



HAL
open science

Deep Bidirectional LSTM Network Learning-Based Sentiment Analysis for Tunisian Dialectical Facebook Content During the Spread of the Coronavirus Pandemic

Samawel Jaballi, Manar Joundy Hazar, Salah Zrigui, Henri Nicolas, Mounir Zrigui

► To cite this version:

Samawel Jaballi, Manar Joundy Hazar, Salah Zrigui, Henri Nicolas, Mounir Zrigui. Deep Bidirectional LSTM Network Learning-Based Sentiment Analysis for Tunisian Dialectical Facebook Content During the Spread of the Coronavirus Pandemic. Ngoc Thanh Nguyen; László Gulyás; Manuel Nunez; Jan Treur; Gottfried Vossen; Adrianna Koziarkiewicz; János Botzheim. *Advances in Computational Collective Intelligence. 15th International Conference, ICCCI 2023, Budapest, Hungary, September 27–29, 2023, Proceedings, 1864*, Springer Nature Switzerland, pp.96 - 109, 2023, *Communications in Computer and Information Science*, 978-3-031-41773-3. <10.1007/978-3-031-41774-0_8>. <hal-04224265>

HAL Id: hal-04224265

<https://hal.science/hal-04224265v1>

Submitted on 2 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Copyright - All rights reserved



Deep Bidirectional LSTM Network Learning-Based Sentiment Analysis for Tunisian Dialectal Facebook Content During the Spread of the Coronavirus Pandemic

Samawel Jaballi^{1,2,3(✉)}, Manar Joundy Hazar^{1,2,4}, Salah Zrigui⁵, Henri Nicolas³,
and Mounir Zrigui^{1,2}

¹ Department of Computer Science, Faculty of Sciences of Monastir, University of Monastir,
Monastir, Tunisia

samawel.jaballi@u-bordeaux.fr

² Research Laboratory in Algebra Numbers Theory and Intelligent Systems, Monastir, Tunisia
³ LaBRI, University of Bordeaux, Talence, France

⁴ Computer Center, University of Al-Qadisiyah, Qadisiyah, Iraq

⁵ Laboratory LIG, CS 40700, 38058 Cedex Grenoble, France

Abstract. Sentiment analysis (SA) is a multidisciplinary field that aims to predict sentiment tone or attitude expressed in a text, SA using social media data has become a popular topic especially during critical events such as natural disasters, social movements and recently the spread of the Coronavirus Pandemic. Sentiments can be expressed explicitly or implicitly in text and identifying these expressions can be challenging. SA in Tunisian dialect is particularly difficult due to the complexity of the language, its morphological richness and the lack of contextual information. Recently, deep learning (DL) models have been widely adopted in the field of SA, especially in the context of Arabic SA. These models, such as Bi-directional LSTM networks (Bi-LSTM) and LSTM networks, have shown to achieve high accuracy levels in sentiment classification tasks for Arabic and dialectal text. Despite the successes of DL models in Arabic SA, there are still areas for improvement in terms of contextual information and implicit mining expressed in different real-world cases. In this paper, the authors introduce a deep Bi-LSTM network to ameliorate Tunisian SA during the spread of the Coronavirus Pandemic. The experimental results on Tunisian benchmark SA dataset demonstrate that our model achieves significant improvements over the state-of-art DL models and the baseline traditional machine learning (ML) methods. We believe that this contribution will benefit anyone working on Tunisian pandemic management or doing comparative work between Tunisian and other jurisdictions, which can provide valuable insights into how the public is responding to the crisis and help guide pandemic management decisions.

Keywords: COVID-19 · Social Media · Sentiment Analysis · Tunisian Dialect · Deep Learning · Bi-LSTM · Text Processing · Pandemic governance · Public Health

1 Introduction

Arabic is the official language of 27 countries and spoken by over 400 million people. The language has two commonly used varieties: Modern Standard Arabic (MSA) and Dialectal Arabic (DA). MSA is used in formal written and oral communication, while DA is used in informal exchanges [1]. DA is divided into two main groups: Middle Eastern (Mashriq) and North African (Maghrebi) dialects. Mashriq dialects include Egyptian, Levantine, and Gulf dialects, while Maghrebi dialects include Algerian, Tunisian, and Moroccan dialects [2].

DA is prevalent on social media and microblogging channels, but it poses challenges for thematic classification and SA [6]. These challenges include a lack of resources for these dialects, as some dialects are better researched than others. Additionally, Arabic is a morphologically rich and complex language with complex morphosyntactic conventions, irregular forms, and numerous dialectal variants with no spelling norms. This makes it difficult to learn robust general models about Arabic text without proper processing. Furthermore, there are fewer resources available for Arabic sentiment analysis, such as sentiment dictionaries and annotated corpora, compared to French or English. These challenges have led to a significant interest in Arabic sentiment analysis [3].

