



ADDIS ABABA UNIVERSITY
ADDIS ABABA INSTITUTE OF TECHNOLOGY (AAiT)
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

**MULTIMODAL AMHARIC FAKE NEWS
DETECTION USING CNN-BILSTM**

BY
MEKDIM TESSEMA

ADVISOR
Dr. FITSUM ASSAMNEW

A thesis submitted to the School of Electrical and Computer Engineering in partial
fulfillment of the requirements for the Degree of Master of Science in Computer
Engineering

June, 2024
ADDIS ABABA, ETHIOPIA

ADDIS ABABA UNIVERSITY
ADDIS ABABA INSTITUTE OF TECHNOLOGY
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

The undersigned have examined the thesis titled:

Multimodal Amharic Fake News Detection using CNN-BiLSTM

BY
MEKDIM TESSEMA

Approval by Boards of Examiners

Dr. Bisrat Derebssa

Dean, SECE, AAiT

Date

Signature

Dr. Fitsum Assamnew

Advisor

Date

Signature

Dr. Bisrat Derebssa

Internal Examiner

Date

Signature

Dr. Sosina Mengistu

External Examiner

Date

Signature

Declaration

I, Mekdim Tessema Teklu, declare that this thesis is my original work. All sources of information in this study have been appropriately acknowledged. I further confirm that this thesis has not been submitted either in part or in full for any other requirements to any other learning institution.

Student Name: Mekdim Tessema Teklu

Signature: _____

Date: _____

June, 2024

Acknowledgments

I would like to express my deepest gratitude to God for providing me with guidance, strength, and wisdom throughout my life, especially during the course of this research. The divine presence has been a constant source of inspiration, and I am truly thankful for the blessings that have illuminated my journey.

I am also profoundly indebted to my family for their boundless love, unwavering support, and enduring patience. Their encouragement has been the cornerstone of my perseverance, and their understanding has been a pillar of strength during the challenges of this research endeavor.

In addition, I extend my sincere thanks to my advisor, Dr. Fitsum Assamnew, for his invaluable guidance and expertise, as well as to Addis Ababa University for providing the necessary resources and facilities.

Lastly, my heartfelt thanks to my friends for their camaraderie and support.

This research journey has been enriched by the blessings of God, the love of my family, and the support of many individuals and institutions. I am truly grateful for the collective contributions that have shaped this work.

Abstract

With the growth of internet accessibility social media users increased rapidly in Ethiopia. This created an easy ground for transmission of information between people. On the flip side it became a hub for fake news fabrication and propagation. Fake news that is available online has the potential to cause significant issues for both individuals and society as a whole. We propose a multimodal fake news detection for Amharic on social media that combines textual and visual features. Genuine and fake news data was collected from social media to create multimodal Amharic news dataset. The collected data was preprocessed to retrieve textual and visual features using Bidirectional Long Short Term Memory (BiLSTM) and Conventional Neural Network (CNN) respectively. Then the two sets of features were concatenated and were used to train our multimodal fake news detection model. Our proposed method achieved a 90% accuracy, 94% Precision. Compared to the state of the art unimodal fake news detection for Amharic, our proposed model achieved 4% accuracy and 7% precision increase in fake news detection performance.

Keywords: Amharic Multimodal Fake News; Unimodal; CNN-BiLSTM

Table of Contents

Declaration	i
Abstract	iii
List of Figures	vii
List of Tables	viii
List of Acronyms	ix
Chapter 1	1
1 Introduction	1
1.1 Statement of the problem	3
1.2 Research Questions	4
1.3 Objectives	4
1.3.1 General Objective	4
1.3.2 Specific Objectives	4
1.4 Contribution	5
1.5 Scope and Limitation	5
1.6 Organization of the study	6
Chapter 2	7
2 Theoretical Background	7
2.1 Fake News	7
2.2 Significance of fake news in Contemporary Society:	8
2.3 Fake News Detection Techniques	10
2.3.1 Existing Technologies and Tools for Fake News Detection	11
2.4 Feature Extraction Methods for the text	13
2.4.1 Bag of Words(BoW)	13
2.4.2 Term Frequency-Inverse Document Frequency(TF-IDF)	14
2.4.3 Word Embeddings (Word2Vec, GloVe, FastText)	14
2.4.4 N-grams	14
2.4.5 Word Frequency-Based Features	14
2.5 Deep Learning Approaches	14

2.5.1	Neural Network(NN)	15
2.5.1.1	Recurrent Neural Network(RNN)	17
2.5.1.2	Feedforward Neural Network(FNN)	18
2.5.1.3	Convolutional Neural Networks (CNN)	19
2.6	Summary	19
Chapter 3		20
3 Literature Review		20
3.1	Feature-Based Fake News Detection Methods	20
3.1.1	Context-Based	20
3.1.2	Content-Based	20
3.2	Related work on Fake News Detection	21
Chapter 4		28
4 Methodology		28
4.1	Dataset Preparation	29
4.1.1	Data Collection	30
4.1.2	Dataset cleaning and filtering	32
4.1.3	Data Annotation	33
4.2	Preprocessing	34
4.2.1	Text preprocessing	34
4.2.2	Image preprocessing	36
4.3	Feature Extraction	36
4.3.1	Word Embedding	37
4.3.1.1	Word2vec	37
4.3.1.2	Fasttext	38
4.3.2	Image Embedding	38
4.4	Proposed models	39
4.4.1	Unimodal Fake News Detection Models	39
4.4.1.1	RNN	39
4.4.1.2	CNN	41
4.4.1.3	Machine Learning Models	43
4.5	Performance Evaluation Metrics	43
4.6	Fake News Detection Model Implementation	45
4.6.1	Preprocessing	45
4.6.2	Feature Extraction	46

4.6.2.1	Word Embedding	46
4.6.2.2	Image Embedding	47
4.6.3	Fake news detection model Implementation	47
Chapter 5		49
5 Result and Discussion		49
5.1	Experiment Setup	49
5.2	Model Performance Metrics	50
5.2.1	Unimodal fake news detection models performance	50
5.2.1.1	Text Only Models	51
5.2.1.2	Image Only Model	57
5.2.2	Multimodal Fake News Detection	59
5.3	Discussion	63
5.4	Summary	64
Chapter 6		66
6 Conclusion and Recommendation		66
	References	67

List of Figures

2.1	Neural Network	15
2.2	Recurrent Neural Network	17
2.3	Feedforward Neural Network	18
2.4	Convolutional Neural Network — CNN architecture	19
4.1	Proposed system flow diagram	28
4.2	Proposed system flow diagram of training and testing	29
4.3	Dataset Preparation steps	30
4.4	Multimodal Dataset	33
4.5	Dataset Preprocessing Steps	34
4.6	List of stop words	35
4.7	Tokenization on Amharic sentence	36
4.8	Word2vec Model	37
4.9	Fasttext on Amharic word	38
4.10	BiLSTM architecture	40
4.11	CNN architecture	42
4.12	Confusion Matrix	44
4.13	CNN-BiLSTM Architecture	48
5.1	Performance Comparison of Unimodal on Word2vec	53
5.2	Performance Comparison of Unimodal Models on Fasttext	56
5.3	Fasttext vs Word2vec	56
5.4	Confusion matrix for CNN Model	57
5.5	Image only Model Performance	58
5.6	Text only Model vs Image only Model Comparison	59
5.7	Multimodal CNN Model Performance	60
5.8	Multimodal CNN-BiLSTM Model Performance	61
5.9	CNN vs Proposed model	62
5.10	Performance Comparison	62

List of Tables

4.1	Facebook pages for potential real news	31
4.2	Facebook pages that are source of fake news	32
5.1	Performance Comparison Of Unimodal Using Word2vec	51
5.2	Performance Comparision Of Unimodal Models Using Fasttext	53
5.3	Summary of Performance Comparison of the Models	63

List of Acronyms

USC	University of Southern California
CNN	Convolutional Neural Network
BiLSTM	Bidirectional Long Short-Term Memory
AFP	Agence France Presse
NLP	Natural Language Processing
SVM	Support Vector Machine
RF	Random Forest
NB	Naïve Bayes
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
BoW	Bag of Words
TF-IDF	Term Frequency-Inverse Document Frequency
TF	Term Frequency
NN	Neural Networks
ReLU	Rectified Linear Unit
BRNNs	Bidirectional Recurrent Neural Networks
GRU	Gated Recurrent Unit
MLP	Multilayer Perceptron
DL	Deep learning
NLPA	Natural Language Preprocessing Architecture
CN	Capsule Networks
PIL	Python Imaging Library

Chapter 1

Introduction

News is newly received or noteworthy information, especially about recent events[1]. There are different ways to get news, including through conventional media like newspapers, radio, and television as well as through contemporary digital platforms like websites, social media, and podcasts. There has been a noticeable shift in people's behavior as a result of the internet's extraordinary ability to provide information readily[2]. The advent of smartphones with their ubiquitous availability and as instruments for easily accessing the internet is making people turning more and more toward their mobile gadgets as a means to access news. As a result the manner how people consume news has undergone a considerable alteration. A growing proportion of people utilize social media platforms as their main way of information access rather than sticking with traditional news sources[3].

The widespread use of social media as a key source for news consumption in the modern world is becoming even more evident[4]. This change can be assigned to how convenient it is to use mobile devices. However, this ease has also brought in a significant rise in the prevalence of false information. There is an alarming lack of control over the spread of false information as a result of the ease with which news can be posted on social media platforms and the possibility of reaching vast populations.

Individuals are now free to widely transmit news without abiding by professional or ethical journalistic standards due to the lack of strict accountability procedures in this field. Furthermore, the absence of pre-publication control makes it much more difficult to combat fake news[5].

Fake news is information that is clearly and demonstrably fabricated and that has been packaged and distributed to appear as legitimate news [6]. Although the problem of false information is not new, it attracted more attention in 2017[7]. In the past, we relied on reliable news sources like credible media outlets and journalists that were bound by rigorous rules and upheld high ethical standards. The internet, however, opens up an entirely different outlet for the creation, distribution, and consumption of information. This frequently happens with minimal oversight and without clear editing standards.

There has been a noticeable shift, and a sizeable section of the public now turns to social media networks and platforms as their main news sources. The difficulty in determining the accuracy of news reports has emerged as a significant barrier as a result of this transition. The overwhelming amount of information available and widespread ignorance of the nuances of the internet only serve to worsen this problem. As a result, these elements have considerably aided in the widespread transmission of false information and manufactured stories.

What motivates these users to fabricate and share false information, wondered the research team made up of people from the University of Southern California (USC) Marshall School of Business and USC Dornsife College of Letters, Arts, and Sciences. Surprisingly, social networking platforms function with a rewards system that encourages users to keep their accounts active by regularly uploading and sharing information, much like the mechanics of games. Users that post and share frequently, especially with sensational and attention-grabbing content, tend to gain more visibility and interaction [8].

Users cultivate habits of sharing news that receives recognition from others because of the reward-based learning systems on social media, according to the researchers. “Once habits take hold, cues on the platform automatically activate information sharing without users taking critical response outcomes, like the spread of misinformation, into account”. It is therefore possible to develop a habit of posting, sharing, and interacting with people on social media. These findings imply that social media platforms can prevent the propagation of false information by pursuing fundamental changes to their reward structure rather than simply controlling what information is uploaded.

The intent of fake news is to deliberately defame people, cause division, and stoke hostility. Ethiopia is currently experiencing the real effects of hate speech and the spread of false information[9]. The negative consequences spread across society and force the closure of educational institutions like colleges, universities, and schools, both public and private. In addition, crucial transit routes used to be closed leading to disruption of key supply chains. The restriction of citizens’ freedom of movement causes widespread displacement and the tragic loss of many lives due to the shortage of food and shelter.

Ethiopians are experiencing social media’s negative influences more severely than people in developing nations [8]. It has been established that protecting lives is inextricably linked to the battle against false information and hate speech [10]. The spread of false information, hate speech, and misinformation has become pervasive in Ethiopia’s media landscape, particularly on internet platforms. This trend has had serious negative effects in the real world, increasing existing tensions and fueling incidents of violence and conflict.

To this point, the Ethiopian government has taken a strict stance against the spread of fake news, misinformation, and hate speech, frequently resorting to carrying out internet shutdowns as a reaction to tense situations. However, the constant expansion of social media platforms like Twitter, YouTube, and Facebook as well as the internet has given rise to new opportunities and obstacles in finding effective solutions. Nonetheless, the method employed for the detection and mitigation of fake news and hate speech on social media is with its limitations, as it is yet to achieve flawless accuracy.

To stop the spread of false information in Amharic, multiple initiatives have been launched. While Agence France Presse (AFP) has adopted fact-checking methods, though mostly for English content, Facebook has expanded its fact-checking program to Ethiopia[11]. Many journalists have also contributed to the effort by conducting fact-checking investigations and looking for false information.

The government has also shown its dedication to tackle this issue by passing regulations and binding laws designed to lessen the impact of false information. These initiatives demonstrate the broad acceptance of the issue and the concerted efforts made by various stakeholders to stop the transmission of false information within the Amharic-speaking community.

1.1 Statement of the problem

With the rise in fake news on social media, utilizing advanced technologies for detection has become essential, as manual identification is both challenging and time-consuming. Many researchers tried to detect fake news using linguistic features [12, 13, 14, 15]. These approaches can be considered as unimodal methods because they only use one type of input data to deal with the task. However, modern news has transitioned from being solely text-based to multimodal news that includes images and videos. Keeping with the trend fake news also incorporates multimodal aspects of news. Researchers have tried to address this type of news for the English language [16, 17, 18, 19, 20, 21].

To the best of our knowledge, no research has yet addressed multimodal fake news detection for the Amharic language. Also, we could not find existing Multimodal dataset for detection of fake news that exist for the Amharic language. As stated by Jawaher et al.[22] low resource languages face limitations due to two factors: sequential length restrictions in pre-trained language models, and the presence of noisy training data.

This research endeavor seeks to address this gap by training feature extraction methods on Amharic corpus with variable dimensions. The dataset issue will be addressed by collecting and preparing a multimodal fake news detection dataset from social media for the Amharic language. We propose to build a multimodal Amharic fake news detection deep learning model using a hybrid of Convolutional Neural Network (CNN)-Bidirectional Long Short-Term Memory (BiLSTM).

1.2 Research Questions

The research questions below are what this study aims to address.

RQ1 What is the performance of the proposed multimodal approach in detecting Amharic fake news compared to the existing unimodal and multimodal approaches?

RQ2 What is the most effective feature extraction method for detecting fake news in Amharic text datasets?

1.3 Objectives

1.3.1 General Objective

This research aims to detect Multimodal Amharic fake news on social media using a hybrid deep learning approach that's CNN-BiLSTM.

1.3.2 Specific Objectives

The research paper has the following specific objectives:

- To Collect multimodal Amharic news from social media and build a dataset.
- To build a hybrid deep learning approach that's CNN-BiLSTM and training it on a given multimodal dataset.
- To compare the fake news detection performance of our proposed approach with the existing unimodal methods and multimodal methods.

- To evaluate the effect of the addition of Amharic feature extractor in the detection of Amharic fake news.

1.4 Contribution

The main contributions of this project for detection of fake news are:

- An Amharic multimodal news dataset, encompassing both textual and visual elements, has been developed to address the issue of fake news.
- A hybrid deep learning model is built and trained to detect Amharic fake news.
- An evaluation of the performance improvements achieved by incorporating visual elements alongside textual content in fake news detection.
- The proposed model has been compared with existing state-of-the-art models to benchmark its effectiveness and performance.

A novel hybrid deep learning model has been designed to effectively detect fake news in the Amharic language. The project evaluates the performance improvements achieved by incorporating visual elements alongside textual content in fake news detection. The proposed model has been rigorously compared with existing state-of-the-art models to benchmark its effectiveness and performance.

1.5 Scope and Limitation

The scope of this research is to detect Amharic fake news on social media. The dataset for identifying fake news was constructed using content scraped from multiple social media sites, and Amharic is the study's chosen language. The goal is to gather a collection of 4000 news articles that includes both textual and visual components.

An additional noteworthy constraint is the extent of the gathered information, which is derived just from a restricted set of social media networks. This restriction suggests that the data-set might not adequately represent the variety and subtleties seen on different online platforms, which could restrict how broadly the study's conclusions can be applied.

Although the study was only able to gather 4,000 news stories, the dataset's size is likewise limited. A bigger and more varied dataset could improve the statistical robustness and universality of the study's findings, even though this dataset might still provide insightful information.

1.6 Organization of the study

The rest of this document is organized as follows. The second chapter discusses definitions, types, theoretical backgrounds related to fake news and its context in Ethiopia, the detection mechanism, features extraction techniques and news classification algorithms used. In addition to that previous research works related to fake news detection and state of the art approaches for fake news detection are also discussed. Our proposed approach for fake news detection is elaborated in Chapter three. The experimental setups, procedures, evaluation metrics, results, and discussion are discussed in Chapter four. The final chapter discusses the conclusion and future research direction on Amharic fake news detection.

Chapter 2

Theoretical Background

2.1 Fake News

Fake News is defined as “false stories that are created and spread to influence public opinion and appear to be true”[23]. This problem has been in society for a long time even before the invention of the internet and social media. People used to fabricate fake news and propagate it for their political or economical gain. However, the effect of it was not visible since it had a very slight chance of reaching the mass. With the growth of the internet and social media, the communication medium through people completely changed and most people use their phone as a source of information and news. This has created an auspicious opportunity for those creating fake news to get their news out there and easily reach millions of people in an instant. This issue has become very critical in our country over the past few years.

Ethiopia has been affected by fake news, unlike other countries[24]. Fake news has been the source of war and loss of people’s lives. Masses were affected by it directly and indirectly throughout the years. Because of the reasons stated above combating fake news on social media is a very important issue. Fake news refers to news stories that are false or nonfactual: the story is manipulated, with no authenticated facts, sources, or quotes[13]. Most of the time it’s intended to deceive people and it is typically created for financial or political gain. Analysis of each fake news story’s sources identified social media as responsible for 73% of the fake stories. Of this, around 80% appeared to originate on Facebook. No examples of fake news appeared to originate from traditional print media (or conventional print media websites) from the sample available[25]. Social media provide a platform for intentional propaganda and trolling. Propaganda, misinformation, disinformation, rumors, hoaxes, and clickbait are some of the terms that are used interchangeably[23]. The terms are discussed below:

1. Propaganda: It is a form of news articles and stories created and disseminated by political parties to shape public opinion.

2. Misinformation: It is purposely crafted erroneous information that is broadcast intentionally or accidentally, without regard for the true intent.
3. Disinformation: It refers to a piece of misleading information that is circulated to distort facts and deceive the intended audience.
4. Rumors and hoaxes: These terms are often used interchangeably to refer to the purposeful fabrication of evidence that is intended to appear legitimate. They publish unconfirmed and false allegations as true claims that are validated by established news outlets.
5. Clickbait: Clickbait headlines are frequently used to attract readers' attention and encourage them to click, redirecting the reader to a different site. More advertisement clicks equal more money.

