Ganfure *Journal of Big Data* (2022) 9:76
https://doi.org/10.1186/s40537-022-00628-w

Journal of Big Data

RESEARCH Open Access

Comparative analysis of deep learning based Afaan Oromo hate speech detection

Gaddisa Olani Ganfure*

*Correspondence:
gaddisaolex@gmail.com

Computer Science, Dire Dawa University Institute of Technology, Sabean, Dire Dawa 1362, Ethiopia

Abstract

Social media platforms like Facebook, YouTube, and Twitter are banking on developing machine learning models to help stop the spread of hateful speech on their platforms. The idea is that machine learning models that utilize natural language processing will detect hate speech faster and better than people can. Despite numerous progress has been made for resource reach language, only a few attempts have been made for Ethiopian Languages such as Afaan Oromo. This paper examines the viability of deep learning models for Afaan Oromo hate speech recognition. Toward this, the biggest dataset of hate speech was collected and annotated by the language experts. Variations of profound deep learning models such as CNN, LSTMs, BiLSTMs, LSTM, GRU, and CNN-LSTM are examined to evaluate their viability in identifying Afaan Oromo Hate speeches. The result uncovers that the model dependent on CNN and Bi-LSTM outperforms all the other investigated models with an average F1-score of 87%.

Keywords: Hate Speech Detection, Deep Learning, Artificial Intelligence, Afaan Oromo, Ethiopian Languages

Introduction

The last few years have witnessed a dramatic increase in the usage of the internet. Notably, the introduction of social media networks such as Facebook, Twitter, Telegram, Skype, and others easies the way people communicate, and exchange information with friends and other family members. As of January 2020, in Ethiopia alone, there are more than 21 million internet users of which 6.2 million actively use different social media platforms. Social media technology allows society to easily communicate and express their opinion. However, in developing countries such as Ethiopia, the usage of social media is changing from posting or tweeting social, financial, and political debates to disseminating hate speech [1]. The flourishing of hate speech and disinformation online can disrupt democratic debate and practices, facilitate gross human rights violations, and further marginalize minority groups. Ethiopia, even with one of the lowest internet connectivity penetration rates in the continent, is not immune to this phenomenon.

Given the increase in online hate speech, it has become increasingly important to examine the ways in which harmful speech influences physical world attitudes and behaviors, such as hate crimes. However, there has been difficulty in empirically

identifying the connections between online racist speech and physical world consequences and thus the research in this area remains limited.

Social media giant like Facebook has been criticized for allowing their platform to be used as a vehicle for hate speech, with minimum moderation given to non-English pages, which has caused recurrent problems in sharply polarized countries like Ethiopia.

In recent years, the increasing propagation of hate speech on social media and the urgent need for effective countermeasures have drawn significant investment from governments, companies, and researchers. Social media platforms rely on a combination of artificial intelligence, user reporting, and staff known as content moderators to enforce their rules regarding appropriate content publication. However, the problem arises when the platform's artificial intelligence is poorly adapted to the under-resourced language and companies have invested little in staff fluent in them.

Aiming to classify textual content into non-hate or hate speech, a large number of methods have been developed for resource reach languages such as English, Chinese, French, and others [2–8] using statistical and machine learning approaches. Some research conducted on those languages [9] depicts that the detection model built using a deep learning model achieves a state-of-the-art accuracy. However, only a few attempts have been made for Ethiopian languages [10–13].

This paper introduces an automatic classification of social media post that contains Afaan Oromo text into hate speech or not hate by leveraging variants of deep learning approaches. First, a corpus of comments and posts retrieved from Facebook and Twitter are built, and then features are extracted using word n-grams and word embedding techniques such as Word2Vec [14]. Then, variants of deep learning models such as convolutional neural networks (CNN), long short-term memory networks (LSTM [15]), bidirectional long short-term memory networks (Bi-LSTM [16]), GRU, and CNN-LSTMs are used for hate speech detection. The experiment result reveals that the model built using CNN and BiLSTM achieves the highest weighted average F1-score of 87%. Likewise, this study also assesses the impact of incorporating augmented data into the training dataset. The result shows that the textual data augmentation enhances the model F1-score up to 3%. data. In the light of the gap in this research area, our contributions described in this paper are the following:

