Bilingual Hate Speech Detection on Social Media : Amharic and Afaan Oromo

Teshome Mulugeta Ababu[1*], Michael Melese Woldeyohannis[2] and Emuye Bawoke[3]

[1]Department of Computer Science, Dire Dawa University Institute of Technology, Dire Dawa, Ethiopia.
[2*]School of Information Science, Addis Ababa University, Addis Ababa, Ethiopia.
[3]Department of Information Technology , Dire Dawa University Institute of Technology, Dire Dawa, Ethiopia.

*Corresponding author(s). E-mail(s): teshome.mulugeta@ddu.edu.et;
Contributing authors: michael.melese@aau.edu.et; emuyeb27@gmail.com;

Abstract

Due to significant increases in internet penetration and the development of smartphone technology during the preceding couple of decades, many people have started using social media as a communication platform. Social media has grown to be one of the most significant components, with several benefits. However, technology also poses a number of threats, challenges, and barriers, such as hate speech, disinformation, and fake news. Hate speech detection is one of the many ways social media platforms can be accused of not doing enough to thwart hate speech on their platform. People in Bilingual and multinational societies commonly employ a code mix in both spoken and written communication. Among these, Amharic and Afaan Oromo language speakers frequently mix the two languages when conversing and posting on social media. The majority of previous study concentrated on identifying either technological favoured language or monolingual hate speech in Ethiopian languages; however, the availability of Bilingual communication in social media hampers the propagation of hate speech via social media. In this work, a Bilingual hate speech detection for Amharic and Afaan

Oromo languages were conducted using four different deep learning classifiers (CNN, BiLSTM, CNN-BiLSTM, and BiGRU) and three feature extraction (Keras word embedding, word2vec, and FastText) techniques. According to the experiment, BiLSTM with FastText feature extraction is an outperforming the other algorithm by achieving a 78.05% percent of accuracy for Bilingual Amharic Afaan Oromo hate speech detection. The FastText feature extraction overcomes the problem of out of vocabulary (OOV). Furthermore, we are working towards including others linguistic features of the languages to detect hate speech and make the resource available to facilitate further research in the area of Bilingual hate speech detection for other under-resourced Ethiopian languages.

Keywords: Amharic, Afaan Oromo, Hate Speech, Bilingual, Under-resourced, Ethiopian languages

1 Introduction

Nowadays, social media has become an integral element of human life interaction, owing to the exponential development in Internet and social media use over the last two decades [1]. Social media is a popular, important, and easy way for people to publicly share their ideas and opinions as well as interact with others [2]. Social media allows people to connect with all over the world who have diverse backgrounds, cultures, languages, ethnicity, lifestyles, and traditions [3]. According to recent global statistics, 63.1% of the world's population is Internet users, with 4.1% not using social media and increasing at an alarming rate [4]. Today's social media platforms are expanding and developing to satisfy a broader range of customers through the technology that enables people across the world to stay connected to social media. Among these, 99% of social media user access resources using websites or apps through the handheld devices.

Social media creates a significant advantage in creating organic content, paid advertising services, branding, performance evaluation, free and viral content that uncover valuable insights [5]. However, social media has also negative feedback, Lack of emotional connection, the spread of hate speech and people spend plenty of time passively consuming information which would need to be moderate or controlled. Among these problems, the online spread of hate speech results in conflict, tension, violence, and stereotyping between the online communities [6, 7]. The genocide of the Rohingya community in Myanmar, the anti-Muslim mob in Sri Lanka, the Pittsburgh shooting, Rwanda genocide in 1994, Kenya post-election in 2008, Burundi election in 2015 and recent south Africa xenophobia is the result of online hate speech [6]. These are only a few of the more serious cases where hate speech-related offenses have escalated in recent years. Similarly, as a result of hate speech, fake news,

and misinformation, several conflicts have emerged throughout Ethiopia, primarily in the Amhara, Oromia and Benishangul regional states as well as the nation's capital, Addis Ababa [8, 9].

Hate speech is a crime in any country and it is a global problem in which the definition of hate speech is neither universal nor individual facets definition fully agreed upon it [10, 11]. Hate speech is the speech content that promotes violence or attack against an individual or group of community-based on protected characteristics such as ethnic origin, disability, religion, gender, race, color, and sexual orientation [12–15]. Similarly, offensive speech is speech that offends someone and causes someone to feel upset, insulting, angry, disgusting, hurt, and annoyed [16]. Offensive speech is also called bad language the wrong choice of words, vulgar language, swear words, expletives, curse words, and profanity, and is used to reflect behavior that is offensive or lacking in respect to others [17]. Detecting hate and offensive speech on social media is very important to reduce crime and protect people's beliefs [5].

Ethiopia is the second largest and most populous country in Africa next to Nigeria with a total population of more than 120 million consisting a 57% of the people age range from 15 to 65, 39% of the people under the age 15, while the remaining 4% for aged one [18]. The country is located in the eastern part of Africa commonly known as the Horn of Africa which contribute around 93 registered language among the 7,154 world languages [19]. Among the registered Ethiopian languages, Amharic and Afaan Oromo are two of the most extensively used languages in Ethiopia. The most widely spoken language in Ethiopia is Afaan Oromo, which has a 33.8% speaker followed by a 29.3% speaker for Amharic [19]. Afaan Oromo belongs to the Cushitic language family group of the Afroasiatic language family, while Amharic is a member of the Semitic language family group [20]. After Arabic and Hausa, Afaan Oromo is the third most frequently used Afro-Asiatic language in the world [21]. Similarly, Amharic is the second most spoken Semitic language family next to Arabic. Both languages are mainly spoken in Ethiopia and neighbouring countries. Along with Amharic, which has been the country's working language until 2020[1], Ethiopia has approved and introduced four other working languages those are: Tigrigna, Afar, Somali, and Afaan Oromo, which is the most widely used language in Ethiopia.

Amharic and Afaan Oromo languages have a substantially higher presence of digital resources in social media and news than other working languages of Ethiopia [13, 14, 20]. The presence of more digital resources on various platforms is due to the vast population in various parts of the country and the geographical distribution of the languages. These digital resources include news, blogs, posts, comments, and others that may contain hate speech that has a variety of effects on the community.

For a variety of reasons, automatic hate speech detection has recently gained attraction in the academic community. In recent years, the European

[1] https://unpo.org/article/21787

Union (EU) commission has contributed to hate speech by combining various measures such as a design program focused at combating hate speech on social media [22]. Recently, automatic hate speech detection become hot and popular in the research field for many reasons. EU commission in recent years, has contributed to hate speech by incorporating different initiatives such as a design program that aimed to fight against hate speech on social media [22]. Many scholars have contributed to the development of the model used to detect hate and offensive speech on social media for technological favored and resourced as well as underrepresented local languages such as English [22, 23] Hind-English [1, 6], Bangla [2], Italian [24], Indonesia [25], local language such as Amharic [14, 16] and Afaan Oromo [26–29].

To the best of the researcher's knowledge, no research on hate speech detection for underrepresented Ethiopian languages that is bilingual or Bilingual has ever been conducted. However, social media users sometimes express their opinions on social media intentionally or unintentionally using derogatory words and by blending different languages in a single sentence, which makes it difficult to detect hate and offensive speech. Hate speech detection efforts have been made in a variety of local languages, but they are monolingual and are insufficient to combat hate speech on Ethiopian social media that contain multiple languages. As a result of the rapid spread of hate and offensive speech on social media, there is a need to develop an automatic and state-of-the-art intervention to manage data on social media to protect hate and offensive speech [5]. In order to serve the community of social media users, it is necessary to find a solution to the problem of code mix hate and offensive speech detection for Ethiopian languages, particularly for Amharic and Afaan Oromo. Accordingly, to determine the optimal solution for hate speech detection a deep learning algorithm and feature extraction approaches were planned in this study, the researcher conducted numerous tests in addition to the Bilingual hate and offensive speech detection model.

