JOIN (Jurnal Online Informatika)
p-ISSN: 2528-1682, e-ISSN: 2527-9165
Volume 10 Number 2 | December 2025: 352-361
DOI: 10.15575/join.v10i2.1631

Cyberbullying Detection in the Libyan Dialect Using Convolutional Neural Networks

Sara M. Elgoud[1], Mustafa Ali Abuzaraida[2], Zainab S. Attarbashi[3], Mohamed Ali Saip[4]
[1]Department of Computer Science, College of Information Technology, Al-Asmaria Islamic University, Libya
[2]Department of Computer Science, Faculty of Information Technology, Misurata University, Libya
[3]Faculty of Information and Communication Technology, International Islamic University Malaysia, Malaysia
[4]School of Computing, Universiti Utara Malaysia, Sintok, Kedah, Malaysia

Article Info

Article history:

Received May 05, 2025
Revised June 21, 2025
Accepted July 13, 2025
Published November 10, 2025

Keywords:

Arabic Dialect
Convolutional Neural Network
Cyberbullying
Deep Learning
Meta-Learning
Natural language processing
Removing stopwords

ABSTRACT

Recently, the widespread use of social media has increased, leading to increased concerns about cyberbullying. It has become imperative to intensify efforts and methods to detect and manage cyberbullying through social media. Arabic has recently received increasing attention to improve the classification of Arabic texts. Given the multitude of Arabic dialects used on social media platforms by Arabic speakers to express their opinions and communicate with each other, applying this approach to Arabic becomes extremely challenging due to its structural and morphological complexity. Analyzing Arabic dialects using Natural Language Processing (NLP) tools can be more challenging than Standard Arabic. In this paper, the impact of using stopword removal and derivation techniques on detecting cyberbullying in the Libyan dialect was presented. The efficiency of text classification was compared when using a Libyan dialect word list alongside pre-generated Modern Standard Arabic (MSA) lists. The texts were classified using Convolutional Neural Network (CNN) classifiers, and the experiments showed that when using Libyan dialect words, the accuracy results were 92% and 83%, and when using only Standard Arabic stop words, the accuracy results were dropped to 91% and 77%. Based on these results, the higher accuracy was obtained when using the presented stop words list which it is specific to the Libyan dialect, and they had a positive impact on the results, better than Standard Arabic stop words.

Corresponding Author:

Mustafa Ali Abuzaraida,
Department of Computer Science, Faculty of Information Technology, Misurata University, Libya
Email: abuzaraida@it.misuratau.edu.ly

1. INTRODUCTION

In recent times, the internet and social media have become essential sources of data as they are platforms for reading and writing by many users, and speakers of different languages access the internet on a daily basis because it is not only used by English speakers. For example, because Arabic is spoken in more dialects, people tend to express their opinions, thoughts, feelings, and comments on various issues posted on social media in their dialects. Dialects are the informal form of language. Every country in the Arab world has its own dialect, and that is why there is a need to process this type of data. This huge amount of data generated has attracted business owners, marketers, government institutions, scientists, and researchers as well [1].

Text classification is the process of labeling test items using training data. It has wide applications in various fields such as news classification, online libraries, author authentication, detecting cyberbullying, and many more. Cyberbullying happens when someone uses the internet to hurt or annoy a celebrity or anyone on a social media platform, game, or app, and includes posts, comments, messages, chats, live broadcasts, photos, videos, and emails [2].

Over the past years, many studies and research have been conducted to detect cyberbullying as a text classification problem, especially in English [3]. Since cyberbullying is a problem facing the whole world, this problem also faces the Arab nation, and only a few studies have been conducted to work on detecting cyberbullying in the Arabic language [4].

Despite this, there is a lack of linguistic resources for Arabic, and most of them are under development. In order to use Arabic in different dialects in modern ways, some text processing, such as word removal, is needed. There are some unofficial word list sources available to the public, but they work on Modern Standard Arabic and not Arabic Dialects. This paper addresses the impact of word removal on Libyan Arabic texts.

Stopwords are the words that are more common in many sentences and do not have a significant semantic relationship to the context in which they appear. There are some researchers who have created word lists. The lists are also available in Modern Standard Arabic (MSA). A reference has created a word list of Modern Standard Arabic from the most common words that appear in its corpus [5]. Improving stopword lists for the Libyan dialect (spoken by more than 7 million people) is essential to ensure that natural language processing of Libyan texts is more accurate, efficient, and aware of the linguistic and cultural nuances of this dialect. Using generic stopword lists can lead to inaccurate or misleading results when dealing with Libyan dialect texts. Therefore, improving password lists specific to the Libyan dialect can improve the model performance and achieve higher accuracy [6] [7].