- Develop the largest labeled Afaan Oromo hate speech classification dataset of his kind.
- This work investigates the accuracy of five state-of-the-art deep learning models at detecting hate speech for resource-scarce languages, i.e., Afaan Oromo. The output of the experiment provides insight into their detection accuracy, and capability in using pre-trained models, and text data augmentation, and offers important guidelines for their deployment in real-world applications.
- Assess the impact of adding augmented textual data on the Afaan Oromo Hate Speech classification performance
- Assess the impact of using pre-trained Word2Vec model with the one directly trained with the hate speech classification model
- Build a pre-trained word embedding model, which is useful for other works in this area.

The rest of the paper is organized as follows. First it explains related works in "Background and related work" section. "Afaan Oromo Corpus creation and annotation" section presents the details of the data set. "Experiment and discussion" section reports on approaches utilized in this study to classify the hate speech content. Section 5 presents the results accompanied by detailed performance analysis. "Conclusion" section, concludes the paper.

Background and related work

Word embedding

In Natural Language Processing (NLP) research, the aim is to make a model that is capable to recognize human languages. However, there is a challenge that leaps out: we, people, communicate with words and sentences; meanwhile, deep learning models expect the number as the input. To make the textual data ready for processing, it must be transformed into a sequence of vectors. One of the techniques commonly employed in NLP to transform textual data into a real-valued vector is called Word embedding. Word embedding techniques learn a real-valued vector representation for a predefined fixed-sized vocabulary from a corpus of text [17]. The learning process of vector representation is done separately or joined with the neural network model on some tasks using document statistics. Pre-trained word embeddings are publicly available for researchers. For instance, Facebook provided the fastText model, Google provided several BERT models for different languages. Word embedding models have been widely used in NLP research such as hate speech detection, sentiment analysis, machine translations, storage optimizations, and security [2, 18–20]. In literature, numerous word embeddings techniques are introduced, for instance, word2vec [17], Glove [21], and FastText [22]. The main idea behind these models is that the word co-occurrence probabilities have the potential for encoding some type of semantic importance between the words (refer to [23] for a detailed description of all those techniques).

Related work

The introduction of social media such as Facebook and other community forums has simplified the manner people communicate and express their thoughts, however, it also brought the issue of hate speeches. Online platforms provide a space for discourses that are harmful to certain groups of people. Recently, hate speech can be considered a serious problem by different country authorities. To minimize the impact of hate speech, various studies have been conducted at detecting hate speech. The majority of the research works have been done in resource reached languages like English [2, 3], Spanish [24], and Chinese [7, 8]. The most recent studies on hate speech detection have proposed the use of deep learning techniques for classification [4–6].

Aimed at distinguishing the instance of hate speech, Davidson et al. [7] utilized bigram, unigram, and trigram features with TF-IDF with a part-of-speech tagger as a feature for the machine learning model. The Twitter dataset that constitutes 33,458 English tweets was used for investigation. Their experiment using Logistic regression with a Linear Support Vector Machine (SVM) yielded an overall precision of 0.91, recall of 0.90, and F1 score of 0.90. Considering the simplicity of their model their finding is interesting. Authors in [25] also investigate the variety of hate categories to distinguish the kind of hate in Italian

text. By leveraging morpho-syntactical features, sentiment polarity, and word embedding lexicons, they design and implement a classification model using SVM and Long Short Term Memory (LSTM). To train the LSTM architecture, they represent each word with a 262-dimensional vector. Their finding reveals that the detection model built using an LSTM has an F1-score of 72%.

A study introduced in [26] developed an Apache Spark-based model to classify Afaan Oromo language Facebook posts and comments into hate and not hate. They employed Random forest and Naïve Bayes as learning algorithms and Word2Vec and TF-IDF for feature extraction using 6120 (4882 to train the model and 1238 for testing). In their experiment, Naïve Bayes and Random Forest outperform with an accuracy of 79.83% and 65.34% with the word2vec feature vector modeling approach respectively. However, they recommended expanding the classification category with different aspects of hate and increasing the corpus size including other sources.

