

Review

Not peer-reviewed version

NLP-Based Approach to Multilingual Fake News Detection Through Social Media in Low-Resource Languages: A Review

[Sibgha Munir](#)* and [Haris Munir](#)

Posted Date: 9 September 2025

doi: 10.20944/preprints202509.0710.v1

Keywords: natural language processing; fake news detection; multilingual nlp; low-resource languages; social media misinformation; text classification; named entity recognition; transfer learning; multilingual embeddings; cross-lingual models; data scarcity; language diversity; misinformation detection



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

NLP-Based Approach to Multilingual Fake News Detection Through Social Media in Low-Resource Languages: A Review

Sibgha Munir ^{1,*} and Haris Munir ²

¹ University of Engineering and Technology (UET), Pakistan
² COMSATS University, Pakistan; fa22-bcs-069@students.cuisahiwal.edu.pk
* Correspondence: 2024msaie14@student.uet.edu.pk

Abstract

Rapid dissemination of fake news via social media threatens society's stability, especially in multilingual and low-resource languages. This paper reviews the state-of-the-art NLP techniques applied in fake news detection across multiple languages, with a particular focus on low-resource languages. It discusses the main methods of text classification, sentiment analysis, and NER and their applications, advantages, and challenges in such settings. This paper also investigates how far techniques such as transfer learning, multilingual embeddings, and cross-lingual models can go toward meeting the challenges of linguistic diversity and scarcity of annotated data. Lastly, it identifies where the current state of the art fails and the implications of new directions for future improvement in model accuracy and adaptability for multilingual fake news detection. The findings of this paper not only enlighten the different challenges faced and innovations in this crucial area but also contribute to enabling the development of robust and inclusive solutions to mitigate the adverse effects of misinformation around the world.

Keywords: natural language processing; fake news detection; multilingual nlp; low-resource languages; social media misinformation; text classification; named entity recognition; transfer learning; multilingual embeddings; cross-lingual models; data scarcity; language diversity; misinformation detection

1. Introduction

The ever-changing trend in sharing information worldwide that has resulted from the development of many social media platforms has changed into a potent activity for wild propagation. Equally, in the other light, this fast development has been marked by the wild spread of misinformation and fake news, which has dramatically become a significant challenge to societal well-being, democracy, and public safety [1]. Also, detecting and mitigating fake news have become a vital area of research in the multilingual social media setting. If significant progress has been achieved today for other major languages, such as English, the current situation is significantly different for low-resource languages where linguistic resources and advanced natural language processing tools are not expected [2].

Hundreds of millions worldwide speak languages with a relative lack of large, annotated datasets, fewer linguistic tools, and sparse research focus. This, in turn, leads to a gap between high-resource languages about fake news detection and fighting [3,4]. The development of robust, scalable, and adaptable NLP-based approaches for the complexities of multilingual fake news detection is

overnight in view, particularly for low-resource languages, as social media is used increasingly across different linguistic groups.

The challenge that lies ahead is appropriately put into perspective by the following statistics in Table 1 on global language distribution and social media usage:

Table 1. Global Language Distribution and Social Media Statistics Applicable to the Detection of Fake News.

Statistics	Value
Number of languages spoken worldwide	Approximately 7,000
Percentage of languages classified as low-resource	Over 90%
Number of people using social media globally	4.7 billion (as of 2023)
Percentage of social media users in low-resource language regions	25%
Increase in fake news incidents on social media (2019-2023)	78% increase
Languages with substantial NLP resources	Fewer than 100
Number of research papers on fake news detection in low-resource languages (2018-2023)	Less than 5% of all studies

The techniques of NLP have been very promising in application domains, ranging from sentiment analysis to machine translation, and hold immense promise for detecting fake news in several languages [5,6]. However, developing such systems in low-resource languages is not without challenges. Those challenges include the rarity of big, annotated corpora, the complexity of linguistic diversity, and the requirement for models to be cross-linguistically adaptable. The handling of such topics calls for new methods based on recent advances in transfer learning, multilingual embeddings, and the exploitation of domain-specific knowledge.

This paper aims to fully review state-of-the-art NLP-based multilingual fake news detection methodologies, mainly focusing on low-resource languages [7]. We will present a few of the significant challenges and existing approaches in the area and study the potential of new emerging technologies in closing the gap across fake news detection on linguistic grounds. Finally, how social media acts as a facilitator and a battleground for fake news underlines the need to develop culturally and contextually aware detection systems.

It synthesizes present knowledge in this rapidly evolving field but also points out the gaps and opportunities for further research. The authors intend to contribute to developing more inclusive and practical tools to fight fake news, hence ensuring all linguistic communities can benefit from these protections as advanced NLP techniques have to offer [8].

This review answers the following questions:

RQ1: Which NLP technique performs best for fake news detection in low-resource languages?

RQ2: How have current models overcome issues related to linguistic diversity in multilingual fake news detection?

RQ3: What are the contributions of transfer learning in enhancing fake news detection for low-resource languages?

RQ4: How does exploiting multi-lingual embeddings improve the performance of fake news detection across many languages?

RQ5: What are the main limitations of the current state-of-the-art in multilingual fake news detection with respect to low-resource languages?

RQ6: What are the most important promising future directions for developing NLP-based systems against fake news, considering the low-resource language environment?

A. Topology of Review

Section 2 presents the review methodology of NLP-based approaches in multilingual fake news detection. Section 3 elaborates on the key NLP techniques, Text Classification, Sentiment Analysis, and Named Entity Recognition, that are basic in identifying fake news. Section 4 discusses state-of-the-art approaches designed explicitly for low-resource languages, examining the roles of Transfer Learning, Multilingual Embeddings, and Cross-linguistic Adaptation. Section 5: Critical analysis of the existing models with the presentation of strengths, limitations, and areas of potential improvements. Section 6 underlines the main challenges in multilingual fake news detection, outlining what the authors call a Scarcity of Annotated Data, Linguistic Diversity, and Model Adaptability problem. Section 7 outlines the future directions in the field. Section 8 concludes the review by outlining the main findings and implications for future research. The depiction of the topology of the review is shown in Figure 1.



Figure 1. Organization of the review.

2. Literature Review

Alnabhan et al. conducted a systematic review on fake news detection using deep learning. After screening 1642 articles, 393 papers were analyzed, with 178 selected for detailed review. Commonly used models like CNN and Bi(LSTM) showed high accuracy, with 47% using BERT and hybrid models for practical purposes. Class imbalance was highlighted as a challenge, impacting detection accuracy. Transfer learning was recommended to enhance performance. The study identifies areas for future research, emphasizing the importance of advanced deep learning techniques and addressing class imbalance in fake news detection systems [9].

Alghamdi et al. analyzed a framework's ability to model inter-domain relationships and intra-domain relationships within the same language. The study examined training and testing in the same language, as well as the transfer of domain-independent knowledge across different domains. The hybrid extractive-abstractive summarization method showed effectiveness in detecting fake news across languages and domains, improving model flexibility and accuracy in classification [10].