2.2 Significance of fake news in Contemporary Society:

1. Erosion of trust on media: As false news becomes more common, people's faith in established media outlets may be damaged since they start to doubt the accuracy and authenticity of the news they read. When fake news is unrestrained, legitimate journalism finds it difficult to remain credible.
2. Impact on Public Perception: People may come to believe things based on inaccurate or skewed information as a result of fake news' ability to propagate misinformation. It has the power to mold people's perceptions of certain events, problems, and people in general.
3. Political and Social Consequences: By promoting preexisting ideas and fabricating conflicting narratives, fake news can aggravate polarization in politics. In severe circumstances, the spread of false information may be a factor in political and social discontent.
4. Manipulation of Public Discourse: People who propagate false information may be doing it with a specific goal in mind, utilizing false information to sway public opinion and legislative choices. By spreading misleading information about candidates or issues, fake news has been linked to attempts to sway political outcomes.
5. Economic Impacts: A few phony news articles aim to publicize financial frauds that take advantage of people for financial benefit. Inaccurate information regarding businesses or economic indicators may cause financial losses and market volatility.

6. Challenges for Information Consumers: The ubiquity of false information highlights the significance of media literacy, since people use critical thinking abilities to separate reliable information from false information. Confirmation bias may be strengthened by consumers' susceptibility to bogus news that supports their pre-existing opinions.
7. Technological Challenges: Fake news can spread quickly due to the popularity of social media platforms. Online platform algorithms may unintentionally encourage the dissemination of false information by raising its visibility.

A complex strategy that includes media literacy instruction, ethical journalism, technical advancements, and public awareness campaigns is needed to address the problems caused by fake news. Understanding the role that fake news plays in modern society is crucial to creating solutions that lessen its effects and encourage the people to become more knowledgeable and resilient. The false narratives disseminated as fake news strategically employ attention-grabbing semantics in their headlines, enticing readers with captivating language. This is often coupled with the inclusion of eye-catching phrases, enhancing the allure of these stories. As a significant number of individuals tend to unquestionably embrace such news items, consuming and subsequently sharing them across social media platforms, the rapid diffusion of misleading information to a vast audience becomes an alarmingly straightforward process. Numerous studies tried to solve this issue by using new technologies such as Natural Language processing. Linguistic resources are vital in the creation of fake news and hate speech detection approaches. However, "low-resource" languages, primarily African languages, lack such tools and resources. These languages remain among the world's "low resource" languages, lacking the tools and resources required for natural language processing applications and other techno-linguistic activities[8, 26].

2.3 Fake News Detection Techniques

The rapid growth and expansion of social media platform has filled the gap in information exchange in day-to-day life. Social media is the main arena for disseminating manipulated information in a high range and exponential rate[13]. Ethiopia like other countries is influenced by the growth and advancement of social media. There is a vibrant online news ecosystem that covers Ethiopia. Although there are estimated to be only 21.14 million internet users in Ethiopia (January 2020), many individuals and organizations can boast of well over a million followers on Facebook alone[8]. Any content can be posted online and instantly reach millions of people. That's one of the reasons why people misuse social media and post fake news.

The spread of fake news within social media landscapes possesses the potential to trigger inter-ethnic and inter-religious conflicts in Ethiopia. Consequently, the ramifications of fake news reverberate across diverse sectors, impacting both citizens and the country's governmental institutions and commercial entities.

Historically, the impact of fake news on the nation's well-being has been so dire that the government resorted to censoring news outlets and imposing nationwide internet shut-downs. Even though formal internet access rates remain relatively low, a substantial number of individuals utilize mobile phones to connect online, often relying on platforms like Facebook and other social media channels for news consumption. Consequently, enforcing internet shutdowns disproportionately targets those who disseminate false information, hate speech, and misinformation.

Such actions not only severely curtail people's access to information and their capacity to communicate and report but also cast a shadow over economic activities[8]. In light of these considerations, the urgency of fake news detection cannot be underestimated; it demands immediate attention and action rather than being deferred to a later time.

The headlines of this fake news are primarily semantic to pique the reader's interest, and eye-catching words are added to make it more interesting. Because the majority of people blindly believe the news, they read on social media without examining the facts and sharing it, it has become effortless to quickly spread misleading information to a huge audience. Fake news on social media is a potential cause to spark ethnic and religious clashes in Ethiopia[27]. As a result, people, governments, and companies were affected by fake news in the country.

In the past, fake news had such a detrimental effect on the country that the government was forced to censor the news and shut down internet access nationwide. Although theoretical internet access remains low, many people have access to the internet through mobile phones, and many get their news from Facebook and other social media platforms. Therefore, shutting down the internet denies those that spread 'fake news,' hate speech, and misinformation to their audience, severely restricts people's freedom of information and ability to communicate and report, and affects economic life[8]. Because of the aforementioned factors, detecting fake news is crucial and can't stay for tomorrow.

Amharic has its ancient script and is a very powerful language. The working language of courts, the military, trade, and everyday communications since the late 12th century, in addition to its early existence and uses, is still the official language of the Ethiopian government today. Most of the Ethiopian Jewish communities in Ethiopia and Israel speak Amharic. In Washington DC, Amharic became one of the six non-English languages in the Language Access Act of 2004, which allows government services and education in Amharic [28]. Despite Amharic being highly powerful, it is still one of the "low-resource" languages in the world. A lack of sufficient datasets and linguistic tools to help use technology makes the language disadvantaged in this regard.

Researchers tried to detect fake news by applying Natural Language Processing (NLP) and different machine learning approaches in their research to detect Amharic fake news. The earliest approaches[29, 12, 15] used various machine learning techniques like Support Vector Machine (SVM), Random Forest (RF), and Naïve Bayes (NB) but with the increase in the amount of data on social media platforms a shift to deep learning approaches [30, 31, 32, 16, 17, 33] can be seen which includes the use of CNN, Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM) based approaches.

2.3.1 Existing Technologies and Tools for Fake News Detection

- **Natural Language Processing (NLP) Algorithms:** NLP approaches look for anomalies or patterns suggestive of fake news by analyzing linguistic factors including text patterns and language structure. It has been demonstrated that NLP algorithms are capable of identifying linguistic clues linked to false data. They are able to evaluate the text's lexical choices, tone, and sentiment. NLP models might have trouble picking up on irony, context, and linguistic nuances.

- **Machine Learning Models:** Utilizing features obtained from the content, news stories can be classified using a variety of machine learning techniques, including supervised and unsupervised learning. When trained on a variety of data-sets, machine learning models have shown promise in distinguishing between fake and authentic news. But they might find it difficult to adjust to the fast changing varieties of fake news, and they might be susceptible to biases in training data. The following are a few machine-learning approaches:

1. Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for both classification and regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is the number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well.
2. Random Forest (RF) is one of the many machine learning algorithms used for supervised learning, this means learning from labeled data and making predictions based on the learned patterns. RF can be used for both classification and regression tasks.
3. Naive Bayes (NB) classifier is a probabilistic machine learning model that's used for classification tasks. The crux of the classifier is based on the Bayes theorem. Bayes Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.1)$$

Using Bayes theorem, we can find the probability of A happening, given that B has occurred. Here, B is the evidence and A is the hypothesis. The assumption made here is that the predictors/features are independent. That is presence of one particular feature does not affect the other. Hence it is called naive.

- **Social Network Analysis:** This method focuses on examining the dynamics and structure of social networks to find trends in the spread of information and the reliability of sources. Social network analysis may identify influential nodes and show how bogus news spreads. Nonetheless it might not be able to identify nuanced and context-specific false information. It also depends on having full access to social network data.

- **Fact-Checking Platforms:** Fact-checking organizations thoroughly review news reports to determine their correctness before labeling them as true or false. Fact-checking is a dependable technique that offers human judgment in determining the veracity of material. The limitations of Fact-checking is it requires a lot of time and resources, and it might not be able to keep up with the speed at which false information is spreading on social media.

To summarize, current technologies and tools for identifying false news use a variety of techniques, including fact-checking, multimedia analysis, linguistic analysis, machine learning, and social network insights. Even though these technologies have proven useful, there are still issues to be resolved, such as dealing with biases, adjusting to new types of fake news, and striking a balance between accuracy and speed trade-offs. Comprehensive approaches and interdisciplinary collaboration are necessary for effective false news detection tactics.

2.4 Feature Extraction Methods for the text

Feature extraction methods for text are techniques that transform raw text data into numerical representations (features) that can be used as input for machine learning models or other text analysis tasks. These features capture essential information from the text, allowing algorithms to learn patterns and make predictions. Here are some common feature extraction methods for text:

2.4.1 Bag of Words(BoW)

Bag of Words (BoW) represents a document as a vector of word frequencies. Each unique word in the entire corpus is treated as a feature, and the value of each feature is the count of the corresponding word in the document. BoW is a simple and interpretable representation but doesn't capture word order or context.

2.4.2 Term Frequency-Inverse Document Frequency(TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) represents the importance of a word within a document relative to its importance across the entire corpus. It combines Term Frequency (TF), which measures word frequency in the document, and inverse document frequency (IDF), which measures how unique a word is across documents. TF-IDF is useful for capturing important keywords while down weighting common words.

2.4.3 Word Embeddings (Word2Vec, GloVe, FastText)

Word embeddings represent words as dense vectors in a continuous space. These vectors are learned from large corpora of text data and capture semantic relationships between words. Pre-trained word embeddings can be used as features or fine-tuned for specific tasks. They are useful for capturing word context and semantics.

2.4.4 N-grams

N-grams are sequences of adjacent words or characters in a document. For example, unigrams (single words), bigrams (two-word sequences), and trigrams (three-word sequences). N-grams capture local word order and can be used to represent text for tasks like text classification or sentiment analysis.

2.4.5 Word Frequency-Based Features

Features based on word statistics, such as word count, character count, average word length, and unique word count.

2.5 Deep Learning Approaches

Advanced neural network architectures are used in deep learning models for fake news detection to automatically recognize and evaluate complex patterns in textual content with the goal of identifying characteristics suggestive of disinformation[34].

By utilizing many levels of abstraction, these models are able to identify subtleties in language, structure, and context, which allows them foresee the reliability of news reports. The main difference is that the model can automatically extract and process features without the need for explicit feature engineering. This enables a more sophisticated and data-driven method of differentiating between real and fake news.

The following are some important deep learning models and methods for detecting fake news:

2.5.1 Neural Network(NN)

A neural network is a type of computational model that derives inspiration from the composition and operations of the human brain emulating the behavioral patterns of actual neurons. An input layer, one or more hidden layers, and an output layer make up a node layer in Neural Networks (NN). Every node, or artificial neuron, has a weight and threshold that are connected to one another. A node is activated and sends data to the following layer of the network if its output exceeds a given threshold value. If not, no data is transferred to the network's subsequent layer.

Machine learning relies heavily on neural networks, which are applied to a variety of tasks such as regression analysis, pattern recognition, classification, and decision-making. The

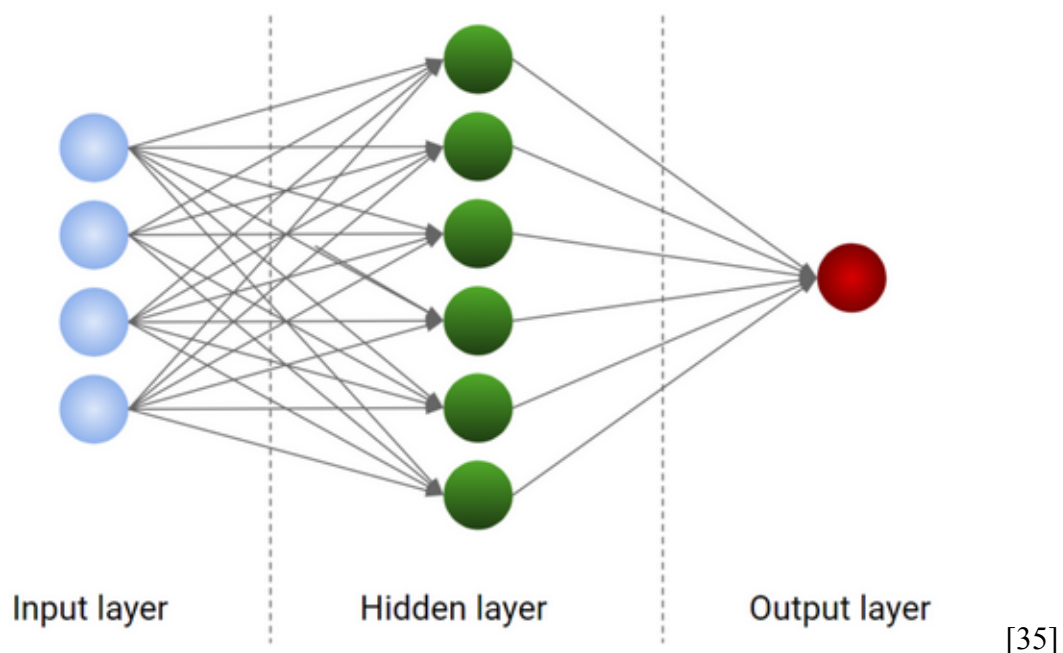


Figure 2.1: Neural Network

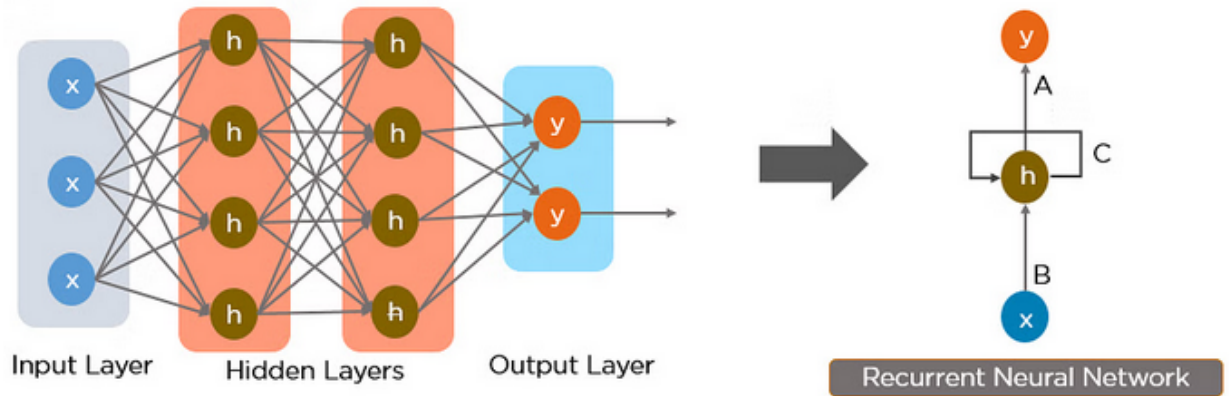
following are the key components and concept underlie neural networks:

- **Neuron:** Neural networks use neurons as fundamental computational units. Every neuron takes in one or more inputs, applies a mathematical operation to them, and outputs something.
- **Layers:** The layers that make up a neural network usually include an input layer, one or more hidden layers, and an output layer. The network's final predictions are generated by the output layer, whereas the input layer represents the features of the incoming data.
- **Weights and Biases:** Every neural connection has a weight attached to it which indicates how strong the connection is. Furthermore, every neuron has a corresponding bias, which enables the network to capture more intricate interactions.
- **Activation Function:** In order to add non-linearity to the model, neurons apply an activation function to the weighted sum of their inputs. The sigmoid, Rectified Linear Unit (ReLU), and hyperbolic tangent (tanh) are examples of common activation functions.
- **Feedforward:** Information passes without cycles or loops from the input layer to the output layer of a feedforward neural network through the hidden layers. It is through this procedure that predictions are made.
- **Backpropagation:** Neural network learning algorithms use backpropagation. It entails modifying the network's weights and biases in response to variations between the expected and actual outputs. The model performs better and the error is reduced by this iterative procedure.
- **Loss Function:** The gap between the expected and actual outputs is measured by the loss function. In order to direct the network to provide precise predictions, the objective of training is to reduce this loss.
- **Training Data:** Labeled training data, in which the inputs are connected to known outputs, is the source of knowledge for neural networks. In order to reduce the gap between the expected and actual outputs on the training set, the network's parameters are changed during the training process.

Neural networks come in a variety of forms, each intended for a particular purpose, such as:

2.5.1.1 Recurrent Neural Network(RNN)

Recurrent neural networks are a kind of artificial neural network intended for processing data sequences. They do particularly well in tasks involving sequences, including voice, natural language, time series data, and other applications. Their competency to use in-



[36]

Figure 2.2: Recurrent Neural Network

formation from previous inputs to influence the present input and output sets them apart. Recurrent neural networks rely on the previous parts in the sequence to determine their output, in contrast to typical deep neural networks, which presume that inputs and outputs are independent of one another. Although the outcome of a given sequence could also be determined by future occurrences, unidirectional recurrent neural networks are unable to include these events into their predictions.

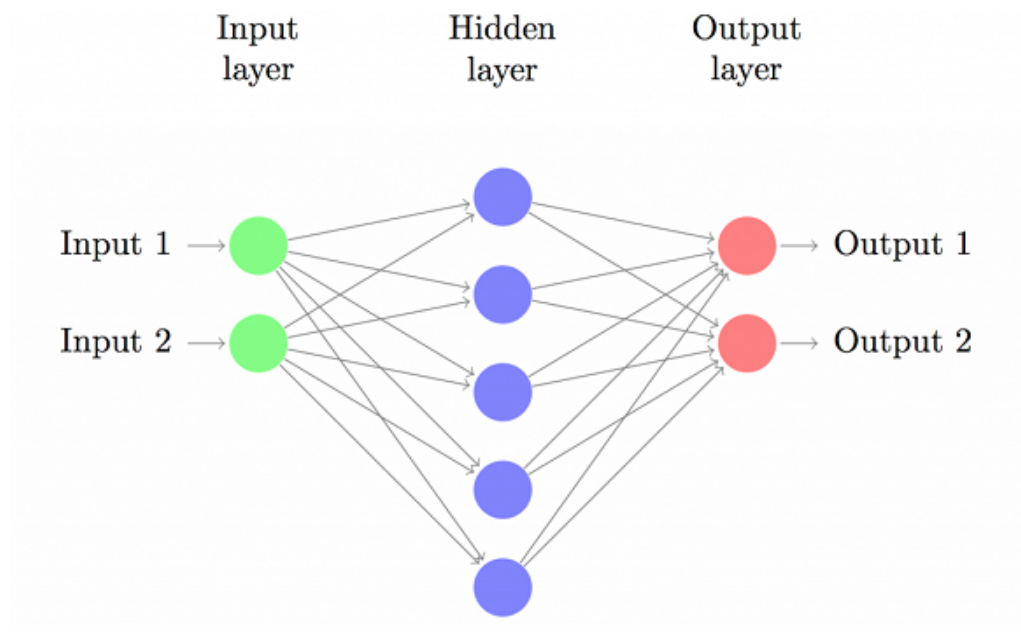
RNN come in a variety of forms, each with unique architectures or adjustments to meet particular needs. These are a few varieties of RNNs:

1. **Bidirectional Recurrent Neural Networks (BRNNs):** These are an alternative RNN network architecture that process the incoming data both forward and backward. This improves the network's comprehension of the sequential dependencies by enabling it to gather data from the context of each time step, both past and future. Bidirectional RNNs incorporate future data to increase the accuracy of their predictions, whereas unidirectional RNNs can only draw from past inputs to do so.
2. **Long Short-Term Memory(LSTM):** is an RNN capable of learning long-term dependencies which is designed to address the issue of the vanishing gradient that arises in conventional RNNs.

