Deep Learning Multi-head Attention Based Fake News Detection for the Amharic Language

By

Beur Admasu Gaga

A Thesis Submitted to the Department of Computer science and Engineering,

School of Electrical Engineering and Computing

Presented in Partial Fulfillment for the Degree of Masters of Science in Computer Science and Engineering

Office of Graduate Studies

Adama Science and Technology University

August 2021

Adama, Ethiopia

Deep Learning Multi-head Attention Based Fake News Detection for the Amharic Language

By: Beur Admasu Gaga

Advisor: Dr. Teklu Urgessa (PHD.)

A Thesis Submitted to the Department of Computer science and Engineering,

School of Electrical Engineering and Computing

Presented in Partial Fulfillment for the Degree of Masters of Science in Computer Science and Engineering

Office of Graduate Studies

Adama Science and Technology University

August 2021

Adama, Ethiopia

APPROVAL SHEET

The advisor of the thesis entitled "**Deep Learning Multi-head Attention Based Fake News Detection for the Amharic Language" and developed by Beur Admasu Gaga.** Hear by certifying that the recommendation and Suggestions made by the board of examiners are appropriately incorporated into the final version of the thesis

Teklu Urgessa (Ph.D.)	_____	_____
Advisor	Signature	Date

We, the undersigned, members of the board of Examiners of the thesis by **Beur Admasu Gaga** Have read and evaluated the thesis entitled "**Deep Learning Multi-head Attention Based Fake News Detection for the Amharic Language"** and examined the candidate during the open defense. This is, therefore, to certify that the thesis is accepted for partial fulfillment of the requirement of the degree of Masters of Science in Computer Science and Engineering.

_____	_____	_____
Chairperson	Signature	Date
_____	_____	_____
Internal Examiner	Signature	Date
_____	_____	_____
External Examiner	Signature	Date

Final approval and acceptance of the thesis are contingent upon submission of its Final copy to the office of postgraduate Studies (**OPGS**) through the Department Graduate Council (**DGC**) and School Graduate Committee (**SGC**)

_____	_____	_____
Department Head	Signature	Date
_____	_____	_____
School Dean	Signature	Date
Legesse Lemeche Obsu (PH.D.)	_____	_____
Office of postgraduate studies. Dean	Signature	Date

DECLARATION

I hereby declare that this MSc Thesis is my original work and has not been presented for a degree in any other university, and all sources of material used for this thesis have been properly acknowledged.

Beur Admasu Gaga	_____	_____
Name Student	Signature	Date

RECOMMENDATION

I, the advisor of this thesis, hereby certify that we have read the revised version of the thesis entitled "**Deep Learning Multi-head Attention Based Fake News Detection for the Amharic Language**" prepared under my/our guidance by Beur Admasu Gaga Submitted in partial Fulfillment of the requirement for the degree of master's degree of Computer Science and Engineering Therefore, I recommend submitting the revised of the thesis to the department following the applicable procedures.

Teklu Urgessa (Ph.D.)	_____	_____
Advisor	Signature	Date

ACKNOWLEDGEMENT

First and above all, I would like to thank to the Almighty of God and St. Marry whose eternal and undying love kept me and for infinite mercy throughout my life and during this thesis work. Next I would like to thank my advisor Dr. Teklu Urgessa for his effortless support and valuable comments throughout this thesis work, his support and encouragement gave me a great inspiration to finish this work. Next, I would like to thank Mr. Biruk Arega a journalist in EBC who helped me in the data collection process and intelligent advice about the proposed topic.

Finally, I would like to thank my beloved mother Ayelech Fufa who sacrificed a lot for me and our family, even though she couldn't make it in the last moment she deserved all the credit and to all my family members, friends and classmates without them this wouldn't be possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

API----------Application Program Interface

Bi-LSTM----------Bi-Directional Long Short-Term Memory

Bow----------Bag of Words

CBOW----------Continuous Bag of Words

CNN----------Convolutional Neural Network

Glove----------Global Vectors for Word Representation

LSTM----------Long Short-Term Memory

MLP----------Multi-Layer Perceptron

NLP----------Natural Language Processing

RNN---------- Recurrent Neural Network

SVM----------Support Vector Machine

TFIDF----------Term Frequency Inverse Document frequency

Word2vec----------Word to Vector

ABSTRACT

Due to the increasing of social media usage, it has become necessary to combat the spread of false information and decrease the reliance on information retrieval from such sources. Social platforms are under constant pressure to come up with an efficient method to solve this problem because users' interaction with fake and unreliable news leads to its spread at an individual level. This spreading of misinformation adversely affects the perception of important activity, and as such, it needs to be dealt with using a modern approach. This research *presents a deep learning based fake news detection using a multi-head attention for Amharic language. This* research *proposes multi head attention-based BI-LSTM, ensemble of CNN-LSTM, and LSTM which are state-of-the-art and NLP approaches. The goal of this research is to analyze Amharic text and classify news taken from social media as fake or real. To evaluate the news classification system, this research* collects *about 2400 news from different social and Main stream Media. This research also tries to investigate the effect of removing Amharic language stop words during preprocessing stage on the Amharic language fake news detection process. The proposed multi-head attention model outperforms other deep learning models such as BI-LSTM, CNN, LSTM, and CNN-LSTM by 7%, 7%, 16%, and 7% respectively. The BI-LSTM, CNN, CNN-LSTM produces almost equal classification accuracy. Results shows that applying stop words drop the accuracy of the fake news detection by 2.83% while examine on propose attention-based.*

***Keywords*:** *Amharic Fake news detection, Natural Language Processing (NLP), Stop word, Self-Attention Model*

CHAPTER ONE
INTRODUCTION

1.1Background of the Study

Internet and social media allow access to news information much easier and comfortable. Often Internet users can follow the events of their interest in online mode, and the spread of mobile devices and Smartphones makes this process even easier [1]. The importance and use of these devices and platforms are enormous. However, this enormous potential becomes a greater threat if it misleads users with false information. As an increasing amount of our lives interacting, online through social media and platforms, people tend to seek out and consume news from these online media rather than traditional news organizations [2]. Many reasons make people to be attracted to these platforms: (i) it can be accessed using simple and handheld devices like smartphones, tablets, and PCs. (ii) it is very easy to write news, share, comment, and discuss any issue and also it is very cost-effective.

The spread of fabricated news results in an instability of a country. For instance, in the past 5 years, Ethiopia has been unsettled due to the spread of information and news through these media made the political revolution which boosts antigovernment protests to achieve their goals faster than expected. These dangers come after the spread of fake, fabricated, manipulated, and imposter news divides the society, and currently, Ethiopia is in danger more than ever to stand as a nation.

In recent years, to help online users identify important and valuable information, there has been extensive research on establishing an effective and automatic framework for online fake news detection. However, Fake news detection on social media presents unique characteristics and challenges that make the detection challenge. Some of these challenges include: (i) Identifying credible social information from millions of messages is challenging, due to the heterogeneous and dynamic nature of online social communication. More specifically, it is difficult to distinguish online truthful signals from fake and anomalous information, since the fake news is intentionally written to mislead readers [2]. (ii) The linguistic-based features extracted from the news content are not sufficient for revealing the in-depth underlying distribution patterns of fake news [2]. Auxiliary features such as the credibility of the news author and the spreading patterns of the news play more important roles for online fake news detection. However, exploiting this

auxiliary information is challenging in and of itself as users' social engagements with fake news produce data that is big, incomplete, unstructured, and noisy. (iii) Online social data is time-sensitive, which means that they occur in a real-time pattern, and represent the trending topics and events.

In Ethiopia, there has been no means of detecting fake news that is fleeing freely without any restrictions, and there should be some mechanism to detect this fake news. This research proposes efficient Amharic language fake news detection.

1.2 Motivation

Over ongoing years, the quick and dangerous improvement of online news stages has seen broad development in the quantity of phony news. These days, counterfeit news is irritating, prominent, diverting, and everywhere. It significantly affects the two people and society. So it is critical for building a viable discovery framework for counterfeit news ID. The fundamental attributes of phony news can be summed up as follows:

The volume of false news: With no confirmation methodology, everybody can undoubtedly compose counterfeit news on the computerized world utilizing various stages. These individuals and associations are intentionally made to disseminate deceptions, purposeful publicity, and disinformation, regularly for monetary or political addition. Hence, a gigantic measure of phony substance is dispersed through the media, even without clients' mindfulness.

The assortment of phony news: There are a few close meanings of phony news, like reports, parody news, counterfeit audits, deception, counterfeit promotions, fear inspired notions, bogus explanation by legislators, and so on, which influence each part of individuals' lives. With the expanding fame of web-based media, counterfeit news can rule the popular sentiments, interests, and choices. Moreover, counterfeit news changes the way that individuals collaborate with genuine news. Some phony news is made deliberately to deceive and befuddle clients, particularly youthful understudies and individuals who have no expectation in light of a social and financial issue (Forbes.com). We can see from the current circumstance of Ethiopia that online phony news is significant and expansive into each part of our day-by-day life.

The following points are the key motivations for this research work

- Counterfeit news via web-based media has been happening for quite a while; notwithstanding, there is no endless supply of the term counterfeit news". To all the more

likely guide the future headings of phony news location research, suitable explanations are fundamental.

- Web-based media has ended up being an amazing hotspot for counterfeit news dispersal. There are some arising designs that can be used for counterfeit news recognition in web-based media. An audit of existing phony news location techniques under different web-based media situations can give a fundamental comprehension of the cutting-edge counterfeit news recognition strategies.
- Counterfeit news location via online media is as yet in the early time of advancement, and there are as yet many testing issues that need further examination. It is important to examine potential exploration bearings that can improve counterfeit news identification and relief capacities.

1.3Statement of the Problem

In the current digital world, information sharing becomes through different platforms but not everything in these platforms the fake news is indeed the most challenging issue for the public and government. Fake news can destroy many things if it is not controlled. One of the best examples is the Arab revolution which spread within just a few months and destroyed Libya and other Arab countries. This fake news can lead to negative effects or even manipulation of public events. One of the unique challenges of detecting fake news is how to identify fake news about recent events in the Amharic language. The task of detecting fake news has tested a variety of labels from misinformation to rumors; to spam. There has been a large body of work surrounding text analysis of fake news and similar topics such as rumors or spam. Language in Ethiopia especially the Amharic language has characters (fields) more than 380 Unicode representations (U+1200-U+137F) this language is morphologically complex. There are two directions to detect text deception. The first of which is based on separate handheld features, which can capture linguistic and psychological causes, however, these features failed to classify text well, which limits performance. The second approach is based on a convolutional neural network model that learns document-level representation to discover deception text. Neural network models have been used to learn semantic representations for NLP tasks and it reaches the highest competitive result According to [3] and also in NLP. But the most difficult thing is datasets which is the reason why news detection fake was not successful in the past because it was small and included

unrealistic news. One of the gaps that will be addressed in this research is multilingualism for detecting fake news.

1.4 Research Questions

This research on Amharic fake news detection addresses the following research questions.

RQ1 [performance enhancement]; Can Multi-head attention model improve the performance of Amharic fake news detection?

RQ2 [Annotation]: How to identify fake news that is published in the Amharic language?

RQ3 [Effect of stop word removal]: What is the effect of stop word removal on the Amharic language fake news detection?

1.5 Objectives

1.5.1 General Objectives

The general objective of this research is to develop fake news detection mechanisms for Amharic languages using state of the art approach.

1.5.2 Specific Objectives

The following specific objectives are identified to realize the general objective:

- Proposing an efficient way of detecting fake news that is posted in the Amharic language
- Preparing a fake news dataset for Amharic language.
- Suggesting the best fake news detection mechanism.
- Experiment on the performance of detecting fake news.
- Comparing the performance of NLP, deep learning mechanisms for detecting fake news.

1.6 Scope and Limitation

Fake news is very difficult and sophisticated problem to tackle, as of nowadays because even the term fake has a controversial definition because to identify news content whether it is fake or not is unconditional. This research is mainly concerned with fake news detection on local languages which addresses the current research gap but the lack and shortage of data set on the languages to be used is one of the limitations of the research.

1.7Contributions

This research ***will have a contribution*** to the analysis of Amharic fake news review by classifying the news as fake or real. This **research** uses Long Short-Term memory based fake news detection with multi head attention mechanism that can handle the context of a text. This research work has the following contributions

- Performing analysis to identify the gaps on the previous developed fake news detection system.
- This research work collects Amharic news data set for future researches.
- Comparing the state-of-the-art deep learning approaches to identify the best approach for Amharic language.

1.8 Application of Results

The spread of fake news has a cynical impact on our society in the past decade, it is clear that the impact it's putting in the norms and how we observe information. The development of fake news detection mechanisms is vital for the entire digital community and also for future researchers can be benefited from the datasets by improving into large data corpus it can be achieved a more accurate detection mechanism.

1.9 Thesis Organization

The rest of this document is organized as follows. **Chapter Two** discusses a general overview of Amharic Language, challenges in developing Amharic fake news detection, and how the LSTM and Word2vec algorithm works.

Chapter Three discuss previous research works on both resource-rich and Amharic language state-of-the-art works for fake news detection. Our proposed work for Amharic fake news detection will be elaborated on in **Chapter four** of the document. The experimental setup, experiments evaluation metrics, results, discussions on results are discussed in **Chapter five** of the document. Finally, **Chapter six** discusses the conclusion and future research directions on Amharic fake news.

CHAPTER TWO

LITERATURE REVIEW AND RELATED WORKS

2. 1 Fake News Analysis Approaches

Fake news analysis is a process of developing an automated fake news detection system from sources like unstructured text, images/videos, and references to other sources. There are many tools and techniques to develop an automated fake news detection system from user generated data. Fake news detection approaches may use the following possible solutions and techniques.

(i) Gathering and indexing of information published in the internet in order to cross-reference current news with the previous one. (ii) Image processing techniques for image comparison and content analysis. (iii) Text analysis techniques, Reputation scoring (to identify reliability of person and/or information source) providing the news, Reputation of the content. (iv) Webpage providing/forwarding news (v) Machine learning techniques for content features analysis and Analysis of semantics by means of applied ontology, and Comparison, of the similar news published by different information sources –in that case we can use common data mining techniques for text analysis [4]. Figure 2.1 shows suggested approaches for fake news detection approach.

Hereby, in this paper they focus on image processing techniques for photo content verification, as presented in the next sections.

Many machine learning methods are proposed and studied to address the problem of fake news. These are classified by feature extraction techniques and model extraction techniques. In this research, the methods used are NLP, Naïve bays, CNN, and deep learning methods, and compare the results that are conducted in the experiment and come up with the best method to achieve the objectives.

2.1.1 Machine Learning Based Fake News Analysis

Fabricated news can be detected using natural language processing (NLP) techniques by applying term frequency-inverse document frequency (TF-IDF) of bi-grams and probabilistic context-free grammar detection to news articles [5]. The dataset was tested on various classification algorithms such as support vector machines (SVMs), stochastic gradient descent,

gradient boosting, bounded decision trees, and random forests. The algorithm was found that TF-IDF of bi-grams fed into a Stochastic Gradient Descent model results in an accuracy of 77.2%.

Fake news detection using naive Bayes classifier that achieved a classification accuracy of approximately 74%, and whose results suggest that fake news detection problem can be addressed with artificial intelligence methods [6].

2.1.2 Deep Neural Networks Based Fake News Analysis

The model that integrates capture (C), score(S) and integrates (I) modules combine the three generally agreed upon characteristics of fake news such as the text of an article, the user response it receives, and the source users promoting it, for a more accurate and automated prediction (). The first module is based on the response and text that employs the use of a recurrent neural network (RNN) to capture the temporal pattern of user activity. The second one learns the source characteristic based on the users' behavior, and then the third module integrates both of them to classify whether an article is fake or not [7].

Gilda, S., and et al develop a deep learning-based fake news detection using different word representations approach such as TF-IDF on unigrams and bigrams with cosine similarity fed into a dense neural network, BoW without unigrams and bigrams with cosine similarity fed into a dense neural network, and Pre-trained embeddings (Word2Vec) fed into dense neural network [8]. Tf-Idf word vector representations when propagated through a dense neural network give an accuracy of 94.21% which outperform existing model architectures that give 89.23%, 75.67% respectively. The Word2Vec based latent space representation is not able to capture the word semantic level importance if the news article length is very large [8].

2.1.3 Image Processing Techniques for Fake News Detection

Photo and video content is an essential part of media coverage, often more attracting a content consumer’s attraction than information provided as a text. Image processing is become popular due to the development of specific software tools [4]. Fake photos can be linked to both factual (true) news and totally faked information shared online. These fake photos might be posted due to propaganda reasons, but also for strengthening the intended message and for inducing emotions.

Michal Chora´s and et al developed Image processing techniques for verification of photo content and detection of fake images [4]. The paper employed 800 unmodified images (original) and 921 images with additional elements added. The paper employs the SURF algorithm to extract features from blocks, and the FLANN algorithm for matching.

2.2 Fake news analysis

In recent years, fake news (or rumor, misinformation) detection has gained particular attention in the literature. A major track of existing studies aims at developing machine learning-based classifiers to automatically determine whether a news story spreading in a social media environment is fake based on a variety of news characteristics. A few early studies try to detect fake news based on linguistic features extracted from the text content of news stories. [10], [11] they proposed neural network architecture to accurately predict the stance between a given pair of a headline. They have used different model variations like TF-IDF on unigrams and bigrams with cosine similarity fed into a dense neural network and get an accuracy of 94.31%. Using a finely tuned TF-IDF – Dense neural network (DNN) model, they were able to outperform existing model architectures by 2.5% and they were able to achieve an accuracy of 94.21% on test data. Sherry Girgis et al,2018 studied and used a purely deep learning perspective by RNN technique models (vanilla, GRU) and LSTMs. They have shown the difference and analysis of results by applying them to the dataset that was used called LAIR.

It is clear in this study that the results they found are close, but the GRU is the best of the results that reached (0.217) followed by LSTM (0.2166) and finally comes vanilla (0.215). Due to these results, they increase accuracy by applying a hybrid model between the GRU and CNN techniques on the same data set and reached a better result. Monther Aldwairi and Ali Alwahed [12] come up with a solution that utilizes users to detect and filter out sites containing false and misleading information. they use simple and carefully selected features of the title and post to accurately identify fake posts. The experimental result they have shown has 99.4% accuracy and they have used four classifiers: **Bayes Net:** Bayes network learning using various search algorithms and quality measures. Bayes Network classifier provides data structures such as network structure, conditional probability distributions, etc., and facilities common to Bayes network learning algorithms such as *K*2 and *B*. **Logistic:** Class for building and using a multinomial logistic regression model with a ridge estimator. **Random Tree:** Class for

constructing a tree that considers *K* randomly chosen attributes at each node. It performs no pruning and has an option to allow estimation of class probabilities (or target mean in the regression case) based on a hold-out set (back fitting). **Naïve Bayes:** Class for a Naive Bayes classifier using estimator classes. They compared these classifiers using Precision, Recall, F-Measure, and, ROC. And as their result suggest logistic classifier has the highest precision, 99.4% and therefore the best classification quality they have suggested is Logistic and Random Tree classifiers had the best recall that is the best sensitivity of 99.3%. The measure combines precision and recall, the Logistic and Random Tree classifiers outperformed other sat 99.3%. Kelly and et al performs a research that considers previous and current methods for fake news detection in textual formats while detailing how and why fake news exists in the first place [13]. The paper discusses on Linguistic Cue and Network Analysis approaches and proposed a three-part method using Naïve Bayes Classifier, Support Vector Machines, and Semantic Analysis as an accurate way to detect fake news on social media. Arjun Roy et al have developed various deep learning models for detecting fake news and classified them into the pre-defined fine-grained categories at first, they have developed models based on Convolutional Neural Network (CNN) and Bi-directional Long Short Term Memory (Bi-LSTM) networks [14]. The representations obtained from these two models are fed into a Multi-Layer Perceptron Model (MLP) for the final classification. Their experiments on a benchmark dataset show promising results with an overall accuracy of 44.87%, which a high accuracy at that time.