Researchers conducted a study [8] on detecting abusive language or cyberbullying. They created a new balanced Arabic dataset for use in the abusive content detection process, as having a balanced dataset for the model would improve the accuracy of the models. The study used individual classifiers using ensemble machine learning. Applying the study to three Arabic datasets, the results showed that the individual learner machine learning strategy outperformed the ensemble machine learning approach. Outperforming the best single-learner classifier (65.1%, 76.2%, and 98%) for the same datasets, with accuracy scores of 71.1%, 76.7%, and 98.5% for each of the three datasets used. In the same context, researchers [9] presented a study to enhance the accuracy of Arabic social media cyberbullying detection using deep learning approaches, specifically CNN, LSTM, and CNN-LSTM models. The study aimed to bridge the research gap in the field of Arabic social media cyberbullying detection and enhance the accuracy of Arabic social media cyberbullying detection using deep learning models. The study found that LSTM performed better with an accuracy of 96.44%, a recall of 97.03%, and an F1 score of 96.73% for two-class classification. For six-class classification, LSTM demonstrated superior performance with an impressive accuracy of 95.59%. Three deep learning models were implemented to train the data: a CNN, an LSTM, and an RNN. Finally, the researchers presented a study [10] on Arabic dialect detection and classification, comparing the performance of traditional machine learning and deep learning models. The paper reviews various approaches and techniques used to identify Arabic dialects, including frequency-related features, deep learning-based methods, and transfer learning. It also highlights the importance of data cleaning and preprocessing in improving the accuracy of dialect classification.

The proposed hybrid model, equipped with an attention mechanism, achieves the highest accuracy rate of 88.73%, outperforming state-of-the-art models such as RoBERTa, MARBERT, AraBert, and mBERT. The TF-IDF word representation method also performs better than Word2Vec and GloVe. While some research has focused on creating spam word lists, to the best of our knowledge, no one has created a spam word list for Arabic dialects. In [11], they proposed a methodology for preparing Arabic text corpora from online social networking sites (OSNs) and review websites for a sentiment analysis (SA) task. They also proposed a methodology for generating a stopword list from the prepared corpora to investigate the impact of removing stopwords on the SA task. The problem is that previously generated stopword lists were in Modern Standard Arabic (MSA), which is not the common language used in OSNs. In [12], they created a stopword list for conversations and a corpus-based list of words, and compared the effectiveness of the new lists with the previously generated MSA lists. The validity of the conversations is verified using sequential pattern mining. The study collected all possible words or tools that could be considered stopwords from different grammatical categories in Arabic and then divided them into two categories: words that accept suffixes and words that do not accept suffixes.

In [13], they created a stopword list for the central query language from the best nonsense words appearing in their corpus. The study used natural language processing techniques (normalization, stopword removal, and segmentation) to validate the P-Stemmer using different classifiers (NB, SVM,

p-ISSN: 2528-1682
e-ISSN: 2527-9165

RF, KNN, and K-Star) to improve text classification results. A new Arabic dataset was created, and different classifiers (NB, SVM, RF, KNN, and K-Star) were applied with and without the P-Stemmer. The performance of the classifiers was evaluated using metrics such as the F1 measure, precision, and recall. The experimental work demonstrated that the Naive Bayes classifier is the best classifier to use with the P-Stemmer, with a classification accuracy of 0.75, followed by support vector machines. The results of the study showed that combining the P-Stemmer with the NB classifier achieves the best performance with an accuracy of 0.75, and that using the P-Stemmer increases accuracy by up to 6%. The results also showed that the performance of the classifiers varies depending on the classification category, with some classifiers performing better in some categories than others. [14] in the field of sentiment analysis in the Algerian dialect by creating a custom dataset and utilizing an advanced deep learning model called BERT. The research achieved an F1 score of 78.38% and an accuracy of 81.74% on the test set, demonstrating the effectiveness of the approach and the potential of BERT for sentiment analysis in the Algerian dialect. Key findings include the optimal preprocessing tasks, model architecture, and hyperparameters that achieve the best F1 score and accuracy on the test set. This research highlights the importance of studying the Algerian dialect and the potential of using state-of-the-art deep learning models for natural language processing in this field.

