Journal of Information Systems and Informatics

Vol. 6, No. 4, December 2024 e-ISSN: **2656-4882** p-ISSN: **2656-5935**

DOI: 10.51519/journalisi.v6i4.931 Published By DRPM-UBD

Misinformation Detection: A Review for High and Low-Resource Languages

Seani Rananga[1], Bassey Isong[2], Abiodun Modupe[3], Vukosi Marivate[4]

[1,2]Computer Science Department, North-West University, Mafikeng, South Africa
[1,3,4]Data Science for Social Impact & Computer Science Department, University of Pretoria, Pretoria, South Africa
Email: [1]seani.rananga@up.ac.za, [2]bassey.isong@nwu.ac.za, [3]abiodun.modupe@up.ac.za, [4]vukosi.marivate@up.ac.za

Abstract

The rapid spread of misinformation on platforms like Twitter, and Facebook, and in news headlines highlights the urgent need for effective ways to detect it. Currently, researchers are increasingly using machine learning (ML) and deep learning (DL) techniques to tackle misinformation detection (MID) because of their proven success. However, this task is still challenging due to the complexity of deceptive language, digital editing tools, and the lack of reliable linguistic resources for non-English languages. This paper provides a comprehensive analysis of relevant research, providing insights into advanced techniques for MID. It covers dataset assessments, the importance of using multiple forms of data (multimodality), and different language representations. By applying the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) methodology, the study identified and analyzed literature from 2019 to 2024 across five databases: Google Scholar, Springer, Elsevier, ACM, and IEEE Xplore. The study selected thirty-one papers and examined the effectiveness of various ML and DL approaches with a focal point on performance metrics, datasets, and false or misleading information detection challenges. The findings indicate that most current MID models are heavily dependent on DL techniques, with approximately 81% of studies preferring these over traditional ML methods. In addition, most studies are text-based, with much less attention given to audio, speech, images, and videos. The most effective models are mainly designed for high-resource languages, with English datasets being the most used (67%), followed by Arabic (14%), Chinese (11%), and others. Less than 10% of the studies focus on low-resource languages (LRLs). Therefore, the study highlighted the need for robust datasets and interpretable, scalable MID models for LRLs. It emphasizes the critical need to prioritize and advance MID research for LRLs across all data types, including text, audio, speech, images, videos, and multimodal approaches. This study aims to support ongoing efforts to combat misinformation and promote a more informed understanding of under-resourced African languages.

Keywords: Misinformation Detection, Low-Resource Languages, High-Resource Languages, African Languages.

1 INTRODUCTION

The advent of online social media platforms has transformed interpersonal communication [1, 2,]. These platforms, such as social media, blogs, and websites, provide individuals with easy access to news and information. However, this ease of spreading information has also facilitated the dissemination of false news. Anyone using these platforms can create and propagate false news for personal or professional purposes [1]. While users engage in communication, information sharing, and news consumption, a significant portion of the content that goes viral is unsatisfactory and sometimes harmful. To address this issue, researchers and governments are employing various methods to detect misinformation promptly. Machine learning (ML) and deep learning (DL) have garnered significant attention for their wide range of applications, including misinformation identification [2, 4-7]. Although effective, traditional ML techniques are becoming less suited to detect misinformation due to the increasing complexity of the data [2]. DL models are more suitable for this task, but there is a lack of research on DL for misinformation detection (MID) in African languages [2, 8, 9].

Recent studies [2, 9-11] highlighted two primary concerns: the lack of a clearly defined boundary between misinformation, disinformation, and mal-information, and the difficulty in evaluating reviews using DL techniques for MID in low-resource languages (LRLs). LRLs, not confined to Africa, include languages with limited linguistic resources globally. Thus, only about 20 out of more than 7000 languages spoken worldwide have extensive text collections [14, 20]. Systems capable of processing multiple languages are notably scarce within the context of MID, particularly for African languages due to the lack of data and resources [10, 19]. Moreover, misinformation is unintentionally false information spread without intent to deceive, while disinformation is deliberately false information meant to mislead [11, 12, 17]. Mal-information involves genuine information disseminated to cause harm [2] while distinguishing between these types is crucial for fake news detection research and designing information distribution platforms [14]. To address these menace, social context-based approaches, such as post-based and propagation-based methods, have been proposed to tackle fake news [2, 16]. However, reputable news outlets sometimes distribute false news without thorough verification, further eroding trust in information ecosystems. Confirmation bias also exacerbates this issue as people tend to accept and share information that aligns with their beliefs [10, 18].

As confirmed in this study, previous studies emphasized high-resource languages (HRLs) and existing techniques rather than addressing core issues in LRLs. This complicates the task of current models in detecting misinformation from social media comments, particularly considering the complexity of social media language, such as idioms, acronyms, slang, sarcasm, and irony. These nuances require sophisticated natural language processing (NLP) methods to accurately identify and eliminate misinformation. In addition, the traditional ML methods struggle to adequately address the complexity of MID. Advances in large-scale pre-trained models such as BERT and GPT-3 [2, 5], and adversarial learning techniques have made it more difficult to identify [2, 5]. This requires high-capacity models such as DL to combat the evolving landscape of misinformation effectively. Researchers and governments are increasingly focusing on DL techniques to detect misinformation before it causes harm. Traditional ML models, which often require manual feature extraction, are gradually being outperformed by DL models. These DL models automatically learn hierarchical features, making them far more effective for tasks like MID that involve unstructured data and complex patterns [14]. Despite the rise of DL methods, the spread of false information continues to grow at an alarming rate. Identifying misinformation, disinformation, and mal-information is essential to maintaining the integrity and trustworthiness of information sources. The precise distinction between these types is fundamental for effective detection and minimising the harmful effects of fake news [8, 14].

In this paper, we delved into how ML/DL has been harnessed to detect misinformation by compiling state-of-the-art research in MID. The study emphasizes the importance of employing both traditional ML methods and advanced DL models to combat the ever-evolving nature of fake news, allowing for a nuanced and effective approach to tackling digital misinformation [16, 22]. We explored the various ML/DL techniques used to detect false information, evaluated their effectiveness, and analysed important components such as data types, datasets, and performance metrics. It also highlighted the challenges of detecting misinformation in LRLs, particularly African languages, and suggested future research directions to overcome these issues and improve detection accuracy and effectiveness. Given the rapid proliferation of misinformation, especially in developing countries, it is imperative to develop robust and adaptable tools to combat this issue. However, in spite of the significant progress in MID for HRLs like English, Arabic, and Chinese, there's a stark gap in research and resources for LRLs, which make up most of the world's languages. This disparity is due to the lack of datasets, tools, and tailored methods for LRLs, and the overemphasis on HRLs in publicly available datasets and advanced MID models. Furthermore, multilingual data presents further challenges, as differences in syntax, semantics, and cultural contexts complicate the application of models developed for HRLs to LRLs. For African languages, these issues are exacerbated by the lack of standardized writing systems, no or limited annotated datasets, and the prevalence of code-switching in many communities and so on. This motivates the urgent need for robust, scalable, and interpretable MID models tailored to LRLs. This study provides a comprehensive understanding of the current state of MID research for both HRLs and LRLs. The following research questions (RQs) were addressed.

RQ1: What are the latest trends in research on detecting misinformation in LRLs?
RQ2: How are researchers working to control the spread of misinformation?
RQ3: How are these models tested and evaluated?
RQ4: What are the emerging trends in combating misinformation?

The key contribution of this paper is, therefore, summarised as follows:

1) We comprehensively explored the latest trends in MID research and analysed how researchers addressed it focusing on HRLs and LRLs.
2) We bring together different methodologies, datasets, and evaluation metrics used in MID.
3) The study identified emerging trends in combating misinformation for HRLs and LRLs and highlighted potential areas for future research to bridge the gap such as LRLs for African languages. This is because, LRLs which are spoken by millions, lack representation in research and technology. Misinformation across text, audio, images, and video needs culturally inclusive solutions. Hence, focusing research on LRLs, especially African languages can bridge the research gap and equip the underrepresented in combating misinformation.

The remaining parts of the paper are structured as follows: Section 2 discusses the related works, Section 3 presents the research methodology employed, Section 4 presents the results of the review based on the RQs and Section 5 presents the paper discussion, future directions, and the validity threats of this study. In addition, Section 6 presents the paper's conclusion.

2 RELATED WORKS

Recent years have seen significant advancements in detecting misinformation within LRLs. Traditional ML techniques, such as support vector machines (SVM), decision trees (DT), random forests (RF), Naive Bayes (NB), and logistic regression (LR), rely heavily on manually engineered features to make predictions or classifications [2, 5]. While effective in various applications, these methods may struggle with the nuances of misinformation, especially with complex data and natural language processing (NLP) tasks [2]. DL models, on the other hand, utilize neural networks with multiple layers to extract features from raw data, identifying intricate patterns [2]. Unlike traditional ML models, DL models like convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer-based models (e.g., BERT and GPT) can handle unstructured data such as text, images, and speech [2, 5]. These models excel at understanding context, semantics, and syntactic structures within text, making them highly effective in detecting misleading or false information [5].

A study by [23] emphasized the need for advanced methodologies to combat misinformation in low-resource contexts, focusing on tasks like offensive language detection, fake news detection, and rumour detection. Similarly, Ricketts [18] highlighted the low rates of bias detection and the need for more specific detection instructions. Another study [20] suggested using summarization models for feature extraction to identify central claims in texts, crucial for rumour detection. Other efforts include utilizing HRL resources for low-resource hypernymy detection [24] and constructing annotated datasets for detecting COVID-19-related fake news in multiple Indic languages [25].

