Caliphate Journal of Science & Technology (CaJoST)

ISSN: 2705-313X (PRINT); 2705-3121 (ONLINE) Research Article

Open Access Journal available at: https://cajost.com.ng/index.php/files and https://www.ajol.info/index.php/cajost/index

DOI: https://dx.doi.org/10.4314/cajost.v7i3.16

Article Info

Received: 13th September 2024

Revised: 19th August 2025

Accepted: 26th August 2025

Exploring Emotion Detection and Sentiment Analysis of Texts in Low-Resource Languages: Techniques Challenges, Emerging Trends and Future Directions

Joy A. Emmanuel [1*], Taiwo Kolajo [2] and Joshua B. Agbogun[3]

1,2,3 Department of Computer Science, Federal University Lokoja, Nigeria.

***Corresponding author's email:**

joy.emmanuel-pg@fulokoja.edu.ng

Emotion detection and sentiment analysis in low-resource languages have gained significant attention in natural language processing (NLP). This is due to their importance in understanding user-generated content, especially on social media platforms. However, these tasks face numerous challenges due to the scarcity of labelled data, complex linguistic structures, and limited computational resources. This paper explores the state-of-the-art methodologies for emotion detection and sentiment analysis in low-resource languages. We identify key open problems such as poor analytical accuracy due to limited datasets, the complicated morphology of native languages, and the lack of suitable linguistic resources. Additionally, we discuss the use of machine learning and deep learning models, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), deep belief networks (DBNs), and lexicon-based approaches, as potential solutions to these challenges. These approaches leverage techniques such as transfer learning, pre-trained embeddings, and hybrid models to mitigate the limitations of small datasets. Finally, to further address these problems, we present emerging trends in research in this field, a combination of advanced models and data augmentation techniques to enhance the accuracy and robustness of emotion detection and sentiment analysis in these languages.

Keywords: Emotion detection, Sentiment analysis, Low-resource languages, Machine learning, Deep Learning models.

Cite this: *CaJoST*, 2025, 3, 429-446

1. Introduction

Natural language processing, or NLP, has advanced significantly, making it possible for machines to understand and communicate with human language. Sentiment analysis and emotion detection are closely related fields at the forefront of this advancement [1]. Emotion detection, sometimes referred to as emotion recognition, is the process of using technology to identify and decipher the range of human emotion from both spoken and unspoken cues. This includes determining a person's emotions through data analysis using artificial intelligence (AI), machine learning (ML), and natural language processing (NLP), which enables more individualized interactions and profound insights. In order to do this, emotion detection systems use sophisticated algorithms to analyze communication signals such as speech patterns, vocal tones, facial movements, body language, and physiological reactions. AI-driven systems categorize emotions like surprise, joy, happiness, frustration, and sadness. They frequently incorporate methods like text sentiment analysis, voice analysis and facial recognition to obtain a deeper comprehension of human emotions and actions. Identifying and extracting emotions expressed in spoken or written language is the specific goal of emotion detection within natural language processing (NLP) [2]. This is accomplished by looking for patterns in text that indicate emotional states, such as words or phrases that are linked to particular feelings. Sentiment analysis, frequently referred to as opinion mining, complements emotion detection by focusing on determining the emotional tone or polarity expressed within a piece of text. This involves utilizing NLP, text analysis, and statistical methods to analyze customer sentiment, understanding what people are saying, how they are saying it, and what their words imply. Sentiment analysis classifies the expressed opinion in a document, sentence, or even at the feature/aspect level as positive, negative, or neutral. This process

often involves algorithms that employ text analysis and NLP to categorize words based on their sentiment. While sentiment analysis primarily identifies the overall positive, negative, or neutral leaning of a text, emotion detection delves deeper to recognize specific emotions such as joy, sadness, anger, or fear. In essence, sentiment analysis can be seen as a broader category, with emotion detection representing a more granular level of analysis focused on discrete emotional states [3]. Due to its broad range of applications, the field of emotion detection has gained significant popularity. As a result, emotion-driven technology plays a crucial role in decision-making, benefiting various application domains [4]. These domains include artificial intelligence, opinion mining, social media monitoring, recommendation systems, human-computer interaction, psychology, medicine, and marketing. Historically, research in this area has predominantly centred on high-resource languages like English and Chinese, benefiting from extensive annotated datasets and advanced linguistic tools, which have facilitated the development of highly accurate models.

In contrast, many of the world's low-resource languages like Yoruba, Igbo, Igala, Hindi, Bangla, Roman-Urdu and many others are underrepresented and are still emerging in NLP research [5]. These low-resource languages, despite being spoken by millions, often lack sufficient labelled corpora, standardized orthographies, NLP tools, and language-specific pre-trained models. Consequently, implementing emotion detection and sentiment analysis in these languages presents distinct challenges, including limited model generalization, significant dialectal variations, and cultural differences in emotional expression.

Early approaches to sentiment and emotion analysis relied heavily on rule-based systems and manually curated lexicons. While these approaches provided a foundation, they often fell short in capturing the nuanced, context-dependent nature of human emotions, especially in morphologically rich languages. The emergence of machine learning and deep learning techniques, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers, marked a shift towards data-driven approaches. These models have demonstrated effectiveness when trained on large-scale datasets; however, their applicability to low-resource languages remains constrained by data scarcity.

Despite these advancements, challenges persist in achieving high accuracy and cultural relevance in emotion and sentiment classification for low-resource languages. The ongoing scarcity of annotated corpora, complex linguistic structures, and limited computational resources continue to impede progress.

However, several surveys and review articles have been published on Text-Based Emotion Detection (TBED), each emphasizing different aspects of the field. For example, some works have broadly reviewed text-based emotion detection techniques, while others have narrowed their focus to Deep Learning (DL)-based approaches or transformer-based methods and some focused on the state of the arts techniques of TBED. Also, researchers have investigated various classifiers, algorithms, lexicons, and datasets related to TBED. While some have examined the essential components required for effective emotion detection, many existing reviews have concentrated on either emotion detection [3], or sentiment analysis of texts in low resource settings [5]. Only a few studies focused on emotion detection and sentiment analysis of texts in low-resource languages.

Recognizing these limitations, this paper builds upon existing research by examining the current landscape of emotion detection and sentiment analysis in low-resource languages. It also identifies existing challenges and open problems in this domain. Finally, it explores the state-of-the-art methods, the emerging trends, improvements and future research directions aimed at developing more inclusive and representative NLP systems.

1.1 Research questions with objectives

Table 1 displays the research questions along with the related objectives that direct this study. These questions are designed to establish a systematic approach for investigating low-resource languages within the context of emotion recognition and sentiment evaluation. Every objective corresponds with its relevant research question to guarantee clarity, focus, and consistency during the research process.

Table 1: Research questions and objectives

Research question number	Research question	Objective
RQ1	What are Low-resource languages and what are their challenges in NLP?	The aim is to study low-resource languages and identify their challenges in NLP
RQ2	What are the different state of the art techniques used for Emotion detection and sentiment analysis of texts in Low-resource languages?	The aim is to study the different state of the art techniques used for Emotion detection and sentiment analysis of texts in Low-resource languages.
RQ3	What are the challenges and open problems in Emotion detection and sentiment analysis of texts in Low-resource languages?	The aim is to identify open problems in Emotion detection and sentiment analysis of texts in Low-resource languages.
RQ4	What are the emerging trends and improvements and future directions in Emotion detection and sentiment analysis of texts in Low-resource languages?	The aim is to identify emerging trends and improvements and future directions in Emotion detection and sentiment analysis of texts in Low-resource languages.

1.2 Contribution to Knowledge

The contribution of this research work is as follows;

i. It provides the background of low-resource languages.

ii. It explores the state-of-the-art techniques and methods for research in text-based emotion detection and sentiment analysis for low-resource languages.

iii. It also identifies different types of open problems associated with text-based emotion detection and sentiment analysis in this domain.

iv. It presents emerging trends improvements and future directions in research of text-based emotion detection and sentiment analysis in low-resource languages.

v. It contributes to the body of literature.

1.3 Low-Resource Languages

Low-resource languages are those that lack adequate support from digital technology [6], and they do not have the necessary linguistic resources, such as pre-trained models, annotated corpora, and lexicons, which are essential for developing sentiment analysis and emotion recognition systems. These languages often have limited online presence, scarce linguistic resources, and insufficient linguistic expertise [7]. Examples of low-resource languages include Urdu, Uyghur, Telugu, Tamil and Tulu, Hausa, Yoruba, and Igbo (Nigerian), Amharic (Amharic), Tigrinya (Tigrinya, Oromo), Oromo (Ethiopian), Swahili, Algerian Arabic dialect, Twi (Ghanaian), Portuguese, and Darija [8], among others.