In [15], a sentiment analysis study was conducted on Arabic comments in the Moroccan dialect using various machine learning and deep learning algorithms, including logistic regression, decision tree, support vector machine, naive Bayesian polynomial, XGBoost, and neural network. The study found that the best performance was achieved using the [Unigram/TF - IDF] and [Unigram + Bigram/TF - IDF] combinations, regardless of the algorithm used. Logistic regression with Unigram extraction and TF - IDF weighting was the most efficient in terms of accuracy and reliability. The ARABERT model achieved an accuracy of approximately 88%.

Finally, a CNN model was chosen to achieve the study's goal: improving the excluded word lists for the Libyan dialect, ensuring that Libyan texts are processed in a more accurate and efficient natural way, while being aware of the linguistic and cultural nuances of this dialect. Using generic excluded word lists may lead to inaccurate or misleading results when dealing with Libyan texts.

The rest of this paper is organized as follows: Section 2 describes the methodology, Section 3,4 present the results and discussions, and finally, the conclusion is summarized in Section 5.

2. METHOD

The methodology followed in this study will be done by four stages. In the first stage, data is collected, and in the second stage, the collected data and the annotated data will go through data pre-processing and data cleaning to remove unwanted symbols and codes. In the third stage, a deep learning model, a CNN model, is implemented to train the data, while in the final stage, the evaluation and analysis of the results are carried out. Figure 1 illustrates the methodology of this study.

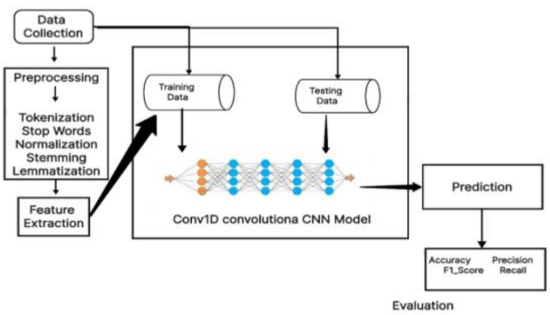

Figure 1. Proposed methodology Phases

2.1 Data collection and analysis

In this study, the first dataset was collected from YouTube, Twitter, and Facebook from Libyan news websites and channels covering political topics related to Libyan affairs during the period from

2020 to 2023. The dataset totaled approximately 7,500 comments divided into two categories (bullying and neutral). This is a sentence-level dataset for the Libyan dialect, and it is the first publicly published and manually annotated dataset for this task.

This dataset is considered balanced. Unfortunately, there is a clear shortage of Arabic datasets in the field of Arabic natural language processing. Consequently, providing this dataset will contribute to supporting efforts in this direction. Table 1 shows the number of samples for each category, while Table 2 highlights some samples for the (bullying and neutral) categories.

Table 1. Number of samples in the dataset (1)

Label	No of Samples
Bullying	3843
Neutral	3648
Total	7491

Table 2. Examples of the dataset

Comments	Comments translation	label
يا عميل ياعيل	you agent, you child	Bullying
التركي لي من مصراته يبي قطع راس وتهبيك	Someone from Misrata wants to cut off your head and humiliate you	Bullying
مشاء الله مشاءالله يا استاذ ياليثني انت	God willing, God willing, you impressed me	Neutral
اسير اتشاد المللعون	ırsed Chad	Bullying
نسأل الله ان يحفظه بطل وأسد نسأل الله ان يطول في عمره	We ask God to protect the horse, a hero and a lion, we ask God to prolong his life	Neutral
الصراحه اسلوبه الواطي يمشي مع مستوى القرامطه	His low style matches the level of the Qaramita	Bullying
مريض نفسي	Mentally ill	Bullying

The second set of data is publicly available data and is also in the Libyan dialect, with a total of 8560 comments divided into two categories (0, 1), negative and positive, and is considered an unbalanced dataset. Table 3 shows the number of samples for each category, and Table 4 highlights some samples of the Negative and Positive categories.

Table 3. Number of samples for each category

Number of samples	Class
6406	0
2154	1

Table 4. Examples of the second dataset

Comments	English Translation	label
كنه نت المدار	It's like Al-Madar net	1
نت المدار كنه ضعيف هك	Al-Madar net is so weak	0
نت المدار يشفط في شفط طيب	Al-Madar net is sucking up data well	0
المدار حست بيا في هالثانية قالتلي تبقى لديك 5% يعني في الحالتين حتقعدي من غير نت	Al-Madar sensed me in this moment and told me you have 5% left, meaning in both cases you'll be without internet	0
هو نت المدار ضعيف اليومين هذينا ولا نا نتخايل	Is Al-Madar net weak these days or am I imagining things?	0
نت المدار زبالة بككل	Al-Madar net is completely trash	1
نت المدار رددد	Al-Madar net is bad	1