Ghafoor et al. [26] explored the impact of translating datasets from HRLs to LRLs via multilingual text processing. Their findings underscored the contributions of DL methods and word embeddings towards developing automated fake news detection mechanisms. This led to the creation of the Amharic fake news detection model, a general-purpose Amharic corpus, a novel fake news detection dataset (ETH_FAKE), and Amharic fast text word embeddings (AMFTWE) [14]. Despite efforts by major platforms to debunk COVID-19 misinformation, much fact-checking information remains predominantly in English [27]. To tackle misinformation in other languages, Du et al. [28] attempted to detect COVID-19 misinformation in Chinese using English fact-checked news. Another study [29] introduced a multilingual dataset of COVID-19 vaccine misinformation from Brazil, Indonesia, and Nigeria, using domain-specific pre-training and text augmentation. Finally, a comparative analysis by [30] evaluated the effectiveness of chatbots like ChatGPT and Bing Chat in discerning political information accuracy across languages.

These discussions reflect ongoing reviews and surveys in MID, accenting the need for more systematic literature reviews to enhance advancements in the field. This paper aims to bridge that gap, providing a comprehensive overview of current MID research and methodologies.

3 METHODOLOGY

This study conducted a systematic analysis of the existing literature on MID to answer the defined RQs. The RQs were designed to highlight the publication trends of MID for HRLs and LRLs, the different ML/DL techniques employed in evaluation and validation methodologies, and opportunities for future research directions in this field. To examine the publication trends of MID, we utilized the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement template to elucidate the overall process of selection and exclusion of articles for review in this study [31]. The PRISMA statement aids in improving the reporting of systematic reviews, and the SLR focuses exclusively on peer-reviewed publications [31]. The comprehensive PRISMA process for this study involved reviewing publications from the years 2019 to 2024, as depicted in Figure 1. For this study, a research strategy was developed to identify pertinent literature across five reputable electronic databases, including Google Scholar, Springer, Elsevier, ACM ScienceDirect, and IEEE Xplore. The search terms used were "Misinformation detection," "MID" "Fake news detection," "high-resource languages", and "low-resource languages," guided by Boolean operators.

3.1 PRISMA Study Selection

The identification and selection of relevant articles followed the PRISMA process, as shown in Figure 1. Initially, 200 relevant articles were identified through database searches, with an additional 10 articles identified via snowballing techniques. After removing duplicates (n=5), 195 articles were screened for eligibility. A total of 166 studies were excluded for various reasons, including failure to address future research directions, lack of relevant models or data types, or duplicate contributions. Ultimately, 31 studies met the inclusion criteria and were selected for inclusion in this review. The inclusion criteria for this study were the model used in terms of the applied ML/DL models for MID and the relevant datasets for evaluating MID, the research focused on both developed and developing countries, with more interest on LRLs and future work where the study discussed potential future directions for research in MID. Moreover, the the exclusion criteria included studies that: did not discuss future work, did not use relevant models or data types for MID and duplicated contributions from other included articles.

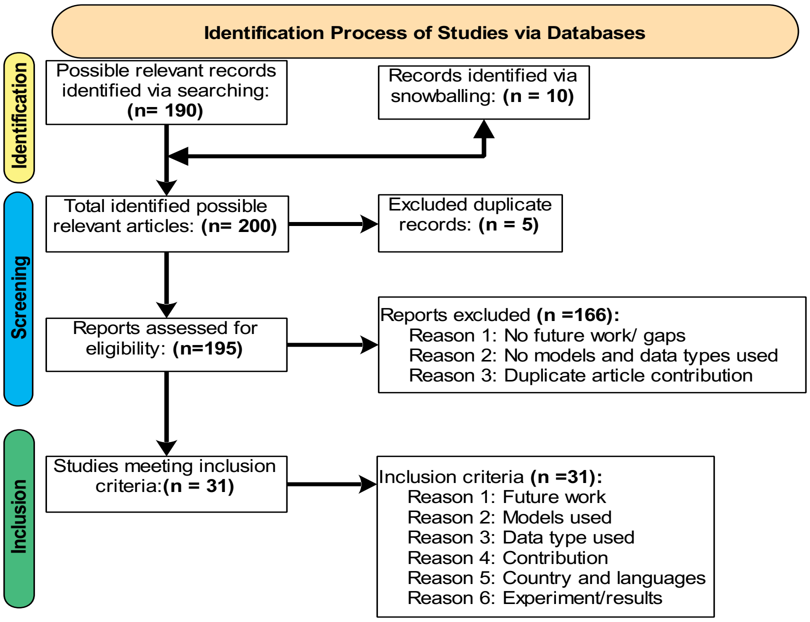
Identification Process of Studies via Databases
Identification
Screening
Inclusion
Possible relevant records identified via searching: (n= 190)
Records identified via snowballing: (n = 10)
Total identified possible relevant articles: (n= 200)
Excluded duplicate records: (n = 5)
Reports assessed for eligibility: (n=195)
Reports excluded (n =166):
Reason 1: No future work/ gaps
Reason 2: No models and data types used
Reason 3: Duplicate article contribution
Studies meeting inclusion criteria:(n = 31)
Inclusion criteria (n =31):
Reason 1: Future work
Reason 2: Models used
Reason 3: Data type used
Reason 4: Contribution
Reason 5: Country and languages
Reason 6: Experiment/results

Figure 1. PRISMA Workflow

In addition, the key studies evaluated MID methods across HRLs and LRLs, assessing methodologies based on language resources, modalities (e.g., text, audio, images, speech), and models applied. Findings were summarized in Table 1 and categorized to address RQs, revealing trends, methods, and evaluation metrics like accuracy and F1-score, mostly for HRLs.

3.2 Data Extraction and Analysis

We carefully examined each study and extracted key information, including:

1. Study type by where it focused on evaluating models, transferring learning across languages, or combining different approaches.
2. The research focus includes main topics, such as detecting misinformation in multiple languages, identifying AI-generated fake news, and using large language models.
3. Key contributions involved innovative ideas, especially those addressing misinformation in languages with limited resources.
4. Research gaps are areas where more research is needed, such as developing stronger DL models for these languages.
5. Quality assessment evaluates each study's methodology and its connection to previous research.

We used a systematic approach of PRISMA to analyse the collected data and answer the defined RQs. This helped us identify trends, evaluation techniques, challenges, and opportunities for future research in MID, particularly for LRLs. Furthermore, to ensure the focus of our study, we only included research papers written in English, while five articles that were not written in English were excluded from the study. The rationale for defining the inclusion criteria, as depicted in Figure 1, is to be able to answer the research questions for this study. The ability to identify the future work mentioned in a research publication enables the outlining of potential gaps to be answered based on a specific study per the researcher's interest. Similarly, understanding the model used in the research allows for the identification of relevant models and the data used for our study from the body of knowledge. The subsequent section outlines the outcomes derived after the utilization of the PRISMA process, explaining comprehensively how each RQ was addressed within the ambit of this study. This explanation holds paramount significance as the results section represents the core of this study.

4 STUDY ANALYSIS FROM THE SELECTED PAPERS

This section presents the findings of the analysis conducted in this paper to answer the RQs defined. The PRISMA chart presented in Figure 1 delineated the selection strategies employed for selecting the relevant studies considered for this study, thereby addressing RQ1 as the publication trends of MID for LRL. Moreover, we analysed in-depth the contributions, evaluations, and effectiveness of these studies to address RQ2 and RQ3. By analysing the different strategies employed to combat misinformation, we aim to provide a comprehensive overview of the field.

4.1 What are the latest trends in research on detecting misinformation in LRLs (RQ1)

This section analyses the trends in MID research for HRLs and LRLs from 2019 and 2024. Figure 2 highlights the significant gap between LRLs and HRLs, especially for African languages. From 2019 to 2024, MID research for LRLs showed varied progress. Initially, interest was minimal, with only 4 publications in 2019 and none in 2020 which could potentially be due to inclusion criteria or database limitations. A modest increase in 2021 and 2022, with 4 and 2 publications respectively, indicated a growing recognition of MID's significance in LRLs. In 2023, the field saw a significant increase in the number of publications, resulting in heightened awareness, collaborative initiatives, and advancements in NLP technologies. The most significant rise occurred in 2024, with 15 publications, highlighting a growing focus on addressing misinformation in LRLs through improved international collaboration, increased funding, and extensive datasets and tools.

Despite these positive developments, the ongoing gap between HRLs and LRLs indicates a need for continued efforts. The challenges such as limited resources and infrastructure, particularly in areas such as Africa, continue to impede progress. To address this issue, improving data availability, infrastructure, and funding is essential. This will lead to effective solutions to combat misinformation across a variety of languages, addressing the global challenge more effectively.

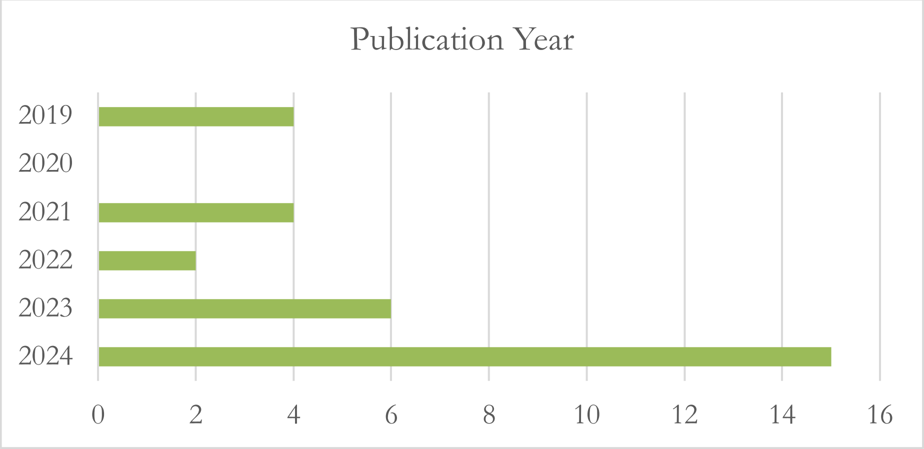

Figure 2. Year-wise publication trend

4.2 How are researchers working to control the spread of misinformation (RQ2)

This section presents the key contributions of each study included in this paper which aimed at addressing misinformation. We focused on the methods and techniques employed and provided a detailed evaluation of their approaches. The summary of the findings is presented in Table 1.