SA is the process of identifying and extracting specific sentiments from text using Natural Language Processing (NLP), computational linguistics, and data mining techniques. It can be applied at different levels of analysis, such as document level, sentence level, aspect level, and word level. In addition, SA in Arabic dialectal text is an important and timely topic, as Arabic is one of the most widely spoken languages in the world and is the official or native language of many countries [4].

Tunisian dialect is a subset of the Arabic dialects spoken in the Maghreb region. The goal of this work is to develop an efficient SA tool for Tunisian textual productions on Facebook networks during the spread of Coronavirus Pandemic. Despite recent progress in the study of this dialect, there is still important work to be done. However, the development of these tools faces several challenges, including a limited number and availability of resources and a lack of standard linguistic rules for this dialect [5]. This makes the task of sentiment classification more complicated as there is no agreement on semantic analysis conventions [3].

In our approach, we have proposed a deep Bi-LSTM sentiment classification model to improve the state-of-the-art performance of Tunisian dialectal SA. Our contribution can be summarized as follows:

1. We examine the benefits of our pre-processing method such as tokenization, punctuation removal, light stemming and Latin character removal.
2. Our work has used the bidirectional Deep LSTM with the ability to extract contextual information from the feature sequences of Tunisian dialectal sentences.
3. Our Bi-LSTM model significantly outperforms other DL models and the baseline traditional ML algorithms in terms of accuracy and F1-measure on Tunisian benchmark SA dataset.

The rest of this paper is organized as follows. We will first provide an overview of some related recent work on approaches and methods of Arabic and Tunisian dialectal

SA. Then, in Sect. 3, we will detail our Tunisian dialectical SA system using the deep Bi-LSTM learning model. Section 4 provides the experimental study, the results obtained and the discussion. Finally, Sect. 5 gives the conclusion of this paper and highlights future works.

2 Related Work

SA aims to classify subjective texts into two or more categories. The most obvious are negative and positive. We refer to this problem as Binary Sentiment Analysis. Some works contain a third category for neutral text. We refer to this problem as Ternary Sentiment Analysis. A final option is to consider sentiment based on a rank score or rating system, such as the 5-star rating system. This is known as Multi-Way Sentiment Analysis [6]. Thus, ML, lexicon-based, hybrid and DL approaches are among the existing research on Arabic and Tunisian dialectical SA approaches.

2.1 Machine Learning Approach

ML is the most applied approach in SA, where algorithms are trained on labeled data to predict the sentiment of a given text. Shoukry et al. [7] performed SA of tweets written in Egyptian dialects and MSA. They have collected 1,000 tweets, consisting of 500 positive and 500 negative tweets, and they have used standard n-gram features. Then, they have experimented with two classifiers, Support Vector Machines (SVM) and Naive Bayes (NB), to classify the sentiment of the tweets.

In [8], authors have created a SA method for both informal Arabic and MSA to assess the sentiment in social media. They used the K-Nearest Neighbors (KNN) algorithm as a classifier to classify 1,080 Arabic reviews from social media and news websites. In [9], authors have used two ML algorithms, SVM and NB to perform SA on Arabic reviews and comments from the Yahoo Maktoob website. Abdul-Mageed et al. [10] have created a supervised system for Arabic social media named SAMAR. The system was built using an SVM classifier.

In [11], authors have presented the Tunisian SA Corpus (TSAC¹), which focuses on the Tunisian dialect. They have collected 17,000 user comments from the official Facebook accounts of Tunisian radio and television stations. To test the performance of their collected corpus, they have employed three ML algorithms, Multilayer Perceptron, Multinomial Naive Bayes and SVM.

Salamah et al. [12] have collected 340,000 tweets in the Kuwaiti dialect. To measure the performance of the suggested corpus, they have used three ML algorithms, Random Forests (RF), SVM J48, and Decision Trees (DT).

2.2 Lexicon-Based Approach

The lexicon-based approach also known as knowledge-based approach is another commonly used method for Arabic sentiment analysis, especially when there is limited

¹ <https://github.com/fbougares/TSAC>.

labeled data available. It involves using pre-existing sentiment dictionaries containing words and their associated sentiment scores to identify and classify the sentiment of a given text. The sentiment dictionaries used in this approach are created manually by annotating words with sentiment polarity (positive, negative, neutral) or assigning them a sentiment orientation score (score between 0 and 5). The overall sentiment of a text is then computed by aggregating the sentiment scores of the individual words in the text [13]. This method can be fast and effective, but its accuracy may be limited by the quality and completeness of the sentiment dictionary used.