For ethical considerations, the data used in this study consists of publicly available user comments from platforms such as YouTube, Twitter, and Facebook. All data was anonymized to ensure that no personal identifiers were retained during the collection and annotation processes. Ethical

p-ISSN: 2528-1682
e-ISSN: 2527-9165

considerations were taken into account throughout the research. Comments were reviewed by language experts under guidelines designed to minimize subjective judgment and bias during annotation. Although ethical approval was not required due to the use of publicly accessible data, the aware of the sensitivity surrounding cyberbullying content. Also, recognizing the potential bias introduced by focusing on a specific dialect and socio-political context, future work will aim to broaden the dataset to include other Arabic dialects and domains while incorporating formal ethical review procedures.

2.2 Data Pre-Processing

The pre-processing process consists of five steps: removing words, tokenization, normalization, stemming, and lemmatization. However, the pre-processing stage faces many challenges and obstacles, especially when dealing with emotions expressed in unstructured languages such as Arabic dialects. The pre-processing techniques [16][23] have been used in this study, including the following steps:

2.2.1. Removing stopwords

Removing stop words is one of the steps in pre-processing texts before analysis. It is based on excluding common words that have no value in the text and do not help the user to analyze texts and extract the meaning of the text. A stop list is a list that contains all the words of a sentence such as: Pronouns (أنا، هو، هي، أنت، نحن، هم) Conjunctions: (...في، على، إلى،) Prepositions: (..ثم، حتى، أو، ولكن), words that are not useful in classification such as (...بعض، أيضا، حاليا،), directions: (أمام، خلف، بجانب، قبل، بعد، يسار يمين) and any word that does not add any meaning to the text: such as (أمس، اليوم، غداً) in addition to the additional words that have been added to the list in the Libyan dialect which are different from other Arabic dialects like Moroccan and Egyptian, as shown in Table 5.

Table 5. Examples of the stop words list in the Libyan dialect

Libyan dialect	Moroccan	Egyptian	Meaning in Arabic	Meaning in English
معش	معادش	ماتعملش	لا تفعل	Don't do
شوف	دا اتشوف	بص	انظر	Look
هلبة	بالزاف	كثير	كثيراً	Much
كنك	شنو وقع	مالك	ماذا حدث	What happened
هكي	هاكا	كدا	هكذا	Thus
شني	شنو	ايه	ماذا	What
خلاص	بركة	كفاية	يكفي	Enough
خلي	خلي	سيب	اترك	Leave it

Stopword removal techniques were applied based on a Libyan dialect-specific list. A total of 1300 stopwords were extracted for the Libyan dialect, 500 of which were extracted from [13] and 500 from [14]. Only about 300 words were added as stopwords in line with the dataset of this study, and to avoid deleting important words that might affect the accuracy of the classifier. These stopwords were used during the preprocessing stage, specifically for the stopword removal technique. In addition, a comprehensive MSA list containing all possible prefixes and suffixes was used in this analysis[16][26].

2.2.2. Tokenization

At this stage, sentences or strings are divided into words or symbols using a separator, which can be any punctuation character or space, usually words or phrases. Using the "hash" interface of the Keras library, comments are converted into sequences of integers, creating a numerical representation of the text, which is a prerequisite for processing by neural network models[17].

2.2.3. Normalization

During this process, words are normalized by removing any possible misinterpretations of letters. Then, replacing some Arabic letters was done with their official form due to a common spelling mistake in some words, such as replacing a letter at the end of a word with the letter [ه], and replacing the letter [أ إ آ] with [ا][18].

2.2.4. Stemming

Aims to reduce words to their basic uninflected forms. Sometimes, the stem differs from the root, but it is useful because related words are often linked to the same stem, even if the stem is not a proper root. In this study, the Tashaphyne library was used to apply the Stemming technique to the dataset. Farasa root was used as an additional tool: Farasa provides data extraction because words have different structures, especially in Arabic [19][21].

2.2.5. Lemmatization

Lemma analysis is a linguistic method that involves analyzing the form of a word, removing its inflectional suffix, and creating its basic form. This technique is used to return words to their original root based on the morphological analysis of the sentence [19][20].