Raja et al. [10] developed a novel method to detect fake news in Indian languages such as Telugu, Tamil, Kannada, and Malayalam. They utilized a powerful language model called MuRIL to comprehend complex language patterns. Their model which is a combination of CNNs and LSTMs, captured both short-term and long-term dependencies in the text. When evaluated on a large dataset, the model significantly outperformed existing methods, achieving a high accuracy rate of over 99% for each language. Similarly, Alghamdi et al. [3] focused on creating a multilingual fake news detector. They utilized a transformer-based model and a hybrid approach to summarize text to improve accuracy and minimize data noise. Their model outperformed popular language models like mBERT and XLM-RoBERTa, with accuracy and F1 scores between 93% and 95%. Rashid et al. [9] also tackled Bengali fake news by developing a large dataset of fact-checked news articles. They used advanced language models to detect fake news, and although a traditional LSTM model performed well, a more advanced model called Bangla BERT, based on the transformer architecture, achieved an accuracy of almost 99%.

Malik and Kumar [15] suggested a hybrid approach to detect fake news by combining Word2Vec and LSTM. Word2Vec converts words into meaningful vectors, while LSTM processes these vectors to understand the text's context. Their model was evaluated on several datasets and consistently outperformed other methods, including pre-trained models like BERT, achieving excellent accuracy rates of over 99%. The success can be attributed to the effective use of Word2Vec and LSTM in extracting key information from the text. Authors in [13] addressed the problem of COVID-19 misinformation in Arabic by developing a system called AraCovTexFinder, specifically designed for Arabic text. AraCovTexFinder outperformed other language models, including mBERT and XLM-RoBERTa, with an accuracy of nearly 99%. To increase the model's performance, they created two new datasets for Arabic language models. Hashmi et al. [12] presented a robust fake news detection approach using FastText word embeddings in combination with both ML and DL techniques. Tested on three datasets such as WELFake, FakeNewsNet, and FakeNewsPrediction, the hybrid CNN-LSTM model, improved with FastText embeddings, outperformed other models, achieving accuracy and F1-scores of 0.99, 0.97, and 0.99, respectively. Transformer models like BERT, XLNet, and RoBERTa were also tested, and the tuned models consistently topped traditional RNN-based approaches. In the same vein, Luvembe et al. [22] introduced a new model called Complementary Attention Fusion and an Optimized DNN (CAF-ODNN) to detect fake news using multiple types of information, such as text and images. The CAF-ODNN model excelled at combining different types of information to make more accurate predictions, outperforming other models on several datasets and achieving an impressive accuracy rate of up to 90% on the Fakeddit dataset. Similarly, Yan et al. [32] focused on improving the quality of training data for fake news detection models by developing techniques to select the most relevant synthetic data. Fine-tuning with this selected synthetic data produced strong results, with SemSim achieving an F1 score of 0.687 on the MediaEval dataset and DisSim reaching 0.813 on the Snopes dataset.

Al-Zahrani et al. [11] also tackled the challenge of detecting fake news in Arabic. They utilized various language models and discovered that combining multiple models significantly enhanced the accuracy of fake news detection in Arabic. Their ensemble model achieved an F1 score of 94% on the Arabic Multisource Fake News Detection (AMFND). Similarly, the authors in [17] proposed a framework that combines traditional ML techniques with DL to analyze text, images, and videos. This method enabled a more comprehensive analysis of fake news, resulting in more accurate detection. The model was tested on several datasets which showed promising results, with RF achieving 99% accuracy on the text and the multimodal approach improving accuracy by 3.1% over existing models. Moshen et al. [16] also presented an innovative method to detect fake news by combining traditional methods with more advanced ones. They used a combination of language-based features and ML models, such as NB and Gradient Boosting, to identify false news. This approach was particularly effective in minimizing the training time while maintaining high accuracy. Bernoulli NB achieved 89% accuracy and notable reductions in training time. Similarly, Zeng et al. [7] improved the detection of fake news that includes both text and images by developing a method to select the most relevant synthetic data to train their model. This led to improved performance, even surpassing more complex models such as GPT-4V on real-world datasets. SemSim achieved an F1 score of 0.687 on the MediaEval dataset and DisSim achieved 0.813 on the Snopes dataset after fine-tuning with selected synthetic data.

Farhangian et al. [2] developed a new technique that combines CNNs with Dynamic Time Warping (DTW) to identify patterns indicative of fake news. By incorporating additional techniques like Bottle-Neck Features (BNFs) and a Contractive Auto-Encoder, they significantly improved their model's accuracy compared to traditional methods. The results showed a 5% relative improvement in performance over MFCCs, with BNFs providing a 27% improvement in ROC and AUC, leading to significantly more accurate top-10 retrievals. Another study [3] focused on using both text and images to detect fake news. They experimented with models like ResNet50, VGG16, and EfficientNet, and found that combining text features from BERT with image features from ResNet50 provided the best results, achieving an average accuracy of 80.7%. Similarly, Salau et al. [6] aimed to detect fake news early on using a DL model based on CNNs to analyze various data types. This model achieved over 99% accuracy in detecting misinformation before it could spread widely.

Authors in [35] combined multiple ML models to enhance fake news detection through an ensemble learning approach, including LR, SVM, linear discriminant analysis, stochastic gradient descent, and ridge regression. This ensemble achieved an accuracy of over 98%. Hansrajh et al. [36] identified new features to enhance fake news detection. When combined with traditional classifiers, these features helped more accurately distinguish between real and fake news. However, it's important to note that while this approach was successful, it misclassified 40% of real news as fake in their study [36]. Reis et al. [37] used a combination of Bidirectional Long Short-Term Memory (BiLSTM) and CNN models to detect rumours on Twitter. By taking the context of the tweets into account, their model achieved an accuracy of 86.12% in classifying tweets as rumours or non-rumours. Similarly, Asghar et al. [38] employed a combination of RNNs and semantic models to detect fake news on Twitter, achieving a remarkable accuracy of 99% in distinguishing between fake and real news. Another approach [39] combined content analysis with social context to detect fake news. This model analyzed both the content of news articles and their spread on social media, and when tested on real-world datasets like BuzzFeed and PolitiFact, it achieved high training accuracies of 99.04% and 99.31%, and validation accuracies of 86.49% and 88.64%, respectively. Jadhav et al. [39] developed a method using capsule networks to detect fake news, utilizing various techniques to extract features from the text. This model outperformed existing methods on several datasets, achieving validation accuracies of 7.8% on the ISOT dataset and 3.1% on the LIAR dataset, with a final test accuracy of 1%.

In another study, researchers focused on detecting fake news in Hindi, a language with limited resources. They used an ensemble learning technique, which combines multiple models to improve accuracy, achieving over 90% accuracy in detecting fake news in Hindi [41]. Similarly, studies in the Arabic language [23, 33] explored using both text and images from tweets to detect rumours but found that using just the text was more effective, achieving an accuracy of 89.64%. Another study [21] introduced the MCred model, which analyzes the meaning of words and sentences to determine the authenticity of the news, achieving impressive accuracy rates on various datasets, including 99.46% on the Kaggle dataset. A recent study [34] used a combination of techniques, including topic modelling and DL, to improve fake news detection. Similarly, Lin et al. [14] focused on detecting rumours on social media, particularly in low-data situations. They developed a method combining different techniques to improve accuracy, achieving 89.9% accuracy on Chinese COVID-19 datasets and 77.3% on English COVID-19 datasets. Another study [19] aimed at detecting fake news in Hindi combined CNN and Bidirectional Long Short-Term Memory (BLSTM) models to improve accuracy, achieving significant improvements of 5.8% for CNN and 10% for DNN. Similarly, Dlamini et al. [43] worked on detecting fake news in languages with limited resources like Zulu. They used a transfer learning technique, training a model on a language with more data (like English) and adapting it to a language with less data. The results were promising, with an accuracy of 54.5%, although there's still room for improvement.

Van der Westhuizen et al. [35] combined techniques like CNNs and DTW to improve the accuracy of fake news detection, particularly in identifying patterns within data. Other studies [4, 8] emphasized the importance of using both text and images to detect fake news. By integrating models such as BERT and ResNet50, these studies achieved high accuracy rates. One study [8] focused on early fake news detection, using a CNN-based model to analyze various data types, and achieving over 99% accuracy. Another method [36] combined multiple ML models to enhance accuracy, resulting in an ensemble method that achieved over 98% accuracy. Reis et al. [37] identified new features that, when combined with traditional classifiers, significantly improved fake news detection. This approach detected all fake news data but misclassified 40% of true news. Asghar et al. [38] combined BiLSTM and CNN techniques to detect rumours on Twitter, achieving an accuracy of 86.12% in classifying tweets as rumours or non-rumours. Jadhav et al. [39] utilized a DL model to analyze Twitter reviews, achieving an impressive 99% accuracy in distinguishing fake news from real news. Concurrently, the DeepFakE model [40] was developed to detect fake news by analyzing both content and its spread on social media. This model achieved high training accuracies of 99.04% and 99.31% on datasets like BuzzFeed and PolitiFact, with validation accuracies of 86.49% and 88.64%.