Adapting emotion detection and sentiment analysis methods to low-resource languages presents several challenges. A primary obstacle is the lack of annotated data, as these models require large labelled datasets for accurate classification. Due to the scarcity of resources and the variety of dialects and cultures within these languages, gathering such data can be costly, time-consuming, and difficult [9].

Another challenge is the lack of language-specific linguistic resources, such as emotion dictionaries or sentiment lexicons. These tools are essential for assigning words and phrases to emotional categories or identifying their polarity. Creating lexicons for low-resource languages requires significant effort, linguistic knowledge, and manual annotation, which may not always be feasible. It is important to note that low-resource languages are not categorized as such because they lack speakers but rather due to their underrepresentation in digital spaces.

Despite these challenges of dataset scarcity and limited resources, efforts are ongoing to improve the situation for some of these languages. For example, languages like Bengali, Hausa, Igbo, Yoruba, Oromo, Telugu, Tigrinya, Twi, Uyghur, and Urdu are now available on Google Translate. This progress signifies a step forward, resulting from intensive efforts in NLP research where datasets have been created for tasks like sentiment analysis, emotion detection, and machine translation. The inclusion of these low-resource languages in Google Translate

suggests a growing digital presence and the availability of translated text data.

Although these languages have transitioned from low-resource to well-represented languages in NLP, they are yet to be elevated to high-resource languages. Continued efforts in data creation, digitization, and further NLP research are necessary to close this gap. However, innovative methods specifically adapted for languages with limited resources are required to address these challenges. To tackle data scarcity, researchers have explored various transformations and approaches that can be applied to existing datasets to increase the volume and diversity of training data. For example, the effectiveness of cross-lingual emotion classification was demonstrated by [3]. In the study, models were trained on English and knowledge was transferred to languages like Farsi, Arabic, and Odia, achieving superior performance compared to random baselines. Also, [2] applied data augmentation methods to enhance sentiment analysis in Marathi, a low-resource Indian language, resulting in significant performance improvements. Authors used semi-supervised learning combined with a small amount of labelled data with a larger set of unlabelled data to improve model performance.

1.4 Emotion Detection and Sentiment Analysis

Emotion detection and sentiment analysis are key subfields of NLP. While both involve analyzing human emotions expressed in text, they focus on different aspects of emotion understanding. Sentiment analysis (also known as opinion mining) is the process of identifying sentiment or subjective polarity in a piece of text [5]. The goal of sentiment analysis is to determine whether the sentiment expressed is positive, negative, or neutral. The primary focus of sentiment analysis is on the opinion, attitude, or evaluation expressed in the text, rather than specific emotions. For example, in the sentence "I loved the meal, it was sumptuous!" sentiment analysis would classify it as positive. Sentiment analysis is also applied in social media analysis, brand reputation monitoring, market research, and customer feedback analysis, helping organizations gauge public sentiment and satisfaction trends.

On the other hand, emotion detection attempts to identify specific emotions expressed in text, going beyond sentiment analysis. It focuses on recognizing and classifying feelings such as happiness, sadness, anger, fear, surprise, or disgust [10]. For example, emotion detection would identify the sentence "I am very happy to see you!" as happiness. Applications of emotion detection can be seen in areas like social media analysis, customer feedback analysis, chatbot interactions, and mental health monitoring [2]. It provides a better understanding of people's emotional states and responses.

Sentiment analysis and emotion detection can complement each other. Some methods integrate both sentiment analysis and emotion detection to offer a more thorough examination of text, identifying both the sentiment and the specific emotion conveyed [10]. Both sentiment analysis and emotion detection utilize a variety of Natural Language Processing (NLP) techniques, such as machine learning algorithms, deep learning models, lexicon-based approaches, and rule-based methods to analyze and interpret text data in terms of the sentiments or emotions expressed.

2. Related Works

The rapid growth of social networking sites and the resulting generation of massive amounts of unstructured data on the internet has significantly increased research interest in emotion detection and sentiment analysis of texts, as noted by [11]. This section presents related research efforts in the field.

Maruf et al. [12] conducted a literature review of text-based emotion detection (TBED). The authors reviewed previous works and survey papers on the subject. The review specifically focused on Emotional models and their architecture, Machine learning (ML) and deep learning (DL) techniques for TBED, Text preprocessing techniques, Feature extraction methods, Datasets and lexicons used for TBED and the challenges faced by current emotion detection systems. Authors also examined TBED in languages other than English, along with any language-specific considerations.

Jim et al. [13] presented a survey on sentiment analysis (SA) within the field of Natural Language Processing (NLP), focusing particularly on recent advancements involving Machine Learning (ML), Deep Learning (DL), Large Language Models (LLMs), and pre-trained models. The study employed a Systematic Literature Review (SLR) methodology based on PRISMA guidelines to curate 162 high-quality articles published between 2020 and 2023. Authors focused on categorizing applications of sentiment analysis, examine datasets, preprocessing techniques, and evaluation metrics, analyze algorithms from ML, DL, LLMs, pre-trained models, review recent experimental results, identify research challenges and future directions.

Daneshfar [14] outlined the process of collecting and annotating dataset based on tweeter for sentiment analysis in Central Kurdish. The study employed

classical machine learning and neural network-based techniques for the task. Author also adopted a transfer learning approach to enhance performance by leveraging pre-trained models for data augmentation. Results demonstrated that despite the challenging nature of the task, data augmentation contributes greatly to achieving high accuracy.

Wang and Zhang [15] proposed a cross-modality fusion model to addresses the limitations of traditional emotion detection methods that primarily rely on textual analysis. The authors develop Emotion Fusion-Transformer, a model that integrates EEG signals and textual data to enhance emotion detection in English writing. The model utilizes the Transformer architecture to capture contextual relationships within the text and processes EEG signals to extract underlying emotional states. The Emotion Fusion-Transformer preprocesses EEG data, extracts features, and fuses these modalities within a unified Transformer framework. The study demonstrates that the proposed model outperforms text-only and EEG-only approaches, showing improvements in accuracy and F1-score across various emotional categories.

Ahmed et al. [16] proposed using deep learning techniques for sentence-level Urdu sentiment analysis within the context of the Multimedia Internet of Things (MIoT). Dataset for this task were scraped from different Urdu blogs and social media platforms. Some parts of the IMDB data set were used after translating them to Urdu language. Dataset were annotated by Native Urdu speakers and various preprocessing Techniques were applied. Two deep learning models, i.e., Convolutional Neural Network (CNN) and Long Short-term Memory (LSTM), were trained on the Urdu review dataset. The performance of both models was evaluated using accuracy and F1 score and results shows that the LSTM model performed better than the CNN model.

Kaur & Sharma [17] presented a deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. A hybrid approach for analyzing sentiments on review-related features and aspect-related features was introduced for constructing the distinctive hybrid feature vector corresponding to each review. The sentiment classification was performed on three different datasets using the deep learning classifier LSTM and the model achieved good result.

Muhammad & Burney [18] conducted research on Innovations in Urdu Sentiment Analysis Using Machine and Deep Learning Techniques for Two-Class Classification of Symmetric Datasets. Machine learning algorithms; Naïve Bayes and support vector machine and deep learning algorithms such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Bidirectional LTSM, Bidirectional GRU, and Attention-based Bidirectional GRU were applied on datasets of more than 10,000 sentences collected from various blogs, websites, and tweets. Then, a combinational algorithm was also tested on the datasets and result showed that a combination of two algorithms, RNN +CNN outperformed all other applied algorithms.

Cahya et al. [19] investigated the use of SMOTE and Augmentation techniques to handle this imbalance and improve classification performance. The study utilizes the IndoBERT model for classification and employs TF-IDF for numerical representation of text data in conjunction with the SMOTE technique. The authors compare the performance of the IndoBERT model using both the SMOTE and Augmentation techniques. The results indicate that both techniques can address data imbalance and enhance the classification model's performance. The Augmentation technique improved performance by up to 20%, achieving an accuracy of 78%. However, the SMOTE technique demonstrated the best results, with a model accuracy of 82%.

Hung & Alias [20] examined the approaches and challenges associated with text sentiment analysis and emotion detection over the previous five years. They concluded that the trend in text-based emotion detection has shifted from early keyword-based methods to the use of machine learning and deep learning algorithms, which have outperformed unsupervised algorithms.

Deng and Ren [21] conducted a thorough review of recent developments in Text Emotion Recognition (TER), with a focus on methods utilizing deep neural networks. They reviewed the word embeddings, architecture, and training methods of various TER approaches, and addressed four key challenges: incompletely extractable emotional information in texts, TER in dialogues, fuzzy emotional boundaries, and the lack of large-scale, high-quality datasets. The authors concluded by discussing the challenges and future directions in this area.

