polarity flips near negators, and performing morphology normalization before tokenization to reduce misclassification.

6) Linguistic resource engineering: Studies in P6, P16, and P22 highlight the continued importance of linguistic resources such as sentiment lexicons, pivot-based alignment, and corpus construction in supporting sentiment analysis for low-resource languages. These resources, often considered "traditional", remain high-leverage tools when PLMs underperform, as they provide interpretable anchors for sentiment recognition. However, a key limitation is coverage drift, as lexicons quickly become outdated due to the rapid evolution of language, especially in social media contexts. To address this, researchers recommend semi-automating lexicon expansion through distributional similarity or retrofitting techniques, supported by human-in-the-loop validation to ensure quality. Furthermore, documenting update cadence and monitoring out-of-vocabulary sentiment drift over time are necessary to maintain the reliability and relevance of these resources.

7) Fine-tuning strategies: Fine-tuning strategies, examined in studies such as P1, P11, P13, and P17, adapt pre-trained multilingual models to sentiment tasks through end-to-end optimization. While full fine-tuning often achieves strong results in high-resource and medium-resource settings, its effectiveness in true low-resource languages is limited. With very little annotated data, full fine-tuning risks catastrophic forgetting of cross-lingual generalization and is computationally demanding, making it less accessible for researchers with limited resources. A more sustainable alternative is parameter-efficient adaptation methods, such as adapters, low-rank adaptation (LoRA), or prefix-tuning, which require fewer parameters and are less computationally intensive. Researchers also recommend performing ablation studies to identify which layers benefit most from adaptation and reporting computational budgets alongside accuracy scores to promote reproducibility, equity, and fair comparison across studies.

Overall, the synthesis of findings for RQ2 suggests that no universal winner for all the methods. The most credible pipelines are composite, matching technique to language typology, script, morphology, and domain, where they report robustness, cost, and portability, and not just peak scores.

C. Challenges and Research Gaps in Cross-Lingual Sentiment Analysis

Addressing RQ3, this review identifies seven recurring obstacles in cross-lingual sentiment analysis (CLSA) for low-resource languages (LRLs). Importantly, many of these barriers are structural as well as technical, reflecting deeper issues of fairness, inclusivity, and sustainability in the field.

1) Annotated data scarcity: A recurring challenge across multiple studies (P1, P3, P5, P14, P20, P22) is the limited availability of annotated datasets in low-resource languages. This scarcity severely constrains the development and evaluation of supervised models and encourages over-reliance on translation or synthetic data generation, which are often imperfect substitutes. Without high-quality annotated resources, model performance cannot be reliably compared, and advances risk being skewed toward well-resourced languages. To address this, researchers highlight the importance of community-driven dataset creation, underpinned by fair compensation and participatory governance. Complementary strategies, such as active learning to maximize annotator efficiency or simplifying task designs through pairwise preference judgments rather than fixed multi-class labels, can further reduce annotation costs while maintaining data quality.

2) Semantic misalignment across languages: Studies such as P4, P7, and P18 report difficulties with semantic misalignment, where shared embedding spaces fail to capture culturally specific sentiment cues. This is particularly problematic for languages with rich morphology or those that rely heavily on honorifics, pragmatic markers, or cultural idioms to express sentiment. In cross-lingual evaluations, aligning syntactic and semantic representations remains one of the most pressing issues in current research [31]. As a result, models trained on generic multilingual embeddings often misinterpret or dilute language-specific sentiment signals. Research is therefore needed to develop context- and morphology-aware representations, introduce language-cluster adapters to better capture typological variation, and incorporate contrastive alignment methods anchored by small bilingual lexicons to preserve semantic nuance across languages.

3) Translation noise in pipeline methods: Translation-based approaches, used in studies P2 and P9, often suffer from translation noise, where idioms, sarcasm, or polarity shifters are mistranslated, leading to distorted sentiment predictions. Such noise contaminates the downstream classification process, as models are forced to learn from inaccurate or incomplete representations. This problem is particularly acute in informal or domain-specific registers such as social media. To mitigate these risks, research has suggested noise-aware training, where models are trained with intentionally corrupted translations to increase robustness; quality-estimation filtering to remove unreliable translations; and domain-specific machine translation tuning, which prioritizes sentiment-sensitive phenomena during training.