3. **Gated Recurrent Unit (GRU):** GRUs solve the vanishing gradient issue in a manner similar to LSTMs. Unlike LSTMs, they have a more straightforward structure and employ a gating mechanism to regulate the information flow between the previous concealed state and the current time step.

2.5.1.2 Feedforward Neural Network(FNN)

Information passes from the input layer through one or more hidden layers and ends up at the output layer in a FNN, sometimes referred to as a Multilayer Perceptron (MLP). Classification, regression, and pattern recognition are just a few of the many applications for feedforward neural networks, which are the most basic type of artificial neural networks.



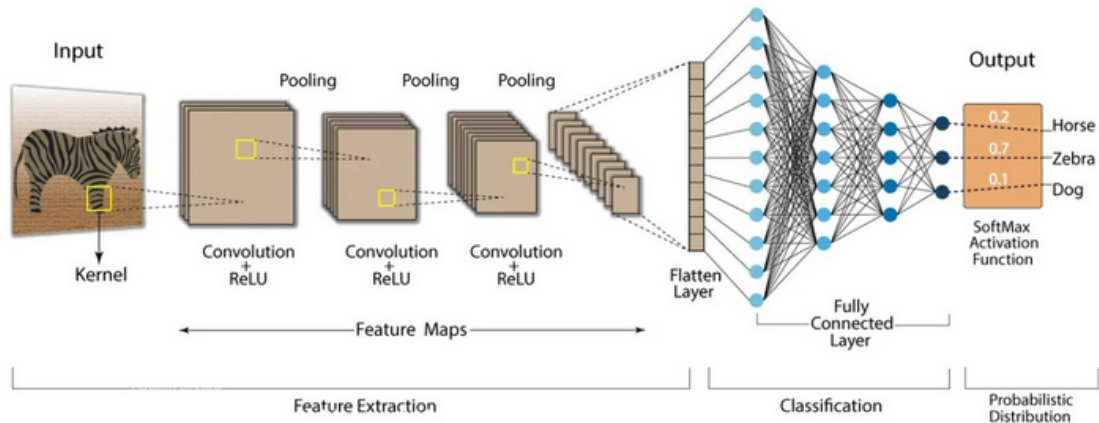
[37]

Figure 2.3: Feedforward Neural Network

Three distinct types of layers make up a feedforward neural network's architecture: input, hidden, and output layers. Neurons, which make up each layer, are connected by weights between the layers.

2.5.1.3 Convolutional Neural Networks (CNN)

Convolutional neural network is a form of artificial neural network intended for visual perception tasks like object detection, image recognition, and image classification. When dealing with grid-like data, where the spatial relationships between adjacent items are important, CNNs do especially well[38]. They have been adapted for other areas, such as natural language processing, and have shown great success in computer vision applications. An input image is first fed into a convolutional neural network, which uses a



[39]

Figure 2.4: Convolutional Neural Network — CNN architecture

sequence of convolutional and pooling layers to convert it into a feature map. The input image is subjected to a series of filters by the convolutional layer, each of which generates a feature map that emphasizes a distinct component of the input image. The feature map is then downsampled by the pooling layer to minimize its size while preserving the most significant information.

The convolutional layer generates a feature map, which is subsequently processed by several further convolutional and pooling layers, each of which learns ever more complex characteristics from the input image. Depending on the assignment, the network's ultimate output is a probability score or projected class label for each class.

2.6 Summary

We covered terms, classifications, theoretical underpinnings, related to fake news and it's context in Ethiopia, detection mechanisms, feature extraction strategies, and news classification algorithms in this chapter.

Chapter 3

Literature Review

In this chapter, we will discuss about the various feature-based fake news detection methods in the first section. The subsequent section of this chapter gives review on the different papers that were done by researchers on the topic of fake news detection.

3.1 Feature-Based Fake News Detection Methods

Feature based fake news detection methods rely on extracting and analyzing specific features from news articles, social media posts, or other textual information to distinguish between genuine and fake content. These features capture various aspects of the content, and machine learning algorithms are often employed to make predictions based on these features.

3.1.1 Context-Based

Context based fake news detection is a more sophisticated method of detecting and countering fake news, also known as contextual analysis or context-aware fake news detection. Context-based approaches evaluate the reliability of information by taking into account a wider range of factors and context, as opposed to content-based methods, which primarily focus on the text and content of news stories. Users, networks, or temporal data can all be used.

3.1.2 Content-Based

The term "content-based fake news detection" describes the method to identify if information provided in articles is purposefully fake, misleading, or both by analyzing its textual and visual content. Instead than depending on outside features or metadata, this method examines the news stories' actual content.

The information can be text-based, visual, or multimodal.

1. Text-Based: Text-based news is defined as news stories, articles, or information that mainly consists of text and does not heavily depend on multimedia or visual components. One of the conventional forms of news that readers are given in the context of journalism and media is text-based news.
2. Image-Based: the term "image-based news" describes news reporting in which information is primarily conveyed through the use of visual components including photos, infographics, charts, and other graphical representations. In contrast to conventional text-based news, image-based news primarily uses visual components to report on events or tell a story.
3. Multimodal-Based: in order to provide a more comprehensive and engaging storytelling experience, news content that incorporates several modes of communication—such as text, photographs, videos, audio, and interactive elements—is referred to as multimodal-based news. A multimodal approach combines many media channels to communicate information, leveraging each one's advantages to improve audience comprehension and engagement.

3.2 Related work on Fake News Detection

Researchers tried to detect Amharic fake news on social media and we will review the studies below.

Gereme et al.[30] built fake news detection model in the Amharic language based on news content and using Deep learning (DL) approaches. The research undertook a notable research endeavor encompassing various stages. Initially, they constructed an innovative Amharic word embedding, characterized by distinct word vectors. Additionally, they curated an Amharic news dataset sourced from social media platforms. Subsequently, leveraging their collected data, they devised a unique deep learning model tailored for detecting fake news in the Amharic language, yielding an impressive accuracy rate.

A noteworthy aspect of the study lies in the superiority demonstrated by their proprietary Amharic word embedding. This achievement can be attributed to the utilization of a self-generated corpus for training purposes.

However, a noticeable limitation pertains to the absence of a comparative analysis against existing Amharic fake news models. Furthermore, their research solely concentrated on a singular modality specifically, text-based Amharic fake news detection. It is noteworthy that their work remains unpublished online, precluding the opportunity for result validation through replication. Despite this constraint, Gereme et al.'s research constitutes a significant contribution to the domain of Amharic fake news detection.

Ibrahim N. [40] integrates a lexicon-based representation with a stance-based feature alongside the page score. The stance feature primarily encompasses the headline-to-headline cosine similarity, the headline-to-another-headline cosine similarity, and the page score that's calculated by reflecting the perspective of each page regarding fake news topics. The researcher meticulously collected a dataset of 3000 Amharic news articles from social media.

Through experimentation involving the combination of lexicon-based and stance-based features, remarkable advancements were achieved. The combined model exhibited a noteworthy enhancement of 4.1% accuracy, 3% precision, 4% recall, and a 4% increase in the F1 score in comparison to the lexicon-only model. Intriguingly, the experimentation also revealed that the addition or removal of stop words within news articles did not significantly influence the outcomes.

Ibrahim N.'s research underscores the effectiveness of integrating lexicon and stance-based components in the realm of Amharic fake news detection, showcasing a quantifiable boost in performance.

Gurmesssa D. [41] built Afan Oromo fake news detection system. The proposed approach includes preprocessing such as tokenization, normalization, stop word removal, and abbreviation resolution, feature extraction such as term frequency-inverted document frequency (TF-IDF), term frequency (TF), and hash to determine word importance in the news and the corpus, and N-grams, a powerful natural language processing technique for capturing semantic and syntactic information. With a passive-aggressive classification system, all conceivable combinations of feature extraction techniques and natural language processing approaches were applied.

According to the study, passive-aggressive (PA) outperforms ensemble methods like gradient boosting, random forest, and linear classifiers like MNB. The PA outperforms with 97.2% and an error of 2.8%. Finally, utilizing the TF-IDF feature extraction using Unigram and PA classification approaches, a Python Django was utilized for the web-based deployment of the model.

Despite the dataset's shortcomings, the linear PA with TF-IDF vector and unigram model outperforms the competition with 97.2% of precision, 97.9% of recall, and 97.5% receiver operating characteristic area under the ROC Curve (ROC AUC) f1-score.

According to [41], DL approaches have recently gained a lot of attention and have improved the state-of-the-art for many difficulties that artificial intelligence and machine learning approaches have faced for a long time. The goal of the research was to provide a method for detecting fake news on social media using the DL approach for Afan Oromo news text. A model to predict and classify Afan Oromo news text must be preprocessed and trained on the sample dataset. As a result, the researchers looked at one hot encoder for mapping category integers and used it in the context of word embedding by training it with Bi-LSTM and a cosine similarity measure, which are supplied as input features to the neural network (NN). After the classifier was trained to classify, a 0.5 threshold was applied to the output score to decide whether it was true or fake, and statistical analysis, a confusion matrix was used to compare across different thresholds, and the suggested model necessitated a large amount of data.

However, when compared to the dataset created for the English language, the dataset in the Afan Oromo language is a major concern; the model is trained on very minimal data. Boosting the consistency of the performance by adding data to the news dataset would increase user trust in the system. On a benchmark dataset, the model can predict with an accuracy of 90%, precision of 90%, recall of 89%, and an f1-score of 89%, outperforming the current state of the art applying the Bi-LSTM model. Finally, they concluded that the Bi-LSTM system prototype can be used as a foundation for future work with the Afan Oromo news text datasets and other Ethiopian local languages.

T. Tewodros et al. [13] were able to craft Amharic fake news dataset from verified news sources and various social media pages. To detect fake news six different machine learning classifiers Naïve bays, SVM, Logistic Regression, SGD, Random Forest and Passive aggressive Classifier model were built. They compared all the models and Naïve bays and Passive Aggressive Classifier surpass the remaining models with accuracy above 96% and F1- score of 99%. Social media platforms' quick growth and expansion have filled the information-sharing gap in everyday life.

The Amharic language fake news dataset was created using verified news sources and social media. The experimental results show a precision of 100% RF for both TF-IDF and Count Vectorizer (CV), a recall of 95% using the PA classifier for TF-IDF, and an f1-score of 100% in NB and LR classifier for TF-IDF vectorizer using PA classifier. The research has substantially contributed to slowing the spread of misinformation in vernacular languages.

Hailemichael et al. [10] applied deep learning approach to detect Amharic language fake news. This paper employed a newly acquired dataset to complete the research because there were no previously available resources in the area. They used the graph application programming interface (API) to collect data from the Facebook platform, and two journalists annotated the dataset. With an accuracy of 93.92%, a precision of 93%, recall of 95%, and an f1-score of 94%, the convolutional neural network (CNN) model outperforms all other models. The impact of morphological normalization on Amharic fake news identification was investigated using the top two performing models, and the results demonstrated that normalization harms classification performance, lowering both models' f1-score from 94 to 92%. Finally, CNN was shown to be the most effective model in the research paper.

Furthermore, contrary to their expectations, the attention mechanism used in the sequential models performs worse than the baseline model. Another finding of the study was that in the Amharic language fake news dataset, morphological normalization was not always helpful in improving model performance. According to this study, evaluating different approaches from other disciplines, such as Capsule Networks (CN), would be a good idea. The CN is doing better in the world of computer vision, and applying their strength to the Natural Language Preprocessing Architecture (NLPA) challenges could assist in improving the Amharic language fake news and hate speech detection model. Furthermore, they recommend that researchers interested in this field should have to train their embeddings with domain-specific data to obtain a more semantically strong embedding model, which could lead to better detection.

Most of the researches concerned on linguistic analysis of news content to identify its credibility,[14] used a feature fusion of linguistic and social context feature of the publisher information to detect Amharic fake news. The experimental result of feature fusion-based experiment shows at least 94.13% and at most 98.7% with a high relative error reduction over the content-based approaches.

The study only investigated posts and comments in textual documents. The posts and comments in the form of images or photos, audio, or video data have not been considered.

It's important to detect and acknowledge fake news before it spread to the masses. The multimodal aspect of the news article makes the content look much more credible than its counterparts [23]. Most of the existing studies that were done on Amharic fake news detection focus on text only content. However, some researchers tried to work on multimodal datasets that were prepared for fake news detection in English language.

Isabel et al. [16] in particular, multimodal approaches which combine text and visual data are emphasized in the research as a means of developing tools for automatic fake news detection.

A multimodal technique utilizing a Convolutional Neural Network (CNN) architecture produces most accurate results, with an accuracy of 87%, according to the experiments done on the Fakeddit dataset. While a unimodal approach using only text obtains an accuracy of 78% using the Bidirectional Encoder Representations from Transformers (BERT) model, the inclusion of visuals increases the detection of certain kinds of fake news. With an emphasis on the fine-grained categorization of several categories of fake news, the research examines both unimodal and multimodal approaches for the classification of fake news.

The research does not, however, disclose exact recall, precision, and F1 scores for each class or extensive details on how each model performed for various fake news categories. The particular implementation details of the models utilized, such as hyperparameters or training techniques, are not covered in detail in the paper. Because the results for each model and class are not fully broken down in the study, it is difficult to assess how well each model performs for different types of fake news. The findings may not be as applicable as they may be because the research does not address the generalizability of the suggested models to other datasets or fields.

The generalizability of the findings to real-world circumstances may be impacted by any biases or restrictions in the Fakeddit dataset used for evaluation, which are not addressed in the research. The paper does not examine the models' interpretability or offer insights into the characteristics or trends that help identify fake news. S. Singhal et al. [17] introduced Spotfake a multi-modal framework for fake news detection. The paper discusses fake news on social media and proposes a multi-modal framework called SpotFake in order to identify fake news without requiring additional subtasks. SpotFake uses the textual and visual characteristics of an article, utilizing language models such as BERT for textual features and VGG-19 that has already been trained on the ImageNet dataset for visual features. The suggested method outperforms the current state-of-the-art on the Twitter and Weibo datasets by 3.27 and 6.83 percent, respectively.

The research highlights the necessity of a multimodal strategy to address this problem and highlights the limitations of depending solely on textual assertions for detecting deception. The research highlights the necessity of a multimodal strategy to address this problem and highlights the limitations of depending solely on textual assertions for detecting deception.

On the Twitter dataset, SpotFake surpasses MVAE by 12.17%, 3.27%, and 14.93%, 6.83%, respectively. The research does not address the SpotFake framework's scalability and computing efficiency, which could represent important aspects for real-world use. The study does not address the applicability of the suggested methodology to other languages or platforms outside Twitter and Weibo.

Yang et al. [33] proposed a model dubbed TI-CNN (Text and Image Information Based Convolutional Neural Network), which incorporates explicit and latent properties from both text and image data to accurately detect fake news. The explicit and latent characteristics are projected into a single feature space by the TI-CNN model, which is trained using both text and image data at the same time. Numerous tests performed on real-world fake news datasets show how well TI-CNN handles the false news identification challenge.

The challenge of detecting fake news in online social networks is discussed in the study. This problem has grown in importance as a result of the extensive propagation of false information

The research emphasizes how crucial it is to take into account both text and image information in fake news detection since fake news can spread quickly over social networks and cause many people to report the same false observations. The TI-CNN (Text and Image information based Convolutional Neural Network) model is suggested in the paper as a method for identifying false news. To extract latent features from both textual and visual data, it uses two simultaneous CNNs.

The text and photos utilized in the fake news are used to identify the explicit features. In order to create novel representations of texts and images, the TI-CNN model mixes explicit and latent features by projecting them into the same feature space. The text branch and the image branch are the two main branches of the model. Exact and latent features are extracted for each branch in order to make final predictions.

In order to identify hidden patterns in the text and images used in fake news, the model is trained using both text and image information concurrently. To assess the success of the TI-CNN model in addressing the fake news detection challenge, extensive experiments are carried out on real-world fake news datasets. When using both text and image data, the TI-CNN model surpasses numerous other baseline techniques for spotting fake news. The model combines latent data obtained through many convolutional layers with explicit features obtained from text, phrases, and images used in fake news. Numerous tests using false news datasets from the actual world show how well TI-CNN handles the fake news identification issue.

The model incorporating text and image information significantly beats models that just use image or text information, which is why the article emphasizes the value of taking both into account when detecting fake news. The findings imply that the TI-CNN model may successfully identify underlying patterns in the text and pictures used in fake news, resulting in increased detection precision. Depending on the unique properties of the false news datasets utilized in the tests, the TI-CNN model's efficiency may change.

The quality and variety of the explicit and latent features retrieved from the text and image data may have an impact on the model's performance. Further research may be required to determine the TI-CNN model's adaptability and generalizability to various languages, domains, and false news subtypes.

In general, different researchers worked to detect fake news on social medias using both machine learning and deep learning methods. With the growth in data on social medias the usage of machine learning approaches decreased. However, since deep learning approaches do an excellent job in learning large datasets and can be used for detection of fake news researchers changed their focus on applying them for their studies.

It was impossible to find research that was done on detecting multimodal fake news in Amharic language. People use images in addition to text to create fake news on social media since it will make their article more believable when it adds visual feature. So, detecting the multimodal news on social medias is critical. Due to unavailability of on-line labelled fake news dataset for Amharic makes it harder to conduct research and provide convenient solution. We built a fake news dataset and classification model from the scratch to combat this issue. This proposed paper incorporated different deep learning methods and chose the model that works best for our research.

Chapter 4

Methodology

The methods and strategies used to accomplish the research objectives are the primary subjects of this chapter. The acquisition, preparation, and annotation of data which are crucial for use by various deep learning models are covered in depth in the first section. The actual procedure for creating and refining these models, including data preparation, feature extraction methods, and model designs, is discussed in the next section. Finally, the chapter describes the method utilized to evaluate the models included in the study.

