Evaluating the effectiveness of hybrid features in fake news detection on social media

1st Mesay Gemeda Yigezu
Dept. Computer Science, NLP lab.
Centro de Investigación en Computación
Instituto Politécnico Nacional, Mexico
mgemedak2022@cic.ipn.mx

2nd Moges Ahmed Mehamed
School of Computer Science and AI
Wuhan University of Technology, China
moges.ahmed@whut.edu.cn

3rd Olga Kolesnikova
Dept. Computer Science, NLP lab.
Centro de Investigación en Computación
Instituto Politécnico Nacional, Mexico
kolesolga@gmail.com

4th Tadesse Kebede Guge
Dept. Computer Science
Haramaya University, Ethiopia
taddeekb@gmail.com

5th Alexander Gelbukh
Dept. Computer Science, NLP lab.
Centro de Investigación en Computación
Instituto Politécnico Nacional, Mexico
gelbukh@cic.ipn.mx

6th Grigori Sidorov
Dept. Computer Science, NLP lab.
Centro de Investigación en Computación
Instituto Politécnico Nacional, Mexico
sidorov@cic.ipn.mx

***Abstract*—This research focuses on the detection of fake news on social media, specifically in Amharic language posts. The study highlights the importance of utilizing hybrid features, which include both news content and social content features, to improve the accuracy of fake news detection. We evaluated the effectiveness of these hybrid features using state-of-the-art methodologies and explored methods that optimize detection accuracy and reduce latency. Our research revealed that we achieved an impressive F1-score of 0.99 by utilizing a BERT-based uncased model. This outcome was obtained by incorporating a combination of textual content, publication dates, and page creation dates as hybrid features.**

***Index Terms*—fake news, hybrid features, pre-trained model, social context, and news content.**

I. Introduction

With the growing prevalence of social media platforms in our daily lives, an increasing number of individuals tend to turn to and consume news from social media instead of conventional news outlets [1], [2] [1] Among those social media platforms, Facebook, Twitter, and Reddit are the most widely used platforms to deliver information. These social media platforms allow people to access the news and share news at less cost and short time. This dramatically changed the way people access news information. Though social media platforms have made news sharing more accessible, they are also vulnerable to various issues such as hate speech, rumor-mongering, the spread of fake news, conspiracy theories, xenophobia, racism, and prejudice. In developing countries, this is particularly problematic as the dissemination of fake news through social media has been identified as a cause of conflict and war [3], [4]. According to [5], fake news refers to intentionally crafted, sensational, emotionally charged, misleading, or entirely fabricated information that resembles mainstream news. The widespread prevalence of fake news has negative repercussions for social media users and the wider community, including the potential to disrupt the balance of authentic news reporting, influence people's beliefs with false information, and alter the way people understand and react to legitimate news [6], [7]. To mitigate these effects, developing advanced techniques for automatically detecting fake news on social media can benefit the community and the news environment. However, detecting fake news on social media is difficult, especially when relying solely on textual features. This is because fake news on social media includes various features such as news content, social networking, user-based, and action-based features, which collectively determine the authenticity of the news. The use of linguistic features alone is not sufficient for detecting fake news [8]. Therefore, combining the appropriate news and social context features can improve the accuracy and detection speed of fake news [9]. According to [10], social media news can be classified into content-related and social content features. The former encompasses valuable features derived from linguistics, while the latter includes features sourced from user profiles, posts, shares, and social networks. News from social media can further include sub-features such as user-related, propagation-related, and action-related features. To detect fake news on social media, additional information extracted from social context is required [8]. Despite the increased prevalence of fake news in Amharic language social media posts, there is still a lack of attempts to use hybrid features, i.e., news content and social content features. Thus, the target of this research is:

- To evaluate effectiveness of hybrid features in fake news detection on social media using state of art methodology.
- To investigate the appropriate methods that optimize fake news detection accuracy and decrease detection latency.