4) Domain and dialect mismatch: Another significant gap arises from domain and dialect mismatch, reported in studies such as P10, P19, and P22. Models trained on generic domains (e.g., product reviews) often perform poorly when transferred to specialized domains like healthcare or law, where sentiment expressions differ significantly. Similarly, dialectal variation within languages remains underserved, with models frequently failing to adapt to non-standard forms or code-switching. Addressing this requires domain adaptation techniques such as adapter modules or feature alignment, as well as dialect-sensitive evaluation splits to ensure fair testing. In addition, code-switch stress tests should be integrated into benchmarks to better reflect the realities of multilingual online communication.

5) Bias in PLMs and unfair performance distribution: Bias embedded within pre-trained multilingual models is another critical challenge, as highlighted in (P23, P26). PLMs tend to disproportionately benefit high-resource languages, while low-resource ones face significantly higher error rates. This unequal distribution of performance amplifies digital inequity, leaving already underrepresented communities further behind. To address this, researchers recommend systematic bias audits, which assess per-language and per-dialect performance disparities, alongside re-weighting strategies during fine-tuning to rebalance training. Furthermore, fairness metrics should be reported alongside accuracy, ensuring that performance is evaluated not only in aggregate but also across linguistic subgroups.

6) Few-shot performance instability: While few-shot and zero-shot learning promise inclusivity for languages without annotated data, studies (P25, P26) consistently report performance instability. Results fluctuate depending on prompt design, choice of verbalizers, or random seeds, making them unreliable in production or high-stakes contexts. This instability undermines confidence in few-shot learning as a dependable solution for CLSA in LRLs. Suggested remedies include adopting prompt sets instead of single prompts, systematically reporting variance across runs, introducing consistency regularization to stabilize predictions across paraphrased inputs, and exploring meta-learning approaches to improve generalization from small support sets.

7) Evaluation inconsistency and weak comparability: Finally, a persistent obstacle identified in studies such as P6, P12 and P24 is the lack of standardized evaluation benchmarks. The use of heterogeneous datasets, task formulations, and metrics makes it difficult to compare results across studies, weakening claims of "state-of-the-art" performance. This inconsistency fragments the field and slows cumulative progress. To resolve this, the community should prioritize the development of shared multilingual evaluation suites covering both high- and low-resource languages, incorporate robustness benchmarks for negation, sarcasm, and code-switching, and mandate cost/compute reporting to contextualize performance claims. Future studies could also draw from ongoing advances in multilingual hate speech detection and emotion-based sentiment tasks to ensure fairer evaluation across diverse language contexts [32].

Across all challenges, three cross-cutting practices are recommended to elevate CLSA research from reporting scores to producing scientific evidence. First, papers should include robustness and fairness audits to demonstrate reliability across languages and domains. Second, cost–benefit disclosures (e.g., annotation hours, compute budgets) should accompany accuracy reports to ensure equitable comparisons. Third, error taxonomies should be included to highlight common failure modes such as negation, intensifiers, irony, dialect variation, or code-switching. Overall, the analysis of RQ3 indicates that progress in CLSA for LRLs depends less on any single algorithmic innovation and more on achieving method–task fit, culturally valid label spaces, and transparent reporting practices. By addressing structural as well as technical barriers, the field can move toward more equitable and robust sentiment analysis systems that reflect the linguistic diversity of global communities.

VI. Discussion

This systematic review of 27 studies on cross-lingual sentiment analysis (CLSA) for low-resource languages (LRLs) reveals a field that is advancing in technical scope yet uneven in inclusivity and robustness. Synthesizing across the three research questions (RQ1 to RQ3), three thematic insights emerge: task diversification, methodological trade-offs, and persistent structural barriers.

A. Diversification of Sentiment Tasks (RQ1)

In addressing RQ1, the findings show that while polarity detection continues to dominate, there is a gradual shift toward fine-grained and socially relevant tasks, such as aspect-based sentiment analysis (ABSA), emotion detection, and hate speech recognition. This diversification signals maturity, as researchers recognize that polarity alone cannot capture the richness of human affect or the complexity of multilingual discourse.

However, the evidence also suggests that this diversification is uneven and shallow. ABSA and emotion detection remain marginal in CLSA studies, largely because of annotation challenges and cultural divergence in sentiment expression. For example, most ABSA implementations borrow English-based aspect taxonomies, which may not align with domain-specific discourse in LRLs, while emotion detection often relies on

Western-centric psychological categories. These practices risk semantic distortion, producing models that are technically functional but culturally misaligned. Recent studies continue to highlight the challenges of extending CLSA frameworks to low-resource languages, where semantic alignment and transfer learning remain key strategies for improving sentiment transfer across language pairs [33].