Kusal et al. [22] conducted a systematic literature review of text-based emotion detection (TBED) from 2005 to 2021, studying 63 research papers from IEEE, Science Direct, Scopus, and Web of Science databases. They examined various TBED applications across different research domains, along with research challenges and future directions.

Salahudeen et al. [23] presented the findings of SemEval-2023 Task 12, a shared task on sentiment analysis for low-resource African languages using

Twitter dataset. Three subtasks were introduced subtask A is monolingual sentiment classification with 12 tracks which are all monolingual languages, subtask B is multilingual sentiment classification using the tracks in subtask A and subtask C is a zero-shot sentiment classification. the focus of their task was to leverage low-resource tweet data using pre-trained Afro-xlmr-large, AfriBERTa-Large, Bert-base-arabic-camelbert-da-sentiment (Arabic-camelbert), Multilingual-BERT (mBERT) and BERT models for sentiment analysis of 14 African languages. These subtasks contain gold standard multi-class labelled Twitter datasets from these languages. Results showed that Afro-xlmr-large model outperformed other models in most of the language datasets. Also, Nigeria Languages of Hausa, Igbo and Yoruba due to large volume of data in the languages performed better compared to other Languages.

[1] presented in detail, the methodology for generating a comprehensive code-mixed dataset which consists of 48,324 comments in low-resource languages of Roman Urdu, Roman Punjabi including English. Traditional natural language processing tools alongside the BERT tokenizer and contextual embedding modules developed were developed. These was used to preprocess this unstructured dataset. Dataset were trained, tested and evaluated on CMSA-mBERT model, machine learning models, deep learning models and transformer-based models. Results showed that CMSA-mBERT performed better than machine learning and deep learning models. Authors recommend that the CMSA-mBERT model's effectiveness and generalizability can be improved by increasing dataset to ensure balanced representation of all classes.

[3] presented a cross-lingual emotion classifier that transfers learning from resource-rich languages (English) to low and moderate resource languages. Authors compared two transfer learning approaches: one using parallel corpora to project annotations and another using direct transfer. The efficacy of these approaches is demonstrated across six languages: Farsi, Arabic, Spanish, Ilocano, Odia, and Azerbaijani. The authors also created annotated emotion-labeled resources for Farsi, Azerbaijani, Ilocano, and Odia; they investigated which emotion tags are universal, using a subset of Ekman's model (anger, fear, and joy) based on data availability. The study uses parallel corpora, including religious texts (Bible and Quran), to transfer emotions from English to the target languages. The results indicate that direct cross-lingual transfer outperforms annotation projection, and the study emphasizes the importance of overlap-genre training for better emotion classification.

In another study [10], authors addressed the challenges of sentiment and emotion detection in low-resource, code-switched languages, specifically focusing on Kenyan data. They analyzed a Kenyan code-switched data and evaluated them on four state-of-the-art (SOTA) transformer- based pre-trained models for sentiment and emotion classification, using supervised and semi supervised methods. Authors created a dataset, RideKE, which consist of a blend of Kenyan-accented English with Swahili and Sheng. The dataset contains over 29,000 tweets, labelled for sentiment analysis (positive, negative, neutral) and emotion detection (frustration, happy, angry, sad, empathy, fear, love, and surprise). Results showed that XLM-R outperformed other models; for sentiment analysis, XLM-R supervised model achieves the highest accuracy. In emotion analysis, DistilBERT supervised performed better while mBERT semi-supervised and AfriBERTa models presented the lowest accuracy.

[11] conducted a literature review of sentiment analysis in low-resource settings investigating which approach was mostly used for sentiment analysis tasks in low resource languages. Authors reported that social media is the most used platform for data collection. They reviewed previous works and survey papers on deep learning approaches applied for low-resource sentiment analysis, the most common sources of data for low-resource sentiment analysis and the challenges affecting this area. Authors also analyzed 56 relevant studies from 2018 – 2023 and found that the transfer learning approach is the most frequently used, followed by word embedding learning and machine translation systems.

Table 2 offers a comparative summary of emerging techniques and well-established state-of-the-art (SOTA) models utilized in emotion detection and sentiment analysis for low-resource languages. It emphasizes crucial elements like methodological approaches, dataset sizes, and evaluation metrics, facilitating a better comprehension of how contemporary techniques compare to conventional models in various linguistic settings.

Table 2: Comparison of emerging techniques with state-of-the-arts

Paper	SOTA model	Emerging techniques	Dataset/ size	Accuracy	Recall	F1 score
[3]	-	Cross-lingual transfer	English to low-resource Large	62%	61%	67%
[14]	-	Transfer learning	Kurdish Very small	57%	56%	57%
[10]	Deep learning, transformer-based XML-R	-	Code-switched small	69.2%	73%	66.1%
[15]	Transformer-based EEG	-	EEGEyeNet signal small	94.73%	92.55%	91.32%
[16]	Deep learning CNN LSTM	-	Urdu review large	96%	92%	91%
[1]	Machine learning, Deep learning, Transformer-based CMSA-mBERT	-	Code-mixed large	22.55%	21.06%	22.55%
[17]	Deep learning LSTM HFV+LSTM	-	SemEval-2014, Sentiment140, STS-Gold, very large	92%	93%	94%
[18]	Deep learning BiLSTM,BiGRU,CNN, RNN, LSTM+GRU RNN+CNN	-	Code-mixed, large	84%	85%	88%
[19]	Transformer-based IndoBERT, TF-IDF	Augmentation using SMOTE	Small	82%	85%	86%
[23]	-	Transformer-based AfriBERTa, AfroxLM-R, ARABIC BERT, mBERT	Code-mixed, large	69.3%	62%	64%

Currently, there is a surge in research in emotion detection and sentiment analysis of texts in low-resource languages. From review of past literatures, it is seen that traditional machine learning and deep learning state-of -the- arts techniques were mostly explored by researchers [1],[10],[15],[16],[17],[18]. Although, it is noted that authors who utilized deep learning techniques had good results but results would have been better if not for the problem of scarcity of data, imbalance dataset etc. However, it is worthy of note that, recent researches are utilizing the emerging new techniques [3], [14],[19],[23]. These techniques have been able to address some of the challenges encountered in this area. The emerging techniques when fine-tuned on small datasets have shown great improvements in model performances. They have also tried to tackle the problem of scarcity of dataset and solved the problem of imbalance dataset [19] for emotion detection and sentiment analysis.

3. Methodology

The aim of this paper is to explore the state-of-the-art techniques used for emotion detection and sentiment analysis of texts in low-resource languages. It identifies the open challenges in this domain. It also investigated emerging trends, improvements and proposes future directions. This paper is constructed using scholarly articles and reports published between 2019-2025 from reliable databases, especially Google Scholar, ScienceDirect, ACM Digital Library, Springer, and IEEE, among others. In each of these databases, the search string "emotion detection and sentiment analysis of texts in low-resource languages OR “emotion detection and sentiment analysis of texts" was employed to retrieve relevant articles. These articles were selected based on the criteria that they were written in English, related to the field of Natural Language Processing (NLP), and were published in reputable journals or conference proceedings.

4. State-of-the-Art Techniques for Emotion Detection and Sentiment Analysis of Texts in Low-Resource Languages

Emotion detection and sentiment analysis are vital tasks in NLP. However, expressing emotions through text lacks a standardized approach, meaning that individuals can express emotions in various ways—some in an open manner, others in a more reserved fashion, and some through rational expressions [24]. As a result, researchers face

challenges in developing a universal strategy that works effectively across all contexts. Emotion detection and sentiment analysis have a broad range of applications and can be implemented using diverse techniques such as deep learning, machine learning, hybrid approaches, and lexicon-based methods.

4.1 Machine Learning-based Approaches

Machine learning algorithms are often used for classification of feelings. Within artificial intelligence (AI), machine learning is the branch of study that focuses on enabling computers or machines to learn and make decisions or predictions without explicit programming. The process involves creating models and algorithms that can examine and construe vast quantities of data to reveal patterns, connections, and revelations. [25] notes that machine learning (ML) techniques are also used extensively in statistical analysis. However, machine learning methods for low-resource language sentiment analysis and emotion detection are examined subsequently.

a. Decision Tree Algorithm

Decision trees are widely used supervised learning techniques for classification, regression, and prediction tasks. These models categorize emotional and sentimental content in texts for emotion detection and sentiment analysis in low-resource languages. The internal nodes represent attributes, while the leaf nodes represent class labels. Decision rules are typically based on if-then-else conditions. The more branches a tree has, the more complex the rules, which can improve model performance [26].