2.3 Fake News Analysis for Amharic Language

Fantahun Gereme and et al develop fake news detection for the Amharic language. The author collected ETH_FAKE: a novel Amharic fake news detection dataset with fine-grained labels, which was collected from various multi-domain sources. Considering the lack of quality Amharic word embedding, the paper has prepared AMFTWE: Amharic fast text word embedding with sub-word information [15]. The developed fake news detection model performed very well using word embedding, cc_am_300 and AMFTWE. The model exhibited a higher performance when AMFTWE was used compared to cc_am_300, which could be attributed to the quality of our word embedding, which was trained on a relatively huge corpus [15]. While using the 300- and 200-dimension embedding, the developed model scored

validation accuracy above 99%. This good performance could be attributed to both the fake news detection model and the quality of the word embedding.

2.4 Challenges in Developing Fake News

"Fake News" is a term used to represent fabricated news comprising misinformation communicated through traditional media channels like print, and television as well as non-traditional media channels like social media. The general motive to spread such news is to mislead the readers, damage the reputation of any entity, or gain from sensationalism [8]. Fake news is greatly being shared via social media platforms like Twitter and Facebook. These platforms offer a setting for the general population to share their opinions and views in a raw and unedited fashion [8].

Technologies such as Artificial Intelligence (AI) and Natural Language Processing (NLP) tools offer great promise for researchers to build systems that could automatically detect fake news [8]. However, the classification of fake news is challenging due to its vague definition and tensions related to freedom of speech [1]. Even trusted news agencies that are readily available in content production systems, e.g. Associated Press, are not easy to verify [4]. In addition, comparing proposed news with the original news itself is a daunting task as it's highly subjective and opinionated [8].

2.4.1 Challenges in Developing Amharic Fake News Detection

Amharic (አማርኛ, Amarəññа), with the only African-origin script named Ethiopic/Fidel, is the second most spoken Semitic language in the World, next to Arabic, and it is the official working language of the Ethiopian government [15]. As more Ethiopians are living outside their home, the Amharic language speakers in different countries of the world are also growing. In Washington DC, Amharic has gotten the status as one of the six non-English working languages [15]. Despite this, the Amharic language is one of the "low-resource" languages in the world, which lacks the tools and resources important for NLP (natural language processing) and other techno-linguistic [15].

Linguistic resources play crucial roles in the development of fake news detection tools. However, "low-resource" languages, mostly African languages, such as Amharic, lack such resources and tools [15]. To the best of our knowledge, for the Amharic language, there are no

fake news detection datasets and we find only one work done to detect fake news written in the Amharic language. In addition, there is a lack of quality of Amharic word embedding. The available Amharic corpora are not sufficient and some of them are not freely open to the public. This is a total disadvantage, not being able to benefit from technology solutions [15].

The CNN models are widely used for image classification and detection purposes while also being used for sentence classification [16].

Different types of ensemble methods are bagging, boosting, and stacking. Another factor we use is the attention mechanism. In recent years, attention has emerged out as a widely used and important tool in field of deep learning [17]. Attention can be defined as a vector derived as output of dense layer of network using the **soft max** function. Before attention, one needed to compress all the input information into a fixed length vector. However, a sentence with large number of words would result in loss of important information. Attention partially fixes this problem by helping the network in learning as to where it should pay attention in the input sequence for each item in the output sequence.

2.4.2 Summary of Related Works

Table 2.1 Summary of Related Work

Authors & year	Title	Method and algorithm	Result
Granik, M., & Mesyura, V. [6]	*Fake news detection using naive Bayes Classifier*	Naive Bayes classifier	Achieved an accuracy of 74%.
Aswini Thota [8].	*Deep learning-based fake news detection using different word representations G*	Tf-Idf word vector representations	Word2Vec achieved an accuracy of 94.21%.
A. Thota, P. Tilak, S. Ahluwalia, N. Lohia, S. Ahluwalia, and N. Lohia, [10]	"*Fake News Detection: A Deep Learning Approach,"*	TF-IDF on unigrams and bigrams with cosine similarity fed into a dense neural	TF-IDF on unigrams and bigrams achieved an accuracy of 94.31%, and on finely tuned TF-

		network, Using a finely tuned TF-IDF – Dense neural network (DNN)	IDF-DNN achieved an accuracy of 94.21%.
Fantahun Gereme [15].	*"Low-Resource" Languages: Amharic Fake News Detection Accompanied by Resource Crafting. Information.*	Word embedding, cc_am_300 and AMFTWE.	Using the 300- and 200-dimension of the proposed model they achieved an accuracy of 99%.

CHAPTER THREE
RESEARCH METHODOLOGIES

3.1 Research Methodology

The key goal of this chapter is built the methodology to carry out the study. This study followed experimental research design and applied different methods starting from problem identification, define the objectives, Design and prototype development, testing and evaluating the performance. This is the procedure used in evaluating the various predictive Deep learning approach and algorithms using the imported custom data sets from social media. The chapter also deals with the Explanation of the software tool used to apply various algorithms to data preprocessing, feature selection classification, and prediction. Also observation and analysis of Amharic language raw data, back ground and limitation case study natural language processing outlying done.

In this research iteration requires several data collection instruments or strategies that range in complexity, design, etc. interpretation, and administration; each technique is suitable to get certain type of information [18]. Data collection techniques such as interviews, questionnaire, direct observation, and document analysis have been used in different studies. Those common approaches help us to collect sufficient and effective data sources. Mostly collected from public service provider, organizations, digital platforms, are the popular once. [18] [19] [20] For this study data collected from Facebook API. Figure 3.1 shows the overall research methodology followed in this research work.

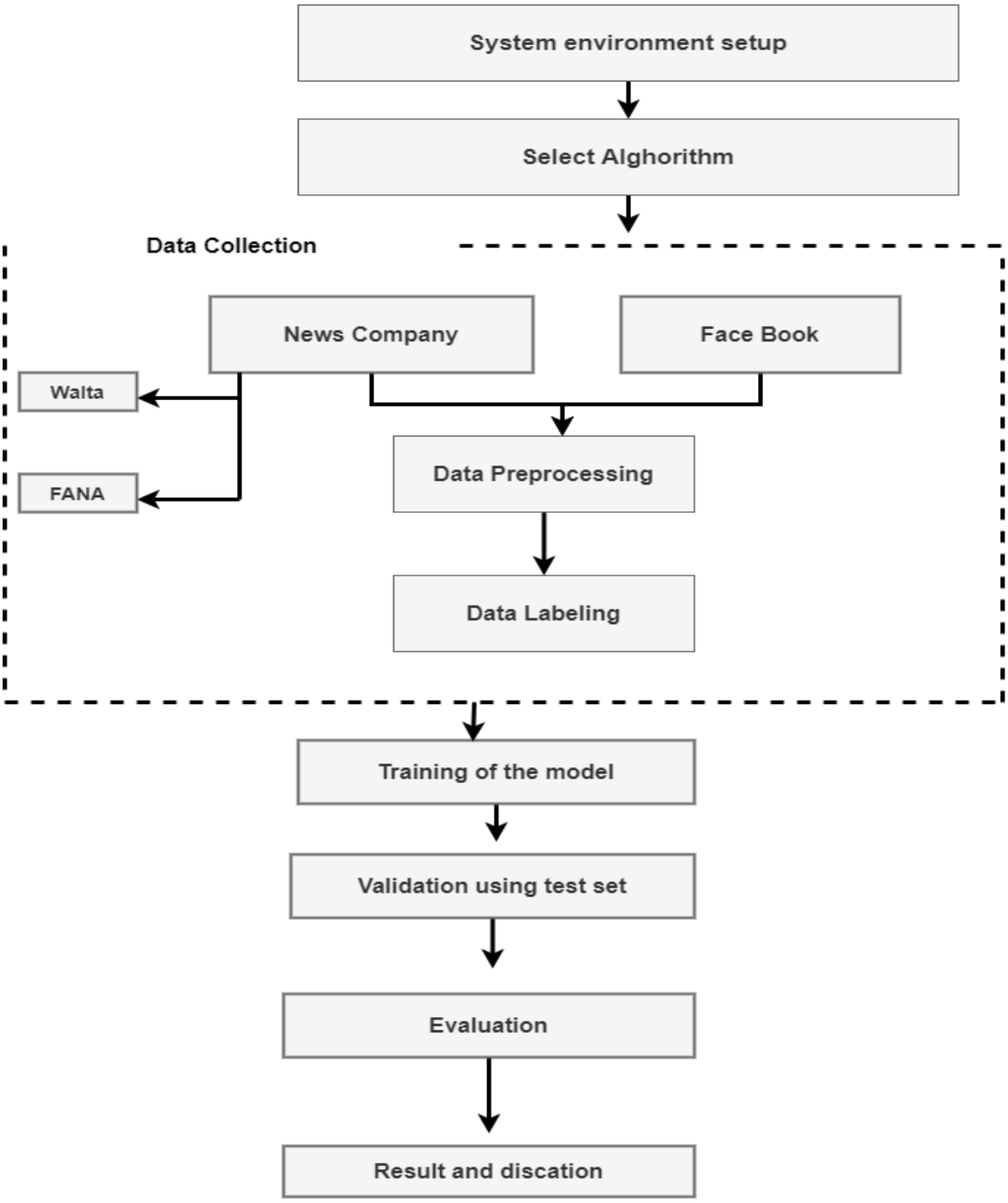


Figure 3. 1 Research Methodology

3.2 Research Design

The research design relates to the overall methodology to coherently and logically combine the different components of the analysis. In order to meet the general and specific aims of the research, and to solve the problem by answering the research questions, different methods are

used. After deciding the area of interest, the first step was performing literature review to get previous works Fake news on the area of different languages and methods.

3.2.1 Understanding the Problem domain

Since political language on TV interviews, Facebook updates, and Twitter posts is mainly short sentences, detecting fake news is more difficult than detecting misleading reviews. Fake news can have a wide-ranging negative impact, and it can also influence or even exploit major public events. Various Amharic text lexicon data are currently inaccessible, structured, or supported by NLP. The most difficult aspect of this research is filtering selected raw data into a proposed model that is both accurate and answers the research questions After identifying the problem, the requirement or data collection was the first step in developing the proposed system. The needed data and user requirements were gathered in this study using a python code to extract raw data from the Facebook API and news organizations.

Data interpretation involves a closer look at the data. This move involves gathering sample data and determining which data depending on form and scale would be necessary. In terms of the models, it attempts to illustrate for this analysis, and check the usefulness of the results required.

Data must be checked for completeness, redundancy, missing values, and attribute value plausibility. It is best if the assumptions made during the project understanding phase are justified in terms of representativeness, formativeness, data accuracy, and the presence or absence of external variables by the end of the data understanding process.

3.2.2 Preparation of the data

Good data planning makes it easier to manipulate raw data, which is often unstructured and chaotic, into a more organized and usable form that is ready for further study, for accurate analysis, limits errors and inaccuracy that can occur to data during processing. The entire planning process consists of a number of main operations including data profiling, cleaning, integration, and transformation, and makes all processed data more formal and easier to achieve prepared model targets.

One of the main advantages of establishing a systematic method for data preparation is that consumers can waste less time identifying and structuring their records. This is the main step on which the progress of the whole method of exploration of information depends; it typically

absorbs about half of the entire effort of study. It can entail data sampling, running, data cleaning, such as testing the completeness of data records, eliminating or correcting noise, etc. meeting clear input criteria to be included in our model as expected.

3.2.3 Data Source

Identifying source of data is one of the primary tasks in fake news detection. This research work identifies source of real and fake news from several social and news agencies some of these are: (i) Fana broadcasting is an organization the stream news and different multimedia program through local and international stream platforms Fana Broadcasting Corporate began operations in 1994 with obsolete equipment and insufficient manpower, but with a unique and innovative style in the country's broadcast media industry. (ii) Walta broadcasting has added a number of values since its inception in 1986. It is the owner of a massive library that houses text and audio files on a broad range of Ethiopian issues. And other data were collected from Facebook and other social Media.

3.2.4 Software Tools

This research mainly uses python, which is a high-level programing language that's interpreted by a python interpreter. Many features make python the best choice of language for machine learning and Artificial Intelligence (AI). Among the essential features textual analytics, its open source, and efficient in solving machine learning criteria like the handling of massive data sets and statistical computing.

3.3 Word Representations Approaches

The aim of language modeling is to learn a common probability of a given word series occurring in texts. The preprocessed data should be changed to a format that is suitable to the neural network classier. Mainly, there are most common-word representation approaches; like bag of words and word embedding.

3.3.1 Bag of Words Model

The Bag of Words (Bow) model is extremely easy to understand, execute, and it provides tons of versatility to configure your unique text. In Bag of words, word lists are paired with their amount of occurrence per document [21] [22]. The Bag of words representation model may use Unigram, Bigram, Trigram, or N-gram for creating a vocabulary of a document that is used to classify the

document [23]. During a BOW model, sentences are represented as multi-set of their words that disregards the order and disregards the context during which the word occurred [8].

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

Term Frequency-Inverse Document Frequency (TF-IDF) weight gives us an indication of how important a given word is for a sentence or document in relation to the entire corpus [8].

3.3.3 Word Vector Representation

Word embedding is essentially a type of word representation that bridges the human comprehension of language to that of a computer. Word embedding is real-valued representations for text by embedding both the semantic and syntactic meanings obtained from an unlabeled large corpus and are perhaps one of the key advances for the remarkable performance of deep learning methods in challenging NLP (natural language processing) problems, such as content-based fake news detection [15].

Word embedding is a dispersed representation of text in an n-dimensional space. These are important for the resolution of most NLP problems [24]. Word Embedding often referred to as distributed semantic model or distributed represented or semantic vector space or vector space model [24]. When you read these names, you come across the word semantic, which means that related words can be categorized together [25]. There is a well-known word embedding algorithms such as Word2Vec [25] [24], Glove, and Fast Text.

Glove: Global vectors for Word Representation (GloVe) are learned by first constructing a co-occurrence matrix of all the words in the corpus, and the dimensions are then reduced by matrix factorization methods. Researchers have trained and created GloVe models on multiple corpora and made them available to the public [8].

3.3.4 How Word2Vec Works?

Word2vec is a two-layer neural net that takes a text corpus and produces a set of vectors by "vectorizing" the words. Word2Vec is a predictive model as the model is designed to predict the word given a context window [8]. Word2vec is not a deep neural network; instead, it turns text into a numerical form that deep neural networks can understand. The purpose of word2vec is to group vectors of similar words together in a vector space.

For improved word representation, Word2vec is the technique/model to build word integration. It catches several exact syntactic and semantic words. The idea of word2vec method is that appear in similar ways have related implications. It takes feedback from large corpora and produces a word vector for each word. Embedding's: continuous-bags-of-words (CBOW) and skip-gram models [25].

In the representation of uncommon terms or phrases, the Skip Gram model is more accurate and works very well with a limited amount of details. To create a model such that the word in the middle is given "jumped", the model would be able to predict or produce the terms surrounding it. "The", "cat", "over", "the", "puddle". Here, the word "jumped" is considered the context. This sort of model called a Skip Gram model[25] [24]. Every word has been trained against the meaning in the CBOW template. One way is to treat {"The", "cat", 'over", "the', "puddle"} be able to forecast or produce the central word as a context and from these words "jumped". This type of model is called the Continuous Bag of Words (CBOW) Model. Depending on the amount of training data Skip-gram models are efficient for a limited number of training details whereas CBOW is efficient with a huge volume of data for training. **Fig 3.2** shows the CBOW and Skip-gram word2vec representations.

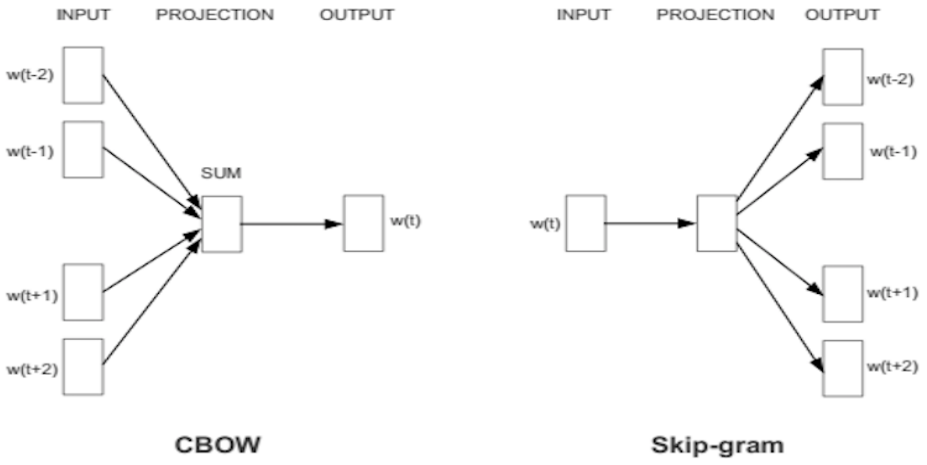


Figure 3. 2 Two forms of word2vec architectures [43]

3.4 Introduction to Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) are machine simulation tools that have recently appeared and are commonly recognized in many fields for modeling complex real-world problems. ANNs can be characterized as structures consisting of densely interconnected adaptive basic processing elements (called artificial neurons or nodes) capable of performing massively parallel data computing. Basically, the ANN was influenced by the human nervous system. It is based on hypotheses of vast interconnection and parallel processing architecture of the biological system human neurons have various types of phases, method, even stages of producing certain observable output, beginning from observing and extracting information through the neural process and eventually determining what the output will be. An ANN model is a data-driven mathematical model that has the potential to solve problems with machine learning neurons[26] [27].

Artificial neural networks (ANNs) are relatively modern computing methods that have been commonly used to solve many complex real-world problems. The attractiveness of ANNs derives from their exceptional information processing characteristics, which are primarily related to non-linearity, high parallelism, fault and noise resistance, and learning and generalization capacities. ANN learning is done iteratively, as the network provides samples of instruction, close to the way to gain experience. Most researches mentioned ANN to properly learn if it is capable of managing.

