ISSN 2007-9737

Pre-trained Model Sentiment Analysis of Tunisian Telecommunications Operators' Comments on Social Media

Abir Masmoudi*, Nour Aridhi, Lamia Hadrich Belguith

University of Sfax,
MIRACL Laboratory,
Tunisia

{masmoudiabir, nouraridhi2905}@gmail.com, l.belguith@fsegs.rnu.tn

Abstract. Sentiment analysis (SA) has emerged as a crucial computational method for extracting subjective information from text, facilitating organizations to transform unstructured opinions through actionable insights that drive strategic decision-making across domains covering from business intelligence to public policy formation [46]. Pre-training models for SA have gained significant attention for improving opinion extraction from text. In recent years, social media has become a crucial platform for customer engagement, with SA playing a key role in maintaining client loyalty. Extracting sentiments from comments and reviews is particularly challenging for under-resourced languages like the Tunisian Dialect (TD), which is written in both Arabizi and Arabic scripts. Despite advancements in SA, processing TD remains complex. In this study, BERT and CNN-Bidirectional LSTM models are employed to perform SA on unstructured data collected from Facebook. The dataset, TUNisian TElecom Sentiment Analysis (TUNTESA), consists of 27,080 Arabizi and 17,816 Arabic comments sourced from official telecommunications operators' Facebook pages. The comments are labeled as positive, negative, or neutral. The results demonstrate high accuracy (Acc), with the BERT Arabic model achieving 0.99 and the BERT Arabizi model reaching 0.94, outperforming existing studies. These findings highlight the practical applications of SA for businesses leveraging social media interactions. By effectively analyzing sentiments, telecom operators can enhance customer satisfaction, manage relationships, and extract valuable feedback, ultimately maintaining a competitive edge.

Keywords. Sentiment analysis, social media, telecom operators, Tunisian dialect.

1 Introduction

With the widespread adoption of the Internet and the transformative impact of social networks in recent decades, an increasing number of individuals are deeply engaged in social networks [19]. This evolution has significantly facilitated global connectivity, enabling people around the world to express themselves and share a wide array of content, including posts, articles, links, videos, and photos, in an increasingly accessible manner.

In the Arab region, there has been a marked growth in the adoption of social media platforms. The number of Internet users amounts to more than 226 million and those who access Facebook exceed 141 million, and Arabic is considered the fourth most popular language on the Web [42]. The literature speaks of 22 Arabic dialects (AD) that represent the official Arabic speaking countries [20], and the study focuses particularly on Tunisia, where internet penetration is 79%, comprising approximately 10 million users. This statistic places Tunisia third in terms of internet penetration among African nations [47]. Since April 2024, Facebook has remained the dominant social media platform in Tunisia, compared to Instagram and Twitter [21], a trend that has persisted since January 2024 [47]. This dynamic trend underscores the profound influence of social networks, especially Facebook, in reshaping communication patterns and cultural exchanges. Foster unprecedented levels of digital engagement and interaction, both locally and across borders.

SA is a critical subfield of natural language processing (NLP) that focuses on categorizing text into three key sentiments: positive, negative, and neutral [44] and it is based on a multiphase process involving data retrieval, data extraction, data pre-processing. SA has proven to be a valuable tool for gauging public opinion in various areas of use such as hospitality [8][24], healthcare [9][10], stock market news [7], and educational institutions [11]. It has been used successfully in the prediction of the financial market, health issues such as the COVID-19 pandemic [5], customer analytics, commercial valuation evaluation, brand marketing, politics, crime prediction, and emergency management [40].

However, most SA research has focused on Indo-European languages, leaving under-represented languages like Arabic significantly less explored. Despite the increasing presence of Arabic content on social networks and ongoing advances in Arabic NLP tools, research on the Arabic language faces challenges due to the scarcity of annotated corpora [23]. For example, Arabic emotional speech datasets constituted only 4.76 % of global SER resources between 2000 and 2021 [25].

These difficulties largely stem from the inherent complexities of the Arabic language, which exists in three main variants: Classical Arabic, used in the Quran; Modern Standard Arabic (MSA), the formal version of the language; and Dialectal Arabic , which encompasses various regional dialects [44]. Among these dialects, the Tunisian Dialect (TD) is particularly complex and unique, spoken by around 12 million people and incorporating linguistic variations between regions. This dialect includes multilingualism, morphological, syntactical, and lexical differences, as well as a variety of word types [30, 37].

In this context, the study introduces a novel sentiment analysis methodology that synchronously processes Tunisian dialectal content through both Arabizi (Latin script) and Arabic script datasets, focusing particularly on consumer attitudes toward Tunisia's three major telecommunications operators: Orange Tunisia, Ooredoo, and Tunisie-Telecom. This dual-corpus approach addresses the special code-switching patterns universal in Tunisian digital communication, capturing sentiment expressions that would be missed by single-script analysis methods. By evolving analytical frameworks tailored to the linguistic complications of Tunisian dialect across different writing systems, new methodological standards are established for sentiment analysis in dialectal Arabic contexts with multiple orthographic representations.

To achieve this, comments in the Tunisian Dialect (TD) were gathered from official Facebook pages, with the aim of better capturing the linguistic characteristics of TD and improving sentiment analysis for this under-represented language. This study addresses a significant gap in the literature, as most existing research has been dedicated to Indo-European languages, and emphasizes the necessity of developing robust NLP tools tailored for Arabic and its dialects.

The major contributions of this research are summarized as follows:

— A comprehensive corpus of Tunisian Dialect in both Arabizi and Arabic scripts was compiled from Facebook comments on telecommunications operatorsâ€™ pages. This focused dataset provides a representative example of everyday language usage among Tunisian social media users.

— An in-depth study was conducted on Arabizi and Arabic corpora related to telecommunications operators, highlighting the specific challenges and linguistic features of the Tunisian dialect in the context of sentiment analysis

— A meticulous methodology was established for data selection, preprocessing, and annotation. The process involved cleaning the collected content to ensure consistency and quality. Each comment was carefully labeled, resulting in a high-quality annotated dataset suitable for model training and evaluation.

— A CNN-BiLSTM and a fine-tuned BERT models are proposed and implemented to develop sentiment analysis using comments from telecommunications operators. These models were constructed to improve the accuracy of SA for the Tunisian dialect.

— Extensive experimentation was conducted to evaluate the performance of the proposed models against current state-of-the-art methods. The results demonstrate that these models significantly outperform previous approaches for sentiment analysis in the Tunisian Dialect, achieving higher accuracy, precision, and recall. The comparative analysis highlights the effectiveness of the proposed methodology in addressing the unique linguistic challenges posed by the Tunisian dialect.

The paper is structured as follows: Section 2 reviews related work, while Section 3 describes the key characteristics of the TD. Section 4 details the adopted methodology by highlighting data construction and data processing. Section 5 outlines the experimental setup for CNN-BiLSTM and Bert models. The results are represented in Section 6, followed by discussion in Section 7. Finally, Section 8 concludes the paper and explores prospective directions for future research.

2 Related Work

Sentiment encompasses people's opinions or emotions regarding entities, events, or ideas. It reflects subjective content that indicates a positive, negative, or neutral polarity within a written text. SA, or opinion mining, involves the development of automated methods to examine the opinions expressed in a text. SA can be performed at three levels: document, sentence, and aspect levels [36, 4].

At the document level, the objective is to categorize the entire document and extract a general perspective. The text can be processed distinctly by SA: Multiple opinions are present within a document, or the whole document has a single opinion. At the sentence level, the task involves analyzing each sentence to determine if it conveys a positive, negative, or neutral sentiment. At the aspect level, the focus is on identifying sentiments related to specific aspects of words. This approach helps pinpoint exactly what people like or dislike, concentrating on the features' characteristics rather than the overall sentiment of paragraphs. Extracting implicit or explicit aspects is considered central to sentiment analysis [36, 4, 48].

Recent years have seen a rapid expansion in studies focusing on SA of Arabic dialects on social networks, exploring various aspects and levels within this field. For example, [6] developed the SDCT corpus for annotation of sentiment and emotion in the Saudi dialect, annotating 32,063 tweets in positive and negative classes. Their study compared LSTM, Bi-LSTM, and SVM models, highlighting Bi-LSTM's effectiveness with an impressive F-score of 94%. [35] investigated CNN, LSTM, and RCNN models for Arabic SA using a dataset of 40,000 tweets, achieving the highest accuracy of 88% with LSTM.

Similarly, efforts to develop sentiment analysis models for TD have involved creating essential resources such as annotated corpora, lexicons, and advanced models. [33] pioneered the Tunisian Sentiment Analysis Corpus (TSAC) by gathering 17,000 comments from official Facebook pages of Tunisian radios and TV channels, all in Arabic script. Annotated by a native speaker, the corpus identified 8,215 positive and 8,845 negative comments, leading to an MLP classifier achieving an accuracy of approximately 78%. In another study [31], researchers selected data from Facebook, specifically identifying comments from five official supermarket pages, totaling 44,000 comments. They applied a BiLSTM model, which achieved an accuracy of 82%. These initiatives illustrate ongoing advancements in adapting sentiment analysis techniques tailored to the nuances of TD across diverse datasets and social media contexts.

Furthermore, TunBERT [34] was trained on a dataset that incorporated TSAC and the Tunisian Election Corpus (TEC) [43], consisting of 3,042 tweets related to Tunisian elections in 2014, encompassing both Modern Standard Arabic (MSA) and Tunisian content. TunBERT demonstrated significant performance advancements, achieving an accuracy of 96.98% on TSAC compared to 81.2% on TEC.