The process for using decision trees in emotion detection involves several steps. First, a labelled dataset of texts in the target language is collected, indicating the emotional or sentimental content. Next, features such as word frequencies, part-of-speech tags, and sentiment scores are extracted from the texts. These features are then used to train the decision tree model. The model learns to classify the text into different emotional or sentimental categories based on the extracted features. After training, the model is tested using a different dataset to evaluate its performance. Metrics like accuracy, precision, and recall are commonly used for assessment. Once trained and tested, the model can be deployed to classify the emotional and sentimental content of new texts.

Authors in [27] demonstrated that decision trees perform well for sentiment analysis and emotion detection in low-resource languages like Telugu, outperforming other machine learning models such as Naive Bayes and Support Vector Machines.

One advantage of decision trees is their ability to handle noisy and sparse data, a common characteristic of low-resource languages. Additionally, decision trees are interpretable, making it easier to understand how the model makes decisions and which features are critical for classification. However, decision trees may overfit if the model becomes too complex or the dataset is too small. Pruning and regularization techniques are essential to prevent overfitting and ensure robust performance.

b. Support Vector Machines (SVM)

SVM is a supervised machine learning technique applied to both classification and regression tasks [24]. SVM is effective for emotion detection and sentiment analysis in low-resource languages, particularly when dealing with high-dimensional, sparse data typical of text classification tasks. The process begins with collecting a labelled dataset of texts and extracting features such as word frequencies, part-of-speech tags, and sentiment scores. The SVM model is then trained using these features to identify a hyperplane that separates the texts into different emotional or sentimental categories. SVM's kernel trick allows it to handle non-linearly separable data by mapping it to a higher-dimensional feature space.

Authors in [2] showed that SVM achieves high accuracy in sentiment analysis and emotion detection with even small amounts of labeled data. The study found that SVM outperformed other models like Random Forest and Naive Bayes for Bengali text analysis. However, SVM models can be computationally expensive and require careful hyperparameter optimization to achieve optimal performance. Regularization techniques should also be applied to prevent overfitting, especially when working with limited data in low-resource languages.

c. Naive Bayes Algorithm

The Naive Bayes (NB) algorithm is commonly employed for classification tasks. It can classify emotional and sentimental content in texts by estimating conditional probabilities for each feature given the class label. The process involves collecting datasets, extracting features such as word frequencies, part-of-speech tags, and sentiment scores, and training the model to classify texts into various emotional categories. The model is then evaluated based on performance metrics such as accuracy, precision, and recall.

The Naive Bayes (NB) algorithm is commonly employed for classification tasks. It can classify emotional and sentimental content in texts by estimating conditional probabilities for each feature given the class label. The process typically involves first, collection of datasets, performing feature extraction on the texts that will be fed into the Naive Bayes model with common features like word frequencies, part-of-speech tags, and sentiment scores. After learning the conditional probability of each feature given each class label and the prior probability of each class label, the model is tested on a different dataset, and performance metrics like accuracy, precision, and recall are used to assess the model's effectiveness.

[24] demonstrated that Naive Bayes performed well in sentiment analysis and emotion detection in languages with limited labelled data, outperforming models like Decision Trees and SVM in Kannada language.

d. Random Forest

Random Forest (RF) is an ensemble learning technique that constructs multiple decision trees for classification tasks. It is effective in handling various data types and often produces robust results, even in low-resource settings [22]. Random Forest is less prone to overfitting than individual decision trees and can handle non-linear relationships between features and target variables.

A study by [28] shows that Random Forest outperformed Naive Bayes and SVM in Malayalam language sentiment analysis.

e. eXtreme Gradient Boosting (XGBoost)

XGBoost is an advanced machine learning algorithm based on the gradient boosting technique, designed to improve model efficiency and speed [29]. XGBoost builds an ensemble of decision trees, iteratively adding new trees to correct the errors made by the previous ones. It can be parallelized for faster training and uses regularization to prevent overfitting.

XGBoost has shown promise in sentiment analysis and emotion detection tasks, but it requires careful hyperparameter tuning to ensure optimal performance.

f. Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) are machine learning algorithms inspired by the human brain, consisting of interconnected nodes that process and transmit data. ANN can be applied to sentiment analysis and emotion detection by selecting appropriate architectures, such as feedforward, recurrent, or convolutional neural networks. ANN models are trained on labeled datasets, adjusting weights and biases to minimize errors between predicted and actual labels. This was demonstrated in [30],[43].

4.2 Deep Learning Algorithms

According to [11], deep learning embodies the very best of artificial intelligence (AI). Deep learning is a subfield of machine learning that leverages multi-layered neural networks to automatically extract features from data, making it ideal for tasks like emotion detection and sentiment analysis in low-resource languages. Deep learning models can learn abstract representations of data and have achieved state-of-the-art performance in NLP tasks such as text classification and sentiment analysis. Below are some examples of deep learning techniques.

a. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a class of learning algorithms widely used in applications such as computer vision, sentiment analysis, and natural language processing (NLP) tasks. Transfer learning can modify pre-trained models to adapt to new languages, especially in low-resource languages with limited labeled data [31].

To apply CNNs to sentiment analysis and emotion detection in low-resource languages, a labeled dataset must first be collected, with each text sample annotated with the corresponding emotion or sentiment polarity. Afterward, data preprocessing is performed to remove stop words and apply tokenization, stemming, or lemmatization. Word embeddings are then employed to transform the pre-processed text into low-dimensional word representations that capture the syntactic and semantic relationships between words. For low-resource languages, pre-trained word embeddings such as GloVe or Word2Vec can be utilized, if available. Otherwise, custom word embeddings can be trained using the available dataset.

Convolutional layers are employed to extract features from the text data, applying filters to the word embeddings. These filters convolve over the input data to create feature maps. Pooling layers are then used to down-sample the feature maps, retaining the most important information despite the reduced spatial size. The pooled feature maps are passed through one or more fully connected layers, which are responsible for mapping the feature vectors to their corresponding sentiment polarity or emotion for classification. The model is trained using

a loss function like cross-entropy and an optimizer like stochastic gradient descent (SGD) on the labelled dataset. Finally, the CNN's performance is evaluated using a held-out test set, this was presented in [16], [32], [43], with results outperforming baseline models.

In summary, CNNs can effectively be used for sentiment analysis and emotion detection in low-resource languages by adapting pre-trained models through transfer learning and training them on labeled datasets. The use of pre-trained word embeddings and convolutional layers enables the model to extract meaningful features, even in the absence of large amounts of labelled data.

b. Deep Belief Networks (DBN)

Deep Belief Networks (DBNs) leverage the hierarchical structure of training data to construct abstract representations of datasets. A classic DBN is composed of stochastic models with visible and hidden layers. Under specific conditions, lower-level units, other than those in the top two layers, may be interdependent [24]. DBNs are deep learning neural networks used for unsupervised tasks such as feature learning, dimensionality reduction, and generative modeling. DBNs can automatically extract relevant features from text data for emotion detection and sentiment analysis in low-resource languages, without the need for explicit feature engineering.

DBNs are typically trained using an energy-based generative model, the Restricted Boltzmann Machine (RBM). Through training, DBNs develop the ability to represent input text data in a distributed, hierarchical manner, where each network layer captures progressively more abstract features. Once trained, DBNs can perform emotion detection and sentiment analysis by passing new text data through the network. The output of the DBN can then be used to predict the sentiment or emotion of the input text [33]. DBNs are particularly valuable in low-resource languages due to their robustness in handling noisy, missing, or varied text data, making them suitable for languages with limited linguistic resources and less standardized grammatical structures [34].

c. Deep Auto Encoder (DAE)

In a Deep Auto Encoder (DAE), the number of neurons in the output layer is equal to the number of neurons in the input layer [35]. The DAE consists of an encoder network, which maps the input data to a lower-dimensional latent space, and a decoder network, which reconstructs the original input from the latent representation. During training, the DAE minimizes the difference between the input data and its reconstruction by adjusting the network's weights and biases.

Once trained, the encoder network can be used to extract features from new text data for sentiment analysis and emotion detection. These extracted features can be fed into a classifier network, such as a random forest or support vector machine (SVM), to predict the sentiment or emotion of the input text. The DAE's primary goal is to reconstruct its inputs while performing dimensionality reduction, simplifying the process of feature extraction. This was presented in [13]

d. Recurrent Neural Networks (RNN)

Recurrent Neural Networks (RNNs) are particularly well-suited for natural language processing tasks because they can capture temporal dependencies in word sequences, such as sentence structure and context [36]. RNNs are used to model relationships between words in a sentence and predict the associated sentiment or emotion. Long Short-Term Memory (LSTM) networks, a type of RNN, excel in handling long-term dependencies in sequences, making them ideal for capturing complex relationships in text data.

