



## A Deep Learning-Based Approach for Detecting Afan Oromo Fake News on Social Media

Kedir Lemma Arega<sup>1</sup>, Ashish Bagwari<sup>2</sup>, Kula Kekeba Tune<sup>1</sup>, Asrat Mulatu Beyene<sup>3</sup>, Ciro Rodriguez<sup>4\*</sup>,  
Pedro Lezama<sup>5</sup>, Ayodeji Olalekan Salau<sup>6,7</sup>

<sup>1</sup> Department of Software Engineering, Addis Ababa Science and Technology University, Addis Ababa 16417, Ethiopia

<sup>2</sup> Department of Electronics and Communication Engineering, Uttarakhand Technical University, Dehradun 248007, India

<sup>3</sup> Department of Electrical and Computer Engineering, Addis Ababa Science and Technology University, Addis Ababa 16417, Ethiopia

<sup>4</sup> Department of Software Engineering, Universidad Nacional Mayor de San Marcos UNMSM, Lima 15081, Peru

<sup>5</sup> Department of Computer Science, Universidad Nacional Mayor de San Marcos UNMSM, Lima 15081, Peru

<sup>6</sup> Department of Electrical/Electronics and Computer Engineering, Afe Babalola University, Ado-Ekiti 360101, Nigeria

<sup>7</sup> Department of Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai 600124, India

Corresponding Author Email: [crodriguezro@unmsm.edu.pe](mailto:crodriguezro@unmsm.edu.pe)

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/mmep.120930>

### ABSTRACT

**Received:** 10 May 2025

**Revised:** 18 July 2025

**Accepted:** 25 July 2025

**Available online:** 30 September 2025

#### Keywords:

*Afan Oromo, convolutional neural networks (CNNs), deep learning, fake news detection, natural language processing, neural networks, social media*

The research aims to develop a deep learning system for detecting fake news on social media using Afan Oromo news text. The study reviews related research, introduces phony news detection, and develops a methodology. The research explores linguistic challenges in the Afan Oromo language, focusing on identifying and categorizing misleading news content. It evaluates deep learning models and provides insights for stakeholders. The study empowers local communities to combat misinformation, considering cultural context and ethical considerations. It also addresses biases and biases that may affect accurate information dissemination. The pre-trained model is evaluated using metrics, and a system prototype is created. This research used a meticulously trained dataset to identify fake news in the Afan Oromo language. Convolutional neural networks (CNNs) were used to accurately categorize the data, with a feature set technique. The study found that convolutional networks achieved an impressive accuracy rate of 93%, surpassing traditional neural networks' 92% accuracy. The proposed strategies achieved a 91% accuracy rate. The research suggests using deep learning to identify fake news on social media, highlighting the impact of false information on society. The approach involves scraping Afan Oromo posts, cleaning and filtering data, and consolidating and annotating the data.

## 1. INTRODUCTION

Deep-learning algorithms are being used to detect bogus news in the Afaan Oromo language on social media sites, a difficult task owing to the fast spread of misinformation. These strategies improve detection accuracy and efficiency, as social media is a key source of fake news [1].

A fundamental challenge is the absence of annotated data for training deep learning models to detect and categorize fake news in Afaan Oromo. The lack of data, along with the language's complicated features, makes it difficult to create exact and reliable models. The paucity of research and development efforts on deep learning-based algorithms for false news detection on social media networks exacerbates these issues, leading to possible biases and limits in current datasets [2]. The creation of deep learning models for identifying and categorizing fake news in the Afaan Oromo language on social media networks is difficult due to a lack of established benchmarks and assessment measures, as well as issues concerning interpretability and explainability. These issues raise ethical concerns, particularly regarding freedom of

speech and censorship. The sheer volume of data and computational resources makes it difficult to implement these models effectively. To overcome these challenges, research efforts, computational resources, and transparent models are needed. The research aims to improve detection accuracy, provide real-time monitoring tools, and promote media literacy among Afaan Oromo speakers. A real-time monitoring system will flag potential fake news stories, allowing for timely interventions.

Selecting and scraping social media sources involves several steps to ensure relevant and useful data. These steps include selecting relevant platforms, defining data requirements, using Application Programming Interfaces (API), web scraping, data cleaning and storage, and analyzing and visualizing the collected data. Identifying the target audience, data availability, and data requirements are crucial steps in this process. APIs allow developers to access data programmatically, but authentication and endpoints are essential. If APIs are not available, web scraping tools or libraries can be used to extract data from web pages. Data cleaning and storage are also essential steps, with cleaned data

stored in suitable formats for further analysis.

The study looks at the application of deep learning models to detect bogus news in the Afan Oromo language and on social media. However, it has limitations, such as a narrow emphasis on a single dataset, the possibility of bias, and the lack of a full comparison of deep learning and other approaches. The study also ignores additional criteria, such as recall, F1 score, and interpretability, which might give a more detailed assessment of deep learning models' performance. Ethical considerations such as privacy concerns and potential biases in algorithmic decision-making are also overlooked. The study contributes to global efforts in combating misinformation.

The research endeavor meticulously underscores the multifaceted linguistic challenges that exist while simultaneously advocating for deeper intercultural comprehension, all the while engaging in a thorough investigation of the mechanisms involved in the identification and categorization of misleading or fraudulent news content, specifically within the framework of the Afan Oromo language. This thorough analysis not only examines the state-of-the-art deep learning models currently in use in the field, but it also offers a wealth of information about the complex linguistic issues at hand, as well as practical and perceptive recommendations meant for different stakeholders who are actively involved in this field. Moreover, this scholarly work serves to empower local communities in their efforts to combat the proliferation of misinformation, taking into consideration the unique cultural context in which they operate, while simultaneously addressing pertinent ethical considerations and striving to alleviate inherent biases that may adversely affect the dissemination of accurate information.

## 2. LITERATURE REVIEW

The paper [3] presents a detailed system architecture for fake news detection, outlining its structure and behavior. It discusses performance evaluation metrics like precision, recall, F-measure, and accuracy. The paper reviews existing literature on fake news detection, highlighting gaps and providing context for future research. The findings and data analysis provide insights into the system's performance and practical implications. The author expresses gratitude to individuals and institutions for their contributions, emphasizing the collaborative nature of academic work. However, significant issues include the lack of fact-checker websites for Afaan Oromo news on social media, a focus on textual news articles, and limited resources.

The paper [4] explores methods for fake news detection in social media and the Afaan Oromo language. Key methods include Linguistic Features-Based Models and Predictive Modeling, which analyze the language used in news articles to identify fake or credible news. However, these methods may not be as effective as more advanced models. The paper details the architecture of a model designed for automatic fake news detection in Afaan Oromo, including techniques, algorithms, and data preparation processes for experimental validation. The aim is to enhance fake news detection in social media, especially for Afaan Oromo speakers. However, the study has limitations, including its focus on the Afaan Oromo language and specific news sources, which may limit its applicability to other languages or broader contexts. Data constraints, such as limited or biased datasets, could also hinder the effectiveness

of machine-learning models in real-world scenarios. The difficulty humans face in distinguishing between real and fake news on social media presents a challenge for automated detection systems.