Other research has also explored transfer learning with transformer architectures like BERT, evaluating LSTM, CNN, and BERT models on Algerian tweets. Here, CNN and LSTM models achieved the highest accuracy at 76% and 75%, respectively, while BERT lagged behind at 68%. Additionally, [2] implemented ARBERT and MARBERT models for various Arabic text classification tasks, including SA, achieving an F-score of 71.50% with the MARBERT model. [3] focused on sentiment detection in Arabic dialects using the MARABERT model, obtaining an F1-score (F1) of 86%. [1] fine-tuned MARBERT for sentiment classification across different Arabic dialects, achieving an accuracy of 69.57%. Moreover, [18] proposes an approach to create a well-annotated corpus with positive or negative sentiment, comparing the effectiveness of machine learning (ML) and transfer learning models for constructing fine-grained sentiment analysis models. The BERT model, fine-tuned with TD data, achieved the highest accuracy of 85%.

For AraBert, it showed good performance in papers such as [41], which used it with pretrained embeddings and achieved better results with BiLSTM, reaching an accuracy of 93.97%, and BERT in trained embeddings with BiLSTM, achieving an accuracy of 94.25%. This paper tested five datasets: HARD, Khooli, AJGT, ArSAS, and ASTD, but both models excelled with the HARD dataset, which contains 106k comments from MSA, Egyptian, and Gulf dialects.

In [22], the AlgBert model was developed by using AraBert and was compared with several models like lexicon combined with RF and SVM combined with CBOW. However, it achieved a better accuracy of 92.6% with a dataset of 110K comments and tweets in various Arabic dialects, including 54k comments and tweets written in Algerian Arabic or Arabizi.

In [17], the ASA-Medium Bert was developed by combining three BERT-based models, namely Arabic Bert, AraBert, and Mbert, achieving an accuracy of 96.11%. To reach the best accuracy, Data Augmentation was applied, improving accuracy from 77% to 96% with the Arabert multiplied by two model and the AJGT dataset. However, this was not the case in [15], which integrate the pretrained arabic Bert models which are AraBERT, QARIB, ALBERT, AraELECTRA, and CAMeLBERT in SVM and CNN. Qarib gave the best accuracy of 96% with the MSDA-MAC dataset. This dataset combined two imbalanced datasets and applied under-sampling to reduce positive comments while over-sampling increased negative comments, focusing on the Moroccan dialect.

The study also compared BERT results with CNN and LSTM, concluding that BERT-based models yield superior results compared to their application with CNN or SVM. In [12], five models were developed, including three for SA, namely QST, QSR, and QSRT, and two for emotion classification, namely QE3 and QE6, all based on QARIB. QSRT achieved the best macro F1 score of 97.45% with Twitter-AB, a sentiment dataset containing 2000 tweets, while QE3 reached the best macro F1 score of 90.10% with the EATD emotion dataset, which consists of 2021 tweets written in multiple dialects, including Jordanian, Levantine, Arabic, Syrian, and Lebanese dialects. Additionally, Retrieval-Augmented Generation was integrated into generative LLM models such as Llama3-8B, Llama3-8BInstruct, Gemini-1.0-pro-latest, and Ace-gpt7B in [29] to reduce hallucinations.

This paper also tested results by keeping the neutral data and removing it, showing an accuracy of 64% in the first case and 82% in the second case with the SemEval dataset. Finally, in [16], a comparison was made between machine learning approaches such as Logistic Regression, Support Vector Machine, Decision Tree, Multinomial NaÃ¯ve Bayes, and XGBoost and deep learning (DL) approaches such as the Neural Network approach and AraBert. For the first classification, Logistic Regression achieved the best accuracy of 82% with the TF-IDF configuration, while for the second classification, AraBert was the most performant with an accuracy of 88% using the Bert configuration. [27], which focused on TD and compared Bi-LSTM with other machine learning algorithms such as SVM, DT, and NB, as well as deep learning models like LSTM and CNN. The results showed that Bi-LSTM performed the best, achieving an accuracy of 78.10% on the CTSA database.

In addition, while existing research has made meaningful strides in SA, specifically for under-resourced languages like the TD, several difficulties remain unaddressed. Recent methodologies frequently struggle with the nuances of informal expressions and dialectical modifications present in social media data. The reliance on traditional machine learning models may limit the ability to capture complex linguistic features inherent in dialectical forms. Additionally, the scalability and generalizability of existing approaches to handle large-scale unstructured data from platforms like Facebook need further enhancement. Accordingly, there is an important need for advanced models that can efficiently process and analyze sentiment from multiple linguistics sources, such as Arabizi and Arabic scripts.

3 Tunisian Dialect Characteristics

In Tunisia, while MSA serves as the official language, TD, commonly known as ”Darija” or ”Tounsi,” is the primary language for daily communication. It is widely used in informal interactions, digital platforms, and broadcast media [13]. This dialect exhibits distinct linguistic and cultural characteristics, along with unique specificities [26].

3.1 History of Tunisian Dialect

TD is deeply shaped by a rich history of linguistic influence, stemming from its integration into various civilizations, including the Phoenicians, Romans, Byzantines, Arabs, and Ottomans. Each of these cultures has left a distinct linguistic imprint. For example, French colonization introduced numerous loanwords, while the Berber (Amazigh) substratum remains significant despite its decline in everyday usage. The dialect is, in essence, a product of continuous interaction between Berber, Arabic, and several other languages such as French, Italian, Turkish, and Spanish. This linguistic fusion is evident in its vocabulary, which incorporates words from diverse origins. Illustrative examples of these influences across languages and civilizations are provided in Tables 1 and 2 [30, 37].

TD is different from MSA in phonological, lexical, morphological, and syntactic variations, it is considered as the low variety and an under-resourced language. It has neither a standard orthography nor dictionaries. It has been used in its oral form for a long time. Therefore, TD can be classified into two oral linguistic registers [14]:

— the intellectualized dialect found in the conversations of scholars and radio and television broadcasts. This level of dialect is defined by the use of a literal lexicon and borrowings but keeping all the structure of the dialect.

— the popular (familiar) dialect which conveys daily needs.

3.2 Challenges of Tunisian Dialect

Sentiment analysis and its underlying platforms empower NLP researchers to achieve significant advancements across multiple areas of the field, resulting in a positive practical impact within their research domains. Nevertheless, the difficulties faced by classification techniques in sentiment analysis, as shown in Fig 1 should not be overlooked. To address the main challenges, particular attention was given to obstacles such as code-switchingâ€”where multiple languages are alternated within the same discourse and the scarcity of available corpora. In fact, research on Arabic dialects struggles with the limited accessibility of data resources collected by researchers. Some organizations specialized in creating linguistic resources, such as the LDC (Language Data Consortium) and ELRA (European Language Resources Association), give corpora for Middle Eastern dialects like Saudi Arabic and Egyptian Arabic, which are available to the entire scientific community. However, for the Tunisian dialect, no transcribed and annotated corpus has been made available by these organizations.

Table 1. The origin and meaning of some borrowed words from other languages used in the TD

Source language	Word in source language	Word in TD	Word in English
spanish	calles	kayes	street
italian	cucina	koujina	kitchen
french	carabat ”voiture de transport public”	karaba-karahba	car
turkish	cesme	shishma	water tap
	bakrac	ba9raj	jug

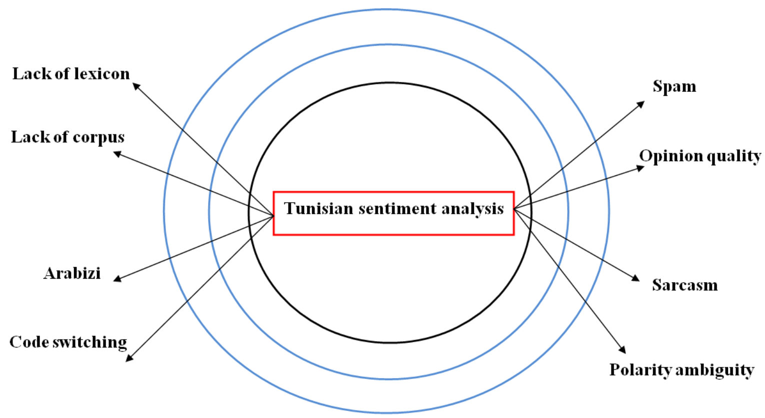


Fig. 1. Challenges of TD

Table 2. The origin and meaning of some borrowed words some civilizations used in the TD

Source civilization	Word in source	Word in TD	Word in English
malte	qaT uws	kattous	cat
berber	3alloush	3alloush	sheep
	fakroun	fakroun	turtle

3.3 Lexical Differences

Socially, it is common to distinguish two sub-dialects in the dialect of each region: the north lexical and the south lexical which adds several different words from different backgrounds and having the same meaning. It is important also be noted that at the lexicon level, dialectal Arabic includes words specific to cities in the South and other cities in the North [30, 37].

3.4 Morphological Differences

At the morphological level, Gender distinction is not uniformly marked across all Tunisian regions.

Table 3. Lexical differences

Northern Dialect	Southern Dialect	Meaning in English
شوف /$wf/	آرَا /ArA/	see
عملت /Emlt/	درت /drt/	do

Table 4. Example of conjugation of the verb عمل/دَار [to do] according to the northern and southern dialects

Pronouns	Northern dialect	Southern dialect	Translation
2MS	aamalt عملت	درت	you did
FS2	aamalt عملت	درتي	she did

For example, this feature is less prominent in the northern dialects compared to those of southern regions. The dialects of the North do not distinguish between the second person singular masculine and feminine [32].

This table shows the conjugation of the verb عمل/دَار[to do] according to these two local dialects.