The remaining sections of the paper are structured as follows: Section II provides an extensive review of relevant literature. Section III presents a description of the utilized dataset. Challenges associated with fake news detection are discussed in Section IV. The proposed framework is elucidated

in Section V. Section VI outlines the conducted experiments for fake news detection. Model evaluation is expounded upon in Section VII, and the results of the experiments are discussed in Section VIII. Finally, a summary of the study's findings and conclusions is provided in section IX.

II. Related work

This section covers various studies on fake news detection for the Amharic language. Numerous research studies have been carried out to detect fake news in Amharic, utilizing diverse methodologies. The source of data, experimental procedures, performance metrics, and challenges associated with using different features were also taken into account.

In his research, [14] utilized a Dense Neural Network to detect fake news based on stance for Amharic news. The dataset comprised 5,000 news articles, with 1,546 labeled as fake news and 3,454 as genuine news. The study focused solely on news content features for fake news detection, achieving an accuracy of 95.12% using a finely-tuned Tf-IDF - Dense Neural Network (DNN). However, social features, which also play a significant role in identifying fake news, were not taken into account in this analysis.

The detection of fake social media news in the Amharic language has been developed by [15] using a machine learning model. The objective of this research was to improve the accuracy of detecting fake accounts on social media. However, the study resulted in an insufficient level of accuracy in detecting fake news in Amharic. Additionally, the study relied on the English MIB fake news dataset and translated it into Amharic, while the detection of fake posts was not included, despite the use of textual content and some social content features such as the date and time of publication, language, and geo-positioning.

The study in [16] explored the use of deep learning to detect fake news in the Amharic language. The dataset used in the research consists of news articles containing false context, clickbait, parody, or satire, which are labeled as fake. The study employs FastText and word embedding feature extraction techniques. A noteworthy achievement of this work is the collection of 12,000 fake news instances from social media. While the focus of the research is solely on detecting fake news based on the content of the articles, the authors suggest that future work should also consider feature extraction and dataset collection.

Gereme F. et al. [17] put forward a deep learning model to tackle the issue of fake news in the Amharic language. To this end, they developed the first-ever Amharic fake news detection dataset, which included 3417 real news articles and 3417 fake news articles obtained from various Amharic Facebook pages and online newspapers, resulting in a total of 6834 articles. Notably, the dataset came with fine-grained labels and the authors also created Amharic FastText word embedding, which is a significant contribution. However, the study's limitation lies in the fact that the Amharic fake news dataset was manually crafted and only news content features were considered.

Tewodros T. et al. [12] developed a dataset aimed at identifying fabricated news in the Amharic language. The research involved gathering approximately 7547 news articles and utilizing techniques such as TF-IDF, CountVectorizer, and N-gram to extract features from textual content. However, the scope of the study did not encompass social media characteristics.

Ibrahim N et al. [13] devised a method for detecting fake news in Amharic by using a lexicon-stance-based approach. The approach aimed to identify fake news by analyzing its content, without relying on external sources or prior knowledge about the news. The researchers proposed incorporating important news stance features, such as headline-to-article similarity, headline-to-headline similarity, and page-generated score, into a state-of-the-art text representation lexicon-based approach, in order to enhance the performance of fake news detection. The study found that incorporating these stance features extracted from news attributes improved the detection of fake news using a lexicon-based approach. However, the paper's limitation is that it solely focuses on content-based analysis and does not utilize external sources of news. Additionally, the study did not examine network analysis or social context features for detecting Amharic fake news.

Worku M et al. [19] were concerned with Amharic fake news detection on social media by using feature fusion and used NLP-based feature extraction techniques such as TF-IDF and word2vec for the social context. They collected 4590 news data which is Facebook and obtained an accuracy of 99.67% with the Random Forest classifier. The author did not make use of the most recent and cutting-edge strategies that are currently considered to be state-of-the-art in the field.

Overall, most previous research on the topic of detecting fake news in Amharic has focused on news content. Another issue is that the models used to identify fake news, which was discussed earlier, do not attain an accuracy that is promising. In order to build a method of detecting fake news in Amharic that is both highly accurate and has a short detection latency time, it is essential to apply features drawn from news and social content.