In [14], authors have presented a system consisting of two parts. The first part is a free online game that aims to build a lexicon of positive and negative words, while the second part is a SA system that classifies reviews based on their sentiment, using the lexicon created in the first part. This approach leverages the knowledge of users who participate in the online game to build the sentiment dictionary, which can be used for SA in the second part of the system. Al-Ayyoub et al. [15] have created a lexicon of 120,000-word Arabic linguistic terms by collecting a large number of articles from the Arabic news website. This lexicon includes Arabic linguistic features, which can be used in SA to classify the sentiment of text written in Arabic, as it provides a comprehensive list of words and their sentiment scores.

In [16], authors have proposed to investigate the impact of pre-processing techniques on SA of Tunisian dialect. They have compared the performance of two SA models, a lexicon-based model and a supervised ML-based model (SVM and NB). The results showed that including pre-processing methods such as emoji handling, negation tagging, and stemming improved the performance of the SA of Tunisian dialect.

In [17], authors have proposed a subjectivity SA system for Egyptian tweets and have developed a new Arabic lexicon by combining two existing standard Arabic lexicons (ArabSenti and MPQA) with two Egyptian Arabic lexicons. The new lexicon consisted of 900 tweets (300 positive, 300 negative, and 300 neutral) and the results showed that their lexicon-based model outperformed with an accuracy of 87%. This study highlights the importance of considering dialect-specific lexicons in SA and the potential improvement it can bring to the performance of SA models.

2.3 Hybrid Approach

The hybrid approach merges the strengths of both ML and lexicon-based approaches. Briefly, it leverages the high accuracy of ML and the stability of the lexicon-based approach to improve overall performance. Authors in [18] have proposed a hybrid SA method that combined lexicon-based and ML approaches. They used the SentiStrength English lexicon for translation and applied Maximum Entropy (ME) and KNN for classification. In [19], authors have developed a semantic model called ATSA (Arabic Twitter Sentiment Analysis) that uses supervised ML algorithms (NB, SVM and semantic analysis) and a vocabulary built from freely available dictionaries like Arabic WordNet. The performance of the model improved compared to the basic bag-of-words representation with an increase of 5.78% for the NB classifier and 4.48% for the SVM classifier.

Soliman et al. [20] have tested and evaluated different configurations to develop an Arabic SA tool. They have conducted three experiments. The first experiment used the SVM-based approach, the second experiment used only the lexicon-based approach with

SSWIL lexicon, and the third experiment combined both the SSWIL lexicon and the SVM-based approach. In [21], authors have tackled the SA of dialectical Arabic words and MSA by using the SVM and NB classifiers. To handle dialectical words, they have employed a dialect vocabulary that maps dialectical words to their corresponding MSA words.

2.4 Deep Learning Approach

DL models are based on neural network architectures to model high-level abstractions in data and to automatically extract features and perform complex tasks such as SA [22]. These models often outperform traditional ML methods in terms of accuracy and can handle noisy and unstructured data, making them well suited for SA tasks. Only a few studies have explored DL models in Arabic dialect text and Tunisian dialect.

Masmoudi et al. [23] have created a corpus of 43K comments in Tunisian dialect from Tunisian supermarkets' Facebook pages and have tested its performance using CNN, LSTM, and Bi-LSTM models. The best results were achieved by LSTM and Bi-LSTM, with an F-Measure of 87%. In [24], Jerbi et al. have proposed a DL approach for SA of code-switched Tunisian Dialect. They have evaluated the performance of four variants of RNNs: LSTM, Bi-LSTM, deep LSTM, and deep Bi-LSTM. The results, based on the TSAC corpus [11], showed that a deep LSTM model achieved the highest accuracy of 90% compared to other DL models.

Authors in [25] have introduced the TUNIZI² corpus as a SA dataset for Tunisian Arabizi³. The corpus consists of 17K comments collected from YouTube, written using Latin characters and numerals. The dataset is balanced with 47% positive and 53% negative comments and has been preprocessed for analytical studies and NLP tasks. In [26], Bsir et al. have explored the Bi-LSTM learning model for Arabic SA with the ability of extracting the semantic information to predict the sentiment of Arabic text. Experiments showed that their proposed model achieves significant improvements in the Accuracy and F1-measure results over the existing models and it can handle both forward and backward dependencies from feature sequences.

3 Proposed Methodology

Basically, our approach consists of three major phases: Firstly, the data exploration phase, where we explore an online corpus of Tunisian Facebook comments during the spread of Coronavirus Pandemic. During this phase, we thoroughly analyze the data to gain insights and understanding of the content. Secondly, the data pre-processing phase, where we apply various techniques to preprocess the collected data and make it suitable for further analysis. This involves tasks such as cleaning the text, removing irrelevant information, handling dialectical words, and performing other necessary transformations. Finally, the model training phase where we train and test our deep Bi-LSTM model and evaluate it as compared to several ML classifiers (e.g., SVM, DT, NB, etc.) and DL models (e.g., LSTM, CNN, etc.)

² Arabizi: Arabizi refers to Arabic written using the Roman script.

³ <https://github.com/chaymafourati/TUNIZI-Sentiment-Analysis-Tunisian-Arabizi-Dataset>.