ISSN 2007-9737

Table 5. Example of how transliterating some Arabic letters with numerical format in Arabizi

مُتَأَكِّد	A	أ
met2akd / metaked [Sure]	2	
عَجلَة	Aa	ع
3ajla / Aajla [Wheel]	3	
خَايِب	Kh	خ
5ayeb / khayeb [Bad]	5	
حلَالِم	H	ح
7lelim / hlelim [Hlelim]	7	
غنَاية	Gh	غ
8neya / ghneya [Song]	8	
قلقَال	K	ق
9al9al / kalKal [Rolling pin]	9	

3.5 Change of Code

Due to the evolution of communication and social media, and to the luck of Arabic keyboards in some devices, the Arabic chat alphabet, Arabizi ”Franco-Arabic” [47], seems to be the solution for Tunisian to facilitate communication over the Internet, social media, in particular, or for sending messages via cellular phones. This method is a transliterating Arabic with Latin letters and numerals to represent letters [45] as it is shown in the Table below. it’s faster, more informal, trendier, and easier to type than formal Arabic.

Also, an example of a Facebook comment written in the Tunisian Dialect Arabizi format is provided in Fig 3.

3.6 Code Switching

This phenomenon, which refers to the practice of alternating between multiple languages within the same discourse, is widespread in multilingual societies, including Arab countries. Given its global prevalence, it is crucial to equip language technologies with the ability to handle code-switching effectively, ensuring the development of inclusive and user-friendly tools that meet the needs of multilingual communities. In the Arab region, code-switching takes the following forms [23] :

— Switching between MSA and dialects.

Fig. 2. Example of Facebook comment written in Arabizi and Arabic letters

Fig. 3. Example of Facebook comment written in Arabizi

— Switching between Arabic and foreign languages.

And the three main types of code-switching include:

— Inter-sentential: Switching at the sentence level.

— Extra-sentential or tag-switching: The integration of elements from another language such as fillers, interjections, tags, or idiomatic expressions within another language.

— Intra-sentential: Switching within a sentence on the word level, but ensuring that the segments maintain their syntactic rules of both languages.

Fig2 shows an example of code switching (Arabic and French). It also highlights the use of the French word ”illimite,” while the rest of the comment is written in Tunisian Dialect using Arabic script.

3.7 Diacritization

Diacritization in Arabic is the operation which consists in assigning diacritics to the letters of the words not diacritized. Generally, writings in TD are non-diacritic and it is up to the reader to guess the diacritics of the texts at the time of reading [21]. Although diacritics are intended to remove ambiguities during automatic processing, most

ISSN 2007-9737

Fig. 4. Example of ambiguity in the decomposition of words in TD

of the Arabic and TD morphosyntactic parsers only analyze non-diacritic texts due to the lack of diacritical Arabic resources. For example, the non-diacritic word تلقى can get three meanings; the first تَلقَى [find], the second تَلَقَّى [receive] or [throw] تُلقَى [44].

3.8 The New Clitics

The TD introduces new clitics that do not exist in MSA, such as the clitic of negation [مَا + ش]/mA sh/ [not] which manifests in MSA with different particles, namely, مَا /ma/ ,la/ لَا and /lan/ لَن /lam/ لم [32].

Generally, these particles in MSA are often located in front of the verb and can sometimes modify the conjugation such as the verb [أَكَلَ] /aKala/ [to eat] when it is preceded by a negation particle such as [لم]/lam/ [not], it gives rise to the expression of the negation [لم أكِل] /lam Aakol/ [I won't eat]. However, in the TD, this verb changes to [مَا كليتِش] /mA Klytish/ by adding the negation particle [مَا] /mA/. In another example from MSA, the verbal interrogation clitic [أَ] /A/ and the particle [هل] /hal/ are replaced by the clitic [شيء] /shy/ in the TD.

TD exhibits a remarkable linguistic diversity shaped by historical influences, regional variations, and social factors. Its phonological, lexical, morphological, and syntactic distinctions set it apart from MSA, making it a unique and dynamic form of expression. Additionally, the absence of standardized orthography and linguistic resources further highlights its complexity. The interplay between different languages, code-switching, and the rise of Arabizi in digital communication illustrate its evolving nature. Given these specificities, ongoing research and technological advancements are essential to better understand, analyze, and process TD, ensuring its integration into modern language technologies.

4 Methodology

SA in TD faces unique challenges stemming from its oral nature, regional linguistic variations, and the lack of standardized resources. To overcome these hurdles, our methodology, as illustrated in Fig5, follows a comprehensive three-step process: data construction, data preprocessing, and pattern classification. This approach ensures a balance between linguistic accuracy and the incorporation of both Arabic and Arabizi data, effectively bridging the gap between the informal, dynamic nature of TD and the structured demands of NLP tools.

4.1 Data Construction

A corpus is essential for reliable opinion analysis, as it serves to train and evaluate the models developed. To achieve an efficient dataset that ensures optimal performance, specific steps must be followed in its creation and refinement.

4.1.1 Data Source

As in Tunisia, it is possible to make fair use of copyrighted materials according to Article 23, particularly within a research framework. Leveraging this provision, the dataset of this study is primarily built from the Facebook pages of Tunisian telecommunication operators such as "Ooredoo[1], Tunisie-Telecom[2], and Orange Tunisia[3]. To export comments from Facebook and build this dataset, a free version of the website called EXPORT COMMENTS was used [4].

This tool enables the extraction of hundreds of comments from any public Facebook post within seconds by simply providing the postâ€™s

[1] https://www.facebook.com/ooredootn
[2] https://www.facebook.com/TunisieTelecom
[3] https://www.facebook.com/orange.tn/
[4] https://exportcomments.com/export-facebook-comments/