1) Imprecise, blurry, noisy and probabilistic information without significant adverse effects on the level of response.
2) The mission it had learned from unfamiliar ones was generalized.

A special aspect belonging to intelligent systems, biological or otherwise, is the capacity to learn. Training is used in artificial systems as the method of modifying the system's internal representation in response to external stimuli so that it can execute a specific task, there are three key layers in the typical form of ANNs, namely: **an input layer, a hidden layer, and an output layer**. In an ANN, the data layers for input and output are practically independent of each other. Very frequently, it is apparent that one or more hidden layers are often interconnected by weight matrices, biases, and multiple activation functions between the input and output layers [26]. ANN models have their own limitations like sequential data problems, time series problem on

text competition problem ANN models predict the next word is related to the position this could be on the word or sentences in the previous position.

Based on this ANN model drawback recurrent neural network introduces to maintain the model RNN is also a class of ANN [26].

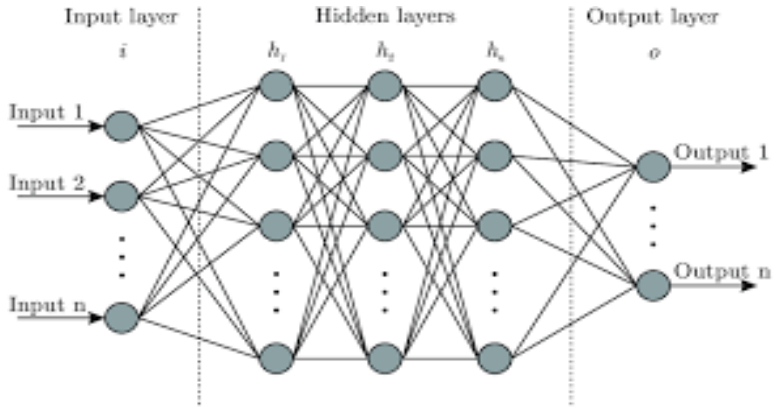


Figure 3. 3 Typical Artificial Neural Network (ANN) architecture [29]

3.5 Recurrent Neural Network (RNN)

Application of machine learning have a lot of attraction in last few years there are a couple of big categories that able to magnify and being popular specially identifying pictures and sequence to sequence translation that could be speech to text or one language to another, most of the former are done in conventional neural network (CNN) and most of the latter are done In the recurrent neural network(RNN) especially long-term memory(LSTM) for this analysis, also this research try to deploy the LSTM neural network.

Neural network-based approaches have achieved a significant range of natural language workflows. Recently incredible improvement on different problems such as recognition of expression, language modeling, translation, captioning of images, and recurrent Neural Networks are a class of artificial neural networks that are applicable to different problems.

RNN has a "memory" that remembers all the details about what has been determined. It uses the same parameters for each input as it executes the same task on all inputs or hidden layers to create the output. This reduces the complexity of the parameters as opposed to other neural networks. Decisions made by RNN are dependent on what is learned in the past and things learn during training. It has an input layer, one or more hidden layers and an output layer in structure.

RNNs have chain-like systems with repeated modules with the idea to store essential data from previous processing phases using these modules [28] [26].

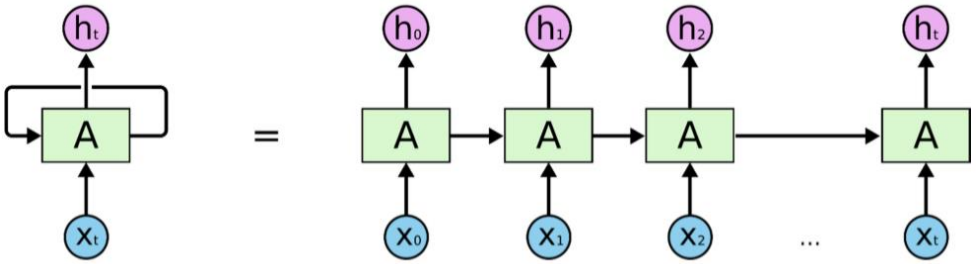


Figure 3. 4 sequential processing in RNN [26]

LSTM introduced by Hoch Reiter & Schmidhuber (1997) [26] Long Short Term Memory is the basic persistent RNN architecture that can benefit from our experience of classifying, processing and predicting time series of uncertain time lags. LSTMs have proved more precise than traditional RNNs for modeling temporal sequences and long-range dependencies [29] [26].

The back-propagation algorithm is the most widely used learning algorithm to train NNs. The principle of back propagation is to distribute the error from the output into the input, where the weights are periodically adjusted to a default value.

LSTM's have the capability of learning long term dependencies, and this enables RNN's take a long time to recall the inputs. The LSTM contain in a memory of their information can be considered as a gated cell where the cell decides whether to save or delete information based on the importance the LSTM network assigns to the information [29] [28]

3.5.1 Long Short -Term Memory (LSTM) Architecture

The LSTM network can store, write, read, and delete information on its memory. The memory of LSTM is similar to computers memory LSTM networks are embedded in a hidden layer(s) composed of so-called memory cells. There is a hidden process for every phase and feature to perform well with the neural network Each of the memory cells has three gates to retain and change its cell state to forget the gate, the input gate, and the output gate [27].

There are some researches doing improve LSTM architecture [44] they try to maintain some part of LSTM model and called LSTM like architecture forget gate about the gate is substituted by a

practical input-output gates. These LSTM improvements, now abbreviated LSTWM, are named with working memory [44].

There are different approaches applied to reshape LSTM architecture on different researches they also difficult and complex to manipulate when it compared to the base LSTM architecture so it is preferred to continue use standard LSTM architecture. As it mentioned before LSTM contain different units called memory blocks. That have memory cell with each self-connection temporary information that still in progress with those gate list input gate, forget gate, output gate so each memory block contains those three objects to work effectively. The structure of a memory cell is represented as follow figure

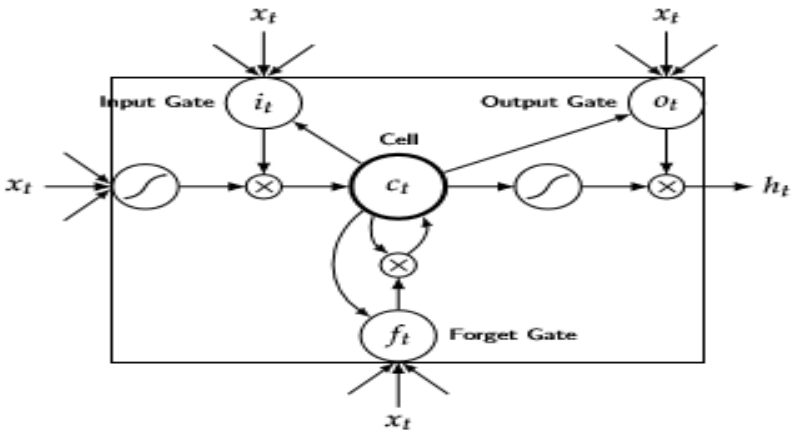


Figure 3. 5 Structure of LSTM neural unit [27] [45]

3.5.2 How Long Short-Term Memory (LSTM) Works

The first move will be building an LSTM network is to classify information that is not needed and will be excluded from the cell in that step. Sigmoid function is deciding and identifying also excluding data that takes the last LSTM unit output (ht−1) at time t−1 and with the current input (Xt) at time t. In addition, the sigmoid function decides which portion of the old output can be removed.

σ (Sigmoid layer): sigmoid function decides which part of the old output would be eliminated or kept for current state. To perform this operation, it uses forget gate value range between 0 and 1 step in constructing an LSTM network

$$ft = \sigma(Wf\,[ht-1,Xt] + bf)\ldots.[26]$$

1) Decide the information delete

The first step of LSTM is to determine which information can release from the cell state. This decision is made by the sigmoid layer, this layer looks at the value of ht-1 output and Xt input, determines the value in the range from 0 to 1 for each Ct-1 state. If the layer is returned 1, this means that this value should be left (remember) if 0 is excluded from the cell state. To use only relevant data and pass for the next state it matters wheatear the information deleted or kept for this specific task it uses sigmoid function namely called forget layer (or ft). Finally, the sigmoid function scales all activation values to the range between 0 (forgotten completely) and 1 (completely remember).

$$ft = \sigma (Wf [ht-1, Xt] + bf) \ldots\ldots\ldots\ldots\ldots\ldots (3.1)$$ [27] [45]

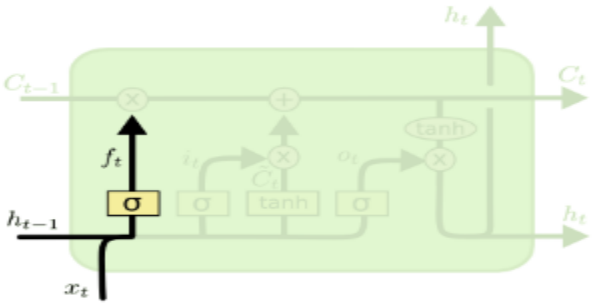


Figure 3. 6 Forget layer of LSTM Network [27] [45]

2) Deciding what new information store

The next step is to specify which new knowledge can store in cell state perform this firstly sigmoid function needed to decides which values are going to update when each iteration new information flow those connected networks. This operation it has two main sections of the sigmoid layer and second half of the tanh layer. Layer of Sigmoid this is called the "input gate layer “which determines which values we'll updated or ignored (0 or 1), and second, tanh layer (establish new values for candidates who should be applied to the network).

$$it=\sigma (Wi. [ht1, xt] + bi) \text{ ----------(3,2)}$$

$$Ćt=tanh (WC. [ht-1, xt] + bC) \text{ ----------(3,3)}$$

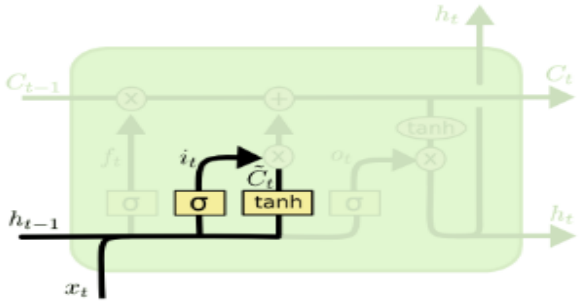


Figure 3. 7 New information store

3) Update the old cell state into the new cell state

Change the old cell state to the new cell state Update the old cell state by multiplying the new cell state with what you have forgotten before, and then adding the latest candidate cell state. The new candidate values are scaled by how much each state value wanted to change [26] [27].

Ct=ft *Ct-1+it* Ćt----------(3.4) ……. [26] [27]

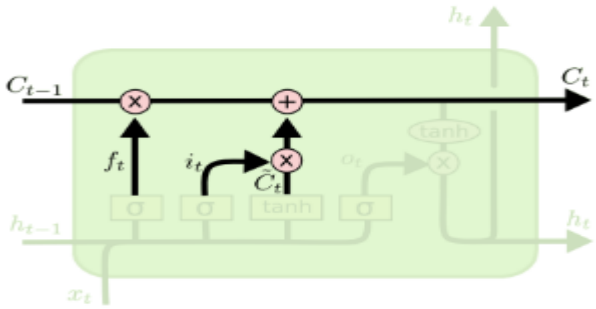


Figure 3. 8 Memory cell update layer of LSTM Network

4) Final step, decide the output values

The neural network output layer captures and transmits the information accordingly in the manner it has been programmed to provide. The output value (ht) step is based on the state (Ot) output cell, but is filtered. A sigmoid layer first specifies which parts of the cell state are formed by the cell to output, the output of the sigmoid gate (Ot) is multiplied by the new values created by the tanh layer from the cell state (Ct), with a value ranging between−1 and 1.

Ot = σ (Wo [ht−1, Xt] +bo) ---------- *(*3.5*)*

ht = Ot * tanh (Ct) ----------(3.6)

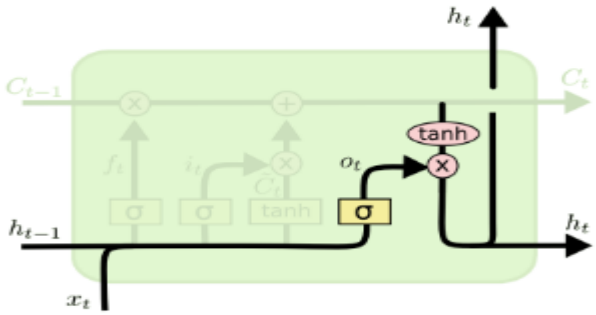


Figure 3. 9 Output layer of LSTM Network [26] [27].

The forward layer output sequence, $\boldsymbol{h}$ {{→, is iteratively calculated using inputs in a positive sequence from time $T-n$ to time $T-1$, while the backward layer output sequence, $\boldsymbol{h}^{\leftarrow}$ {{, is calculated using the reversed inputs from time $T-n$to $T-1$ [30].

One specific detail that differences when comparing the LSTM and BILSTM methods is that the consistency is achieved at a slower rate in the case of the BILSTM models compared to the LSTMs. In fact, when referring to the BILSTMs, two LSTM networks are applied to input data as follows: first, the input sequence is processed via the first LSTM network, in the so-called "forward layer" while, second, the inverted input sequence is entered through another LSTM model, thus receiving the "backward layer".

3.6 Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) are very similar to ordinary Neural Networks: they are made up of neurons that have learnable weights and biases. A CNN consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of convolutional layers, pooling layers, fully connected layers and normalization layers

CNN has been successful in various text classification tasks. In **"Kim Y. Convolutional Neural Networks for Sentence Classification. 2014",** Kim showed that a simple CNN with some hyper parameter tuning and static vectors like Glove vector and word2vec CNN can achieve excellent results on text classification.

• The use of word vectors allows CNN to use its state of art parameter sharing for reduction in calculation and govern the overall text features.

- **ReLU**: stands for Rectified Linear Unit for a non-linear operation. The output is $f(x) = \max(0, x)$. REL U's use is to introduce non-linearity in our CNN.
- **Pooling**: The pooling layer reduces the height and width of the input to the layer. It helps reduce computation, as well as helps make feature detectors more invariant to its position in the input
- **Padding**: It allows you to use a CONV layer without necessarily shrinking the height and width of the volumes. This is important for building deeper networks, since otherwise the height/width would shrink as you go to deeper layers.

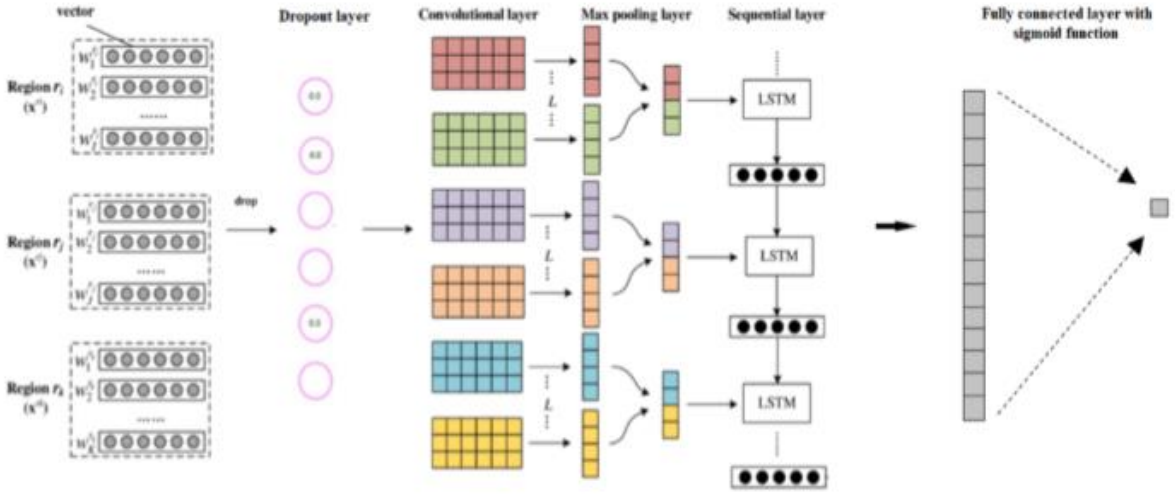


Figure 3. 10 Convolutional Neural Network Architecture

3.7 Multi-Head Attention Model

Emergence of Attention improves the performance of neural machine translation applications [31]. Self-attention or intra-attention has been used successfully in a variety of tasks including reading comprehension, abstractive summarization, and textual entailment and learning task-independent sentence representations [32]. There are two most commonly used self-attention functions such as scaled dot product attention, and multi-head attention [32].

Given the same set of queries, keys, and values, Self-Attention model combines the knowledge from different behaviors of the same attention mechanism, such as capturing dependencies of various ranges (short range or long range) within a sequence. In language modeling, Self-

Attention focuses on words that form queries [32]. Self-attention focuses on relationship among words that form the sentence that calculates attentions, or relevance between queries, and keys.

Multi-Head Attention is a type of attention that allows you to do exactly the same attention process or procedure in every head at the same time [32]. This procedure enables parallelization of Self-Attention mechanism. Multi-Head Attention concatenates the outputs of the different representation heads, and you put the concatenated matrix through fully connected layers [32]. In practice, dot-product attention (multi-head) attention is much faster, and more space space-efficient [32]. Figure 1 below on the left side shows a scaled-dot product self-attention. Whereas, the figure on the right side is a Multi-Head Attention consisting of h number of self-attention layers that runs in parallel [32]. Figure 2 below shows parallel execution of several self-attention models.