After the text cleaning is completed, the features are extracted, and each word is converted into features using the inverse frequency of the document (TF-IDF). The final stage of preprocessing is to initialize the embedding layer within the neural network. The embedding dimension is defined, which specifies the size of the vector space in which the words will be represented in vectors [22,23]. Each word has a length of (100) dimensions provided by Keras [24]. This functionality allows the text collection to be routed by converting each text into a series of integers, which facilitates the classification efforts in the model by providing a clear and measurable target for the algorithmic prediction, as in Figure 2. The next step is to split the data into 80% for training and 20% for testing the proposed model.

```
Vector for 'المسئول':
[ 6.11062385e-02 -3.81000564e-02 -4.14297432e-02  6.06251787e-03
 -3.42502519e-02  3.20238620e-02  2.63297204e-02  2.24426738e-03
 -4.45121974e-02  1.33695668e-02  3.37054841e-02  6.89330138e-03
  1.73648559e-02  2.12060288e-02 -1.56488195e-02  2.25194320e-02
  2.13651620e-02 -5.50456941e-02 -2.20914222e-02  3.55730690e-02
 -3.77184935e-02  3.53689305e-03 -9.31204180e-04 -2.86220517e-02
  3.57988440e-02  4.50025126e-02 -1.91386025e-02 -3.16882245e-02
 -5.66606596e-03  4.58574779e-02 -1.59256794e-02 -1.17114629e-03
 -2.45661605e-02  5.17028337e-03  3.66011783e-02  2.58637145e-02
 -3.51013988e-02  4.00364101e-02  1.24253202e-02  1.34703526e-02
 -3.42814531e-03 -1.10814609e-02 -3.51301059e-02 -1.74806789e-02
 -3.13031906e-03 -3.32164094e-02 -3.07346564e-02  1.70916691e-02
  3.32419872e-02 -2.19186880e-02  3.73957716e-02 -1.38753941e-02
 -2.38303430e-02 -1.89642888e-02  1.11646615e-02 -1.48536516e-02
 -1.58486608e-02 -3.03600002e-02 -3.36106354e-03  4.42349724e-03
 -1.76656581e-02 -1.13670155e-02 -1.56557932e-02 -4.47290614e-02
 -1.23928417e-03  1.01283165e-02  1.24929463e-02  3.04365736e-02
 -2.09224150e-02 -2.65362114e-03  2.61894036e-02  1.17815873e-02
  5.01988688e-04 -1.18590351e-02  5.78230321e-02 -8.07859283e-03
  6.58528879e-02  2.28522718e-02  1.70195736e-02  1.99959464e-02
 -3.50253396e-02  6.06385730e-02 -4.07775268e-02  5.59688583e-02
 -8.42902227e-05 -2.50058919e-02  3.89913726e-03  2.74165962e-02
  5.32119581e-03  4.37708795e-02 -1.84921380e-02  7.49670388e-03
  6.62231222e-02 -2.64729075e-02 -5.47405072e-02 -3.41401771e-02
  5.34560718e-02  2.54622251e-02 -2.92192353e-03  3.35700586e-02]
```

Figure 2. Representing words into vectors

2.3 CNN architecture

The architecture used in this research work is built on the Keras sequential model framework [24]. The sequential model is a linear combination of layers. A convolutional neural network (CNN) is created with the following layers:

- Embedding layer: maps words to numerical representations, capturing semantic relationships.
- 1D convolutional layers: extract features from sequences, and identify patterns within text.
- Global maximum pooling layer: summarizes the extracted features and provides a summary representation.
- Dense layers: perform further processing and classification, turning features into predictions.
- Dropout layer: introduces randomization to prevent overfitting, improving the model's generalizability.

Figure 3 shows the structure of a CNN model (layers, kernel sizes, flow, activation functions). Then, Table 6 shows the parameters of the CNN model.