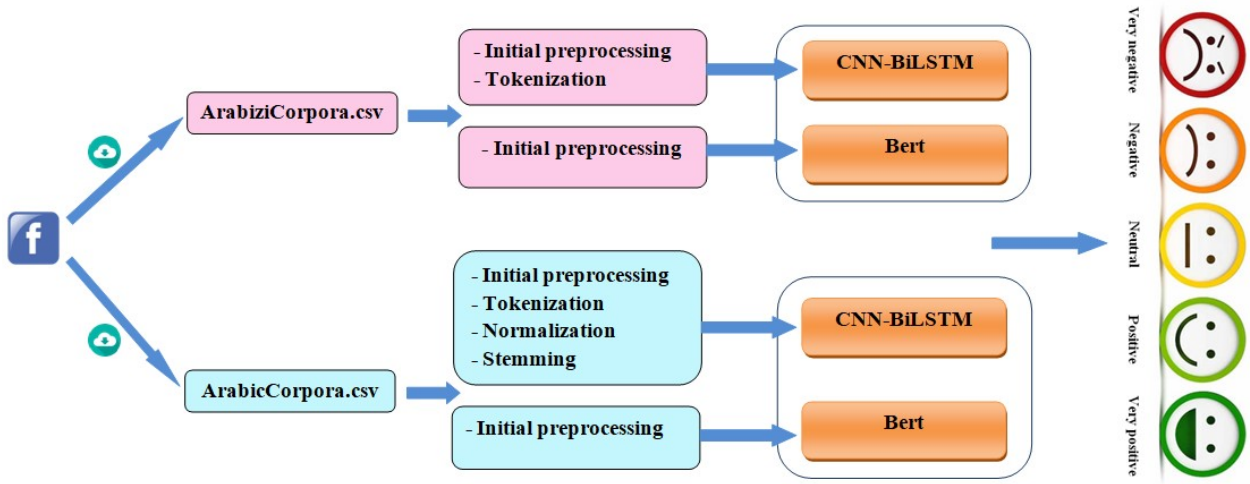


Fig. 5. General structure of our model

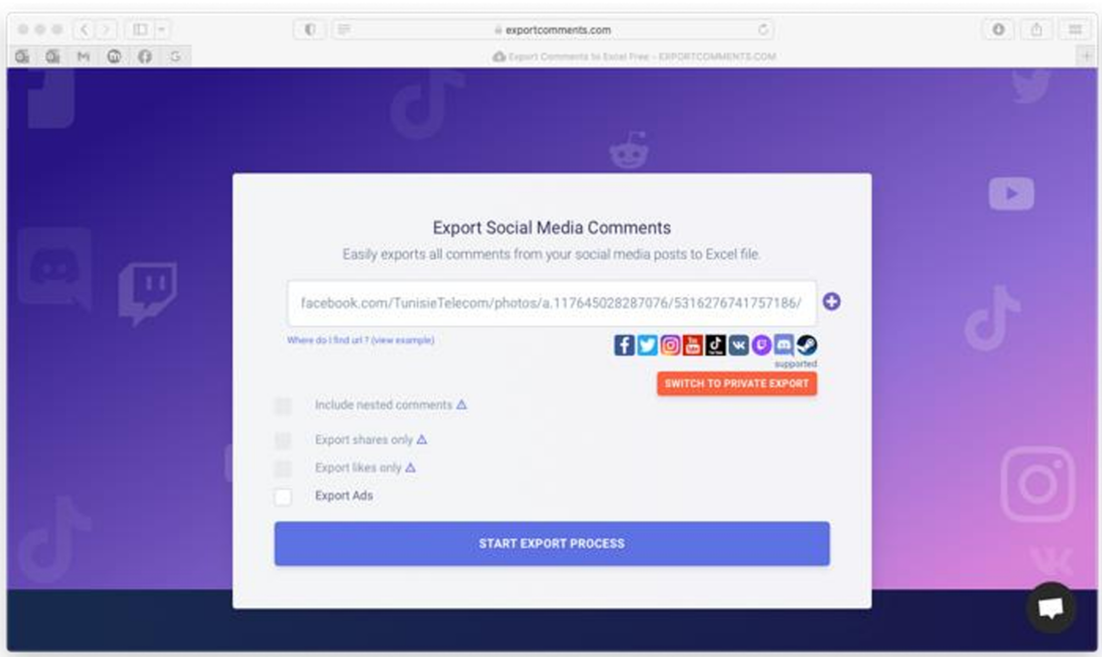
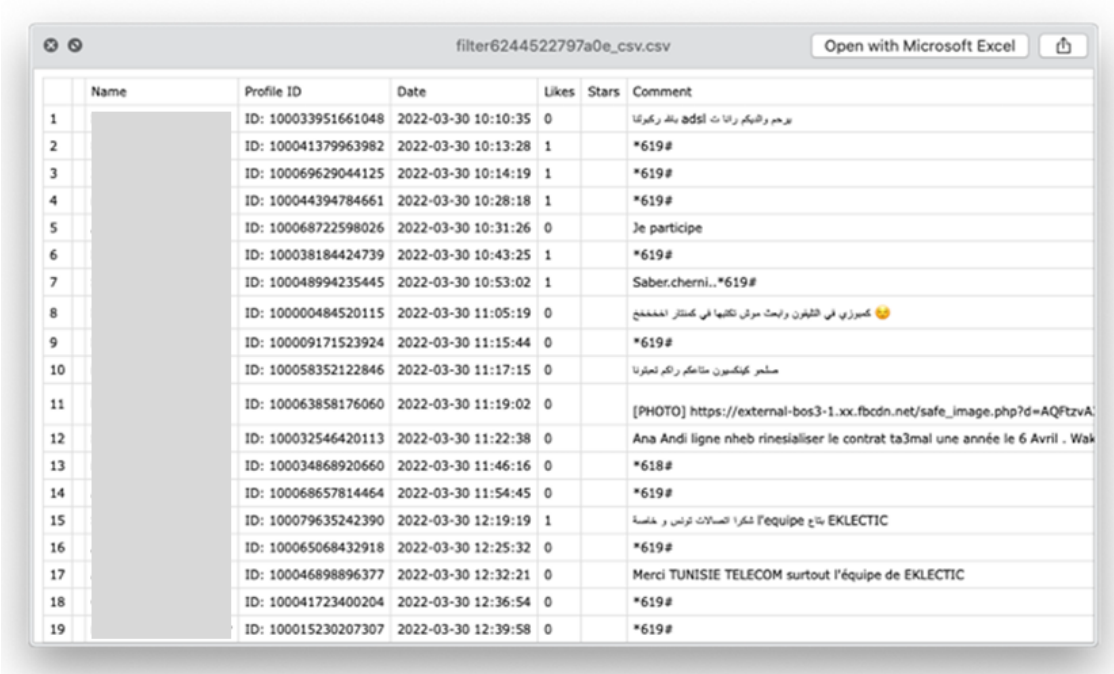


Fig. 6. Interface of the EXPORT COMMENTS tool

URL. As shown in Fig6, this tool can download comments from Facebook, Instagram, Twitter, YouTube, etc. Moreover, it allows the data collected to be saved in various formats, such as CSV or Excel files, as illustrated in Fig7.

The advantages of this tool include not needing to own the Facebook, Instagram, Twitter, TikTok, or YouTube page that published the post. It doesn't matter whether the post is a status, a photo, a video, or a URL; as long as it is public, the tool can

ISSN 2007-9737



	Name	Profile ID	Date	Likes	Stars	Comment
1		ID: 100033951661048	2022-03-30 10:10:35	0		[illegible] adsl [illegible]
2		ID: 100041379963982	2022-03-30 10:13:28	1		*619#
3		ID: 100069629044125	2022-03-30 10:14:19	1		*619#
4		ID: 100044394784661	2022-03-30 10:28:18	1		*619#
5		ID: 100068722598026	2022-03-30 10:31:26	0		Je participe
6		ID: 100038184424739	2022-03-30 10:43:25	1		*619#
7		ID: 100048994235445	2022-03-30 10:53:02	1		Saber.cherni..*619#
8		ID: 100000484520115	2022-03-30 11:05:19	0		[illegible]
9		ID: 100009171523924	2022-03-30 11:15:44	0		*619#
10		ID: 100058352122846	2022-03-30 11:17:15	0		[illegible]
11		ID: 100063858176060	2022-03-30 11:19:02	0		[PHOTO] https://external-bos3-1.xx.fbcdn.net/safe_image.php?d=AQFtzvA
12		ID: 100032546420113	2022-03-30 11:22:38	0		Ana Andi ligne nheb rinesialiser le contrat ta3mal une année le 6 Avril . Wak
13		ID: 100034868920660	2022-03-30 11:46:16	0		*618#
14		ID: 100068657814464	2022-03-30 11:54:45	0		*619#
15		ID: 100079635242390	2022-03-30 12:19:19	1		[illegible] l'equipe بتاع EKLECTIC
16		ID: 100065068432918	2022-03-30 12:25:32	0		*619#
17		ID: 100046898896377	2022-03-30 12:32:21	0		Merci TUNISIE TELECOM surtout l'équipe de EKLECTIC
18		ID: 100041723400204	2022-03-30 12:36:54	0		*619#
19		ID: 100015230207307	2022-03-30 12:39:58	0		*619#

Fig. 7. Example output of the ”Export Comments” tool applied on the official page of ”Tunisie Telecom”

assist in exporting the comments. Simply enter the URL of the article, and it will export all accessible comments. Additionally, most foreign languages, including Arabic, can be exported.

However, the output of the tool contains a lot of unwanted information, necessitating a preliminary step of cleaning the files extracted. Only the comments column was extracted, excluding other unnecessary information such as name, date, and profile ID.

These comments were then compiled into the main corpus file in preparation for preprocessing. A limitation was also identified: only 500 comments can be exported from any post for free at a time, which made building a substantial dataset time-consuming.

4.1.2 Data Specificities

The corpus was created from Facebook comments, chosen primarily because Tunisians express their opinions on this platform significantly more than on other social networks. Tunisians, in fact, are less active on Instagram and Twitter, particularly in daily issues such as sports, fashion, and so on. Furthermore, Facebook allows for longer comments than other social networks, such as Twitter, where the maximum length for a message is 280 characters. This data corpus was built with the help of two technologies. The telecommunications operators in Tunisia are the subject of this corpus study.

The official Facebook accounts of three major Tunisian telecommunications operatorsâ€””Tunisie Telecom”, ”Orange Tunisie”, and ”Ooredoo Tunisie” were examined. The corpus contains comments written in both Latin script (Arabizi) and Arabic script.This is owing to Tunisians’ propensity of writing in Latin, as well as the ease with which they incorporate French terms into their writings and discussions (standard form or SMS). The corpus was collected from the official Facebook pages of telecommunication operators. The presence of Arabic writing and Arabizi writing identifies this corpus. Additionally, the dataset includes informal and non-standard vocabulary, featuring

Table 6. Corpora Statistics

Type of Corpora	Size of Corpora	Total number of words
Arabic TD	17816	194325
Arabizi	27080	245226
Total	44896	439551

spontaneous writing styles such as repeated letters, non-standard abbreviations, misspellings, and language mixing.

In this study, 2 main corpora the first in ”Arabizi” with data size of 27080 comments and the second in ”Arabic TD” with data size 17816 comments were collected. Table 6 provides the statistics for each corpus in terms of comment and word count.

4.1.3 Data Annotation

Manual annotation is thus required to create a learning corpus for SA, particularly to address the multi-layered nature of sentiment interpretation, the corpus was annotated by three native Tunisian speakers over a period of three months. They cooperatively defined five sentiment classes adapted to the specificities of the Tunisian dialect and the nature of the collected data. The annotation involved meticulously reading each remark and assigning it to one of the predefined sentiment categories. Habitual discussions and validation steps were conducted among the annotators to solve complexities and guarantee a high level of inter-annotator agreement. This rigorous process was essential to improve the reliability and quality of the corpus, which serves as a foundation for subsequent sentiment analysis tasks. This process generates five classes which are :

— Very positive (Label = 2) :

When the feedback comprises, statements expressing complete pleasure with a service or product. For example, [service excellent/ excellent service], [yfatta9 telifoun/], [يَهبَل الأَوُفر/ the offer is fantastic] and so on.

— Positive (Label = 1):

Table 7. Results of inter- annotators agreement in terms of LER

	2 judges	3 judges
LER	7.6%	7.9%

If the comment reflects a positive emotion, such as satisfaction or enthusiasm, it is acceptable. For instance, [cv hlowa/ it’s ok], [bien/ good], [سرفيسِهم بَاهي/ their service is good] and so on.

— Neutral (Label = 0):

When the statement is informative or without emotion. For example, the statement [telecom fih telifounet/ there are mobile phones in telecom] is informed but lacks sentimentality.

— Negative (Label = -1):

If the comment communicates a negative emotion, such as displeasure, unhappiness, regret, or any other negative emotion, it should be removed. For example, [ma5yebhom/ they are not good], [معجبنيش/ It doesn’t appeal to me] and so on.

— Very negative (Label = -2):

When a statement communicates a strong negative emotion, such as rage or disappointment .For example, [chay y7achem/ it’s shameful], [يعيف/ disgusting] and so on.

The Inter-Annotator Agreement (IAA) to measure the extent to which independent annotators agree on the diacritics chosen for each label. It was assessed by averaging the Label Error Rate (LER) across all pairs of annotators. Table 7 presents the IAA results.