Thus, CLSA research must move beyond convenience-driven task choices toward context-driven task design. Future work should incorporate culturally grounded annotation schemes, involve community participation in defining aspects and emotion categories, and focus on safety-critical tasks like hate speech detection, where the stakes are highest for marginalized communities.

B. Methodological Diversity and Trade-Offs (RQ2)

In relation to RQ2, the reviewed studies demonstrate a wide methodological spectrum: pre-trained multilingual models (PLMs), few-shot and zero-shot learning, translation-based pipelines, contrastive/self-supervised methods, hybrid approaches, and linguistic resource engineering.

Each approach offers distinct strengths and weaknesses. PLMs provide scalability and strong baselines but disproportionately benefit HRLs due to biased pre-training data. Translation pipelines remain practical but introduce translation noise, which distorts sentiment signals in idiomatic or informal language. Few- and zero-shot approaches promise inclusivity but suffer from instability, with results swinging according to prompt design or dataset composition. Few-shot learning techniques have gained momentum as an effective solution for data-scarce settings by optimizing inter-sample relationships and improving class separation [34]. Hybrids and linguistic resources appear "old-fashioned", yet they deliver robustness and interpretability in morphologically complex or dialectally diverse languages, precisely the settings where PLMs struggle.

The key insight is that there is no universal solution. Instead, the most promising direction lies in composite or adaptive frameworks, where methods are tailored to the typology, morphology, and domain of each language. For example, PLMs can provide general cross-lingual representations, while rule-based components enforce negation handling, and lexicons anchor culturally specific sentiment cues. Studies that experimented with such combinations (e.g., P14, P20, P22) highlight the value of methodological hybridity, yet systematic evaluations of these trade-offs are still rare.

Going forward, CLSA research should explicitly report trade-offs—not only accuracy but also robustness under domain/dialect shift, annotation cost, computational budget, and fairness outcomes. Without such transparency, the field risks celebrating narrow performance gains while ignoring broader inclusivity and sustainability concerns.

C. Structural Challenges and Research Gaps (RQ3)

With respect to RQ3, the most striking finding is that progress in CLSA for LRLs is constrained less by algorithmic capacity than by structural limitations. Data scarcity remains the single greatest bottleneck, forcing reliance on synthetic augmentation or translation proxies. While useful stopgaps, these substitutes cannot fully replicate the richness of authentic, in-language annotations. Semantic misalignment persists as a fundamental challenge, with PLMs failing to capture culturally embedded sentiment cues in typologically distant languages. Translation noise undermines the reliability of pipeline methods, especially in informal or idiomatic domains. Domain and dialect mismatch continues to degrade model robustness, revealing that general-purpose CLSA systems lack adaptability. Bias in PLMs introduces unfairness, privileging HRLs at the expense of underrepresented languages, while few-shot instability limits the practical reliability of low-label methods. Hybrid cross-lingual approaches have shown promise in handling dialectal variations by combining multilingual representations with localized embeddings [35]. Finally, evaluation inconsistency fragments the field, making results across studies incomparable and slowing cumulative progress. These gaps underline that CLSA for LRLs is not only a technical challenge but also a socio-structural one, tied to questions of fairness, inclusivity, and cultural validity.

D. Broader Implications and Future Directions

Three implications for the future of cross-lingual sentiment analysis (CLSA) research emerge: Firstly, task relevance matters, where research agendas must expand beyond polarity detection to emphasize fine-grained and socially impactful tasks such as ABSA, emotion detection, and abusive language detection. These tasks are harder to annotate but yield greater societal value. Secondly, adaptivity is essential. Rather than pursuing "one-size-fits-all" methods, researchers should embrace hybrid and adaptive pipelines, combining PLMs with rule-based cues, lexicons, and contrastive learning, depending on the linguistic and domain context. Finally, fairness and inclusivity must be central. Without deliberate attention to dataset creation, bias mitigation, and evaluation standardization, CLSA risks reinforcing digital inequality. Low-resource language communities will continue to lag unless they are directly involved in defining sentiment tasks, building resources, and validating models.

In summary, the synthesis across RQ1 to RQ3 reveals that cross-lingual sentiment analysis (CLSA) for low-resource languages (LRLs) is progressing, but in fragmented and uneven ways. The technical toolkit is expanding—PLMs, few-shot learning, and contrastive methods all offer promise—but structural barriers persist. True progress will come not from chasing higher accuracy on polarity benchmarks, but from building inclusive, adaptive, and culturally grounded cross-lingual sentiment analysis CLSA systems that reflect the realities of the languages and communities they aim to serve.