Fagundes et al. conducted a study reviewing 20 papers to address specific research questions. Most excluded articles focused on syntax without delving into primary research. Shallow syntactic representations like parts of speech offer minimal improvements in fake news detection, while deeper representations like context-free grammar show promise. Combining syntactic information with other techniques enhances detection accuracy. Limitations include search engine precision and publication scope up to October 2021. This review highlights the value of syntactic information in enhancing automated fake news detection [11].

Krasadakis et al. introduced a paper discussing challenges and developments in NLP within the legal domain. The focus is on tasks such as NER, EL, RelEx, and Coref, important for creating a legal citation network and law consolidation system. The study highlights the complexities of legal texts and the challenges in replicating results due to lack of standardization. There is a call for more research in multilingual NLP, particularly for low-resource languages, and the need for benchmarks like LEXTREME to measure progress [12].

Meesad et al. developed a framework to identify fake news in Thai, with three main modules: Information Retrieval, Natural Language Processing, and Machine Learning. They collected data from Thai news websites using web crawlers, resulting in 41,448 samples. Various machine learning models were tested, with the Long Short-Term Memory model proving most effective. The study also highlighted challenges in Thai text processing, such as word segmentation issues [13].

Bahmanyar et al. introduced a transformer-based deep learning pipeline for detecting fake news in multiple languages. By fine-tuning the model on a multilingual dataset with 16 languages, they achieved an accuracy of 97.38% and an F1 score of 0.9734. Zero-shot testing demonstrated the model's ability for cross-lingual transfer and understanding, even for languages not seen in the training data. The study also developed a filtering model with 99.8% accuracy to remove impractical data during dataset enlargement, ensuring high-quality data for detecting fake news across various languages [14].

Schmidt et al. conducted a study using the Snopes dataset to tune LLaMA and baseline models, with the best results seen with the LLaMA-2 model with 70 billion parameters. Performance differences were significant between SVM models relying on stylistic features and LLaMA-2 models leveraging pretrained knowledge. Experimentation on the FCTR1000 dataset showed high F1-macro scores of 0.89 and 0.828 from the LLaMA-13B model when both claim statements and summaries were input. However, LLaMA models performed poorly in Turkish, particularly when only claim statements were used, indicating the need for additional context for improved results [15].

Abboud et al. developed a fake news detection system for low-resource languages using a knowledge base from high-resource languages. Their system overcomes the challenge of limited training data, with a significant accuracy improvement of 3.71% compared to baseline. By incorporating a cross-lingual source information feature, accuracy increased by an additional 3.03%, allowing for automated credibility checks on articles. This approach can be applied to various low-resource languages, making it a versatile solution for fake news detection [16].

Masis et al. conducted a study using a dataset of 19,000 entries, splitting it 80/20 for training and testing. Logistic regression was employed, with results showing that news titles alone were sufficient for training the model effectively. The study highlighted the importance of news titles in identifying fake news, as demonstrated by the model's performance in classifying articles. However, the model's performance varied when tested on articles from different sources, dates, or agendas, indicating a need for fine-tuning or algorithm adjustments. The authors recommended using a hybrid dataset to improve performance and minimize data drift in machine learning models [17].

Ziyaden et al. found that text augmentation methods, especially through translation approaches like Google Translate API, significantly improved model performance, with an accuracy rate of 87%. The baseline approach failed for certain classes, prompting the introduction of augmentation techniques that greatly increased F1 scores by at least one-third for augmented classes. Comparing back-translation and neural network-based translations, Google Translate proved most effective with an accuracy of 0.87. Augmentation techniques were deemed crucial for addressing class imbalances and enhancing model generalization for low-resource languages, providing valuable gains for text classification tasks. The authors also made their Azerbaijani language dataset and pre-trained Roberta model publicly available for future NLP research. [18].

3. Methodology

The steps adapted to perform this systematic review are as follows:

A. *Collection of Articles*

A structured search was performed across major databases on aspects of NLP-based [19] approaches to multilingual fake news detection [20,21], focusing on low-resource languages. This search was conducted in June 2025 using the following databases: Google Scholar, Scopus, and Web of Science. These keywords were, therefore, targeted to capture relevant studies: fake news detection, multilingual NLP [22], low-resource languages [23], transfer learning, multilingual embeddings, cross-linguistic adaptation, and social media. This search returned 987 peer-reviewed articles. Further screening excluded those studies which had a focus on detecting fake news in high-resource languages. Further, only the articles after the year 2014 were considered for this review since the development in NLP techniques for fake news detection happened at a very fast pace in the last decade.

B. *Search Strategy*

Clear inclusion and exclusion criteria were set to assess the quality of the selected literature. These criteria are crucial in ensuring the reliability of findings and consistency in the review process. The evaluation will consider the title of the article, the abstract, and the full text. The inclusion criteria that have been considered for the article review include the following:

- Only studies published between 2015 and 2025 were included to reflect recent research trends.
- It includes various research works on the methods of NLP-based multilingual fake news detection [24] focusing on low-resource languages [25,26].
- More preferences were given to those papers which discussed technical aspects-feature extraction, model architecture, preprocessing of data, and transfer learning techniques for cross-linguistic adaptation.
- Those papers were more preferred which reported the results based on performance metrics like accuracy, precision, recall, F1-score, and AUC.
- Only peer-reviewed journal articles and conference papers were included to maintain the integrity of the review. In this context, editorials, posters, and technical reports were not considered to be peer-reviewed sources.

Each study was critically evaluated for inclusion, and only those studies that met the quality criteria were included in the final review. A schematic overview of the general plan for this review is summarized in Figure 2 (hypothetical).

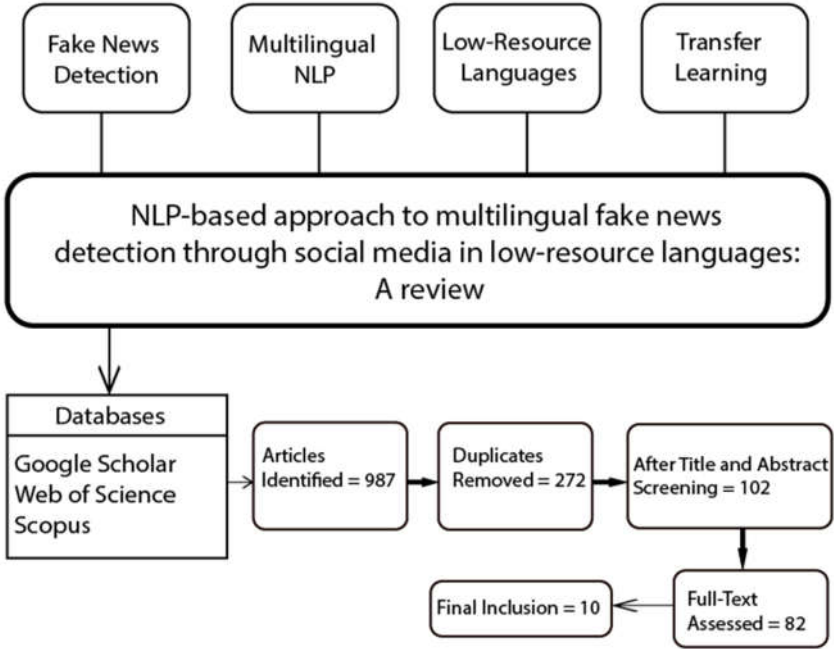


Figure 2. Strategy for articles screening and inclusion.