2 Related Works

Social media platform provides a space for targeted attack to particular group of community based on protected characteristics: such as race, gender, and religion [29]. The majority of research is focused on detecting monolingual hate speech on social media for various technologically favored foreign and underrepresented local languages. On the contrary, there have been few attempts to detect hate speech in bilingual or multilingual languages such as Hindi-English [30, 31], English-Italian-German [32], English-Malay [33], Arabic-English-German-Indonesian-Italian-Portuguese-Spanish-French [5], English-French [34]. However, no bilingual or multilingual hate speech detection research has been done for languages with limited resources, particularly for Ethiopian languages like Afaan Oromo and Amharic.

The model used to identify hate speech from 9 distinct languages was developed using 16 publicly available hate speech datasets [6]. A total of 159,753

dataset were collected by the researcher using unbalanced dataset employing a LASER embedding with logistic regression perform best in low-resource situations while BERT model performs better in high-resource situations. They conducted experiments by classified data into 70%, 10%, and 20 % for training, validation, and testing respectively. However, the researcher experiments with an unbalanced dataset that may lead the model to be biased. Overall for low resourced language, LASER with logistic regression is more effective while for high resource BERT models are more effective.

Another study effort has been devoted to establish a model for combating hate speech in the Hindi-English language [30]. For the study, the researcher used 4,575 publicly available datasets, of which 1,661 are hate speech and 2,944 are non-hate speech. Using a transformer and translation based technique, the author outperformed accuracy by 73% through a classical machine learning suggesting a deep neural network with Transformer-based interpretation and feature extraction model (TIF-DNN).

In addition to technological favored language attempts, apachesparkbased model has reportedly been created to detect and classify Amharic Facebook comments and post into the free and hate classes [13]. Due to the limited number of data, the research used a 10 cross-validations employing a total of 6,120 datasets. The accuracy of word2vec feature extraction using Naive Bayes and Random forest approaches in their experiment was 79.83 % and 65.34%, respectively. The study also recommend the expansion of the dataset and the number of class for the hate speech detection. Similarly, another attempt has also been made to create a model for detecting hate speech in Amharic in posts and comments employing Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) [14]. In the attempt, the researcher collected a total of 30,000 datasets from Facebook and classified them as hate or non-hate speech using deep learning with a word2vec feature extraction techniques for Amharic language. The result obtained from the experiment reported as 97.9% accuracy.

In addition to this, a multilingual based offensive language identification methods for the Indian language [35]. They experimented were conducted with six native Indian languages families: Indo-Aryan and Dravidian, and the English language families. The author used nine recently released offensive datasets collected from Twitter and YouTube. Their experiment result illustrated that for all the languages, multilingual models outperform monolingual models in terms of identifying offensive language.

In addition, according to [26], they developed the model used to detect and classify Afaan Oromo hate speech on social media. They are conducted the experiment by using a different machine learning classifier (classical, ensemble and deep learning) with different feature extraction techniques such as BOW,TF-IDF,word2vec and Keras Embedding. The author developed the model that classifies the text in to octal and binary classes. As their experiment result illustrated that, BiLSTM with pre-trained word2vec feature extraction, achieved slightly superior accuracy of 0.84 and 0.88 for eight classes and binary

classes, respectively. But it is limited to monolingual hate speech detection only for Afaan Oromo language.

Furthermore, 13,600 hate speech datasets from Facebook and Twitter were collected to develop Afaan Oromo hate speech detection model [28]. The researcher employed classical and ensemble machine learning techniques with N-gram and TF-IDF text features extraction. The classical ML is composed of Liner SVM, SVM, Multinominal NB, Logistic regression, whereas the ensemble ML is constitute Random Forest classifier and Decision tree. The experiment's findings were underwhelming since hate speech detection performed worse than anticipated, with precision and recall scores of 64% and an F1 measure of 64% using linear SVM.

The research attempt made in hate speech detection is either monolingual or multilingual considering a technological favoured and resourced language. On the contrary, underrepresented and Bilingual hate speech detection research were not attempted either due to the unavailability of language resource or complexity of the language for the hate speech detection. Thus the purpose of this research is to design and develop a Bilingual hate speech detection for Amharic and Afaan Oromo languages.

3 Ethiopian Languages

Ethiopia is a multilingual and multinational country with 93 registered languages. of these, 92 are living consisting of 87 indigineous and 5 non-indigenous while 1 extinct language [19]. Ethiopian languages belongs to the Afro-Asiatic language phy consisting of Semitic, Cushitic, Omotic of the Nilo-Saharan language family [36]. Among those languages, this study inspects Amharic from Semitic and Afaan Oromo from Cushitic language families. This is because Amharic and Afaan Oromo are widely spoken languages in Ethiopia. Ethiopia is a unique uncolonized independent African country that share borders with Somalia, Djibouti, Kenya, Eritrea, Sudan, and South Sudan allow for the exchange of a variety of languages and dialects. Like the other under-resourced African languages, Ethiopian languages also suffers from the availability of digital language resources [37].

Amharic is the Ethiopian federal government's working language and the working language for other regional states while Afaan Oromo is the working language of the Oromia regional state [19]. In addition to Amharic, recently Ethiopia has introduced and proved the policy that introduced four widely spoken and regional language such Afaan Oromo, Tigrigna, Afar, and Somali, as the working language of the Ethiopian federal government [37][2]. Section 3.1 and 3.2 presents the detail of Amharic and Afaan Oromo language characteristics, respectively.

[2]https://unpo.org/article/21787

3.1 Amharic Language

Amharic is the second most spoken Semitic language next to Arabic serving as the official and working language of the Federal Democratic Republic of Ethiopian [38]. In addition to this, the language is used as inter regional communication with in the country. According to world language [19], the language Amharic has more than 56 million speakers of these more than 32 million as a first language (L1) and more than 25 million as a second language (L2) speaker. Though there are Amharic speakers in a number of other countries, including Italy, Canada, the United States, and Sweden, Ethiopia is home to the majority of Amharic speaker.

Amharic is written using Fidel (ፊደል/fīdeli) script which is derived from Ge'ez (ግእዝ) which is thought to be the historic center, classical and ecclesiastical language [38]. The language follow Subject + Object + Verb (SOV) sequences agreement. The script is written from left to right using Ge'ez with the nonexistence of capitalization consisting of a total 276 distinct symbols in four categories [39]: 231 core characters, 20 labiovelar symbols, 18 labialized consonants and 7 labiodental characters. In addition to this, 26 numerals and seven punctuation are there in Amharic language. Table 1 presents detail category of Amharic character, numerals and punctuation.

	Character List	Count	Orders	Total
Core	ሀ ለ ሐ መ ሠ ረ ሰ ሸ ቀ በ ተ ቸ ኀ ነ ኘ አ ከ ኸ ወ ዐ ዘ ዠ የ ደ ጀ ገ ጠ ጨ ጰ ጸ ፀ ፈ ፐ	33	7	231
Labiodental	ቨ	1	7	7
Labiovelar	ኰ ቈ ጐ ጕ	4	5	20
Labialized	ሏ ሟ ሯ ሷ ሿ ቧ ቷ ቿ ኗ ኟ ዟ ዧ ዷ ጇ ጧ ጯ ጷ ኧ	18	1	18
Numerals	፩ ፪ ፫ ፬ ፭ ፮ ፯ ፰ ፱ ፲ ፳ ፴ ፵ ፶ ፷ ፸ ፹ ፺ ፻ ፼	19	1	19
Punctuation	፡ ። ፣ ፤ ፥ ፦ ፧	7	1	7
Total				302

Table 1: Distribution of the Amharic script, numerals and punctuation adopted and modified from [38, 39].