Taken together, the answers to RQ1 to RQ3 reveal that CLSA research for low-resource languages is marked by imbalances in task focus, tensions in methodological design, and persistent structural barriers. These findings highlight that technical progress alone will not guarantee equitable outcomes. Instead, the field must reorient toward questions of relevance (task selection), reliability (methodological robustness), and responsibility (fairness and inclusivity). Building on this synthesis, the next section outlines the broader implications of these findings and proposes future directions that can guide a more inclusive and sustainable research agenda.

E. Synthesis Framework for Cross-Lingual Sentiment Analysis (CLSA) in Low-Resource Languages

To consolidate the findings of this review, this subsection introduces a Synthesis Framework for Cross-Lingual Sentiment Analysis (CLSA) in low-resource languages (see Fig. 3). The framework reconceptualizes the challenges identified across the reviewed studies—not as limitations, but as *drivers of research innovation*. Specifically, three fundamental challenges which are data scarcity**,** domain mismatch**,** and cultural bias**,** emerge as key factors shaping how CLSA methods and tasks evolve. Moreover, dialectic preference bias persists even in large-scale PLMs, raising concerns about fairness and representational balance across dialects [37].

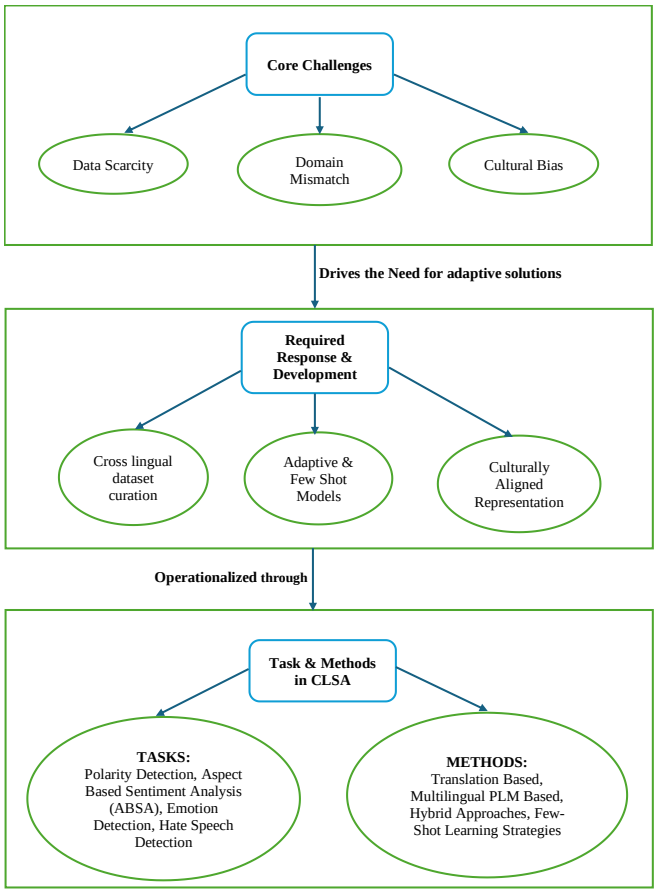

Fig. 3. Synthesis framework linking core challenges, development needs, CLSA tasks and methods in low-resource languages.

As illustrated in Fig. 3, these challenges generate distinct research and development needs, including cross-lingual dataset curation to address the lack of annotated resources, adaptive and few-shot learning techniques to enable knowledge transfer under limited supervision, and culturally aligned semantic representations to mitigate linguistic and contextual bias. Each of these needs directly influences the design of current and future CLSA tasks and methodological approaches.

The lower layer of the framework connects these needs with the practical implementation space—encompassing core CLSA tasks (polarity detection, aspect-based sentiment analysis (ABSA), emotion detection, and hate speech identification) and methodological strategies (translation-based, multilingual PLM-based, hybrid, and few-shot learning approaches). This linkage demonstrates how theoretical constraints are translated into actionable design choices. For instance, data scarcity stimulates the development of few-shot learning models, while cultural bias drives hybrid and knowledge-injected approaches that integrate linguistic and contextual knowledge. The importance of community-driven dataset development is evident in recent work on low-resource languages such as Bhojpuri, Maithili, and Magahi, which demonstrates how localized corpus design strengthens model generalization [38].

In summary, the proposed framework synthesizes empirical evidence from the reviewed literature into a structured, three-layer model that explains why certain CLSA methods emerge and how they align with the evolving research landscape. Rather than presenting CLSA as a static taxonomy of techniques, this synthesis reframes it as a dynamic, challenge-driven ecosystem, in which methodological innovation continuously responds to the constraints of low-resource linguistic environments.