In the examination of inter-annotator agreement, the overall agreement between the three annotators was 92.31%. All annotation disagreements were meticulously analyzed, and several illustrative examples are presented. For instance, the comment ”bellahii yeziw bla khedheb” / [Seriously, enough with the lies!] was labeled as very negative by two annotators, while the third annotated it as negative. Similarly, the comment ” ملَا كوفرتير في تونس” /mlA kwfrtyr fy twns/ [What a network in Tunisia] received divergent annotations: positive by one annotator and negative by the others. An additional example involves a word whose sentiment varies depending on the context. For instance, the word يّاسر /yAsr/ [a lot] can communicate either a positive or negative word based on how it is used. These inconsistencies emphasize the inherent subjectivity in sentiment analysis in TD.

4.2 Data Preprocessing

Once the corpus is gathered, it must go through a pre-processing procedure before it can be used. First, initial preprocessing methods were applied to clean both corpora and remove non-sentimental content, followed by tokenization. For the Arabic corpus only, two additional steps were performed: rooting and normalization

4.2.1 Initial Pre-Processing

These are first steps in opinion analysis for data preparation, and they include many sub-steps that aim to leave out a lot of content, namely:

Punctuation marks: The most common punctuation marks in the language are used to indicate pauses and divisions in a sentence or text. They include complete stops, quotation marks, exclamation marks, question marks, commas, parentheses, dashes, square brackets and a variety of other signs. These trademarks were ignored in this work since they have no impact on the task of sentiment analysis.

URLs: Since URLs do not impact sentiment analysis, they were completely removed.

Additional white spaces: The extra spaces in the comments have an effect on the analysis task. As a result, an effective analysis of a comment should exclude any extraneous whitespace.

Words that contain single character: All single-character words were removed due to their lack of meaningful content.

Identification of users: Each Facebook user has a unique username that allows them to use the network and identify their posts. And since it lacks a symbol that denotes usernames, such as "" in Twitter, the identification had to be done manually or was often overlooked.

Repetitive comments: A comment may appear multiple times in the corpus, but repeated comments do not add value to sentiment analysis. Therefore, only one instance of each comment was retained. Additional initial preprocessing steps were then applied exclusively to the Arabic corpus.

The numbers, including the dates: All numbers were removed from the Arabic corpus, as they appear as part of words in the Latin (Arabizi) corpus. In the Arabizi corpus, numbers attached to letters were replaced with their corresponding letters.

The missing words: These words, such as pronoms, prepositions, conjunctions, and so on, will not provide information for sentiment analysis. As a result, their removal simplifies the analysis process.

4.2.2 Tokenization

Tokenization is an important process in NLP. It breaks a piece of text into meaningful units, called tokens. These tokens may be words, phrases, or other significant linguistic elements. In English, a non-agglutinative language, tokenization is relatively straightforward: words are usually separated by spaces. However, this poses a greater challenge in languages like Arabic, which have a rich morphology. In both Modern Standard Arabic and dialectal Arabic, a unique token can often contain multiple words, making the process more intricate. This intricacy comes from the presence of affixes, clitics, and morphological varieties that are dealt with in a manner distinct from the simple space-based tokenization.

4.2.3 Arabic-Specific Steps

Normalization: The TD is distinguished using informal writing that does not adhere to strict orthographic rules. This makes automatic processing difficult. In fact, several corpus words have more than one form. As a result, a special preprocessing step was implemented to address these issues and provide consistent data.

Table 8. Example of Arabic TD comments

Labels	Total number of Comments	Comment in TD	Meaning in English
2	311	أَحلَى أَفَار رَاقِي من عند توب نَات و حَالفة مَا نعمل ابونِي لدَارِي كَان من عندكم ان شَاء الّله متفوقة فِي المجَالَات الكل	Such an excellent and classy offer from Topnet! I swear I'll only subscribe to home internet from you. Hoping you continued success in all fields, God willing.
1	1980	أَكِيد كلنَا نحبوهَا برَافو لِيهَا عملت مجهود و ان شَاء الّله المرة الجَاية	Of course, we all love it. Well done for the effort made, and wishfully next time!
0	8372	اِبرِّي وِين عنوَانهَا فِي سوسة	Please, where is its address in Sousse?
-1	5477	كَان تشوفوا أُوبِيرتور حمَام الأَنف تقول عُطرِية	If you see the Ooredoo store in Hammam Lif, you'd think it's a perfume shop
-2	1678	بِالَله لكرِيم تعمل العَار و الأَسوَام أَغلَا برشة من أَورونج مسِيكن اللِّي مَا يعرفش الأَسوَام قبل مَا يمشِي	Honestly, it's shameful and the prices are much higher than Orange. Poor person who doesn't check the prices before going

Table 9. Example of Arabizi comments

Labels	Total number of Comments	Example of Comment
2	396	Bravo kol mara tothboutlna elli entom fel mostawa
1	2893	kol am tunisie telecom fi ta9adim izdihar
0	17698	brabi amloulna des promo fi les telephons tawa wakt iphone pro
-1	5340	bellahii yeziw bla khdeb
-2	753	hethoukom lithnin kil kleb ma3doumin il insania houma mistensin sir9a wi trafik ybi3ouh wou yna7iou 7a9ou lihom ya9smouh

This step involves normalizing Arabic words by converting various forms of a given word into a single orthographic representation.

Stemming: Words written in TD are often formed of several words, underlining the importance of the root word task. The goal of light stemming is to preserve the root of the word while removing all prefixes and suffixes. Here is an example illustrating the deletion of a suffix [نَا], the term [بوتيكنَا] is substituted by [بوتيك].

4.3 Classification

4.3.1 CNN BiLSTM

The CNN-BiLSTM model combines strengths from Convolutional Neural Networks (CNNs) as well as Bidirectional Long Short-Term Memory (Bi-LSTM) networks, thus making it highly efficient within Natural Language Processing (NLP) tasks, most particularly in sentiment analysis.

4.3.2 BERT

BERT is a transformer-based language model that uses multiple encoder layers along with self-attention heads to learn contextual word embeddings. It is mostly pre-trained upon very large unlabeled datasets, including BooksCorpus and even English Wikipedia, in order to fully catch deep linguistic patterns. BERT can still be fine-tuned now for various NLP tasks, like with sentiment analysis, question answering, as well as text classification, by leveraging its contextual embeddings in order to understand all of the subtleties and meaning of words in context.

5 Experiments

In this section, the entire pipeline is described, from dataset collection and preprocessing to model evaluation. The two datasets used are arabicCorpora with 17,816 comments and arabiziCorpora with 27,080 comments. Each dataset was collected using the free EXPORT COMMENTS website, and the comments are sorted into three sentiment categories (negative, neutral, and positive). The pretraining stage begins with manual initial preprocessing, followed by normalization, which was performed using a semi automatic tool [13] .

Tokenization was carried out using the camel_tools module [38] , while stemming was also performed manually to prepare the text data for further processing. For the CNN-BiLSTM model, a complete preprocessing workflow was executed. For arabiziCorpora, only initial preprocessing and tokenization were applied, whereas arabicCorpora underwent initial preprocessing, tokenization, normalization, and stemming. For BERT, the only initial preprocessing took place on both datasets because the tokenization is inherently done by the BERT model itself. After the preprocessing, the datasets were split with a 75% training and 25% testing split, and 10% of the training data was set aside for hyperparameter validation. This is to ensure the model can be tuned properly, and to avoid overfitting.

For the CNN-BiLSTM model, Tables 10 and 11 offer the optimal configurations. The sequence length was set to 300, which corresponds to the maximum tweet length in the corpus, ensuring that all input text fits within the model's sequence limits. The embedding dimension was set to 100 to capture the semantic meaning of words, and the hidden size of the BiLSTM layer was set to 64. The vocabulary size was 37,428, which guarantees an effective representation of textual data. In terms of model sophistication, there are 3,742,800 non-trainable parameters corresponding to the pre-trained GloVe embeddings, which endures

Table 10. Hyper-parameters values for CNN-BiLSTM model

Learning rate	Batch size	Dropout rate	Number of epochs
0.001	200	0.7	100

Table 11. Configuration of CNN-BiLSTM parameters

Sequence length	300
Embedding dimension	100
Hidden size	64
Vocab size	37,428
Non-trainable params	3,742,800
Trainable params	32,197
GRU parameters	31,872
Total parameters	3,774,997

fixed during training to conserve their learned semantic knowledge.

The 32,197 trainable parameters include the weights and biases of the layers, permitting the model to adjust and fine-tune during training. Among these, 31,872 parameters belong to the GRU layer, which is responsible for capturing the temporal dependencies in the sequence. In total, the model has 3,774,997 parameters, signifying its complexity and ability to capture intricate patterns in the data.

For the BERT model, Tables 12 and 13 represents the best hyperparameters and parameters identified during the experimentation phase. The BERT architecture used here consists of a single layer with an input dimension of 28,996, which matches with the chosen vocabulary size. Each word in the vocabulary is embedded into a dense vector of 768 dimensions, allowing the model to capture deep semantic relationships between words. The input length is fixed at 512 to maintain stability across input sequences while assuring that essential contextual information is protected. This configuration guarantees a balance between model efficiency and predictive accuracy, allowing BERT to perform well on various downstream tasks. Finally, for model evaluation, several metrics were utilized to assess performance, including Accuracy, Precision, Recall, F1 Score, Training Accuracy, and Testing Accuracy. These metrics give a comprehensive view of the models' effectiveness in handling

Table 12. Hyper-parameters values for BiLSTM model

Learning rate	Batch size	Dropout rate	Number of epochs
0.00005	80	0.1	5 10

Table 13. Configuration of BiLSTM parameters

Sequence length	512
Embedding dimension	768
Hidden size	768
Vocab size	28,996

sentiment analysis tasks, assuring that the models are evaluated thoroughly in terms of both general performance and task-specific capabilities.