III. Dataset

During our experimental research, we used a publicly available data set. According to [19], the data set is available to the public through a GitHub page [1] for fake news detection. It was gathered from Facebook pages from various categories, including news media, broadcasting pages, bloggers, and politicians, which were selected based on the frequency of news post content. The pages were analyzed by domain experts (journalists) to differentiate between real and fake news.

The data was collected using a web scraper tool, covering Amharic news content. In total, 4,590 data with six attributes were collected, including news content, posted date, published page, created date, verification of published Facebook pages, and the number of followers and likes of published

[1] https://github.com/MenbereHailu/Fake$_{n}ews_{D}etection$

pages. Moreover, the required pre-processing such as removing irrelevant characters, and Amharic character normalization was done. The data was labeled as fake when the statement contradicted the truth, seemed true but was interpreted in the wrong context, the news was mismatched, or the date of the event and the news did not match. Out of 4,590 instances, 2,211 news items were labeled as fake and 2,379 as real [11]. Figure 1 shows the distribution of fake and real news.

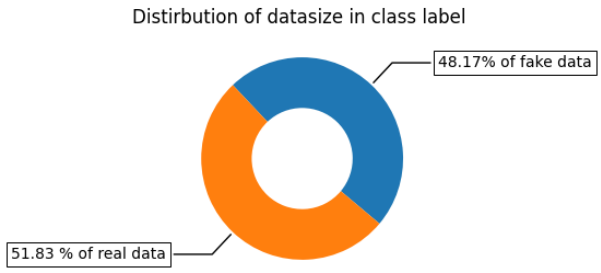

Fig. 1. Dataset size in class label

.

IV. Challenges of Fake News Detection

Social media sites have become a popular source of information due to their accessibility and affordability, but they have also become a breeding ground for misinformation and fake content. The proliferation of fake news on social media platforms has become a significant challenge in recent years [20]. The detection of fake news is a complex and challenging task that requires a comprehensive understanding of the underlying issues. In this section, we explore the challenges associated with detecting fake news and propose a novel approach to detect fake news on social media.

One of the primary challenges of fake news detection is the exponential growth of information in online news portals and social media sites which makes it difficult to distinguish between real and fake. Besides, the speed at which fake news travels is unprecedented, and the outcome is its wide-scale propagation which makes it challenging to detect and address before it goes viral [21]. Another obstacle to detecting fake news is the increasing sophistication of fake news creators. They use advanced techniques such as deep fakes [22], and AI-generated content [23]. These algorithms enable content to be both synthetically generated and seamlessly edited. Deepfakes swap the attributes (e.g., face, voice, skin tone, gender, accessory colors, fashion designs) of the source with those of the target by training a deep neural network [24], [25]. Those technologies create convincing fake news stories that can be difficult to detect. This requires advanced tools and techniques to identify and analyze these types of content.

Another issue that we want to highlight here is the scarcity of labeled fake news data (news labeled as real or fake) in real-world scenarios. Existing state-of-the-art works generally use fully labeled data to classify fake news. However, the real-world data is likely to be largely unlabeled [21].

In conclusion, detecting fake news on social media is a complex and challenging task that requires a multifaceted approach. Our proposed hybrid feature approach aims to address some of the challenges associated with fake news detection by combining different features to improve the accuracy of the detection process. However, further research is needed to develop more effective and efficient methods for detecting fake news on social media.