4. Research Findings

The results of the research in this review are presented below:

A. NLP Techniques for Fake News Detection

In that respect, various techniques of NLP have been applied in the multilingual detection of fake news to identify challenges associated with spreading misinformation on social media. These techniques include text classification, sentiment analysis, and named entity recognition. All these techniques are central in detecting and classifying fake news. We shall describe each of these techniques by focusing on their applications, advantages, and challenges in this domain.

a) Text Classification

One of the most basic NLP techniques in fake news detection is text classification [27,28], where the given text is classified into pre-defined classes like 'fake' or 'real'. This is vital in filtering and identifying misleading information across various languages. In low-resource languages, there is an acute scarcity of annotated data. This problem has, however, been circumvented by techniques like transfer learning [27] and data augmentation [28], as well as the utilization of synthetic data generation among others. For example, pre-trained models in high-resource languages may enhance their performance on the classification task by fine-tuning them on small datasets in low-resource languages.

Other robust features are multilingual embeddings, which project words from different languages on a shared vector space. As the model can make better generalizations, this improves text classification across languages even when trained on small amounts of data from low-resource languages [29]. While there have been considerable gains in handling code-switching or dialectal variations—quite common on social media—there are still challenges to overcome.

b) Sentiment Analysis

Among the most critical techniques employed is sentiment analysis , which infers the feeling or emotion the piece of text conveys regarding the intent behind the information [30,31]. For instance, most fake news stories are overcooked or couched in emotionally loaded language to worm their way

into readers' sensibilities. In low-resource languages, sentiment analysis is a significant challenge due to the lack of annotated datasets and the complexity of their linguistic structure. The intricacy of handling nuanced language features such as puns and wordplay adds another layer of difficulty, as these elements can dramatically alter the sentiment conveyed.

Recent progress in cross-lingual transfer in sentiment analysis uses models trained in high-resource languages to classify sentiment in low-resource languages directly. This might be based on a method like zero-shot learning, wherein a model trained in one language can make inferences regarding another untrained language. However, the results are not as promising when low-resource languages are considered, as accuracy is generally lower than in high-resource languages, which usually indicate the state of the art and thus warrant actual further advances in research and development [32,33].

c) *Named Entity Recognition (NER)*

It identifies and classifies the entities mentioned in the text, such as people, organizations, locations, and so on [34,35]. Fake news detection would leverage NER to identify from the text key players and locations, which it would then cross-check against fact-based databases for authenticity. Nevertheless, NER [36,37] for low-resource languages is quite difficult because no annotated corpora exist and language-specific entities are pretty complicated.

Recent approaches have relied on transfer learning from high-resource languages to improve the performance of NER in low-resource conditions. Similarly, multilingual models of NER [38] trained on texts from multiple languages have proved quite effective for identifying entities from low-resource languages. They, however, usually lose their way on language-specific nuances and cultural references that creep in, hence leading to inaccuracies [39].

B. *State-of-the-Art Approaches in Low-Resource Languages*

The challenges in fake news detection for low-resource languages are very peculiar, mainly due to the lack of availability of annotated datasets, scarcity of linguistic resources, and diversity in dialects and variants. Traditional NLP techniques often struggle in working within such constraints, and more advanced or innovative approaches will have to be explored. The chapter looks in depth at three of the major methodologies dealing with these challenges: Transfer Learning [40,41], Multilingual Embeddings [42,43], and Cross-linguistic Adaptation. These constitute preliminary pointers to areas that are quite promising in terms of potential effectiveness for low-resource languages in fake news detection.

a) *Transfer Learning*

Recently, transfer learning has been very successful in NLP. In most cases, especially for low-resource languages, this has been the case. Essentially, transfer learning applies knowledge from high-resource-language or domain-trained models to help in the performance on tasks in the low-resource setting [44]. This typically means pre-training a model on a large corpus from a high-resource language like English and fine-tuning it on a smaller dataset from the target low-resource language.

Transfer learning has shown reasonable promise in the domain of fake news detection. Pretraining on huge amounts of data enables models to learn linguistic patterns, syntactic structures, and semantic nuances common across languages [45,46]. These models have been demonstrated to detect fake news with much better performance than from-scratch models, after fine-tuning with targeted data in a low-resource language. Besides, transfer learning removes the requirement of large annotated datasets, which are very hard to come by in low-resource languages; hence, the solution is pretty workable for multilingual fake news detection.

b) *Multilingual Embeddings*

Another important component dealing with low-resource languages challenges is multilingual embeddings. Word representations learn semantic relationships across languages. Multilingual

embedding spaces map words from different languages into a common vector space and let models make use of cross-lingual information, enabling transfer from high- to low-resource languages [47].

In fake news detection, multilingual embeddings enable the integration of data from several languages to be utilized with greater generalization across linguistic boundaries. This can become very helpful in multilingual societies where fake news is propagated in many languages at once. In this way, a model could handle fake news in low-resource languages using the rich information available from high-resource languages and improve the system performance and strength for fake news detection in general [48].

c) Cross-linguistic Adaptation

Cross-linguistic adaptation is the method for adapting or tuning existing NLP models [49,50] for good performance over a different portion of language—usually those for low-resource languages. Based on linguistic similarity, either in base roots, cases of grammar, or cognates, among others, then such models trained over one language will have great performance over another. When detecting fake news, the linguistic subtleties of misinformation may vary from one language to the next but will still share some core features.

The first proposes that training should be done on a pivot language near the target low-resourced language. The model can thereafter be fine-tuned [51] to the target language with minimal additional effort to ensure that the approach is cost-effective in fake news detection. Also, other techniques, such as zero-shot learning and a few-shot, which allow generalization from very few examples, can be applied to enhance cross-linguistic adaptation further. Thereby, the model would be improved with respect to fake news detection in low-resource languages.

The flowchart diagram of approaches in Low-Resource Languages is shown in Figure 3.