The framework illustrates how major challenges such as data scarcity, domain mismatch, and cultural bias drive the need for adaptive solutions, leading to the development of new research responses and methodological innovations in cross-lingual sentiment analysis.

TABLE IV. COMPARATIVE SUMMARY OF TASKS, METHODS, AND CHALLENGES IN CROSS-LINGUAL SENTIMENT ANALYSIS (CLSA) FOR LOW-RESOURCE LANGUAGES

Task Type	Dominant Methods	Key Challenges	Research Gap / Opportunity
Polarity Detection	Translation + PLM	Translation noise, imbalance	Domain-adaptation across dialects
Aspect-Based SA (ABSA)	Hybrid + Few-Shot Learning	Aspect misalignment	Need culturally grounded aspect extraction
Emotion Detection	PLM + Contrastive	Semantic bias	Lack of local emotion lexicons
Hate Speech	PLM + Knowledge Injection	Bias & toxicity lexicon gaps	Fairness and bias auditing frameworks

As summarized in Table IV, the comparative overview highlights how each task category aligns with dominant methods, key challenges, and potential research opportunities within CLSA for low-resource languages. These patterns are consistent with prior large-scale reviews on cross-lingual sentiment analysis that documented similar task-method trends and evaluation issues [36].

VII. IMPLICATIONS AND FUTURE DIRECTIONS

The findings of this review carry theoretical, methodological, and practical implications for advancing cross-lingual sentiment analysis (CLSA) in low-resource languages (LRLs). The synthesis of RQ1 to RQ3 highlights that the field is at a crossroads, while technical capacity has expanded through pre-trained multilingual models (PLMs) and few-shot learning, progress remains constrained by imbalanced task focus, uneven methodological robustness, and structural barriers to inclusivity.

A. Theoretical Implications

From a theoretical perspective, the review underscores that CLSA research has been overly anchored in polarity detection,

reinforcing a narrow view of sentiment as binary or ternary classification. These risks oversimplifying human affect and ignoring culturally specific sentiment categories. A theoretical implication is the need to reconceptualize sentiment in multicultural, multilingual contexts. Instead of importing Western-centric taxonomies, future work should build frameworks that emerge from local discourse—for example, through community-driven aspect induction or culturally validated emotion ontologies. This would strengthen the linguistic and psychological validity of CLSA tasks in LRLs.

B. Methodological Implications

Methodologically, the findings imply that no single technique is sufficient for CLSA. PLMs provide scalability but are biased toward high-resource languages. Translation pipelines are pragmatic but prone to semantic drift. Few- and zero-shot learning are inclusive but unstable. Hybrids and linguistic resources offer robustness but demand manual effort.

The implication is clear, adaptive, composite frameworks should be the methodological standard, not the exception. Future CLSA studies should report trade-offs explicitly, including robustness under domain and dialect shift (not just peak accuracy), cost–benefit ratios (annotation hours, compute budgets vs. performance gains), and fairness metrics (performance gaps between HRLs and LRLs). Such transparency will enable the field to move beyond fragmented "accuracy chasing" and toward scientific accumulation and practical reliability.

C. Practical Implications

Practically, this review reveals that the current trajectory risks widening the digital divide between high- and low-resource languages. Without deliberate intervention, CLSA advances will disproportionately benefit languages already well-represented in pre-training corpora. This has real-world consequences: LRL communities may remain excluded from tools for opinion mining, public health monitoring, or content moderation.

To counter this, future CLSA must be inclusive by design. This means establishing community-driven dataset creation initiatives with fair compensation and participatory governance, designing culturally sensitive benchmarks that capture local discourse phenomena such as code-switching, dialectal variation, and culture-specific sentiment markers, and building bias-aware evaluation frameworks that surface disparities across language families, domains, and social groups. These steps would ensure that CLSA contributes not only to academic progress but also to equitable digital participation.