V. Proposed Framework

Our proposed method for enhancing the accuracy of fake news detection involves integrating a variety of features, including linguistic features, into a comprehensive evaluation approach. We aim to develop a robust framework that can effectively distinguish between authentic and fake news by considering multiple factors. Our strategy involves an extensive examination of various features, including news content features such as the publication date, as well as social context features such as the number of followers, the page creation date, and the number of likes. By combining these hybrid features, we have created a more comprehensive and reliable model for detecting fake news, As Figure 2 shows, we considered the following steps:

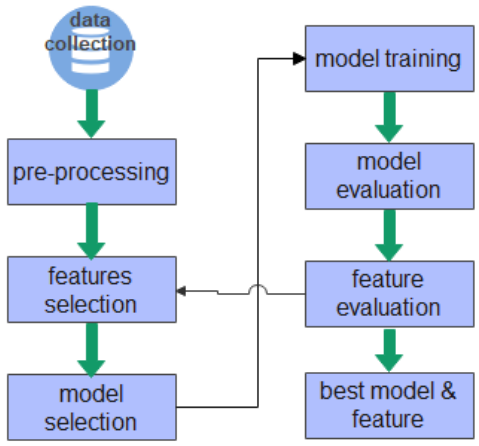

Fig. 2. Proposed frame work

.

Step-1 **Data Collection:** gather a comprehensive dataset consisting of both real and fake news articles.

Step-2 **Pre-processing:** clean and pre-process the collected data to remove irrelevant information, such as HTML tags, punctuation, and stopwords.

Step-3 **Feature Selection:** Select appropriate features to detect fake news.

Step-4 **Model Selection:** Choose an appropriate pre-trained model for fake news detection.

Step-5 **Model Training:** Use the training dataset to train the selected model using the extracted features.

No. features	Features
Feature - 1	Text only
Feature - 2	Text + the number of followers
Feature - 3	Text+ the date of publication + the date of page created

TABLE I
COMBINATION OF FEATURES

Step-6 **Model Evaluation:** evaluate the trained model using the testing set. Metrics such as accuracy, precision, recall, and F1-score were used to assess the performance of the model.

Step-7 **Feature Evaluation:** evaluate the features based on the model performance then go to **Step-3** Iteratively improve until the model achieves better performance.

Step-8 **Announcing best model and feature:** publish the model and features which achieve a better result.

VI. EXPERIMENTS

We carried out comprehensive experiments aimed at assessing the efficacy of distinct features and models in effectively differentiating between fake and real news. Throughout the experiments, we explored a range of non-linguistic features by integrating them with linguistic features. To enhance the performance in combating fake news, we employed diverse pre-trained models including BERT base, XLM-RoBERTa base, and AfriBERTa base. For each model, we assessed the effectiveness of the features by combining them with textual data, specifically user comments.

VII. MODEL EVALUATION

Our objective was to identify the most effective model for distinguishing between real and fake news by evaluating different features. We evaluated the performance of the models using a combination of precision, recall, and F1 scores and determined the best-performing model based on weighted scores. Furthermore, we examined the effectiveness of the features used in the classification process by considering the F1 score results of the models. The F1 score was chosen as an appropriate metric for this evaluation because it considers both precision and recall, providing an overall assessment of the features' ability to accurately classify real and fake instances.

VIII. RESULTS AND DISCUSSION

According to Table I and II, the initial experiment aimed to evaluate Feature 1 using three distinct pre-trained models: BERT, AfriBERTa, and XLM-RoBERTa. Precision, recall, and F1 score were used to measure the performance of each model. The results showed that BERT achieved a precision, recall score, and F1 score of 0.73. Conversely, AfriBERTa outperformed the other techniques with a precision of 0.87, a recall of 0.87, and an F1 of 0.87. Lastly, XLM-Roberta yielded a precision score of 0.26, a recall of 0.50, and an F1 of 0.34 when used with Feature 1. These findings demonstrate that AfriBERTa performed better than the other techniques in terms of precision, recall, and overall F1 score in this particular experiment.

In the second experiment, our focus was on utilizing Feature 2 which is text + the number of followers. BERT achieved exceptional performance with a precision of 0.99, a recall of 0.98, and an F1 of 0.98. On the other hand, AfriBERTa obtained a precision of 0.87, a recall of 0.87, and an F1 of 0.87. When XLM-RoBERTa base was used in combination with the feature, it resulted in a precision of 0.26, a recall of 0.52, and an F1 of 0.35. Based on these results, it is evident that BERT outperformed the other pre-trained models.