This research [5] proposes a hybrid strategy for improving automated prediction accuracy by combining social context-based and content-based methodologies. The study methodology involves data preparation, feature extraction techniques, and model designs. The method reduces duplication and dimensionality using text mining algorithms, whereas social context-based approaches detect bogus news using social interaction data. For a less biased estimate, the models are assessed using k-fold cross-validation. Dataset difficulties, a lack of data, computing resources, morphological richness, and problems with hyperparameter tuning are some of the study's drawbacks. Optimizing model performance may be impacted by the enormous datasets, small number of journalists, and low-end GPU resources. The grammatical richness of the Amharic language may also make it difficult for deep learning systems to effectively recognize bogus news articles.

The paper [6] examines fake news detection models' generalization ability, highlighting that they often fail across different datasets. The authors propose a new framework using the Hierarchical Discourse-level Structure of Fake News (HDSF) and introduce a hybrid deep learning model that combines convolutional and recurrent neural networks (RNNs). The research emphasizes the importance of model complexity and training for better generalization across datasets and tasks. However, challenges such as overfitting, dataset size limitations, and lack of interpretability in the hybrid CNN-RNNs model are discussed. Further experiments are suggested to explore bio-inspired optimization techniques for hyperparameter tuning.

The paper [2] uses deep learning models, specifically a Bi-directional Long Short-Term Memory (Bi-LSTM) network, to improve fake news detection in the Afaan Oromo language. The models are trained on a dataset from social media platforms, focusing on the language's unique characteristics. The performance is measured using the F1 score, with the model achieving a 90% score. The implementation is carried out using Python Spyder IDE for practical application. However, the study on fake news detection in the Afaan Oromo language faced limitations due to a lack of a standard text corpus, significant training challenges, and the vanishing gradient problem. RNNs require significant training data and time, which could hinder efficiency. The study also did not incorporate features like article source or author and user responses, which could improve performance. These limitations underscore the need for further research and development in fake news detection.

Using data from many sources and preprocessing techniques, the research [7] proposes a method for identifying bogus news in the Afaan Oromo language on social media. To determine the significance of words in news stories, feature extraction techniques are employed. To increase accuracy, a hybrid model that combines LSTM networks and convolutional neural networks (CNNs) is employed. Although the model's accuracy was 67.1%, it had a 32.9% error rate, possible bias, and reliability problems. The intricacy of the model may restrict its ability to comprehend the subtleties of the Afaan Oromo language, indicating areas in need of development and further study in the identification of false information.

To find hate speech material, the study [8] used a sizable dataset of Afaan Oromo social media posts and comments. Deep learning models such as CNN, LSTM, Bi-LSTM, GRU, and CNN-LSTM combinations were employed. With the highest weighted average F1-score of 87%, the CNN and Bi-LSTM model demonstrated its efficacy in identifying hate speech. The model's drawbacks, however, are its small sample size, possible misclassification problems, cultural bias, and absence of a uniform corpus. To further reflect the intricacy of hate speech detection, the authors recommend investigating classifier ensembles and meta-learning techniques. Results from a bigger dataset could be more reliable.

The paper [9] presents a deep learning model for detecting disinformation and fake news on Twitter, utilizing a Distil BERT transformer, frequent item sets, association rules, graph theory, and parameter-efficient fine-tuning. The model analyzes tweet texts, user behavior patterns, and sentiments associated with named entities, which can be used in supervised machine-learning models. However, the model's limitations include unsubstantiated claims, contradictions, emotional manipulation, personal attacks, and a limited scope of analysis, raising concerns about its reliability, credibility, and generalizability.

The paper [10] presents a method for identifying fake news on social media platforms using a fake news dataset created by fact-checking sites, specifically those specific to India. The model, Debunking Multi-Lingual Social Media Posts using Deep Learning (DSMPD), classifies news items into five categories: real, could be real, fabricated, or unknown. The researchers compare different machine learning and deep learning algorithms to improve the accuracy of identifying fake news. However, the BERT model, which does not update in real-time, may not perform well on political content. Data tagging accuracy is crucial for the model's effectiveness, and the LSTM model's performance is based on a small dataset.

The paper presents a deepfake detection system using three deep learning models: Inception ResNetV2, EfficientNet, and VGG16. These models achieve 97% accuracy in detecting deepfakes. GANs generate fakes, while CNNs and RNNs extract features from videos, images, and audio. However, the study highlights the constraints of these models, such as resource-intensiveness, generalization concerns, data quality dependencies, and spatiotemporal analysis constraints. The authors also warn of potential adversarial assaults, where bad actors create fakes to avoid detection [11].

A study [12] on detecting fake news on social media in Turkish and English has achieved high accuracy rates using deep learning systems using CNN and RNN-LSTM models. The Turkish systems outperformed previous studies, with accuracy rates ranging from 87.14% to 92.48%. However, the study acknowledges the need for larger datasets and testing to ensure generalizability and robustness. The study also notes that Turkish is an agglutinative language, presenting unique text classification challenges. Future research should focus on improving fake news detection on social media platforms.

The study [13] used a dataset of 44,898 real and fake articles, tokenized and converted into numerical data using a count vectorizer. The researchers tested classification models like Logistic Regression, Support Vector Machine (SVM), and Linear Support Vector Classifier (SVC) to classify articles as real or fake. Data preprocessing techniques were applied to improve model accuracy. The Linear SVC model achieved the highest accuracy of 99.97%, while the Logistic Regression model achieved 98.4%. The study also noted that punctuation

marks were more common in real news articles, so they were removed from the dataset to enhance the model's applicability beyond the training data.

The paper [14] presents four deep learning models for hoax detection on social media: Bi-LSTM, RNN, hybrid RNN-Bi-LSTM, and hybrid Bi-LSTM-RNN. These models compare different approaches to detecting hoaxes on social media. The Bi-LSTM model achieved the highest accuracy of 96.60%, indicating its effectiveness. The research emphasizes the importance of dropout in improving model accuracy and using a similarity corpus to transform zero-value words into meaningful TF-IDF values. However, the paper has limitations, such as the dataset from GitHub and Indonesian news websites, the complexity of the models, and the reliance on TF-IDF and GloVe for feature extraction. Future research should explore other methods, such as Restricted Boltzmann Machines (RBMs) and hyperparameter adjustments, and address other metrics like F1-Score or AUC-ROC for a more comprehensive evaluation of model performance.