A study in [27] designed a system for hate speech text classification on Twitter using, the CNN model with 6,655 total datasets. The classifier predicts each Tweet to one of four predefined categories as racism, sexism, both (racism and sexism), and neither. They have created two CNN models based on different input vector sets. using, word2vec, and randomly generated vector baseline. The system based on word2vec word vectors performed best overall, with an F1-score of 78.3%. However, their result is not sufficient and the system wrongly identified some non-hate speech tweets as hate speech. In particular, the system was not able to identify properly the category of both racism and sexism.

Likewise, research introduced in [28] used a Convolutional Neural Network classifier with word embedding as a feature using the Hate Speech Identification dataset distributed via Crowd Flower. They used 24,783 English tweets that have been classified into three classes hate, offensive language, and neither. And a Publicly available Word2Vec word embedding with 300 dimensions pre-trained on the 3-billion-word from Google News with a skip-gram model. The final model resulted in an F-measure of 90%. However, the model incorrectly identified some non-hate speech as hate speech. Also, the majority of the hated class is misclassified, while the majority of the offensive class is correctly identified.

A study in [29] utilized the Bidirectional Long Short Term Memory method and the word2vec feature extraction method with Continuous bag-of-word (CBOW) architecture to detect hate speech for Indonesian tweets. After applying the word2vec model, and by setting the epoch to 10, the learning rate to 0.001, and the number of hidden neurons to 200 their model achieves an F1-score of 96.29%. They also found that the addition of more LSTM layers can increase the accuracy by 2.27%. Gated Recurrent Unit (GRU) was implemented in [30] to classify Bengali comments on Facebook as Hate Speech, Communal attacks, Inciteful, Religious Hatred, Political Comments, and Religious Comments. They introduce an annotated Bengali corpus of 5,126 Bengali comments belonging to six classes. Their experiment shows that a classification model built using GRU can achieve an accuracy of 70.10%.

Table 1 Summary of public pages to retrieve comments and posts from Facebook and twit from Twitter

Page/account names	Page/account names
FBC Afaan Oromoo TV	Ethiopian Press Agency/Bariisaa
BBC Afaan Oromo TV	Oromia Democratic Part/ODP
OMN TV	Kush Media
Fanabc Afaan Oromo	OBS
Jawar Mohammed	Taye Dendea Aredo

Table 2 Summary of Balanced data distribution in four classes

Class	Class label	No of texts
Neutral	0	10,525
Hate	1	10,525
Offensive	2	10,525
Both	3	10,525

Afaan Oromo Corpus creation and annotation

The Afaan Oromo Text dataset for Afaan Oromo hate speech detection (AHSD), which is the main focus of analysis in this paper, is retrieved from comments and posts published on Facebook and Twitter from January 2019 to June 2019 by the authors of this paper.

Corpus collection

This work targets Facebook pages and Twitter accounts that are open to suspected hate speech rather than focusing on websites or blogs that already have specific agenda. In Ethiopia, it's common for social network communities are commonly posting on political and religious issues. The summary of Facebook pages and Twitter pages that was utilized to build the corpus is provided in Table 1. Those pages listed in Table 1 typically post discussions on political, social, economic, religious, and environmental issues that took place in Ethiopia.

In total, 35,200 posts and comments were collected. In order to remove the noise from the data set, rigorous preprocessing was carried out, which resulted in the removal of HTML, URLs, tags, emoticons, and other language scripts. By applying the data augmentation approach outlined in [31], the total dataset size was increased to 42,100. Detailed statistics of the balanced Afaan Oromo dataset categorized into 4 classes (i.e., NEUTRAL, HATE, OFFENSIVE, and BOTH) are provided in Table 2.