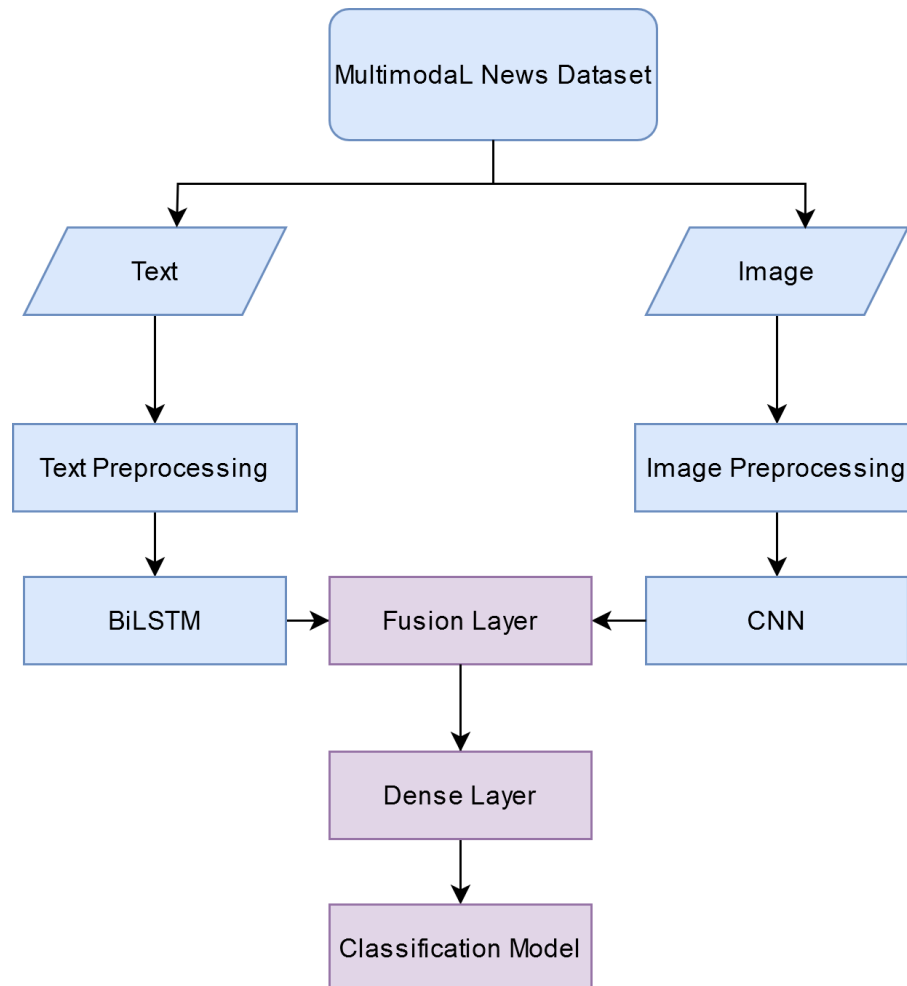


Figure 4.1: Proposed system flow diagram

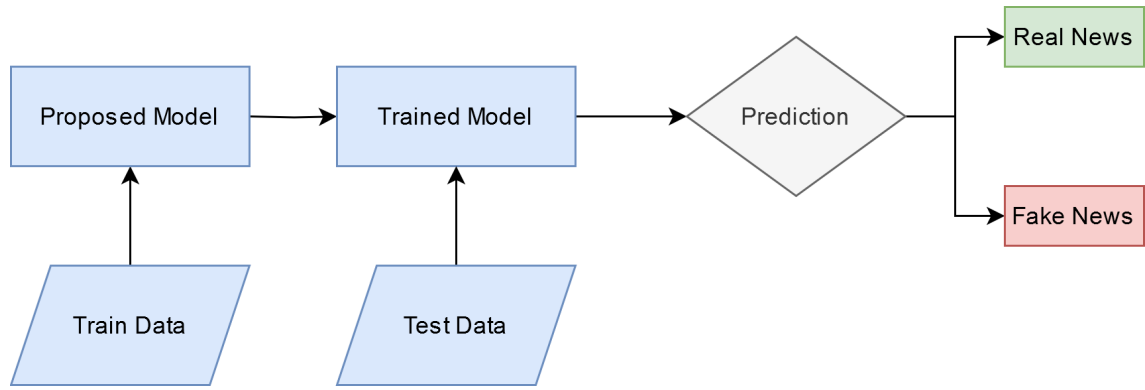


Figure 4.2: Proposed system flow diagram of training and testing

4.1 Dataset Preparation

Even though Amharic is the official language of Ethiopia, there are very few easily accessible news datasets, which presents a significant difficulty for researchers working in this field. Even with Amharic’s linguistic importance, there are still little datasets available for research, which limits the capacity to examine and evaluate news articles in the context of this official language. This scarcity could have an effect on the scope and depth of studies that look into many facets of news, language use, and communication in Ethiopian contexts. In order to promote thorough research and a deeper comprehension of news dynamics in the language and cultural context of Ethiopia, it is imperative that this gap in the available datasets be filled.

Due to the lack of readily available resources, we were compelled to deliberately collect and organize a dataset that was customized to our unique language and cultural environment. This allowed us to conduct a deeper analysis of news-related events in Ethiopia.

We went through a number of procedures in the dataset collection process. We started by choosing particular news outlets from which to compile relevant data. Then, we started the process of eliminating any unnecessary components or noises from the news content. Finally, we created a single dataset out of all the data we had gathered from different news sources.

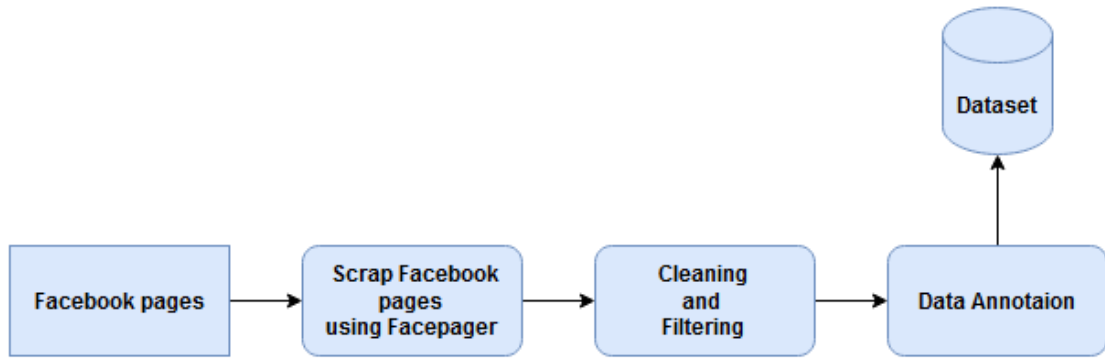


Figure 4.3: Dataset Preparation steps

4.1.1 Data Collection

Facebook is a notable social media platform that enables social networking, content sharing, and communication between people. A significant portion of the population in Ethiopia uses Facebook as their main information source. Hence, we decided to use it as the platform for gathering news data.

We used Facepager, a software application that made it easier to collect news stories from several Amharic Facebook pages, in our data collection procedure. Researchers can obtain content, comments, and other pertinent information from public Facebook pages by using Facepager, a data scraping and exploration tool that is built to work with the Facebook Graph API.

Before presenting the data to the professional journalists, we tackled the arduous task of sorting through numerous Facebook pages. Collecting data from all these pages was exhausting and time-consuming. To streamline our efforts and ensure efficiency, we focused on selecting pages that could potentially be sources of both real and fake news. This selective approach aimed to prioritize pages likely to have a substantial impact and relevance in accurately reporting information and addressing misinformation. However, the pages we selected may not always match our initial expectations or assessments, which underscores the importance of involving journalists in the process. Their expertise is crucial in evaluating and interpreting the data effectively.

We used the following criteria to choose Facebook pages that are the source of real news.

- If the page has a follower exceeding 500,000 followers and having engagement with the audience

- If the page update regularly and maintains a consistent standard of reporting
- If the news they post has relevant evidence
- If the page demonstrates unbiased reporting by providing perspectives from various sides of an issue

Based on our criteria the following are the well-known Facebook pages for potential genuine news sources in Amharic language.

Facebook pages	Likes	Followers	Number of collected news
Zehabesha	1.8M	2.1M	500
BBC News Amharic	813K	1M	1000
Al Ain Amharic	594K	702k	1000
Fast Mereja	745K	809K	1000
Total number of news collected			3500

Table 4.1: Facebook pages for potential real news

Our data collection process was hampered by the difficulty of locating Amharic Facebook pages that spread false information. The difficulty of locating sources spreading false information was increased by the secretive character of these pages and any discrepancies in the authenticity of the content.

We used the following criteria to choose Facebook pages that are the source of fake news.

- Inconsistency with facts
- Violence, Hate Speech, or Harassment
- Non-Genuine Content
- Lack of transparency and accountability
- Community report and feedback

The following are the list of Facebook pages that are the source of potential fake news.

Facebook pages	Likes	Followers	Number of collected news
Awramba Times	49K	63K	500
Terara network	34K	43k	500
Awlo media	123K	133K	1000
Ethiopian dj	4.2K	24K	500
Ethio360	358K	464K	500
Total number of fake news collected			3000

Table 4.2: Facebook pages that are source of fake news

The realization that a significant portion of the posts collected did not fit the requirements to be categorized as news created a hurdle for the data scraping and collection process. The extraction of pertinent information was made more difficult by the fact that there were times when visual content was missed or improperly captured during the data gathering. The challenge of selecting a dataset that was exclusively comprised of news-related content was made more challenging by the requirement to discern between news and non-news content, as well as the possibility of visual aspects being overlooked.

4.1.2 Dataset cleaning and filtering

Cleaning and filtering datasets are crucial steps in the detection of fake news since they guarantee the accuracy, representative, and quality of the training data. Addressing problems with data quality, eliminating noise and inconsistencies, managing class imbalances, reducing biases, and standardizing formats are the main goals. The cleaning process aids in the production of a trustworthy dataset for machine learning model training by eliminating duplicates, managing false information, and taking ethical issues into account.

We addressed missing columns, eliminated repeated data, excluded news lacking visual content or text, standardized inconsistent formats, and removed duplicate entries.

4.1.3 Data Annotation

Our dataset was established with 6500 entries after news data from several public Facebook pages was collected. We selected three journalists to ensure that if one exhibited bias, the other two would provide impartiality. This strategy aimed to uphold objectivity and deliver a balanced perspective in our annotation work. Each journalist brought unique expertise and a commitment to unbiased reporting, which is essential for thorough and credible material analysis. By employing this approach, we minimized the risk of any single bias influencing the annotations, thereby maintaining the integrity and reliability of our work. The journalists are independent, not affiliated with mainstream or government media, and work individually.

After a thorough assessment by professional journalists, some entries were deemed unfit for publication as news, while others were excluded from our dataset because they did not have the necessary visual elements. Three experts participated in the annotation process; if two out of the three thought a piece of news was fake, it was classified as such; if not, it was regarded as authentic. Only 4000 of the 6500 news submissions that were first gathered satisfied the strict requirements to be included as news. 2353 entries were verified as genuine news within this improved dataset, while 1647 entries were classified as fake news.

Multimodal dataset

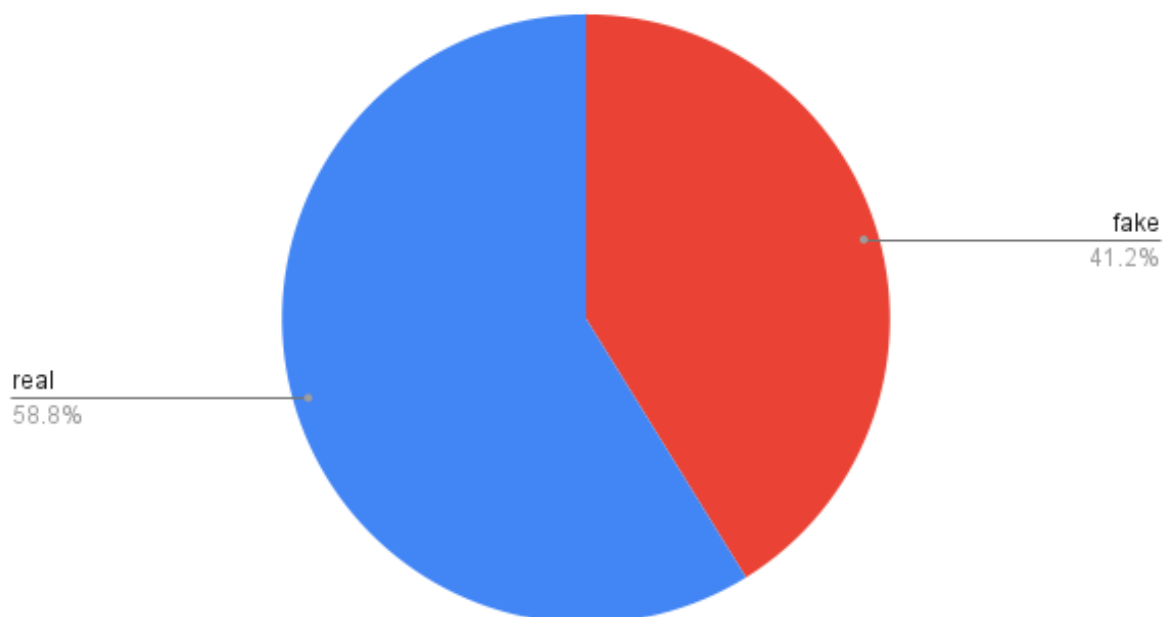


Figure 4.4: Multimodal Dataset

4.2 Preprocessing

The dataset we prepared after the above steps has both the visual and text content. The next step in our research was preprocessing the dataset.

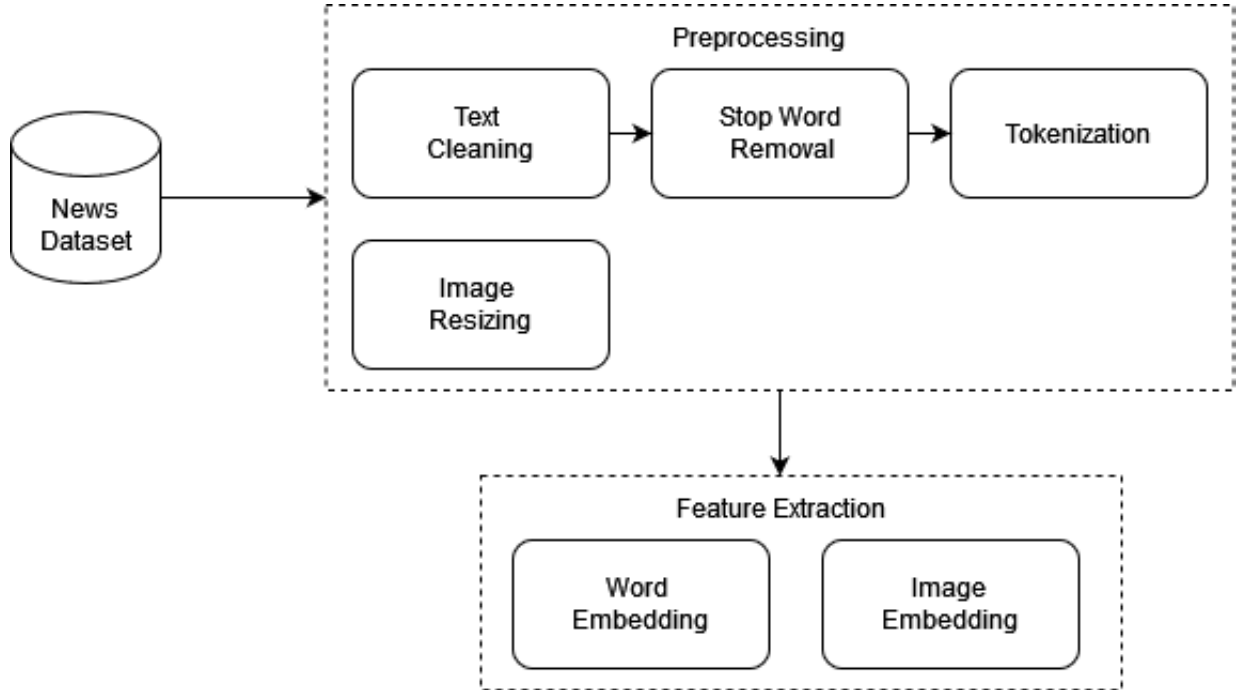


Figure 4.5: Dataset Preprocessing Steps

4.2.1 Text preprocessing

Text preprocessing is an approach for cleaning and preparing text data for use in our model.

The following are the steps for text preprocessing:

1. **Text Cleaning:** We removed numbers, symbols, different URL links and words in another language that exist in the dataset to improve the preprocessing task.
2. **Stop Word Removal:** Stop words are words that are typically filtered out before processing natural language. These are the most commonly used words in any language (like articles, prepositions, pronouns, conjunctions, etc.) and do not add much to the text. There are common stop words in Amharic which are used for grammatical purposes, which are non-informative to identify documents. In addition to the common stop words, there are also news specific stop words and their use is for elaboration and common to all news [42]. We prepared a text document that has a list of stop words in Amharic language.

Below in figure 4.8 is list of some of the stop words we collected.

ትናንት	አይደለም	እያሉ	ወደፊት	ጥቂት	ሲል	ማን	ያንዳንዳችው
የተሟላ	እስከ	እያንዳንዱ	ቢል	ፊት	በምን	ባክዎ	እባኩ
ነበረ	እርሱም	እያንዳንዳችው	ውስጥ	ባሉ	እንዲሆን	ባክ	ባኩሽ
ነበረች	እባኩህ	እያንዳንዱ	ውጪ	ይሁን	እነሆም	ዉጪ	ያ
ነበሩ	እባኩዎ	ከላይ	የታች	ብል	ባይሆንም	ዉስጥ	ተለያዩ
ነኝ	እና	ከታች	የት	ቢኖር	ሲሆን	ኋላ	ተለያየ
ነው	እኔ	ከውስጥ	የዚህ	ያለ	ሁለት	መካከል	ነይ
ነገር	እንደ	ከዚያ	የዚያ	ስለዚህም	እዚህ	ሰሞን	እክል
ነገሮች	እንደገና	ከፊት	ይህ	ያደርጋል	ባለፉት	ትናንትና	ሆነ
ናት	እንዲሁም	ወዘተ	ይላል	ይሆናል	የውስጥ	ጋራ	ሆና
ናቸው	እንዲህ	ወይም	ይባላል	አንተ	ስለ	እነሱ	በሆነ
አለበት	እንዲያውም	ወይንም	ደግሞ	ቢሆንም	በርካታ	እሱ	ከሆነ
አለባችሁ	እንጂ	ወይዘሪት	ድረስ	እንደዚህ	አንዳንድ	እሷ	የሆነ
አሉና	እኛ	ወይዘሮ	ጋር	በጣም	አንዳንዳችው	እናንተ	የሆኑ
አቶ	እያለ	ወደ	ግን	በዚህም	ገና	ወይ	በመሆኑ

Figure 4.6: List of stop words

We remove the low-level information from our text by removing these words, allowing us to focus on the important information. We can conclude that removing such words has no negative impact on the model we train for our task. The removal of stop words significantly reduces the dataset size and thus the training time due to the fewer tokens involved in the training.

3. **Tokenization:** It is the next step in our research which is splitting the sentence into words or breaking down a piece of text into small units called tokens. It's a very important step in preprocessing.

Amharic language has its own punctuation marks which separate texts or sentences into a stream of words. Amharic punctuation marks include "Hulet netb" or colon, "arat netb ", "netela serez", "derib sereze", question mark and exclamation mark '!' Or ' ' were used as sentence delimiter or as white space.

In our research we used all punctuation marks and white space as a delimiter of words to break the sentence into words based on the white space that exist between words. NLTK (Natural language toolkit) was used for this purpose. Below in figure 4.7 we can see an example of tokenization applied on fake news that's in our data.