D. Future Research Directions

Based on the synthesis, several concrete research directions emerge. Firstly, task diversification. Researchers need to move beyond polarity detection toward ABSA, emotion detection, and abusive language detection in cross-lingual and low-resource settings. Prioritize tasks with high societal impact (e.g., detecting harmful speech, understanding public health discourse). Secondly, culturally grounded task design. Researchers may develop annotation schemes and sentiment categories that reflect local cultural contexts rather than importing English-based taxonomies. Thirdly, adaptive hybrid frameworks. Future research may combine PLMs with linguistic rules, lexicons, and contrastive/self-supervised objectives to improve robustness in morphologically rich or dialectally diverse low-resource languages (LRLs). Fourthly, fairness-aware model development where researchers can integrate bias detection, fairness metrics, and bias mitigation as core evaluation components, ensuring that CLSA does not exacerbate inequalities between high resource languages (HRLs) and low-resource languages (LRLs). Next, standardized multilingual benchmarks. Establish shared, community-maintained evaluation suites that cover HRLs and LRLs, include robustness tests (negation, sarcasm, code-switching), and track performance parity across language families. Finally, apply transparent reporting practices. Encourage CLSA publications to disclose not only accuracy metrics but also annotation cost, computational budget, variance across seeds/prompts, and error taxonomies. This would promote replicability and responsible claims of state-of-the-art performance.

In conclusion, the implications of this review are clear to state that CLSA for LRLs is not just a technical challenge but a socio-technical enterprise, where fairness, inclusivity, and cultural grounding are as important as model performance. Future research must adopt adaptive, hybrid, and community-driven approaches to ensure that sentiment analysis truly serves the linguistic diversity of the world. By embracing this agenda, the field can shift from incremental performance gains toward building equitable, culturally resonant, and globally relevant sentiment analysis systems.

VIII. CONCLUSION

This review has systematically examined recent advances in cross-lingual sentiment analysis (CLSA) for low-resource languages, focusing on the interrelations between sentiment tasks, methodological approaches, and the challenges that shape them. The findings reveal that despite substantial progress in multilingual pre-trained models and translation-based pipelines, CLSA still faces persistent issues related to data scarcity, domain mismatch, and cultural bias.

To address these gaps, the study proposed a Synthesis Framework for CLSA (Fig. 3), which integrates the connections between core challenges, emerging research needs, and corresponding methodological responses. This framework illustrates how each limitation serves as a catalyst for innovation—driving the development of adaptive and few-shot learning models, cross-lingual dataset curation, and culturally aligned representation strategies. By interpreting challenges as enablers of progress, the framework transforms CLSA from a descriptive research field into a dynamic, challenge-driven ecosystem**.**

In summary, the review highlights that future CLSA research should move toward adaptive, ethically informed, and context-aware models that not only achieve cross-lingual transferability but also preserve the cultural and linguistic diversity of low-resource communities. The proposed synthesis framework can serve as a conceptual guide for these future directions, promoting more inclusive, robust, and equitable sentiment analysis across languages and dialects.

ACKNOWLEDGMENT

This research was funded by a grant from Ministry of Higher Education of Malaysia Scheme [FRGS/1/2022/ICT02/UNISZA/02/2].

REFERENCES

[1] F. T. Johora Faria, M. B. Moin, R. I. Mumu, M. M. Alam Abir, A. N. Alfy and M. S. Alam, "Motamot: A Dataset for Revealing the Supremacy of Large Language Models Over Transformer Models in Bengali Political Sentiment Analysis," *2024 IEEE Region 10 Symposium (TENSYMP)*, New Delhi, India, 2024, pp. 1-8, doi: 10.1109/TENSYMP61132.2024.10752197.

[2] Almahdi, A. J., Mohades, A., Akbari, M., & Heidary, S. (2025). Enhancing cross-lingual hate speech detection through contrastive and adversarial learning. *Engineering Applications of Artificial Intelligence*, *147*. https://doi.org/10.1016/j.engappai.2025.110296

[3] Masih, S., Hassan, M., Fahad, L. G., & Hassan, B. (2025). Transformer-Based Abstractive Summarization of Legal Texts in Low-Resource Languages. *Electronics (Switzerland)*, *14*(12). https://doi.org/10.3390/electronics14122320

[4] Dai, W., Han, D., Tohti, T., Liang, Y., Zuo, Z., Liao, Y., & Yang, Q. (2025). TFT-TL: Token-Level Filter Training Transfer Learning for Low-Resource Neural Machine Translation. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *24*(6). https://doi.org/10.1145/3732779

[5] Chen, L., Shang, S., & Wang, Y. (2025). Bridging resource gaps in cross-lingual sentiment analysis: adaptive self-alignment with data augmentation and transfer learning. *PeerJ Computer Science*, *11*. https://doi.org/10.7717/peerj-cs.2851

[6] Salleh, A., Osman, M. H., Hassan, S., Said, M. Y., Sharif, K. Y., & Wei, K. T. (2025). A hybrid model for low-resource language text classification and comparative analysis. *Knowledge-Based Systems*, *326*. https://doi.org/10.1016/j.knosys.2025.114068