These studies highlight significant advancements in the field of fake news detection. Researchers have explored a variety of techniques, including traditional ML, DL, and hybrid approaches, to identify and mitigate misinformation. Key findings include the effectiveness of combining multiple models, such as CNNs and LSTMs, for improved accuracy. Additionally, integrating social context and linguistic features has proven beneficial in detecting fake news. While challenges remain, particularly in LRLs and multimodal settings, ongoing research continues to push the boundaries of fake news detection, offering promising solutions to combat the spread of misinformation.

5 ANALYSIS

This research investigated the landscape of MID, particularly focusing on LRL. We analysed 31 studies published between 2019 and 2024, as detailed in Table 1.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

This section provides an analysis of the findings across all papers considered. We analysed the findings based on misinformation types, inputs, languages, ML/DL models, benchmarked datasets, and performance evaluation metrics used.

5.1 Misinformation types and impacts

Recent research has explored various methods for detecting misinformation, categorized into types such as rumours (unverified stories spread from person to person) [25, 41], fake news (intentionally false news articles) [1, 3, 11, 27, 28], and spam (unwanted messages often used for malicious purposes) [2]. Disinformation, which involves spreading false information deliberately to deceive, differs from misinformation in its intent [2]. As shown in Figure 3, the analysis shows a bias toward English-language studies (67%), followed by Arabic (14%), Chinese (11%), Hindi (5%), and a variety of other languages (3%) including Zulu, Luganda, Tamil, Cantonese, Swahili, Bengali, Afan Oromo, Vietnamese, Malayalam, Telugu, Kannada, and Indonesian. This bias leaves speakers of LRLs more vulnerable to misinformation due to the lack of resources.

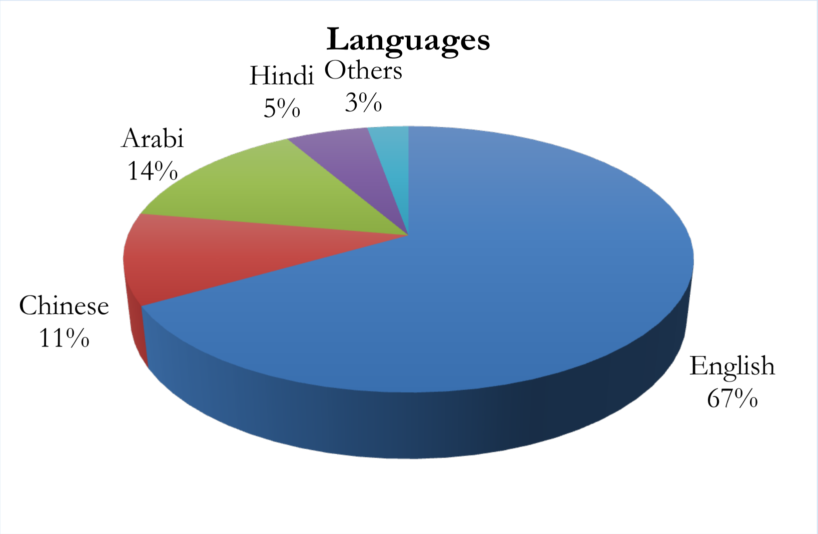

Figure 3. Different languages used for MID

Moreover, most studies (73%) relied on text data, with only 2% exploring speech/audio, 5% analyzing video, and 20% involving image analysis as shown in Figure 4. Non-English studies often lacked proper validation (3%), raising questions about their effectiveness. However, 80% of the datasets used were publicly available, which aids future research.

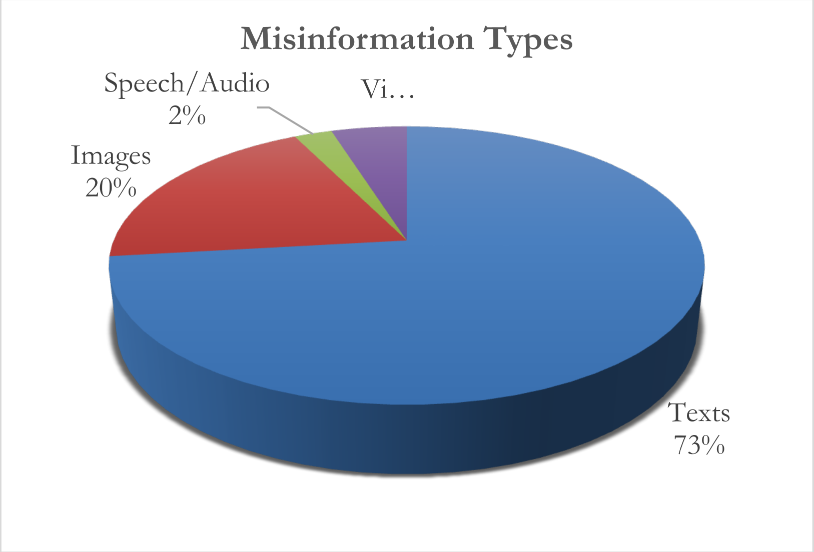

Figure 4. Misinformation types

The research identifies various misinformation types, with fake news and rumours being the most studied (77% and 11%, respectively). (See Figure 5) Other forms like spam and disinformation, including hate speech and offensive posts, received less attention. Misinformation can have significant consequences in social, political, economic, and public safety domains, manipulating public opinion and influencing elections. Therefore, detecting misinformation, particularly in LRLs, is crucial for promoting a more informed and resilient society. Despite the current focus on English, diversifying research efforts to include LRLs is essential to effectively address the global challenge of misinformation.

Despite numerous techniques for detecting misinformation, not all methods are effective [42]. Text-based misinformation spreads quickly through social media, forums, and news sites, using persuasive language and emotional appeals to influence public opinion [42]. It often manipulates reader perceptions by selectively presenting facts [42]. Furthermore, image-based misinformation involves deceptive images that convey false narratives, commonly on social media. Visual misinformation exploits human susceptibility to visual cues, making it challenging and challenging to detect [41]. Audio-based misinformation includes false or misleading information in recorded sound. With advanced audio editing technology, it's becoming more difficult to distinguish real audio from manipulated recordings. This can spread rumours, defame individuals, or fabricate events, resulting in significant threats to public trust and social cohesion [21]. Speech-based misinformation is spread through spoken communication, such as speeches, interviews, and podcasts. It can reach large audiences and shape public discourse, often using rhetorical techniques like repetition to reinforce false beliefs and undermine factual accuracy [34]. Each type of misinformation utilizes different aspects of human perception, making it a complex issue to address.

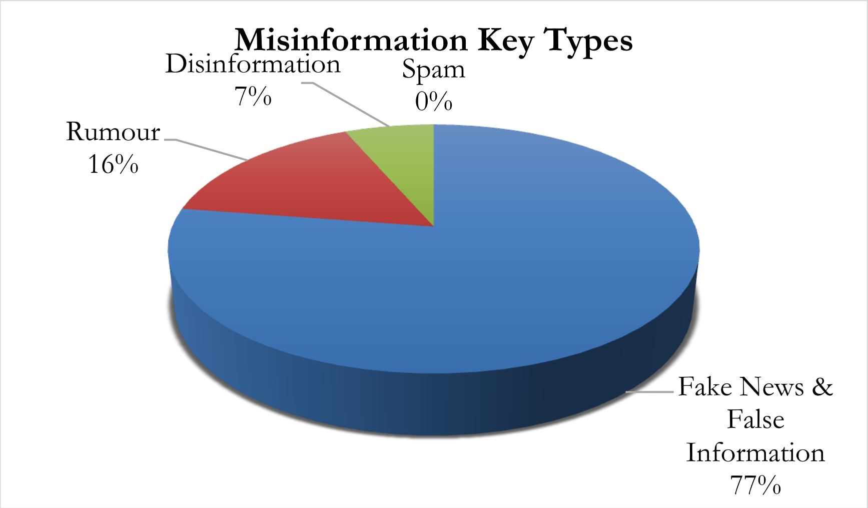

Figure 5. Misinformation input types

5.2 MID detection models

The analysis performed highlights the variety of models and classifiers used in MID to tackle the different formats misinformation can take, such as text, images, audio, and speech. Each format requires distinct models for effective detection due to its unique features and patterns. Studies utilized various models and classifiers, with some using a single model and others combining multiple models to enhance accuracy. Based on Table 1, Figures 6 and 7 summarize the different models based on input types, whereas Figure 7 revealed DL models make up about 81.25% of the models used in MID. These DL models, including BERT [21], DNN [9], RoBERTa [21], LSTM [34], CNN [14], VGG [2], Inception, EfficientNet, Transformers, Automatic Speech Recognition (ASR), and RNN [14], are highly effective at handling unstructured data like text, images, and speech. They automatically learn hierarchical features from the data, making them well-suited for the complexities of misinformation detection.