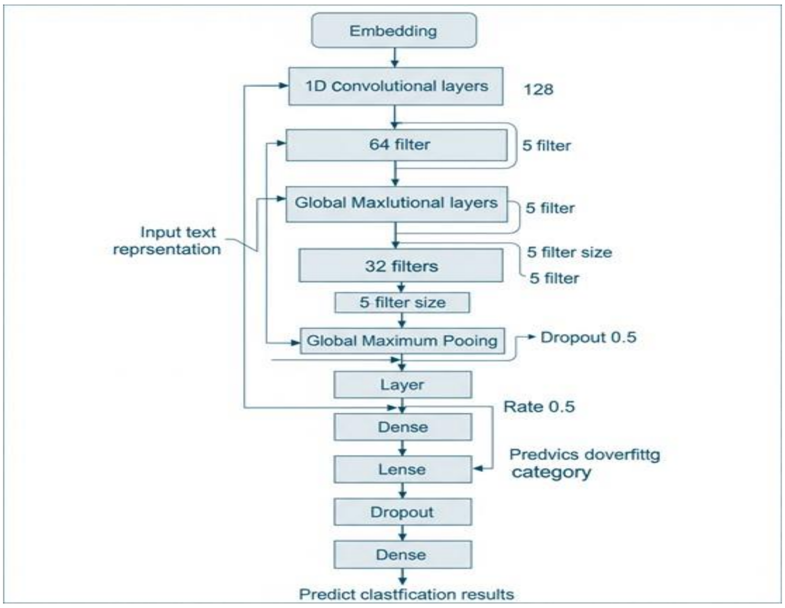

Figure 3. The diagram shows the structure of a CNN model

Table 6. Details the training parameters for the CNN model.

Parameter	Value
Optimizer	Adam
Loss Function	Categorical Cross entropy
Metrics	Accuracy
Number of Epochs	10
Validation Split	0.2
batch size	32
Number of filters	128 ,64 ,32
Filter size	5
Dropout layer rate	0.5

3. RESULT AND DISCUSSION

In this study, four experiments were conducted to test the effect of removing unwanted words from different lists by testing the pre-generated MSA list separately and the list containing the Libyan dialect stop words, including Stemming, Lemmatization and TF-IDF. Each dataset has two experiments, and these experiments were conducted using a CNN classifier. Each dataset was divided into training and testing folds with 80% for training and the rest for testing. The precision, accuracy, recall and F1-score metrics were used to evaluate the models.

In the first experiment for balanced data, each word was treated as a feature and assigned a weight based on its importance in the dataset. Alternatively, this experiment was conducted to study the effect of removing unwanted words from the list containing stop words of the Libyan dialect. Table 7 shows the results of the first experiment.

Table 7. First Experiment Results

Model	Categories	Precision	Recall	F1-Score	Accuracy
	CB	83%	82%	84 %	
CNN	Not CB	84%	81%	83 %	84%

In the second experiment, the balanced dataset was used with the pre-generated MSA list to see how removing stop words affected the Libyan dialect data. Table 8 shows the results of the second experiment.

Table 8. Second Experiment Results

Model	Categories	Precision	Recall	F1-Score	Accuracy
CNN	CB	76%	75%	77 %	77%
	Not CB	76%	73%	76 %	

As mentioned earlier, in the third experiment, unbalanced Libyan data were used. In this study, the problem of unbalanced data is out of the study scope and will not include any use of stratified weighting or oversampling techniques. Instead, the dataset is used for comparison purposes only, and the undesired words were used from the list of stop words in the Libyan dialect. Table 9 shows the results of the third experiment..

Table 9. Third Experiment Results

Model	Categories	Precision	Recall	F1-Score	Accuracy
CNN	CB	91%	93%	92 %	92%
	Not CB	86%	83%	85 %	

In the fourth experiment, the unbalanced Libyan dataset was used with the pre-generated MSA list to see how removing stop words affects the Libyan dialect data. Table 10 shows the results of the fourth experiment.

Table 10. Fourth Experiment Results

Model	Categories	Precision	Recall	F1-Score	Accuracy
CNN	CB	91%	90%	90 %%	90%
	Not CB	81%	85%	83 %	

According to Table 9, CNN achieved the highest accuracy of 92% in the third experiment, while the first experiment achieved an accuracy of 84%. It's worth noting that the effect of stop words was evident, and this may be attributed to the nature of stop words. When stop words of the Libyan dialect is used, the results were higher, while CNN achieved a lower accuracy of 77% when the list of Standard Arabic stop words is used. This underscores the importance of using proper stop words.

Unlike other available, imbalanced datasets that were either not publicly available or were largely automatically annotated, the collected dataset is the first of its kind to be publicly published and undergo a comprehensive manual annotation process by a team of specially trained auditors to identify instances of cyberbullying. This ensures a higher level of classification accuracy and allows other researchers to rely on reliable data to train and evaluate their models. Furthermore, the diversity of the data, in addition to the Libyan dialect stop word list, also affected accuracy. This was due to its diversity and the inclusion of words with the same character, an aspect not adequately addressed in the standard Arabic stop word list. This was in addition to the unique value provided by the CNN model and the architecture used to achieve cyberbullying detection results.