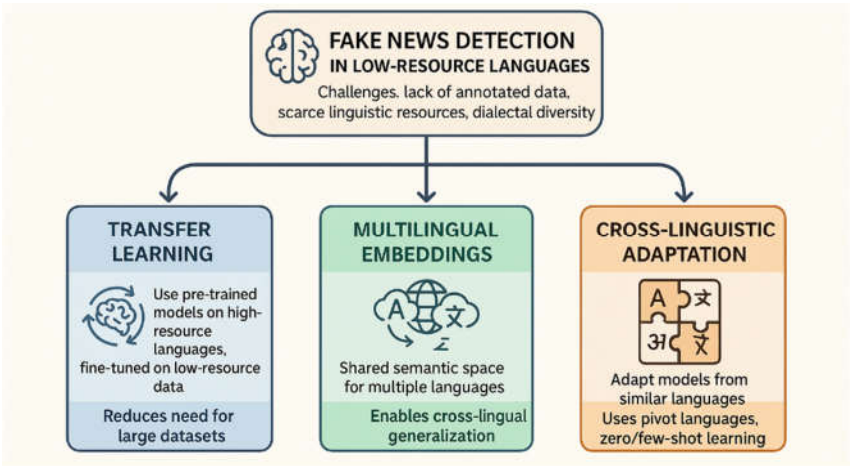


Figure 3. Advanced NLP Approaches for Low-Resource Languages in Fake News Detection.

C. Critical Analysis of Existing Models

We now critically review the present models that were designed for fake news detection in multilingual and low-resource language settings. The outline of this analysis falls into the following three categories: strengths, limitations, and opportunities for improvement. This helps to go through, systematically, how the current state of NLP models stands within this domain and hence point out what works, where these models fall short, and how future research can address those gaps.

a) Strengths

i. Cross-Linguistic Generalization

One of the major strengths of current NLP models for fake news detection in low-resource languages rests in their ability to generalize across languages. Transfer learning and multilingual embeddings equip such models to borrow knowledge from high-resource languages and apply it in a low-resource setting, hence improving performance. Cross-linguistic transfer has contributed decisively to the enrichment of semantic knowledge and has thus provided better accuracy of detection in target languages, even when limited training data is available [52,53]. AI-driven techniques will further enhance the ability of capturing complex linguistic patterns, thus helping cross-linguistic generalization.

ii. Scalability and Flexibility

Most contemporary fake news detection models are both scalable and flexible. As such, they are easily adaptable across many languages and social media platforms. Modular architectures and resorting to pre-trained language models like BERT [54,55] and XLM-R allow fine-tuning in language-specific and domain-specific scenarios, hence applicability in very diversified linguistic settings without the need for extensive retraining from scratch. The integration of AI enables these models to learn and adapt continuously, making them scalable to new languages and platforms with minimal intervention.

iii. Improved Textual Feature Extraction

The integration of advanced NLP techniques such as sentiment analysis, named entity recognition (NER), and contextualized embeddings has significantly enhanced the capability of models to extract relevant textual features from multilingual data. These features are crucial for distinguishing between fake and real news, especially in languages where traditional lexical resources are scarce or non-existent [56,57]. AI enhances this by making the process deeper and more subtle, which allows for capturing subtleties that would be lost through traditional methods. XAI, in turn, enhances that with insight into what features were most influential to the model's decision-making process, offering a level of transparency and thus trustworthiness.

iv. Enhanced Model Robustness

A great number of strategies to improve the robustness against adversarial attacks and noise in social media data are already included in existing models. Those models are more and more powerful in maintaining accuracy across different languages and dialects, due to focusing on feature stability and resilience against linguistic variations [58,59]. AI plays a critical role in enhancing this robustness, as it allows models to learn from diverse and noisy data environments. XAI contributes by making the decision-making process more interpretable, helping identify and mitigate potential vulnerabilities in the model, thereby improving overall robustness and reliability.

b) Limitations

i. Data Scarcity and Quality

Even after all the developments, high-quality annotated data remains one of the most important limitations to NLP models in low-resource languages [60,61]. In this respect, a lack of comprehensive datasets hinders not only model training but also generalizability and reliability with regard to fake news detection across different linguistic contexts. This is much more pressing when dealing with underrepresented languages, wherein not even the basic linguistic resources may be available.

ii. Limited Cultural Context Understanding

Current models often miss an understanding of the cultural and contextual nuances underlying such fake news content. This becomes even more critical in multilingual and low-resource settings where cultural references, idiomatic expressions [62,63], and local knowledge are crucial in correctly detecting fake news. Their inability to understand such subtleties properly might hence result in misclassification and loss in effectiveness. Moreover, cross-culture challenges arise when models

trained on one cultural context are applied to another, often failing to capture the unique socio-cultural dynamics that influence how fake news is spread and perceived.

iii. *Inadequate Adaptation to Low-Resource Languages*

While transfer learning and multilingual embeddings have enhanced model performance, this is not without limitations. For instance, adaptation may still be out of reach for truly low-resource languages, in which knowledge transfer from high-resource languages still proves too inadequate to capture the richness of the language [64]. This often results in suboptimal performance and thereby puts a pointer to the fact that more tailored approaches are thus needed. Additionally, uneven economic development across regions exacerbates these issues, as countries with limited access to technological resources are unable to deploy and benefit from advanced NLP models, leading to a digital divide in fake news detection capabilities.

iv. *High Computational Requirements*

However, one drawback of using state-of-the-art models with sophisticated architectures is their high computational costs, making them less deployable in resource-constrained environments. This particularly falls under relevant conditions in the low-resource language setting, where low access to state-of-the-art computational infrastructure clearly hinders the practical application of those models [65]. Political turmoil in certain regions can further complicate the deployment of these models, as instability may limit the availability of resources and the ability to maintain consistent data collection efforts, further hampering the effectiveness of fake news detection initiatives.

c) *Opportunities for Improvement*

i. *Development of Multilingual Datasets*

The most promising line of improvement lies in the creation of large, high-quality multilingual datasets for fake news detection. Because of where models are at the moment—both performing and being evaluated badly—the extension of the availability of annotated data to a greater number of languages will help researchers better train and evaluate their models [66].

ii. *Incorporation of Cultural and Contextual Knowledge*

Increasing the models with cultural and contextual understanding can improve their accuracy in fake news detection [67]. For instance, it can be done by incorporating some forms of external knowledge sources, such as cultural ontologies or archives of local news, within the training process for a model. On the other hand, the creation of models able to learn dynamically from context-specific cues at inference time will further increase their adaptability.

iii. *Optimization for Low-Resource Environments*

In view of computational constraints, it is future research that shall orient NLP models to become more suitable for low-resource environments. This can include more efficient algorithms, reducing model sizes without losing too much in performance, or new methods for training that require less of the same resource. These efforts will make advanced techniques in NLP available to a wider audience and more applicable in low-resource scenarios [68].

The summary of a literature review is shown in Table 2.

Table 2. Summary of literature review.