3.1 Data Exploration

The CTSA⁴ dataset (Corona Tunisian Sentiment Analysis) constitute a good representative sample of the Tunisian Facebook comments and posts during the COVID-19 pandemic. The dataset covers the prevalent topics discussed on Facebook during the critical period from 28th June to 27th August 2021, when the Pandemic started to spread in Tunisia. We have analyzed the textual content of Tunisian Facebook posts; in particular, we have identified the most-frequent words, hashtags, posts and comments posted in this period and tracking their frequency over time, as the virus started to disperse in many Tunisian governorates. Figure 1 shows the time series (over days) of the top 10 most frequent words.

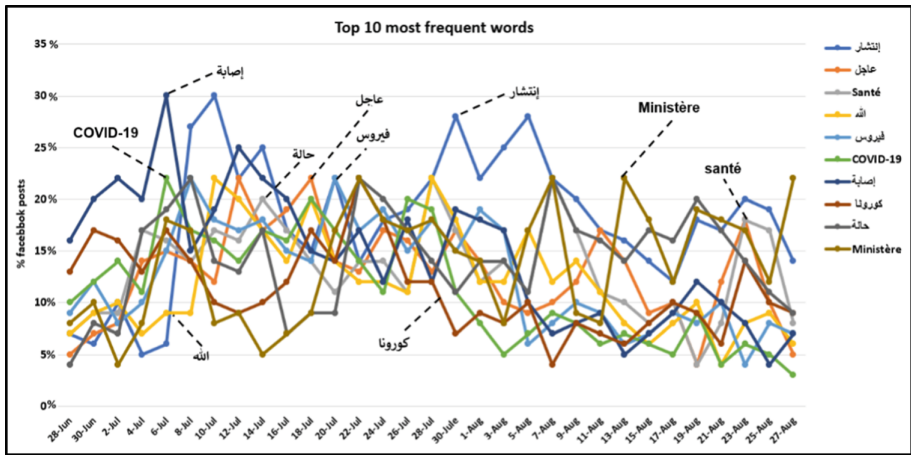


Fig. 1. Time series of the frequency (in percentage of Facebook posts) of top 10 most frequent word for the period of 28th June to 27th August 2021.

The 10 most-frequent words shown in Fig. 1 appear to belong to two distinct categories: those related to COVID-19 (such as “santé”, “COVID-19”) and those related to prayers and supplications (such as “الله”, “God”). The frequency of the word “الله” rose early on when news about the virus started to spread, as it may reflect discussions about whether the pandemic is a punishment from God. The frequency of the word declines over time but rises again as the virus starts to spread more widely in Tunisia. Thus, we have collected more than 40,000 posts and comments from Facebook. We obtained 9,396 Tunisian French code-mixed posts after cleaning the corpus. Table 1. shows the distribution of Emotion tags in the code-mixed corpus. We used hashtags related to the Covid-19 pandemic, recent sanitary trends and words which depict emotions in Tunisian dialect like (“فرحان”, “far7an” for happy, etc.).

Afterwards, we have utilized language identification for each word to have a better understanding of the corpus, using the tool⁵ from the study [27]. We have shown the

⁴ <https://www.kaggle.com/naim99/tsnaimmhedhbiv2>.

⁵ <https://github.com/irshadbhat/litcm>.

distribution of words present in the corpus between Tunisian Dialect (Arabic and Arabizi) and Latin languages in Table 2., which provides insight into the code-mixing nature of the data.

Table 1. Distribution of Emotion tags in the code-mixed corpus.

Emotion	Related emojis	Sentences
Happy	😊	725
Sad	😞	1457
Angry	😡	1118
Disgust	🤢	349
Fear	😱	937
Surprise	😲	83
Total	-- --	4669

Table 2. Distribution of total and unique words in the code-mixed corpus.

Language	Word Count	Unique Words
Tunisian Dialect	17631	8517
Latin	11309	6031
Total	28940	14548

3.2 Data Pre-processing

The collected corpus must be preprocessed for use. Figure 2 depicts the data pre-processing pipeline.

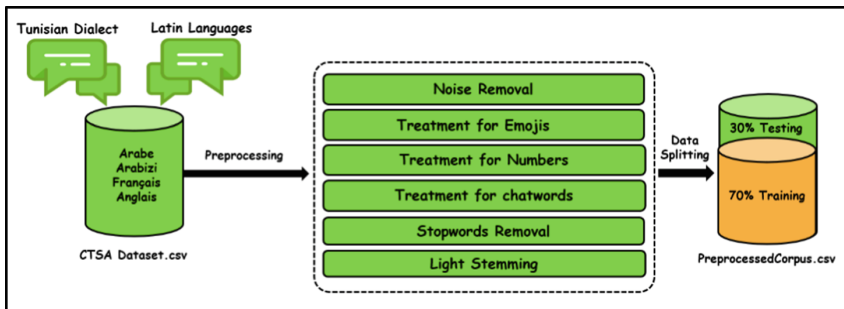


Fig. 2. The pipeline of the data pre-processing phase.