As depicted in table 1, Amharic core and labiodental character possess 7 orders consisting a consonant having one basic (ʔ/አ) and six non-basic orders (ə/አ, u/ኡ, i/ኢ, a/ኣ, e/ኤ, o/ኦ) to indicate the vowel which comes after the consonant to represent CV syllables. The labiodental character

only appears in modern loaned words borrowed from foreign languages like ቪዛ/visa/ and ቫይረስ /virus/.

Similarly, the labiovelar category consists of 4 character (ቈ/k'^{w}/, ኀ/h^w/, ኰ/k^w/ and ጐ/g^w/) with 5 orders (ə'/ኍ/, i/ኌ/, a/ኋ/, e/኎/, ʔ/ኈ/) that generates a total of 20 distinct graphemes. Unlike the others, labialized has 18 characters with single order (^{w}ä); this includes, ሏ/l^{w}ä/, ሟ/m^{w}ä/, ሯ/r^{w}ä/ and ሷ/s^{w}ä/ that are used in the Amharic writing system. Aside from the Amharic characters, the Amharic language has 19 different digits that can be combined to give any number, in addition to the 7 punctuation marks used in the language. In general, there are 302 unique symbols that are required for the representation of Amharic orthography.

3.2 Afaan Oromo Language

Afaan Oromo is one of the Afro-Asiatic languages from the Cushitic language family. It is a Horn of Africa language that is indigenous to the Ethiopian state of Oromia and is mostly spoken by the Oromo people and other nearby ethnic groups in the horn of Africa [21]. Afaan Oromo is the third most widely spoken Afro-asiatic language in the world next to Arabic and Hausa [21]. Most of Ethiopia neighboring countries such as Kenya, Somalia, Tanzania, and Sudan speaks Afaan Oromo language as a first and second languages [21, 26]. Afaan Oromo is the mother tongue language for the Oromo ethnic group, the language is also the largest ethnic group in Ethiopia. According to language of the world [19], Afaan Oromo has more than 74 million speakers in Ethiopia. It is written in Latin Script called 'Qubee Afaan Oromo'. After 1991 Afaan Oromo is officially, adopted and started to use Latin script previously it is written in four different scripts such as Arabic alphabet, Shaykh Bakri Sapalo's orthography, Roman alphabet, and Ge'eez script [21]. Table 2 presents list of consonant, compound consonant, short and long vowels used in Afaan Oromo languages.

Categories	Character List	Count
Consonants	b, c, d, f, g, h, j, k, l, m,n, p, q, r, s, t, v, w, x, y, z	21
Compound consonants	ch, ts, zh, dh, sh, ny, ph	7
Vowels	a, e, i, o, u	5
Total		38

Table 2: Distribution of the Afaan Oromo script adopted and modified from [26, 40]

.

As depicted in Table 2, Afaan Oromo writing system has 33 distinct alphabets including five basic Vowels (5 short and 5 long), 7 double consonants, and single consonants (21) likewise English Language. In addition to this, there are

13 basic Punctuation marks almost common in the English language except for apostrophes. The apostrophe mark (‘) in the English language is used to show possession, However, in Afaan Oromo it is used to represent a glottal sound (’Hudhaa’) sound. unlike the Amharic language, Afaan Oromo has Capital and small letter like English Language [40]. In the Afaan Oromo writing system, consonants are also germinated to form compounded consonant as presented in Table 2. Germinated consonants and long vowels are represented by double letters in the Afaan Oromo writing System.

Afaan Oromo language borrowed single constants (p, v, and z) and double consonants (ts and zh) from different languages used to write borrowed (loan) words from different languages such as English, Amharic, and Roman language. However, there are no original Afaan Oromo words that start with these borrowed consonants [21]. Afaan Oromo language has also five long and five short vowels, which can occur at the beginning, middle, or end of a word the sort vowels is used to articulate short sound likewise the long vowels is used to articulate long sounds. In general, there are 33 unique symbols or characters that are essential for the representation of Afaan Oromo orthography.

4 Methodology

The systematic approach to solving a research problem through data collection utilizing various approaches, offering an interpretation of the data collected, and drawing inferences from the study data is known as research methodology.In this section, we discussed different methods and techniques used to achieve the research objective. Research methodology explain and discusses the data collecting, Data annotation, data pre-processing, feature extraction, normalization and analysis techniques carried out while doing the research.

4.1 Dataset Collection

NLP uses machine learning to understand, analyze, and extract meaning from language resources, making human language intelligible to machines employing a huge amount of language resources. For these, there are different publicly available datasets for hate speech detection but most of them are for foreign languages such as the English language [3, 41], Arabic [42] and multilingual with nine Englllish, European and Asian [6] languages. In addition to resourced language, underrepresented hate speech detection data are collected for different Ethiopian languages Amharic [14], Afaan Oromo [26, 28] and Tigrigna [43]. Hence, our focus is on two under-resourced Ethiopian languages specifically Amharic and Afaan Oromo. However, due to the lack of availability and accessibility of Bilingual hate speech detection data for Amharic and Afaan Oromo, the researcher opted to collect and prepare data from social media.

Accordingly, the researcher collected data from social media post and comment from January 2019 to July 2022 for both Amharic and Afaan Oromo language. The data is collected from broadcast media such as Fana Broadcast

Corporate (FBC), Oromia Broadcast Network (OBN), Oromia Media Network (OMN), Voice of America (VoA), and Ethiopian Press Agency (EPA). In addition to broadcast media, personal accounts, pages, activist pages, journalist pages, and other minor accounts or pages which have greater than 500 followers for both Afaan Oromo and Amharic [15].

The data is collected from accounts or pages with more than 500 followers for three primary reasons. The first is that it guarantees the dataset's legitimacy without bias. Second, the posts made by this account or page are seen by many social media users, which deserves special attention, and third, the Ethiopian constitutional orientation has a large number of followers. The dataset for Afaan Oromo is collected from Facebook, whereas the dataset for Amharic is collected from both Facebook and Twitter platforms. The data is gathered employing a Facepager and Data minor tools consisting of five fundamental attributes such as Object Id, Level, Query date, Label, and Message. Table 3 presents the distribution of Amharic and Afaan Oromo language datasets for hate speech detection.

Table 3: Distribution of Amharic and Afaan Oromo hate speech dataset.

Languages	Hate	Neutral	Offensive	Total
Afaan Oromo	5,000	5,000	5,000	15,000
Amharic	5,000	5,000	5,000	15,000
Total	10,000	10,000	10,000	30,000

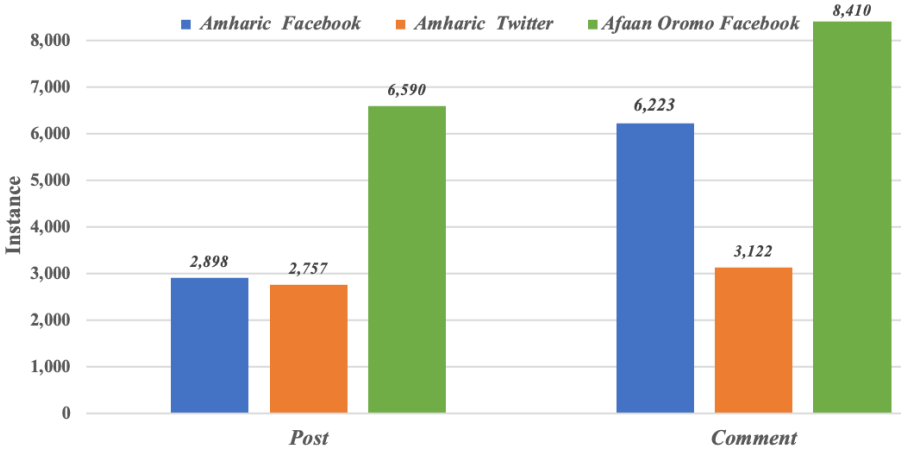
As depicted in Table 3, a total of 30,000 datasets were collected for Amharic and Afaan Oromo languages each contributing 15,000 instances. The extracted dataset consists of three labels: neutral, hate, and offensive speech from each language contributing 5,000 instances. The hate speech data for Afaan Oromo is collected from Facebook in addition to the previous data collected by Ababu [26] while for the language Amharic the from both Twitter and Facebook. Figure 1 presents the distribution of Amharic and Afaan Oromo hate speech detection data against the data source.