A thorough approach for identifying phony Facebook profiles in Afaan Oromo is presented in the research. Data gathering, feature extraction, labeling, preprocessing, model selection, feature engineering, training, assessment, optimization, and deployment are all part of it. The objective is to use the Afaan Oromo language to efficiently identify phony profiles on social networking sites. The approach is constrained, nevertheless, by issues with generalizability, cross-validation, feature restrictions, and dataset imbalance. The 10-fold cross-validation approach might not be entirely successful in identifying phony accounts because the original dataset was highly skewed. The results might not translate well to other languages or social networking sites [15].

Using preprocessing, deep learning, word embedding, and single-layer and three-layer LSTM architectures for training, the study [16] investigates multi-label news text categorization approaches in Afaan Oromo. The model's shortcomings include its modest dataset size, concentration on main classes, and absence of a standard corpus, despite its excellent accuracy levels. It's possible that the model won't work well for increasingly complicated classification problems. Furthermore, the model's performance is affected by the computing resources that are available; models with a high number of neurons and epochs are advised to have a high processing power.

The study [17] examines various techniques for identifying false information, such as deep learning models like CNNs, RNNs, LSTM, and Bidirectional LSTM, as well as machine learning algorithms like Random Forest, Decision Tree, and Multinomial Naïve Bayes. 99% accuracy is attained via a hybrid model known as the Fake News Detections (FND) model. Data pre-processing is done using the ISOT dataset, which has over 12,600 items of both false and true news. The processing power needs, dataset number and quality, and the possibility of performance degradation due to lengthy input sequences are some of the drawbacks of classical models.

The study [1] focuses on detecting fake news using text features and user feedback, highlighting the role of social media and online news in spreading misinformation. It examines how recurring and non-recurring events affect news classification. Different models like LSTM, BERT, and CNN-BiLSTM were tested, with BERT showing the best results; it was found that 70% of true news is recurring while 30% is non-recurring. The paper discusses the challenges faced by existing fake news detection solutions, including low precision

and high computational complexity. It also highlights the neglect of temporal aspects, which are crucial for understanding the context and relevance of news articles. The proposed framework relies on user feedback to determine news authenticity, but this may introduce biases or inaccuracies. The study suggests that models trained on imbalanced datasets perform better but may not generalize well to real-world scenarios. Additionally, the paper may not fully explore all temporal features that could enhance fake news detection.

The study [18] detects and categorizes bogus news in the Afan Oromo language using machine learning methods such as SVM, KNN, and CNNs. Facebook provided the data, and

data collecting, cleansing, and segmentation are all part of the system design. The classifiers' efficacy is evaluated by measures such as F1 score, recall, accuracy, and precision. In identifying bogus news, the study demonstrates the higher accuracy rates of SVM and naïve Bayesian classifiers. Nonetheless, the study highlights places for more study and advancement in false news identification while addressing the difficulties of gathering a sizable dataset, language subtleties, and feature engineering. Table 1 presents a summary of the introduction, strategy or method employed, results, and limitations. in addition to the many perspectives and contributions [19-25].

**Table 1.** The presented showcases the various perspectives, contributions, methods employed, and constraints

| References                  | Perceptions                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Contributions                                                                                                                                                                 | Approach/Technique Used                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | Limitations                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|-----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Abera and Tegegne [19]      | In order to detect fake news in the Afan Oromo language, the study investigates the creation and use of an information extraction model. By removing important details, the goal is to make Afan Oromo news items easier to read and understand. However, a critical component of news consumption and distribution in the digital age, identifying false news in Afan Oromo is not discussed in the research. In order to better understand and develop successful strategies to counteract misinformation and disinformation and promote an educated and empowered community, the article recommends that future research investigate techniques for identifying and countering false news in the Afan Oromo language. | - AOTIE's F-measure is 80%, its accuracy is 79.5%, and its recall is 80.5%.<br>- The grammatical structure of Afan Oromo is crucial to AOTIE's functionality.                 | Learning and extraction, postprocessing, and document preprocessing Agazetteer invention and Afan Oromo grammatical structure research.                                                                                                                                                                                                                                                                                                                                                                 | Information extraction (IE) from news articles published in Afan Oromo is the main topic of this essay. 3169 tokens make up the short dataset used for training and testing. Evaluation metrics like as recall, accuracy, and F-measure are used to objectively evaluate the model's performance; however, contextual relevance and semantic comprehension are not taken into account. The two experimental scenarios most likely do not cover all possible factors influencing the model's performance. The study does not address the scalability and generalizability of the proposed model to larger datasets or additional Afan Oromo linguistic regions. |
| Tazeze and Raghavendra [20] | Establishing a thorough data collection and categorization system to detect fake news in Amharic is the focus of the work that is being presented. However, identifying fake news in the Afan Oromo language is not covered in the report.                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Creation of classifiers using machine learning to identify bogus news. Establishment of a dataset on false news in Amharic.                                                   | Using information from social media accounts and reliable sources, researchers produced an Amharic false news dataset. Based on accuracy and F1-score, they assessed six machine learning classifiers, including Nave Bayes and Passive Aggressive Classifier. With an accuracy of over 96% and an F1-score of over 99%, Nave Bayes and Passive Aggressive Classifier outperformed the others. The project's goal is to use machine learning techniques to automatically identify fake news in Amharic. | The size and representativeness of the Amharic fake news dataset, the machine learning classifiers' specific features, the potential biases or limitations of using verified news sources and social media pages, the difficulties in detecting fake news in Amharic, and the effect of the selected classifiers on computational resources and scalability for real-time fake news detection are all missing from the paper. The reliability of the model and the results' generalizability may be impacted by several factors.                                                                                                                               |
| Güler and Gündüz [12]       | The study is devoid of any discussion or information on the subject, raising concerns about its scope and emphasis. The report fails to address the rising concern about disinformation and fake news on social media platforms, and it makes no mention of the                                                                                                                                                                                                                                                                                                                                                                                                                                                          | The establishment of SOSYalan, the first real-world public dataset of Turkish false and legitimate news tweets. - The creation of deep learning-based methods for identifying | RNNs-LSTM-CNNs                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Manually identifying fake news is subjective and difficult. - An automated approach needs knowledge of complicated NLP and may struggle to produce correct results.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |

|                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                          |                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                    | significance of deep learning techniques in addressing this issue. The absence of investigation into the Afaan Oromo language in the identification of fake news on social media is a key deficiency that requires additional attention.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | bogus news in English and Turkish.                                                                                                                                                                       |                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Al-Tai et al. [21] | The important issue of "Fake News Detection on Social Media by Utilizing Deep Learning for the Afan Oromo Language" is regrettably neither discussed or mentioned in the supplied paper.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Use deep learning techniques to identify false information. Examining attention strategies and word embedding models.                                                                                    | The use of deep learning techniques (CNNs, RNNs, and multimodal approaches). Text-to-vector conversion word embedding models. | Manually detecting fake news is subjective and difficult. An automated approach needs knowledge of complicated NLP and may struggle to produce reliable results.                                                                                                                                                                                                                                                                                                  |
| Gereme et al. [22] | The study works to tackle misleading information in low-resource languages by establishing a cutting-edge algorithm for detecting fake news in Amharic. The model is predicted to be useful in detecting false information in Amharic text. The researchers also developed tools to aid in the detection of fake news in Amharic, such as a large Amharic corpus for training and testing algorithms and word embeddings that record semantic links between words. These contributions not only illustrate the difficulties in combatting false news in low-resource languages but also provide actual answers and tools to overcome these difficulties. The study, however, makes no mention of the Afan Oromo language. | - Model for detecting fake news in Amharic<br>GPAC stands for General-Purpose Amharic Corpus.                                                                                                            | Deep learning techniques.<br>Embeddings of words.                                                                             | The research emphasizes the difficulties in detecting fake news due to insufficient datasets and word embeddings, especially in low-resource African languages like Amharic. It also lacks a full description of the Amharic fake news detection model's weaknesses, such as potential biases or areas of struggle. The assessment is confined to the ETH_FAKE dataset and the AMFTWE word embedding, with no additional datasets or embeddings being considered. |
| Singh [23]         | Fake News Detection on Social Media by Using Deep Learning for Afaan Oromo Language is not mentioned. The goal of the project is to identify fake news by analyzing news items and social network user activity.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Deep learning methods for identifying false news comparison, analysis of results, and investigation of the reasons behind the results.                                                                   | Deep learning techniques comparison.<br>In a vector space, new occurrences are represented.                                   | The information provided has no specified constraints.                                                                                                                                                                                                                                                                                                                                                                                                            |
| Ahmed et al. [24]  | Unfortunately, Fake News Detection on Social Media by Utilizing Deep Learning Specifically for the Afaan Oromo Language or offer any information on it.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | Evaluating a model's ability to identify fake news.<br>Comparing different extractors for text features.                                                                                                 | - TD-IDF vectorizer - Glove embeddings - BERT embeddings.                                                                     | Both fake and real news might come from the same sources. The language used may be misleading.                                                                                                                                                                                                                                                                                                                                                                    |
| Rawat et al. [25]  | In the provided paper, Fake News Detection on Social Media by Using Deep Learning for Afaan Oromo Language is not mentioned. The study uses deep neural networks to evaluate and detect fake news across a variety of modalities.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Looking into different methods and manifestations of dissatisfaction in the administration of online news. An explanation of the traits of fake news and the need for a system to assess its likelihood. | Machine learning for the identification of false news using natural language recognition techniques.                          | It is possible to calculate the classifier's performance. Precision, recall, and f1 scores are used to measure prediction accuracy.                                                                                                                                                                                                                                                                                                                               |

### 3. METHODOLOGY

#### 3.1 Data collections

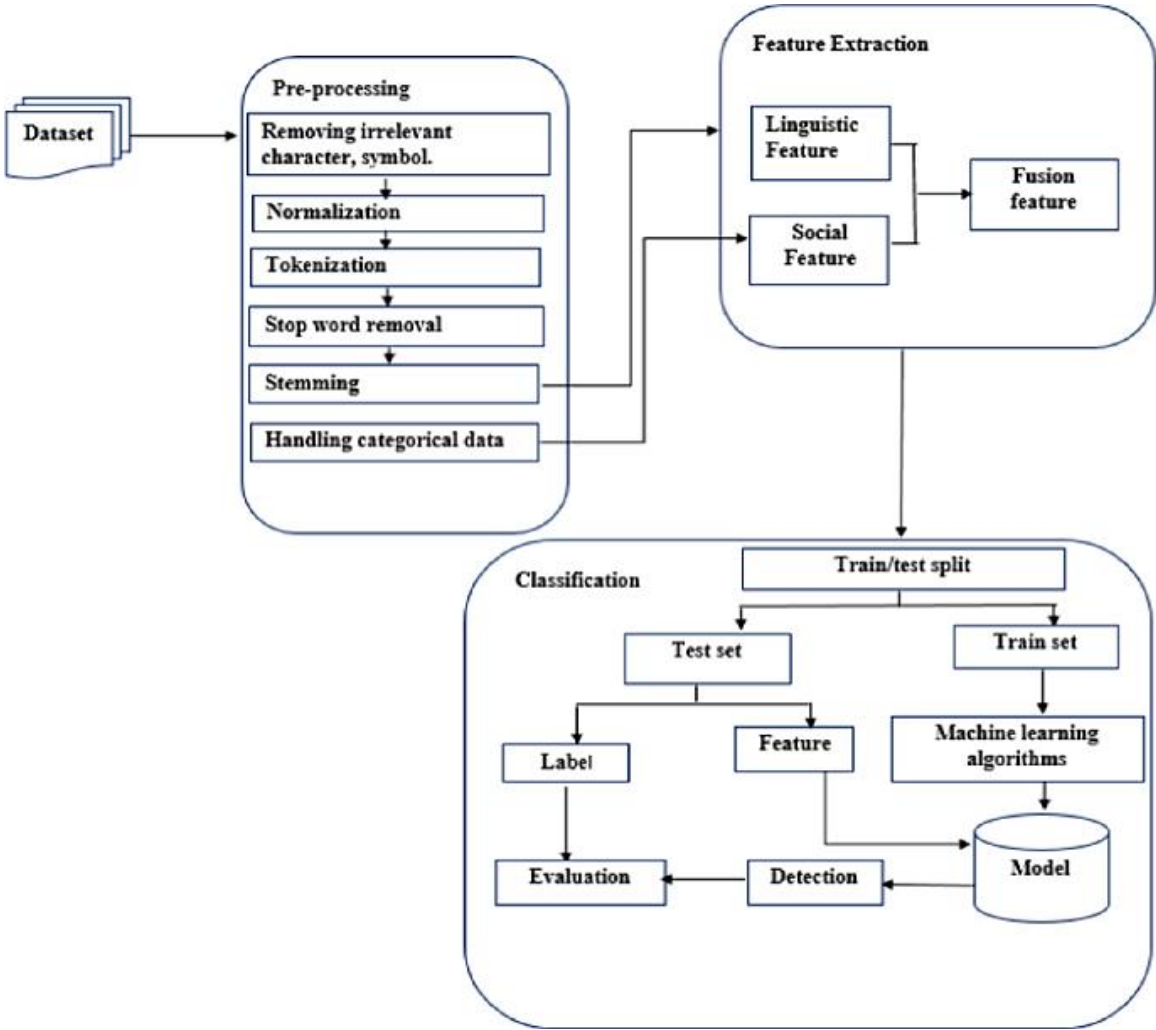
Fake news identification is a burgeoning area of computer science, yet there are few datasets available for classifier training in Afaan Oromo. The primary problem is developing

a big, consistent dataset for training and testing algorithms. The study employed social media as data sources, with data extracted using Octo Parse, an automated web scraping program. Ethiopia's national language, Afaan Oromo, is a low-resource language for Natural Language Processing (NLP) study due to restricted computer resources. To solve this issue, a new dataset for false news identification is required, which

will include extracting Afan Oromo postings from social media, cleaning and filtering each datum, then consolidating and annotating the data into a single dataset [2].