Noise Removal

As the corpus's comments are web scrapped, chances are they contain some URLs, HTML tags, and links. Since these tags are useless, it is better to remove them for effective classification results. Thus, Noise removal involves eliminating special characters, mentions, hashtags, and multiple whitespaces.

Treatment for Emojis & Numbers

Tunisian Facebook users frequently use emoji and emoticons instead of words to express emotions. To preserve this information, a small emoji-to-word converter has been developed. As for our treatment of numbers, two steps are taken: The first step is to remove numbers that are not part of words, as numbers do not carry sentimental information in SA. The second step is to replace numbers concatenated with letters with substitute letters (e.g., 5 to kh and 7 to h, as seen in the case of Arabizi words.)

Treatment for Chat Words and Stopwords.

The treatment for chat words involves removing character redundancy while preserving meaning (e.g., “خطررررر” to “خطر”, “Urgenceeeee” to “Urgence”). Also, common chat words will be converted back to their original forms (e.g., “Slm” to “Salem”). We have used the Unigram model to identify the frequent chat words. Recall that in our contribution, we are interested in the SA of Tunisian French code-mixed posts and since publicly available stopwords extraction tools work only with monolingual corpora such as the example of NLTK⁶ library, that's why we have developed a multilingual stopwords extractor. To do this, we have utilized two N-gram models (Unigrams and Bigrams) for stopwords extraction and then manually reviewed the results to eliminate words that do not carry significant meaning. As an illustration, some examples of stopwords in our research are (“cette”, “بعد”, “eni”, “انت”, “inty”, “عن”, etc.)

Light Stemming

The light stemming approach involves removing a small set of suffixes and/or prefixes to find the root of a word. Our contribution uses the ISRI stemmer from the Information Science Research Institute (ISRI). It operates similarly to the Khoja stemmer, but without utilizing a root dictionary. If a word cannot be rooted, the ISRI stemmer normalizes it by removing specific end patterns and determinants instead of leaving it unchanged. It also recognizes a set of diacritical marks and affix classes, as demonstrated by the example (“استمرض - مرض”, “إنتشار- نشر”).

3.3 Model Training

After performing the processing of the Facebook posts, a vocabulary size was defined to convert the words into vectors. The sentiment labels (positive, negative and neutral) were encoded as well, and padding step was done to bring all the vectors to a specific length. The Bi-LSTM network architecture included the following layers: the Embedding layer to convert incoming words into vectors, the Bi-LSTM layer for processing the word vectors, the Dropout layers to avoid overfitting, the Conv1D layer to summarize the

⁶ <https://www.nltk.org>.

vector along one dimension, the Global MaxPooling1D layer to find the maximum value in each feature map and finally the last layer as a fully connected layer with a Softmax activation function and three neurons for the output predicted label (positive, negative and neutral) producing a one-hot encoded result. Thus, Bi-LSTM combines the two LSTM hidden layers to the output layer, improving the learning of long-term dependencies and better model performance. Previous studies [22] have demonstrated that bidirectional networks are more effective than standard networks in various areas, including SA. The structure of a Bi-LSTM layer, made up of a forward and backward LSTM, is depicted in Fig. 3. As the forward LSTM layer output sequence \vec{h} , is obtained in a common way as the unidirectional one, the backward LSTM layer output sequence \overleftarrow{h} , is measured using the reversed inputs from time $t - 1$ to $t - n$. Then, these output sequences fed to σ function to combine them into an output vector y_t [19]. Similar to the LSTM layer, the final output of a Bi-LSTM layer can be represented by a vector, $Y_t = [y_{t-n}, \dots, y_{t-1}]$, in which the last element, y_{t-1} , is the estimated analysis label for the next iteration [28].

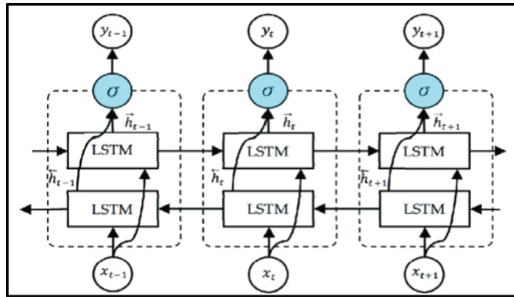


Fig. 3. The workflow of Bidirectional LSTM layer

4 Experiments and Results

4.1 Baseline and Evaluation Metrics

We have randomly divided our collected dataset into two corpora, 70% of the dataset for training and 30% for testing. The number of training iterations (epochs) is fixed at 10 for all experiments. To prevent overfitting, Dropout regularization is applied with a dropout rate of 40%. The goal of the experiments is to find the optimal parameters for the model. Notice that only the parameters that yield the best results are reported. To evaluate the performance of our proposed SA system, we have conducted experiments using various classification algorithms. The first set of experiments is designed to measure the performance of traditional ML methods such as NB, SVM, RF and Logistic Regression (LR) classifiers. These classifiers are applied to different combinations of preprocessed data. For feature extraction, we have used three embedding models: Bag Of Words (BOW), Term Frequency-Inverse Document Frequency (TF-IDF), and GloVe (Global Vectors for