As depicted in Figure 1, a total of 12,245 post and 17,485 comment instances are collected from Facebook and Twitter. From these, 15,000 Afaan Oromo instances are collected from Facebook consisting instance of 6,590 posts and 8,410 comments. Similarly for Amharic, 9,121 and 5,879 instance from Facebook and Twitter, respectively.

4.2 Dataset Annotation

Annotation is the process of adding extra information to a collected dataset or it is the technique of labeling data in numerous formats so that machines can understand it [44]. The definition of hate and offensive speech slightly differs by country and even by a platform on social media. So, in this study,

Fig. 1: Detail dataset distribution per source and language.



we focused on the definition of hate speech provided by two social media sites, Facebook[3] and Twitter[4], as well as the Ethiopian government's declaration on hate speech and disinformation prevention [15].

To provide clear and concise data annotation guidelines, the researcher developed rules to classify posts or comments into three categories (Neutral, Hate, and Offensive) to solve the ambiguity problem during data annotation, as suggested by various researchers and social media platforms [13, 14, 29].

- The post or comment is labeled as Hate if the post or comment:
  - Encourages hatred or incites violence by making disparaging remarks about protected characteristics such gender, age, serious illness, race/ethnicity, disability, national origin, physical appearance, and others.
  - Imitates someone to superfluous objects, animals, or device to undermine psychological morality through discrimination against protected traits like religion, race, gender, age, and handicap, among others.
  - Promote hatred or attacks on others directly or indirectly by discriminating against protected characteristics.
  - Has stereotyped a specific target group.
  - Has blames or Reproachful individuals or groups based on their target group.
- The post or comment is labeled as Offensive if the post or comment:
  - Is insulted an individual or group make upset, annoyed, and angry.
  - Contains upsetting words but the particular subject of the sentence is unknown

[3]Source: https://transparency.fb.com/policies/community-standards/hate-speech/
[4]Source : https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy

  - Contains malign or offensive words but may or may not promote to take violence or attack against an individual or group.
  - Is contain common offensive words without specifying race gender, religion, and other protected characteristics.

- The post or comment is labeled as Neutral:
  - The post or comment is labeled as Neutral if it does not fulfill the criteria of hate and offensive speech.

Using the rule generated, the dataset is annotated independently by the domain expert in each language. Each data is annotated by three different domain experts to assign to one of three classes from native speakers of the respective language. Since data annotation is the subjective matter and sometimes it is difficult to decide whether the post or the comment is neutral, hate or offensive speech. To accept the correct annotation of each post or comment, the researcher uses the following equation 1.

$$Label = Mode(Ann_1, Ann_2, Ann_3) \tag{1}$$

Where: Ann is the Annotator expert of the respective language.

Inter-Annotator Agreement is a metric that measures how well a group of annotators can agree on an annotation option to analyze quantitative data [45]. The inter annotators reliability statistic can be used to assess the level of agreement between individual annotators. Accordingly, the annotators are purposely selected by the researcher based on their language skill and willingness to do the tasks. To accept the correct annotation of each post or comment, the researcher used Fleiss Kappa inter-annotator agreement to compute the inter-annotator agreement for more than two annotators. The Kappa value is computed using equation 2.

$$K = \frac{P - P_e}{1 - P_e} \tag{2}$$

where P is the mean proportion of agreement by chance, Pe is the mean proportion of agreement by analytical reasoning. Table 4 presents a sample Amharic and Afaan Oromo annotated dataset against the defined hate speech class beside the transliteration and English translation.

As depicted in Table 4, the annotation is done for both languages independently and Fleiss Kappa is computed individually deciding the class label using equation 1. The Fleiss Kappa's inter-annotator agreement of 0.75 for Afaan Oromo and 0.73 Amharic languages using the equation 2. The Fleiss Kappa's inter-annotator agreement result for both languages is in the substantial class of agreement.

4.3 Dataset Pre-processing

It is evident that social media data contains noise, and a number of pre-processing techniques were introduced to the text field to remove unnecessary noise from a dataset [30]. It is also essential to cleanse or filter the unstructured

Table 4: Sample Amharic and Afaan Oromo annotated dataset.

Language	Type	Post/Comment	Class
Amharic	Sentence	**'X' ብሔር ለኢትዮጵያ ግን በሽታ ነው**	Hate
	Transliteration	'X' biḥēri le'ītiyop'iya gini beshita newi	
	English	"X" nation is a disease for Ethiopia	
	Sentence	**አንተ ውሸታም ሌባ**	Offensive
	Transliteration	ānite wishetami lēba	
	English	you are a lying thief	
	Sentence	**አንተ ጥሩ ነህ**	Neutral
	Transliteration	ānite t'iru nehi	
	English	You are good	
Afaan Oromo	Sentence	Sabanni X Diina Saba Yti	Hate
	English	Nation X is antagonist of nation Y	
	Sentence	Namni gadheen akka kee hin jiru!	Offensive
	English	There is no wicked man like you!	
	Sentence	Oromoon sirna bulchiinsaa gadaa qaba	Neutral
	English	Oromo has Gadaa ruling system	

data before feeding it into the machine learning algorithm for natural language processing tasks. Several pre-processing operations were performed in this study, including data cleaning, character normalization, stop-word removal, and extending abbreviation.

The first and most important stage in data processing for machine learning is cleansing the data. Data collected from social media networks like Facebook and Twitter include noise like hashtags, emoticons, special characters, punctuation marks, URLs, and other unnecessary symbols that are important for human interpretation but not for machine learning. As a result, after gathering the data, we removed all superfluous noise from the Amharic and Afaan Oromo hate speech detection datasets.

In addition to cleaning, normalization is the most popular text preprocessing technique used to reduce the diversity of letters or words in a dataset [39]. In the Afaan Oromo language, character normalization is utilized in the same way that English is used to minimize differences in word forms to a common form. Character normalization in Amharic languages, as opposed to Afaan Oromo, refers to characters used interchangeably without bearing the meaning difference utilizing various orthographic representations. Table 5 presents the character normalization used in Amharic language.

Table 5: Amharic character normalization with their variant adopted and modified from [39].

Variants	Unnormalized	Normalized
4	ሀ, ሐ, ኸ, and ኀ	ሀ /h/
2	ስ and ሥ	ስ /s/
2	አ and ዕ	አ /ʔ/
2	ፀ and ጸ	ፀ /ts'/

As depicted in 5, Amharic characters with the same homophonic but multiple variants or orthographic representation have to be normalized into one equivalent character. As a result of normalizing the characters **ሀ, ሐ, ኸ, and ኀ** normalized to **ሀ/h/**, **ስ and ሥ** to **ስ/s/**, **አ and ዕ** to **አ/ʔ/**, and **ፀ and ጸ** to **ፀ/ts'/**. In addition to this, **"ሐ", "ኀ", "ኃ", "ሃ", "ሓ", "ኻ"** are replaced with **"ሀ", and "ኣ", "ዐ", "ዓ'** are replaced with **"አ"**. The equivalent character normalization is selected based on the most commonly used characters in the Amharic writing system.