Figure 1 presents a conceptual framework for an innovative

system architecture designed to detect and counteract the spread of fake news, especially within the Afan Oromo language and cultural context [26].



**Figure 1.** Conceptual framework of an innovative system architecture for detecting and countering fake news in the Afan Oromo language and cultural context

3.2 Data preprocessing

Data preprocessing and annotation are essential steps in categorizing news articles as genuine or fraudulent. These processes involve systematic data collection, thorough cleaning, normalization of text, filtering of irrelevant information, and establishing labeling criteria. Data collection involves gathering unprocessed information from various sources, while data cleaning removes duplicate entries, handles missing values, and transforms text into a uniform lowercase format. Data annotation requires explicit criteria to differentiate between real and fake news articles, involving fact-checking, source evaluation, and content analysis. Methodologies for annotation can be manual or automated, with advanced techniques like NLP helping identify patterns associated with fake news propagation. Quality control measures ensure the reliability and accuracy of annotations.

Fake news detection using Octoparse is a complex task that involves analyzing text for credibility, sources, and misinformation signals. To implement fake news detection using Octoparse, define objectives, set up the tool, create a new task, and export data. Analyze the data for fake news

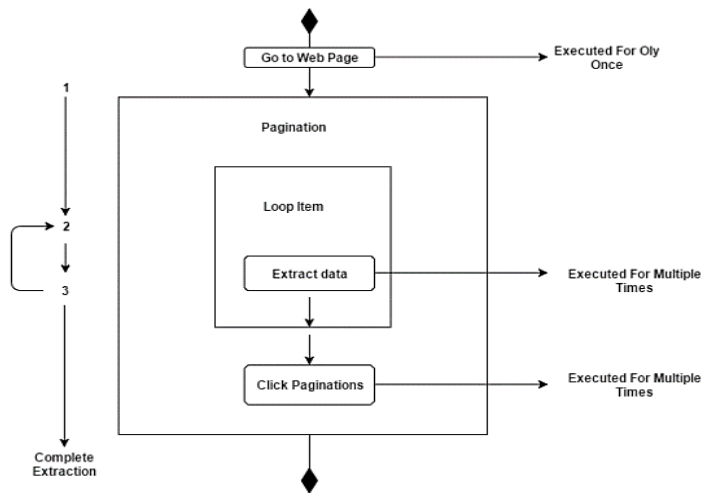
detection, including text preprocessing, feature extraction, and machine learning models. Figure 2 shows that Octaparse implementations for data management and analytics, enable efficient and structured data extraction from various online sources. Regularly schedule scraping and real-time analysis, consider ethical scraping, use APIs for structured data access, and stay updated with misinformation tactics.

3.3 Model

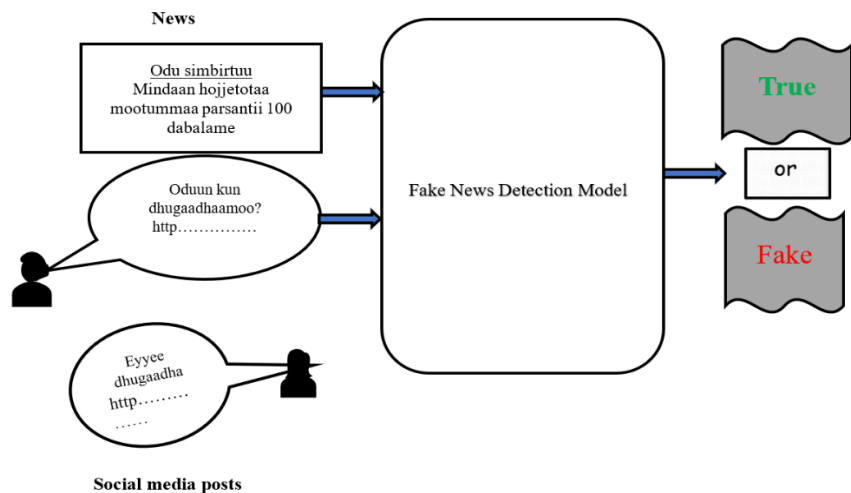
A deep learning model for identifying genuine or fake news articles requires careful consideration of its architecture, training process, hyperparameter selection, and data-splitting strategies. Common deep learning architectures include CNNs, RNNs, and Transformers, particularly BERT models. The training process involves dataset preparation, tokenization, loss function selection, and model optimization. Hyperparameter tuning, dataset splitting, and stratified techniques are crucial. Implementation workflows involve data loading, preprocessing, tokenization, encoding, and rigorous validation. By selecting the right architecture, managing the training process, fine-tuning hyperparameters,

and executing data splitting protocols, a highly effective model can be constructed for news classification. Figure 3 presents a comprehensive framework for identifying and categorizing

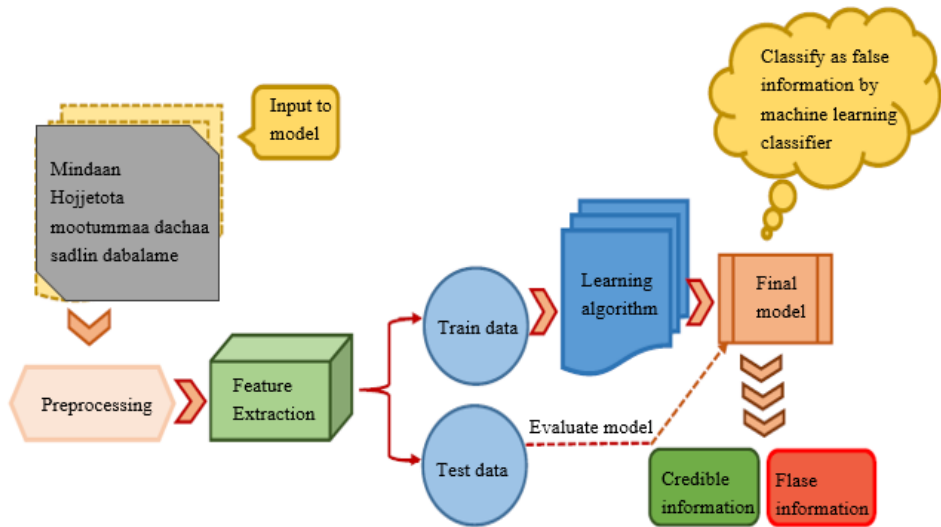
fake news, a type of misleading or deceptive content designed to improve the accuracy and efficiency of media platform detection.