To use RNNs for sentiment analysis and emotion detection, the network is trained on a labelled dataset, where sentiment or emotion labels act as the target output. During training, the RNN learns to predict the corresponding sentiment or emotion label by identifying relationships between words in the input text. Once trained, the RNN can predict the sentiment or emotion of new, unseen text by processing each word sequentially, updating its hidden state at each step, and producing a final sentiment or emotion prediction. This is demonstrated in [18],[37].

e. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are a specific type of RNN frequently employed in natural language processing tasks such as sentiment analysis and emotion detection [31]. LSTMs are capable of selectively remembering or forgetting past inputs over time, which makes them particularly effective for processing sequences like text.

For sentiment analysis and emotion detection in low-resource languages, LSTM models are trained on labelled datasets, where each text sample is annotated with a sentiment or emotion label. The text data is pre-processed and converted into a numerical format using techniques like padding (to ensure consistent input length) and tokenization (which converts words into numerical tokens). After training, the LSTM can predict the sentiment or

emotion of new text by applying the same preprocessing steps and passing the text through the model. Authors in [17], [38] demonstrated this technique.

f. Restricted Boltzmann Machines (RBM)

Restricted Boltzmann Machines (RBMs) are a type of generative neural network used for unsupervised learning tasks, including sentiment analysis and emotion detection in low-resource languages. The hidden units in an RBM learn to represent higher-level features of input data by reducing the system's energy through a training process called contrastive divergence.

To apply RBMs to sentiment analysis or emotion detection, the network is trained on a large corpus of text data in the target language. The RBM learns to identify patterns in the text that indicate sentiment or emotion. After training, the RBM can predict the sentiment or emotion of new text samples by propagating the input through the network and using the hidden layer's output as features for a classification model, such as logistic regression or SVM.

One advantage of RBMs is their ability to be stacked to form deep belief networks (DBNs). They can also adapt pre-trained models using transfer learning and be trained on labelled datasets. This approach has been shown to be effective in various NLP tasks, including sentiment analysis and emotion detection [39].

4.3 Lexicon-Based Approach

The lexicon-based method uses a lexicon where each word is associated with a sentiment value, either positive or negative. The overall sentiment of a document is determined by calculating the mean or total of these sentiment values [2]. This method is divided into two parts: the dictionary-based approach, which looks up opinion words in a dictionary and finds synonyms and antonyms, and the corpus-based approach, which starts with a list of opinion words and expands it by using a large corpus to find context-specific words. Semantic or statistical techniques can aid this expansion [20]. The NRC lexicon approach, used by [40] for classifying Nigerian emotions during the COVID-19 lockdown, and [49] demonstrated the effectiveness of this method in NLP tasks.

4.4 Hybrid Approach

A hybrid approach enhances performance in low-resource languages by combining various techniques, such as statistical, rule-based, and machine learning methods. This combination helps address the limitations of individual approaches. One way to do this is to combine machine learning or deep learning models with lexicon-based techniques (using manually made sentiment dictionaries). Integrating transfer learning with deep learning. Using rule-based systems in conjunction with neural networks to handle code-switching or capture particular linguistic nuances. By using the advantages of several approaches, such as the interpretability of lexicons with the generalization capacity of neural networks, hybrid models can frequently attain greater accuracy. For example, [17] introduced a hybrid approach for analyzing sentiment in reviews by combining different models and incorporating aspect-related features to create unique feature vectors. Additionally, [34] combined fastText, BERT, and a 1D-CNN+BiLSTM model for sentiment analysis in low-resource Albanian. The best results were achieved when combining BiLSTM with an attention mechanism. [25] also employed a hybrid model combining a lexicon-based approach with a sentiment rule-based engine (VADER) for sentiment classification. This approach was also utilized [16], [1], [18].

5. Open Problems in Low-Resource Languages

As social media continues to be the primary source of datasets for emotion recognition and sentiment analysis [34], researchers often focus on emotions like happiness, anger, sadness, and fear, as these are the most frequently studied. While ML and deep learning models have shown strong performance, there are still areas for improvement [24] Below are some open challenges in this area.

5.1 Poor Analytical Accuracy Due to Limited Datasets

Low-resource languages typically suffer from limited data sources, insufficient research, and inadequate computational resources, which impede the use of Deep Neural Network (DNN) models for detecting harmful and offensive content on social media [41]. Achieving accurate analytical results remains a challenge due to small dataset sizes and limited classification accuracy, as highlighted by [42].

5.2 Complicated Morphology of Native Languages

Some native languages present challenges in emotion detection due to their complex morphology. For instance, detecting emotions from Urdu text is more challenging than from English due to Urdu's complex structure and its absorption of influences from Arabic, Turkish, English, Persian, and Sanskrit

[37]. This complexity adds to the difficulty of linguistic analysis, as Urdu shares morphological similarities with Hindi and includes over 40 verb and noun forms. [5] also mentioned these issues in their work.

5.3 Lack of Labeled, Unlabeled, and Imbalanced Datasets

[40] noted that there are very few online datasets for research in Emotion detection and sentiment analysis. Many of them contain imbalanced dataset and do not meet accurate standards for machine learning algorithms. The lack of labelled and unlabelled datasets hampers emotion detection and sentiment analysis research, especially in low-resource languages [11].

5.4 Context, Misspellings, Mocking, and Multiple Emotions in a Single Utterance

The rise of unstructured text data on the internet has led to challenges in emotion and sentiment analysis, as misspellings, slang, and complex contexts make it difficult to analyze mood and emotion [43]. [24] pointed out, texts may include ambiguities, such as "y are u soooo lazy?" where the tone is unclear, and it is hard to determine if the individual is angry. Therefore, detecting emotions and sentiments from texts poses a lot of difficulties because of ambiguities of the texts. This was also mentioned in [44],[45].

5.5 Limited Availability of Language-Specific Toolkits and Annotated Data

[46] highlighted that the identification of emotions and sentiment analysis tasks face challenges due to insufficient resource availability. Certain statistical methods, require a large dataset that has been annotated. Collecting data is not difficult, but labelling a large dataset is time-consuming and unreliable. Another challenge that emerges with materials is that the majority of them are only readily available in the English language [47]. Conducting sentiment analysis and emotion recognition in languages other than English presents a significant challenge as well as a potentially beneficial opportunity for scholars [22]. Furthermore, certain lexicons and corpora are limited to just one specific domain, making their reuse in different areas challenging.

5.6 Ambiguity in Lexical and Syntactical Terms

The ambiguity in language, especially with slang/abbreviation/acronym and context-dependent meanings, makes sentiment and emotion detection even more difficult [40], [46]. In addition, [48] suggest that in contrast to facial expression and speech recognition, a text sentence lacks the ability to define itself due to its lack of flavour. Due to the complexity and ambiguous nature of the text, determining the emotions within it proves to be a challenging endeavour. Recognizing the emotion of a text is challenging because each word can carry a different meaning and morphological form, making it a difficult task.

6. Emerging Trends and Improvements in Emotion Detection and sentiment Analysis of texts in Low-resource Languages

Emotion detection and sentiment analysis in low-resource languages present unique challenges due to a scarcity of annotated data, linguistic tools, and established research [49]. However, significant progress is being made to address these challenges basically through the evolution of emerging trends in deep learning and transfer learning, that leverage various computational and linguistic strategies.

6.1 Transfer Learning and Pre-trained Multilingual Models

Models such as mBERT (multilingual BERT), XLM-R, and AfroBERTa are pre-trained on large text corpora in a variety of languages. This enables them to acquire general linguistic representations and patterns that can be refined for particular low-resource languages even in the absence of a large amount of target-language data. The use of extensive, language-specific dataset is greatly reduced by this method. With competitive results in sentiment and emotion detection, it bootstraps performance in low-resource environments by utilizing the knowledge gained from high-resource languages. For instance, mBERT has demonstrated effectiveness in multilingual sentiment analysis tasks [1], outperforming conventional machine learning techniques. This method has been employed in numerous research projects [3], [14], [23] and have achieved good performance.