Besides the data cleaning and normalization, another important data preprocessing task is expanding the acronym, abbreviation and contraction from the shorter version of the words. So in this study, common bilingual contractions, abbreviations, and acronyms are replaced with their expansion to reduce variation between them. Table 6 presents sample abbreviations, contractions, and acronyms along with the expanded forms presented for Amharic and Afaan Oromo languages.

Table 6: Sample abbreviations, contractions, and acronyms with expanded forms for Amharic and Afaan Oromo languages

		Word Type	Expanded Form
Amharic	Abbreviation	**ምዕ.** mi'i	**ምዕራፍ** mi'irafi
	Acronym	**ኢዜአ** īzē'ā	**ኢትዮጵያ ዜና አገልግሎት** ītiyop'iya zēna āgeligiloti
	Contraction	**ት/ሚኒስተር** ti/mīnīsiteri	**ትምህርት ሚኒስተር** timihiriti mīnīsiteri
Afaan Oromo	Abbreviation Acronym Contraction	hub. BBO M/Barnootaa	hubachiisa Biiroo Barnoota Oromiyaa Mana barnootaa

Furthermore, stop words are words that appear often and frequently in datasets. The removal of stopwords from datasets has no effect on the semantic content of posts or comments. For this study, the researcher used Afaan Oromo

stopwords in the hate speech detection [26] while for Amharic language, a manually constructed stopwords are used.

4.4 Feature Extraction Techniques

Text feature extraction uses real-valued values or vectors that can be processed by a computer to extract the essential data or hidden patterns from text for machine learning [46]. The two most popular text feature engineering methods, Keras embedding and Pre-trained models were used in this work. In Keras embedding, the embedding operations applied in the input layer with a fixed size vector and it offers an embedding layer that is frequently used for neural networks on multimodal data [47]. The Embedding layer starts with random weights and learns an embedding for each word in the training dataset. The first hidden layer of a neural network is called the embedding layer. It has consisting of three basic parameters those are vocab_size, embeeding_dimension, and input_length. The embedding layer uses one-hot encoding to convert categorical features into real numbers that can be processed by a machine learning algorithm.

Unlike the Keras embedding, the pre-trained model uses text embedding representation is carried out using word2vec and FastText model utilizing a free and open source Python framework called Gensim [48]. Word2vec is one of the popular text feature extraction technique as an extension of TF-IDF with a shallow neural network algorithms for learning word embedding and its representation vector. Word2vec captures the semantic similarity of words in the document based on the sequence presence of the word in a document. Due to the lack and unavailability of bilingual pre-trained word2vec for the Ethiopian languages (Amharic and Afaan Oromo), we created custom pre-trained word2vec using data derived from the Amharic and Afaan Oromo language dataset by 100 dimensions with the skip-gram model. The pre-trained bilingual word2vec model developed from Amharic and Afaan Oromo has a total 158,476 word type while the data used for hate speech detection has only 60,139 word type.

Similarly, as an extension of word2vec, the FastText model has recently been proven to be state-of-the-art for word embedding and text classification tasks that consider the internal words structure with character n-gram [48, 49]. FastText is a very user-friendly, tremendously rapid, and extremely effective learning tool for the word vector format when compared to other word embedding technique. Unlike word2vec, Facebook has released Amharic FastText word embedding with 300 dimensions, but not for Afaan Oromo. Similarly, we constructed a bilingual pre-trained FastText model using 100 dimensions with a skip-gram model that has 158,476 unique words likewise pre-trained word2vec model.

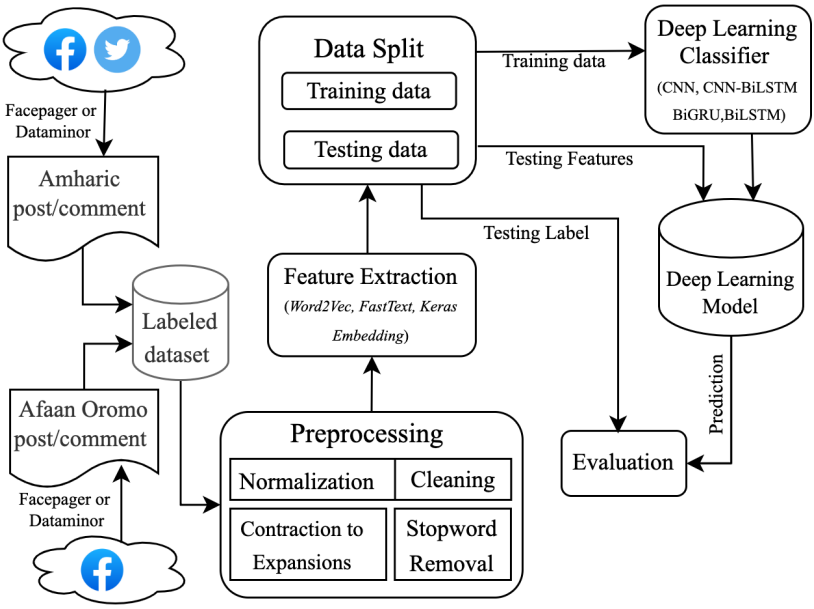
In this study, we used Five evaluation metrics most commonly used in machine learning model evaluation; such as Precision, Recall, F1-measure, Confusion matrix, and Accuracy. One of the simplest Classification metrics to

use is accuracy, which is calculated as the proportion of accurate predictions to all other predictions.

5 Proposed Architecture

The proposed system architecture includes several processes like pre-processing, feature extraction, and creating a model for predicting a bilingual Amharic Afaan Oromo hate speech detection. The architecture of Amharic Afaan Oromo hate speech detection utilizing several deep learning techniques is presented in Figure 2. As depicted in Figure 2, the author first gath-

Fig. 2: Proposed Bilingual hate speech detection architecture



ered datasets from Facebook (for Afaan Oromo and Amharic) and Twitter (for Amharic), combined them into a single file, and then the data is annotated using the respective domain expert. During the pre-processing stage, the collected and annotated data is cleaned, normalized, expanded contraction, and stopwords are removed for both languages. Following pre-processing, the data set is divided into train tests, with 80% of the dataset utilized for training and 20% for testing the hate speech model.Each tokenized word is transformed into its corresponding embedding vector using the feature extraction methods. In this work, three feature extraction techniques were followed: Word2vec, FastText, and Keras Embedding. After feature extraction, the extracted feature is fitted with different parameters into four distinct

different deep learning algorithms.To solve the problems of manual hyperparameter search, We employed Keras tuner to select a scalable hyperparameter optimization framework.

In selecting model, four different deep learning algorithms for hate speech detection such as CNN, BiLSTM, CNN-BiLSTM, and BiGRU. Though CNN is frequently used to evaluate visual data, it has also been successfully applied in the field of text classification [5, 50, 51]. LSTM is initially introduced to address the standard RNN's gradient vanishing problem [52]. Bidirectional LSTM is the special type of LSTM in which it is trained in forwarding and backward direction [53]. BiLSTMs effectively improve the quantity of information available to the network, improving the context provided to the algorithm. CNN-BiLSTM is the combination of both CNN and BiLSTM algorithms. A bidirectional GRU is a sequence processing model that consists of two GRUs, one takes the input in the forward direction, and the other take it in the backward direction. GRU is a simple variant of LSTM and which is contains two gates these are forget gate and the update gate.

The model is trained utilizing features extracted from Bilingual hate speech detection data, the optimal hyperparameters chosen, and deep learning approaches. Finally, the model is evaluated to assess the correctness of models on test data. The test data consists of data points that have not been seen by the model before.

6 Experimental Setting and Setup

A total of 30,000 posts and comments were used in the experiment to detect bilingual hate speech in Amharic and Afaan Oromo, with 15,000 instance of posts or comments from each languages yielding 5,000 instances of hate, free, and offensive speech. The experiment uses 80% of the data for training while the remaining 20% for testing a bilingual hate speech detection.