**Figure 2.** Octaparse implementations for data management and analytics, enabling efficient and structured data extraction from various online sources [2]



**Figure 3.** A comprehensive framework for identifying and categorizing fake news, a type of misleading or deceptive content designed to improve the accuracy and efficiency of media platform detection



**Figure 4.** Deep learning model

CNNs, originally designed for computer vision, have demonstrated exceptional performance in text classification and other NLP tasks, even with simple one-layer models. Researchers have explored the inner workings of CNNs, particularly in text classification, to understand their suitability for certain tasks. The study [27] found that filters with different activation patterns and global max-pooling can capture different semantics of grams.

3.4 Feature selections

The study explores advanced feature extraction methods using text mining techniques to reduce redundant detection problems, features, and dimensionality in datasets. Word embeddings are crucial for encapsulating textual data but require significant computational resources, especially in low-resourced languages like Afan Oromo. To address these challenges, the study uses fastText, a library developed by Facebook's AI Research lab, which accommodates unsupervised and supervised learning algorithms for word embeddings. fastText's neural network architecture facilitates embedding generation and serves as a robust classification model. The study also integrates other deep learning algorithms to create a comprehensive analytical framework. The integration of these methods aims to improve text mining performance and contribute to knowledge on feature extraction in low-resourced languages. Figure 4 elaborates the deep learning models and feature extractions process.

4. EVALUATIONS

The study thoroughly compares a variety of sophisticated deep learning models, including well-known architectures such as RNNs and CNNs, in order to gain a comprehensive understanding of the nuances and implications embedded within their confusion matrix. This highlights the performance measures and forecasting capabilities included in each model.

A confusion matrix illustrates the performance of a machine learning model on test data by presenting the number of accurate and incorrect occurrences depending on the algorithm's predictions. It is often used to evaluate the performance of classification models, which are designed to predict a category label for each input occurrence. The matrix shows the number of instances created by the model on the test data, labeled True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). A confusion matrix is critical for evaluating a classification model's performance, especially when there is an unequal class distribution in a dataset, because it goes beyond conventional accuracy measurements. Figure 5 classifications generated by the predictive model, which are commonly known as the predicted labels, thereby providing a comparative analysis, and Figure 6 Confusion Matrix model 1 versus model 2 shows respectively.

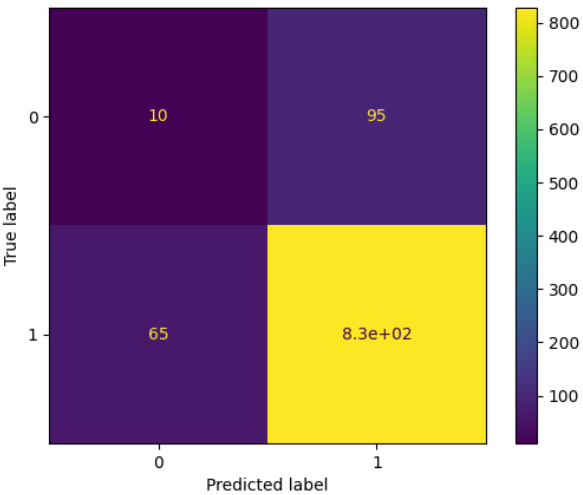


Figure 5. Classifications generated by the predictive model, which are commonly known as the predicted labels, thereby providing a comparative analysis

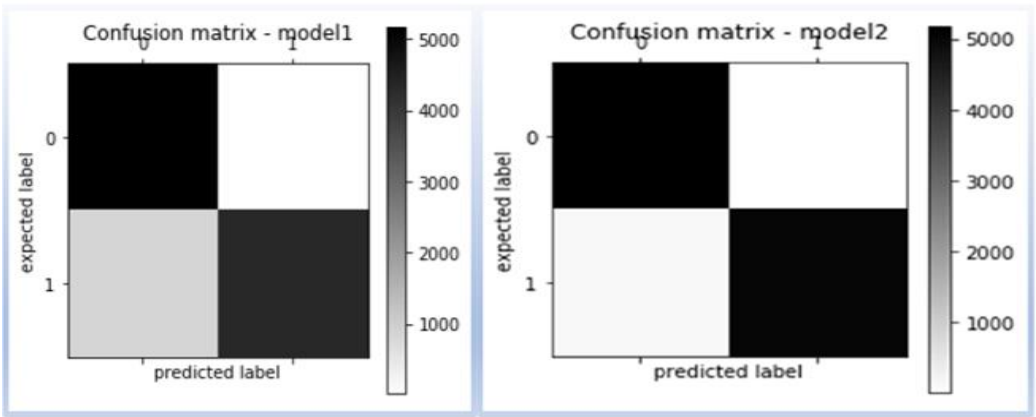


Figure 6. Confusion Matrix model 1 versus model 2

5. RESULTS

The study aims to create a deep learning system for detecting fake news on social media using Afan Oromo news text. The system uses CNNs to accurately categorize data, with a 0.93% accuracy rate. The Bi-LSTM model serves as a foundational prototype, with the CNN emerging as the most effective model. Future research could focus on knowledge-

based strategies to enhance user trust and overcome challenges in detecting fake news. The confusion matrix represents four possible outcomes, and measures like accurate predictions are used to assess the model's performance. The study aims to identify false news and improve the accuracy of machine learning models in detecting fake news. Figures 7 and 8 show model accuracy and model loss, respectively.

5.1 CNNs

The detection of false news using CNNs is a potential deep learning application in NLP. CNNs are commonly employed for image data, but they can also be useful for text categorization problems. To develop a CNN for false news detection, a labeled dataset of news items and their accompanying labels is required. Data must be preprocessed, which includes tokenization, padding, and label encoding. The CNN model will include embedding layers, convolutional layers, pooling layers, and dense layers. This step-by-step explanation explains how to use a CNN to detect false news.

5.2 Comparing the results with the existing methods

The study compares the efficacy of neural networks versus CNNs for data processing and deep learning. It compares the results of modern procedures against older techniques, offering a thorough grasp of their efficacy. The research compares the parameters and time needed by the current approach to the suggested hybrid technique, providing operational efficiency and performance measures. The study also investigates the accuracy rates of the neural network and convolutional

The success of the hybrid approach is heavily dependent on neural networks. The comparison is an invaluable resource for academics and practitioners, exposing each technique's strengths and flaws and highlighting possible areas for development. The findings might have a substantial impact on decision-making processes in the implementation of these technologies across several fields. The findings increase knowledge in this field by laying a strong platform for further investigation and refinement of computational methodologies. Figure 9 presents a detailed comparison of the results with previous.