[7] Almalki, S. (n.d.). Sentiment Analysis and Emotion Detection Using Transformer Models in Multilingual Social Media Data. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 16, Issue 3). www.ijacsa.thesai.org

[8] Alvarado, F. H. C. (2025). The impact of linguistic variations on emotion detection: A study of regionally specific synthetic datasets. *Applied Sciences*, *15*(7), 3490. https://doi.org/10.3390/app15073490

[9] Man, C. Z., Si Mar Win, S., & Lai Khine, K. L. (2025). Exploring Transfer Learning in a Bidirectional Myanmar-Tedim Chin Machine Translation with the mT5 Transformer. *Proceedings of the 22nd IEEE International Conference on Computer Applications, ICCA 2025*. https://doi.org/10.1109/ICCA65395.2025.11011103

[10] Pandya, H., & Bhatt, B. (2024). Enhancing Low-Resource Question-Answering Performance Through Word Seeding and Customized Refinement. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 15, Issue 3). www.ijacsa.thesai.org

[11] Shafikuzzaman, M., Islam, M. R., Rolli, A. C., Akhter, S., & Seliya, N. (2024). An Empirical Evaluation of the Zero-Shot, Few-Shot, and Traditional Fine-Tuning Based Pretrained Language Models for Sentiment Analysis in Software Engineering. *IEEE Access*, *12*, 109714–109734. https://doi.org/10.1109/ACCESS.2024.3439450

[12] Přibáň, P., Šmíd, J., Steinberger, J., & Mištera, A. (2024). A comparative study of cross-lingual sentiment analysis. *Expert Systems With Applications*, *247*, 123247. https://doi.org/10.1016/j.eswa.2024.123247

[13] A, A. G., & V, V. (2024). Sentiment analysis on a low-resource language dataset using multimodal representation learning and cross-lingual transfer learning. *Applied Soft Computing*, *157*, 111553. https://doi.org/10.1016/j.asoc.2024.111553

[14] van Thin, D., Quoc Ngo, H., Ngoc Hao, D., & Luu-Thuy Nguyen, N. (2023). Exploring zero-shot and joint training cross-lingual strategies for aspect-based sentiment analysis based on contextualized multilingual language models. *Journal of Information and Telecommunication*, *7*(2), 121–143. https://doi.org/10.1080/24751839.2023.2173843

[15] Tareq, M., Islam, M. F., Deb, S., Rahman, S., & Mahmud, A. A. (2023). Data-Augmentation for Bangla-English Code-Mixed Sentiment Analysis: Enhancing cross linguistic contextual Understanding. *IEEE Access*, *11*, 51657–51671. https://doi.org/10.1109/access.2023.3277787

[16] Kumar, P., Pathania, K., & Raman, B. (2023). Zero-shot learning based cross-lingual sentiment analysis for sanskrit text with insufficient labeled data. *Applied Intelligence*, *53*(9), 10096–10113. https://doi.org/10.1007/s10489-022-04046-6

[17] Laifa, M., & Mohdeb, D. (2023). Sentiment analysis of the Algerian social movement inception. *Data Technologies and Applications*, *57*(5), 734–755. https://doi.org/10.1108/dta-10-2022-0406

[18] Ghasemi, R., & Momtazi, S. (2023). How a deep contextualized representation and attention mechanism justifies explainable Cross-Lingual Sentiment analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22*(11), 1–15. https://doi.org/10.1145/3626094

[19] An, B. (2023). Prompt-based for Low-Resource Tibetan text classification. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22*(8), 1–13. https://doi.org/10.1145/3603168

[20] Raja, E., Soni, B., & Borgohain, S. K. (2023). Fake news detection in Dravidian languages using transfer learning with adaptive finetuning. *Engineering Applications of Artificial Intelligence*, *126*, 106877. https://doi.org/10.1016/j.engappai.2023.106877

[21] Eronen, J., Ptaszynski, M., Masui, F., Arata, M., Leliwa, G., & Wroczynski, M. (2022). Transfer language selection for zero-shot cross-lingual abusive language detection. *Information Processing and Management*, *59*(4). https://doi.org/10.1016/j.ipm.2022.102981

[22] Mao, C., Man, Z., Yu, Z., Wu, X., & Liang, H. (2022). Burmese Sentiment Analysis Based on Transfer Learning. *Journal of Information Processing Systems*, *18*(4), 535–548. https://doi.org/10.3745/JIPS.04.0249