Annotation guideline

Annotation is an integral part of the development of Text classification. Annotated data provides useful quantitative information about the occurrence of certain contextual features. For the Afaan Oromo, there is no standardized and labeled corpus for hate speech detection. In hate speech detection, dataset annotation can be performed either manually

or crowd-sourcing. A research work in [32] shows annotations generated by expert annotators outperform crowd-sourcing strategy. The reason is that cultural norms play an important role in how hate is expressed and whether or not people perceive something as hate speech. Thus, authors also believe that testing for inter-annotator reliability may help alleviate these biases in some cases but will not necessarily neutralize the impact of views on what is hateful or not, which may be shared by the majority of annotators. Toward this, Afaan Oromo language speakers from different corners of Ethiopia were engaged in the annotation process. Each annotator labeled a text into one of the four classes as shown in Table 2. To ensure consistency among annotators they are provided with the following guidelines (or rules) for each annotator: A post has been marked as hate:

- If a post/comment uses references to the alleged inferiority or superiority of some target groups.
- If a post/comment affects different characteristics of the person and motivates audiences to take action or make violation.
- If a post/comment contains stereotype which means over-generalized belief about a given target.
- If a post/comment Accusing or Condemning people based on their target groups.

A post has been marked as Offensive:

- If a post or comment contains violent or insulting words but not possible to explicitly identify a target group in the post/comment.
- If a post or comment contains defamation, which is a false accusation a person or attack on a person's character.
- If a post or comment contains insulting, dirty, disgusting, or upsetting words but does not motivate the people to take action.

A post has been marked as Both (i.e, Hate and Offensive):

- If there is a combination of hateful expression, and use an insult, threats, or derogatory terms toward a target's groups.

The structure of the Afaan Oromo Hate Speech Detection system is depicted in Fig. 1. As shown in Fig. 1, the system has three main components, i.e., Data preprocessor, Feature extractor, and deep learning models. First, the system accepts a sequence of the token as the input, and then it goes through a data preprocessor module that discards URLs, tags, and other unnecessary inputs. Using a pre-trained Word2Vec or a Word-2Vec trained with the model, the feature engineering component will extract the relevant feature vectors that are used for classification. Finally, the deep learning model will take the vector that corresponds to the input to output the corresponding class label after repetitive training.

Data preprocessor

As shown in Fig. 1, the data preprocessor module is responsible to remove Twitter profile tags, hashtags, URLs, emojis, and stop words (words that do not add any meaning or value to the sentence) in addition to the tokenization task.

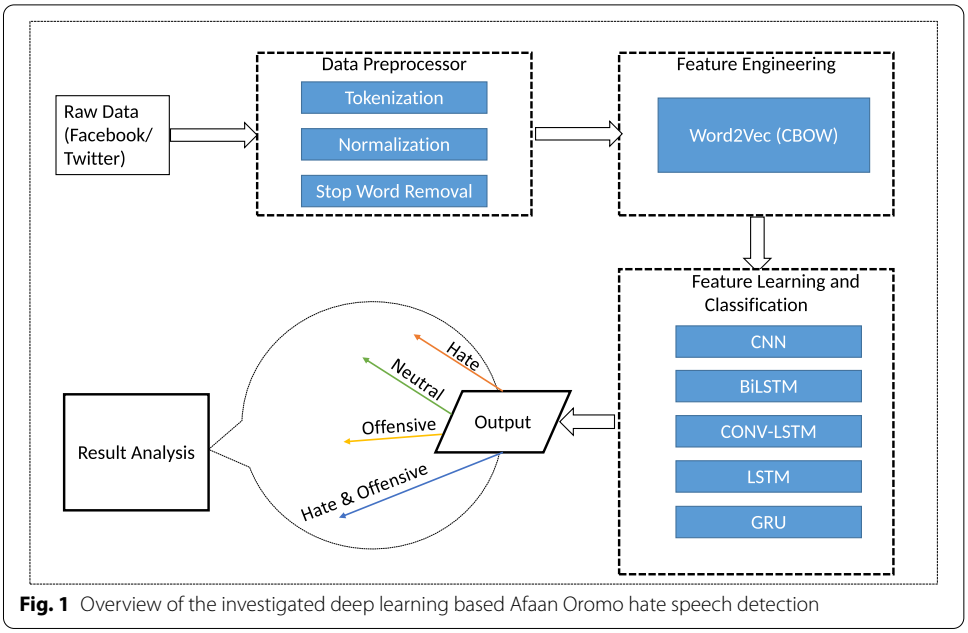

Fig. 1 Overview of the investigated deep learning based Afaan Oromo hate speech detection