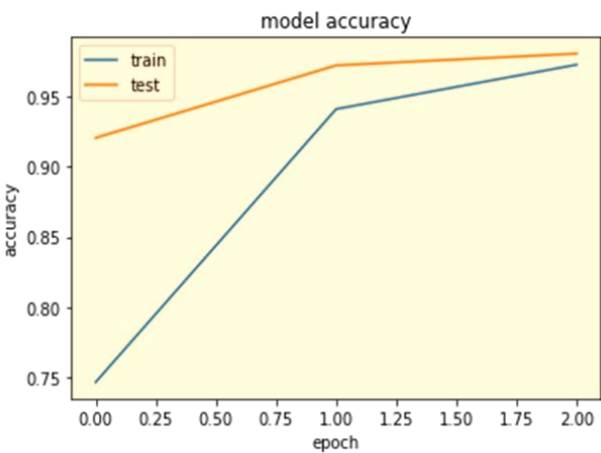


Figure 7. Model accuracy

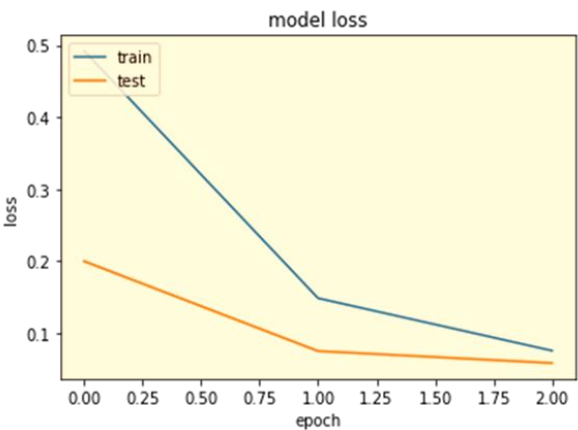


Figure 8. Model loss

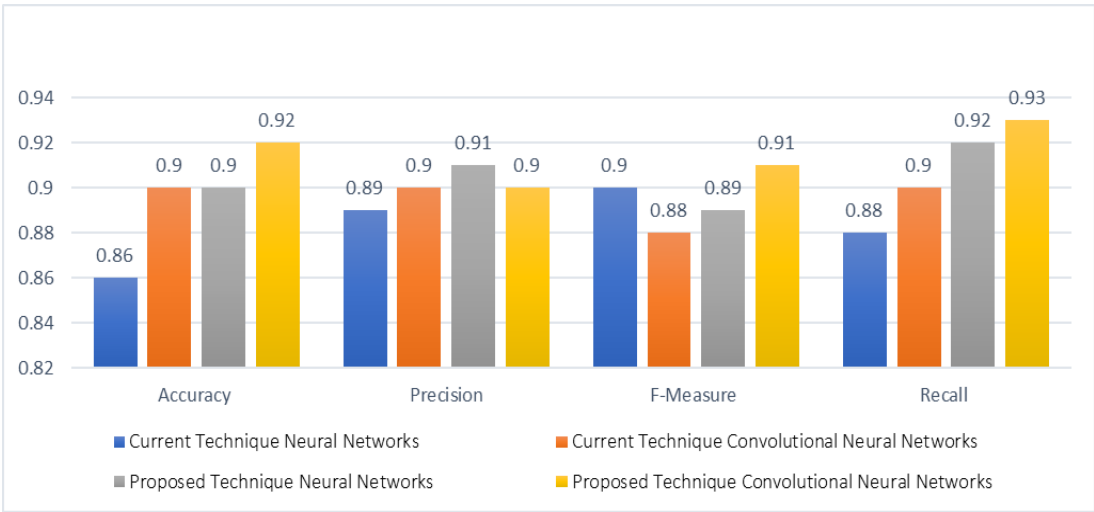


Figure 9. Comparing the results with the existing methods

6. CONCLUSIONS

The objective of this research is to create a deep learning system for detecting false news on social media by employing Afan Oromo news texts. It investigates linguistic issues in the Afan Oromo language and concentrates on recognizing and classifying false news items. The study evaluates deep learning models and provides insights for stakeholders,

empowering local communities to combat misinformation while considering cultural context and ethical considerations.

The pre-trained model is tested using metrics, and a system prototype is developed. CNNs were utilized to categorize the data, with an amazing 93% accuracy rate. The proposed strategies achieved a 91% accuracy rate. The primary objective is to create an advanced automated system capable of proficiently identifying instances of fake news disseminated

across various social media platforms, specifically focusing on content published in the Afan Oromo language through sophisticated ensemble approaches.

The Bi-LSTM model has been conceptualized as a foundational prototype that will serve as a critical reference point for subsequent research initiatives. The findings indicate that the CNN emerged as the most effective and highest-performing model, while the attention mechanism exhibited a performance level that was notably inferior compared to its corresponding baseline model.

Future research could focus on knowledge-based strategies to enhance user trust in the system and overcome the multifaceted challenges associated with detecting fake news that is predominantly reliant on supervised models of textual content.

## ACKNOWLEDGMENTS

We acknowledge the Advanced and Innovative Research Laboratory (AAIR Labs: [www.aairlab.com](http://www.aairlab.com)), India for technical support.

## FUNDING

This research was supported by Universidad Nacional Mayor de San Marcos – R. R. No. 004305-2024-R, Project C24202791 – Project PCONFIGI, for the year 2024.

## REFERENCES

- [1] Raza, A., Usmani, R.S.A., Khan, S.U.R. (2024). A hybrid deep learning based fake news detection system using temporal features. *The Asian Bulletin of Big Data Management*, 4(2): 264-273. <https://doi.org/10.62019/abbdm.v4i02.179>
- [2] Waldesanbet, A. (2021). Afaan Oromo fake news detection on social media by using deep learning approach. Master Thesis, Ambo University Hachalu Hundessa Campus, Ethiopia. [https://www.researchgate.net/publication/392324643\\_Afaan\\_Oromo\\_Fake\\_News\\_Detection\\_on\\_social\\_media\\_by\\_using\\_Deep\\_Learning\\_Approach](https://www.researchgate.net/publication/392324643_Afaan_Oromo_Fake_News_Detection_on_social_media_by_using_Deep_Learning_Approach).
- [3] Regesa, H.B. (2020). Detection and classification of Afaan Oromo fake news on social media: Using a machine learning approach. Master Thesis, Ambo University Woliso Campus School. <https://pdfcoffee.com/detection-and-classification-of-afaan-oromo-fake-news-on-social-media-using-a-machine-learning-approach-pdf-free.html>.
- [4] Kadir, A. (2020). Course: Advanced topic in information system fake news detection using machine learning approach (The case of social media by Afaan Oromo). Dire Dawa University.
- [5] Abebe, M. (2021). Fake news detection for Amharic language using deep learning by Ermias Nigatu Hailemichael Office of graduate studies. [https://www.academia.edu/77248002/Fake\\_News\\_Detection\\_for\\_Amharic\\_Language\\_Using\\_Deep\\_Learning\\_by\\_Ermias\\_Nigatu\\_Hailemichael\\_Office\\_of\\_Graduate\\_Studies](https://www.academia.edu/77248002/Fake_News_Detection_for_Amharic_Language_Using_Deep_Learning_by_Ermias_Nigatu_Hailemichael_Office_of_Graduate_Studies).
- [6] Nasir, J.A., Khan, O.S., Varlamis, I. (2021). Fake news