Author, Year	Target Variable	Input	Architecture	Pre-Processing	Dataset	Outcome	Output Result
Alnabhan et al. – 2023	Fake News Detection	Social Media Text	CNN, Bi(LSTM),	Data Augmentatio	178 articles	Fake News Detection	High Accuracy with Class

			BERT, Hybrid Models	n, Class Balancing			Imbalance Issues Addressed
Alghamdi et al. – 2024	Fake News Detection	Social Media Text	Hybrid Extractive-Abstractive Summarization	Domain Transfer, Data Reduction	Various Domains	Cross-Domain Transfer	Improved Model Versatility
Fagundes et al. – 2024	Fake News Detection	Social Media Text	Deep Syntactic Representations	Feature Combination	20 papers	Syntactic Information Usage	Enhanced Detection with Deep Syntactic Features
Krasadaki s et al. – 2024	Fake News Detection	Legal Documents	NER, EL, RelEx, Coref	Custom Dataset Challenges	Legal Texts	Legal Citation Management	Improved Multilingual NLP Applications
Meesad et al. – 2021	Fake News Detection in Thai	Thai News Articles	Naïve Bayes, Logistic Regression, LSTM	Word Segmentation	41,448 samples	Fake News Detection	LSTM Model Most Accurate, Addressed Thai Language Challenges
Bahmanyar et al. – 2024	Fake News Detection in Multiple Languages	Multilingual Text	Transformer-based Models	Data Filtering, Zero-Shot Learning	16 Languages	Multilingual Fake News Detection	97.38% Accuracy, Cross-Lingual Transfer
Schmidt et al. – 2024	Fake News Detection	Statement Text	LLaMA Models (Various Sizes)	Stylistic Feature Tuning	Snopes and FCTR1000 Datasets	Fake News Detection	High Accuracy with Contextual Improvement
Abboud et al. – 2022	Fake News Detection in Low-Resource Languages	Multilingual Text	Cross-Lingual Transfer Learning	Machine-Translated Data	Various Low-Resource Languages	Low-Resource Fake News Detection	Improved Accuracy with Cross-Lingual Features

Masis et al. – 2024	Fake News Detection	News Titles	Logistic Regression	Title-Based Feature Extraction	19,000 entries	Fake News Detection	Effective with Title-Only Input
Ziyaden et al. - 2024	Fake News Detection	Social Media Text	Text Augmentation Techniques	Back-Translation, Neural Network-Based	Various Languages	Enhanced Model Performance	Improved Accuracy with Augmentation Strategies

5. Challenges

These numerous challenges are represented in the detection of this fake and unreliable news moving across multilingual social media platforms, especially for low-resource languages, due to the complexities of language processing, data availability, and the adaptability of models with respect to diverse linguistic contexts. As misinformation spreads unrestrained, it becomes critical to understand and address those hurdles that absolutely impede NLP-based fake news detection in such under-resourced languages.

A. Scarcity of Annotated Data

Multilingual fake news detection using multilingual approaches is a big pain in the neck because of data scarcity with manual annotation [69]. High-quality labeled datasets are the backbone for training and validating NLP models; however, in big parts of the world, they remain limited or nonexistent for languages with low resources. The process of coming up with manual annotations is time-consuming and costly, and most importantly, for effective machine learning, it is difficult to achieve huge datasets [70]. Such an absence of annotated data leads to requirements of either transfer learning or unsupervised methods, causing generally low attainment of desired accuracies in low-resource settings [71]. This also prevents robust models from being developed.

B. Linguistic Diversity

Another huge challenge to multilingual fake news detection is linguistic diversity. Most of the low-resource languages indicate high variation at the dialect, orthographic, and syntax levels, making the development of NLP models really challenging [72]. Another way of looking at this is that diversity implies a model trained on one dialect or variant of a language may not work well on another. Thirdly, many low-resource languages are mainly oral, and hence there is little in terms of written forms; this therefore presents a challenge to the development of text-based models in NLP. With the languages varying both intra- and interlingually to handle, high flexibility is required from the models when considering general phenomena covered by a model against the limited resources available [73].

C. Model Adaptability

This has been a critical challenge in fake news detection: the adaptability of NLP models to low-resource languages. The effectiveness of most state-of-the-art models developed and fine-tuned for high-resource languages like English is quite limited when applied to low-resource languages [74]. This is because it is hard to generalize across the models due to the linguistic differences and scarcity of training data in the target language [75]. Moreover, models trained on high-resource languages often exhibit degraded performance when transferred to low-resource languages. This can be improved by techniques that make the models more effective learners of limited data, as well as by architectures that accommodate features unique to low-resource languages.

Table 3. Key Challenges in Multilingual Fake News Detection in Low-Resource Languages.

Challenge	Description	Implications	Potential Approaches
Scarcity of Annotated Data	Limited availability of manually labeled datasets in low-resource languages due to high costs and time required for annotation.	Hinders model training and validation; leads to poor performance and low accuracy.	Use of transfer learning, data augmentation, and unsupervised learning methods.
Linguistic Diversity	High variation in dialects, orthography, syntax, and presence of primarily oral languages.	Models trained on one dialect may not generalize to others; limited written data hinders the development of these models.	Develop dialect-specific resources; leverage multilingual embeddings and speech-to-text technologies.
Model Adaptability	Difficulty in adapting NLP models trained on high-resource languages to low-resource contexts.	Reduced performance and generalization; failure to capture unique linguistic features.	Design lightweight, adaptable architectures; train models on multilingual datasets; integrate meta-learning.

6. Future Directions

Inevitably, the future of multilingual fake news detection lies in low-resource languages. As the digital world deepens, the urge for sophisticated and accurate tools to fight misinformation will heighten across diverse linguistic contexts [76]. Certainly, one of the promising avenues of future research lies in developing more advanced models of NLPs that can be fine-tuned seamlessly to the nuances of low-resource languages. This will require similar efforts in the creation of new datasets and in augmenting models with techniques such as transfer learning and multilingual embeddings.

Another important line of research is to make techniques in transfer learning more effective for enabling models fit for use in high-resource languages to be adapted for use in low-resource languages. This could be in the form of more efficient algorithms that transfer knowledge at minimal loss of accuracy or the development of intermediate languages that bridge the gap between high- and low-resource languages. Otherwise, unsupervised learning methods would reduce dependency on large annotated datasets often unavailable for low-resource languages [77].

Other key directions would be to include techniques for cross-linguistic adaptation that would enable models to learn from and generalize in one go across multiple languages. This may involve the development of totally new architectures that would capture language similarities and differences better or, conversely, meta-learning strategies whereby models learn how to adapt rapidly to new languages with minimum retraining [78]. This way, enhancing the models' ability to generalize across languages, researchers will then establish much more robust systems capable of detecting fake news in most linguistic and cultural contexts.

Future work must also take into socio-cultural considerations regarding the detection of fake news in low-resource languages. This is generally done by creating more contextually sensitive models that consider, first of all, the regional cultural and political environments in which they must work. By doing so, it can make sure that these tools designed by researchers are effective and culturally sensitive [79]. This collaboration between academia, industry, and government would play an important role in the future of shaping multilingual fake news detection. Such collaborative efforts

are considered to enable the sharing of resources, data, and expertise that are necessary to address the specific challenges of low-resource languages [80]. At the same time, international cooperation can be very instrumentally important for setting global standards and best practices on fake news detection to ensure that benefits from these technologies are realized on a global scale.