In addition to this, a total of 12 different experiments are conducted employing four deep learning algorithms: CNN, BiLSTM, CNN-BiLSTM and BiGRU with three feature extraction techniques those are Keras Embedding layer, word2vec, and Fasttext. For the deep learning techniques, a filter size 3x3, a batch size of 64, 100 embedding dimension, 0.5 dropout rate , adam optimizer, sparse categorical cross entropy for loss, and Rectified Linear Unit (Relu) for hidden layer activation function, softmax for output layer activation function, 0.01 learning rate, and 20 epochs is used as optimal hyper-parameters.

Likewise, deep learning model to develop our pre-trained word2vec and fast text model the researcher used 100 embedding dimension, 10 window, minimum count of one words1 min, 4 threads, 1 sg. In creating a pre-trained model 120 iteration for word2vec while 20 iteration for fast text were used. In addition to deep learning and feature extraction, python modules like Flair, Natural language processing toolkit (NLTK), Scientific computing (SciPy), and Multidimensional processing (NumPy) have been used at different degrees of deep

learning to carry out the tests. For data pre-processing and model development, the Jupyter notebook is used as an integrated development environment with Python.

The bilingual Amharic and Afaan Oromo hate speech detection experiment is carried out using a personal computer equipped with an Intel(R) Core(TM) i5-2520M CPU @ 2.50GHz 2.50 GH processor, 8.00 GB RAM, an x64-based processor, and a 1TB hard disk storage capacity. In addition to our personal PC, we trained our custom word2vec and Fast Text models using Google Colaboratory.

7 Experiment Result and Discussion

Given the experimental setting and setup in Section 6, this section provides the experimental result and discussion for each experiment against the three text feature extraction for deep learning. Table 7 presents the experimental result of four deep learning algorithms with three different feature extraction techniques.

Table 7: Macro average score of selected deep learning algorithm with three feature extraction techniques

Feature	Algorithms	Precision	Recall	F1-Score	Accuracy
Keras embedding	CNN	0.7340	0.7282	0.7267	0.7295
	BiLSTM	0.7409	0.7413	0.7405	**0.7413**
	CNN-BiLSTM	0.7394	0.7353	0.7342	0.7363
	BiGRU	0.7412	0.7395	0.7397	0.7402
Word2Vec	CNN	0.7321	0.7298	0.7290	0.7308
	BiLSTM	0.7379	0.7380	0.7375	**0.7378**
	CNN-BiLSTM	0.7260	0.7262	0.7260	0.7262
	BiGRU	0.7348	0.7340	0.7326	0.7338
FastText	CNN	0.7554	0.7451	0.7442	0.7467
	BiLSTM	0.7836	0.7811	0.7803	**0.7805**
	CNN-BiLSTM	0.7370	0.7350	0.7336	0.7353
	BiGRU	0.7488	0.7489	0.7485	0.7493

As depicted in Table 7, the average accuracy achieved with Keras embedding feature extraction and Pre-trained word2vec feature extraction is statistically competitive but slightly there is fluctuation with per-trained Fast text feature extraction. With Keras embedding feature extraction, the highest accuracy which is 0.7413 is achieved by BiLSTM algorithm. Similarly, 0.7295, 0.7363, and 0.7402 of accuracy are achieved by CNN, CNN-BiLSTM, and BiGRU algorithms respectively.

Likewise the Keras embedding feature extraction, the second experiment is conducted under the same setting with a change of feature extraction to a custom pre-trained word2vec and the highest accuracy 0.7378 is achieved

again by the BiLSTM algorithm whereas 0.7308, 0.7262, and 0.7338 of accuracy is achieved by CNN, CNN-BiLSTM, and BiGRU algorithms respectively. In a similar way, the third experiment is conducted using Fast Text as feature extraction and the highest accuracy 0.7805 is achieved by BiLSTM algorithm likewise the first and second experiments. Similarly, 0.7467, 0.7353, and 0.7493 of accuracy is achieved by CNN, CNN-BiLSTM and BiGRU algorithms respectively. Figure 3 presents the comparison of the four deep learning techniques against the three feature extraction technique for bilingual Amharic Afaan Oromo hate speech detection.

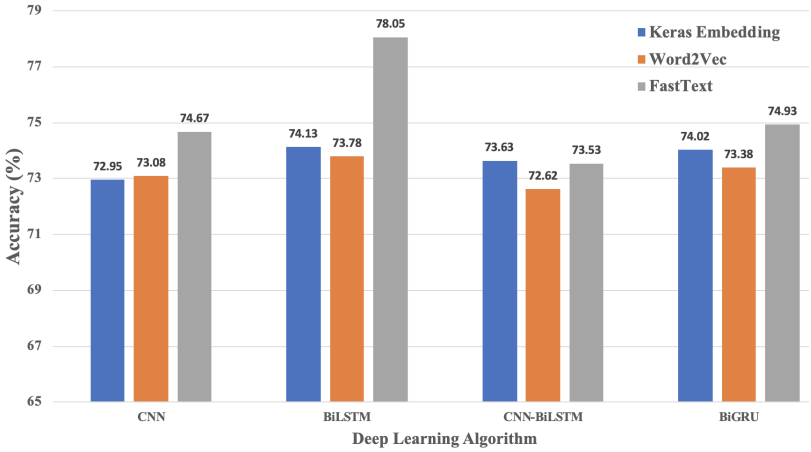


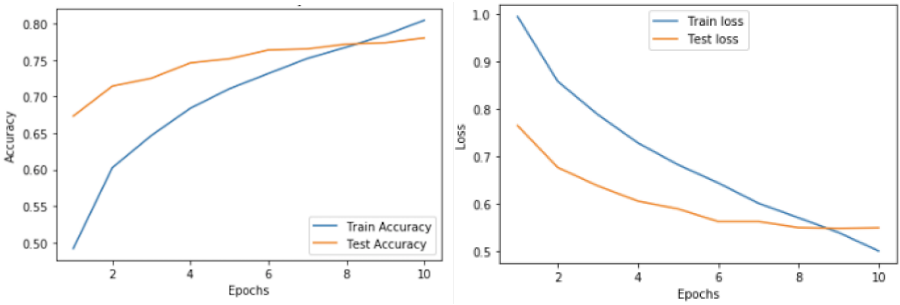
Fig. 3: Comparison of Bilingual Amharic Afaan Oromo hate speech detection.

As depicted in Figure 3, among the three feature extraction technique FastText feature extraction achieved better performance over word2vec and Keras Embedding. This is because the FastText can generate a words vector that is not available in training datasets or can overcome out of vocabulary problems as depicted in Figure 3. FastText feature extraction slightly improved bilingual hate speech detection classification performance over keras embedding and word2vec model by 1.62 and 2.08 average percent of accuracy. However, The average accuracy achieved by pre-trained word2vec is lower than keras word embedding this is because it is based on the quality of our custom pre-trained word2vec model. Since we are trained our word2vec and FasText model with the same dataset the accuracy achieved with FastText is outperformed over word2vec this because the fast text model can generate the vector of words that is out of vocabulary but word2vec only uses complete words from the training corpus to learn vectors representation.

Overall, we experimented with four different deep learning algorithms with three different feature extraction techniques. Our experiment result is in line with some other studies showing that the BiLSTM algorithm is the best algorithm for Afaan Oromo hate speech detection [26, 29] and Amharic [13, 14]

languages. So, BiLSTM algorithm with fast text feature extraction techniques is the best algorithm and feature extraction techniques for bilingual (Amharic and Afaan Oromo) hate speech detection on social media. As a result, the BiLSTM algorithm with fast text feature extraction techniques is the best algorithm and feature extraction techniques for detecting hate speech on social media in Bilingual (Amharic and Afaan Oromo).