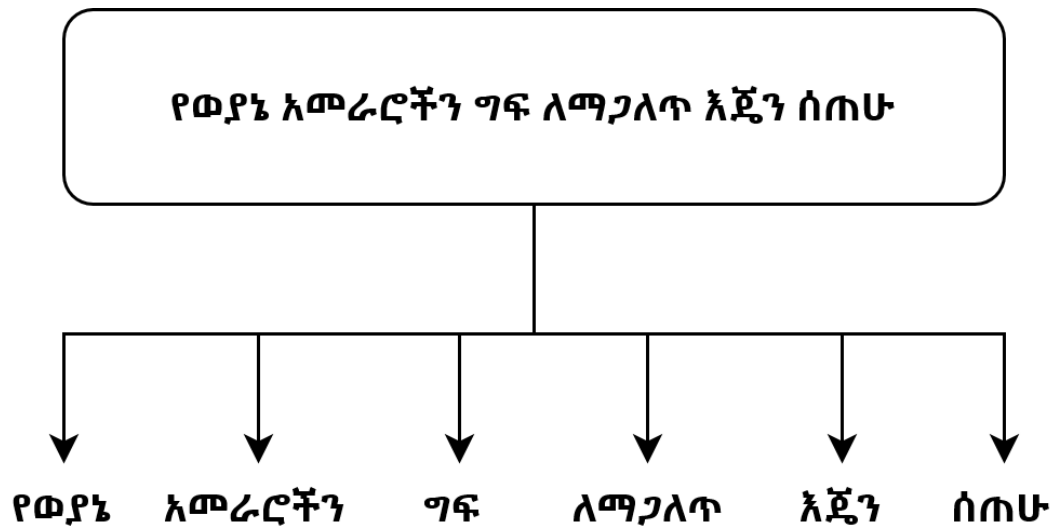


Figure 4.7: Tokenization on Amharic sentence

4.2.2 Image preprocessing

Image processing is an essential step before feeding images to a model. This process helps to enhance the image quality, reduce noise, and standardize the data, which can significantly improve the performance of the model.

The images were resized to a specific size (224x224 pixels). This is done to standardize the image size and reduce computational complexity that in return ensure uniformity in a dataset.

4.3 Feature Extraction

Feature extraction in the context of fake news detection is finding and picking pertinent traits from textual or other data sources that can aid in differentiating between real and fake news. The intention is to identify any patterns or traits that might point to false or misleading information.

4.3.1 Word Embedding

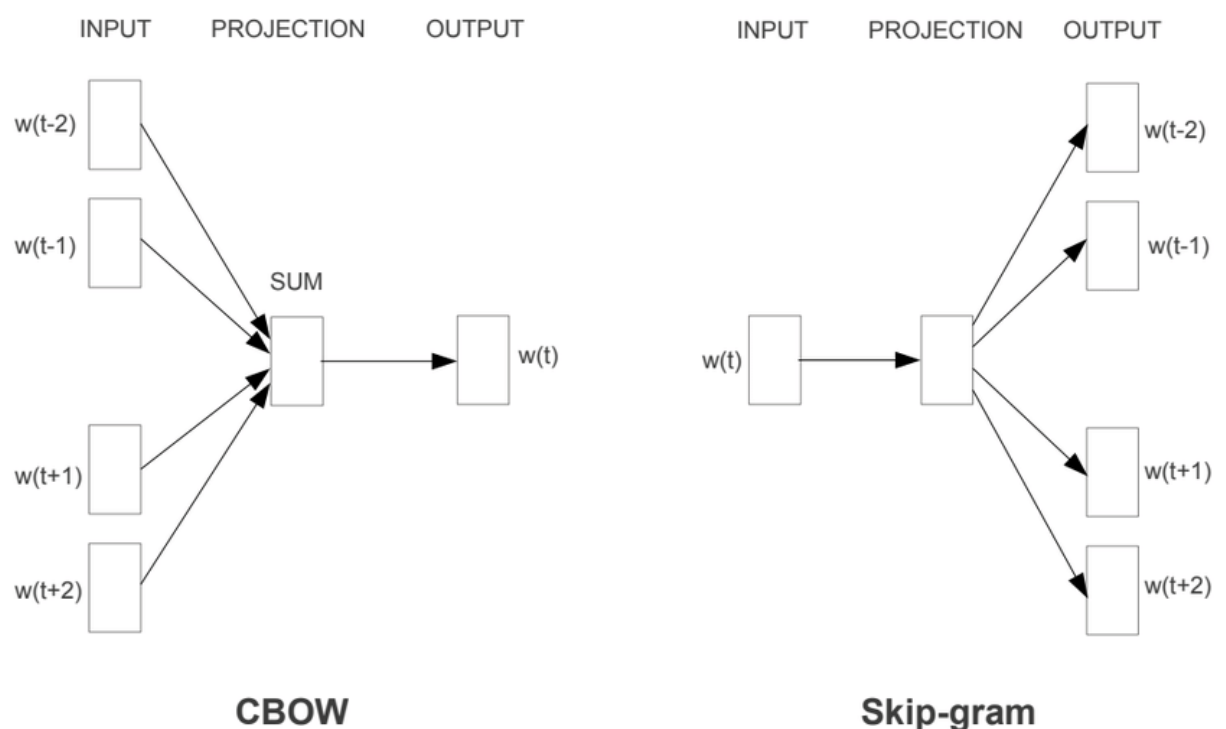
Word Embedding or vectorization is the process of converting text data to numerical vectors. To convert the text data into numerical data, we need some smart ways which are known as vectorization, or in the NLP world, it is known as Word embedding. There exist different word embedding techniques for text feature extraction in text.

4.3.1.1 Word2vec

word2vec as the name indicates it operates on word level. It gives similar vector representation for words with similar meaning. All words in the dataset are represented by its own vector and has distinct representation. Which will cause the ignorance of the internal structure in a language. Word2Vec model is used to find the similarity of words across the model. This model uses two main architectures to find the vectors: Skip-gram and CBOW.

Skip-gram: this architecture uses a word to predict the context.

CBOW: this method is the reverse of Skip-gram, it uses the context to predict the missing word.



[43]

Figure 4.8: Word2vec Model

4.3.1.2 Fasttext

Fasttext is an extension of word2vec but works on character n-grams. It operates on character and subword levels. Fasttext handles out of vocab words better than word2vec by considering subword information. A FastText-based word embedding model, named *cc_am_300*, has been developed for the Amharic language. This embedding utilizes 300 dimensions for word representation. Notably, the model has a restricted number of word vectors and includes uncleaned English tokens. To solve this issue we trained a fasttext model using Amharic-corpus with variable dimensions.

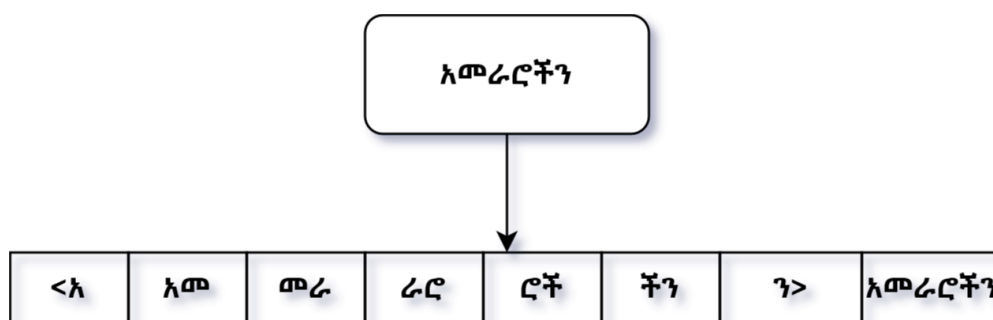


Figure 4.9: Fasttext on Amharic word

4.3.2 Image Embedding

Most machine learning algorithms for classification, regression or clustering cannot work on images directly. To perform the analysis, we need to transform them into more suitable representation vectors of numbers named embeddings. Image embedding is a vector representation of an image in which images with similar motives have similar vector profiles. Image embedding is lower-dimensional representation of the image. Alternatively put, it is a detailed vector representation of the image that can be applied to a variety of tasks, including classification. In this study, the image embedding was produced by a Convolutional Neural Network (CNN).

4.4 Proposed models

Deep learning belongs to the branch of machine learning and is an advancement of classic machine learning. Classic machine learning requires human assistance in extracting its feature while deep learning has the advantage of automatically learning raw data features. By forming more general features, more specific features are obtained and Deep learning uses a Deep Neural Network (DNN) which consist of a convolutional layer, pooling layer, and fully connected layer to perform this task . This study uses a deep learning method to detect fake news.

4.4.1 Unimodal Fake News Detection Models

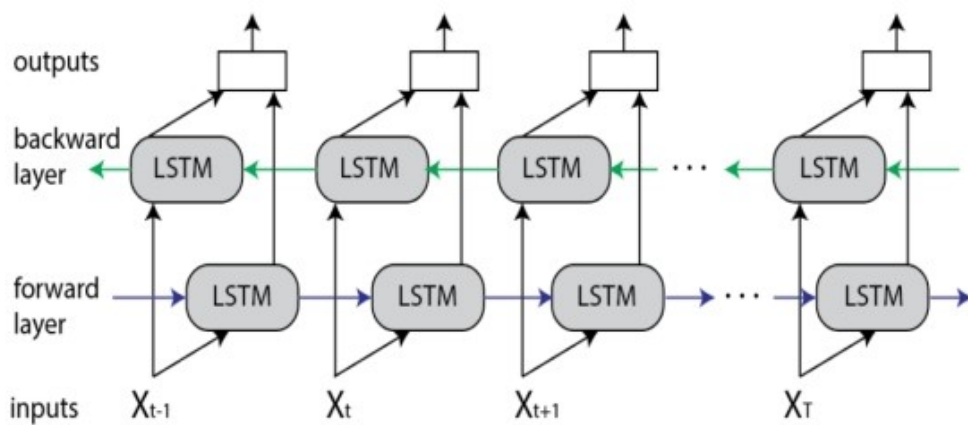
4.4.1.1 RNN

Recurrent Neural Network is a kind of artificial neural network intended to identify patterns in data sequences, including time series, voice, text, and other types of data. Because RNNs may retain information about past inputs through their internal state, they are especially useful for jobs where context and order are important. This explains why NLP uses them the most.

Long Short-Term Memory (LSTM) networks which are type of RNN efficiently overcome problems like disappearing gradients by capturing long-term dependencies in sequential data through the use of memory cells and gating methods[44]. Time series forecasting, speech recognition, and natural language processing are three common applications for LSTMs. This is extended by Bidirectional LSTMs (BiLSTMs), which analyze sequences in both forward and backward directions. This gives a richer context and enhances performance in tasks like named entity identification and machine translation, where knowing both past and future information is critical.

Bidirectional LSTM (BiLSTM) is a recurrent neural network used primarily on natural language processing. Unlike standard LSTM, the input flows in both directions, and it's capable of utilizing information from both sides[45]. It's also a powerful tool for modeling the sequential dependencies between words and phrases in both directions of the sequence. In summary, BiLSTM adds one more LSTM layer, which reverses the direction of information flow. Briefly, it means that the input sequence flows backward in the additional LSTM layer. Then we combine the outputs from both LSTM layers in several ways, such as average, sum, multiplication, or concatenation. BiLSTM will have a different output for every component (word) of the sequence (sentence). As a result, the BiLSTM model is beneficial for our task.

Using bidirectional will run the inputs in two ways, one from past to future and one from future to past and what differs this approach from unidirectional is that in the LSTM that runs backwards you preserve information from the future and using the two hidden states combined you are able in any point in time to preserve information from both past and future.



[46]

Figure 4.10: BiLSTM architecture

Sequential dependencies and contextual information are extracted from the text by the BiLSTM model through processing of the text part of the data. In order to extract visual features, the CNN processes the image component of the data simultaneously. Following processing, concatenation takes place in a fusion layer that combines the outputs from the dense layers of the two models (BiLSTM and CNN). The text and image features are combined into a single representation via this fusion layer. The combined features are further refined by passing the concatenated output through another dense layer. The classification layer receives the output from the dense layer at the end and uses it to determine if the news is authentic or fake.

4.4.1.2 CNN

Convolutional Neural Network is the most popular deep learning architecture that uses convolutional layers to map input data features. The layers are arranged by applying different filter sizes to produce different feature mappings. CNN can obtain information about the input data based on the feature mapping results. A pooling layer usually accompanies the convolutional layer to produce the exact output dimensions even with different filters. The pooling layer also lightens the computational load by reducing the output dimensions without losing essential information.

The following are the layers of CNN.

1. Convolutional Layer is the first layer used to extract the various features from the input. This layer performs the mathematical operation of convolution between the input and a filter of size $M \times M$. The dot product between the filter and the parts of the input with respect to the filter size is calculated by sliding the filter over the input ($M \times M$).
2. Pooling Layer follows a Convolutional Layer. This layer's primary goal is to reduce the size of the convolved feature map in order to reduce computational costs. This is accomplished by reducing the connections between layers and operating independently on each feature map. There are various types of Pooling operations depending on the method used. It essentially sums up the features produced by a convolution layer. The largest element from the feature map is used in Max Pooling. Average Pooling computes the average of the elements in a predefined image section size. Sum Pooling computes the total sum of the elements in the predefined section. The Pooling Layer is commonly used as a link between the Convolutional Layer and the FC Layer.
3. The Fully Connected (FC) layer is made up of weights and biases as well as neurons and is used to connect neurons from different layers. These layers are typically placed prior to the output layer and constitute the final few layers of a CNN Architecture. The previous layers' input images are flattened and fed to the FC layer in this step. The flattened vector is then passed through a few more FC layers, where the mathematical function operations are typically performed. At this point, the classification procedure is initiated. The reason for connecting two layers is that two fully connected layers outperform a single connected layer. These CNN layers reduce the need for human supervision.

4. Dropout Overfitting in the training dataset is common when all features are connected to the FC layer. Overfitting occurs when a model performs so well on training data that it has a negative impact on the model's performance when applied to new data. To address this issue, a dropout layer is used, in which a few neurons are removed from the neural network during the training process, resulting in a smaller model. When a dropout of 0.3 is reached, 30 percent of the nodes in the neural network are randomly removed. Dropout improves the performance of a machine learning model by preventing overfitting by simplifying the network. It also removes neurons from the neural.
5. Activation Function The activation function is a critical parameter in the CNN model. They are used to learn and approximate any type of continuous and complex relationship between network variables. In other words, it determines which model information should be ejected forward and which should not at the network's end. It increases the network's nonlinearity. There are several popular activation functions, including the ReLU, Softmax, tanH, and Sigmoid functions. In a CNN model, activation functions determine whether or not a neuron should be activated. It determines whether or not the input to the work is significant enough to predict utilizing mathematical operations.

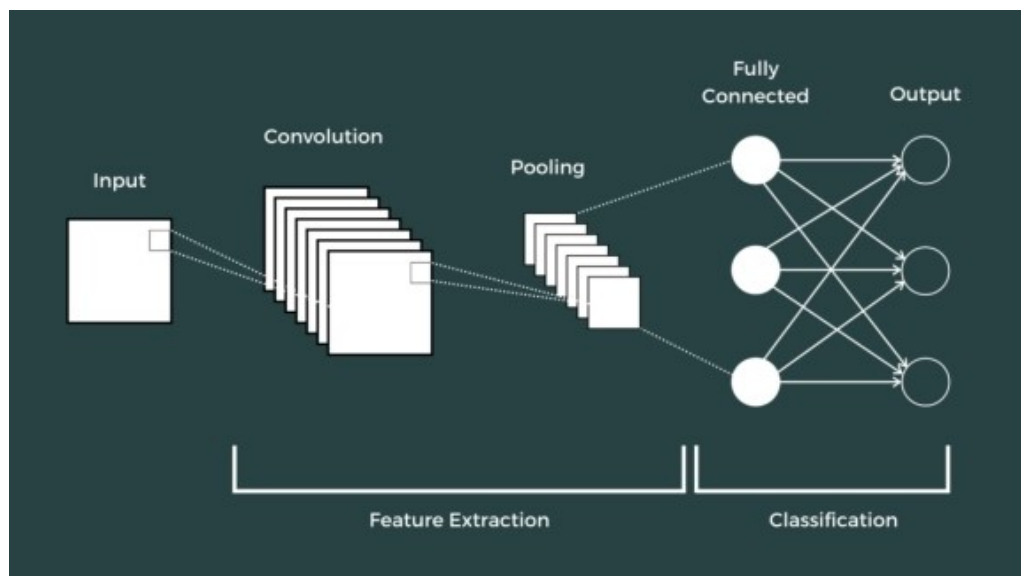


Figure 4.11: CNN architecture

[47]

CNN mostly works by getting an image, designating it some weightage based on the different objects of the image, and then distinguishing them from each other. CNN requires very little pre-process data as compared to other deep learning algorithms. One of the main capabilities of CNN is that it applies primitive methods for training its classifiers, which makes it good enough to learn the characteristics of the target object. Many studies showed that CNN perform better than other machine learning and deep learning models to detect fake news so we will use it to build the proposed model.

4.4.1.3 Machine Learning Models

Traditional machine learning algorithms are essential for categorizing and recognizing false information in fake news detection. Regression models, both logistic and linear, can be used to determine the correlation between textual characteristics and the probability that a news article is fraudulent. Using a variety of variables, such as word frequency and source credibility, decision trees and random forests offer interpretability and resilience in their prediction processes. Support Vector Machines (SVM) are useful for determining the best borders between authentic and fraudulent news. When it comes to text classification, Naive Bayes is especially helpful since it utilizes the likelihood of words appearing in phony versus authentic news items.

4.5 Performance Evaluation Metrics

These four assessment criteria will be used as performance measures to assess the suggested approaches' efficiency. Accuracy, recall, f1-score, and precision are the metrics used to evaluate the approach for proposed fake news detection system constructed using DL approaches.

1. **Accuracy** measures how much accurately the model learns to classify the data. It is computed by dividing True Positive to the overall number of instances in the dataset.

$$Accuracy = \frac{TP}{TR} \quad (4.1)$$

2. **Precision** is a measure of the likelihood of getting correct positive class classification. It is computed as the number of True Positives divided by the number of True Positive plus False Positive.

$$Precision = \frac{TP}{TR + FP} \quad (4.2)$$

3. **Recall** is the measure of how sensitive our model is in identifying the positive class. A recall is computed as the number of True Positive divided by the number of True Positives plus False Negatives.

$$Recall = \frac{TP}{TR + FN} \quad (4.3)$$

4. **F1-score** is the measure of how sensitive our model is in identifying the positive class. A recall is computed as the number of True Positive divided by the number of True Positives plus False Negatives..

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4.4)$$

5. **Confusion matrix** is a table that is used in statistics and machine learning to assess how well a classification system performs. It gives an overview of the outcomes of a classification model applied to a test set of data, displaying the quantity of true positive, true negative, false positive, and false negative predictions.

		Actual	
		Fake (1)	Fake (0)
Predicted	Fake (1)	TP	FP
	Real (0)	FN	TN

Figure 4.12: Confusion Matrix

Where,

- **TP (True Positive)** represents the number of correctly categorized fake news in the real news category.
- **FP (False Positive)** represents the number of incorrectly classified fake news in the negative news category.