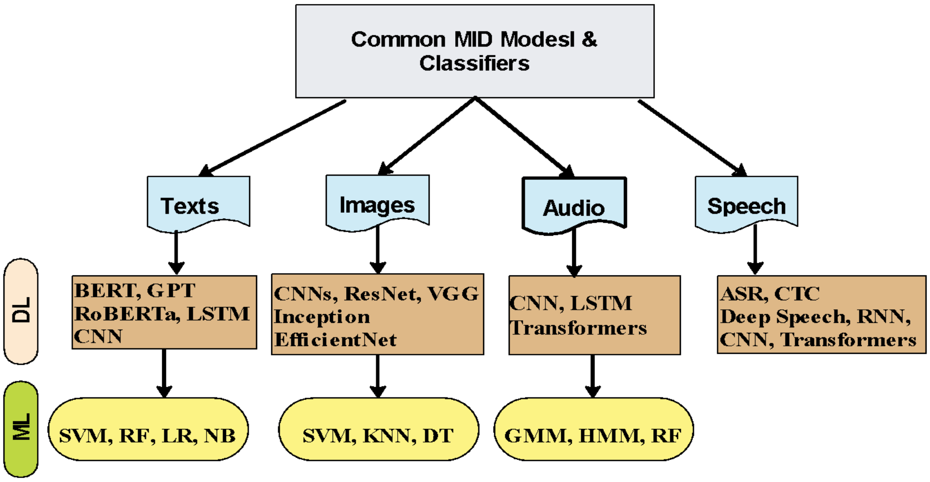

Figure 6. Commonly used MID models

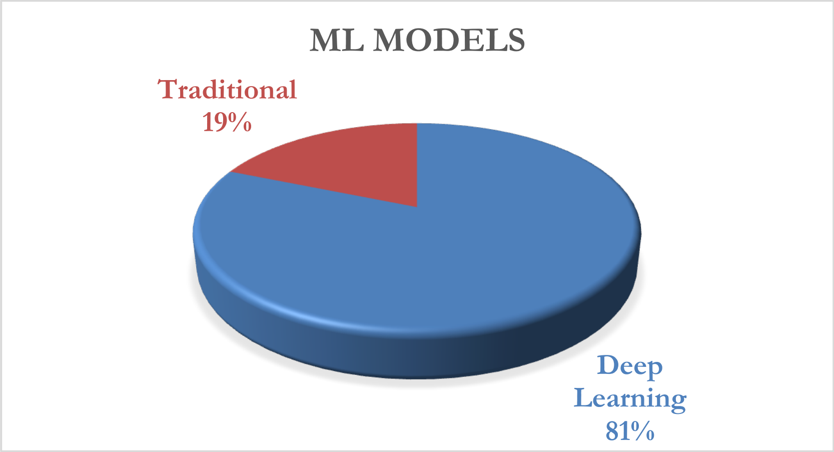

Figure 7. DL vs traditional ML models for MID

Conversely, traditional ML models account for 18.75% of the models used for MID. These models, such as SVM [5], RF [5], LR [27], and NB [12], are known for their simplicity and effectiveness in various applications. However, they often require manual feature extraction, which can be detrimental to their ability to handle the diverse and complex nature of misinformation. Transfer learning techniques are frequently employed in LRL settings, where pre-trained models from HRLs are fine-tuned on smaller datasets in LRLs. Models such as multilingual BERT (mBERT) and XLM-RoBERTa [32] are preferred for their language-agnostic abilities.

By combining multiple approaches and models, researchers can further enhance MID. A multimodal approach that incorporates textual analysis, image processing, and audio/speech recognition provides a more comprehensive understanding of the content, enhancing detection accuracy. The use of both traditional ML methods and advanced DL models helps researchers develop more robust MID systems capable of addressing the multifaceted nature of misinformation. This approach is crucial for tackling the complexities and ensuring the effectiveness of MID across various formats and languages, helping to keep pace with the evolving landscape of misinformation and creating a more resilient information ecosystem.

5.3 Benchmark datasets

The development of effective MID tools is heavily dependent on the availability of high-quality datasets. In Table 1, numerous valuable datasets are available for MID research, enabling researchers to evaluate the effectiveness of various ML models for detecting misinformation. However, existing MID techniques vary widely in their approaches and data collection methods. The datasets span a range of topics and languages but show a significant bias towards HRLs like English, Arabic, and Chinese. This bias makes it challenging to obtain data for LRLs, as researchers often have to rely on machine or manual translation, which can be expensive and time-consuming. Additionally, small and low-quality datasets can hinder the ability of ML models to capture language nuances and accurately identify misinformation patterns.

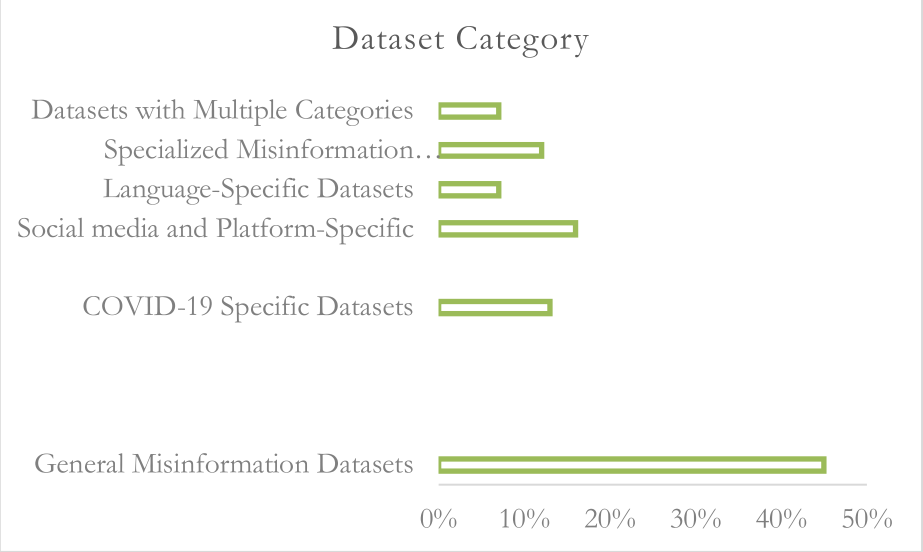

Figure 8. Commonly used MID dataset categories

Based on the analysis of the studies considered, the available dataset types used by researchers are shown in Figure 8. The analysis revealed that about 45% of the datasets focus on general misinformation detection, including Kaggle, TICNN, ISOT, SMS Spam, WELFake, FakeNewsNet, FakeNewsPrediction, Fakeddit, Pheme, GossipCO, PolitiFact datasets, ReCOvery, GossipCop, MR2, AMFND dataset, FA-KES, fake and Real Dataset, trimmed-WELFake, trimmed-Scraped, trimmed-Kaggle1, Snopes (O+), LIAR dataset, SemEval-2016 dataset, BuzzFeed, George Mclntire, KaggleMclntire, FakeNews, ELFake, and GitHub. Around 13% are specifically tailored to COVID-19 related misinformation, such as English-COVID19, Chinese-COVID19, Cantonese-COVID19, Arabic-COVID19, Twitter COVID-19, Weibo-COVID19, and COVID, Twitter-COVID19. Moreover, 16% are derived from social media and various platforms, including Facebook, MediaEval 2016, MediaEval, Weibo-hybrid, Weibo, Twitter, and Twitter-COVID-19. About 7% cater to specific languages, highlighting the underrepresentation, including datasets like Arabic Fake News Dataset, South African Broadcast News, Radio broadcasts in Kampala, and Artefacts. Moreover, 12% are specialized misinformation datasets such as the Pheme rumour dataset, Fake-news, Fake_real_news, PolitiFact, and FEVER, while 7% fall into multiple categories, like ISOT and Kaggle.

The underrepresentation of LRLs poses several challenges for developing tailored MID techniques. Creating labelled datasets for these languages is often expensive and time-consuming, especially for languages spoken by smaller communities. Furthermore, ethical considerations, such as privacy and bias, must be addressed when collecting data from different languages. Despite the various methods available for detecting misinformation, no one is effective. Text-based misinformation, for instance, spreads quickly through social media, forums, and news sites, using persuasive language and emotional appeals to influence public opinion. Image-based misinformation, involving deceptive or altered images, is common on social media and utilizes the brain's ability to perceive visual cues, making it highly impactful and challenging to detect. Audio-based misinformation includes misleading content in recorded sound, which can be increasingly difficult to comprehend due to advanced audio editing techniques. Furthermore, speech-based misinformation, spread through spoken communication, shapes public discourse and often reaffirms false beliefs.

Although the availability of diverse datasets has significantly contributed to understanding and combating misinformation, there is still a need to expand resources and efforts towards LRLs. It is essential to develop robust MID tools that can cater to a wider range of languages, ensuring a more informed and resilient global society.

5.4 Evaluation metrics

In the realm of NLP, MID presents a formidable challenge, and ML and DL techniques are extensively utilized to combat it. The effectiveness of these methods is evaluated using multiple critical metrics, providing a comprehensive overview of their performance in various situations. These evaluation metrics include:

1. *Accuracy:* Computes the ratio of correctly identified instances, whether they are incorrect or not.
2. *Precision:* Measures the fraction of incorrect misinformation cases among all instances labelled as such by the model.
3. *Recall:* Determine the percentage of accurate misinformation cases accurately identified by the model.
4. *F1 Score:* The harmonic mean of precision and recall which provides a balanced single metric for evaluation where a higher score indicates better model performance [41, 42].

The F1 score is commonly used for binary classification tasks, while the micro F1 score is preferred for multi-classification tasks. High accuracy indicates the model's ability to distinguish between true and false information. High recall and precision values suggest that the model is proficient in capturing a large amount of false information and accurately labelling instances, minimizing false alarms. In addition, the Area Under the Curve (AUC), utilized with Receiver Operating Characteristic (ROC) curves, evaluates the model's performance across various decision thresholds. The ROC curves demonstrate the trade-off between the true positive rate (TPR) and the false positive rate (FPR) at various thresholds, while AUC provides a single value indicative of the classifier's overall effectiveness. The confusion matrix depicts the model's performance, showing true positives (TP), false negatives (FP), true negatives (TN), and false negatives (FN) [41]. The Matthews Correlation Coefficient (MCC) is a balanced measure of classification performance, especially for imbalanced datasets. This offers insights into how well the model identifies false information while minimizing false alarms and ensuring high precision. Furthermore, the equal error rate (EER) indicates the point where the model is likely to make an FP error as it is to make an FN error, with a lower value reflecting improved performance.