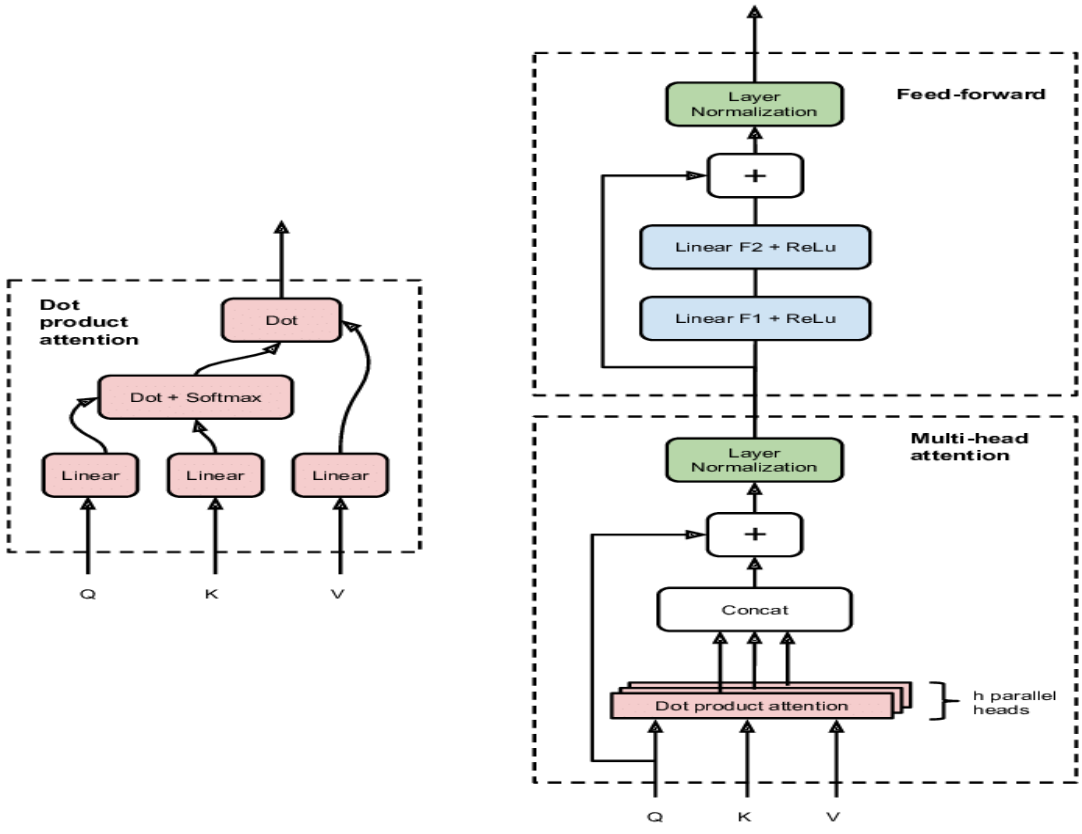


Fig 3.11: Details of multi-head attention building blocks **(Afouras, T, and et al., 2018)**

CHAPTER FOUR

PROPOSED FAKE NEWS DETECTION

This research develops fake detection models for the Amharic language. The developed models are i) LSTM model, ii) LSTM + Bi directional LSTM ensemble model, iii) CNN+ LSTM ensemble model, and LSTM model with attention mechanism.

Additionally, comparison with machine learning algorithms specifically with Naïve Bayes, and Support Vector Machine (SVM) has also been done in the continued subsection to measure their relative performance with deep learning methods on the same datasets. Figure 4.1 shows the steps followed from data collection up to classification of news as or real.

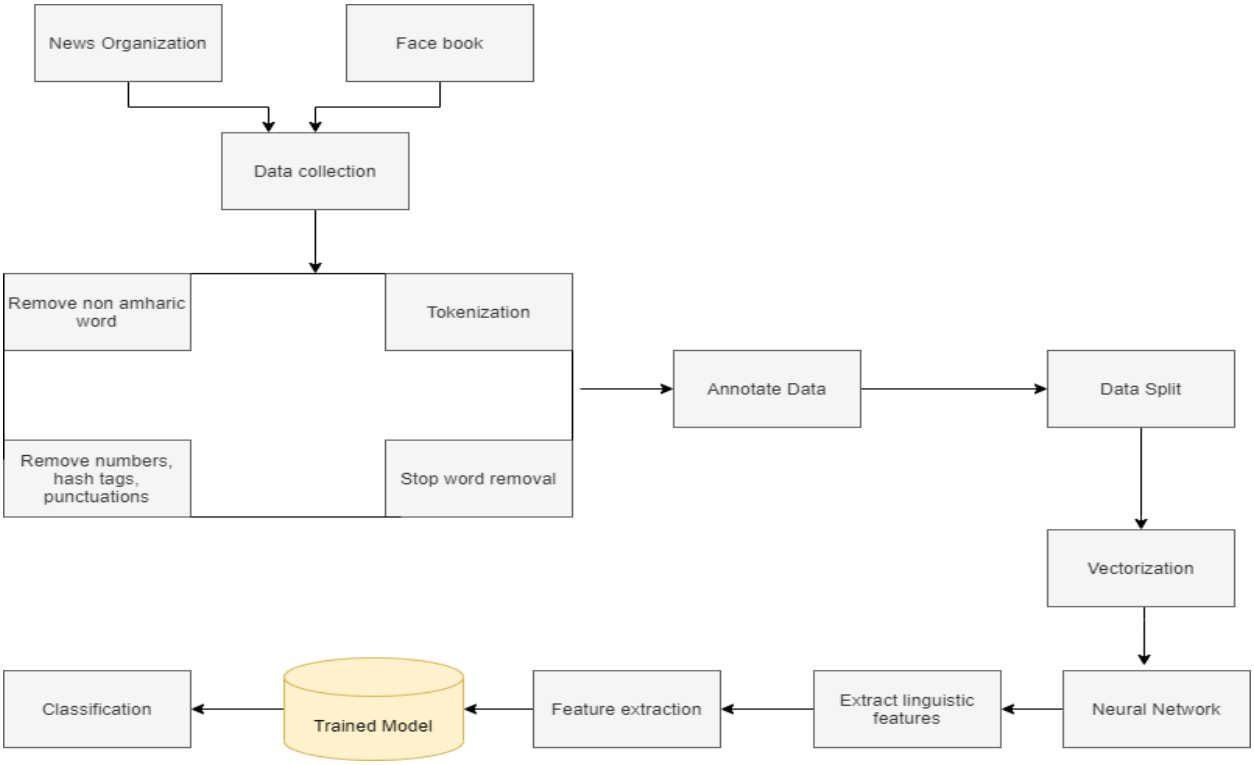


Figure 4. 1 Data Collection and Annotation with development process

4.2 Data Collection

To develop a fake news detection model, data collection is the first and fundamental step that encompasses collecting a user-generated contents from blogs or social media like Facebook, Twitter, and others. Scraping is the process of collecting user-generated content from social media and blogs. This research uses a python code for collecting raw data in the form of

unstructured data because they are expressed in different vocabularies, ways of writing and others.

4.3 Data Preparation (Preprocessing)

Data preparation is a major step in data analysis that encompasses many techniques mainly data transformation, data cleaning, and data reduction to make the collected data clean and suitable for analysis. In data preparation, non-textual contents and contents that are unimportant for the analysis are first identified and eliminated.

Natural Language applications use consecutive data preparations approaches due to many reasons, for example, real-world data is full of impurity and incompleteness, data preparation generates a small amount of quality data set than the original one which improves the efficiency of data mining algorithms and results in high-performance mining systems, this research come up with the following approaches for cleaning the collected data.

- Eliminating non-textual contents and symbols like hashtags, URLs, links and others [21] [33].
- Removing non-Amharic words
- A consistent tokenization rule should be applied to split sentences into words and treat each punctuation as a separate token word tokenizer from NLTK is used in this thesis. Also this research assumes that punctuation elements, including capitalization, the presence of question marks and exclamation marks, etc., tend to identify hateful expression, and cannot simply be dismissed.

4.3.1 Tokenization

Tokenization is one of the preprocessing steps in NLP. Tokenization is splitting the text into smaller units and each unit is called tokens. Tokenization mostly is the first step in pre-processing of the input review. The input for this activity is the actual review which is going to be categorized. This activity reads a sequence of characters as a string and tokenizes them using predefined list of delimiters such as new lines and space [20].

Separate tokens are properly separated, such as punctuation marks, words, email addresses, links, numbers, abbreviations, etc. This process detects the boundaries of a written text; Tokenizing of a given text depends on the characteristics of language of the text which it is written. The

Amharic language has its own punctuation marks that demarcate words in a stream of characters. The tokenization of text on this component uses Amharic punctuation marks and white space.

4.3.2 Normalization

Preprocessing tasks such as normalization, tokenization, stop word and number elimination, stemming, weighting words, and dimension reduction are performed. Dimension reduction is performed by moving varying Amharic characters with similar sound to a standard type, and by changing punctuation marks to space. After all these preprocesses, datasets are prepared in a matrix form, which is accompanied by the preparation and testing of training and evaluation data sets. Normalization is one of the steps used to get clean data from unstructured textual data. After tokenization, then normalization of homophones is followed. Amharic writing system has homophone characters, characters with same pronunciation but different symbols; for example, it is common that the character ስ and ሥ are used interchangeably as ስራ and ሥራ to mean work, also ሀ, ሐ or ኀ should be changed to the same form because they have the same punctuation all represent one alphabet called "ha". Such types of inconsistencies in writing words are handled by replacing characters of the same sound by a common symbol.

There are two levels of normalization these are character level normalization, and word-level normalization. Short forms of characters that are usually written using forward slash "/") and period ("."), for example, ጠቅላይ ሚንስተር can be written as ጠ/ሚ ስተር, አዲስ አበባ as አ.አ and ዶክተር as ዶ/ር.

4.3.3 Punctuation Removal

Punctuation in natural language provides the grammatical context to the sentence. Punctuations such as a comma, question mark, full stop, and others might not add much value in understanding the meaning of the sentence [8].

Amharic text analysis demonstrates that different Amharic Punctuations markers are employed for distinct purposes. There are roughly 17 punctuation marks of which just a few of them are regularly used and have representations in Amharic software [8]. Some local and foreign punctuation is used in the Amharic writing system. However, only a few of these are employed in practice, particularly in computer-generated writing. The word-separator is a character that separates two words. (: ሁለት ነ ጥብ), two square dots arranged like colon (:), and sentence-separator (። አራት ነ ጥብ), four square dots arranged in a square pattern (: :), are the basic punctuation marks in Amharic writing system that are used consistently. In Amharic language 34 Words are separated by two dots (: ሁለት ነ ጥብ). However, it is not

seen in modern typesetting. In typesetting its place is almost completely taken over by blank space. The end of the sentence is marked by a square-formed four dots. Lists in Amharic text are separated by an equivalent of comma (፣ ነ ጦላ ሰረዝ) and (፤ ድርብ ሰረዝ) which is the equivalent of semi-colon. Moreover, the language borrows some punctuation marks from foreign languages such as (? ,!, “ , ”, „, /, \, etc.).

4.3.4 Stemming

Stemming After tokenizing the data, is a process of changing the words into their original form, and decreasing the number of word types or classes in the data. Stemming is a technique to remove prefixes and suffixes from a word, ending up with the stem. Using stemming we can reduce inflectional forms and sometimes derivationally related forms of a word to a common base form [8].

For example, argue, argued, argues and arguing will be reduced to argu [21]. The trouble with stemming is that multiple words may be simplified to the same word. Amharic is morphological rich language having many words with the attachment of different affixes to a stem. Stemming can be applied to both inflectional morphology and derivational morphology or on either of the two. Derivational morphology usually results in a change in class of word which may result in some loss of semantic. This semantic loss of a word may create a negative effect on the performance of system. For example, the Amharic word ዳኛ (a judge or an arbiter) and ዳኝነት (judging) has the same stem ዳኝ. However, these two words have different meanings. Amharic language includes some prefixes and combination of prefixes and suffixes which create negative meaning when they are applied to a given stem. This kind of semantic loss is not usually witnessed during inflectional morphology, which usually involves grammatical features such as, singular/plural, tense. The work **on effect of preprocessing on Amharic sentiment analysis** has proved that stemming has negatively affect sentiment analysis [34]. Hence, this research work does not incorporate stemming for preprocessing task. Figure 4.3 depicts a paragraph before and after stemming is applied.

4.3.5 Stop Word Filtering

Stop words are insignificant words that do not provide much context in a language or hold less information. These words will create noise when used as features in text classification. These are words commonly used a lot in sentences to help connect thought or to assist in the sentence structure. Articles, prepositions and conjunctions, and some pronouns are considered as stop words. For example, conjunctions like “and”, “or” and “but”, prepositions like “of”, “in”,

“from”, “to”, etc. and the articles “a”, “an”, and “the”. Such stop words which are of less importance may take up valuable processing time, and hence removing stop words as a part of data preprocessing is a key first step in natural language processing.

Removing stop words helps to reduce the dimensionality of a term space in Amharic, there are common stop words which are used for grammatical purposes, such as ነዉ, ነበር, ሆኖም, እና, ነገር ግን, etc, to classify records which are non-informative. In addition to the standard stop words, there are also news relevant stop words including ገለፁ, ዘግበዋል, አስታወቀ, etc. This research use Natural Language Toolkit – (NLTK) library to remove stop word [8]. Figure 4.4 shows a paragraph before and after stop word removal is applied.

4.4 Developing Neural Network Model

4.4.1 Development Step

Neural net is a machine learning system that uses a network of functions, normally in another form, to interpret and convert a data input from one form into a desired output.

Using neural networks or matrix factorization, the learning of word integration can be achieved. One widely used word embedding method is Word2Vec, which is basically a computationally efficient neural network prediction model that learns word embedding from text. This step encompasses building LSTM model**,** an ensemble CNN and LSTM with and without attention mechanism, and ensemble LSTM and Bi directional LSTM.

LSTM network is learned using these word embedding’s and classifies the words into the pre-defined label categories as it mentioned model was utilized to classify the words based on the word vectors generated by word2vec, LSTM is really effective on capturing long and short term memory, takes vectors of word2vec as input from the dense layer. LSTM layer has a memory unit that helps in adding and removing previous information. LSTM model, the network is learned by using training data and tested data. The network consists of an embedding/input layer, two hidden layers, and an output layer [2].

4.4.2 Sampling Technique

We split our data into Train, Validation and Test data sets. We allocated 80 % of our data to the train set and the remaining 20% to the test set. The training data is further divided into validation sets.

4.4.3 Mult -head Attention approach

In this section, the multi head Self-Attention-based BiLSTM neural networks model is proposed for the sentiment polarity classification for short texts. The proposed model consists of a word encoder layer, a BiLSTM layer, an attention layer and a softmax layer and relu activation. This simultaneous and independent calculations is to decompose the attention in numerous heads. Its more advance as several "linear views" from the same sequence.

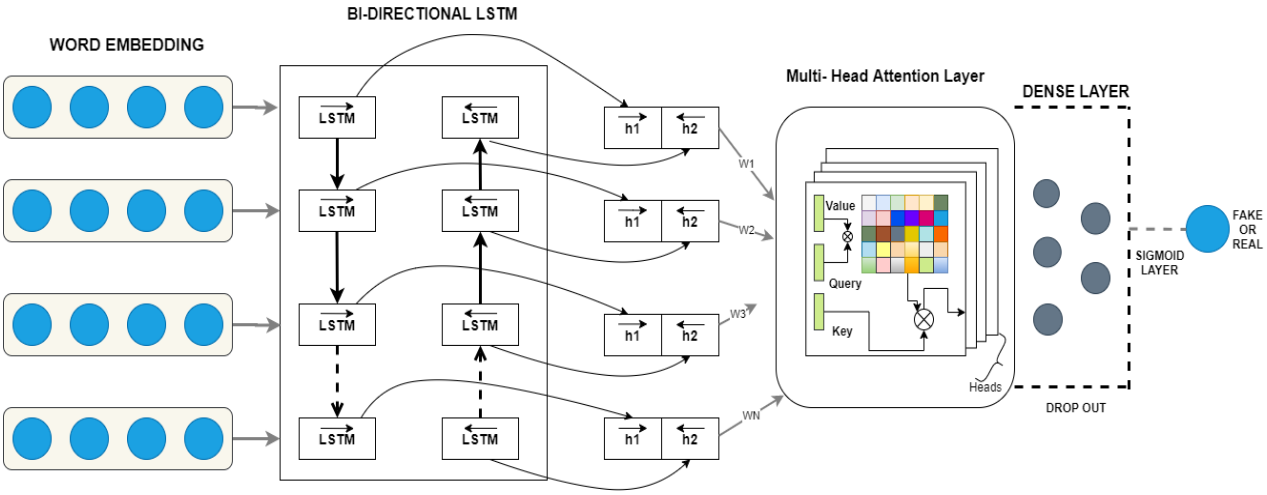


Figure 4. 2 Proposed multi head attention based BI-LSTM model

4.5 Evaluation

After classification has finished, the performance of the classier is expressed using different performance metrics such as recall, precision, F-score and accuracy. The evaluation metrics are displayed using Scikit-learn python library. Training and validation ac curacies and losses are also used as evaluation metrics.

CHAPTER FIVE
RESULTS AND DISCUSSION

5.1 Performance Measures

Accuracy: Accuracy is often the most used metric representing the percentage of correctly predicted observations, either true or false. To calculate the accuracy of a model performance, the following equation can be used [35].

$$Accuracy = \frac{TP + TN}{TP + FP + FN} - - - -(eq\ 1)$$

In most cases, high accuracy value represents a good model, but considering the fact that we are training a classification model in our case, an article that was predicted as true while it was actually false (false positive) can have negative consequences; similarly, if an article was predicted as false while it contained factual data, this can create trust issues.) Therefore, we have used three other metrics that take into account the incorrectly classified observation, i.e., precision, recall, and F1-score [35].

Recall: Recall represents the total number of positive classifications out of true class. In our case, it represents the number of articles predicted as true out of the total number of true articles.

$$\boldsymbol{Recall = \frac{TP}{TP+FN} - - - - - (eq\ 2)}$$

Precision: Conversely, precision score represents the ratio of true positives to all events predicted as true. In our case, precision shows the number of articles that are marked as true out of all the positively predicted (true) articles:

$$\boldsymbol{Precision = \frac{TP}{TP+FP} - - - -(eq\ 3)}$$

F1-Score: F1-score represents the trade-off between precision and recall. It calculates the harmonic mean between each of the two.) Us, it takes both the false positive and the false negative observations into account. F1-score can be calculated using the following formula:

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} - - - (eq\ 4)$$

Table 5. 1 depicts the confusion matrix

	Predicted true	**Predicted false**
Actual true	True positive (TP)	False negative (FN)
Actual false	False positive (FP)	True negative (TN)

True Positive (TP): when predicted fake news pieces are annotated as fake news.

True Negative (TN): when predicted true news pieces are annotated as true news.

False Negative (FN): when predicted true news pieces are annotated as fake news.

False Positive (FP): when predicted fake news pieces are annotated as true news.

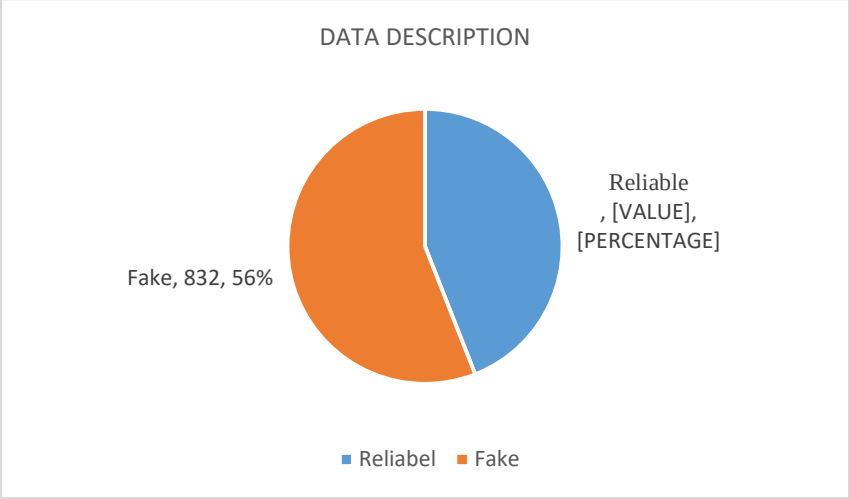


Figure 5. 1 data properties

5.2 Experimental setup

This research work uses one personal computer for all experiments. Table 5.2 shows the hardware and software specification of the machine used in all experiments

Table 5. 2 Experiment setup

Manufacturer	LENOVO
Model	G-50
Processor	Intel® Core™ i7-4510U CPU @ 2.00GHz × 4
Memory(RAM)	8 GB
Operating System	Ubuntu 20.04 LTS

Model Configuration: In addition to hardware and software specifications, configure the LSTM network required for developing fake news detection model. In the experiment,

LSTM network is configured with the following parameters: one LSTM layer with 8 neurons, 20 epochs, 512 batch, 8 NB_WORDS (Embedding size), 'Adam' optimizer, 90% train size, and 10% test size. main problem in training neural networks is the over-fitting and under fitting [36] . Over fitting happens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data. Generally processing on training task it's not improving to increase the performance, but its adding and continue randomly learn training pattern.