In addition to this, classification model successfully predicts by dividing the total number of predictions as model accuracy beside the precision, recall and F1-score. Furthermore, assessing a model's performance certainly not the only one. The error that occurs on the training and testing data are also important criteria to evaluate the performance in terms of training and testing loss, respectively. The value that a deep neural network seeks to minimize the loss by maximizing the accuracy in a given dataset. Figure 4 presents Amharic Afaan Oromo hate speech detection on social media using BiLSTM model accuracy and loss with FastText feature extraction under 4a and 4b, respectively.



(a) Model Accuracy (b) Model Loss

Fig. 4: Accuracy and loss of bilingual Amharic and Afaan Oromo hate speech detection using BiLSTM.

As depicted in Figure 4, the accuracy of training and testing shows a good fit with slightly over-fitted because both graphs are slightly converged to each other as presented in figure 4a. In the same way, training and the testing accuracy is increasing at the decreasing rate while the loss is decreasing at the decreasing rate for the testing and increasing rate for the training loss as depicted in figure 4b.

Taking the relative accuracy and loss of the Amharic Afaan Oromo hate speech detection, a flask framework used to create user friendly interface to detect hate speech by selecting BiLSTM model because of the performance. In this study, BiLSTM model integrated with the flask framework to detect and classify bilingual hate speech from social media. The framework is implemented on a local host with port number 5001. Figure 5 presents a sample

prototype developed for the Amharic Afaan Oromo hate speech detection model.

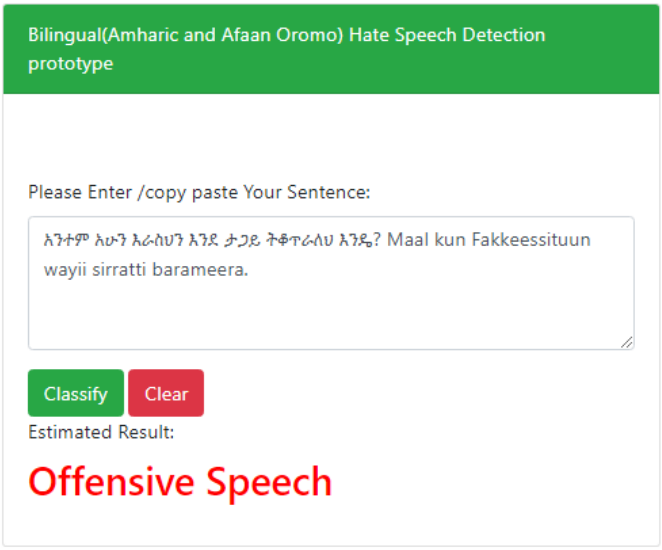


Fig. 5: Proposed Prototype

As we observed in Figure 5 the model correctly detects and classifies the comment **"አንተም አሁን እራስህን እንደ ታጋይ ትቆጥራለህ እንዴ?**/ānitemi āhuni irasihini inide tagayi tik'ot'iralehi inidē?/ **Maal kun Fakkeessituun wayii sirratti barameera."** to Offensive speech. The message of above comment **"አንተም አሁን እራስህን እንደ ታጋይ ትቆጥራለህ እንዴ?Maal kun Fakkeessituun wayii sirratti barameera."** is despise the notion of particular person. So it make upset particular person as the reward of the post. So the model is correctly detect and classify the comment as it is expected.

8 Conclusion and Recommendation

In this paper work, the researcher provide the first attempt to detect and classify bilingual (Amharic and Afaan Oromo) hate speech. In the attempt, the researcher collected a total of 30,000 annotated Amharic and Afaan Oromo bilingual datasets for hate speech detection and classification. Employing the collected data, four different deep learning (CNN, BiLSTM, CNN-BiLSTM, BiGRU) algorithms with three different feature extraction techniques (Keras Layer embedding, word2vec, fast text) used to look up and experiment and develop a deep learning model. The BiLSTM with Fatsttext feature extraction gave the best performance with 78.05% accuracy.

Furthermore, as part of future research, the experiment described in this paper might be reproduced and applied to other Ethiopian languages with

limited resources. It is crucial given Ethiopia's diversity of ethnic groups and languages to combat Bilingual hate speech detection. In addition to this, the expansion to other local language to build hate speech detection employing linguistic and social information beside the multi-modal representation through different techniques.

9 Abbreviations or Acronyms

BiLSTM: Bidirectional Long Short Term Memory
CNN: Convolutional Neural Network
EPA: Ethiopian Press Agency
fastText: Fast Text
LSTM: Long Short Term Memory
ML: Machine Learning
NLP: Natural Language Processing
OBN: Oromia Broadcast Network
OMN: Oromia Media Network
VoA: Voice of America
Word2vec: Word to Vector

Declarations

- **Funding** :Not Applicable
- **Conflict of interest/Competing interests** :Not Applicable
- **Ethics approval**
  :Not Applicable
- **Consent to participate**
  :Not Applicable
- **Consent for publication**
  :Not Applicable
- **Availability of data and materials**
  :Not Applicable
- **Code availability**
  :Not Applicable
- **Acknowledgments**
  :Not Applicable
- **Authors' contributions**
  TMA: Prepared the manuscript including analysis, data curation, visualization, conceptualization, experimenting, methodology and Prototype development. In addition to this, the collection of the data for Afaan Oromo hate speech detection. **MMW**: Prepared the manuscript and performs the tasks including conceptualization, validation, experimenting, writing, review, and editing of the final work. **EB**: Contributed on the collection and development of the Amharic hate speech data Design and Develop a bilingual hate speech detection model for two Ethiopian under-resourced language. All the three authors read and approved the final manuscript for submission.

References

[1] Vashistha, N., Zubiaga, A.: Online multilingual hate speech detection: experimenting with hindi and english social media. Information **12**(1), 5 (2020)

[2] Das, A.K., Al Asif, A., Paul, A., Hossain, M.N.: Bangla hate speech detection on social media using attention-based recurrent neural network. Journal of Intelligent Systems **30**(1), 578–591 (2021)

[3] Watanabe, H., Bouazizi, M., Ohtsuki, T.: Hate speech on twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection. IEEE access **6**, 13825–13835 (2018)

[4] Chaffey, D.: Global social media statistics research summary 2022. Accessed: 2022-09-24

[5] Alkomah, F., Ma, X.: A literature review of textual hate speech detection methods and datasets. Information **13**(6), 273 (2022)

[6] Aluru, S.S., Mathew, B., Saha, P., Mukherjee, A.: Deep learning models for multilingual hate speech detection. arXiv preprint arXiv:2004.06465 (2020)

[7] Guiora, A., Park, E.A.: Hate speech on social media. Philosophia **45**(3), 957–971 (2017)

[8] Chekol, M.A., Moges, M.A., Nigatu, B.A.: Social media hate speech in the walk of ethiopian political reform: analysis of hate speech prevalence, severity, and natures. Information, Communication & Society, 1–20 (2021)

[9] EIP: Fake News Misinformation and Hate Speech in Ethiopia : A Vulnerability Assessment. European Institute of Peace (EIP), Rue des Deux Eglises 25 1000 Brussels, Belgium. Accessed: 2022-09-24. https://www.eip.org/wp-content/uploads/2021/04/Fake-News-Misinformation-and-Hate-Speech-in-Ethiopia.pdf

[10] Ahammed, S., Rahman, M., Niloy, M.H., Chowdhury, S.M.H.: Implementation of machine learning to detect hate speech in bangla language. In: 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART), pp. 317–320 (2019). IEEE

[11] MacAvaney, S., Yao, H.-R., Yang, E., Russell, K., Goharian, N., Frieder, O.: Hate speech detection: Challenges and solutions. PloS one **14**(8), 0221152 (2019)