More important in the future will be the rise of user-centric approaches to fake news detection, where end-users take an active role in identification and reporting [81]. It may involve the development of new interfaces and tools that empower people to be part of the detection process or even the creation of community-driven platforms using collective intelligence naïve ways to stand against fake news on a regular basis [82]. By more directly including users in the process, researchers can create more effective and scalable solutions to the problem of misinformation.

In other words, the technical innovation on the method of detecting multilingual fake news, considering ethical aspects, the socio-culture awareness, and collaborative effort, charts the future directions of this field of low-resource languages. These will be extremely important efforts at the core as researchers continue to push boundaries on what is possible to make sure the digital information space remains reliable and trustworthy to all, no matter their linguistic background.

7. Conclusions

We outline some of the important, open questions identified from the literature relating to NLP-based approaches for multilingual fake news detection through social media in low-resource languages. Even though transfer learning has greatly helped make strides in porting knowledge between well-resourced and low-resource languages, wide chasms still remain. Further research is required in developing sophisticated models and techniques that can rise to the challenges presented by low-resource languages. Our review underlines the need to create larger and more diverse annotated datasets to improve accuracy and generalizability in fake news detection systems. There is a striking scarcity of publicly available sets for multilingual fake news detection, and interest in the development and sharing of such resources is growing. Practitioners are invited to test their methods against several datasets to guarantee robustness and effectiveness on different contexts. There is a heavy focus on certain NLP techniques and language pairs in the existing literature. In view of this fact, one should extend research toward understudied languages and dialects and look into the application of advanced techniques like cross-linguistic embeddings and multilingual models. However, there is a dearth of studies that investigate the integration of multiple techniques in NLP for a holistic methodology to detect fake news. In the future, this may focus on integrating text classification, sentiment analysis, and NER with the newer techniques such as transfer learning and domain adaptation to improve detection power. Addressing these gaps will significantly advance the field and improve the effectiveness of NLP-based fake news detection systems in different languages and on social media.

References

1. Dr. B. Dhiman, "The Rise and Impact of Misinformation and Fake News on Digital Youth: A Critical Review," *SSRN Electronic Journal*, 2023, doi: 10.2139/ssrn.4438362.
2. A. De, D. Bandyopadhyay, B. Gain, and A. Ekbal, "A Transformer-Based Approach to Multilingual Fake News Detection in Low-Resource Languages," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 21, no. 1, 2022, doi: 10.1145/3472619.
3. H. Jand, "Natural Language Processing with Speech Recognition," *International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)*, 2018.
4. A. Bhardwaj, P. Khanna, S. Kumar, and Pragya, "Generative Model for NLP Applications based on Component Extraction," in *Procedia Computer Science*, 2020. doi: 10.1016/j.procs.2020.03.391.
5. C. M. Tsai, "Stylometric Fake News Detection Based on Natural Language Processing Using Named Entity Recognition: In-Domain and Cross-Domain Analysis," *Electronics (Switzerland)*, vol. 12, no. 17, 2023, doi: 10.3390/electronics12173676.
6. R. Mohawesh, X. Liu, H. M. Arini, Y. Wu, and H. Yin, "Semantic graph based topic modelling framework for multilingual fake news detection," *AI Open*, vol. 4, 2023, doi: 10.1016/j.aiopen.2023.08.004.

7. Y. Zhang *et al.*, "Stance-level Sarcasm Detection with BERT and Stance-centered Graph Attention Networks," *ACM Trans Internet Technol*, vol. 23, no. 2, 2023, doi: 10.1145/3533430.
8. K. Sharifani, M. Amini, Y. Akbari, and J. A. Godarzi, "Operating Machine Learning across Natural Language Processing Techniques for Improvement of Fabricated News Model," *International Journal of Science and Information System Research*, vol. 12, no. 9, 2022.
9. M. Q. Alnabhan, P. Branco, and M. Alnabhan, "Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000. Fake News Detection Using Deep Learning: A Systematic Literature Review", doi: 10.1109/ACCESS.2023.DOI.
10. J. Alghamdi, Y. Lin, and S. Luo, "Fake news detection in low-resource languages: A novel hybrid summarization approach," *Knowl Based Syst*, vol. 296, Jul. 2024, doi: 10.1016/j.knosys.2024.111884.
11. M. J. G. Fagundes, N. T. Roman, and L. A. Digiampietri, "The Use of Syntactic Information in fake news detection: A Systematic Review," *SBC Reviews on Computer Science*, vol. 4, no. 1, pp. 1–10, Mar. 2024, doi: 10.5753/reviews.2024.2718.
12. P. Krasadakis, E. Sakkopoulos, and V. S. Verykios, "A Survey on Challenges and Advances in Natural Language Processing with a Focus on Legal Informatics and Low-Resource Languages," Feb. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/electronics13030648.
13. P. Meesad, "Thai Fake News Detection Based on Information Retrieval, Natural Language Processing and Machine Learning," *SN Comput Sci*, vol. 2, no. 6, Nov. 2021, doi: 10.1007/s42979-021-00775-6.
14. K. Agras and B. Atay, "A Novel Transformer-Based Deep Learning Pipeline for Multilingual Fake News Detection," 2024, doi: 10.36838/v6i4.12.
15. R. F. Cekineli, P. Karagoz, and C. Coltekin, "Cross-Lingual Learning vs. Low-Resource Fine-Tuning: A Case Study with Fact-Checking in Turkish," Mar. 2024, [Online]. Available: <http://arxiv.org/abs/2403.00411>
16. S. Han, "Cross-lingual Transfer Learning for Fake News Detector in a Low-Resource Language," Aug. 2022, [Online]. Available: <http://arxiv.org/abs/2208.12482>
17. M. Zovikoğlu and U. Çetin, "Detecting misinformation on social networks with natural language processing," 2024. [Online]. Available: <https://dergipark.org.tr/>
18. A. Ziyaden, A. Yelenov, F. Hajiyeve, S. Rustamov, and A. Pak, "Text data augmentation and pre-trained Language Model for enhancing text classification of low-resource languages," *PeerJ Comput Sci*, vol. 10, 2024, doi: 10.7717/peerj-cs.1974.
19. P. Mookdarsanit and L. Mookdarsanit, "The covid-19 fake news detection in thai social texts," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 2, 2021, doi: 10.11591/eei.v10i2.2745.
20. F. Farhangian, R. M. O. Cruz, and G. D. C. Cavalcanti, "Fake news detection: Taxonomy and comparative study," *Information Fusion*, vol. 103, 2024, doi: 10.1016/j.inffus.2023.102140.
21. B. Hu, Z. Mao, and Y. Zhang, "An overview of fake news detection: From a new perspective," 2024. doi: 10.1016/j.fmre.2024.01.017.
22. J. A. Nasir and Z. U. Din, "Syntactic structured framework for resolving reflexive anaphora in Urdu discourse using multilingual NLP," *KSII Transactions on Internet and Information Systems*, vol. 15, no. 4, 2021, doi: 10.3837/tiis.2021.04.012.
23. S. Ranathunga, E. S. A. Lee, M. Prifti Skenduli, R. Shekhar, M. Alam, and R. Kaur, "Neural Machine Translation for Low-resource Languages: A Survey," *ACM Comput Surv*, vol. 55, no. 11, 2023, doi: 10.1145/3567592.
24. B. Ramesh Reddy, T. Tejaswi, P. Naveen, J. Giresh, and B. Vamsi, "Optimising the Detection of Fake News in Multilingual Exposition using Machine Learning Techniques," in *Proceedings of the 7th International Conference on Intelligent Computing and Control Systems, ICICCS 2023*, 2023. doi: 10.1109/ICICCS56967.2023.10142901.
25. R. Wijayanti, M. L. Khodra, K. Surendro, and D. H. Widyantoro, "Learning bilingual word embedding for automatic text summarization in low resource language," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 4, 2023, doi: 10.1016/j.jksuci.2023.03.015.
26. T. Mahmud, M. Ptaszynski, J. Eronen, and F. Masui, "Cyberbullying detection for low-resource languages and dialects: Review of the state of the art," *Inf Process Manag*, vol. 60, no. 5, 2023, doi: 10.1016/j.ipm.2023.103454.
27. S. Na, P. Na, T. Na, and Dr. Z. Nasim, "Verify: Breakthrough accuracy in the {U}rdu fake news detection using Text classification," in *Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation*, 2022.
28. S. A. F. Azhar, F. Hidayat, M. H. Azfarezat, G. Z. Nabiilah, and Rojali, "EFFICIENCY OF FAKE NEWS DETECTION WITH TEXT CLASSIFICATION USING NATURAL LANGUAGE PROCESSING," *J Theor Appl Inf Technol*, vol. 101, no. 22, 2023.
29. Y. Yang *et al.*, "Multilingual universal sentence encoder for semantic retrieval," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2020. doi: 10.18653/v1/2020.acl-demos.12.
30. W. Cuenca, C. González-Fernández, A. Fernández-Isabel, I. Martín de Diego, and A. G. Martín, "Combining Conceptual Graphs and Sentiment Analysis for Fake News Detection," in *Studies in Computational Intelligence*, vol. 955, 2022. doi: 10.1007/978-3-030-88817-6_15.