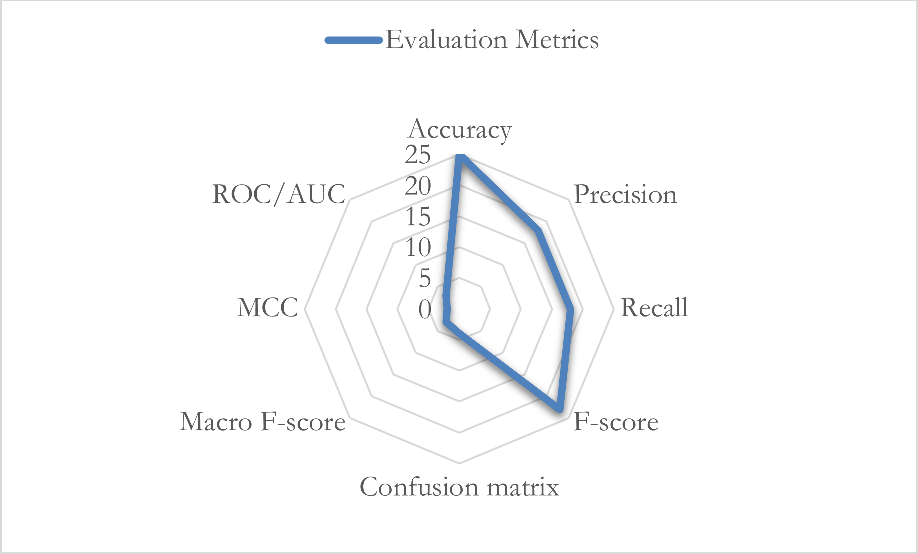

Figure 9. Frequency of performance measures

In examining various studies on MID, Table 1 and Figure 9 highlight the commonly used metrics and their frequencies. About 26.0% of studies used accuracy to evaluate their models, 18.8% utilized both precision and recall, and 24.0% relied on the F1 score. Other metrics included the confusion matrix (4.2%), macro F1 score (3.1%), MCC (2.1%), and ROC/AUC (3.1%). The choice of metrics typically aligns with the specific needs and error impacts of the application. For balanced datasets, accuracy, precision, recall, and the F1 score are effective, but for imbalanced datasets, accuracy can be misleading, making precision, recall, F1 score, and MCC more reflective of the minority class's performance. In binary classification, precision, recall, and F1 score are particularly useful, while the macro F1 score offers a comprehensive evaluation in multi-class scenarios. ROC curves and the AUC metric help analyze a model's behaviour across different decision-making thresholds.

6 DISCUSSION AND FUTURE OPEN ISSUES

6.1 Discussion

MID is a critical field of research due to the rapid spread of inaccurate or misleading information, particularly on social media platforms. The objective of MID is to identify and counter false narratives that could negatively impact public opinion and decision-making processes. This paper reviewed existing MID research, offering insights into the achievements, involved elements, and future research directions aimed at advancing the field. The findings of this study, based on the RQs defined in Section 1 of this paper are as follows.

Regarding RQ1, the analysis reveals a significant lack of publications on MID for African LRLs. Most existing research concentrates on HRLs such as English, Chinese, and Arabic. This underscores the need for more MID research in African languages to effectively counter misinformation in these communities. For RQ2, the study identifies various forms of misinformation, including text-based, image-based, audio-based, and speech-based, each presenting unique detection challenges and necessitating different approaches and technologies. The findings revealed that ML/DL techniques are commonly used for MID. Many researchers have employed DL techniques due to their effectiveness in processing large amounts of data and accurately identifying misinformation patterns. However, there is a disproportionate focus on contributions to MID control in HRLs, with a lack of attention towards developing methods for African languages. The lack of large, labelled datasets for training in LRLs underscores the need to expand research to include African language-specific approaches for effective MID.

In response to RQ3, MID models utilized for detecting misinformation were evaluated using various performance metrics such as accuracy, recall, precision, and F1 score. These models were evaluated using different benchmarked datasets, including text-based, image-based, audio-based, or speech-based misinformation. Although most models possess superior accuracy for MID in HRL settings, there is a significant lack of validated models specifically designed for African languages. This highlights the critical need to verify MID models in the context of African languages to ensure their effectiveness and reliability. The study also identified popular models and classifiers that demonstrate promise for MID in African languages, providing valuable insights for future research in this area. For RQ4, the study highlighted the important gaps and research direction in MID, guiding researchers to develop robust models for African languages. The lack of publicly available datasets for MID in LRLs underscores the need for more extensive data generation efforts in this area. The future research directions provided will enable researchers, policymakers, and stakeholders to address the pressing issue of misinformation in African language contexts.

Overall, the dominance of HRLs such as English, Chinese, and Arabic in datasets impacts the effectiveness of MID models. These languages have extensive datasets from different sources including fake news, COVID-19, Twitter, Facebook, etc. that ensure high accuracy, but this focus creates a bias, limiting models' adaptability to LRLs. Consequently, models trained on HRLs often underperform on LRLs due to linguistic and cultural differences, and insufficient training data. Current techniques for LRLs face limitations like scarce datasets, lack of tailored pre-trained models, and minimal multimodal research. These challenges make it hard to develop accurate MID solutions for LRLs. Thus, improvements could include creating multimodal datasets, using transfer learning from HRLs, and developing low-resource-friendly architectures. In the same vein, advancements in HRL-focused techniques, like transformer-based architectures and cross-lingual transfer learning, can benefit LRL approaches if adapted correctly. Practical applications of improved MID for LRLs could include early-warning systems for misinformation in African languages, offering real-time alerts and tools to verify information credibility, thereby empowering communities to combat misinformation. The discussion outlines the current state of research in MID in LRLs and HRLs. Despite advancements, MID remains challenging due to the subtlety of deceptive language and the complexity of digital editing tools. Therefore, continuous research and development are essential to maintain a steady pace with the evolving nature of misinformation.

6.2 Open Research Directions

In addressing the future research trends to combat misinformation (RQ4), many key gaps and opportunities have been identified for improving MID in LRLs. First of all, there is a strong need to develop multimodal and multilingual approaches that incorporate text, images, audio, and speech for LRLs. This comprehensive strategy will provide deeper insights into the nuanced methods of misinformation spread in these languages.

Researchers should also focus on incorporating code-switched text data and cross-lingual embeddings to enhance the robustness of MID models. These techniques will help in developing more resilient systems that can cope with the complexities of multilingual and multimodal data. Furthermore, more machine translation models specifically designed for LRLs are essential. These models can translate content from HRLs, making MID tools more accessible to LRL speakers. There is an urgent need to generate and publicly share more misinformation datasets for LRLs. Increasing the diversity and volume of these datasets through data augmentation and transfer learning will provide more effective model training. Also, the validation of MID results should be considered to ensure the reliability and accuracy of detection outcomes.

Future research should investigate ways to predict the occurrence of fake news on social media for LRLs, enabling proactive measures to prevent misinformation before it spreads widely. The development of early detection techniques for MID on social media should also be implemented to address misinformation pre-emptively. Using existing models on LLMs to incorporate MID capabilities for LRLs is another essential area of research. This approach utilizes the power of pre-trained models to address the specific linguistic nuances of target languages. The importance of enhancing the contextual understanding of misinformation in LRLs is crucial, considering socio-cultural factors, regional dialects, and linguistic intricacies that influence the spread and perception of misinformation. Finally, implementing education and awareness campaigns to enhance the ability of individuals with critical media literacy skills can enhance technical solutions. Such initiatives will enable communities to recognize and combat misinformation effectively, fostering a more informed and resilient society. By addressing these gaps and focusing on these future research directions, the field of MID can make significant contributions to combating misinformation across diverse linguistic contexts.

6.3 Validity Threats

During the systematic review, we explored numerous relevant articles to gather the necessary information to tackle the RQs outlined in Section 1. We are confident that our review accurately covers MID research published from 2019 to the present. However, we recognize a few potential obstacles that could impact our findings. There's always a chance we might have missed some relevant articles due to the specific search terms we used, or there could have been errors during data extraction from the studies we included such as having no publication in 2020. To address these issues, we conducted an extensive and systematic search across multiple electronic databases and other sources, as stated in Section 3. We also utilized a standardized data extraction method and ensured that all reviewers independently extracted data. While some information may have been overlooked, we believe that these omissions don't significantly affect the overall conclusion of this study.

7 CONCLUSION

This paper presented a systematic analysis of the current state of MID across various languages: HRLs and LRLs, shedding light on contributions, challenges, and future research directions. To achieve this, we carefully selected and analyzed numerous relevant articles based on predetermined criteria. The findings highlighted significant progress in methodologies and technologies for MID, including trends in publications, types of misinformation, ML/DL models, benchmarked datasets, and evaluation metrics. Despite substantial efforts towards HRLs like English, Chinese, and Arabic, the analysis revealed significant gaps and constraints such as the lack of datasets for LRLs, the absence of language-specific tools, and the need for culturally sensitive approaches. The study focuses on a critical need for more specialized MID research in African languages to effectively combat misinformation in these communities. Therefore, future research should focus on the development of robust datasets and language-specific models, utilizing advancements in NLP and ML/DL techniques. MID for LRLs can further be enhanced by integrating multimodal data (such as text, audio, and images) using transfer learning. Techniques like fine-tuning pre-trained multilingual transformer models (e.g., mBERT, XLM-R) could mitigate the lack of labelled data. This is because multimodal approaches are viable solutions to solve the complexity of misinformation in underrepresented languages. In addition, fostering partnerships between academia, industry, and local communities is critical, as is acknowledging the socio-political implications of misinformation in African languages. Effective MID does not only protect the integrity of information ecosystems but also improves democratic processes and enhances social resilience.