6.2 Data Augmentation and Synthetic Data Generation

To address the lack of data, researchers are now using methods to artificially generate new datasets or enlarge existing ones. Creating new paraphrases by translating text from a low-resource language to a high-resource language and back again is known as back-translation. Synonym replacement: Creating diverse sentences by substituting words with their synonyms. Using Large Language Models (LLMs): LLMs can produce synthetic datasets by being asked to produce text in the target low-

resource language with particular sentiment or emotional labels. In terms of producing relevant and varied datasets, this holds great promise. Authors in [43] demonstrated this in technique. Synthetic Minority Over-sampling Technique (SMOTE): Synthetic Minority over-sampling Technique is a statistical technique for increasing the number of cases of imbalance data in a dataset in a balanced way. This can help improve class imbalance by augmenting dataset to balance [42]. By reducing overfitting brought on by small datasets, artificially expanding the amount and variety of training data aid models in learning more resilient representations and improving generalization. Further improving model performance can be achieved by incorporating regional language nuances into prompts for the generation of synthetic data. This was shown in [19].

6.3 Cross-Lingual Approaches

These techniques seek to impart knowledge across languages without translating the target text directly. Learning word representations that are consistent across languages, or cross-lingual word embeddings, enables models developed in one language to be applied to another [3]. Both zero-shot and few-shot learning involve training a model on a high-resource language (zero-shot) or using a limited number of examples (few-shot). [50], [51] demonstrated the application of this model and also in [49]. Nevertheless, ZSL is a fast-evolving field, and advancements are constantly enhancing its accuracy and efficiency.

6.4 Focus on Code-Switching and Dialectal Variations

Owing to the complexity of code-switched text and regional dialects, researchers are now creating specialized preprocessing methods and models. Language identification in mixed sentences is one example of this. For words written in a different script, utilize transliteration. In mixed-language contexts, contextual embedding models are more effective at capturing word meanings. With code-switching common in informal communications and real-world social media, customized methods for these linguistic phenomena result in more accurate sentiment and emotion detection [11].

6.5 Development of Language-Specific Resources

Although transfer learning has advanced, efforts are still being made to create foundational resources such as sentiment lexicons, annotated datasets, and monolingual corpora for particular low-resource languages. This frequently calls for cooperation and community involvement. By providing useful bases for model evaluation and training, these resources will enable more in-depth linguistic comprehension and tailored solutions.

Lastly, even though there are still many obstacles to overcome, new developments in sentiment analysis and emotion detection for low-resource languages are shifting toward utilizing the strength of pre-trained multilingual models, creative data augmentation strategies, and hybrid approaches to close the resource gap. It is becoming more and more possible to create efficient NLP applications for a larger variety of linguistic communities due to these developments. Authors in [52] presented this technique.

7. Future Directions in Emotion Detection and Sentiment Analysis in Low-resource Languages

Sentiment analysis and emotion detection in low-resource languages is a dynamic field with immense potential. Building on present trends, future directions will probably concentrate on resolving outstanding issues and increasing the technologies' usefulness in real-world scenarios. Key directions for the future are as follows:

7.1 Improved Cross-Lingual and Multilingual Transfer Learning

Deeper Cross-Lingual Alignment: More investigation into more complex methods for lining up representations between languages, going beyond shared embeddings to accurately capture subtle emotional and semantic correspondences. This might entail investigating more complex contrastive learning or adversarial training techniques for cross-lingual alignment. Also, creating "Language-Agnostic" models using meta-learning techniques, could be able to make these models comprehend and extract sentiments and emotions from any language with little to no prior exposure. Continuous Learning for New Languages: Rather than requiring total retraining, system designers can create systems that can continuously learn and adapt to new low-resource languages as they come across more data.

7.2 Advanced Data Generation and Augmentation

Leveraging Generative AI for Synthetic Data: Using large generative models, such as advanced LLMs, to generate diverse and incredibly realistic synthetic datasets that are specific to low-resource languages and their distinct linguistic characteristics (for example, slang, cultural references, code-switching patterns). This involves generating texts with particular emotional nuances. This technique will help increase the number of dataset and improve performance.

7.3 Semi-supervised and Self-supervised Learning

In order to reduce the need for expensive human annotation, more advanced semi-supervised and self-supervised learning approaches are being investigated. These methods can efficiently utilize large amounts of unlabeled text data in low-resource languages.

7.4 Multimodal Data Integration

To give emotion detection a richer context, other modalities, such as speech and images, can be combined with text. Analyzing spoken language intonation or facial expressions in video clips, for example, can be done even when the primary data is text in a low-resource language.

7.5 Towards More Nuanced Emotion Detection

Fine-grained Emotion Classification: In order to detect a greater variety of emotions, fine-grained emotion classification should go beyond just simple positive, negative, and neutral emotions (eg., joy, anger, sadness, surprise, fear, disgust, anticipation, trust) as well as even more detailed emotional states that are pertinent to particular cultural contexts. Contextual Emotion Understanding: Creating models that can comprehend a document's or conversation's emotional context rather than just individual sentences is crucial for capturing irony, sarcasm, and implicit emotions in low-resource environments. Developing models that can different between irony and sarcasm from a piece of text will be a step forward in this field. Cultural Nuance in Emotion: Recognizing and demonstrating how emotions are expressed and interpreted differently in various languages and cultures. This calls for potentially language specific emotion taxonomies as well as culturally sensitive annotation guidelines.

7.6 Enhanced Code-Switching Handling

Creating increasingly complex models that can process and comprehend code-switched text with ease, even when there are different degrees of language mixing and structural changes, is one way to demonstrate robustness to linguistic challenges. Dialect and Colloquial Language Adaptation: Developing models that can comprehend and adjust to regional dialects, slang, and informal language that are common in low-resource communities and frequently differ greatly from formal written forms. Managing Ambiguity and Sarcasm: sophisticated methods for identifying and deciphering intricate linguistic phenomena, such as irony, sarcasm, and subtle humor, which can be difficult to understand even in high- resource languages.

7.7 Explainable Artificial Intelligence (XAI)

Developing techniques to improve the transparency and interpretability of sentiment and emotion model predictions, particularly for stakeholders in low-resource contexts, is known as Explainable AI (XAI) for Low-Resource NLP. In addition to facilitating improved debugging and comprehension of model limitations, this fosters trust. It is possible to modify and improve attention-based explanation techniques, such as Local Interpretable Model-agnostic Explanation (LIME) and Shapley additive explanation (SHAP) for low-resource languages [53].

7.8 Applications for Real-time and Edge Computing

Creating more compact, effective model architectures that can be implemented on devices with limited resources (e.g. mobile phones, Internet of Things devices (IOT) in low-resource settings, allowing for sentiment and emotion analysis in real-time without continuous cloud connectivity. Also Federated Learning: Investigating federated learning strategies that allow for privacy-preserving learning from data stored on separate devices in various locations by training models cooperatively on decentralized data. An ensuring that tools for sentiment analysis and emotion detection can operate efficiently even in places with limited or no internet connectivity.

By focusing on these directions, the field can advance towards sentiment analysis and emotion detection systems for the great majority of languages spoken worldwide that are more reliable, accurate, moral, and practically useful. This will ultimately encourage greater digital inclusion and empower diverse communities.

8. Conclusion

In conclusion, the rise of internet technology has significantly accelerated the development of new trends in computing, particularly in the fields of sentiment and emotion analysis. This paper has explored the state-of-the-art methodologies for emotion and sentiment analysis in low-resource languages, highlighting both the challenges and potential improvements in this upcoming area of natural language processing (NLP). The scarcity of labeled data, the intricate linguistic structures, data imbalance and the lack of computational resources remain key barriers to achieving high-accuracy emotion and sentiment detection in these languages. However, substantial progress has been made by leveraging advanced machine learning techniques, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), deep belief networks (DBNs), and lexicon-based approaches. The application of transfer learning, pre-trained embeddings, and hybrid models has proven particularly valuable in overcoming the limitations of small datasets, improving both the efficiency and effectiveness of sentiment analysis. Additionally, data augmentation techniques have enhanced model robustness, allowing for better generalization and performance in languages with

complex morphology. These new trends have led to significant improvements in emotion and sentiment detection, even in low-resource languages, and pave the way for more inclusive NLP applications.

Despite the challenges, this study demonstrates that progress is being made in developing more accurate models for low-resource languages, enabling these languages to participate more fully in the rapidly advancing field of sentiment and emotion analysis. As technology continues to evolve, we can anticipate the emergence of more sophisticated processing tools and models that will not only benefit high-resource languages but also drive further advancements in low-resource language processing.

This article provides a comprehensive overview of the current state of the art techniques for emotion detection and sentiment analysis of text in low-resource languages, challenges, open problems and emerging trends and improvements for NLP tasks emotion. It serves as a valuable resource for researchers seeking to understand the latest methodologies, classification models, available resources, and existing limitations in the field. The insights presented here aim to support and guide future research and development efforts.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] M. K. Nazir, C. N. Faisal, M. A. Habib and H. Ahmad, "Leveraging Multilingual Transformer for Multiclass Sentiment Analysis in Code-Mixed Data of Low-Resource Languages," in *IEEE Access*, vol. 13, pp. 7538-7554, 2025, doi: 10.1109/ACCESS.2025.3527710.