When comparing the experiments, it can be noted that the first experiment outperformed the other experiments, perhaps due to the suitability of the Libyan stop words to the dataset, which takes into account how much they affect the accuracy of the model. This is also supported by the third experiment, where the accuracy scores were low at 92%. On the other hand, in the third experiment for non-parallel data, the model made some errors as it incorrectly classified non-bullying cases as bullying and bullying cases as non-bullying. This indicates that the model is not well-balanced.

4. COMPARISON OF THE RESULTS

During this study, a dataset from a previous study was used [8]. They used machine learning classifiers, models, namely: support vector machine, logistic regression, naïve Bayes, K-nearest neighbor, and decision tree. They used in their study several preprocessing steps, including (Stemming, Lemmatization, and Cleaning) on the Libya Telecom dataset. The achieved results of these models are compared with the proposed model, as shown in Table 11. It is clear that this study achieved higher accuracy through the proposed model that relies on the Convolutional Neural Networks (CNN) algorithm using the same data, with the addition of preprocessing using TF-IDF Stemming, Lemmatization.

Table 11. A comparison between the obtained results and with study [16]

	Model	Precision	Recall	F1-Score	Accuracy
Libya Telecom and Technology	SVM	73%	52%	55%	50%
	NB	78%	62%	71%	63%
	DT	73%	59%	61%	59%
	KNN	78%	58%	73%	59%
	LR	76%	50%	38%	43%
	CNN (proposed Model)	91%	89%	90 %	90%

5. CONCLUSION AND FUTURE WORK

In this paper, a sentence-level dataset for the Libyan dialect was presented, which is the first publicly released and manually annotated dataset for this task. More than 7500 comments or sentences from social media were collected. A list of stopwords in the Libyan dialect from social media groups was proposed and was compared with the MSA list. The list used in the comparison was: a pre-generated list from MSA, and the Libyan dialect list, in addition to comparing it with the unbalanced data. Several preprocessing and cleaning steps were used to prepare it for classification. Then, the results of this study was compared with a previous study that used machine learning models, as they were able to obtain higher accuracy results through the deep learning model of the convolutional neural network algorithm CNN.

The results demonstrated the effectiveness of removing nuisance words using the Libyan dialect word list in experiments, as it performed better than using stopword lists in Standard Arabic alone. They also demonstrated that balanced datasets performed better than imbalanced datasets for text classification in Arabic dialects, especially the Libyan dialect. This research opens the door to further studies on this topic and can be used as a basis for future research in this field on other dialects. Future work offers significant scope for enhancing the power of the CNN model and its integration with transfer learning models through recent studies, in addition to expanding a more diverse and comprehensive database representing different Arabic dialects and exploring more advanced and diverse models to ensure their effectiveness. Focusing on diverse BERT models will open new avenues for performance and efficiency in work, and results can be improved by addressing the problem of imbalanced datasets using techniques such as SMOTE or under/oversampling.

REFERENCES

[1] W. Medhat, A. H. Yousef, and H. Korashy, "Corpora preparation and stopword list generation for Arabic data in social network," arXiv preprint, arXiv:1410.1135, 2014. [Online]. Available: https://arxiv.org/abs/1410.1135

[2] B. Haidar, M. Chamoun, and A. Serhrouchni, "Arabic cyberbullying detection: Using deep learning," in Proc. 7th Int. Conf. Comput. Commun. Eng. (ICCCE), Kuala Lumpur, Malaysia, 2018, pp. 1–6.

[3] M. Khairy et al., "Comparative performance of ensemble machine learning for Arabic cyberbullying and offensive language detection," Lang. Resour. Eval., vol. 58, no. 2, pp. 695–712, 2024, doi: 10.1007/s10579-023-09683-y.

[4] A. Habberrih and M. A. Abuzaraida, "Sentiment Analysis of Arabic Dialects: A Review Study," in *Computing and Informatics (ICOCI 2023)*, N. H. Zakaria, N. S. Mansor, H. Husni, and F. Mohammed, Eds., vol. 2001, Singapore: Springer, 2024, pp. xx–xx. doi: 10.1007/978-981-99-9589-9_11.

[5] M. Jarrar et al., "Lîsan: Yemeni, Iraqi, Libyan, and Sudanese Arabic dialect corpora with morphological annotations," arXiv preprint, arXiv:2212.06468, 2022. [Online]. Available: https://arxiv.org/abs/2212.06468

[6] S. Almutiry and M. Abdel Fattah, "Arabic cyberbullying detection using Arabic sentiment analysis," Egypt. J. Lang. Eng., vol. 8, no. 1, pp. 39–50, Apr. 2021, doi: 10.21608/ejle.2021.50240.1017.