- **FN (False Negative)** represents the number of fake news incorrectly classified in the negative news category.
- **TR** represents the total number of the language's news and speech in the test data

4.6 Fake News Detection Model Implementation

We will go through all the processes that took place in fake news detection model implementation in our research.

4.6.1 Preprocessing

1. **Dataset Loading** the first step in the implementation of our model was to load the collected amharic news dataset to the python environment. We used the pandas library to load the data.
2. **Dataset cleaning:** the loaded dataset was cleaned for better performance of our model by using 're' python module and also we used a 'clean_irrelevant' function which is designed to preprocess and clean our text data by removing unwanted characters, URLs, punctuation, numbers, emojis, and other ,and removing extra spaces.
3. **Stop word removal:** Importing the NLTK library was our next step in making text processing easier. We extracted stop words from the Amharic news dataset we had assembled by utilizing the pre-existing Amharic stop word corpus. This stage consisted of eliminating frequently appearing words that have little bearing on the news stories' context, in order to prepare the dataset for further analysis and modeling.

4. **Tokenization:** We imported the Tokenizer and pad_sequences modules from the TensorFlow.keras.preprocessing.text package in TensorFlow for the tokenization process. The Tokenizer class in Keras was used for text tokenization, which is the process of converting a sequence of text into a sequence of tokens. It allowed us to fit a tokenizer on a given dataset (e.g., a list of text documents) and then convert the text data into sequences of integers, where each integer represents a unique token. The pad_sequences function was used to ensure that all sequences in a list have the same length by padding shorter sequences or truncating longer sequences. It is applied after tokenization to prepare input data for a neural network that expects fixed-length sequences.
5. **Image Preprocessing:** The TensorFlow image preprocessing module's load_img and img_to_array methods were imported. When working with image data for deep learning tasks, these functions are frequently utilized. The load_img function was used to load an image file into a Python Imaging Library (PIL) image and img_to_array function was used to convert a PIL image or a list of PIL images to a NumPy array. An image is loaded, resized to the required input size (224x224 pixels), and preprocessed to match the format expected by the feature extractor. We used preprocess_input function from Tensorflow to maximize the pixel value.

4.6.2 Feature Extraction

We have both the text and visual content in our dataset so we had to extract the features from each.

4.6.2.1 Word Embedding

We imported the fasttext library, which is a popular library for training and using word embeddings and text classification models. We used our own word embedder using fasttext and amharic corpus.

4.6.2.2 Image Embedding

We imported the VGG16 model from the Keras Applications module of TensorFlow. A well-liked convolutional neural network (CNN) architecture that was first shown for image classification is the VGG16 model. It is frequently utilized for a variety of computer vision tasks because it has been pre-trained on the ImageNet dataset. The VGG16 model is loaded with pre-trained weights from ImageNet and we used it for image feature extraction.

4.6.3 Fake news detection model Implementation

CNN-BiLSTM : The input layer for the image model is determined by the image input layer, whose shape matches the dimensions of the image's data. First we applied a 2D convolutional layer (Conv2D) with ReLU activation, 64 filters, and a kernel size of (3, 3). It then use a 2D max pooling layer (MaxPooling2D) to downsample the spatial dimensions with a pool size of (2, 2). We then applied a second Conv2D layer with ReLU activated, 128 filters, and a kernel size of (3, 3). Another layer of max pooling is applied on the output of the previous layer which is conv2. Flatten layer flattens out the output from the last layer of the CNN into a one-dimensional vector.

Text Input Layer defines a text model input layer that's shape matches the text sequences' embedding vector lengths. The next layer was embedding layer that uses pre-trained embeddings (text_embeddings) to transform the input integer sequences into dense vectors. To maintain the pre-trained weights, the layer is configured to be non-trainable (trainable=False). The Bidirectional Long Short-Term Memory (BiLSTM) layer is applied using 64 units.

After concatenation of the image and text features from the CNN and BiLSTM models the next step was adding dense layer. Furthermore, sigmoid activation function and dense layer with 1 unit. This is the output layer for binary classification. A binary cross-entropy loss function, Adam optimizer, and accuracy metric are then used to construct the model. Using np.expand_dims(), the training data is subsequently reshaped to match the Conv1D layer's input shape. Then, using a batch size of 16 and 20% of the entire dataset for validation data, the model is trained on the training data for 100 epochs. In conclusion, the model's accuracy and classification report are assessed using test data.

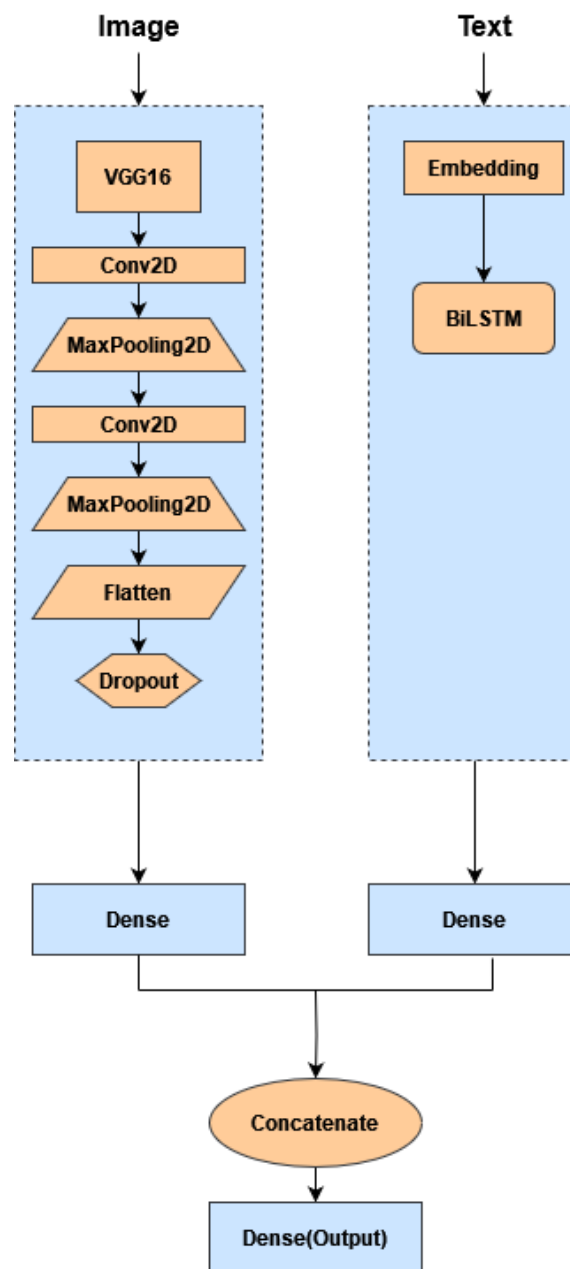


Figure 4.13: CNN-BiLSTM Architecture

Chapter 5

Result and Discussion

In this chapter we will discuss about the experiment setup, the tools we used to complete the research, the results we got after experiments and its respective discussion.

5.1 Experiment Setup

The experiments were conducted using Python 3.9.16, TensorFlow 2.12.0, and Keras 2.12.0, installed on a Dell Precision 7920 Tower server. This server is equipped with an Intel(R) Xeon(R) Gold 6230R CPU, 64GB of RAM, and an NVIDIA RTX A4000 GPU, running the Ubuntu 22.04.2 LTS Server operating system.

The following are software and tools used in the research experiment:

- **Python:** is high level and general-purpose programming language that has abundance of libraries and frameworks that facilitate coding and save development time.
- **TensorFlow:** is a free and open-source software library for machine learning and Artificial intelligence. It can be used across a range of tasks but has a particular focus on training and inference of deep neural network.
- **NLTK:** The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing in the Python programming language.
- **Keras:** is an open-source software library that provides a Python interface for artificial neural networks. Keras acts as an interface for the TensorFlow library.
- **NumPy:** is a robust numerical computation package for the Python programming language. Large, multi-dimensional arrays and matrices are supported, and a number of mathematical operations can be performed on these arrays. A key library for scientific computing in Python, NumPy is extensively utilized in domains like data analysis, scientific research, and machine learning.

- **Panadas:** is a prominent and powerful Python programming language and data manipulation and analysis toolkit. Along with tools for data cleansing, exploration, and analysis, it offers data structures for effectively storing and handling huge datasets.
- **Scikit-learn:** is a well-known machine learning library for the Python programming language. It offers a variety of machine learning tools and utilities for tasks like classification, regression, clustering, and dimensionality reduction, along with easy-to-use and effective tools for data analysis and modeling.
- **Jupyter Notebook:** an open-source web application for creating and sharing live code and explanatory text documents.

To autotune the model we experimented using five hyperparameters interchangeably on our dataset.

1. Batch size :which is the number of samples that will be propagated through the network.In our experiment we used 16,32,64 batch sizes interchangeably.
2. Epochs: refers one whole iteration of the training dataset through the learning process.Different number of epochs were used in our experiment such as 10,50,100.
3. Dropout: is a data, or noise, that is purposefully removed from a neural network in order to speed up processing and result turnaround .For the purpose of our experiment we used 0.1,0.01,0.001 interchangeably.
4. Output Layer: 64,128,252
5. Optimizer: adjust model parameters cyclically during training to minimize a loss function, enabling neural networks to learn from data.Some of the optimizer used are Adam, SGD, RMSprop.

We got the best accuracy and F1-score for Adam optimizer with 64 output layer batch size of 16, on 100 epochs and dropout of 0.1 on our model.

5.2 Model Performance Metrics

5.2.1 Unimodal fake news detection models performance

In this section, we will test various machine learning and deep learning models on our test dataset and then discuss the performance evaluation metrics values of each model.

5.2.1.1 Text Only Models

The first experiment in this section was to choose the feature extraction method and compare the results on different unimodal proposed models.

Experiment 1: Word2vec on different unimodal proposed models performance

Feature Extractor	Models	Accuracy	Precision	Recall	F1-score
Word2vec	CNN	0.78	0.79	0.95	0.86
Word2vec	LR	0.76	0.77	0.94	0.85
Word2vec	BiLSTM	0.81	0.88	0.84	0.86

Table 5.1: Performance Comparison Of Unimodal Using Word2vec

We can see in figure 5.1 performance metrics for a text classification model that utilizes Word2vec for feature extraction and a BiLSTM model for classification.

This model achieves an accuracy of 0.81, indicating that it correctly predicts the class of 81% of the instances in the dataset. The precision score of 0.88 suggests that when the model predicts a positive instance, it is correct 88% of the time. The recall score of 0.84 indicates that the model identifies 84% of all actual positive instances.

The F1-score, which balances precision and recall, is 0.86. This score provides a single measure of the model's overall performance, indicating a good balance between precision and recall in classifying the text data.

LR which stands for Logistic Regression is a machine learning model. We can see on figure 5.1 performance metrics for a text classification model that combines Word2vec for feature extraction and Logistic Regression (LR) for classification.

The model achieved an accuracy of 0.76, indicating it correctly predicted the class of 76% of the instances in the dataset. This metric gives an overall measure of the model's correctness in its predictions.

Precision, which measures the accuracy of positive predictions, was calculated at 0.77. This means that when the model predicted an instance as positive, it was correct 77% of the time. Recall, which measures the model's ability to identify positive instances, achieved a score of 0.94. This indicates that the model identified 94% of all actual positive instances present in the dataset.

The F1-score, which is the harmonic mean of precision and recall, was reported as 0.85. This score provides a balanced assessment of the model's performance, taking into account both the precision and recall metrics.

The high recall suggests that the Word2vec embeddings effectively capture semantic similarities in the text, enabling the model to identify positive instances with high accuracy. The combination of Word2vec and Logistic Regression proves effective in achieving a good balance between accuracy, precision, and recall, making it suitable for various text classification tasks.

In figure 5.1 we can see the performance metrics of a text classification model that utilizes Word2vec for feature extraction and a Convolutional Neural Network (CNN) for classification.

The model achieved an accuracy of 0.78, indicating it correctly predicted the class of 78% of the instances in the dataset. This metric provides an overall measure of how accurate the model's predictions are.

Precision, which measures the accuracy of positive predictions, was reported at 0.79. This means that when the model predicted an instance as positive, it was correct 79% of the time.

The recall score, which measures the model's ability to correctly identify positive instances, achieved a high value of 0.95. This indicates that the model identified 95% of all actual positive instances present in the dataset.

The F1-score, which combines precision and recall into a single metric, was calculated as 0.86. This score provides a balanced assessment of the model's performance, reflecting both its precision and its ability to capture positive instances effectively.

BiLSTM achieved 3% higher accuracy than the other two models, showcasing its superior performance in this evaluation. This highlights BiLSTM's effectiveness in handling the complexity of the dataset and its ability to capture dependencies in both forward and backward directions.

Word2vec + BiLSTM, LR and CNN

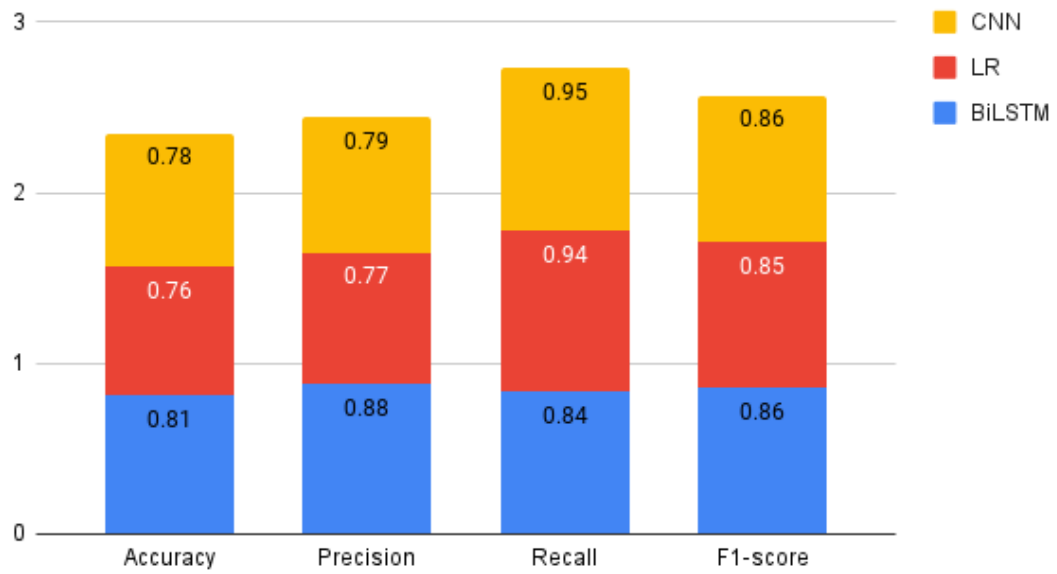


Figure 5.1: Performance Comparison of Unimodal on Word2vec

Experiment 2: Fasttext on different unimodal proposed models performance

Feature Extractor	Models	Accuracy	Precision	Recall	F1-score
Fasttext	CNN	0.84	0.86	0.92	0.89
Fasttext	LR	0.81	0.80	0.98	0.88
Fasttext	BiLSTM	0.86	0.87	0.94	0.9

Table 5.2: Performance Comparision Of Unimodal Models Using Fasttext

Figure 5.2 below show the performance metrics for a text classification task using FastText for feature extraction and BiLSTM as the model.

The model's accuracy of 0.865 shows that it performed well in predicting the correct class labels for the majority of instances in the dataset.

Precision (0.87): Precision measures how accurately the model predicts positive instances. A precision score of 0.87 means that 87% of the instances predicted as positive by the model were correct. With a precision of 0.87, the model effectively minimized false positives by accurately identifying positive instances

Recall (0.94): Recall measures the model's ability to identify all actual positive instances. A recall score of 0.94 suggests the model correctly identified 94% of all positive instances present in the dataset.

The high recall score of 0.94 indicates that the model captured nearly all positive instances present in the dataset, indicating its effectiveness in identifying relevant information.

F1-score (0.9): The F1-score combines precision and recall into a single metric, providing a balanced assessment of the model's performance. A score of 0.9 indicates a good balance between precision and recall.

The performance metrics for a text classification model using FastText for feature extraction and Logistic Regression (LR) for classification.

The model achieved an accuracy of 0.81, indicating it correctly predicted 81% of the instances in the dataset. This metric provides an overall view of the model's correctness in its predictions.

Precision, which measures the accuracy of positive predictions, was reported at 0.80. This means that when the model predicted an instance as positive, it was correct 80% of the time.

The recall score, which measures the model's ability to correctly identify positive instances, achieved a high value of 0.98. This indicates that the model identified 98% of all actual positive instances present in the dataset.

The F1-score, which combines precision and recall into a single metric, was calculated as 0.88. This score provides a balanced assessment of the model's performance, reflecting both its precision and its ability to capture positive instances effectively.

The performance metrics are in figure 5.2 for a text classification model utilizing FastText for feature extraction and a Convolutional Neural Network (CNN) for classification.

FastText is employed to generate embeddings that capture semantic meanings of words and phrases in the text. These embeddings serve as input to the CNN, which utilizes convolutional layers to extract features from the text data, capturing local patterns and relationships.

Accuracy (0.84): The model achieves an accuracy of 84%, indicating it correctly predicts the class labels for 84% of the instances in the dataset. This metric reflects the overall correctness and effectiveness of the model's predictions.

Precision (0.86): Precision measures how accurately the model identifies positive instances. A precision score of 0.86 means that when the model predicts an instance as positive, it is correct 86% of the time. This highlights the model's ability to minimize false positives.

Recall (0.92): Recall assesses the model's capability to correctly identify all actual positive instances in the dataset. A recall score of 0.92 indicates that the model successfully identifies 92% of all positive instances. This metric underscores the model's strength in capturing relevant information.

F1-score (0.89): The F1-score, which harmonizes precision and recall into a single metric, provides an overall measure of the model's performance. With an F1-score of 0.89, the model demonstrates a balanced performance in both precision and recall, indicating its effectiveness in classifying text data.

The FastText and CNN model combination leverages FastText's ability to encode semantic meanings efficiently, enhancing the CNN's capability to learn and extract intricate features from the text data. This approach is particularly effective in capturing both local and global dependencies within the text, contributing to the model's high performance across accuracy, precision, recall, and F1-score metrics.

BiLSTM outperformed other models in the comparison, achieving a 2% higher accuracy and 1% higher F1-score. This suggests that BiLSTM's ability to capture bidirectional dependencies in sequential data contributes to its superior performance in accurately predicting and classifying text data.