As part of our future research, we intend to develop and implement a MID model specifically for LRLs with a focus on South African languages. This will involve developing multimodal datasets or annotating large datasets that are tailored to these languages to effectively identify and counter misinformation. Also, we will adapt pre-trained models, and create lightweight systems integrated with an early-warning. This initiative aims to create a more inclusive, informed, and resilient community across Africa and beyond.

REFERENCES

[1] De, A., Bandyopadhyay, D., Gain, B., and Ekbal, A.: 'A transformer-based approach to multilingual fake news detection in low-resource languages', *Transactions on Asian and Low-Resource Language Information Processing*, 2021, 21, (1), pp. 1-20

[2] Farhangian, F., Cruz, R.M., and Cavalcanti, G.D.: 'Fake news detection: Taxonomy and comparative study', *Information Fusion*, 2024, 103, pp. 102140

[3] Alghamdi, J., Lin, Y., and Luo, S.: 'Fake news detection in low-resource languages: A novel hybrid summarization approach', *Knowledge-Based Systems*, 2024, 296, pp. 111884

[4] Nakamura, K., Levy, S., and Wang, W.Y.: 'r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection', *arXiv preprint* arXiv:1911.03854, 2019

[5] Hosseini, M., Sabet, A.J., He, S., and Aguiar, D.: 'Interpretable fake news detection with topic and deep variational models', *Online Social Networks and Media*, 2023, 36, pp. 100249

[6] Salau, A.O., Arega, K.L., Tin, T.T., Quansah, A., Sefa-Boateng, K., Chowdhury, I.J., and Braide, S.L.: 'Machine learning-based detection of fake news in Afan Oromo language', *Bulletin of Electrical Engineering and Informatics*, 2024, 13, (6), pp. 4260-4272

[7] Zeng, F., Li, W., Gao, W., and Pang, Y.: 'Multimodal Misinformation Detection by Learning from Synthetic Data with Multimodal LLMs', arXiv preprint arXiv:2409.19656, 2024

[8] Gereme, F.B., and Zhu, W.: 'Early detection of fake news" before it flies high"', in Editor (Ed.)^(Eds.): '*Book Early detection of fake news*" before it flies high"' (2019, edn.), pp. 142-148

[9] Rashid, M.R.A., Roy, R., Rahman, D.M.S., Saleh, M.A., Khan, A.A.H., Rayhan, A., Ahmed, K.F., Monsoor, N., and Hasan, M.: 'A Comprehensive Dataset and Deep Learning Approach for Misinformation Detection on Social Media in Bangladesh', *International Journal of Computing and Digital Systems,* 2024, 16, (1), pp. 1-10

[10] Raja, E., Soni, B., Lalrempuii, C., and Borgohain, S.K.: 'An adaptive cyclical learning rate based hybrid model for Dravidian fake news detection', Expert Systems with Applications, 2024, 241, pp. 122768

[11] Al-Zahrani, L., and Al-Yahya, M.: 'Pre-Trained Language Model Ensemble for Arabic Fake News Detection', *Mathematics*, 2024, 12, (18), pp. 1-17

[12] Hashmi, E., Yayilgan, S.Y., Yamin, M.M., Ali, S., and Abomhara, M.: 'Advancing fake news detection: hybrid deep learning with fast text and explainable AI', *IEEE Access*, 2024

[13] Hossain, M.R., Hoque, M.M., Siddique, N., and Dewan, M.A.A.: 'AraCovTexFinder: Leveraging the transformer-based language model for Arabic COVID-19 text identification', *Engineering Applications of Artificial Intelligence,* 2024, 133, pp. 107987

[14] Lin, H., Ma, J., Yang, R., Yang, Z., and Cheng, M.: 'Towards low-resource rumour detection: Unified contrastive transfer with propagation structure', *Neurocomputing*, 2024, 578, pp. 127438

[15] Mallik, A., and Kumar, S.: 'Word2Vec and LSTM based deep learning technique for context-free fake news detection', *Multimedia Tools and Applications*, 2024, 83, (1), pp. 919-940

[16] Mohsen, F., Chaushi, B., Abdelhaq, H., Karastoyanova, D., and Wang, K.: 'Automated Detection of Misinformation: A Hybrid Approach for Fake News Detection', *Future Internet*, 2024, 16, (10), pp. 352

[17] Al-Alshaqi, M., Rawat, D.B., and Liu, C.: 'Ensemble Techniques for Robust Fake News Detection: Integrating Transformers, Natural Language Processing, and Machine Learning', *Sensors*, 2024, 24, (18), pp. 6062

[18] Ricketts, J.A.: 'Powers-of-ten information biases', MIS Quarterly, 1990, pp. 63-77

[19] 19 Praseed, A., Rodrigues, J., and Thilagam, P.S.: 'Hindi fake news detection using transformer ensembles', *Engineering Applications of Artificial Intelligence*, 2023, 119, pp. 105731

[20] Su, Q., Wan, M., Liu, X., and Huang, C.-R.: 'Motivations, methods and metrics of misinformation detection: an NLP perspective', *Natural Language Processing Research*, 2020, 1, (1-2), pp. 1-13

[21] Verma, P.K., Agrawal, P., Madaan, V., and Prodan, R.: 'MCred: multi-modal message credibility for fake news detection using BERT and CNN', *Journal of Ambient Intelligence and Humanized Computing*, 2023, 14, (8), pp. 10617-10629.

[22] Luvembe, A.M., Li, W., Li, S., Liu, F., and Wu, X.: 'CAF-ODNN: Complementary attention fusion with optimized deep neural network for multimodal fake news detection', *Information Processing & Management*, 2024, 61, (3), pp. 103653

[23] Kumar, A., Esposito, C., and Karras, D.A.: 'Introduction to special issue on misinformation, fake news and rumour detection in low-resource languages', in Editor (Ed.)^(Eds.): 'Book Introduction to special issue on misinformation, fake news and rumour detection in low-resource languages' (ACM New York, NY, 2021, edn.), pp. 1-3

[24] Yu, C., Han, J., Zhang, H., and Ng, W.: 'Hypernymy detection for low-resource languages via meta learning', in Editor (Ed.)^(Eds.): 'Book Hypernymy detection for low-resource languages via meta learning' (2020, edn.), pp. 3651-3656

[25] Kar, D., Bhardwaj, M., Samanta, S., and Azad, A.P.: 'No rumours please! A multi-indic-lingual approach for COVID fake-tweet detection', in Editor (Ed.)^(Eds.): 'Book No rumours, please! A multi-indic-lingual approach for COVID fake-tweet detection' (IEEE, 2021, edn.), pp. 1-5

[26] Ghafoor, A., Imran, A.S., Daudpota, S.M., Kastrati, Z., Batra, R., and Wani, M.A.: 'The impact of translating resource-rich datasets to low-resource languages through multi-lingual text processing', *IEEE Access,* 2021, 9, pp. 124478-124490

[27] Gereme, F., Zhu, W., Ayall, T., and Alemu, D.: 'Combating fake news in "low-resource" languages: Amharic fake news detection accompanied by resource crafting', *Information*, 2021, 12, (1), pp. 20

[28] Du, J., Dou, Y., Xia, C., Cui, L., Ma, J., and Philip, S.Y.: 'Cross-lingual covid-19 fake news detection', in Editor (Ed.)^(Eds.): 'Book Cross-lingual covid-19 fake news detection' (IEEE, 2021, edn.), pp. 859-862

[29] Kim, J., Bak, B., Agrawal, A., Wu, J., Wirtz, V., Hong, T., and Wijaya, D.: 'Covid-19 vaccine misinformation in middle income countries', in Editor (Ed.)^(Eds.): 'Book Covid-19 vaccine misinformation in middle income countries' (Association for Computational Linguistics, 2023, edn.), pp.