Word Representation). Furthermore, to evaluate the performance of our system as well as the other classifiers, we have used the most widely popular measures retrieval namely, Accuracy, Precision and Recall Their corresponding formulas are recalled respectively in 1, 2 and 3, where tp , tn , fp and fn represent True Positive⁷, True Negative⁸, False Positive⁹ and False Negative¹⁰ respectively.

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \quad (1)$$

$$Recall = \frac{tp}{tp + fp} \quad (2)$$

$$F_1measure = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3)$$

4.2 Experimental Results

In the experiment, DL models were performed using Keras¹¹ library, while ML models were performed using Scikit-Learn¹² library. Table 3. displays the results of the ML classifiers after applying the pre-processing method. Table 4. presents the results of the second set of experiments, which were done using different configurations of DL models during the training phase.

4.3 Discussion

Our study aims to present a new method for Tunisian Dialectical SA to surpass the limitations of feed-forward models by considering multilingual contextual features and Tunisian French code-mixed information in both directions (forward and backward) of the Tunisian Dialectical written comments. From the results showed in Sect. 4.2, we can highlight that our proposed method yielded the highest sentiment prediction accuracy results. According to Tables 4. and 5., deep Bi-LSTM and Bi-LSTM outperformed other DL models (CNN and LSTM) on the benchmark CTSA dataset. This is due to the fact that Bi-LSTM networks can better capture context in the text by combining both forward and backward hidden layers. This allows it to access both preceding and succeeding contextual features, resulting in richer semantic information and the ability to make full use of contextual information than the Unidirectional LSTM networks. Additionally, according to Table 3., in which we have performed five ML algorithms applied with three embeddings models (BOW, TF-IDF, GloVe), we can clearly see that DL algorithms typically have better accuracy than traditional ML methods and as noted in [29], the SVM becomes inefficient and costly when dealing with large training data as it is not capable of

⁷ The number of correctly classified positive comments.

⁸ The number of correctly classified negative comments.

⁹ The number of incorrectly classified positive comments.

¹⁰ The number of incorrectly classified negative comments.

¹¹ <https://keras.io>.

¹² <https://scikit-learn.org>.

Table 3. Experimental results obtained with machine learning classifiers on CTSA Corpus

Label prediction after pre-processing						
Classifier + Embedding	Metrics	SVM	RF	LR	NB	DT
Classifier + BOW	Recall F1- Measure Accuracy	65.3264.53 71,14	69.16 65.78 70,09	72.1769.6774,91	73.82 72.23 74,57	71.6570.42 71,78
Classifier + TF-IDF	Recall F1- Measure Accuracy	65.7462.58 69,94	70.0969.60 71,49	72.24 71.89 75,12	71,13 72.12 75,84	68.8664.97 69,29
Classifier + GloVe	Recall F1- Measure Accuracy	68.8965.47 70,32	71.9669.64 72,82	69.83 70.79 74,43	72.19 71.51 73,43	70.69 68.21 71,03

Table 4. Experimental results obtained with Deep Learning models on CTSA Corpus

Model Configuration	Recall	F1 score	Accuracy
CNN (fully connected layer size = 100)	72.81	72.52	74.09
LSTM (dropout rate = 0.4)	69.54	70.89	72.89
RNTN	70.49	72.29	74.97
RNN	72.82	73.82	75.12
CNN + LSTM	75.42	76.08	77.62
Combined LSTM	75.12	74.17	76.79
Bi-LSTM	76.32	77.58	78.10

Table 5. Comparison of accuracy results of different DL models after applying light stemming

Classifier + Embedding	Metrics	Deep Bi-LSTM	LSTM	CNN
Without Stemming	Recall F1- Measure Accuracy	75.79 74.24 77.89	69.05 75.78 75.92	74.17 75.42 74,40
Arabic Light Stemming (ISRI Stemmer)	Recall F1- Measure Accuracy	79.25 77.93 79,80	77.57 78.15 77,49	78.21 77.04 78,61

handling such big size data. Moreover, when the training data is unbalanced and noisy, it can negatively impact the SVM performance due to its high computational cost and low generalization error [30]. For the RF algorithm, the complexity increases as the number of trees in the forest and the number of training samples increase, because more trees and more training samples increases the model's complexity, leading to longer processing times and higher computational resource requirements. Furthermore, our model was improved by adding morphological features, including stems, to overcome the problem of limited vocabulary, ambiguity issue and the lexical sparsity.