Feature engineering

Once the raw data undergoes preprocessing, the Feature engineering part will transform each token into its corresponding embedding vector (a vector that summarizes a given token into a new representation that captures the contextual similarity). Deep neural networks can capture features automatically without any human intervention but they are designed to learn from numerical data or word vectors. Among the available word embedding techniques, a research in [33] reveals that a word2vec [14] model is found to be the effective distribution for hate speech detection research. So, in this study, a type of word embedding technique called Word2Vec is used, which is an algorithm that uses a neural network to learn word embedding. Its goal is to estimate each token's position in a multi-dimension vector space based on the similarity of different tokens.

There are numerous variants of word2vec, among those the Continuous Bag-of-words (CBOW) model was adpoted for this study. In the CBOW model, the model learns the distributed representation by training a feed-forward neural network using word co-occurrence with language modeling to predict the word in the given context. The aim of training the CBOW model is to maximize the log-likelihood probability which is calculated as follows:

$$\frac{1}{t}\sum_{t=1}^{T}\log p(w_t \| w_c) \tag{1}$$

where w in Eq. 1 is the target word, and w_c represents the sequence of words in context. Word2Vec model can be implemented in two ways: (1) pre-training and using it as an input layer at the beginning of the model architecture, or (2) training it with the model itself. The impact of both approaches in detecting Afaan Oromo hate speech is assessed using Python together with Keras and Tensorflow. Overall, the utilization of embedded representation has two implications for the proposed hate detection models

(1) it reduces the dimension of the input into a k-size vector, and (2) it is a more expressive representation as it captures contextual resemblance and semantic sequence of data.

Feature learning and classification using deep leaning models

Once the feature engineering module produces the embedding vectors, the next step is feature learning and classification. In this study, five deep learning models are selected for comparison.

- CNN which is a class of Deep Learning model that use convolutional layers and maximum pooling or max-overtime pooling layers to extract higher-level features.
- LSTM is a powerful kind of RNN used for processing sequential data such as sound, time series (sensor) data or written natural language.
- BiLSTM is is a hybrid bidirectional LSTM and CNN architecture.
- GRU is similar to long short-term memory (LSTM) with a forget gate, but has fewer parameters than LSTM, as it lacks an output gate.
- CONV-LSTM is a type of recurrent neural network for spatio-temporal prediction that has convolutional structures in both the input-to-state and state-to-state transitions.

After grid search, the optimal hyperparameters selected for those models are summarized in Table 3.

Experiment and discussion

Evaluation setup

Five-fold cross-validation is used to train and test the model. That is out of 42,100 total sentences, 33,680 of them are used for training and the remaining one is used for testing at a time. In doing partition, due consideration was given to avoid overlap and to preserve the class distribution of the training set as closely as possible. In each case, the remaining training data (8420) was used as the testing set. Then, for every deep learning model investigated in this study training was made independently using a set for parameter optimization, and a development set for validation purposes. The performance report in this study indicates the decision for each method by averaging the result

Table 3 Architectural hyperparameters

Hyparameter name	Hyparameter value
Number of Convolution Layer	3
Number of Filters in Convolution Layer	250
Filter Size	3×3
Dropout Rate	0.5
Batch Size	128
Embedding Dimesion	300
Hidden Layer Activation Function	Relu
Output Layer Activation Function	SoftMax
Optimizer	AdaGrad
Learning Rate	0.001

Table 4 Evaluation metrics

Metrics	Formula
Precision	TP/(TP+FP)
Recall	$TP/(TP + FN)$
F1-Score	2×((precision × recall)/(precision + recall))

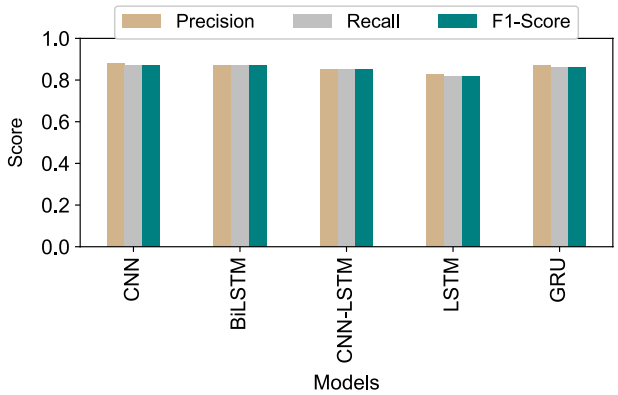

Fig. 2 Weighted average score of deep learning models with pre-trained word embedding

obtained after each fold. To provide the evaluation results, three evaluation metrics are adopted in this work. These are "Precision", "Recall", and "F1-score". The equation of these three performance metrics is shown in Table 4.