31. B. Mohamed, H. Haytam, and F. Abdelhadi, "Applying Fuzzy Logic and Neural Network in Sentiment Analysis for Fake News Detection: Case of Covid-19," in *Studies in Computational Intelligence*, vol. 1001, 2022. doi: 10.1007/978-3-030-90087-8_19.
32. H. Taherdoost and M. Madanchian, "Artificial Intelligence and Sentiment Analysis: A Review in Competitive Research," 2023. doi: 10.3390/computers12020037.
33. A. Lighthart, C. Catal, and B. Tekinerdogan, "Systematic reviews in sentiment analysis: a tertiary study," *Artif Intell Rev*, vol. 54, no. 7, 2021, doi: 10.1007/s10462-021-09973-3.
34. J. Li, A. Sun, J. Han, and C. Li, "A Survey on Deep Learning for Named Entity Recognition," *IEEE Trans Knowl Data Eng*, vol. 34, no. 1, 2022, doi: 10.1109/TKDE.2020.2981314.
35. M. Ehrmann, A. Hamdi, E. L. Pontes, M. Romanello, and A. Doucet, "Named Entity Recognition and Classification in Historical Documents: A Survey," *ACM Comput Surv*, vol. 56, no. 2, 2023, doi: 10.1145/3604931.
36. I. Budi and R. R. Suryono, "Application of named entity recognition method for Indonesian datasets: a review," 2023. doi: 10.11591/eei.v12i2.4529.
37. Z. Sun and X. Li, "Named Entity Recognition Model Based on Feature Fusion," *Information (Switzerland)*, vol. 14, no. 2, 2023, doi: 10.3390/info14020133.
38. H. Wang, L. Zhou, J. Duan, and L. He, "Cross-Lingual Named Entity Recognition Based on Attention and Adversarial Training," *Applied Sciences (Switzerland)*, vol. 13, no. 4, 2023, doi: 10.3390/app13042548.
39. M. S. Hernández, "Beliefs and attitudes of canarians towards the chilean linguistic variety," *Lenguas Modernas*, no. 62, pp. 183–209, 2023, doi: 10.13039/501100011033.
40. M. Iman, H. R. Arabnia, and K. Rasheed, "A Review of Deep Transfer Learning and Recent Advancements," 2023. doi: 10.3390/technologies11020040.
41. Z. Zhu, K. Lin, A. K. Jain, and J. Zhou, "Transfer Learning in Deep Reinforcement Learning: A Survey," *IEEE Trans Pattern Anal Mach Intell*, vol. 45, no. 11, 2023, doi: 10.1109/TPAMI.2023.3292075.
42. K. Hämmerl, J. Libovický, and A. Fraser, "Combining Static and Contextualised Multilingual Embeddings," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2022. doi: 10.18653/v1/2022.findings-acl.182.
43. H. Licht, "Cross-Lingual Classification of Political Texts Using Multilingual Sentence Embeddings," *Political Analysis*, vol. 31, no. 3, 2023, doi: 10.1017/pan.2022.29.
44. B. Dhyani, "Transfer Learning in Natural Language Processing: A Survey," *Mathematical Statistician and Engineering Applications*, vol. 70, no. 1, 2021, doi: 10.17762/msea.v70i1.2312.
45. B. Palani and S. Elango, "CTrL-FND: content-based transfer learning approach for fake news detection on social media," *International Journal of System Assurance Engineering and Management*, vol. 14, no. 3, 2023, doi: 10.1007/s13198-023-01891-7.
46. M. Ghayoomi and M. Mousavian, "Deep transfer learning for COVID-19 fake news detection in Persian," *Expert Syst*, vol. 39, no. 8, 2022, doi: 10.1111/exsy.13008.
47. S. Schwarz, A. Theophilo, and A. Rocha, "Emet: Embeddings from multilingual-encoder transformer for fake news detection," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2020. doi: 10.1109/ICASSP40776.2020.9054673.
48. S. Kasim, "One True Pairing: Evaluating Effective Language Pairings for Fake News Detection Employing Zero-Shot Cross-Lingual Transfer," in *Communications in Computer and Information Science*, 2023. doi: 10.1007/978-3-031-27609-5_2.
49. X. Han, W. Zhao, N. Ding, Z. Liu, and M. Sun, "PTR: Prompt Tuning with Rules for Text Classification," *AI Open*, vol. 3, 2022, doi: 10.1016/j.aiopen.2022.11.003.
50. J. Yang, X. Hu, G. Xiao, and Y. Shen, "A Survey of Knowledge Enhanced Pre-trained Language Models," *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2024, doi: 10.1145/3631392.
51. J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," in *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 2018. doi: 10.18653/v1/p18-1031.
52. S. Gururaja, R. Dutt, T. Liao, and C. Rosé, "Linguistic representations for fewer-shot relation extraction across domains," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2023. doi: 10.18653/v1/2023.acl-long.414.
53. Z. Wang and D. Herscovich, "On Evaluating Multilingual Compositional Generalization with Translated Datasets," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2023. doi: 10.18653/v1/2023.acl-long.93.
54. M. Farokhian, V. Rafe, and H. Veisi, "Fake news detection using dual BERT deep neural networks," *Multimed Tools Appl*, vol. 83, no. 15, 2024, doi: 10.1007/s11042-023-17115-w.
55. J. Alghamdi, Y. Lin, and S. Luo, "Towards COVID-19 fake news detection using transformer-based models," *Knowl Based Syst*, vol. 274, 2023, doi: 10.1016/j.knosys.2023.110642.
56. G. Kuling, B. Curpen, and A. L. Martel, "BI-RADS BERT and Using Section Segmentation to Understand Radiology Reports," *J Imaging*, vol. 8, no. 5, 2022, doi: 10.3390/jimaging8050131.