6 Results

In this study, The transition was made from the CNN-BiLSTM model, used to compare with the BERT model, to evaluate whether semantic classification could improve recall in thought layout tasks. Initial experiments used GloVe embeddings, following [39], who highlighted the advantage of initializing word vectors with those from an unsupervised neural language model when large supervised datasets are unavailable. Due to the medium size of this dataset, several pre-trained models were trained and evaluated.

Before training, Crucial preprocessing steps were performed to improve the quality of the textual data, as detailed in the experiments section. This process contained several steps such removing unnecessary punctuation marks and filtering out repetitive comments to guarantee cleaner input.

Normalization was applied to standardize text variations, tokenization to split sentences into words or subwords, and stemming to reduce words to their root forms.

These steps were essential in improving the model's ability to recognize patterns and accomplish better performance in text classification tasks. Below, some examples of these preprocessing steps are provided in Table 14. The instance used from Arabic Corpora in this table is

بِالَله لكرِيم تعمل العَار و الأَسوَام أَغلَا برشة من أَورونج مسِيكن مسِيكن اللِّي مَا يعرفش الأَسوَام قبل مَا يمشِي

/bAllh lkrym tEml Al EAr w AlAswAm AglA br$A mn Awrwnj msykn msykn Ally mA yAErAf$ AlAswAm qbl mA ym$y/ [Honestly, it's shameful and the prices are much higher than Orange. Poor person who doesn't check the prices before going].

Model performance was monitored through learning curves, stopping training when validation error began to rise to prevent underfitting or overfitting. For the Arabizi corpus, after training for five epochs, the model achieved an accuracy of 89.34%. For the Arabic corpus, after seven epochs, the model reached an accuracy of 88.64%. These results indicated that while the GloVe-based Bidirectional LSTM model performed reasonably well, the testing accuracies were lower than the training accuracies, suggesting overfitting likely due to the limited dataset size.

On the other hand, the latest methodology of attention models was implemented. As detailed in the previous chapter, the pre-trained BERT-Base model was used. Its architecture is as follows: BertForSequenceClassification, which includes the BertModel with its various embeddings, encoder layers, and an output classifier.

The detailed architecture, as figured in Fig 8 ensures comprehensive context understanding and sequence classification capabilities.

After training the BERT-Base model on both corpora, the following results were visualized:

For the Arabic Model: The training loss decreased over time, achieving low error values. The training and validation curves improved significantly, though there was a notable gap between them in the first five epochs with an accuracy of 98.43%, indicating that they operated like datasets from different distributions. However, from epoch five onwards, both curves improved, reaching optimal performance around epoch nine and represents an accuracy of 99.3% in epoch ten.

Table 14. Examples of comments during preprocessing

Dataset	Comment	Initial preprocessing	Normalization	Stemming	Tokenization
ArabicCorpora	بِالَله لكرِيم تعمل العَار و الأَسوَام أَغلَا برشة من أَورونج مسِيكن مسِيكن اللّي مَا يعرفش الأَسوَام قبل مَا يمشِي	بِالَله لكرِيم تعمل العَار و الأَسوَام أَغلَا برشة من أَورونج مسِيكن اللّي مَا يعرفش الأَسوَام قبل مَا يمشِي	بِالَله لكرِيم تعمل العَار و الأَسوَام أَغلَا برشَا من أَورونج مسِيكن اللّي مَا يعرفش الأَسوَام قبل مَا يمشِي	الَله كرِيم عمل عَار و سعر غلَى برش من أَورونج مسكن لِي مَا عرف سعر قبل مَا مشَى	[«الَله»، «كرِيم»، «عمل»، «عَار»، «و»، «سعر»، «غلَى»، «برش»، «من»، «أَورونج»، «مسكن»، «لِي»، «مَا»، «عرف»، «سعر»، «قبل»، «مَا»، «مشَى»]
ArabiziCorpora	Tunisie telecom fahamni ech newa probleme l telifoun dps 17/11 !!	Tunisie telecom fahamni ech newa probleme telifoun	-	-	[”Tunisie Telecom”, ”please”, ”explain”, ”to”, ”me”, ”what”, ”the”, ”issue”, ”is”, ”with”, ”the”, ”phone”]

Table 15. Performance Metrics for Arabic Models

Model	Epochs	Accuracy	Precision	Recall	F1 Score	Training Accuracy	Testing Accuracy
CNN-BiLSTM	7	0.8864	0.8791	0.8864	-	65.38%	60.31%
Bert	5	0.9843	0.9904	0.9843	0.9854	77.23%	76.75%
	10	0.9930	0.9941	0.9930	0.9925	78.46%	77.34%

Table 16. Performance Metrics for Arabizi Models

Model	Epochs	Accuracy	Precision	Recall	F1 Score	Training Accuracy	Testing Accuracy
CNN-BiLSTM	5	0.8934	0.8864	0.8832	-	72.47%	69.92%
Bert	5	0.9424	0.9653	0.9424	0.9472	76.79%	76.62%
	10	0.9448	0.9652	0.9448	0.9486	76.97%	76.72%

For the Arabizi Model: The Table 16 above clearly shows the evaluation metrics for this model at two different training epochs (5 and 10). While the improvements were not as pronounced, the metrics still indicate high performance with an accuracy of 94.24% in five epochs and 94.48% in ten epochs.

To better evaluate the performance of our models beyond global metrics such as accuracy and F1-score, confusion matrices for both the Arabic and Arabizi corpora are presented under different model configurations (CNN-BiLSTM and BERT with varying epochs). As shown in Fig 9, These matrices provide a fine-grained view of the modelsâ€™ ability to correctly classify each sentiment classâ€”Very Positive, Positive, Neutral, Negative, and Very Negative.

For the Arabic corpus, the BERT model trained for 10 epochs showed strong performance, with most predictions concentrated along the diagonal of the matrix, indicating high accuracy across all classes. Particularly, it effectively distinguished between Neutral and Negative comments, which are often confused in sentiment

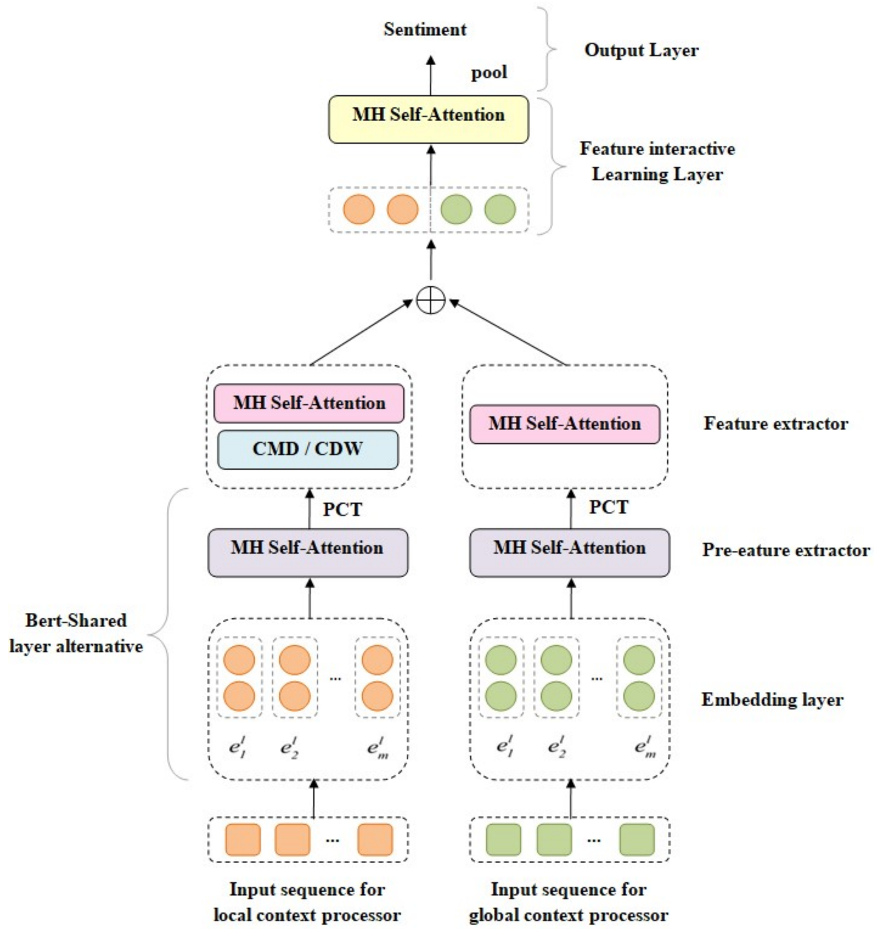


Fig. 8. Decomposition of Bert model architecture for sentiment analysis

Table 17. Comparison of SA Studies in TD

Ref.	Dataset	Accuracy
TSAC [33]	17,000 comments from Tunisian radios and TV channels	MLP, acc: 78%
Supermarket Pages [31]	44,000 comments from supermarket Facebook pages	BiLSTM, acc: 82%
TunBERT [31] [34] [43]	TSAC and TEC datasets	acc: 96.98 % on TSAC and 81.2% on TEC Covid-19 pandemic [27]
CTSA dataset	BiLSTM, acc: 78,10%	
Fine-tuned BERT [18]	TD data	acc: 85% with fine-tuned BERT **Proposed method**
TUNTESA dataset (27,080 Arabizi and 17,816 Arabic comments)	BERT accuracy of 99% for Arabic and 94% for Arabizi	