Features	Models	Weighted score		
		Precision	Recall	F1-score
Feature - 1	bert-base-uncased	0.73	0.73	0.73
	afriberta_base	0.87	0.87	0.87
	xlm-roberta-base	0.26	0.50	0.34
Feature - 2	bert-base-uncased	0.99	0.98	0.98
	afriberta_basse	0.87	0.87	0.87
	xlm-roberta-base	0.26	0.52	0.35
Feature - 3	bert-base-uncased	0.99	0.99	0.99
	afriberta_base	0.89	0.89	0.89
	xlm-roberta-base	0.29	0.61	0.39

TABLE II
EXPERIMENTAL RESULTS

During the third experiment using Feature 3 in Table I, which is text, the date of publication, and the date of page created, we observed impressive results when using BERT, with precision, recall, and an F1 of 0.99, indicating a high level of accuracy and effectiveness. AfriBERTa demonstrated slightly lower performance than BERT, with precision, recall, and F1 of 0.89, which still indicates reasonably good performance. However, Xlm roberta utilizing Feature exhibited significantly lower precision, recall, and F1 score, with values of 0.29, 0.61, and 0.39, respectively. These results suggest that Xlm roberta when combined with Feature 3 that is text, the date of publication, and the date of page created, did not perform as well as the other techniques in the experiment.

In summary, the study aimed to evaluate the effectiveness of different features in identifying fake news using various pre-trained models. The performance of the models varied depending on the features used. AfriBERTa was effective in identifying fake news using only textual content, while Bert-base-uncased performed better with additional features such as the number of followers and dates. XLM-RoBERTa showed weaker performance than the other models. The study concluded that the choice of pre-trained models and feature combinations are important in achieving optimal results in fake news detection and that incorporating feature information enhances accuracy.

IX. CONCLUSION

Our research aimed to assess the effectiveness of different features and pre-trained models in detecting fake news on social media. The findings highlighted the importance of utilizing hybrid features that combine both news content and social content features. As a result, we recognized the need for additional information extracted from the social context to effectively detect fake news.

In the experiments we conducted, three distinct pre-trained models were evaluated: BERT, AfriBERTa, and XLM-RoBERTa.

The study emphasized that the choice of pre-trained models and feature combinations plays a crucial role in achieving optimal results in fake news detection. The incorporation of additional feature information, particularly from the social context, was found to enhance the accuracy of fake news detection. These insights can contribute to the development of more effective methodologies for identifying and combating fake news on social media platforms.

ACKNOWLEDGMENTS

The work was done with partial support from the Mexican Government through the grant A1S-47854 of CONACYT, Mexico, grants 20220852, 20220859, and 20221627 of the Secretaría de Investigación y Posgrado of the Instituto Politécnico Nacional, Mexico. The authors thank the CONACYT for the computing resources brought to them through the Plataforma de Aprendizaje Profundo para Tecnologías del Lenguaje of the Laboratorio de Supercómputo of the INAOE, Mexico and acknowledge the support of Microsoft through the Microsoft Latin America Ph.D. Award.

REFERENCES

[1] Collins, B., et al., Trends in combating fake news on social media–a survey. Journal of Information and Telecommunication, 2021. 5(2): p. 247-266.

[2] Yigezu, M. G., Bade, G. Y., Kolesnikova, O., Sidorov, G., Gelbukh, A. (2023). Multilingual Hope Speech Detection using Machine Learning.

[3] European Institute of Peace. Fake news misinformation and hate speech in Ethiopia: A vulnerability assessment,12 APRIL 2021

[4] Yigezu, M. G., Kolesnikova, O., Sidorov, G., Gelbukh, A. (2023). Transformer-Based Hate Speech Detection for Multi-Class and Multi-Label Classification.