[23] Ghasemi, R., Asli, S. a. A., & Momtazi, S. (2020). Deep Persian sentiment analysis: Cross-lingual training for low-resource languages. *Journal of Information Science*, *48*(4), 449–462. https://doi.org/10.1177/0165551520962781

[24] Tesfagergish, S. G., Kapočiūtė-Dzikienė, J., & Damaševičius, R. (2022). Zero-Shot Emotion Detection for Semi-Supervised Sentiment Analysis Using Sentence Transformers and Ensemble Learning. *Applied Sciences (Switzerland)*, *12*(17). https://doi.org/10.3390/app12178662

[25] Wang, D., Wu, J., Yang, J., Jing, B., Zhang, W., He, X., & Zhang, H. (2021). Cross-Lingual knowledge transferring by structural correspondence and space transfer. *IEEE Transactions on Cybernetics*, *52*(7), 6555–6566. https://doi.org/10.1109/tcyb.2021.3051005

[26] Kumar, A., & Albuquerque, V. H. C. (2021). Sentiment analysis using XLM-R transformer and zero-shot transfer learning on resource-poor Indian language. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *20*(5), 1–13. https://doi.org/10.1145/3461764

[27] Pamungkas, E. W., Basile, V., & Patti, V. (2021). A joint learning approach with knowledge injection for zero-shot cross-lingual hate speech detection. *Information Processing & Management*, *58*(4), 102544. https://doi.org/10.1016/j.ipm.2021.102544

[28] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

[29] Luo, X., Deng, Z., Yang, B., & Luo, M. Y. (2024). Pre-trained language models in medicine: A survey. In *Artificial Intelligence in Medicine* (Vol. 154). Elsevier B.V. https://doi.org/10.1016/j.artmed.2024.102904

[30] Wang, H., Li, J., Wu, H., Hovy, E., & Sun, Y. (2023). Pre-Trained Language Models and Their Applications. In *Engineering* (Vol. 25, pp. 51–65). Elsevier Ltd. https://doi.org/10.1016/j.eng.2022.04.024

[31] Anidjar, O. H., Yozevitch, R., Bigon, N., Abdalla, N., Myara, B., & Marbel, R. (2023). Crossing language identification: Multilingual ASR framework based on semantic dataset creation & Wav2Vec 2.0. *Machine Learning with Applications*, *13*, 100489. https://doi.org/10.1016/j.mlwa.2023.100489

[32] Kia, M. A., & Samiee, D. (2024). From Monolingual to Multilingual: Enhancing Hate Speech Detection with Multi-channel Language Models.

Procedia Computer Science, *246*(C), 2704–2713. https://doi.org/10.1016/j.procs.2024.09.401

[33] Zhou, P., & Zhao, L. (2025). Cross-lingual sentiment analysis for low-resource languages via semantic alignment and transfer learning. *International Journal of Information and Communication Technology*, *26*(33), 1–17. https://doi.org/10.1504/ijict.2025.148657

[34] Chen, X., Wu, W., Ma, L., You, X., Gao, C., Sang, N., & Shao, Y. (2024). Exploring sample relationship for few-shot classification. *Pattern Recognition*, *159*, 111089. https://doi.org/10.1016/j.patcog.2024.111089

[35] Chabane, M., Harrag, F., & Shaalan, K. (2025). Advancing low-resource dialect identification: A hybrid cross-lingual model leveraging CAMeLBERT and FastText for Algerian Arabic. *Expert Systems With Applications*, *284*, 127816. https://doi.org/10.1016/j.eswa.2025.127816

[36] Xu, Y., Cao, H., Du, W., & Wang, W. (2022). A survey of Cross-lingual sentiment Analysis: Methodologies, models and Evaluations. *Data Science and Engineering*, *7*(3), 279–299. https://doi.org/10.1007/s41019-022-00187-3

[37] Hassan, M. F., Khattak, F. K., & Seyyed-Kalantari, L. (2025). Dialectic preference bias in large language models. *Proceedings of the AAAI Symposium Series*, *5*(1), 365–369. https://doi.org/10.1609/aaaiss.v5i1.35613

[38] Mundotiya, R., Kumar, S., Kumar, A., Chaudhary, U., Chauhan, S., Mishra, S., Gatla, P., & Singh, A. K. (2022). Development of a dataset and a deep learning baseline named Entity Recognizer for three low resource languages: Bhojpuri, Maithili, and Magahi. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22*(1), 1–20. https://doi.org/10.1145/3533428