Fasttext + BiLSTM, LR and CNN

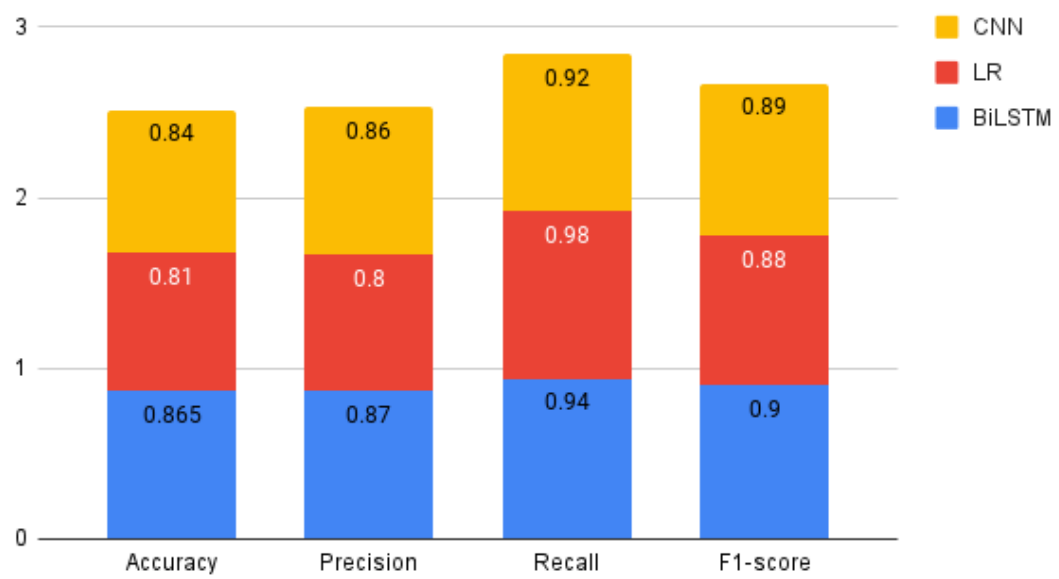


Figure 5.2: Performance Comparison of Unimodal Models on Fasttext

Fasttext/BiLSTM vs Word2vec/BiLSTM

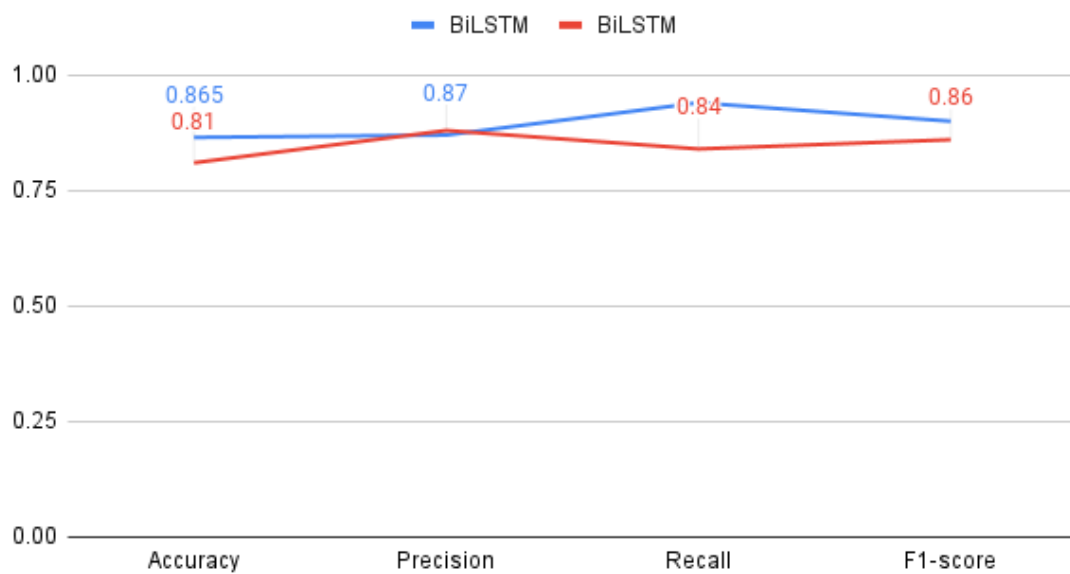


Figure 5.3: Fasttext vs Word2vec

5.2.1.2 Image Only Model

For the image only dataset the CNN model was evaluated on a test dataset consisting of 770 image samples, out of which it accurately identified 558 samples and failed to predict 212 correctly. This results in an accuracy of 72%, meaning the model's predictions were correct 72% of the time. Additionally, the model achieved an f1-score of 83%, which considers both precision and recall, providing a balanced measure of the model's performance. The F1 score indicates that the model is effective in reliably detecting relevant images while minimizing false positives and missed detection. These metrics suggest that while the image model performs reasonably well, there is room for improvement in its predictive accuracy.

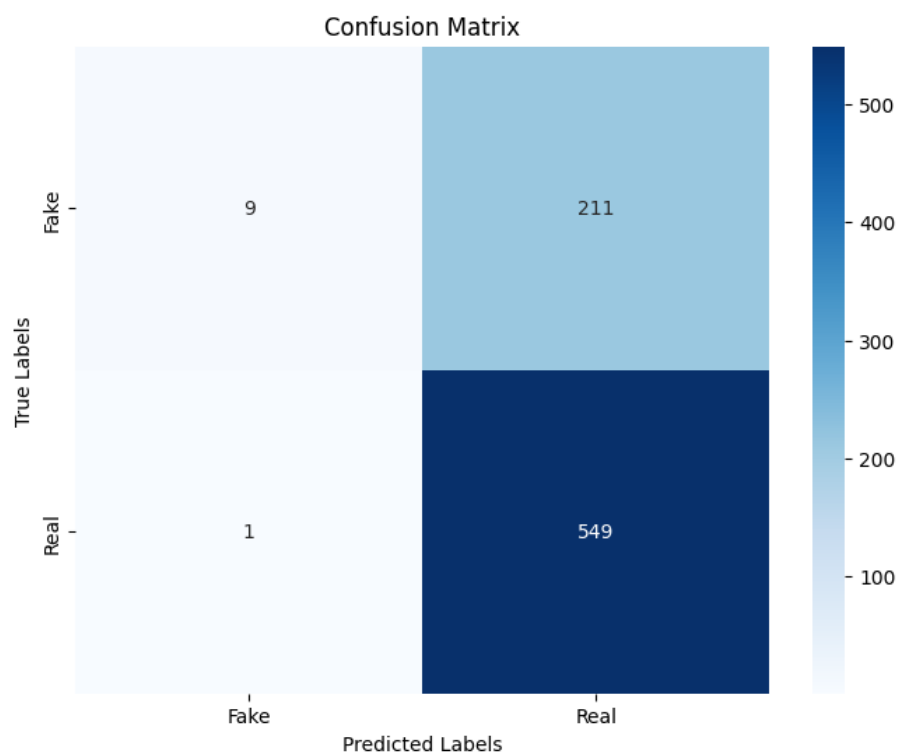


Figure 5.4: Confusion matrix for CNN Model

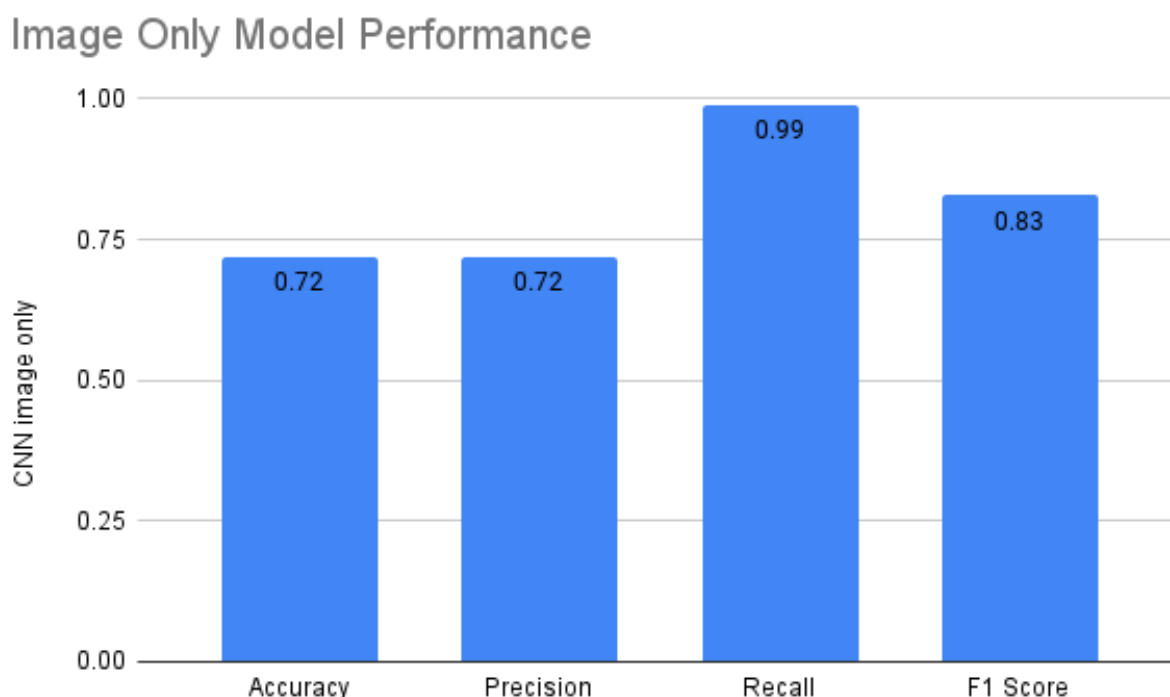


Figure 5.5: Image only Model Performance

The differences in performance between BiLSTM and CNN can be attributed to their respective strengths and the nature of the data they process:

BiLSTM (Text only): BiLSTM excels in processing sequential data such as text by capturing contextual dependencies from both past and future directions (bidirectional). This makes it particularly effective in tasks where understanding the sequence and context of words is crucial, resulting in high accuracy, precision, recall, and F1-score.

CNN (Image only): CNNs are highly effective in extracting spatial features from images through convolutional layers. The high recall of 0.99 suggests CNN effectively identifies nearly all positive instances from image data. However, the lower accuracy and F1-score compared to BiLSTM indicate that while CNN performs well in recognizing positive instances (high recall), it may have challenges in achieving precision and overall accuracy in classifying images alone. In summary, BiLSTM performs strongly in text classification due to its ability to understand sequential dependencies, while CNN excels in image classification by effectively capturing spatial features. The differences in their performance metrics reflect these inherent strengths and the nature of the data each model processes.

BiLSTM Text only and CNN Image only

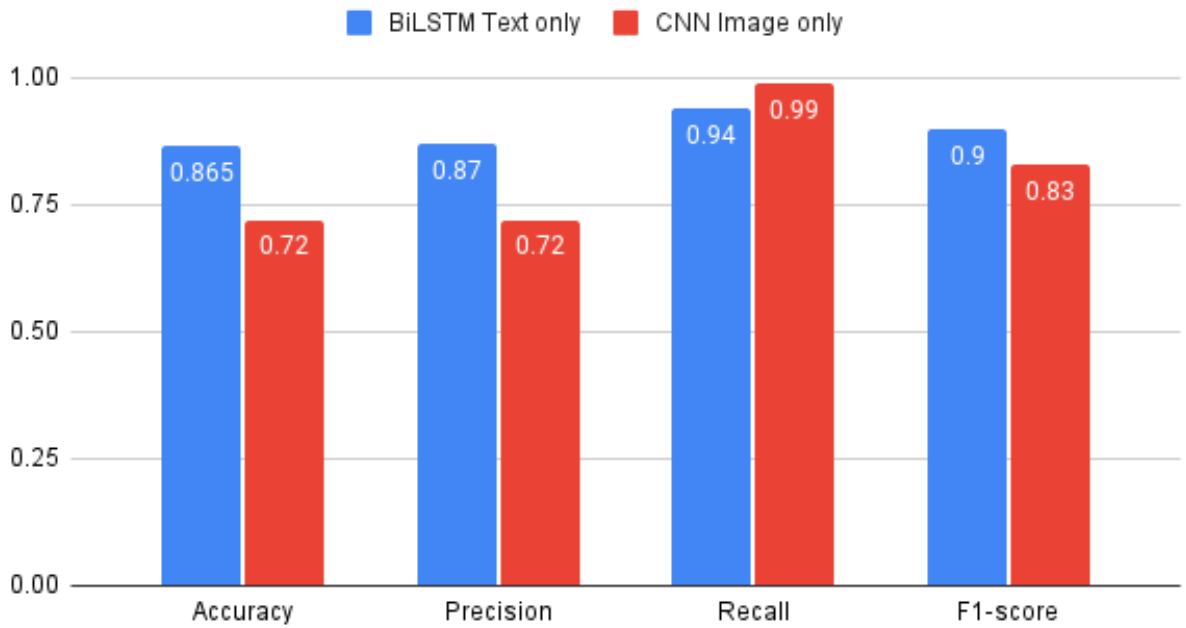


Figure 5.6: Text only Model vs Image only Model Comparison

5.2.2 Multimodal Fake News Detection

CNN: Below in figure 5.7 we can see the performance of CNN multimodal model performance with an accuracy of 72% and f1 score of 82%.

The reported performance metrics of a multimodal CNN model 72% accuracy and 82% F1-score provide a comprehensive evaluation of its effectiveness in integrating both image and text features for predictive tasks. These metrics are indicative of the model's ability to leverage diverse data sources to make accurate predictions. The 72% accuracy signifies that the model correctly predicted outcomes for a substantial majority of instances in the dataset, highlighting its proficiency in handling complex data relationships across modalities. Meanwhile, the F1 score of 82% reflects a robust balance between precision and recall, crucial for tasks where both class identification and minimizing prediction errors are paramount. This suggests that the model performs well in accurately identifying patterns within the data while minimizing false positives and false negatives, essential for applications ranging from image classification to sentiment analysis with multimodal inputs.

These results underscore the model’s suitability for tasks that require holistic data understanding, such as multimodal sentiment analysis or image captioning. By effectively combining image features extracted through CNN layers and text embeddings, the model demonstrates its capability to synthesize information from different domains, enhancing its predictive accuracy. Moreover, the reported metrics serve as benchmarks for further model refinement and comparison against alternative architectures or datasets. This positions the multimodal CNN model as a robust solution in contexts where integrating diverse data types such as images and text is essential for achieving high-performance standards.

While achieving 72% accuracy and 82% F1 score is commendable, ongoing refinement and optimization are crucial for pushing performance boundaries further. Future efforts may focus on fine-tuning model hyperparameters, expanding training datasets to encompass greater diversity, and exploring advanced architectural modifications to enhance the model’s capacity to generalize across different domains and tasks. These endeavors aim to not only improve predictive accuracy but also to broaden the model’s applicability in real-world scenarios where multimodal data integration plays a pivotal role in decision-making and inference processes.

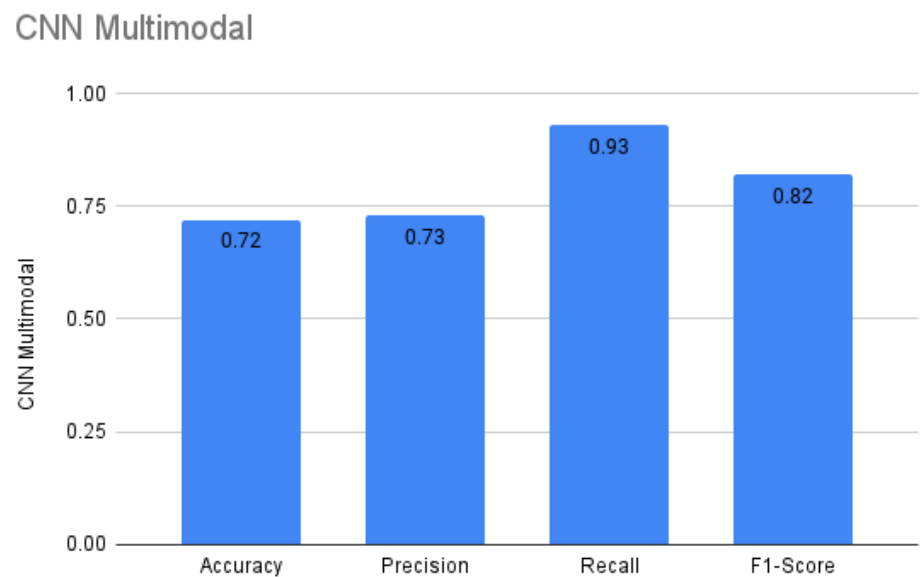


Figure 5.7: Multimodal CNN Model Performance

Although the accuracy of image only and CNN multimodal models is the same, they perform differently in terms of recall and precision. Though it has a little lower precision and a higher recall risk of more false positives, the image-only model is more cautious and captures nearly all positive situations. Though it misses a few true positives than the image-only model, the multimodal model has a better balance between precision and recall, suggesting it is marginally better at identifying false positives.

CNN-BiLSTM We built CNN model for image and BiLSTM for text then concatenated the features. Below in figure 5.8 we can see:

Accuracy (0.9): The model correctly predicts 90% of the instances in the dataset, indicating its overall correctness in classification tasks.

Precision (0.94): Precision measures how accurately the model predicts positive instances. A precision score of 0.94 means that when the model identifies an instance as positive, it is correct 94% of the time.

Recall (0.9): Recall assesses the model's ability to correctly identify all actual positive instances. A recall score of 0.9 indicates that the model successfully captures 90% of all positive instances present in the dataset.

F1-score (0.92): The F1-score is the harmonic mean of precision and recall, providing a balanced measure of the model's performance. With an F1-score of 0.92, the model demonstrates a strong balance between precision and recall.

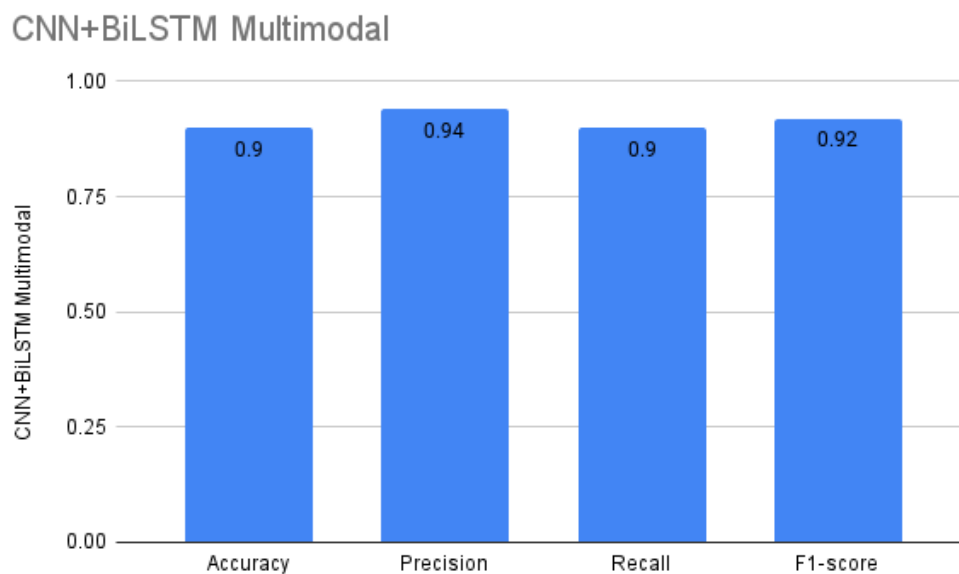


Figure 5.8: Multimodal CNN-BiLSTM Model Performance

Below in figure 5.9 we can see the performance comparison of the proposed model and CNN model for the multimodal fake news detection. CNN-BiLSTM has 92% performance while the CNN was able to achieve 72% accuracy.