[5] White, A. (2022). Review essay: fake news, and online misinformation and disinformation: Fake news: understanding media and misinformation in the digital age, edited by Melissa Zimdars and Kembrew McLeod, Cambridge,Mass. London,The MIT Press,2020,xl+ 395 pp., US 38(paperback), ISBN978-0-262-53836 7

[6] Yigezu, M. G., Tonja, A. L., Kolesnikova, O., Tash, M. S., Sidorov, G., Gelbukh, A. (2022, December). Word Level Language Identification in Code-mixed Kannada-English Texts using Deep Learning Approach. In Proceedings of the 19th International Conference on Natural Language Processing (ICON): Shared Task on Word Level Language Identification in Code-mixed Kannada-English Texts (pp. 29-33).

[7] Arif, M., Tonja, A. L., Ameer, I., Kolesnikova, O., Gelbukh, A., Sidorov, G., Meque, A. G. M. (2022). CIC at CheckThat! 2022: multi-class and cross-lingual fake news detection. Working Notes of CLEF.

[8] Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter, 19(1), 22-36.

[9] Jarrahi, A., Safari, L. (2023). Evaluating the effectiveness of publishers' features in fake news detection on social media. Multimedia Tools and Applications, 82(2), 2913-2939.

[10] Hassan, I., Azmi, M. N. L., Abdullahi, A. M. (2020). Evaluating the spread of fake news and its detection. Techniques on social networking sites. Romanian Journal of Communication and Public Relations, 22(1), 111-125.

[11] Worku, M. H., Woldeyohannis, M. M. (2022). Amharic Fake News Detection on Social Media Using Feature Fusion. In Advances of Science and Technology: 9th EAI International Conference, ICAST 2021, Hybrid Event, Bahir Dar, Ethiopia, August 27–29, 2021, Proceedings, Part I (pp. 468-479). Springer International Publishing.

[12] TT, R. (2021). Building a Dataset for Detecting fake news in Amharic language building a dataset for detecting fake news in Amharic Language. Int J Adv Res Sci Commun Technol, 6(1), 2-9.

[13] Ibrahim N. Sosina M., "Lexicon-Stance Based Amharic Fake News Detection," Addis Ababa University, Addis Ababa, 2022.

[14] T. W. Tefera, "Stance-Based Fake News Detection for Amharic news Using Dense Neural Network," Jimma University, Jimma, 2019.

[15] Kedir, L. "Social Media Fake Account Detection for Amharic Language by using Machine Learning," Global Scientific Journals, vol. 8, no. 6, pp. 604-614, 2020.

[16] Hailemichael, E. N., Ermias, N. (2021). Fake news detection for Amharic language using deep learning. Unpublished, (2021), 2-100.

[17] 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.

[18] TT, R. (2021). Building a Dataset for Detecting fake news in Amharic language building a dataset for detecting fake news in Amharic Language. Int J Adv Res Sci Commun Technol, 6(1), 2-9.

[19] Worku, M. H., Woldeyohannis, M. M. (2022). Amharic Fake News Detection on Social Media Using Feature Fusion. In Advances of Science and Technology: 9th EAI International Conference, ICAST 2021, Hybrid Event, Bahir Dar, Ethiopia, August 27–29, 2021, Proceedings, Part I (pp. 468-479). Springer International Publishing.

[20] Collins, B., et al., Trends in combating fake news on social media–a survey. Journal of Information and Telecommunication, 2021. 5(2): p. 247-266.

[21] Raza, S. and C. Ding, Fake news detection based on news content and social contexts: a transformer-based approach. International Journal of Data Science and Analytics, 2022. 13(4): p. 335-362.

[22] Hussain, S., et al. Adversarial deepfakes: Evaluating vulnerability of deepfake detectors to adversarial examples. in Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2021.

[23] Uzun, L., ChatGPT and academic integrity concerns: Detecting artificial intelligence generated content. Language Education and Technology, 2023. 3(1).

[24] Campbell, C., et al., Preparing for an era of deepfakes and AI-generated ads: A framework for understanding responses to manipulated advertising. Journal of Advertising, 2022. 51(1): p. 22-38.

[25] Becker, C. and R. Laycock, Embracing deepfakes and AI-generated images in neuroscience research. European Journal of Neuroscience, 2023.