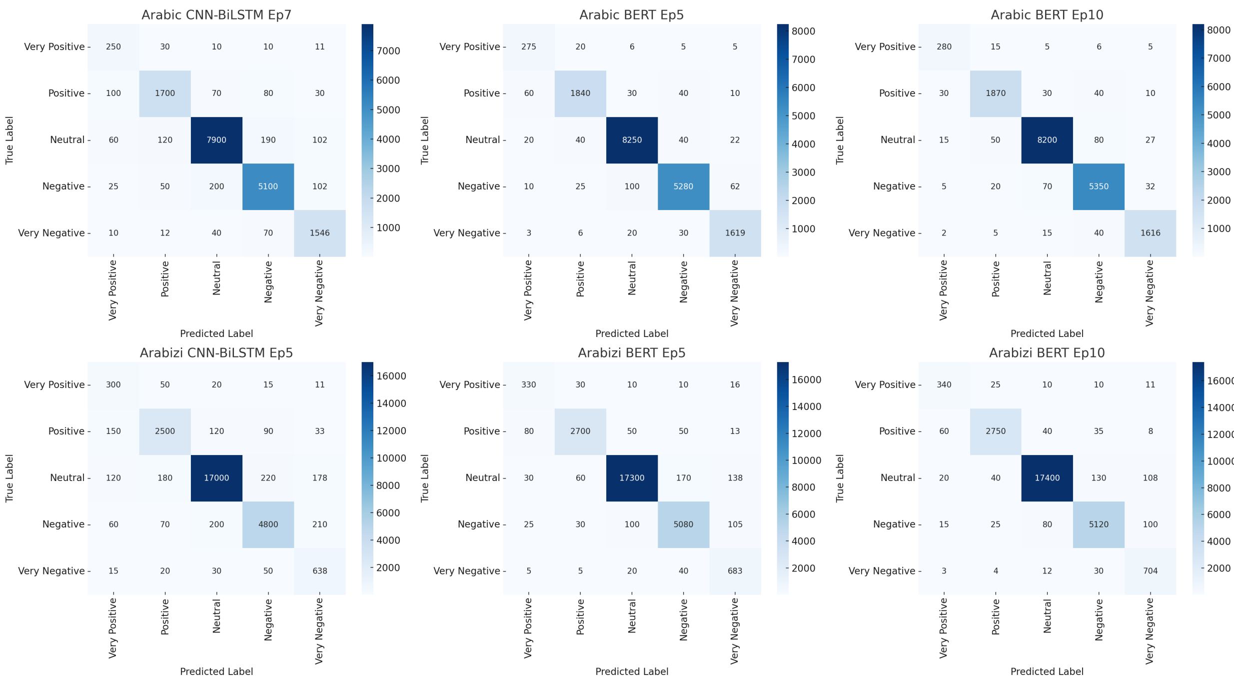


Fig. 9. Confusion Matrices for Arabic and Arabizi Sentiment Classification Using CNN-BiLSTM and BERT

analysis tasks. The CNN-BiLSTM model (7 epochs), while still reasonably accurate, exhibited more misclassifications, especially between Positive and Neutral, and between Negative and Very Negative, which slightly affected its overall precision and recall.

For the Arabizi corpus, a similar trend was observed. The BERT model (10 epochs) again outperformed CNN-BiLSTM by achieving clearer class separation and fewer off-diagonal errors. The CNN-BiLSTM model tended to confuse Neutral with adjacent sentiments like Positive or Negative, likely due to its lower representational capacity compared to BERT.

In contrast, BERT's deeper architecture allowed for better generalization across informal and variable Arabizi text.

These confusion matrices confirm that increasing the number of training epochs helps BERT refine its internal representations, leading to improved classification, especially for the minority classes like Very Positive and Very Negative.

Additionally, they highlight the challenges of sentiment classification in imbalanced and morphologically rich data, underscoring the importance of class-wise evaluation in multilingual and dialectal settings.

7 Discussion

The previous section demonstrate that the BERT-Base models yielded excellent outcomes for both corpora. The BERT models outperformed the traditional Bidirectional LSTM models, indicating that the attention-based approach significantly enhances the accuracy, precision, recall, and F1 score, thereby proving to be a more effective method for semantic classification.

Recent studies in sentiment analysis of TD have employed various methodologies and datasets to tackle challenges posed by dialectical variations and informal expressions in social media texts. As shown in table17, previous research often utilized traditional models such as MLP and BiLSTM, achieving notable results like 78% accuracy

Table 18. Comparison of SA Studies and proposed method

Ref.	Dialects	Datasets	Models	Performances
[6]	Saudi dialect	SDCT corpus :32,063 tweets	LSTM, Bi-LSTM and SVM	Best: Bi-LSTM, F1-score : 94%
[35]	MSA and Egyptian dialect	40,000 tweets	CNN, LSTM, RCNN	Best: LSTM, Accuracy : 88%
[33]	MSA and TD	TSAC : 17,000 comments	SVM, BNB, MLP	Best: MLP, Accuracy : 78%
[31]	TD	Arabic Corpora: 17,816 comments	CNN, LSTM, Bi-LSTM	Best for Binary classification: LSTM and Bi-LSTM, Accuracy: 87%
		Arabizi Corpora: 27,080 comments		Best for Three way classification: LSTM, Accuracy: 87%
[34]	TD	TSAC	TunBert	Accuracy : 96.98%
[43]	MSA and TD	TEC: 3,042 tweets	TunBert	Accuracy : 81.2%
[2]	MSA	ArSarcasm	ARBERT, MARBERT	Best: MARBERT, F-score : 71.50%
[1]	MSA, Egyptian and Maghrebi dialects	ArSarcasm-v2	Multi-headed-LSTM- CNN-GRU, MARBERT	Best: MARBERT, Accuracy : 69.57%
[3]	MSA	ArSarcasm-v2	MARABERT, ArBert, QARIB, AraBert-v02, GigaBert, Arabic Bert, MBert	Best: MARABERT, F1-score: 86%
[18]	TD	72,000 comments	SMO, NB, DT, Bert	Best: Bert, Accuracy : 85%
[41]	MSA, Egyptian and Gulf dialects	HARD : 106,000 reviews	Bi-LSTM+Bert, CNN	Best: Bi-LSTM+Bert, Accuracy : 94.25%
[12]	Jordanian, Levantine, Arabic, Syrian and Lebanese dialects	Twitter-AB: sentiment dataset within 2000 tweets	QST, QSR, QSRT, QE3, QE6	Best for sentiment: QSRT, Macro F1-score: 97.45%
		EATD: emotion dataset within 2021 tweets		Best for emotion: QE3, Macro F1-score: 90.10%
[22]	MSA and Algerian dialect	110K comments and tweets (54k comments in Algerian dialect)	AlgBert, GNB	Best: AlgBert, Accuracy: 92.6%
[17]	MSA and Jordanian dialect	AJGT: 1,800 tweets	ASA-medium Bert, AraBert, mBert, Arabic-Bert	Best: ASA-medium Bert, Accuracy: 96.11%
[29]	Modern Arabic, Khaleeji, Hijazi, Egyptian and Maghrebi dialects	SemEval: 3,355 tweets	Generative LLM	Accuracy with neutral data: 64%
				Accuracy without neutral data: 82%
[15]	Moroccan dialect	MSDA-MAC: 68,000 tweets	AraBERT, QARIB, ALBERT, AraELECTRA, CAMeLBERT, SVM and CNN	Best: QARIB, Accuracy: 96%
[16]	Moroccan dialect	10,254 comments	LR, SVM, DT, MNB, XGBoost, Neural Network approach, AraBert	Best for machine learning: LR + TF-IDF, Accuracy: 82% Best for deep learning: AraBert + Bert, Accuracy: 80%
[27]	TD	CTSA dataset: 40,000 posts and comments	Bi-LSTM, SVM, DT, NB, LSTM, CNN	Best: Bi-LSTM, Accuracy: 78%
[28]	TD	199k comments collected from ASTD, TUNIZI, TSAC and CTSA	Bi-LSTM, SVM, DT, NB, LSTM, CNN	Best for machine learning: NB + TF-IDF and LR + TF-IDF, Accuracy: 85% Best for deep learning: Embedding layer, Bi-LSTM layer and two dropout layers followed by two fully connected layers, Accuracy: 88%
Proposed method	TD	TUNTESA: 44,896 comments	CNN bidirectional LSTM, Bert	Best: Bert, Accuracy for Arabic: **99.3%**, Accuracy for Arabizi: **94.48%**

for TSAC and 82% accuracy for supermarket comments. More recently, approaches using advanced models like BERT have demonstrated exceptional performance, with TunBERT achieving 96.98% accuracy on the TSAC corpus. This method, based on the TUNTESA dataset, employed BERT models that yielded promising results with 99% accuracy for Arabic and 94% for Arabizi, highlighting the effectiveness of this approach for semantic analysis of TD on social media platforms.

A comparative analysis between these studies and this work is presented in Table 18, highlighting dialects, dataset characteristics, and performance metrics.

8 Conclusion and Further Works

In conclusion, this journal article presents a comprehensive investigation into sentiment analysis applied to datasets sourced from Tunisian telecom operators' Facebook pages, focusing on comments written in both Arabizi and Arabic TD. The study employs innovative methodologies such as the EXPORT COMMENTS tool for data collection and specialized preprocessing techniques tailored to manage informal language and dialectical variations prevalent in social media discourse.

The results highlight that while traditional models like Bidirectional LSTM with GloVe embeddings demonstrate reasonable performance, they are surpassed by BERT models, which achieve significantly higher accuracy, precision, recall, and F1 scores across both datasets. These findings underscore the effectiveness of attention-based models like BERT in capturing nuanced sentiment from diverse linguistic contexts found in social media interactions.