[12] Abro, S., Shaikh, S., Khand, Z.H., Zafar, A., Khan, S., Mujtaba, G.: Automatic hate speech detection using machine learning: A comparative study. International Journal of Advanced Computer Science and Applications **11**(8) (2020)

[13] Mossie, Z., Wang, J.-H.: Social network hate speech detection for Amharic language. Computer Science & Information Technology, 41–55 (2018)

[14] Tesfaye, S.G., Kakeba, K.: Automated Amharic hate speech posts and comments detection model using recurrent neural network (2020)

[15] FDRE: Hate Speech and Disinformation Prevention and Suppression Proclamation. (2020)

[16] Yonas, K.: Hate speech detection for Amharic language on social media using machine learning techniques. PhD thesis, ASTU (2019)

[17] Teh, P.L., Cheng, C.-B.: Profanity and hate speech detection. International Journal of Information and Management Sciences **31**(3), 227–246 (2020)

[18] Ian McFarlane: State of World Population 2022. United Nations Population Fund (UNFPA), New York, United States. Accessed: 2022-09-24 (2022). https://www.unfpa.org/sites/default/files/pub-pdf/EN_SWP22%20report_0.pdf

[19] Eberhard, D.M., Simons, G.F., Fennig, C.D.: Ethnologue: Languages of the world, Dallas, Texas (2022). https://www.ethnologue.com/country/ET/languages

[20] Abate, S.T., Tachbelie, M.Y., Melese, M., Abera, H., Abebe, T., Mulugeta, W., Assabie, Y., Meshesha, M., Afnafu, S., Seyoum, B.E.: Large vocabulary read speech corpora for four Ethiopian languages: Amharic, Tigrigna, Oromo and Wolaytta. In: Proceedings of The 12th Language Resources and Evaluation Conference, pp. 4167–4171 (2020)

[21] Degeneh Bijiga, T.: The development of Oromo writing system. PhD thesis, University of Kent, (2015)

[22] Alrehili, A.: Automatic hate speech detection on social media: A brief survey. In: 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), pp. 1–6 (2019). IEEE

[23] Badjatiya, P., Gupta, S., Gupta, M., Varma, V.: Deep learning for hate speech detection in tweets. In: Proceedings of the 26th International Conference on World Wide Web Companion, pp. 759–760 (2017)

[24] Santucci, V., Spina, S., Milani, A., Biondi, G., Di Bari, G.: Detecting hate speech for italian language in social media. In: EVALITA 2018, Co-located with the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), vol. 2263 (2018)

[25] Alfina, I., Mulia, R., Fanany, M.I., Ekanata, Y.: Hate speech detection in the indonesian language: A dataset and preliminary study. In: 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS), pp. 233–238 (2017). IEEE

[26] Ababu, T.M., Woldeyohannis, M.M.: Afaan Oromo hate speech detection and classification on social media. In: Proceedings of the Language Resources and Evaluation Conference, pp. 6612–6619. European Language Resources Association, Marseille, France (2022). https://aclanthology.org/2022.lrec-1.712

[27] Kanessa, L.G., Tulu, S.G.: Automatic hate and offensive speech detection framework from social media: the case of Afaan Oromoo language. In: 2021 International Conference on Information and Communication Technology for Development for Africa (ICT4DA), pp. 42–47 (2021). IEEE

[28] Defersha, N., Tune, K.: Detection of hate speech text in afan Oromo social media using machine learning approach. Indian J Sci Technol **14**(31), 2567–78 (2021)

[29] Ganfure, G.O.: Comparative analysis of deep learning based Afaan Oromo hate speech detection. Journal of Big Data **9**(1), 1–13 (2022)

[30] Biradar, S., Saumya, S., *et al.*: Fighting hate speech from bilingual hinglish speaker's perspective, a transformer-and translation-based approach. Social Network Analysis and Mining **12**(1), 1–10 (2022)

[31] Bohra, A., Vijay, D., Singh, V., Akhtar, S.S., Shrivastava, M.: A dataset of hindi-english code-mixed social media text for hate speech detection. In: Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media, pp. 36–41 (2018)

[32] Corazza, M., Menini, S., Cabrio, E., Tonelli, S., Villata, S.: A multilingual evaluation for online hate speech detection. ACM Transactions on Internet Technology (TOIT) **20**(2), 1–22 (2020)

[33] Moy, T.X., Rahem, M., Logeswaran, R.: Multilingual hate speech detection

[34] Chiril, P., Benamara, F., Moriceau, V., Coulomb-Gully, M., Kumar,

A.: Multilingual and multitarget hate speech detection in tweets. In: Conférence sur Le Traitement Automatique des Langues Naturelles (TALN-PFIA 2019), pp. 351–360 (2019). ATALA

[35] Ranasinghe, T., Zampieri, M.: An evaluation of multilingual offensive language identification methods for the languages of india. Information **12**(8), 306 (2021)

[36] Bender, M.L.: The languages of ethiopia: a new lexicostatistic classification and some problems of diffusion. Anthropological Linguistics, 165–288 (1971)

[37] Demilie, W.B., Salau, A.O.: Detection of fake news and hate speech for ethiopian languages: a systematic review of the approaches. Journal of big Data **9**(1), 1–17 (2022)

[38] Yimam, B.: Amharic grammar. Addis Ababa: EMPDA (1986)

[39] Woldeyohannis, M.M., Meshesha, M.: Usable Amharic text corpus for natural language processing applications. Applied Corpus Linguistics **2**(3), 100033 (2022). https://doi.org/10.1016/j.acorp.2022.100033

[40] Bedane, I.: The origin of afan Oromo: Mother language. Glob. J. Hum. Soc. Sci. G Linguist. Educ **15**(12) (2015)

[41] Davidson, T., Warmsley, D., Macy, M., Weber, I.: Automated hate speech detection and the problem of offensive language. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 11, pp. 512–515 (2017)

[42] Albadi, N., Kurdi, M., Mishra, S.: Are they our brothers? analysis and detection of religious hate speech in the arabic twittersphere. In: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 69–76 (2018). IEEE

[43] Bahre, W.: Hate Speech Detection from Facebook Social Media Posts and Comments in Tigrigna language. St. Mary's University (2022)

[44] Kocoń, J., Figas, A., Gruza, M., Puchalska, D., Kajdanowicz, T., Kazienko, P.: Offensive, aggressive, and hate speech analysis: From data-centric to human-centered approach. Information Processing & Management **58**(5), 102643 (2021). https://doi.org/10.1016/j.ipm.2021.102643

[45] Artstein, R.: Inter-annotator agreement. In: Handbook of Linguistic Annotation, pp. 297–313. Springer, ??? (2017)

[46] Jurafsky, D., Martin, J.H.: Speech and language processing (3rd draft ed.). Stanford University (2019)

[47] Gulli, A., Pal, S.: Deep Learning with Keras. Packt Publishing Ltd, Birminghum, UK (2017)

[48] Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)

[49] Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. Transactions of the association for computational linguistics **5**, 135–146 (2017)

[50] Zhang, Y., Wallace, B.: A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. arXiv preprint arXiv:1510.03820 (2015)

[51] Gitari, N.D., Zuping, Z., Damien, H., Long, J.: A lexicon-based approach for hate speech detection. International Journal of Multimedia and Ubiquitous Engineering **10**(4), 215–230 (2015)

[52] Oljira, M., Sori, K., Kaliyaperumal, K., Natea Sima, B., *et al.*: Sentiment analysis for Afaan Oromoo using combined convolutional neural network and bidirectional long short-term memory. International Journal of Advanced Research in Engineering and Technology **11**(11), 2020 (2021)

[53] Yu, Y., Si, X., Hu, C., Zhang, J.: A review of recurrent neural networks: LSTM cells and network architectures. Neural computation **31**(7), 1235–1270 (2019)