Table 5. 2 data properties

Before and after pre process	Total number of sentences	Total number of words
Before	2200	33,251 words
After	1300	27511 words

Sample of dataset description

Table 5. 3 Reliable News data sample

	Fana	Walta	VOA
1	እንደ አባቶቻችን ከተባበርን ማንምአይደፍረንምጠቅላይ ሚኒስትር ዐቢይ አህመድ አዲስ አበባ፣ ግንቦት 15፣ 2013 (ኤፍ ቢ ሲ) "እንደአባቶቻችንከተባበርን ማንምአይደፍረንም' ሲሉ ጠቅላይ ሚኒስትር ዐቢይ አህመድ ገለጹ።	በ6ኛውሀገራዊ ምርጫበማላውሀገሪቱ ሲካሄድበቆየውየሚራፍቹ ምዘገባከ36ነጥብ2 ሚሊየን በላይ ሚራፍቹ መመዝገባቸው ታውቋል፡፡ ድምጽ ማነጭምሰኔ 14/2013 ዓ.ምእንዲሆን ቀን ተቆርጦለታል።	አንዳንድሀገራት በኢትዮጵያ የውስጥ ጉዳይበእጅጉ ጣልቃ እየገቡማሆኑን የገለጹ በጎ ፈቃደኛ ዜጎች ይሄንኑ የሚቃወም ንቅናቄ ዛሬ አስጀምረዋል።
2	የኦሮሚያ ልማት ማህበር ኮንፈረንስ ተካሄደ አዲስ አበባ፣	ኢትዮጵያ በአለምአቀፉ የድሮን ኮንግረንስ በተጋባዥነት	የዩናይትድስቴትስ የዓለምአቀፍ ልማት

	ግንቦት 15፣ 2013(ኤፍ ቢ ሲ)በአራዳ ክፍለ ከተማየ ኦሮሚያ ልማት ሚህበር ኮንፈረንስ ተካሂዷል።	እየተሳተፈች ነ ውግንቦት 15/2013 (ዋልታ) በቻይና ሽንዝን እየተካሄደ በማ ኘው5ኛውአለም አቀፍ የሰውአልባ አውሮፕላን (ድሮን) ኮንግረንስ 2021 ኢትዮጵያ በክብር ተጋባዥነ ት እየተሳተፈች ትገኛለች።	ተራድዖ ድርጅት/ዩ ኤስ ኤድ/ አስተዳዳሪ ሳማንታ ፓወር የ ትግራይ ክልል ግጭት እጅግ የ ከበደ ረሃብእያስከተለ እንደሆነ እና ከአምስት ሚሊዮን የ ሚበልጥ ህዝብ እርዳታ የ ሚያስፈልገ ው ሚሆኑን አሳስበዋል።
3	በሶማሌ ክልል ፕሬዝዳንት አቶ ሞክጡፉ ሙሃ ሙድ የ ተሙራ ልዑክ ጎደይ ከተማገ ባ አዲስ አበባ፣ ግንቦት 15፣ 2013(ኤፍ ቢ ሲ)በሶማሌ ክልል ፕሬዝዳንት አቶ ሞክጡፉ ሙሃ ሙድ የ ተሙራው የ ፌዴራል እና የ ክልል ከፍተኛ የ ስራ ሀላፊዎች ጎደይ ከተማ ገ ብቷል።	ህብረተሰቡየ አካል ብቃት እንቅስቃሴ በሞራት ጤፍ ሞኽ አምራች ዜጋ ሚሆን እንዳለበት ተጠቆመ ግንቦት 15/2013 (ዋልታ) - ህብረተሰቡየ አካል ብቃት እንቅስቃሴ በሞራት ጤፍ ሞኽ አምራች ዜጋ እንዲሆን የ ወላይታ ዞን ዋና አስተዳዳሪ ዶክተር እንድሪያስ ጌታ አሳሰቡ።	ሞቅሌ ከተማከባለፈው መጋቢት ወር ሜጩፈ ሻ ጀምሮ የ ኮሮና ቫይረስ ምር ሙራ ካደረጉ ሰዎች 37 በሞቶ ቫይረሱ እንደተገ ኘባቸው የ ትግራይ ክልል ጤፍ ቢሮ ገ ልጿል።

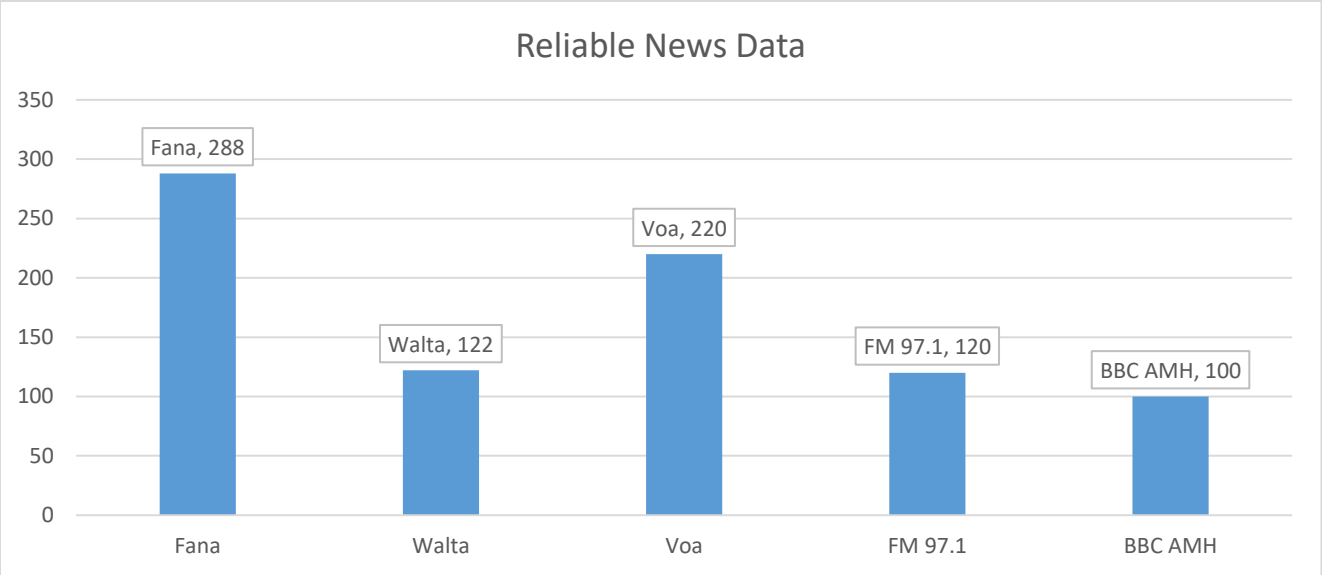


Figure 5. 2 Reliable data sample

Table 5. 4 Fake News data sample

	Hageregna	Seber Zena	Zehabesha
1	በባህር ዳር ከሰሞኑ በተደረገ ው የተቃውሞሠልፍ ወቅት በአምስት ባንኮች ላይ የዝርፊ ሞከራ ተደርጓል ተባለ	ጋዜጡኛና ገጥሚሰለሞን ታይለኢየ ሱስ በአሁኑ ሰአት በኢትዮጵያ ቴሌቪዢን በዜና አቅራቢነ ትና የአለምአቀፍ ጉዳዬች ተንታኝ በሚሆን እየሰራ ይገኛል ።	የ ሚ ስትሮች ምክር ቤት ታሪካዊ ውሳኔን ወስኗል። ከፍተኛ የፈቃድ ክፍያ ዋጋ እና አስተሞሸኝ የ ጫለ ነ ዋይ ፍሰት ዕቅድ ላቀረበው ለግሎባል ፓርትነ ርሺፕ ፎር ኢትዮጵያ የ ኢትዮጵያ የ ኮሚ ኬሽን ባለሥልጣን አዲስ ሀገ ር አቀፍ የ ቴሌኮምፈቃድ እንዲሰጥ በአንድ ድምፅ አጽድቋል።
2	የ ሶ ማሌ ያ ዋና ከተማሞቃዲሾ በወታደሮች ተኩስ ስትና ጥ አድራለች።	ዶ/ር አርከበ ለአፍሪካ፣ ካሪቢያን እና የ ፓስፊክ መንግስታት ድርጅት የ አምባሳደሮች ኮሜቴ ተሞድ የ ኢንዱስትሪ ልማት ድርጅት ዋና ዳይሬክተርነ ት ሆነ ውከተሚጡ ሊሰሯቸውስለሚችሉ ስራዎችና ስኬቶቻቸውዙሪያ ማብራሪያ ሰጥተዋል።	የ ፋሲል ከነ ማደጋፊውለፍቅረኛው ያቀረበውጥያቄ በጨዋታውማዝል ክለቡ ለሚጫሚያ ጊዜ ዋንጫውን በሚያነ ሳበት ቀን እሺታን አግኝቷል ።
3	ከባህርዳር ወደ አዲስ አበባ ተደራጅተውእየ ተጓዙ ነ በሩ የ ተባሉ ወጣቶች ከጎ ሃ ጽዮን እንዲማለሱ ሞደረጉን ፖሊስ ገለጸ	የ ኢትዮጵያ ስትራቴጂክ ወዳጅና ብርቱ አጋር የ ነ በረችውአ ሜሪካ ኢትዮጵያ ላይ ለምን ተነ ሳች ?	የ ሶ ማሌ ክልል ፕሬዝዳን ት አቶ ሞክጡፉ ሙሐመድ ለአር ቲስት አሊ ቢራና ባለቤቱ ወ/ሮ ሊሊ፤ እንዲሁም ለሌላኛውየ አሮምኛ ሙዚቃ አቀንቃኙ ሞክታር አዴሮ የ ምሳ ግብዣ አድርገ ውላቸዋል።

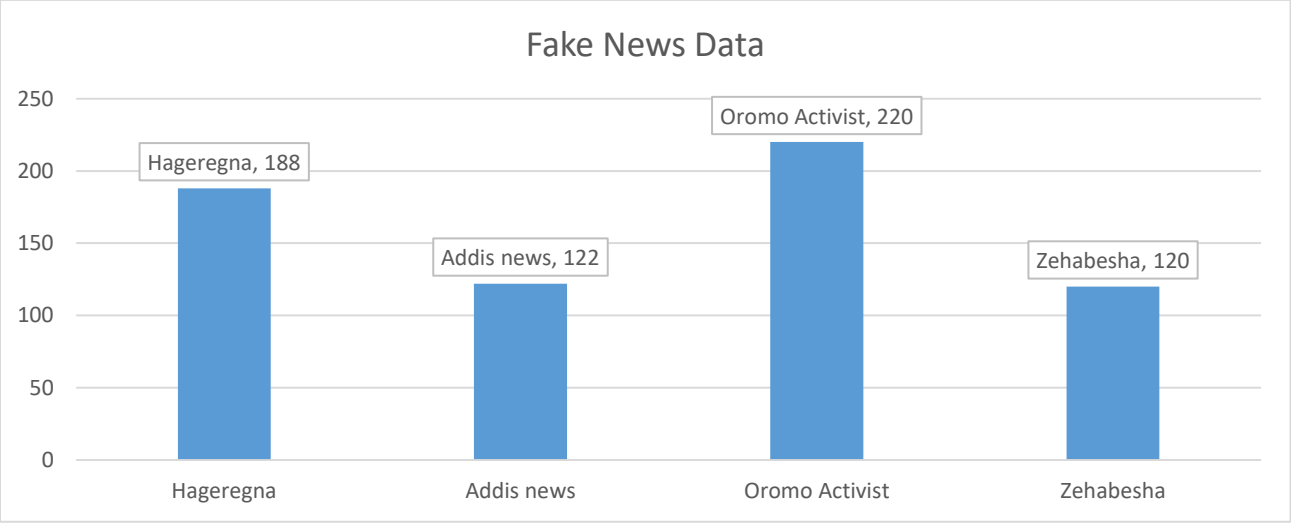


Figure 5. 3 Fake news data characteristics

5.3 Development Tools and packages

These sections reveal development tools used in this research. python programming language commonly building, compare, evaluate, graphical representation, data processing is performing by using python programming language. Python **is Interactive**, **Interpreted** Consistency and simplicity are some beneficial properties. Also reduce development time, because it contains many libraries and framework for machine learning and deep learning architectures.

5.3.1 Libraries and frameworks

Scikit-learn is suitable for processing basic algorithms for machine learning, such as clustering, linear and logistic regression, regression and classifications for small-to medium-sized data sets for feature engineering and classical ML simulation [37]. Evaluate deep neural classifiers by calculating A confusion matrix, which is a table frequently used to explain a classification model's output [37].

Pandas is suitable for advanced data structure and analysis, allowing data to be merged and filtered, and data collected from other external sources such as Excel [37].

Keras Main focus is enabling implementation of deep learning models for research and development as quickly and easily as possible, suitable for deep learning, enabling fast calculations and prototyping. Because the software library uses a GPU in addition to the computer's CPU [37].

Tensor Flow is suitable for deep learning by setting up, training, and using large data sets of artificial neural networks [37].Tensor flow is used as a back end for Keras library which is used for development of deep neural network classifiers.

Matplotlib is suitable to construct 2D plots, histograms, graphs, and other visualization types operations to easily and cleanly data and progress representation based on give para meters [37].

NLTK It is a tool used to create Python programs to work with human language data. Suitable for computational linguistics and the identification and processing of natural languages [37]. NLTK is used for many tasks like tokenization, lemmatization, stemming, parsing and POS tagging. NLTK is developed especially for languages that are rich in resources.

5.4 Experimental scenarios

To answer the research questions that set by the researcher different experiments conducted. Generally, these experiments are grouped into three categories. The second group of experiments aims to show the evaluation of different classifier models with different feature. Analysis and examine different approaches are demonstrate based on proposed classifiers, LSTM, BI-LSTM Attention with BILSTM and CNN-LSTM. And answer RQ1) [Performance] of the deep learning algorithms. The second group of experiment is analyzing and experiment based on other language trend how they justify and discern a news fake or reliable based on those analyses the second experiment examine the basics and the way this research uses to identify Amharic data is fake or reliable and this experiment answer research question two RQ2) How to identify fake news that is published in the Amharic language.

The last group of experiments aims to identify the impact of stop word data could affect fake news detection classifier model. These experiments are conducted to answer RQ3: Effect of stop word removal on preprocessing stage for fake news detection?

5.5 Results

Multi-head Attention Based Bi-LSTM

The model properties for this Attention based Bidirectional long short-term memory change the drop outs defined 0.05 drop out and activation function for BILSTM “sigmoid” activation used for dense 1. For dense 2 we apply relu activation. To train the model for this experiment, at the end of each trial, a complete 2*2 confusion matrix was created, accuracy was calculated for the resulting network, and precision recall and f-measure were calculated for each target neuron detection rate.

With **6** layers and **16** heads per layer trained on. This suggests that lower layers of the Transformer’s decoder are mostly responsible for language modeling, while higher layers are mostly responsible for conditioning on the source sentence.

The number of retained heads for each attention type at different pruning rates. We can see that the model prefers to prune encoder self-attention heads first, while decoder-encoder attention heads appear to be the most important for collected datasets. Obviously, without decoder. encoder attention no translation can happen. The importance of decoder self-attention heads,

which function primarily as a target side language model, varies across domains. These heads appear to be almost as important as decoder-encoder attention heads for Collected data with its long sentences (24 tokens on average), and slightly more important than encoder self-attention heads for Open Subtitles dataset where sentences are shorter (8 tokens on average). Table 5.6 shows results of Amharic fake news detection using BI-LSTM based multi-head attention.

Table 5. 5 Attention based BI-LSTM performance

Performance measure	**Domain of class**	
	Reliable	**Fake**
Precision	0.81	0.78
Recall	0.81	0.80
F Score	0.83	0.82
Macro-Precision	0.84	
Macro-Recall	0.84	
Macro-F score	0.82	
Accuracy	0.86	

Figure 5.4 below shows the training loss vs validation accuracy, and training accuracy vs validation accuracy for BI-LSTM based multi-head attention.

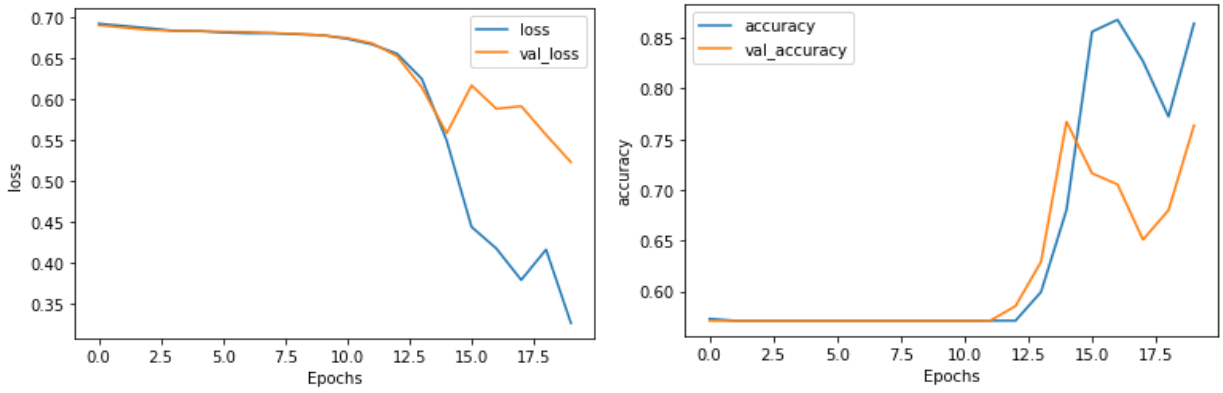


Figure 5. 4 Multi-head Attention BI-LSTM Training Vs Validation accuracy and loss.

There is a previous work that performs Amharic fake news detection using deep learning approach like CNN, LSTM, ad BI-LSTM [15]. These deep learning approaches have limitations such as lack of focusing on certain important contents. Observing drawbacks of previous works,

we propose a Bi-directional Long Short-Term Memory-Based method with attention mechanism for fake news detection. To our knowledge, this research work is the first that applies a multi-head attention mechanism for Amharic language fake news detection.