In Table 4, "TP", "FP", "TN", and "FN" represents True Positive, False Positive, True Negative and False Negative, respectively. In this work, fivefold cross-validation is used, where 4 of the fold are used for training the model, and onefold will be used for testing the model (i.e., at a time). Finally, the average test result after fivefold is reported in Figs. 4, 5.

Results and discussion

This section presents the results of the experiments investigated using different deep learning models. As our primary goal is to compare the performance of models in different circumstances, the results for each model are presented in separate Tables. Three series of experiments are conducted using five different deep learning models. The first one involves the case where the word embedding is pre-trained (see Figure xxx), whereas the second one is the case where word embedding is trained together with the model itself. The third and fourth experiments are conducted to assess the impact of data augmentation on classification performance.

Figure 2 shows the weighted average results of the experiments carried out with five deep learning methods when the pre-trained Word embedding is used for feature extraction. As can be seen in Fig. 2, the BiLSTM and CNN accomplished the best performance (with a weighted average F1-score of 87%). The average F1-score of CNN-LSTM, GRU, and LSTM is 85%, 86%, and 82%, respectively.

Figure 2 shows the result using a word embedding structure that was trained with the model as well. The results of these experiments are listed in Fig. 3. As it can be seen in Fig. 3, training the model with embedded representation slightly increases the classification performance of the investigated model by 1.5% on average. Authors

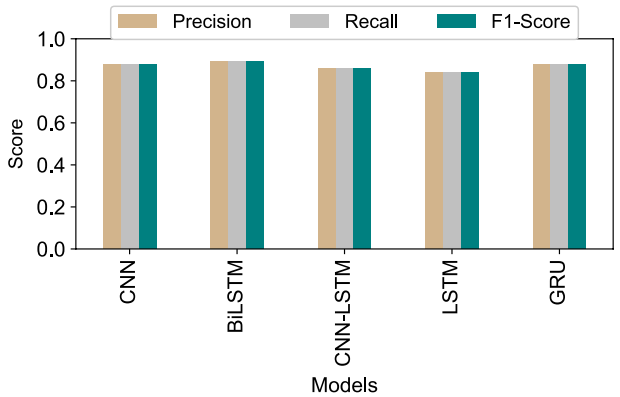

Fig. 3 Weighted average score of deep learning models trained with embedding

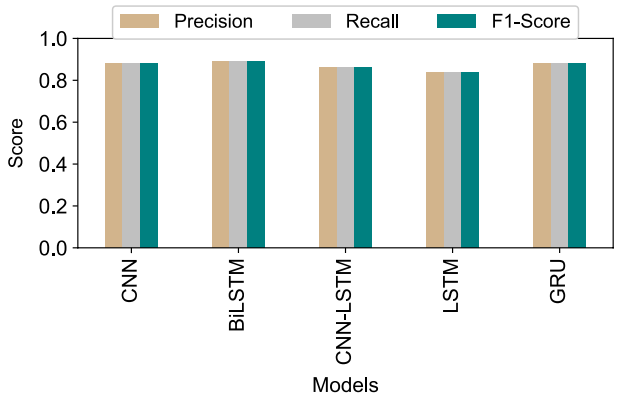

Fig. 4 Weighted average score of deep learning models with pre-trained word embedding and data augmentation