CNN Multimodal vs CNN+BiLSTM Multimodal

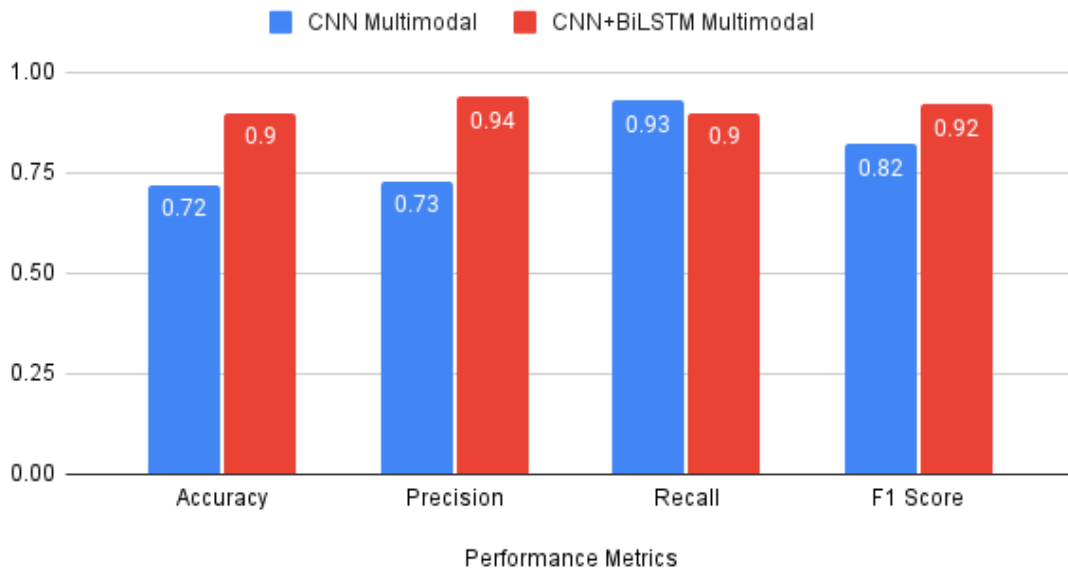


Figure 5.9: CNN vs Proposed model

Finally we compared the unimodal and multimodal models on our dataset. Below in figure 5.10 we can see the performance comparison of the models.

Unimodal vs Multimodal Performance Comparison

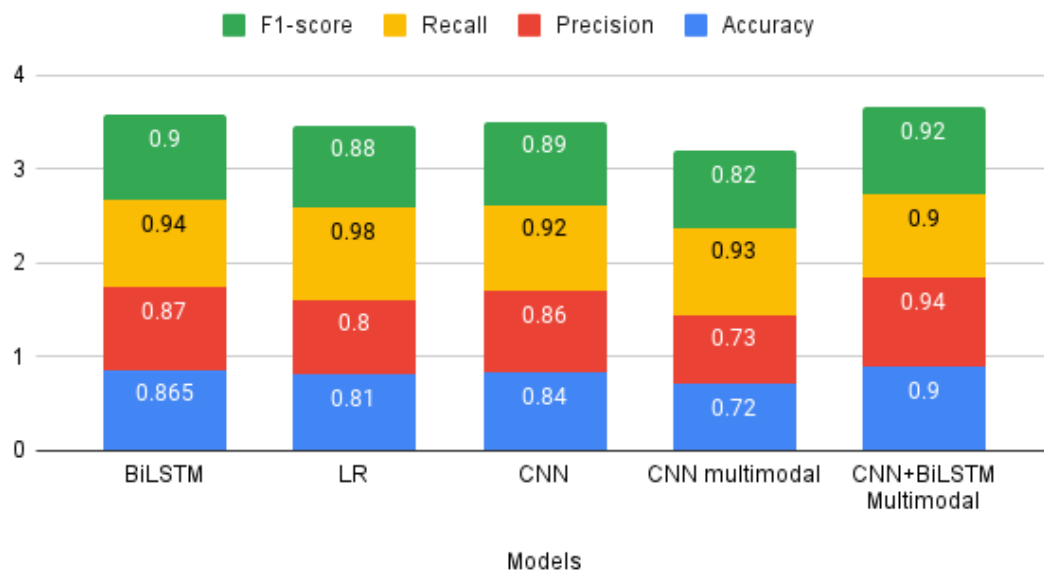


Figure 5.10: Performance Comparison

The multimodal CNN-BiLSTM model outperforms the unimodal version in terms of accuracy (0.90 vs. 0.86) and precision (0.94 vs. 0.87).

The CNN model has lower accuracy and precision compared to the CNN-BiLSTM model, but it achieves a relatively high recall.

In summary, the multimodal CNN-BiLSTM model appears to perform the best overall based on the provided metrics.

Models	Accuracy	Precision	Recall	F1-score
CNN+Fasttext	0.84	0.86	0.92	0.89
CNN+Word2vec	0.78	0.79	0.95	0.86
BiLSTM+Fasttext	0.86	0.87	0.94	0.90
BiLSTM+Word2vec	0.81	0.88	0.84	0.86
LR+Fasttext	0.81	0.80	0.98	0.88
LR+Word2vec	0.76	0.77	0.94	0.85
CNN Multimodal	0.72	0.73	0.93	0.82
CNN+BiLSTM Multimodal	0.90	0.94	0.90	0.92

Table 5.3: Summary of Performance Comparison of the Models

5.3 Discussion

The detection of fake news by multimodal means is currently a hot topic among scholars. Since news is significantly more visually rich than it used to be, research on multimodal news datasets is crucial. As far as we are aware, this is the only study that has tried to investigate this subject using Amharic language.

Subsequently, the primary goal of this study was to determine how incorporating visual elements into the text news dataset improves the ability to identify false news. We created a multimodal news dataset for the Amharic language and trained various deep learning and machine learning models in order to address our first research question.

We first reported on the results of our tests using text-only models. For categorization, these models only used textual data that was taken from news articles. We built different deep learning models for unimodal fake news detection and also we used Machine learning models to compare if the deep learning model has a better performance in text only news dataset. Logistic regression, convolutional neural networks (CNN), and long short-term memory (LSTM) networks were among the many conventional and deep learning techniques that we assessed.

Prior to that, though, we needed to determine which feature extraction for text data technique improved or performed better for the Amharic language. The two most popular embedding methods that were compared in various trained models were Fasttext and Word2vec. As we can see from figure 5.8, Bidirectional Long Short Memory (BiLSTM) performed better according to our performance metrics. The BiLSTM model performs better than the CNN model, showing the importance of sequential information in the task. The CNN model still performs reasonably well but has slightly lower scores compared to BiLSTM. Based on the results, we were able to extract the text feature more effectively on Fasttext, particularly since we trained our own model on a large corpus with less noise. Another reason why Fasttext performed better was that it is best suited for morphologically rich languages, as it studies words at the char level, adding more values by not missing words that are not in the vocabulary.

When it comes to Multimodal fake news detection we compared the state of the art CNN model with our proposed hybrid deep learning model which is CNN-BiLSTM. As we can see from the table which summarize the performance of all the unimodal fake news model and Multimodal models. Even though Khalid et al.[48] stated that image only model performed the best to detect fake news our fake news CNN model had the lowest performance compared to the text only unimodal model. Text perform good at identifying fake news since it contains rich semantic information. However, it lacks visual cues, so we were able to improve our results by adding visual content. Additionally, when we combined picture models with text, we were able to identify modified or deceptive image contents, which ultimately led us to the overall performance on our fake news dataset.

5.4 Summary

RQ1 What is the performance of the proposed multimodal approach in detecting Amharic fake news compared to the existing unimodal approach?

To address this question, we initially trained various deep learning and machine learning models and evaluated their performance on our unimodal dataset. The model with the best performance was then compared to our proposed multimodal fake news detection model, CNN-BiLSTM. The multimodal approach outperformed the unimodal BiLSTM model, achieving **90%** accuracy, **94%** precision, **90%** recall, and a **92%** F1-score, highlighting its superior effectiveness in detecting fake news.

RQ2 What is the most effective feature extraction method for detecting fake news in Amharic text datasets?

Based on our experiments with the Amharic dataset, the **FastText feature extraction** method demonstrated excellent performance. It effectively captures semantic relationships and nuances in the Amharic language, contributing to improved model accuracy and reliability in text classification tasks. Additionally, FastText handles out-of-vocabulary words well, further enhancing its suitability for the Amharic dataset. As shown in Table 5.3, the FastText feature extraction method achieved a performance improvement of more than **5%** on our evaluation metrics values compared to the Word2vec feature extraction method.

Chapter 6

Conclusion and Recommendation

Several significant findings and insights have been made based on a study done on fake news detection for the Amharic language using news content. The dataset was extracted from Facebook and included both text and image. Deep learning techniques were used in the study, namely the combination of CNN and BiLSTM models, which performed better than other approaches. The dataset was thoroughly annotated by three journalists working together on the annotation process, guaranteeing the precision and dependability of the training data. The developed multimodal fake news detection system was evaluated against text-only and image-only (unimodal) approaches. The results showed that the multimodal method outperformed the other approaches, demonstrating the value of combining text and image characteristics for improved detection accuracy.

Additionally, a notable enhancement in performance was shown when comparing the outcomes of the multimodal approach with the most advanced methodologies, highlighting the effectiveness of the suggested methodology in identifying false information within the context of the Amharic language.

In summary, by utilizing deep learning techniques and multimodal data integration, this study has made significant advances to the field of fake news identification, particularly in the context of the Amharic language. The results emphasize how crucial it is to take into account both text and image data in order to have strong detection skills.

It is advised that in order to further improve detection accuracy, future research projects concentrate on enlarging the dataset, improving the annotation procedure, and investigating new modalities. Furthermore, efforts must to be focused on creating user-friendly platforms and tools that make use of the knowledge gathered from this study in order to combat the spread of disinformation within the Amharic-speaking community. For these tools to be successfully implemented and adopted in practice, cooperation with pertinent stakeholders—such as social media platforms and news organizations—is essential. In the long run, we could succeed to build a more informed and robust society in the face of false information and fake news by carrying out more study in this area and fostering partnership.

Bibliography

- [1] The Oxford Pocket Dictionary of Current English. "news". <https://www.encyclopedia.com/humanities/dictionaries-thesauruses-pictures-and-press-releases/news-0>, 2004. accessed December,2023.
- [2] Kai Tai Chan. Emergence of the 'digitalized self' in the age of digitalization. *Computers in Human Behavior Reports*, 6:100191, 2022.
- [3] Nikki Gilliland. 30% of people say social media their main way of getting news. <https://econsultancy.com/reuters-digital-news-report-2023-social-media-versus-direct/>, 2023. accessed December 22,2023.
- [4] Iryna Pentina, Adam Covault, and Monideepa Tarafdar. Exploring the role of social media in news consumption. pages 577–577, 01 2013.
- [5] Damian Tambini. Fake news: public policy responses. *LSE Media Policy Project Series*, 2017.
- [6] Ning Xiang et al. Deep learning-based fake information detection and influence evaluation. *Computational Intelligence and Neuroscience*, 2022, 2022.
- [7] Webwise.ie. What is false information (fake news)? <https://www.webwise.ie/teachers/what-is-fake-news>, 2022. accessed December,2022.
- [8] Wubetu Barud Demilie and Ayodeji Olalekan Salau. Detection of fake news and hate speech for ethiopian languages: a systematic review of the approaches. *Journal of big Data*, 9(1):66, 2022.
- [9] Elias Meseret. Hate speech and disinformation concerns escalate in ethiopia. <https://www.devex.com/news/hate-speech-and-disinformation-concerns-escalate-in-ethiopia-97095>, 2020. accessed September,2022.
- [10] ERMIA NIGATU Hailemichael and N Ermias. Fake news detection for amharic language using deep learning. *Unpublished*, (2021):2–100, 2021.
- [11] Walta Info. Facebook expands third-party fact-checking to ethiopia, more african countries. <https://waltainfo.com/42502>, 2019. accessed September,2022.

- [12] Kedir Lemma Arega. Classification and detection of amharic language fake news on social media using machine learning approach. *Electrical Science & Engineering*, 4(1):1–6, 2022.
- [13] Tewodros Tazeze and Raghavendra Ramesh. Building a dataset for detecting fake news in amharic language. *International Journal of Advanced Research in Science, Communication and Technology*, pages 76–83, 06 2021.
- [14] Menbere Hailu Worku and Michael Melese Woldeyohannis. 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*, pages 468–479. Springer, 2022.
- [15] Okuhle Ngada and Bertram Haskins. Fake news detection using content-based features and machine learning. In *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, pages 1–6. IEEE, 2020.
- [16] Santiago Alonso-Bartolome and Isabel Segura-Bedmar. Multimodal fake news detection. *arXiv preprint arXiv:2112.04831*, 2021.
- [17] Shivangi Singhal, Rajiv Ratn Shah, Tanmoy Chakraborty, Ponnurangam Kumaraguru, and Shin’ichi Satoh. Spotfake: A multi-modal framework for fake news detection. In *2019 IEEE fifth international conference on multimedia big data (BigMM)*, pages 39–47. IEEE, 2019.
- [18] Sabrine Amri, Dorsaf Sallami, and Esma Aïmeur. Exmulf: An explainable multimodal content-based fake news detection system. In Esma Aïmeur, Maryline Laurent, Reda Yaich, Benoît Dupont, and Joaquin Garcia-Alfaro, editors, *Foundations and Practice of Security*, pages 177–187, Cham, 2022. Springer International Publishing.
- [19] Limeng Cui, Suhang Wang, and Dongwon Lee. Same: Sentiment-aware multimodal embedding for detecting fake news. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM ’19*, page 41–48, New York, NY, USA, 2020. Association for Computing Machinery.
- [20] Anastasia Giachanou, Guobiao Zhang, and Paolo Rosso. Multimodal multi-image fake news detection. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 647–654, 2020.

- [21] Toni G.L.A. Van Der Meer Michael Hameleers, Thomas E. Powell and Lieke Bos. A picture paints a thousand lies? the effects and mechanisms of multimodal disinformation and rebuttals disseminated via social media. *Political Communication*, 37(2):281–301, 2020.
- [22] Jawaher Alghamdi, Yuqing Lin, and Suhuai Luo. Fake news detection in low-resource languages: A novel hybrid summarization approach. *Knowledge-Based Systems*, 296:111884, 2024.
- [23] Sakshini Hangloo and Bhavna Arora. Combating multimodal fake news on social media: methods, datasets, and future perspective. *Multimedia systems*, 28(6):2391–2422, 2022.
- [24] Julius Endert. Ethiopia’s uneasy relationship with press freedom: Fighting misinformation at the highest level. <https://akademie.dw.com/en/ethiopias-uneasy-relationship-with-press-freedom-fighting-misinformation-at-the-highest-level/a-54926467>, 2020. accessed November,2022.
- [25] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (ICET)*, pages 1–6. Ieee, 2017.
- [26] Eduri Raja, Badal Soni, and Samir Kumar Borgohain. Fake news detection in dravidian languages using transfer learning with adaptive finetuning. *Engineering Applications of Artificial Intelligence*, 126:106877, 2023.
- [27] Muna Shifa. The interaction of mass media and social media in fuelling ethnic violence in ethiopia. <https://www.accord.org.za/conflict-trends/the-interaction-of-mass-media-and-social-media-in-fuelling-ethnic-violence-in-ethiopia>, 2022. accessed September,2023.
- [28] District of Columbia. District of columbia language access act fact sheet 2004. <https://ohr.dc.gov/publication/know-your-rights-cards-amharic>, 2004. accessed September,2023.
- [29] European Institute of Peace. Fake news misinformation and hate speech in eth iopia: A vulnerability assessment. 2021.
- [30] Fantahun Gereme, William Zhu, Tewodros Ayall, and Dagmawi Alemu. Combating fake news in “low-resource” languages: Amharic fake news detection accompanied by resource crafting. *Information*, 12(1):20, 2021.

- [31] Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M Bronstein. Fake news detection on social media using geometric deep learning. *arXiv preprint arXiv:1902.06673*, 2019.
- [32] Muhammad Umer, Zainab Imtiaz, Saleem Ullah, Arif Mehmood, Gyu Sang Choi, and Byung-Won On. Fake news stance detection using deep learning architecture (cnn-lstm). *IEEE Access*, 8:156695–156706, 2020.
- [33] Yang Yang, Lei Zheng, Jiawei Zhang, Qingcai Cui, Zhoujun Li, and Philip S Yu. Ti-cnn: Convolutional neural networks for fake news detection. *arXiv preprint arXiv:1806.00749*, 2018.
- [34] M. F. Mridha, Ashfia Jannat Keya, Md. Abdul Hamid, Muhammad Mostafa Monowar, and Md. Saifur Rahman. A comprehensive review on fake news detection with deep learning. *IEEE Access*, 9:156151–156170, 2021.
- [35] Francisval Guedes soares. Bank marketing campaign result prediction using neural networks, 2024. Accessed: 2024-06-17.
- [36] Mudasir Ali, Mobeen Shahroz, Muhammad Mushtaq, Sultan Alfarhood, Mejdil Safran, and Imran Ashraf. Hybrid machine learning model for efficient botnet attack detection in iot environment. *IEEE Access*, PP:1–1, 01 2024.
- [37] Niina Pakarinen. Lecture 5: Neural language models, 2024. Accessed: 2024-06-17.
- [38] Mohammad Mustafa Taye. Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. *Computation*, 11(3), 2023.
- [39] B Obermaier, Christoph Guger, Christa Neuper, and Gert Pfurtscheller. Hidden markov models for online classification of single trial eeg data. *Pattern Recognition Letters*, 22:1299–1309, 10 2001.
- [40] Ibrahim Awol and Sosina Gashaw. Lexicon-stance based amharic fake news detection. 03 2023.
- [41] GURMESSA Daraje Kaba, Salau Ayodeji Olalekan, and Gedefa Asrat. 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, 2022.
- [42] Worku Kelemework. Automatic amharic text news classification: Aneural networks approach. *Ethiopian Journal of Science and Technology*, 6(2):127–137, 2013.

- [43] Tomas Mikolov, Ilya Sutskever, Kai Chen, G.s Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26, 10 2013.
- [44] Sima Siامي-Namini, Neda Tavakoli, and Akbar Siامي Namin. The performance of lstm and bilstm in forecasting time series. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 3285–3292, 2019.
- [45] Davi Guimarães da Silva and Anderson Alvarenga de Moura Meneses. Comparing long short-term memory (lstm) and bidirectional lstm deep neural networks for power consumption prediction. *Energy Reports*, 10:3315–3334, 2023.
- [46] Yufis Azhar, M. Anugerah, Muhammad Fahlopy, and Alfin Yusriansyah. Image captioning using hybrid of vgg16 and bidirectional lstm model. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 11 2022.
- [47] Ramesh Kalyan G. Convolution neural networks, 2024. Accessed: 2024-06-17.
- [48] Yasameen Khalid and Nasrolah Charkari. Image fake news detection using efficient netb0 model. *Journal of Information Systems and Telecommunication (JIST)*, 12:41–48, 04 2024.