[30] Kuznetsova, E., Makhortykh, M., Vziatysheva, V., Stolze, M., Baghumyan, A., and Urman, A.: 'In Generative AI we Trust: Can Chatbots Effectively Verify Political Information?', *arXiv preprint* arXiv:2312.13096, 2023

[31] Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., and PRISMA Group*, t.: 'Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement', *Annals of internal medicine,* 2009, 151, (4), pp. 264-269

[32] Yan, F., Zhang, M., Wei, B., Ren, K., and Jiang, W.: 'FMC: Multimodal fake news detection based on multi-granularity feature fusion and contrastive learning', *Alexandria Engineering Journal*, 2024, 109, pp. 376-393

[33] Albalawi, R.M., Jamal, A.T., Khadidos, A.O., and Alhothali, A.M.: 'Multimodal Arabic rumours detection', *IEEE Access*, 2023, 11, pp. 9716-9730

[34] Zheng, P., Chen, H., Hu, S., Zhu, B., Hu, J., Lin, C.-S., Wu, X., Lyu, S., Huang, G., and Wang, X.: 'Few-shot learning for misinformation detection based on contrastive models', *Electronics*, 2024, 13, (4), pp. 799

[35] Van der Westhuizen, E., Kamper, H., Menon, R., Quinn, J., and Niesler, T.: 'Feature learning for efficient ASR-free keyword spotting in low-resource languages', *Computer Speech & Language*, 2022, 71, pp. 101275

[36] Hansrajh, A., Adeliyi, T.T., and Wing, J.: 'Detection of online fake news using blending ensemble learning', *Scientific Programming*, 2021, 2021, (1), pp. 3434458

[37] Reis, J.C., Correia, A., Murai, F., Veloso, A., and Benevenuto, F.: 'Supervised learning for fake news detection', *IEEE Intelligent Systems*, 2019, 34, (2), pp. 76-81

[38] Asghar, M.Z., Habib, A., Habib, A., Khan, A., Ali, R., and Khattak, A.: 'Exploring deep neural networks for rumour detection', *Journal of Ambient Intelligence and Humanized Computing*, 2021, 12, pp. 4315-4333

[39] Jadhav, S.S., and Thepade, S.D.: 'Fake news identification and classification using DSSM and improved recurrent neural network classifier', Applied Artificial Intelligence, 2019, 33, (12), pp. 1058-1068

[40] Kaliyar, R.K., Goswami, A., and Narang, P.: 'DeepFakE: improving fake news detection using tensor decomposition-based deep neural network', The *Journal of Supercomputing*, 2021, 77, (2), pp. 1015-1037

[41] Lin, H., Yi, P., Ma, J., Jiang, H., Luo, Z., Shi, S., and Liu, R.: 'Zero-shot rumour detection with propagation structure via prompt learning', in Editor (Ed.)^(Eds.): 'Book Zero-shot rumour detection with propagation structure via prompt learning' (2023, edn.), pp. 5213-5221

[42] Guo, Z., Zhang, Q., Ding, F., Zhu, X., and Yu, K.: 'A novel fake news detection model for the context of mixed languages through multiscale transformer', *IEEE Transactions on Computational Social Systems*, 2023

[43] Dlamini, G., Bekkouch, I.E.I., Khan, A., and Derczynski, L.: 'Bridging the domain gap for stance detection for the Zulu language', in Editor (Ed.)^(Eds.): 'Book Bridging the domain gap for stance detection for the Zulu language' (Springer, 2022, edn.), pp. 312-325

[44] De Wet, H., and Marivate, V.: 'Is it fake? News disinformation detection on South African news websites, in Editor (Ed.)^(Eds.): 'Book Is it fake? News disinformation detection on South African news websites' (IEEE, 2021, edn.), pp. 1-6

Table 1. Summary of Studies Considered

Year	Models	Language	Modality	Data sets	Problem domain	Evaluation Metrics	Validation	Ref
2024	CNN_BiLSTM; RNN-SVM; AC-BiLSTM; BERT-LSTM-CNN; Adaptive Transfer Learning; Adaptive Hybrid Model	Dravidian languages: Tamil, Telugu, Kannada, Malayalam.	Text	Dravidian	Fake news	Accuracy, F1-score,	yes	[10]
2024	mBERT Baseline, XLM-RoBERTa, mBERT, Semantic graph-based topic modelling	English, Vietnamese, Hindi, Indonesian & Swahili	Text	TALLIP, Multilingual fake news detection (MFND)	Fake news	Accuracy, Precision, Recall, F1-score, F1-macro	yes	[3]
2024	LSTM & Bangla BERT	Bengali	Text	FactWatch	Misinformation	Precision, Recall Accuracy, Confusion matrix	yes	[9]
2024	K-means Clustering, Logistic regression, Multinomial Naïve Bayes, Random Forest, CNN, RNN, BERT, TI-CNN	English	Text and images	Kaggle, TICNN, ISOT, & SMS Spam	Fake news	Accuracy, Precision, Recall, F1 Score & confusion matrix	yes	[15]
2024	mBERT, XLM-RoBERTa, mDeBERTa-V3, mDistilBERT, BERT-Arabic, and AraBERT, Word2Vec, GloVe, and FastText embeddings with CNN, LSTM, VDCNN, and BiLSTM	Arabic	Text	Arabic Fake News Dataset	Fake news	Accuracy, Precision, Recall, Weighted-average, Macro-average, F1-score, Confusion Matrix, MCC, G-mean)	yes	[13]
2024	CNN, LSTM BERT, XLNet, and RoBERTa, RNN	Arabic, English	Text	WELFake, FakeNewsNet, and FakeNewsPrediction	Fake news	Accuracy, F1-scores, Precision, Recall	yes	[12]
2024	CAF, ODNN, USE, CAF-ODNN	English	Text & image	Fakeddit, Pheme, GossipCO, PolitiFact datasets	Fake news	Accuracy, Precision, Recall and F1 score	yes	[22]

2024	BERT, ResNet, BC, RC, BR, FMCBCRC, SpotFake, SAFE, BTIC, CAFÉ, COOLANT & TTEC	English, Chinese	Text & image	ReCOvery, GossipCop & MR^2	Fake news	Accuracy, Precision, Recall and F1 score, MCC	yes	[32]
2024	SVM, KNN, CNN	Afan Oromo	Text	Facebook	Fake news	Precision, Recall, F1-score	yes	[6]
2024	AraBERT, MARBERT, AraELECTRA, AraGPT2, and ARBERT	Arabic	Text	AMFND dataset	Fake news	Accuracy, F1 score	yes	[11]
2024	BERT, CNN, Random Forest, SVM, KNN, Logistic Regression, Multinomial Naïve Bayes, TextLSTM, BERTweet, and Spotfake	English	Text, images, videos	ISOT. MediaEval 2016	Fake news	accuracy, precision, recall, and F1-score		[17]
2024	TF-IDF, Decision Trees (DT), KNN, Gradient boosting classifiers (GBC), Multinomial naïve Bayes (MNB), Bernoulli naïve Bayes(BNB)	English	Text	FA-KES, WELFake, fake and Real Dataset, Kaggle, trimmed-WELFake, trimmed-Scraped, and trimmed-Kaggle1,	Fake news	Accuracy, Precision, Recall, F1 score	yes	[16]
2024	LLaVA-13B, GPT-4V, CLIP, MLLMs SemSim, DisSim	English	Images & Text	MediaEval, Snopes (O+)	Misinformation	F1 score	yes	[7]
2024	CNN, RNN, RvNN. PLAN BiGCN, DANN, UCLR	English, Cantonese, Arabic, Chinese	Text	English-COVID19 Chinese-COVID19 Cantonese-COVID19 Arabic-COVID19	Rumours	Accuracy, Macro F1score	yes	[14]
2024	LDA, Bi-LSTM VAE LDAVAE (LDA + Bi-LSTM VAE+ Classifier architecture)	English	Text	ISOT, Twitter COVID	Fake news	Accuracy, F1-score	yes	[34]
2023	ELECTRA, mBERT, XLM-RoBERTa	Hindi	Text	CONSTRAINT2021	Fake news	F1-Score, Accuracy Recall, Precision MCC	yes	[19]
2023	CNN, LSTM, Transformer BERT, MST-FaDe	Chinese	Text	Weibo-hybrid	Fake news	Precision, Recall F-score, Accuracy	yes	[42]

		Translated to English						
2023	Zero-shot	English Chinese	Text	TWITTER Twitter-COVID19 WEIBO Weibo-COVID19	Rumour	Accuracy, macro, F1-score	yes	[41]
2023	Early Fusion, Late Fusion	Arabic	Text, image	Arafacts	Rumour	Precision, Recall, F1-score, Accuracy	yes	[33]
2023	CNN, BERT MCred (CNN+BERT)	English	Text	Kaggle, McIntire FakeNews, ELFake	Unreliable news & Reliable new	Precision, Recall F1-score, Accuracy	yes	[21]
2023	LDA, MVAE, CNN, RNN, SVM, Logistic regression, Random Forest, Naïve bayes, MLP, KNN	English	Text	ISOT, COVID, Twitter	Unreliable Reliable	F1-score, Accuracy False positive rates, False negative rates	yes	[5]
2022	LSTM	English, Zulu	Text	SemEval-2016 dataset	Rumour	F1 score, accuracy, FAVOR-F1 score, AGAINST-F1 score	yes	[43]
2022	CNN CNN-DTW	English, Luganda	Speech	South African Broadcast News (SABN), Radio broadcasts in Kampala	Fake news	ROC, AUC EER	No	[35]
2021	Logistic regression, LDA, SGDC, Ridge classifier, SVM Blending (BLD) ensemble	English	Text	LIAR, ISOT	Fake news	ROC & AUC, F1-score, precision Recall, Accuracy	No	[36]
2021	CNN, LSTM, BiLSTM-CNN	English	Text	Pheme rumor dataset, Fake-news, Fake_real_news	Rumour	Precision, Recall F-score, Accuracy	yes	[38]
2021	XGBoost, DNN	English	Text	BuzzFeed PolitiFact	Fake news	Precision, Recall F1-Score, Accuracy, TN, FN, TP, FP	yes	[40]
2021	Logistic regression, LSTM	English	Text	US 1 and US2 (Kaggle), GitHub	Fake news	Accuracy, F1 Score	No	[44]
2019	BERT, InferSent, ResNet50, VGG16 EfficientNet	English	Text & Image	FEVER, Fakeddit	Fake news	2-way, 3-way, 6- way classification	No	[4]

2019	Naïve bayes, LSTM, Deep ConvNets	English	Text	Kaggle, George Mclntire, KaggleMclntire	Fake news	Accuracy, Recall Precision, F1 Score	No	[8]
2019	KNN, Naïve bayes, Random Forest, SVM, XGBoost (XGB)	English	Text, image, video	BuzzFeed	Fake news	AUC, F1 score	Yes	[37]
2019	DSSM, RNN, DSSM-LST	English	Text	LIAR dataset	Fake news	Accuracy	Yes	[39]