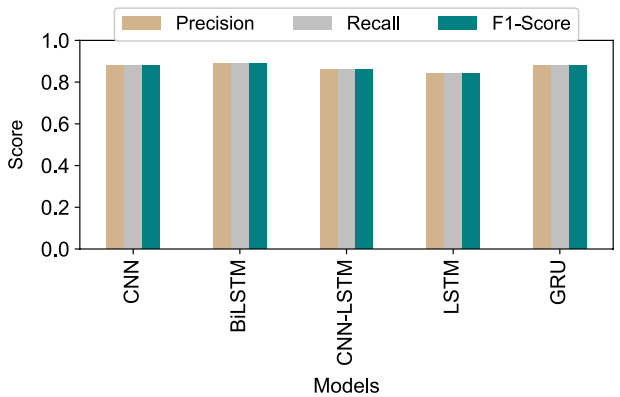

Fig. 5 Weighted average score of deep learning models trained with embedded representation and data augmentation

believe that training the model with embedding representation will allow the model to optimize the trainable parameters, and hence improve the classification performance.

Overall, three main observations can be made by comparing the experimental results of the neural network. First, a model trained with embedding representation is able to capture syntactic and semantic relations of Afaan Oromo words. Secondly, the data augmentation mechanism improves the performance of the hate detection models. Finally, Bidirectional Long Short-Term Memory (BILSTM) achieved the highest F-score of all classifiers used in our experiments.

Conclusion

This paper presents an empirical evaluation of five deep learning models (i.e., CNN, LSTM, GRU, BiLSTM, and CNN-LSTM) for detecting Afaan Oromo hate speech. First, the largest Afaan Oromo Corpus for Hate Speech Detection was prepared. The data used to train and test the model is collected from Twitter and Facebook pages. The finding shows that the best performance was showcased by the BiLSTM with a weighted classification F1 score of 91%. Moreover, the research also compared the effect of using a pre-trained embedded representation with the one training with the model. From the experiment, the authors conclude that training an embedded representation with a model and incorporating augmented samples will enhance the classification performance of all the investigated models. Considering the dataset size investigated in this paper the result performance of the deep learning model at detecting Afaan Oromo hate speech is promising. In future work, we would like to investigate the performance of classifier ensembles and meta-learning for this task. Future research needs to consider a mechanism to incorporate divergent opinions. Also, the performance is not perfect, which means that users will face up with misclassified content. Comparative analysis with models built for other languages would be interesting to know which is also one of the future directions.

Abbreviations

CNN: Convolutional Neural Networks; LSTM: Long Short-Term Memory; Bi-LSTM: Bi-directional Long short-term Memory; GRU: Gated Recurrent Unit; AHSD: Afaan Oromo hate speech detection; CONV-LSTM: Convolutional Long Short-Term Memory.

Acknowledgements

Not applicable.

Author contributions

GOG contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript. The author read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent fo publication

Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 23 January 2022 Accepted: 26 May 2022
Published online: 02 June 2022