According to results in Table 5., experiments clearly indicate that using light stemming leads to an average improvement of 2.56% across the three tested DL models. According to similar studies [31, 32, 33], including morphological features, such as stems, improved performance quality on data containing a mixture of Dialectal Arabic and MSA. This approach offers several benefits, including improvement in sentiment prediction accuracy through the implementation of various Arabic pre-processing techniques like punctuation removal, digits, tokenization, Latin characters removal, light stemming and the ability to consider the code-mixed contextual information by considering both forward and backward dependencies.

5 Conclusion and Perspectives

We tackled in this paper, the SA problem for Tunisian textual productions on Facebook during the spread of the Coronavirus pandemic. To address this challenge, we have proposed an efficient deep Bi-LSTM learning model to extract multilingual and code-mixed contextual features, improving in richer semantic information to overall quality of the sentiment prediction. Experiments on Tunisian benchmark dataset (TSAC) showed the effectiveness of the deep Bi-LSTM model, outperforming state-of-the-art algorithms in terms of accuracy and F1-measure. We believe that our model will contribute to future exploration in the field of SA and text mining especially during critical events.

Our future works will focus on comparing the effect of different recent contextualized word embeddings models such as BERT, ELMO, ULMFiT, XLNet and GPT-3 on their previously presented DL model for Arabic dialectal SA. Additionally, we plan to study the media press and news articles related to COVID-19 in order to identify bias and track the spread of misinformation and fake news.

Acknowledgment. The authors would like to express their deep gratitude towards the members of the Research Laboratory in Algebra, Numbers theory and Intelligent Systems (RLANTIS) for their unwavering support and contribution in the realization of this paper.

References

1. Guellil, I., Saâdane, H., Azouaou, F., Gueni, B., Nouvel, D.: Arabic natural language processing: an overview. *J. King Saud Univ.-Comp. Inf. Sci.* **33**(5), 497–507 (2021)
2. Merhben, L., Zouaghi, A., Zrigui, M.: Lexical disambiguation of Arabic language: an experimental study. *Polibits* (46), 49–54 (2012)

3. Sghaier, M.A., Zrigui, M.: Sentiment analysis for Arabic e-commerce websites. In: 2016 International Conference on Engineering & MIS (ICEMIS), pp. 1–7. IEEE (2016)
4. Mahmoud, A., Zrigui, M.: Deep neural network models for paraphrased text classification in the Arabic language. In: Métais, E., Meziane, F., Vadera, S., Sugumaran, V., Saraee, M. (eds.) *Natural Language Processing and Information Systems*. LNCS, vol. 11608, pp. 3–16. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23281-8_1
5. Merhbene, L., Zouaghi, A., & Zrigui, M.: A semi-supervised method for Arabic word sense disambiguation using a weighted directed graph. In: *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pp. 1027–1031, Oct (2013)
6. Batita, M.A., Zrigui, M.: Derivational relations in Arabic wordnet. In: *Proceedings of the 9th Global WordNet Conference*, pp. 136–144, Jan (2018)
7. Shoukry, A., Rafea, A.: Sentence-level Arabic sentiment analysis. In: 2012 International Conference on Collaboration Technologies and Systems (CTS), pp. 546–550. IEEE (2012)
8. Maraoui, M., Zrigui, M., Antoniadis, G.: Use of NLP tools in CALL system for Arabic. *Int. J. Comp. Process. Lang.* **24**(02), 153–165 (2012)
9. Abdulla, N.A., Al-Ayyoub, M., Al-Kabi, M.N.: An extended analytical study of Arabic sentiments. *Int. J. Big Data Intell.* **1**(1–2), 103–113 (2014)
10. Abdul-Mageed, M., Diab, M., Kübler, S.: SAMAR: Subjectivity and sentiment analysis for Arabic social media. *Comp. Speech Lang.* **28**(1), 20–37 (2014)
11. Medhaffar, S., Bougares, F., Esteve, Y., Hadrach-Belguith, L.: Sentiment analysis of tunisian dialects: linguistic resources and experiments. In: *Proceedings of the third Arabic natural language processing workshop (WANLP)*, Valencia, Spain, pp. 55–61 (2017)
12. Salamah, J.B., Elkhilfi, A.: Microblogging opinion mining approach for Kuwaiti Dialect. In: *Proceedings of the International Conference on Computing Technology and Information Management*, pp. 388–396 (2014)
13. Ayadi, R., Maraoui, M., Zrigui, M.: Intertextual distance for Arabic texts classification. In: 2009 International Conference for Internet Technology and Secured Transactions, (ICITST), pp. 1–6 (2009)
14. Albraheem, L., Al-Khalifa, H. S.: Exploring the problems of sentiment analysis in informal Arabic. In: *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services* (2012)
15. Al-Ayyoub, M., Essa, S.B., Alsmadi, I.: Lexicon-based sentiment analysis of Arabic tweets. *Int. J. Social Netw. Mining* **2**(2), 101–114 (2014)
16. Mulki, H., Haddad, H., Ali, C.B., Babaoğlu, I.: Tunisian dialect sentiment analysis: a natural language processing-based approach. *Comput. y Sist.* **22**(4), 1223–1232 (2018)
17. Stoyanov, V., Cardie, C., Wiebe, J.: Multi-perspective question answering using the opqa corpus. In: *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pp. 923–930, Oct (2005)
18. Brahimi, B., Touahria, M., Tari, A.: Improving Arabic sentiment classification using a combined approach. *Comput. y Sist.* **24**(4), 1403–1414 (2020)
19. Haffar, N., Hkiri, E., Zrigui, M.: TimeML annotation of events and temporal expressions in Arabic texts. In: *Computational Collective Intelligence: 11th International Conference, ICCCI 2019, Hendaye, France, September 4–6, 2019, Proceedings, Part I* 11, pp. 207–218. Springer International Publishing, Cham (2019)
20. Soliman, T.H., Elmasry, M.A., Hedar, A., Doss, M.M.: Sentiment analysis of Arabic slang comments on facebook. *Int. J. Comp. Technol.* **12**(5), 3470–3478 (2014)
21. Zrigui, S., Ayadi, R., Zouaghi, A., Zrigui, S.: ISAO: an intelligent system of opinions analysis. *Res. Comput. Sci.*, **110**, 21–30 (2016)
22. Jaballi, S., Zrigui, S., Sghaier, M.A., Berchech, D., Zrigui, M.: Sentiment analysis of Tunisian users on social networks: overcoming the challenge of multilingual comments in the