[7] M. M. Abubaera and S. M. Jiddah, "Natural language processing and sentiment analysis for Libyan Arabic language dataset," Int. J. Adv. Res. Eng. Sci. Manag., vol. 9, no. 7, pp. 1–6, Jul. 2023.

[8] A. Alhazmi et al., "Code-mixing unveiled: Enhancing the hate speech detection in Arabic dialect tweets using machine learning models," PLoS One, vol. 19, no. 7, p. e0305657, 2024, doi: 10.1371/journal.pone.0305657.

[9] M. Alkhatib et al., "Deep learning approaches for detecting Arabic cyberbullying social media," Procedia Comput. Sci.,

vol. 244, pp. 278–286, 2024, doi: 10.1016/j.procs.2024.10.201.

[10] W. M. Yafooz, “Enhancing Arabic dialect detection on social media: A hybrid model with an attention mechanism,” Information, vol. 15, no. 6, p. 316, 2024, doi: 10.3390/info15060316.

[11] W. Medhat, A. Yousef, and H. Korashy, “Egyptian dialect stopword list generation from social network data,” Egypt. J. Lang. Eng., vol. 2, no. 1, pp. 43–55, 2015.

[12] Y. A.-A. Hazzaimeh, N. M. Norwawi, and N. A. R. Khalaf, “Generating Arabic stop-word for Hadith,” Malays. J. Sci. Health Technol., vol. 4, pp. 1–6, 2019, doi: 10.33102/mjosht.v4iSpecial%20Issue.86.

[13] T. Kanan et al., “Improving Arabic text classification using P-stemmer,” Recent Adv. Comput. Sci. Commun., vol. 15, no. 3, pp. 404–411, 2022.

[14] Z. Benmounah, A. Boulesnane, A. Fadheli, and M. Khial, “Sentiment analysis on Algerian dialect with transformers,” Appl. Sci., vol. 13, no. 20, p. 11157, 2023, doi: 10.3390/app132011157.

[15] E. M. Cherrat, H. Ouahi, and A. Bekkar, “Sentiment analysis from texts written in standard Arabic and Moroccan dialect based on deep learning approaches,” Int. J. Comput. Digit. Syst., vol. 16, no. 1, pp. 447–458, 2024.

[16] T. T. Dien, B. H. Loc, and N. Thai-Nghe, “Article classification using natural language processing and machine learning,” in Proc. Int. Conf. Adv. Comput. Appl. (ACOMP), 2019, pp. 1–6.

[17] A. Omar, M. Essgaer, and K. M. Ahmed, “Using machine learning model to predict Libyan telecom company customer satisfaction,” in Proc. Int. Conf. Eng. MIS (ICEMIS), 2022, pp. 1–5.

[18] A. Habberrih and M. A. Abuzaraida, “Sentiment analysis of Libyan middle region using machine learning with TF-IDF and N-grams,” in Proc. Int. Conf. Inf. Commun. Technol., 2023, pp. 1–10.

[19] A. Habberrih and M. A. Abuzaraida, “Sentiment analysis of Libyan dialect using machine learning with stemming and stop-words removal,” in Proc. 5th Int. Conf. Commun. Eng. Comput. Sci. (CIC-COCOS’24), 2024, pp. 1–8.

[20] A. A. Freihat et al., “Towards an optimal solution to lemmatization in Arabic,” Procedia Comput. Sci., vol. 142, pp. 132–140, 2018.

[21] I. Zeroual and A. Lakhouaja, “Arabic information retrieval: Stemming or lemmatization?,” in Proc. Int. Conf. Intell. Syst. Comput. Vis. (ISCV), 2017, pp. 1–6.

[22] A. S. Alammary, “Arabic questions classification using modified TF-IDF,” IEEE Access, vol. 9, pp. 95109–95122, 2021, doi: 10.1109/ACCESS.2021.3092755.

[23] Charfi, A., et al. (2024). "Hate speech detection with ADHAR: a multi-dialectal hate speech corpus in Arabic." Frontiers in Artificial Intelligence 7: 1391472.

[24] Hashmi, E., et al. (2024). "Enhancing multilingual hate speech detection: From language-specific insights to cross-linguistic integration." IEEE Access.

[25] Daraghmi, E. Y., et al. (2024). "From Text to Insight: An Integrated CNN-BiLSTM-GRU Model for Arabic Cyberbullying Detection." IEEE Access.

[26] Lanasri, D., et al. (2023). "Hate speech detection in algerian dialect using deep learning." arXiv preprint arXiv:2309.11611.