57. Y. Wang *et al.*, "A comparison of word embeddings for the biomedical natural language processing," *J Biomed Inform*, vol. 87, 2018, doi: 10.1016/j.jbi.2018.09.008.
58. X. Shi, M. Hu, J. Deng, F. Ren, P. Shi, and J. Yang, "Integration of Multi-Branch GCNs Enhancing Aspect Sentiment Triplet Extraction," *Applied Sciences (Switzerland)*, vol. 13, no. 7, 2023, doi: 10.3390/app13074345.
59. Y. Zhao, T. Xue, and G. Liu, "Research on the robustness of neural machine translation systems in word order perturbation," *Chinese Journal of Network and Information Security*, vol. 9, no. 5, 2023, doi: 10.11959/j.issn.2096-109x.2023078.
60. I. Mohammed and R. Prasad, "Building lexicon-based sentiment analysis model for low-resource languages," *MethodsX*, vol. 11, 2023, doi: 10.1016/j.mex.2023.102460.
61. I. Haq, W. Qiu, J. Guo, and P. Tang, "NLPashto: NLP Toolkit for Low-resource Pashto Language," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 6, 2023, doi: 10.14569/IJACSA.2023.01406142.
62. Z. Zeng and S. Bhat, "Idiomatic expression identification using semantic compatibility," *Trans Assoc Comput Linguist*, vol. 9, 2021, doi: 10.1162/tacl_a_00442.
63. Seema Singh and Dr. Piyush Pratap Singh, "Transitive and Intransitive Verb Analysis for Idiomatic Expression Understanding: An NLP-based Framework," *International Journal of Advanced Research in Science, Communication and Technology*, 2023, doi: 10.48175/ijarsct-11682.
64. C. Raffel *et al.*, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, vol. 21, 2020.
65. R. Careem, G. Johar, and A. Khatibi, "Deep neural networks optimization for resource-constrained environments: techniques and models," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 3, 2024, doi: 10.11591/ijeecs.v33.i3.pp1843-1854.
66. J. Kreutzer *et al.*, "Quality at a Glance: An Audit of Web-Crawled Multilingual Datasets," *Trans Assoc Comput Linguist*, vol. 10, 2022, doi: 10.1162/tacl_a_00447.
67. C. Fuchs, "Cultural and contextual affordances in language MOOCs: Student perspectives," 2020. doi: 10.4018/IJOPCD.2020040104.
68. L. Wang, M. Feng, B. Zhou, B. Xiang, and S. Mahadevan, "Efficient hyper-parameter optimization for NLP applications," in *Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing*, 2015. doi: 10.18653/v1/d15-1253.
69. D. Dementieva and A. Panchenko, "Fake news detection using multilingual evidence," in *Proceedings - 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020*, 2020. doi: 10.1109/DSAA49011.2020.00111.
70. Z. Tan *et al.*, "Large Language Models for Data Annotation and Synthesis: A Survey," Dec. 2024, [Online]. Available: <http://arxiv.org/abs/2402.13446>
71. L. Alzubaidi *et al.*, "A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications," *J Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00727-2.
72. P. Joshi, S. Santy, A. Budhiraja, K. Bali, and M. Choudhury, "The state and fate of linguistic diversity and inclusion in the NLP world," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2020. doi: 10.18653/v1/2020.acl-main.560.
73. J. Edwards, *Multilingualism: Understanding Linguistic Diversity*. 2012. doi: 10.5860/choice.50-1302.
74. A. Abdedaïem, A. H. Dahou, and M. A. Cheragui, "Fake News Detection in Low Resource Languages using SetFit Framework," *Inteligencia Artificial*, vol. 26, no. 72, 2023, doi: 10.4114/intartif.vol26iss72pp178-201.
75. E. Raja, B. Soni, and S. K. Borgohain, "Fake news detection in Dravidian languages using transfer learning with adaptive finetuning," *Eng Appl Artif Intell*, vol. 126, 2023, doi: 10.1016/j.engappai.2023.106877.
76. F. Gereme, W. Zhu, T. Ayall, and D. Alemu, "Combating fake news in 'low-resource' languages: Amharic fake news detection accompanied by resource crafting," *Information (Switzerland)*, vol. 12, no. 1, 2021, doi: 10.3390/info12010020.
77. B. Zoph, D. Yuret, J. May, and K. Knight, "Transfer learning for low-resource neural machine translation," in *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2016. doi: 10.18653/v1/d16-1163.
78. Y. Chen *et al.*, "Cross-modal Ambiguity Learning for Multimodal Fake News Detection," in *WWW 2022 - Proceedings of the ACM Web Conference 2022*, 2022. doi: 10.1145/3485447.3511968.
79. K. Shu, S. Wang, and H. Liu, "Beyond news contents: The role of social context for fake news detection," in *WSDM 2019 - Proceedings of the 12th ACM International Conference on Web Search and Data Mining*, 2019. doi: 10.1145/3289600.3290994.
80. D. Dementieva, M. Kuimov, and A. Panchenko, "Multiverse: Multilingual Evidence for Fake News Detection +," *J Imaging*, vol. 9, no. 4, 2023, doi: 10.3390/jimaging9040077.

81. M. Park and S. Chai, "Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques," *IEEE Access*, vol. 11, 2023, doi: 10.1109/ACCESS.2023.3294613.
82. Y. Dou, K. Shu, C. Xia, P. S. Yu, and L. Sun, "User Preference-aware Fake News Detection," in *SIGIR 2021 - Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2021. doi: 10.1145/3404835.3462990.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.