CNN-LSTM

The size of the data of the filter is 32 and it have max pooling layer. We applied a max pooling layer after each convolution layer, and then concatenated the different max pooling layers to produce a single fixed size output. There was a LSTM layer that was fully connected layer. The soft-max activation function transformed the vector into probability to define the class of each output. The loss function, essential to compile the model, was "Sparse-categorical-cross entropy". The "Adam" optimizer was used. This combination achieved an accuracy of 79.13% with 200 epochs, which confirmed LSTM's added value compared to CNN.

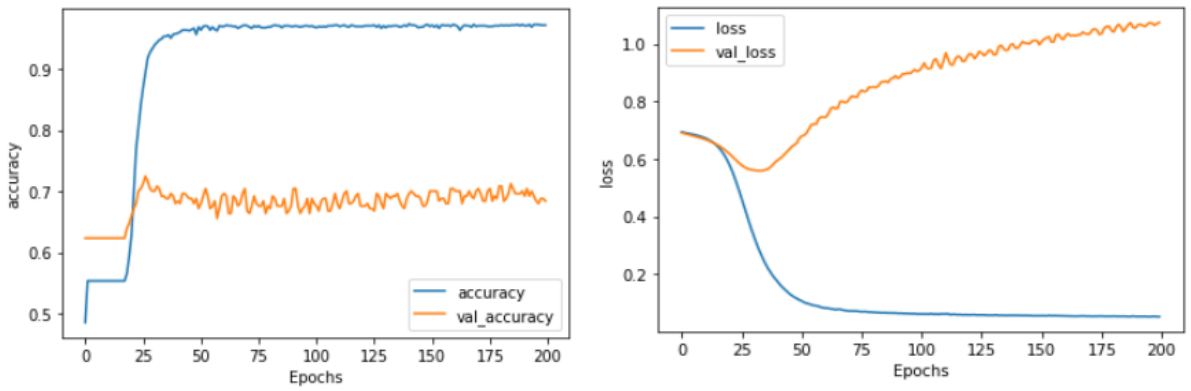


Figure 5. 5 CNN-LSTM Training Vs Validation accuracy and loss.

Table 5. 6 CNN-LSTM performance

Performance measure	Domain of class	
	Reliable	**Fake**
Precision	0.85	0.72
Recall	0.80	0.78
F Score	0.82	0.76
Macro-Precision	0.78	
Macro-Recall	0.79	

Macro-F score	0.78
Accuracy	0.79

Both the above experiment with nearest probability configuration BILSTM is less performance measurement when it compared with the first base LSTM model. Both contain their own feature when it observed from the previous researches. These research list out different point to compare these two models. So, their difference in configuration and feature gives awareness. As the above result describes LSTM better on the numerical measurement. Also, BILSTM on our experiment take a lot of time to train when it compares to the LSTM mode.

LSTM (Long Short-Term Memory)

LSTM network switch with 8 layers and 2 dense defined output layers. 0.2 drop at the end of each trial, a complete 2*2 matrix was formed, and accuracy was determined for the resulting network, as well as precision recall and f-measurement for each target neuron detection rate. For the purpose classification, the target production of the highest numerical value was used. Up to 200 epochs were used to train the networks. The trained network's performance was measured at 200 epochs, with extra epochs added above 200 to see whether performance could be improved 500, 700, and 1000 epochs resulted in no change in training and validation tests above 200 epochs.

This method works well, but there are times when CNN or other deep learning models do not perform well. It's happened to me a couple of times. My data was good, the model's architecture was well-defined, and the loss function and optimizers were appropriately established, however my model continued falling short of my expectations.

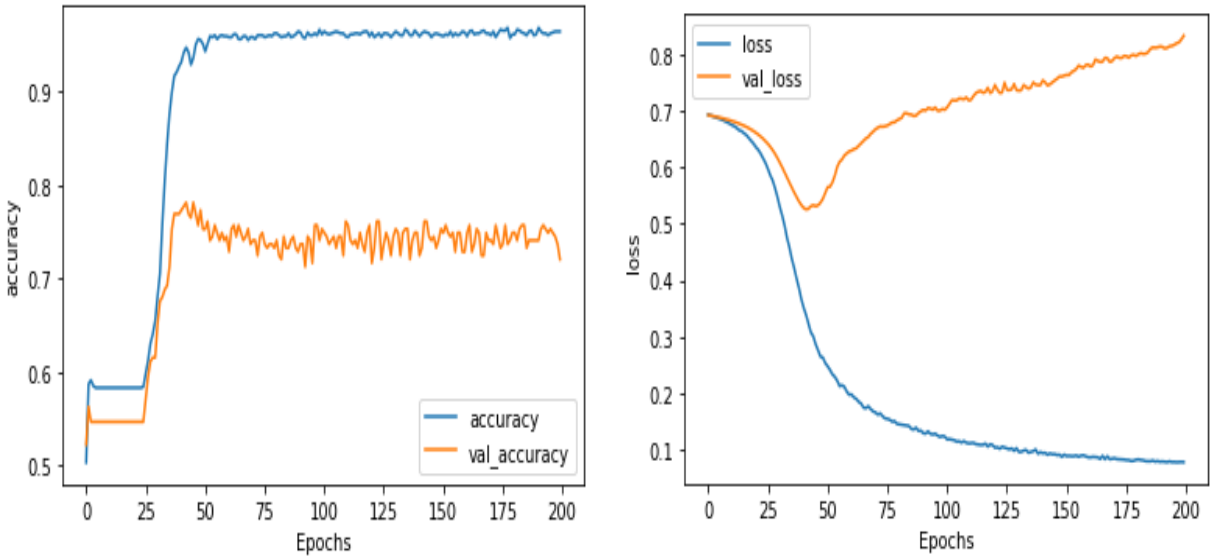


Figure 5. 6 LSTM Training vs Validation accuracy and loss

Table 5. 7 LSTM performance

Performance measure	**Domain of class**	
	Reliable	**Fake**
Precision	0.76	0.65
Recall	0.62	0.78
F Score	0.68	0.71
Macro-Precision	0.71	
Macro-Recall	0.70	
Macro-F score	0.69	
Accuracy	0.70	

A model that has been trained on more full data would automatically generalize better. When that is no longer an option, the next best option is to employ techniques such as regularization. These limit the amount and type of information that model can hold. If a network can only memorize a limited number of patterns, the optimization process will encourage it to focus on the most prominent patterns, which have a better chance of generalizing well. Avoid over fitting is to use more thorough training data. The dataset should include all of the inputs that the model is

supposed to handle. Additional data may be valuable only if it includes new and interesting examples.

However, if you train for an inordinate amount of time, the model will begin to over fit and learn patterns from the training data that do not generalize to the test data. We must establish a balance. Understanding how to train for the right number of epochs, as we'll see later, is a useful ability. For this research we try all patterns epochs regularize, Dropout. But still low performances compared to other selected model we believe data set amount will affect the final result

CNN (convolutional neural network)

The filter's data is 32 bytes in size, and it has a maximum pooling layer of 8 kernels. After each convolution layer, we applied a max pooling layer, and then concatenated the different max pooling layers to create a single fixed-size output. A LSTM layer was present. There was a layer that was completely connected. To define the class of each output, the soft-max activation function translated the vector into probability. The “Sparse-categorical-cross entropy” loss function was used to compile the model. The optimizer "Adam" was utilized. With 200 epochs, this combination obtained 68 percent accuracy, demonstrating LSTM's superiority over CNN.

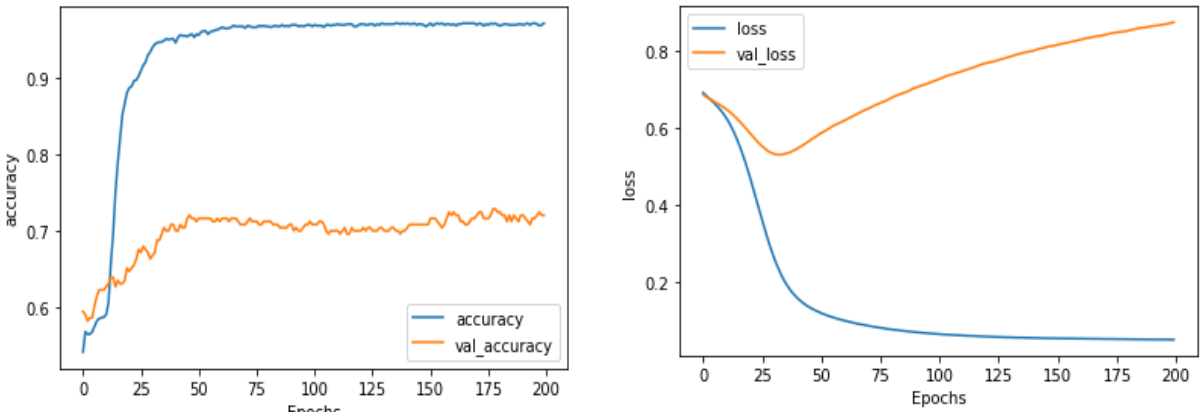


Figure 5. 7 CNN Training Vs Validation accuracy and loss.

Table 5. 8 CNN performance

Performance measure	Domain of class	
	Reliable	**Fake**
Precision	0.76	0.51
Recall	0.76	0.51
F Score	0.78	0.53
Macro-Precision	0.64	
Macro-Recall	0.64	
Macro-F score	0.64	
Accuracy	0.68	

BI-LSTM

The model properties for this Bidirectional long short-term memory change the drop outs defined 0.2 drop out and activation function for BILSTM "soft-max" activation used. To train the model for this experiment, use the. LSTM network switch with 20 layers and two densely defined output layers. At the end of each trial, a complete 2*2 confusion matrix was created, accuracy was calculated for the resulting network, and precision recall and f-measure were calculated for each target neuron detection rate.

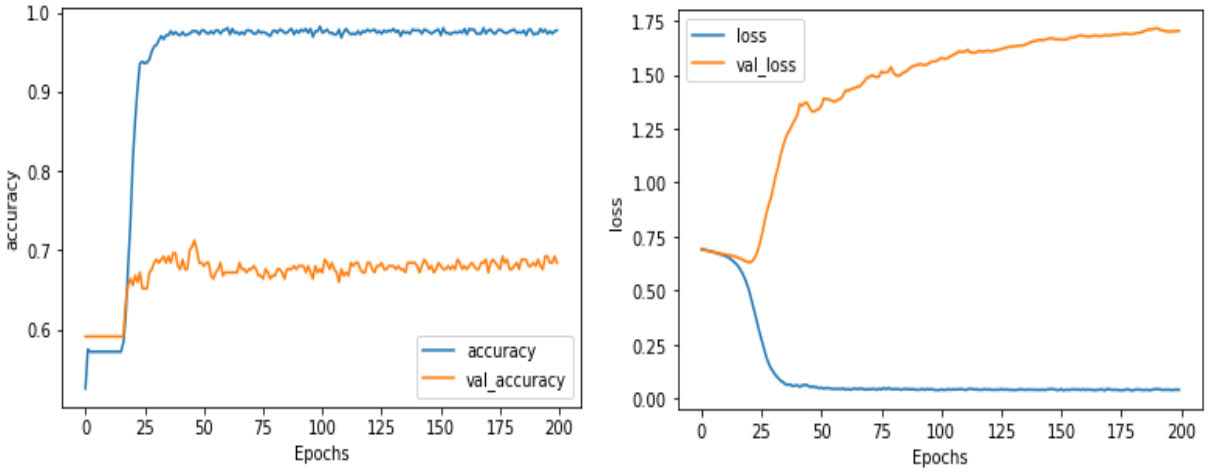


Figure 5. 8 BI-LSTM Training Vs Validation accuracy and loss.

Table 5. 9 BI-LSTM performance

Performance measure	Domain of class	
	Reliable	**Fake**
Precision	0.85	0.79
Recall	0.79	0.85
F Score	0.82	0.81
Macro-Precision	0.82	
Macro-Recall	0.84	
Macro-F score	0.83	
Accuracy	0.82	

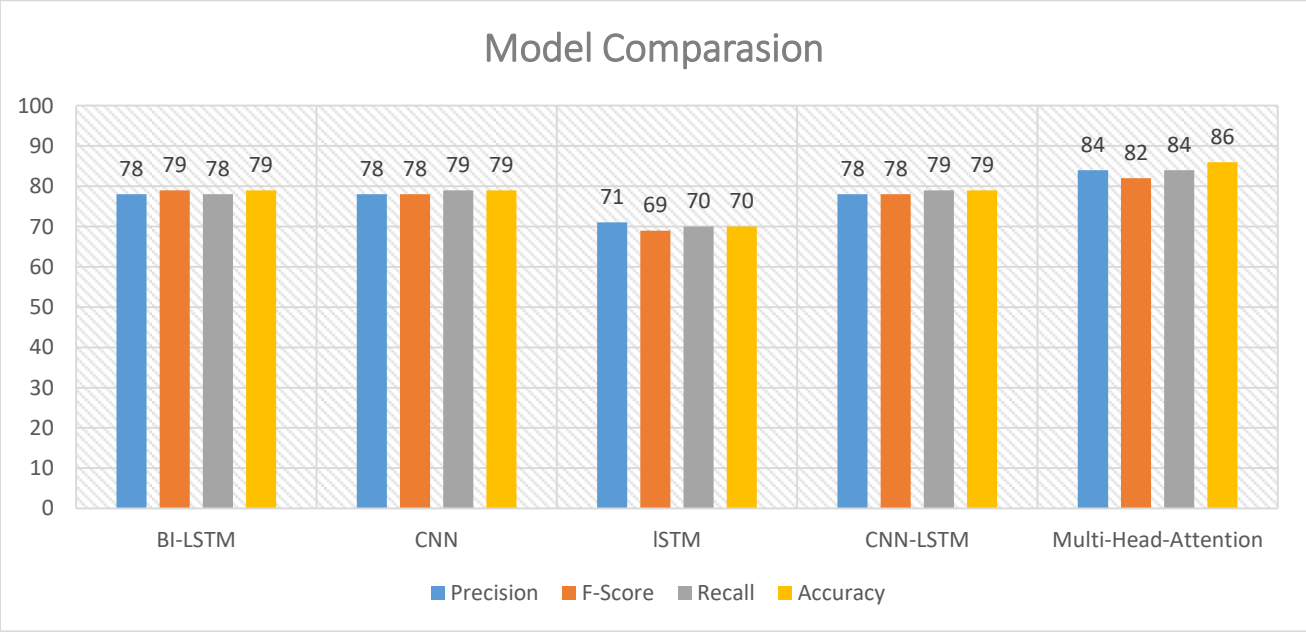


Figure 5. 9 proposed model comparison

RQ2) How to identify fake news that is published in the Amharic language?

Why the Consumers' increasing distrust about media credibility is diverse, but a fundamental component in the avalanche of fake news. Adults in many countries globally mistakenly believed that a news item is true until later it was false and the failure to rely on its honesty was one of the major reasons to avoid the news. Avoid jumping to the conclusion that all content is phony.

Following every rumor or conspiracy idea might be just as dangerous. Based on the finding s and other litrature we are listed the two basics criteria to decide real or fake.

1 **Professional** global news organizations like FANA, WALTA, and the ETV have strict editorial guidelines and vast networks of highly educated reporters, so they're a good bench marks to start a labling our data as real based on those company stracture and profetionalism.

1 **When** an important news event takes place, several media outlets cover the story even if the story is not broken first. Search for further articles on the event or the subject. If no other news outlets are covering the story, be skeptical of the accuracy of the article. Based on these two measures we are define our data label which fake or reliable, that news companies stream and the same content published with the reliable news companies is our measurement of to define the news is reliable.

Table 5. 10 News headlines cover with trusted news organizations and other channels

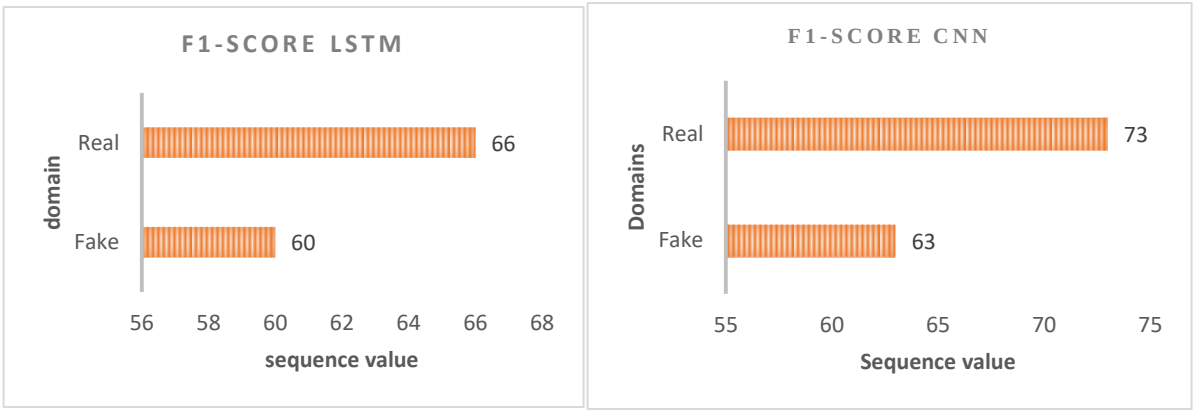
	NEWS	ETV	WALTA	FANA	Hageregna	Seber Zena
1	የኦሊምፒክ ማጣሪያ ተሳታፊዎች የሚለዩበት ቻምፒዮና እየተካሄደ ይገኛል	✓	✓	✓	X	X
2	ጠቅላይ ፍርድ ቤቱ ለፍትሕ ጥራት ትኩረት መስጠቱን ገለጸ	✓	✓	✓	X	X
3	ጠቅላይ ፍርድ ቤቱ ለፍትሕ ጥራት ትኩረት መስጠቱን ገለጸ	✓	✓	✓	X	X
4	ከሳኡዲ አረቢያ 600 ኢትዮጵያዊያን ወደ አገራቸው ተመለሱ	✓	✓	X	X	X
5	ምክር ቤቱ 110 የነበሩ የአስፈፃሚ ተቋማትን ወደ 63 ዝቅ አደረገ	✓	✓	✓	X	X
6	የሶማሌያ ዋና ከተማ ሞቃዲሾ በወታደሮች ተኩስ ስትናጥ አድራለች።	X	X	X	✓	✓

Stop word effect on Amharic fake news detection

Stop and function words must be deleted because they have no discriminating capability for retrieving information. Amharic has its own stopping words because Amharic language has no conventional stop terms.