References

1. Negussie N, Ketema G. The relationship between facebook practice and academic performance of university students. Asian J Humanities Soc Sci (AJHSS). 2014;2(2):1–7.
2. Zhang Z, Robinson D, Tepper J. Detecting hate speech on twitter using a convolution-gru based deep neural network. In: European Semantic Web Conference. Springer. 2018; p. 745–60.
3. Pereira-Kohatsu JC, Quijano-Sánchez L, Liberatore F, Camacho-Collados M. Detecting and monitoring hate speech in twitter. Sensors. 2019;19(21):4654.
4. Aluru SS, Mathew B, Saha P, Mukherjee A. Deep learning models for multilingual hate speech detection. 2020; arXiv preprint arXiv:2004.06465.
5. Alshaalan R, Al-Khalifa H. Hate speech detection in saudi twittersphere: A deep learning approach. In: Proceedings of the Fifth Arabic Natural Language Processing Workshop. 2020; p. 12–23.
6. Zimmerman S, Kruschwitz U, Fox C. Improving hate speech detection with deep learning ensembles. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). 2018;
7. 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, 2017;vol. 11.
8. Das M, Mathew B, Saha P, Goyal P, Mukherjee A. Hate speech in online social media. ACM SIGWEB Newsletter (Autumn); 2020. p. 1–8.
9. 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, 2017; p. 759–60.
10. Wubetu Barud AO. Detection of fake news and hate speech for Ethiopian languages. J Big Data. 2022;9:66.
11. Abebaw Z, Rauber A, Atnafu S. Multi-channel convolutional neural network for hate speech detection in social media. In: International Conference on Advances of Science and Technology. Springer: Berlin, 2021. pp. 603–18.
12. Defersha N, Tune K. Detection of hate speech text in afan oromo social media using machine learning approach. Indian J Sci Technol. 2021;14(31):2567–78.
13. Defersha NB, Kekeba K, Kaliyaperumal K. Tuning hyperparameters of machine learning methods for afan oromo hate speech text detection for social media. In: 2021 4th International Conference on Computing and Communications Technologies (ICCCT), pp. 596–604. IEEE, 2021.
14. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, 2013;pp. 3111–3119.
15. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80.
16. Schuster M, Paliwal KK. Bidirectional recurrent neural networks. IEEE Trans Signal Process. 1997;45(11):2673–81.
17. Le Q, Mikolov T. Distributed representations of sentences and documents. In: International Conference on Machine Learning, 2014;pp. 1188–1196. PMLR
18. Arango A, Pérez J, Poblete B. Hate speech detection is not as easy as you may think: A closer look at model validation. In: Proceedings of the 42nd International Acm Sigir Conference on Research and Development in Information Retrieval, 2019; p. 45–54.
19. Ganfure GO, Wu C-F, Chang Y-H, Shih W-K. Deepprefetcher: A deep learning framework for data prefetching in flash storage devices. IEEE Trans Computer Aided Design Integrat Circuits Syst. 2020;39(11):3311–22.
20. Olani G, Wu C-F, Chang Y-H, Shih W-K. Deepware: Imaging performance counters with deep learning to detect ransomware. IEEE Trans Computers. 2022.
21. Pennington J, Socher R, Manning CD. Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014; p. 1532–1543.
22. Joulin A, Grave E, Bojanowski P, Douze M, Jégou H, Mikolov T. Fasttext. zip: Compressing text classification models. 2016; arXiv preprint arXiv:1612.03651.
23. Camacho-Collados J, Pilehvar MT. From word to sense embeddings: a survey on vector representations of meaning. J Artif Intell Res. 2018;63:743–88.
24. Rodriguez A, Argueta C, Chen Y-L. Automatic detection of hate speech on facebook using sentiment and emotion analysis. In: 2019 international conference on artificial intelligence in information and communication (ICAIIC), 2019; p. 169–174.
25. Del Vigna12 F, Cimino23 A, Dell'Orletta F, Petrocchi M, Tesconi M. Hate me, hate me not: Hate speech detection on facebook. In: Proceedings of the First Italian Conference on Cybersecurity (ITASEC17), 2017; p. 86–95.
26. Mossie Z, Wang J-H. Social network hate speech detection for amharic language. Computer Sci Inf Technol. 2018;9:41–55.
27. Gambäck B, Sikdar UK. Using convolutional neural networks to classify hate-speech. In: Proceedings of the First Workshop on Abusive Language Online, 2017;pp. 85–90
28. Biere S, Bhulai S, Analytics MB. Hate speech detection using natural language processing techniques. Master Business AnalyticsDepartment of Mathematics Faculty of Science. 2018;
29. Isnain AR, Sihabuddin A, Suyanto Y. Bidirectional long short term memory method and word2vec extraction approach for hate speech detection. IJCCS (Indonesian Journal of Computing and Cybernetics Systems). 2020;14(2):169–78.
30. Ishmam AM, Sharmin S. Hateful speech detection in public facebook pages for the bengali language. In: 2019 18th IEEE international conference on machine learning and applications (ICMLA), 2019; p. 555–60.

31. Aroyehun ST, Gelbukh A. Aggression detection in social media: Using deep neural networks, data augmentation, and pseudo labeling. In: Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018), 2018; p. 90–7.
32. Mubarak H, Darwish K, Magdy W. Abusive language detection on arabic social media. In: Proceedings of the First Workshop on Abusive Language Online, 2017; p. 52–6.
33. Gupta S, Waseem Z. A comparative study of embeddings methods for hate speech detection from tweets. 2017.

Publisher's Note