[2] R. Ahamad and K. N. Mishra, “Exploring sentiment analysis in handwritten and E-text documents using advanced machine learning techniques: a novel approach,” *Journal of Big Data*, vol. 12, no. 11, pp. 1–38, 2025. [Online]. Available: https://doi.org/10.1186/s40537-025-01064-2

[3] S. Tafreshi, S. Vatsal, and M. Diab, "Emotion Classification in Low and Moderate Resource Languages," *arXiv preprint arXiv:2402.18424v2*, 2024.

[4] A. Pingle, A. Vyawahare, I. Joshi, R. Tangsali, G. Kale, and R. Joshi, "Robust Sentiment Analysis for Low Resource Languages Using Data Augmentation Approaches: A Case Study in Marathi," *arXiv preprint arXiv:2310.00734*, Oct. 2023. [Online]. Available: https://arxiv.org/abs/2310.00734

[5] A. Ali, M. Khan, K. Khan, R.U. Khan, and A. Aloraini, “Sentiment Analysis of Low-Resource Language Literature Using Data Processing and Deep Learning,” *Comput. Mater. Contin.*, vol. 79, no. 1, pp. 713–733, 2024. [Online]. Available: https://doi.org/10.32604/cmc.2024.048712

[6] R. Hameed, S. Ahmadi, and F. Daneshfar, "Transfer Learning for Low-Resource Sentiment Analysis," ArXiv, abs/2304.04703, vol. 37, no. 4, Apr. 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2304.04703.

[7] 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. 141218-141229, Oct. 2023, doi: 10.1109/ACCESS.2023.3322101.

[8] A. Yusup, D. Chen, Y. Ge, H. Mao, and N. Wang, "Resource Construction and Ensemble Learning based Sentiment Analysis for the Low-resource Language Uyghur," *J. Internet Technol.*, vol. 24, no. 4, pp. 18-27, 2023, doi: 10.53106/160792642023072404018.

[9] K. Chattu and S. D, "Sentiment Classification of Low Resource Language Tweets Using Machine Learning Algorithms," *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, Namakkal, India, 2023, pp. 1055-1061, doi: 10.1109/ICECAA58104.2023.10212247.

[10] N. A. Etori and M. L. Gini, "RideKE: Leveraging Low-Resource, User-Generated Twitter Content for Sentiment and Emotion Detection in Kenyan Code-Switched Dataset," *arXiv preprint arXiv:2502.06180*, 2025.

[11] Y. Aliyu, A. Sarlan, K. U. Danyaro, A. S. B. A. Rahman, and M. Abdullahi, “Sentiment Analysis in Low-Resource Settings: A Comprehensive Review of Approaches, Languages, and Data Sources,” *IEEE Access*, vol. 12, pp. 66883-66909, May 8, 2024. doi: 10.1109/ACCESS.2024.3398635

[12] A. A. Maruf, F. Khanam, M. M. Haque, Z. M. Jiyad, M. F. Mridha, and Z. Aung, “Challenges and Opportunities of Text-Based Emotion Detection: A Survey,” *IEEE*

Access, vol. 12, pp. 18416-18451, 19 January 2024, doi: 10.1109/ACCESS.2024.3356357.

[13] A. R. Jim, M. A. R. Talukder, P. Malakar, M. M. Kabir, A. Nur, and M. F. Mridha, "Recent advancements and challenges of NLP-based sentiment analysis: State-of-the-art review," *Natural Language Processing Journal*, vol. 6, 2024, Art. no. 100059. [Online]. Available: https://doi.org/10.1016/j.nlp.2024.100059

[14] F. Daneshfar, "Enhancing Low-Resource Sentiment Analysis: A Transfer Learning Approach," *Passer Journal*, vol. 6, no. 2, pp. 265-274, 12 May 2024. doi: 10.24271/PSR.2024.440793.1484

[15] J. Wang and C. Zhang, "Cross-modality fusion with EEG and text for enhanced emotion detection in English writing," *Frontiers in Neurorobotics*, vol. 18, 2025, Art. no. 1529880, doi: 10.3389/fnbot.2024.1529880.

[16] N. Ahmed, R. Amin, H. Aldabbas, M. Saeed, M. Bilal, and H. Song, "A Novel Approach for Sentiment Analysis of a Low Resource Language Using Deep Learning Models," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 2024, doi: 10.1145/3696789.

[17] G. Kaur and A. Sharma, "A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis," *Journal of Big Data*, vol. 10, no. 5, Jan.2023. DOI: https://doi.org/10.1186/s40537-022-00680-6

[18] K. B. Muhammad and S. M. A. Burney, "Innovations in Urdu Sentiment Analysis Using Machine and Deep Learning Techniques for Two-Class Classification of Symmetric Datasets," *Symmetry*, vol. 15, no. 5, p. 1027, May 2023, doi: 10.3390/sym15051027.

[19] L. D. Cahya, A. Luthfiarta, J. I. T. Krisna, S. Winarno, and A. Nugraha, "Improving Multi-label Classification Performance on Imbalanced Datasets Through SMOTE Technique and Data Augmentation Using IndoBERT Model," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 5, pp. 290-298, 2023. doi: 10.25077/TEKNOSI.v9i3.2023.290-298

[20] L. P. Hung and S. Alias, "Beyond sentiment analysis: a review of recent trends in text-based sentiment analysis and emotion detection," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 27, no. 1, pp. 84-95, Jan.,2023. doi: 10.20965/jaciii.2023.p0084.

[21] J. Deng and F. Ren, "A survey of textual emotion recognition and its challenges," IEEE Transactions on Affective Computing, vol. 14, no.4, pp. 49-67, Jan.2023. doi:10.1109/TAFFC.2021.3053275. [Online]. Available: https://doi.org/10.1109/TAFFC.2021.3053275

[22] S. Kusal, S. Patil, J. Choudrie, K. Kotecha, D. Vora, and I. Pappas, "A Review on Text-Based Emotion Detection -- Techniques, Applications, Datasets, and Future Directions," Apr.2022. [Online]. Available: http://arxiv.org/abs/2205.03235.

[23] S. A. Salahudeen, F. I. Lawan, A. M. Wali, A. A. Imam, A. R. Shuaibu, A. Yusuf, N. B. Rabiu, M. Bello, S. U. Adamu, S. M. Aliyu, M. S. Gadanya, S. A. Muaz, M. S. Ahmad, A. Abdullahi, and A. Y. Jamoh, "HAUSANLP at SemEval-2023 Task 12: Leveraging African Low Resource Tweet Data for Sentiment Analysis," in *Proc. 17th Int. Workshop on Semantic Evaluation (SemEval-2023)*, Toronto, Canada, Jul. 2023, pp. 50–57.

[24] V. Jain and M. Parmar, "A review on emotion and sentiment analysis using learning techniques," *International Journal for Research in Applied Science & Engineering Technology (IJRASET),* vol. 10, no. 11, pp. 365-405, Nov. 2022. [Online]. Available: https://doi.org/10.22214/ijraset.2022.47337.

[25] A. B. Braiek, Z. Neji, and B. Salem, "Sentiment Analysis Classification for Text In Social Media: Application To Tunisian Dialect," vol. 12, no. 2, pp. 313–326, Apr. 2023. [Online]. Available: https://doi.org/10.5121/ijci.2023.120223.

[26] K. V. Ramana, N. R. Varma, and R. P. Reddy, "Emotion detection and sentiment analysis in Telugu texts using decision trees," *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* Nov. 2019, pp. 47-51, IEEE.

[27] R. Jadhav and V. M. Gadre, "A deep learning approach for sentiment analysis in low-resourced languages using transfer learning," in *P. N. Suganthan, K. Yang, & T. Huang (Eds.), Neural Information*

Processing: 27th International Conference, ICONIP 2020, Bangkok, Thailand, November 23–27, 2020, Proceedings, Part IV, Springer, 2020, pp. 382-393. [Online]. Available: https://doi.org/10.1007/978-3-030-63823-8_36

[28] R. Aswathy, P. T. Anand, and P. M. Shajan, "Emotion detection and sentiment analysis in Malayalam texts using random forest*," in 6th International Conference on Advanced Computing and Communication Systems (ICACCS)* 2020, pp. 1-6.