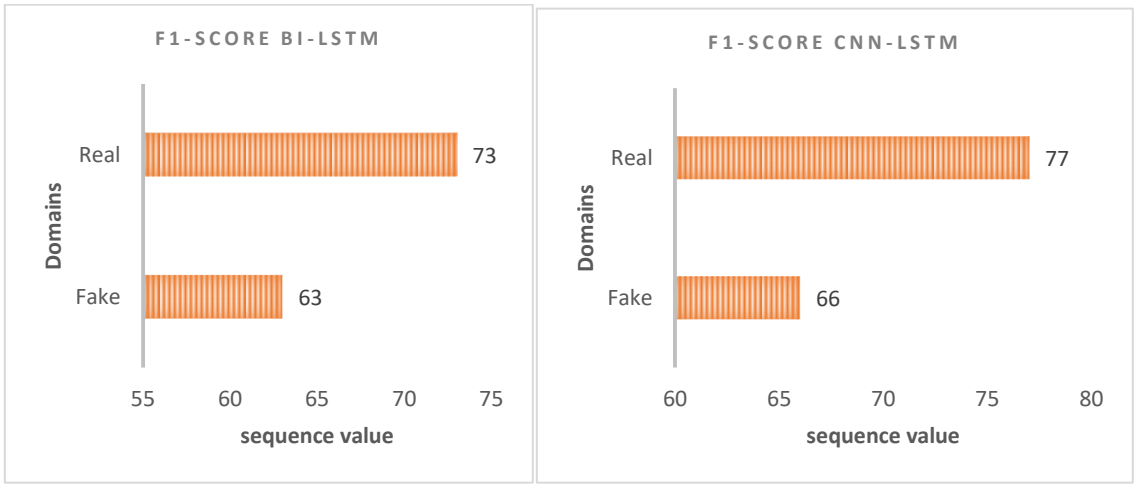
Table 5. 11 sample of Amharic stop word

የሚሆነው	አስከ	ሁኔታ	በሆነው
ከርሱ	ይኸኛው	መሰረት	በአንድ
ስለ	ጀምሮ	ምንም	ከላይ
ከሰላሳ	መሆኑን	አንቀጽ	ብሎ



LSTM after pre stop word removal

CNN after pre stop word removal



BI-LSTM after pre stop word removal

CNN-LSTM after pre stop word removal

Figure 5. 10 After stop word applied model performance

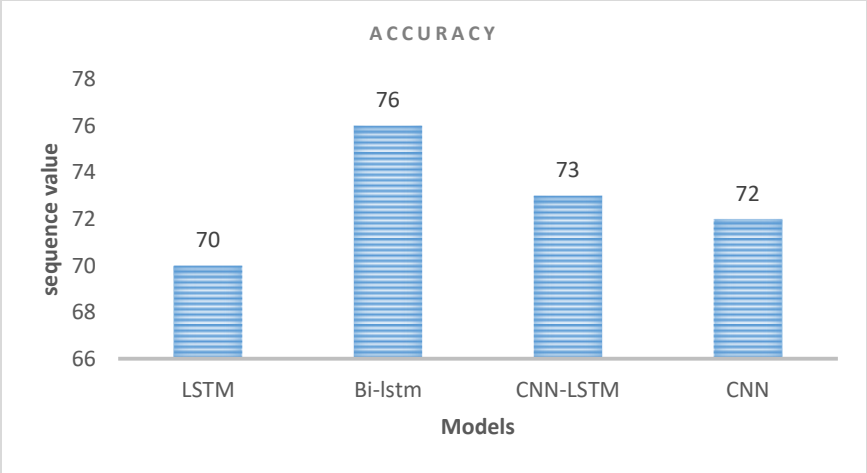


Figure 5. 11 Model comparison after applying stop words process

This research examines the effect of stop word removal implementation on the preprocessing step toward the accuracy of fake news detection in Amharic text documents. The results indicated that the stop word removal technique on the preprocessing stage does not significantly affect the accuracy of fake news detection process with the accuracy neither reduce the accuracy of proposed models on both confusion matrix and k-fold cross-validation test.

There is no universal stop words list because a term can be meaningless depending on the corpus or the problem being studied [46]. Especially on our context in Amharic language is no common collection of stop words and varies to one to other research. Any word can be a stop word. That is why there is so much disagreement regarding whether or not stop words should be eliminated.

Reducing the amount of the data collection is without a doubt a strategy to improve performance. Training models takes time, and if there are fewer tokens to train, the training time should be reduced.

Furthermore, approaches like TF-IDF place a higher emphasis on unusual terms than on excessively repetitious tokens. Assume you're working on a problem for a bank involving one of their products. Because the name of this product may be a word with a high frequency on our corpus, may consider it a stop word.

As a result, stop words are ineffective for evaluations. In any case, preserve these terms and run some tests with and without them to see how they alter the model. In any case, you should never

eliminate stop words without considering their impact on the problem you are attempting to solve.

In this research we conclude that to take assessment on stop word removal analysis which effect the model performance on selected models during data amount and type of data which would be vary results for Amharic corpus we analyze the absence of common stop word lists and amount of data would be consider. Our data is less during the absence of structured data set collection for both reliable organization and other service broadcasts. We conclude that on preprocessing stage adding stop words not affect the models based on the reasons mentioned above amount of data specially considered in this research. For large data set might be an effect on removing stop words

CHAPTER SIX

CONCLUSIONS AND FUTURE WORKS

6.1. Conclusions

The main goal of this research work is to address one of social media's major drawbacks: the rapid spread of fake news, which frequently misleads people, creates false perceptions, and has a negative impact on society. To that goal, an ensemble-based deep learning model is developed to determine if news is fake or reliable. Several preparation approaches were used on the dataset at first. In addition, NLP techniques were used on the statement attribute.

Three deep learning and two hybrid CNN-LSTM and Attention based BI-LSTM model are used. The result achieved by the proposed study is significant with an accuracy of 0.86, when the proposed system compared with other models such as BI-LSTM, CNN, LSTM, and CNN-LSTM by 7%, 7%, 16%, and 7% respectively. This research also concludes that incorporating stop word removal has no any significant contribution. The detection of false news is quite successful. Finally, deep learning for fake news identification is still a new and complex field.

6.2. Future works

Developing economical fully-functional fake news detection needs coordinated team efforts that comprise linguistic skilled, technology skilled and others that may collect additional reliable and fake news from each public pages and streaming network. The research identifies the following directions as a future work.

1. This research is limited only on fake news detection from textual contents. The researchers believe that adding other fake news detection for photos, videos on a single model will have a great role for the system.

2. This research on fake news detection is limited to on Amharic language only. However, the reliable companies such as FANA, WALTA, and ETV also have different local language streaming with Tigrigna, and Afan Oromo. For next works adding these languages will help in developing a full model of fake news detection.

3. When conducting our experiments, small amount of corpus for fake news detection method used. Thus, large data to train classifier believed better results can be achieved in the future.

REFERENCES

[1] M. Granik and V. Mesyura, “Fake news detection using naive Bayes classifier,” 2017 IEEE 1st Ukr. Conf. Electr. Comput. Eng. UKRCON 2017 - Proc., pp. 900–903, 2017, doi: 10.1109/UKRCON.2017.8100379.

[2] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake News Detection on Social Media: A Data Mining Perspective,” vol. 19, no. 1, pp. 22–36, 2017, [Online]. Available: http://arxiv.org/abs/1708.01967.

[3] G. Guibon et al., “Multilingual Fake News Detection with Satire,” no. April, 2019.

[4] Choraś, M., Giełczyk, A., Demestichas, K., Puchalski, D., & Kozik, R. (2018, September). Pattern recognition solutions for fake news detection. In IFIP International Conference on Computer Information Systems and Industrial Management (pp. 130-139). Springer, Cham

[5] Gilda, S. (2017, December). Notice of Violation of IEEE Publication Principles: Evaluating machine learning algorithms for fake news detection. In 2017 IEEE 15th student conference on research and development (SCOReD) (pp. 110-115). IEEE.

[6] Granik, M., & Mesyura, V. (2017, May). Fake news detection using naive Bayes classifier. In 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON) (pp. 900-903). IEEE.

[7] Ruchansky, N., Seo, S., & Liu, Y. (2017, November). Csi: A hybrid deep model for fake news detection. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 797-806).

[8] A. Thota, P. Tilak, S. Ahluwalia, N. Lohia, S. Ahluwalia, and N. Lohia, “Fake News Detection: A Deep Learning Approach,” SMU Data Sci. Rev., vol. 1, no. 3, p. 10, 2018, [Online].Available:https://scholar.smu.edu/datasciencereviewhttp://digitalrepository.smu.edu.Availableat:https://scholar.smu.edu/datasciencereview/vol1/iss3/10.

[9] S. Girgis and M. Gadallah, “’ Hhs / Hduqlqj $ Ojrulwkpv Iru ’ Hwhfwlqj) Dnh 1Hzv Lq 2Qolqh 7H [W,” 2018 13th Int. Conf. Comput. Eng. Syst., pp. 93–97, 2018.

[10] Thota, A., Tilak, P., Ahluwalia, S., & Lohia, N. (2018). Fake news detection: A deep learning approach. SMU Data Science Review, 1(3), 10.

[11] S. I. Manzoor, J. Singla, and Nikita, "Fake news detection using machine learning approaches: A systematic review," Proc. Int. Conf. Trends Electron. Informatics, ICOEI 2019, no. Icoei, pp. 230–234, 2019, doi: 10.1109/ICOEI.2019.8862770.

[12] A. Roy, K. Basak, A. Ekbal, and P. Bhattacharyya, "A Deep Ensemble Framework for Fake News Detection and Classification," 2018, [Online]. Available: http://arxiv.org/abs/1811.04670.

[13] Kelly Stahl (2018), Fake news detection in social media. B.S. Candidate, Department of Mathematics and Department of Computer Sciences, California State University Stanislaus, 1 University Circle, Turlock, CA 95382.

[14] A. Roy, K. Basak, A. Ekbal, and P. Bhattacharyya, "A Deep Ensemble Framework for Fake News Detection and Classification," 2018, [Online]. Available: http://arxiv.org/abs/1811.04670.

[15] Gereme, F., Zhu, W., Ayall, T., & Alemu, D. (2021). Combating Fake News in "Low-Resource" Languages: Amharic Fake News Detection Accompanied by Resource Crafting. Information, 12(1), 20.

[16] Kim Y. Convolutional neural networks for sentence classification. ArXiv: 1408.5882v2 [cs.CL]. 2014

[17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N, & Polosukhin, I. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762.

[18] Rahi, S. (2017). Research Design and Methods: A Systematic Review of Research Paradigms, Sampling Issues and Instruments Development. International Journal of Economics & Management Sciences, 06(02), 1–5. https://doi.org/10.4172/2162-6359.1000403

[19] Wondwossen Mulugeta. (2014). A Machine Learning Approach to Multi-Scale Sentiment Analysis of Amharic Online Posts. HiLCoE Journal of Computer Science and Technology, 2(2), 80–87.

[20] Mihret, M., & Atinaf, M. (2019). Sentiment Analysis Model for Opinionated Awngi Text. IEEE AFRICON Conference, 2019-Septe,49–51. https://doi.org/10.1109/AFRICON46755.2019.9134016

[21] György, G. G. (2017). Using machine learning techniques for deanonymization. Informacios Tarsadalom, 7(1), 72–86. https://doi.org/10.22503/inftars.XVII.2017.1.5

[22] Alfina, I., Mulia, R., Fanany, M. I., & Ekanata, Y. (2018). Hate speech detection in the Indonesian language: A dataset and preliminary study. 2017 International Conference on Advanced Computer Science an Information Systems, ICACSIS 2017, 2018-Janua, 233–237. https://doi.org/10.1109/ICACSIS.2017.8355039

[23] Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. Progress in Artificial Intelligence, 5(4), 221–232. https://doi.org/10.1007/s13748-016-0094-0

[24] Hashemi, H. B., Asiaee, A., & Kraft, R. (2016). Query Intent Detection using Convolutional Neural Networks. WSDM QRUMS 2016 Workshop. https://doi.org/10.1145/1235

[25] Sikdar, U. K., & Gambäck, B. (2018). Named entity recognition for amharic using stack-based deep learning. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10761 LNCS, 276–287. https://doi.org/10.1007/978-3-319-77113-7_22

[26] Le, X.-H., Ho, H. V., Lee, G., & Jung, S. (2019). Application of long short-term memory (LSTM) neural network for flood forecasting. Water, 11(7), 1387.

[27] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

[28] Nowak, J., Taspinar, A., & Scherer, R. (2017). LSTM recurrent neural networks for short text and sentiment classification. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10246 LNAI, 553–562. https://doi.org/10.1007/978-3-319-59060-8_50

[29] Azzouni, A., & Pujolle, G. (2017). A Long Short-Term Memory Recurrent Neural Network Framework for Network Traffic Matrix Prediction. ArXiv Preprint ArXiv:1705.05690. http://arxiv.org/abs/1705.05690

[30] Cui, Z., Ke, R., Pu, Z., & Wang, Y. (2018). Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction. ArXiv Preprint ArXiv:1801.02143.

Cui, Z., Ke, R., Pu, Z., & Wang, Y. (2020). Stacked Bidirectional and Unidirectional LSTM Recurrent Neural Network for Forecasting Network-wide Traffic State with Missing Values. ArXiv Preprint ArXiv:2005.11627.

[31] Bahdanau, D., Cho, K. and Bengio, Y., 2014.Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473

[32] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L. and Polosukhin, I., 2017. Attention Is All You Need. [Online]arXiv.org. Available at: <https://arxiv.org/abs/1706.03762>.

[33] Watanabe, H., Bouazizi, M., & Ohtsuki, T. (2018). Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection. IEEE Access, 6, 13825–13835. https://doi.org/10.1109/ACCESS.2018.2806394

[34] Fikre T. , Lemma S. "Effect of preprocessing on Long Short Term Memory Based Sentiment Analysis For Amharic Language"

[35] Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake News Detection Using Machine Learning Ensemble Methods. Complexity, 2020.

[36] Ahmed, H., Traore, I. and Saad, S., 2017, October. Detection of online fake news using n-gram analysis and machine learning techniques. In International conference on intelligent, secure, and dependable systems in distributed and cloud environments (pp. 127-138). Springer, Cham.

[37] K. Stahl, “Fake news detection in online social media Problem Statement,” no. May, p. 6, 2018, [Online]. Available: https://leadingindia.ai/downloads/projects/SMA/sma_9.pdf.

[38] Alias, N., Razak, S. H. A., elHadad, G., Kunjambu, N. R. M. N. K., & Muniandy, P. (2013). A Content Analysis in the Studies of YouTube in Selected Journals. Procedia - Social and Behavioral Sciences, 103, 10–18. https://doi.org/10.1016/j.sbspro.2013.10.301

[39] Al-Yahya, M., Al-Khalifa, H., Al-Baity, H., AlSaeed, D. and Essam, A., 2021. Arabic Fake News Detection: Comparative Study of Neural Networks and Transformer-Based Approaches. Complexity, 2021.

[40] Argaw, M. (2019). Amharic Parts-of-Speech Tagger using Neural Word Embeddings as Features. Addis Ababa University Addis Ababa.

[41] Awalina, A., Fawaid, J., Krisnabayu, R.Y. and Yudistira, N., 2021. Indonesia's Fake News Detection using Transformer Network. arXiv preprint arXiv:2107.06796.

[42] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078

[43] Nicholson, C. (2019). A beginner's guide to word2vec and neural word embeddings.

[44] Fang, Y., Gao, J., Huang, C., Peng, H. and Wu, R., 2019. Self multi-head attention-based convolutional neural networks for fake news detection. PloS one, 14(9), p.e0222713.

[45] Olah, C. (2015). Understanding LSTM Networks [Blog]. Web Page. https://doi.org/10.1007/s13398-014-0173-7.2

[46] Silva, C. and Ribeiro, B., 2003, July. The importance of stop word removal on recall values in text categorization. In Proceedings of the International Joint Conference on Neural Networks, 2003. (Vol. 3, pp. 1661-1666). IEEE.

APPENDICES

Appendix A

Normalized characters

ዐ=አ	ዖ=ኦ	ኀ=ሀ	ሓ=ሀ	ጸ=ፃ	ሤ=ሴ
ዓ=አ	ዉ=ው	ኽ=ህ	ሔ=ሄ	ጼ:ፄ	ሥ=ስ
ዑ=ኡ	ኃ=ሀ	ሐ=ሆ	ሗ=ሆ	ጸ=ፆ	ሦ=ሶ
ዒ=ኢ	ኄ=ሄ	ኃ=ሀ	ጽ=ፅ	ሠ=ሰ	ሒ=ሂ
ዓ=አ	ኍ=ሁ	ሃ=ሀ	ጸ=ፀ	ሣ=ሳ	
ዔ=ኤ	ኂ=ሂ	ሐ=ሀ	ጹ=ፁ	ሡ=ሱ	
ዕ=እ	ኄ=ሄ	ሑ=ሁ	ጺ=ፂ	ሢ=ሲ	