- detection: A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights*, 1(1): 100007. <https://doi.org/10.1016/j.jjime.2020.100007>
- [7] Kaba, G.D., Olalekan, S.A., Asrat, G. (2022). Afaan Oromo language fake news detection in social media using convolutional neural network and long short term memory. *Journal of Electrical and Electronics Engineering*, 15(2): 37-42.
- [8] Ganfure, G.O. (2022). Comparative analysis of deep learning based Afaan Oromo hate speech detection. *Journal of Big Data*, 9(1): 76. <https://doi.org/10.1186/s40537-022-00628-w>
- [9] Pavlyshenko, B.M. (2023). Analysis of disinformation and fake news detection using fine-tuned large language model. arXiv preprint arXiv:2309.04704. <https://doi.org/10.48550/arXiv.2309.04704>
- [10] Kotiyal, B., Pathak, H., Singh, N. (2023). Debunking multi-lingual social media posts using deep learning. *International Journal of Information Technology*, 15(5): 2569-2581. <https://doi.org/10.1007/s41870-023-01288-6>
- [11] Gayathri, S., Santhiya, S., Nowneesh, T., Sanjana, Shuruthy, K., Sakthi, S. (2023). Deep fake detection using deep learning techniques: A literature review. *Applied and Computational Engineering*, 2: 232-241. <https://doi.org/10.54254/2755-2721/2/20220655>
- [12] Güler, G., Gündüz, S. (2023). Deep learning based fake news detection on social media. *International Journal of Information Security Science*, 12(2): 1-21. <https://doi.org/10.55859/ijiss.1231423>
- [13] Edwards, T., Noel, R.R. (2023). Detecting fake news using machine learning based approaches. In 2023 6th International Conference on Information and Computer Technologies (ICICT), Raleigh, NC, USA, pp. 121-126. <https://doi.org/10.1109/ICICT58900.2023.00027>
- [14] Safira, A.D., Setiawan, E.B. (2023). Hoax detection in social media using bidirectional long short-term memory (Bi-LSTM) and 1 dimensional-convolutional neural network (1D-CNN) methods. In 2023 11th International Conference on Information and Communication Technology (ICoICT), Melaka, Malaysia, pp. 355-360. <https://doi.org/10.1109/ICoICT58202.2023.10262528>
- [15] Arega, K.L., Alasadi, M.K., Yaseen, A.J., Salau, A.O., Braide, S.L., Bandle, J.O. (2023). Machine learning based detection of fake Facebook profiles in Afan Oromo language. *Mathematical Modelling of Engineering Problems*, 10(6): 1987-1993. <https://doi.org/10.18280/mmep.100608>
- [16] Mosisa, S. (2023). Afaan Oromo multi-label news text classification using deep learning approach. *Research Square*. <https://doi.org/10.21203/rs.3.rs-3644142/v1>
- [17] Gohil, H., Joshi, V., Gandhi, S. (2024). Fake news detection using hybrid approach. *Journal of Information and Knowledge*, 61(2): 77-82. <https://doi.org/10.17821/srels/2024/v61i2/171046>
- [18] Salau, A.O., Arega, K.L., Tin, T.T., Quansah, A., Sefa-Boateng, K., Chowdhury, I.J., Braide, S.L. (2024). Machine learning-based detection of fake news in Afan Oromo language. *Bulletin of Electrical Engineering and Informatics*, 13(6): 4260-4272. <https://doi.org/10.11591/eei.v13i6.8016>
- [19] Abera, S., Tegegne, T. (2019). Information extraction model for Afan Oromo news text. In *Information and Communication Technology for Development for Africa*

- (ICT4DA 2019). Communications in Computer and Information Science, vol 1026. Springer, Cham. [https://doi.org/10.1007/978-3-030-26630-1\\_28](https://doi.org/10.1007/978-3-030-26630-1_28)
- [20] Tazeze, T., Raghavendra, R. (2021). Building a dataset for detecting fake news in Amharic language. International Journal of Advanced Research in Science, Communication and Technology, 6(1): 76-83. <https://doi.org/10.48175/ijarsct-1362>
- [21] Al-Tai, M.H., Nema, B.M., Al-Sherbaz, A. (2023). Deep learning for fake news detection: Literature review. Al-Mustansiriyah Journal of Science, 34(2): 70-81. <https://doi.org/10.23851/mjs.v34i2.1292>
- [22] Gereme, F., Zhu, W., Ayall, T., Alemu, D. (2021). Combating fake news in “low-resource” languages: Amharic fake news detection accompanied by resource crafting. Information, 12(1): 20. <https://doi.org/10.3390/info12010020>
- [23] Singh, L. (2020). Fake News Detection: A comparison between available Deep Learning techniques in vector space. In 2020 IEEE 4th Conference on Information & Communication Technology (CICT), Chennai, India, pp. 1-4. <https://doi.org/10.1109/CICT51604.2020.9312099>
- [24] Ahmed, B., Ali, G., Hussain, A., Baseer, A., Ahmed, J. (2021). Analysis of text feature extractors using deep learning on fake news. Engineering, Technology & Applied Science Research, 11(2): 7001-7005. <https://doi.org/10.48084/etasr.4069>
- [25] Rawat, G., Pandey, T., Singh, T., Yadav, S., Aggarwal, P.K. (2023). Fake news detection using machine learning. In 2023 International Conference on Artificial Intelligence and Smart Communication (AISC), Greater Noida, India, pp. 759-762. <https://doi.org/10.1109/AISC56616.2023.10085488>
- [26] Worku, M.H., Woldeyohannis, M.M. (2022). Amharic fake news detection on social media using feature fusion. In Advances of Science and Technology (ICAST 2021). Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Springer, Cham. [https://doi.org/10.1007/978-3-030-93709-6\\_31](https://doi.org/10.1007/978-3-030-93709-6_31)
- [27] Toscano, J.D., Oommen, V., Varghese, A.J. Zou, Z., Daryakenari, N.A., Wu, C., Karniadakis, G.E. (2025). From PINNs to PIKANs: Recent advances in physics-informed machine learning. Machine Learning for Computational Science and Engineering, 1: 15. <https://doi.org/10.1007/s44379-025-00015-1>