- Tunisian dialect. In: Computational Collective Intelligence: 14th International Conference, ICCCI 2022, Hammamet, Tunisia, 28–30 Sept 2022, Proceedings, pp. 176–192. : Springer International Publishing, Cham (2022)
23. Masmoudi, A., Hamdi, J., Belguith, L.: Deep learning for sentiment analysis of Tunisian dialect. *Comp. y Sist.* **25**(1), 129–148 (2021)
 24. Jerbi, M.A., Achour, H., Souissi, E.: Sentiment analysis of code-switched Tunisian dialect: exploring RNN-based techniques. In: Smaili, K. (ed.) ICALP 2019. CCIS, vol. 1108, pp. 122–131. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32959-4_9
 25. Fourati, C., Messaoudi, A., Haddad, H.: TUNIZI: a Tunisian Arabizi sentiment analysis Dataset. arXiv preprint [arXiv:2004.14303](https://arxiv.org/abs/2004.14303) (2020)
 26. Haffar, N., Hkiri, E., Zrigui, M.: Using bidirectional LSTM and shortest dependency path for classifying Arabic temporal relations. *KES* **2020**, 370–379 (2020)
 27. Bhat, I.A, Mujadia, V., Tammewar, A.: IIT-H system submission for FIRE2014 shared task on transliterated search. In: Proceedings of the Forum for Information Retrieval Evaluation on – FIRE ‘14, New York, New York, USA, pp. 48–53 (2015)
 28. Sghaier, M.A., Zrigui, M.: Rule-based machine translation from Tunisian dialect to modern standard Arabic. *Proc. Comp. Sci.* **176**, 310–319 (2020)
 29. Abd Allah, M.A.H., Haffar, N., Zrigui, M.: Contribution to the methods of indexing Arabic textual documents to improve the performance of IRS. In: 2022 International Conference on Innovations in Intelligent Systems and Applications (INISTA), pp. 1–6. IEEE (2022)
 30. Jabnoun, J., Haffar, N., Zrigui, A., Nsir, S., Nicolas, H., Trigui, A.: An image retrieval system using deep learning to extract high-level features. In: Advances in Computational Collective Intelligence: 14th International Conference, ICCCI 2022, Hammamet, Tunisia, September 28–30, 2022, Proceedings, pp. 167–179. Springer International Publishing, Cham (2022)
 31. Slimi, A., Nicolas, H., Zrigui, M.: Hybrid time distributed CNN-transformer for speech emotion recognition. In: Proceedings of the 17th International Conference on Software Technologies ICSOFT, Lisbon, Portugal, pp. 11–13 (2022)
 32. Bellagha, M.L., Zrigui, M.: Speaker Naming in Arabic TV programs. *Int. Arab J. Inf. Technol.* **19**(6), 843–853 (2022)
 33. Trigui, A., Terbeh, N., Maraoui, M., Zrigui, M.: Statistical approach for spontaneous Arabic speech understanding based on stochastic speech recognition module. *Res. Comput. Sci.* **117**, 143–151 (2016)