Appendix B

List of Amharic stop words

አ ንቀጽ,ሀ ያ ,የ ሆኑ ትን ,ስ ድስ ት,እ ስከ ,ይኸኛ ው,አ ንቀጽ,ወይም,ሆነ ,በ ሆኑ ,ምን ም,ብሎ,ከ ር ሱ,እ ንደሆነ ,በ ኍላ ,ጀ ምሮ,በ ሆነ ው,ከ ሰ ላ ሳ ,ሚሆኑ ን ,,ላ ይ,በ አ ንድ,እ ንደዚህ ,ስ ለ ,የ ሚሆኑ ,ለ ዚያ ው,ማን ም,የ ሆኑ ,በ ሚሆን ,ይሁን , ላ ይም,ለ ዚሁ,በ ሚበ ል ጥ,ከ አ ስ ራ,የ ሌለ ው,ከ ዚሁ,የ ሆና ል ,ለ እ ነ ር ሱም,,ማህ ረ ት,የ ሆነ ውን ,በ ማለ ት,በ እ ነ ዚ ህ ,ከ ነ ዚህ ,እ ዚሁ,,,,ሁኔ ታ,ከ ላ ይ,ባ ለ ,ከ ሞ ቸውም,ያ ህ ል ,ቢሆን ም,,ሌላ ,ሁሉ,,ይህ ን ን ,ከ ነ በረ ው,ከ ሆነ ና ,እ ና , ሚሆኑ ,እ ን ዲቆ ይ,በ አ ን ዳ ን ድ,እ ስ ከ ,,ይሆና ል ,ሌላ ውን ,ሌላ ው,በ እ ያ ን ዳ ን ዱ,ለ ሆኑ ት,አ ንቀጽ,ሆኖ ,ከ ሰ ባ ት,የ ሚሆነ ው,ኒ ዜ ም,አ ለ ው,ነ ው,ለ ሌላ ,በ አ ን ዱ,አ ስ ከ ,እ ነ ዚሁ,ከ አ ንቀጽ,ሆን ,,በ ዚህ ,አ ለ በ ት,አ ይሆን ም,የ ሌሎች, እ ንደሆኑ ,,አ ንቀጽ,እ ስ ከ ,ሲል ,ማለ ት,የ ሚሆኑ ት,ስ ለ ሞ ቸውም,,ውስ ጥ,ይሆና ሉ,ሲ ባ ል ,ከ ሆነ ው,ስ ለ ዚሁ,ሆን , ከ አ ን ድ,ያ ል ሆነ ,ሳ ለ ,የ ነ በረ ውን ,ከ አ ን ዳ ን ድ,በ ሞ ቸውም,በ ሞሉ,የ ሆነ ው,ያ ሉ,በ እ ነ ዚሁ,,አ ንቀጽ ሀ ,ወር ,አ ስ ራ,ሚሆና ቸው,ከ ሌሎች,በ አ ምስ ት,ከ አ ምስ ት,አ ስ ር ,በ ዋና ,አ ንዲት,ወይም,በ ላ ይ,እ ን ደ,በ ማቁ ድ,ለ ሌሎች,በ ሆ ኑ ,,ሲሆን ,ቢሆን ም,ኒ ዜ ና ,እ ን ኳ ን ,የ ሆኑ በ ታ ል ,,,በ ሆነ ,አ ን ዱ,ለ ዚህ ,ለ ሆነ ው,ለ ነ ዚህ ,,,ከ ዚህ ,የ ሌላ ውን ,ሶ ስ ተኛ ,አ ን ዳ ን ድ,ለ ማን ኛ ውም,,የ ሆነ ,ከ ሁለ ት,የ ነ ገ ሩ,ሰ ኣ ት,አ ን ደኛ ,,,አ ምስ ት,ሶ ስ ት,እ ን ዲሆን ,እ ን ደ ነ ዚህ , ና ቸው,,,ማን ኛ ውም,ካ ል ሆነ ,እ ን ኳ,የ ሆኑ ት,አ ሁን ,ጋ ር ,ቢ ያ ን ስ ,ከ ሀ ያ ,የ ሞ ቸውም,ሰ ባ ት,,አ ን ድ,ቢሆን ,ከ ሀ ም ሳ ,ይህ ን ን ም,እ ነ ደሆነ ,አ ንቀጽ,,ልዩ ,እ ነ ዚህ ን ,ይኸው,የ አ ን ድን ,እ ነ ደሆነ ,ፋ ር ሚኒ ስ ት,,ከ ሶ ስ ት,ና ቸው,ለ አ ን ድ,በ ሞሉ ም,ይህ ችው,በ ተ ለ ይም,አ ንዱን ,የ ሚቾ ለ ውን ,በ ነ ዚህ ,ከ እ ነ ዚህ ,,,በ ሌላ ,ሁለ ት,ወይም,የ ዚሁ,ከ እ ነ ዚ ሁ,በ አ ንቀጽ,,ሺህ ,ለ ዚሁ,በ ሚ ባ ,ለ እ ያ ን ዳ ን ዱ,የ አ ንቀጹ,ወደ ,ወይዘ ሮ,ይህ ም,ስ ለ ሆነ ,ወይ,,ሆነ ,ማ ቸውን ም,ተ ብሎ,እ ነ ዚህ ,ሚሆና ቸውን ,የ ሆነ ችን ,,ከ አ ስ ር ,ሳ ይሆን ,ከ ዚያ ,ማን ኛ ውን ም,የ ለ ውም,የ ሚበ ል ጥ,እ ን ደሆ ነ ና ,እ ን ዲሆኑ ,ሁለ ቱ,በ ሚቾ ሉ,,ብቻ,ብሎ,ከ ሌላ ,እ ን ጂ,የ ሌላ ቸውን ,እ ን ዲሁም,ከ ብር ,ለ ሆነ ,ከ ስ ምን ት,በ ሶ ስ ተ ኛ ,ሌሎች,ሆኖ ም,በ ሌሎች,ሁለ ቱን ም,በ ቀ ር ,,ይህ ,በ ታች,አ ን ደሆነ ,በ ሁለ ት,በ ነ ሱ,ይህ ን ,የ ሌላ ,እ ን ዲህ ,በ እ ስ ር ,የ አ ን ዱን ,,ከ ሆነ ,,ያ ላ ቸው,በ ነ ዚሁ,በ ሚል ,የ አ ን ዱ,የ ዚህ ,ይህ ን ኑ ,በ እ ን ደዚህ ,ቁ ጥር ,,ማ ቸውም,ሆነ ው,ስ ም ን ት,ባ ሉ,በ ዚህ ,አ ንቀጽ,ከ ስ ድስ ት,በ ስ ተቀ ር ,ሲሆን ና ,ከ ሞ ቆ,በ ዚህ ም,,ሞ ቆ,ሚሆን ,ም ን ጊ ዜ ም,እ ነ ዚህ ም,በ ዚ ህ ና ,,ያ ለ ,ስ ም,ለ ማ ቸውም,ሲ ኖ ር ,ከ ዚህ ም,,,ሚሆኑ ን ,እ ን ደገ ና ,የ አ ን ድ,ሰ ላ ሳ ,በ ሁኔ ታ ው,አ ን ድን ,የ ሚያ ን ስ , እ ነ ዚህ ኑ ,ማ ነ ም,ከ ነ ዚሁ,,ያ ላ ቸውን ,እ ጅግ,ሲሆኑ ,ለ ሆኑ ,ሌሎች,,,ሊሆን ,ግን ,በ ሁለ ቱም,አ ለ ,ይህ ን

Appendix C

Attention based Bi-LSTM model

```
class Attention(tf.keras.Model):
    def __init__(self, units):
        super(Attention, self).__init__()
        self.W1 = tf.keras.layers.Dense(units)
        self.W2 = tf.keras.layers.Dense(units)
        self.V = tf.keras.layers.Dense(1)

    def call(self, features, hidden):
        # hidden shape == (batch_size, hidden size)
        # hidden_with_time_axis shape == (batch_size, 1, hidden size)
        # we are doing this to perform addition to calculate the score
        hidden_with_time_axis = tf.expand_dims(hidden, 1)

        # score shape == (batch_size, max_length, 1)
        # we get 1 at the last axis because we are applying score to self.V
        # the shape of the tensor before applying self.V is (batch_size, max_length, units)
        score = tf.nn.tanh(
            self.W1(features) + self.W2(hidden_with_time_axis))
        # attention_weights shape == (batch_size, max_length, 1)
        attention_weights = tf.nn.softmax(self.V(score), axis=1)

        # context_vector shape after sum == (batch_size, hidden_size)
        context_vector = attention_weights * features
        context_vector = tf.reduce_sum(context_vector, axis=1)
    return context_vector, attention_weights
```

LSTM model

```
NB_WORDS = 20000 # Parameter indicating the number of words we'll put in the dictionary

VAL_SIZE =1000   # Size of the validation set

EPOCHS = 200   # Number of epochs we usually start to train with

BATCH_SIZE = 512 # Size of the batches used in the mini-batch gradient descent

X_train, X_test, y_train, y_test = train_test_split(df.message,

df.id, test_size=0.1, random_state=37)

print('Training Data:', X_train.shape[0])

print('Test Data:', X_test.shape[0])

tweet_df = df[['message','id']]

tk = Tokenizer(num_words=NB_WORDS,split=" ")

tk.fit_on_texts(X_train)

# print('Top 5 most common words are:', collections.Counter(tk.word_counts).most_common(5))

sentiment_label = tweet_df.id

len(df.message)

print(tk.word_counts)

#print(tk.word_index)

#print(sentiment_label)

# using LSTM model

model = models.Sequential()

model.add(layers.Embedding(NB_WORDS, 8, input_length=MAX_LEN))

model.add(layers.LSTM(8))

model.add(layers.Dropout(0.8))
```

```
model.add(layers.Dense(2, activation='softmax'))

model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
```

BI-LSTM MODEL

```
NB_WORDS = 20000 # Parameter indicating the number of words we'll put in the dictionary

VAL_SIZE =1000   # Size of the validation set

EPOCHS = 200   # Number of epochs we usually start to train with

BATCH_SIZE = 512 # Size of the batches used in the mini-batch gradient descent

X_train, X_test, y_train, y_test = train_test_split(df.message,

df.id, test_size=0.1, random_state=37)

print('Training Data:', X_train.shape[0])

print('Test Data:', X_test.shape[0])

#using BI-LSTM

from keras.callbacks import ModelCheckpoint

model =models.Sequential()

model.add(layers.Embedding(NB_WORDS, 8, input_length=MAX_LEN))

model.add(layers.Bidirectional(layers.LSTM(20)))

model.add(layers.Dropout(0.2))

model.add(layers.Dense(5,activation='softmax'))

print(model.summary())

filepath="best_model.hdf5"

model.compile(loss='categorical_crossentropy',optimizer='adam', metrics=['accuracy'])

checkpoint = ModelCheckpoint(filepath, monitor='val_accuracy',
verbose=1,save_best_only=True, mode='max',save_weights_only=False)
```

```
callbacks_list = [checkpoint]
```

CONVOLUTIONAL NEURAL NETWORK (CNN)

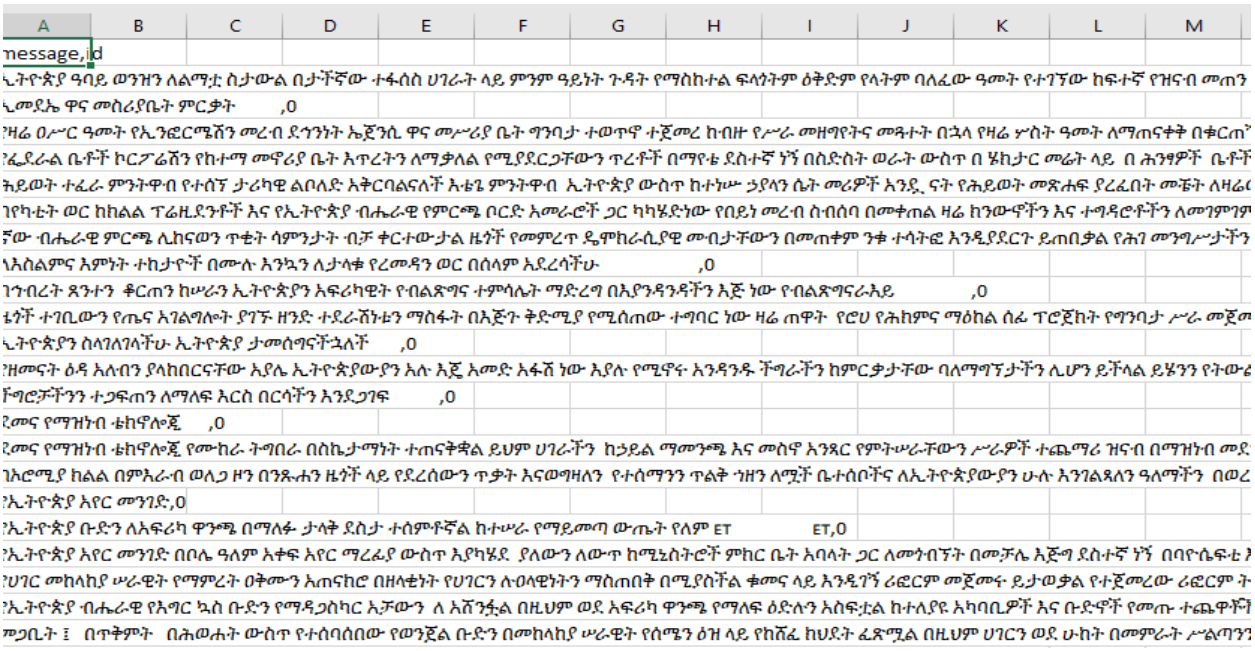
```
NB_WORDS = 20000 # Parameter indicating the number of words we'll put in the dictionary

VAL_SIZE =1000   # Size of the validation set

EPOCHS = 200   # Number of epochs we usually start to train with

BATCH_SIZE = 512 # Size of the batches used in the mini-b  atch gradient descent

X_train, X_test, y_train, y_test = train_test_split(df.message,

df.id, test_size=0.1, random_state=37)

print('Training Data:', X_train.shape[0])

print('Test Data:', X_test.shape[0])

# using convolutional model

model =models.Sequential()

model.add(layers.Embedding(NB_WORDS, 8, input_length=MAX_LEN))

model.add(layers.Conv1D(filters=32, kernel_size=8, activation='relu'))

model.add(layers.MaxPooling1D(pool_size=2))

model.add(layers.Flatten())

model.add(layers.Dense(2, activation='softmax'))

print(model.summary())

model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
```

CNN-LSTM

```
NB_WORDS = 20000 # Parameter indicating the number of words we'll put in the dictionary

VAL_SIZE =1000   # Size of the validation set

EPOCHS = 200   # Number of epochs we usually start to train with
```

```
BATCH_SIZE = 512 # Size of the batches used in the mini-batch gradient descent

X_train, X_test, y_train, y_test = train_test_split(df.message,

df.id, test_size=0.1, random_state=37)

print('Training Data:', X_train.shape[0])

print('Test Data:', X_test.shape[0])

#using CNN-LSTM

model = models.Sequential()

model.add(layers.Embedding(NB_WORDS, 8, input_length=MAX_LEN))

model.add(layers.Conv1D(filters=32, kernel_size=3, padding='same', activation='relu'))

model.add(layers.MaxPooling1D(pool_size=2))

model.add(layers.LSTM(8))

model.add(layers.Dense(5, activation='softmax'))

model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])

print(model.summary())

filepath="weights_best_cnn.hdf5"

checkpoint = ModelCheckpoint(filepath, monitor='val_acc', verbose=1, save_best_only=True,
mode='max',save_weights_only=False)

callbacks_list = [checkpoint]
```

Appendix D



A	B	C	D	E	F	G	H	I	J	K	L	M
message_id												
ኢትዮጵያ ዓባይ ወንዝን ለልማቷ ስታውል በታችኛው ተፋሰስ ሀገራት ላይ ምንም ዓይነት ጉዳት የማስከተል ፍላጎትም ዕቅድም የላትም ባለፈው ዓመት የተገኘው ከፍተኛ የዝናብ መጠን												
ኢመያኤ ዋና መስሪያቤት ምርቃት ,0												
የዛሬ 0ሥር ዓመት የኢንፎርሜሽን መረብ ደኅንነት ኤጀንሲ ዋና መሥሪያ ቤት ግንባታ ተወጥኖ ተጀመረ ከብዙ የሥራ መዘግየትና መጓተት በኋላ የዛሬ ሦስት ዓመት ለማጠናቀቅ በቁርጠ												
የፈደራል ቤቶች ኮርፖሬሽን የከተማ መኖሪያ ቤት እጥረትን ለማቃለል የሚያደርጋቸውን ጥረቶች በማየቱ ደስተኛ ነኝ በስድስት ወራት ውስጥ በ ሄክታር መሬት ላይ በ ሕንፃዎች ቤቶች												
ሕይወት ተፈራ ምንትዋብ የተሰኘ ታሪካዊ ልቦለድ አቅርባልናለች እቴጌ ምንትዋብ ኢትዮጵያ ውስጥ ከተነሡ ኃያላን ሴት መሪዎች አንዷ ናት የሕይወት መጽሐፍ ያረፈበት መቼት ለዛሬ												
ገየካቲት ወር ከከልል ፕሬዚደንቶች እና የኢትዮጵያ ብሔራዊ የምርጫ ቦርድ አመራሮች ጋር ካካሄድነው የበይነ መረብ ስብሰባ በመቀጠል ዛሬ ከንውጥኖን እና ተግዳሮቶችን ለመገምገም												
ኛው ብሔራዊ ምርጫ ሊከናወን ጥቂት ሳምንታት ብቻ ቀርተውታል ዜጎች የመምረጥ ዴሞክራሲያዊ መብታቸውን በመጠቀም ንቁ ተሳትፎ እንዲያደርጉ ይጠበቃል የሕገ መንግሥታችን												
ነእስልምና እምነት ተከታዮች በሙሉ እንኳን ለታላቁ የረመዳን ወር በሰላም አደረሳችሁ ,0												
ገኘበረት ጸንተን ቆርጠን ከሠራን ኢትዮጵያን አፍሪካዊት የብልጽግና ተምሳሌት ማድረግ በእያንዳንዳችን እጅ ነው የብልጽግናሩአይ ,0												
ቴጎች ተገቢውን የጤና አገልግሎት ያገኙ ዘንድ ተደራሽነቱን ማስፋት በእጅጉ ቅድሚያ የሚሰጠው ተግባር ነው ዛሬ ጠዋት የሮሀ የሕክምና ማዕከል ሰፊ ፕሮጀክት የግንባታ ሥራ መጀመ												
ኢትዮጵያን ስላገለገላችሁ ኢትዮጵያ ታመሰግናችኋለች ,0												
የዘመናት ዕዳ አለብን ያላከበርናቸው አያሌ ኢትዮጵያውያን አሉ እጇ አመድ አፋሽ ነው እያሉ የሚኖሩ አንዳንዱ ችግራችን ከምርቃታችው ባለማግኘታችን ሊሆን ይችላል ይሄንን የትውፊ												
ቮግሮቻችንን ተጋፍጠን ለማለፍ እርስ በርሳችን እንደጋገፍ ,0												
የመና የማዝነብ ቴክኖሎጂ ,0												
የመና የማዝነብ ቴክኖሎጂ የሙከራ ትግበራ በስኬታማነት ተጠናቅቋል ይህም ሀገራችን ከኃይል ማመንጫ እና መስኖ አንጻር የምትሠራቸውን ሥራዎች ተጨማሪ ዝናብ በማዝነብ መደ												
ገኦሮሚያ ክልል በምእራብ ወለጋ ዞን በንጹሐን ዜጎች ላይ የደረሰውን ጥቃት እናወግዛለን የተሰማንን ጥልቅ ኀዘን ለሟች ቤተሰቦችና ለኢትዮጵያውያን ሁሉ እንገልጻለን ዓለማችን በወረ												
የኢትዮጵያ አየር መንገድ,0												
የኢትዮጵያ ቡድን ለአፍሪካ ዋንጫ በማለፉ ታላቅ ደስታ ተሰምቶኛል ከተሠራ የማይመጣ ውጤት የለም ET ET,0												
የኢትዮጵያ አየር መንገድ በቦሌ ዓለም አቀፍ አየር ማረፊያ ውስጥ እያካሄደ ያለውን ለውጥ ከሚኒስትሮች ምክር ቤት አባላት ጋር ለመነጋገር በመቻሌ እጅግ ደስተኛ ነኝ በባዮሴፍቲ እ												
የሀገር መከላከያ ሠራዊት የማምረት ዐቅሙን አጠናክሮ በዘላቂነት የሀገርን ሉዐላዊነትን ማስጠበቅ በሚያስችል ቁመና ላይ እንዲገኝ ሪፎርም መጀመሩ ይታወቃል የተጀመረው ሪፎርም ት												
የኢትዮጵያ ብሔራዊ የእግር ኳስ ቡድን የማዳጋስካር አቻውን ለ አሸንፏል በዚህም ወደ አፍሪካ ዋንጫ የማለፍ ዕድሉን አስፍቷል ከተለያዩ አካባቢዎች እና ቡድኖች የመጡ ተጨዋቾች												
ማጋቢት ፤ በጥቅምት በሕወሐት ውስጥ የተሰባሰበው የወንጀል ቡድን በመከላከያ ሠራዊት የሰሜን ዕዝ ላይ የከሸፈ ከህደት ፈጽሟል በዚህም ሀገርን ወደ ሁከት በመምራት ሥልጣንን												