[29] A. Majeed, M. O. Beg, U. Arshad, and H. Mujtaba, "Deep-EmoRU: Mining emotions from Roman Urdu text using deep learning ensemble," *Multimedia Tools and Applications*, vol. 81, pp. 43163–43188, 2022. doi: https://doi.org/10.1007/s11042-022-13783-x

[30] F. Ullah, X. Chen, S. B. H. Shah, S. Mahfoudh, M. A. Hassan, and N. Saeed, "A novel approach for emotion detection and sentiment analysis for low resource Urdu language based on CNN-LSTM," *Electronics (Switzerland),* vol. 11, no. 24, pp.1-19, Dec. 2022. doi: https://doi.org/10.3390/electronics11244096

[31] V. Tejaswini, K. S. Babu, and B. Sahoo, "Depression detection from social media text analysis using natural language processing techniques and hybrid deep learning model," *ACM Transaction on Asian and Low-Resource Language Information Processing,* Just Accepted, 2022. doi: https://doi.org/10.1145/3569580

[32] K. Shanmugavadivel, V. E. Sathishkumar, S. Raja, T. B. Lingaiah, S. Neelakandan, and M. Subramanian, "Deep learnin- based sentiment analysis and offensive language identification on multilingual code-mixed data," *Scientific Reports*, vol. 12, pp.1-12, article no. 13551, Dec. 2022. DOI: https://doi.org/10.1038/s41598-022-26092-3

[33] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: a comparative study," *Electronics,* vol. 9, article no.3, pp. 483, March. 2020. DOI: https://doi.org/10.3390/electronics9030483

[34] Z. Kastrati, L. Ahmedi, and A. Kurti, "A deep learning sentiment analyzer for social media comments in low-resource languages," *Electronics,* vol. 10, article no. 1133, pp.1-19, May.2021. DOI: 10.3390/electronics10101133.

[35] A. Ahmed, D. Abdi, S. Ahmed, and S. Hussein, "An Artificial Neural Network Model for Emotion Detection in Low-Resource Languages: A Case Study on Somali Language," *Journal of Intelligent & Fuzzy Systems,* vol. 43, no. 3, pp. 2853-2865, March.2022.

[36] M. G. Huddar, S. S. Sannakki, and V. S. Rajpurohit, "Attention-based multi-modal sentiment analysis and emotion detection in conversation using RNN," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no.6, pp. 112-121, June. 2021. DOI: https://doi.org/10.9781/ijimai.2020.07.004.

[37] E. Omara, M. Mousa, and N. Ismail, "Character gated recurrent neural networks for Arabic sentiment analysis," *Scientific Reports*, vol. 12, article no. 9779, pp.1-17, June. 2022. DOI: https://doi.org/10.1038/s41598-022-13153-w

[38] M. F. Bashir, A. R. Javed, M. U. Arshad, T. R. Gadekallu, W. Shahzad, and M. O. Beg, "Context aware emotion detection from low resource Urdu Language using Deep Neural Network," *in ACM Transactions on Asian and Low-Resource Language Information Processing,* , vol. 22, no.131, pp. 1-30, May.2023. DOI: https://doi.org/10.1145/3528576.

[39] M. A. Hossain, M. A. Hoque, A. Hossain, and M. S. Hossain, "Emotion detection and sentiment analysis in Bengali text using restricted Boltzmann machines," *IEEE Access*, vol. 8, pp. 175524-175535, Apr. 2020. DOI: https://doi.org/10.1109/ACCESS.2020.3026638

[40] E. Ogbuju, F. Oladipo, V. Yemi-Peters, M. Rufai, T. Olowolafe, and A. Aliyu, "Sentiment Analysis of the Nigerian Nationwide Lockdown due to COVID19 Outbreak," 2020. [Online]. Available: http://dx.doi.org/10.2139/ssrn.3665975.

[41] S. M. Maheen, M. R. Faisal, R. Rahman, and M. S. Karim, "Alternative non-BERT model choices for the textual classification in low-resource languages and environments," in *Proceedings of the Third Workshop on Deep Learning for Low-Resource Natural*

Language Processing, pp. 192–202, 2022. doi: 10.18653/v1/2022.deeplo-1.20.

[42] P. Balachandran, U. Thayasivam, R. Pushpananda, and R. Weerasinghe, "Towards Effective Emotion Analysis in Low-Resource Tamil Texts," *Proc. 5th Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pp. 17–30, May 2025.

[43] M. F. Bashir, A. R. Javed, M. U. Arshad, T. R. Gadekallu, W. Shahzad, and M. O. Beg, "Context-aware emotion detection from low-resource Urdu language using deep neural network," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 22, no. 5, pp. 1-30, 2022. doi: 10.1145/3528576.

[44] S. Kalateh, L. A. Estrada-Jimenez, S. Nikghadam-Hojjati, and J. Barata, "A systematic review on multimodal emotion recognition: Building blocks, current state, applications, and challenges," *IEEE Access*, vol. 12, pp. 103976–104005, 2024, doi: 10.1109/ACCESS.2024.3430850.

[45] T. Kolajo, O. Daramola, A. Adebiyi, and A. Seth, "A framework for preprocessing of social media feeds based on integrated local knowledge base," Information Processing & Management, vol. 57, no. 6, article no. 102348, Nov.2020. Doi: https://doi.org/10.1016/j.ipm.2020.102348.

[46] T. Kolajo, O. Daramola, and A. Adebiyi, "Real-time event detection in social media streams through semantic analysis of noisy terms," *Journal of Big Data*, vol. 9, article no. 90, July. 2022. Doi: https://doi.org/10.1186/s40537-022-00642-y

[47] S. K. Bharti, S. Varadhaganapathy, R. K. Gupta, P. K. Shukla, M. Bouye, S. K. Hingaa, and A. Mahmoud, "Text-based emotion recognition using deep learning approach," *Computational Intelligence and Neuroscience*, vol. 2022, article no. 2645381, Aug.2022. DOI: https://doi.org/10.1155/2022/2645381.

[48] L. S. Meetei and T. D. Singh, "Low resource language specific pre-processing and features for sentiment analysis task," *Language Resources & Evaluation*, vol. 55, no.4, pp. 947-969, Dec.2021. doi: https://doi.org/10.1007/s10579-021-09541-9

[49] F. Koto, T. Beck, Z. Talat, I. Gurevych, and T. Baldwin, "Zero-shot sentiment analysis in low-resource languages using a multilingual sentiment lexicon," in *Proc. 18th Conf. Eur. Chapter Assoc. Comput. Linguistics (EACL 2024)*, pp. 298–320, Mar. 2024.

[50] P. M. Mastel, A. Munezero, E. Namara, R. Kagame, Z. Wang, A. Anzagira, A. Gupta, and J. D. Ndibwile, "Natural language understanding for African languages," in *Proceedings of the AfricaNLP Workshop at ICLR 2023*, 2023.

[51] D. I. Adelani, G. Neubig, S. Ruder, S. Rijhwani, M. Beukman, C. Palen-Michel, C. Lignos, J. O. Alabi, S. H. Muhammad, P. Nabende, C. M. B. Dione, A. Bukula, R. Mabuya, B. F. P. Dossou, B. Sibanda, H. Buzaaba, J. Mukiibi, G. Kalipe, D. Mbaye, A. Taylor, F. Kabore, C. C. Emezue, A. Aremu, P. Ogayo, C. Gitau, E. Munkoh-Buabeng, V. M. Koagne, A. A. Tapo, T. Macucwa, V. Marivate, E. Mboning, T. Gwadabe, T. Adewumi, O. Ahia, J. Nakatumba-Nabende, N. L. Mokono, I. Ezeani, C. Chukwuneke, M. Adeyemi, G. Q. Hacheme, I. Abdulmumin, O. Ogundepo, O. Yousuf, T. M. Ngoli, and D. Klakow, "MasakhaNER 2.0: Africa-centric transfer learning for named entity recognition," *arXiv*, vol. arXiv:2210.12391, Nov. 2022. doi: 10.48550/arXiv.2210.12391.

[52] M. Krasitskii, O. Kolesnikova, L. Chanona Hernandez, G. Sidorov, and A. Gelbukh, "Multilingual approaches to sentiment analysis of texts in linguistically diverse languages: A case study of Finnish, Hungarian, and Bulgarian," in *Proc. 9th Int. Workshop Comput. Linguist. Uralic Lang.*, pp. 49–58, Nov. 2024.

[53] K. B. Zümberoğlu, S. Z. Dik, B. S. Karadeniz, and S. Sahmoud, "Towards better sentiment analysis in the Turkish language: Dataset improvements and model innovations," *Applied Sciences*, vol. 15, no. 4, p. 2062, Feb. 2025, doi: 10.3390/app15042062.