Looking ahead, several avenues for future research emerge from this study. Firstly, expanding the dataset to include a broader range of sources and dialects could enhance the generalizability and robustness of sentiment analysis models, particularly in capturing regional linguistic nuances. Secondly, further refining preprocessing techniques to better handle challenges such as slang, misspellings, and code-switching in informal digital communication would contribute to more accurate sentiment analysis outcomes.

Additionally, exploring hybrid models that integrate BERT with domain-specific embeddings or domain adaptation techniques could potentially yield even higher performance in sentiment classification tasks. Moreover, by effectively analyzing sentiments, operators can enhance customer satisfaction, manage relations, and glean valuable feedback, thereby maintaining a competitive edge in the market. The dataset, named TUNTESA, comprises 27,080 Arabizi and 17,816 Arabic comments sourced from official telecommunications operators' Facebook pages.

Fundamentally, this research contributes profoundly to the advancement of sentiment analysis in the context of under-resourced language types such as Arabic dialects and Arabizi. By addressing key challenges such as code switching, dialectal forms, and the deficiency of annotated resources, this work not only improves the field of Natural Language Processing (NLP) but also builds a solid groundwork for future research projects.

Moreover, the suggested sentiment analysis framework is designed to be adjustable across various domains, including social media, customer feedback, public opinion monitoring, and even healthcare related textual data. This flexibility emphasizes the capacity of this approach to support a wide spectrum of applications where informal and dialectal language is prevalent, eventually contributing to a deeper comprehension of human emotions and opinions in typically occurring text.

In future, this work facilitates the way for building more inclusive, dialect-aware NLP systems capable of capturing linguistic diversity in regions such as North Africa and the Middle East, and inspires interdisciplinary cooperation to further bridge the gap between informal language use and computational text analysis.

References

1. **Abdel-Salam, R. (2021).** Wanlp 2021 shared-task: Towards irony and sentiment detection in arabic tweets using multi-headed-lstm-cnn-gru and marbert. Proceedings of the Sixth Arabic natural language processing workshop, pp. 306–311.

2. **Abdul-Mageed, M., Elmadany, A., Nagoudi, E. M. B. (2020).** Arbert & marbert: Deep bidirectional transformers for arabic. arXiv preprint arXiv:2101.01785.

3. **Abuzayed, A., Al-Khalifa, H. (2021).** Sarcasm and sentiment detection in arabic tweets using bert-based models and data augmentation. Proceedings of the sixth Arabic natural language processing workshop, pp. 312–317.

4. **Al-Qablan, T. A., Mohd Noor, M. H., Al-Betar, M. A., Khader, A. T. (2023).** A survey on sentiment analysis and its applications. Neural Computing and Applications, Vol. 35, No. 29, pp. 21567–21601.

5. **Al-Smadi, B. S. (2024).** Deberta-bilstm: A multi-label classification model of arabic medical questions using pre-trained models and deep learning. Computers in Biology and Medicine, Vol. 170, pp. 107921.

6. **Alahmary, R. M., Al-Dossari, H. Z., Emam, A. Z. (2019).** Sentiment analysis of saudi dialect using deep learning techniques. 2019 international conference on electronics, information, and communication (ICEIC), IEEE, pp. 1–6.

7. **Alasmari, E., Hamdy, M., Alyoubi, K. H., Alotaibi, F. S. (2023).** Arabic stock-news sentiments and economic aspects using bert model. International Journal of Advanced Computer Science and Applications, Vol. 14, No. 1.

8. **Almasri, M., Al-Malki, N., Alotaibi, R. (2023).** A semi supervised approach to arabic aspect category detection using bert and teacher-student model. PeerJ Computer Science, Vol. 9, pp. e1425.

9. **AlMuhaideb, S., AlNegheimish, Y., AlOmar, T., AlSabti, R., AlKathery, M., AlOlyyan, G. (2023).** Analyzing arabic twitter-based patient experience sentiments using multi-dialect arabic bidirectional encoder representations from transformers.. Computers, Materials & Continua, Vol. 76, No. 1.

10. **AlNasser, S., AlMuhaideb, S. (2024).** Listening to patients: Advanced arabic aspect-based sentiment analysis using transformer models towards better healthcare. Big Data and Cognitive Computing, Vol. 8, No. 11, pp. 156.

11. **Alshaikh, K. A., Almatrafi, O. A., Abushark, Y. B. (2023).** Bert-based model for aspect-based sentiment analysis for analyzing arabic open-ended survey responses: a case study. IEEE Access, Vol. 12, pp. 2288–2302.

12. **Alshehri, W., Al-Twairesh, N., Alothaim, A. (2023).** Affect analysis in arabic text: Further pre-training language models for sentiment and emotion. Applied Sciences, Vol. 13, No. 9, pp. 5609.

13. **Besdouri, F. Z., Mekki, A., Zribi, I., Ellouze, M. (2021).** Improvement of the cota-orthography system through language modeling. 2021 IEEE/ACS 18th International Conference on Computer Systems and Applications (AICCSA), IEEE, pp. 1–7.

14. **Boukadida, N. (2008).** Connaissances phonologiques et morphologiques derivationnelles et apprentissage de la lecture en arabe (Etude longitudinale). These de doctorat, Universite Rennes 2.

15. **Bourahouat, G., Abourezq, M., Daoudi, N. (2024).** Improvement of moroccan dialect sentiment analysis using arabic bert-based models. J. Comput. Sci, Vol. 20, No. 2, pp. 157–167.

16. **Cherrat, E. M., Ouahi, H., Bekkar, A. (2024).** Sentiment analysis from texts written in standard arabic and moroccan dialect based on deep learning approaches. International Journal of Computing and Digital Systems, Vol. 16, No. 1, pp. 447–458.

17. **Chouikhi, H., Jarray, F. (2021).** Bert-based ensemble learning approach for sentiment analysis. International Joint Conference on Knowledge Discovery, Knowledge Engineering, and Knowledge Management, Springer, pp. 118–128.

18. **Fsih, E., Boujelbane, R., Belguith, L. H. (2022).** A deep sentiment analysis of tunisian dialect comments on multi-domain posts in different social media platforms. Proceedings of the 5th International Conference on Natural Language and Speech Processing (ICNLSP 2022), pp. 226–233.

19. **Fsih, E., Kchaou, S., Boujelbane, R., Belguith, L. H. (2022).** Benchmarking transfer learning approaches for sentiment analysis of arabic dialect. Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP), pp. 431–435.

20. **Guellil, I., Azouaou, F. (2017).** Asda: Analyseur syntaxique du dialecte alg {\'e} rien dans un but d'analyse s {\'e} mantique. arXiv preprint arXiv:1707.08998.

21. **Guerfali, A., Hajji, I. (2024).** Utilisation des reseaux sociaux en tunisie. Technical report, Ikigai Tunisie. Accessed 22 June 2024.

22. **Hamadouche, K., Bousmaha, K. Z., Bekkoucha, M. A., Hadrich-Belguith, L. (2023).** Algbert: Automatic construction of annotated corpus for sentiment analysis in algerian dialect. ACM Transactions on Asian and Low-Resource Language Information Processing, Vol. 22, No. 12, pp. 1–17.

23. **Hamed, I., Sabty, C., Abdennadher, S., Vu, N. T., Solorio, T., Habash, N. (2025).** A survey of code-switched arabic nlp: Progress, challenges, and future directions. arXiv preprint arXiv:2501.13419.

24. **Hameed, R. A., Abed, W. J., Sadiq, A. T. (2023).** Evaluation of hotel performance with sentiment analysis by deep learning techniques.. International Journal of Interactive Mobile Technologies, Vol. 17, No. 9.

25. **Iben Nasr, L., Masmoudi, A., Hadrich Belguith, L. (2024).** Survey on arabic speech emotion recognition. International Journal of Speech Technology, Vol. 27, No. 1, pp. 53–68.

26. **Ibn Nasr, L., Masmoudi, A., Hadrich Belguith, L. (2025).** Emotion recognition from spontaneous tunisian dialect speech. ACM Transactions on Asian and Low-Resource Language Information Processing, Vol. 24, No. 2, pp. 1–16.

27. **Jaballi, S., Hazar, M. J., Zrigui, S., Nicolas, H., Zrigui, M. (2023).** Deep bidirectional lstm network learning-based sentiment analysis for tdical facebook content during the spread of the coronavirus pandemic. International Conference on Computational Collective Intelligence, Springer Nature Switzerland, Cham, pp. 96–109.

28. **Jaballi, S., Zrigui, S., Sghaier, M. A., Berchech, D., Zrigui, M. (2022).** Sentiment analysis of tunisian users on social networks: overcoming the challenge of multilingual comments in the tunisian dialect. International Conference on Computational Collective Intelligence, Springer, pp. 176–192.

29. **Khaled, S., Mohamed, E. H., Medhat, W. (2024).** Evaluating large language models for arabic sentiment analysis: A comparative study using retrieval-augmented generation. Procedia Computer Science, Vol. 244, pp. 363–370.

30. **Masmoudi, A., Bougares, F., Ellouze, M., Estève, Y., Belguith, L. (2018).** Automatic speech recognition system for tunisian dialect. Language Resources and Evaluation, Vol. 52, pp. 249–267.

31. **Masmoudi, A., Hamdi, J., Belguith, L. H. (2021).** Deep learning for sentiment analysis of tunisian dialect. Computación y Sistemas, Vol. 25, No. 1, pp. 129–148.

32. **Masmoudi Dammak, A. (2018).** Approche hybride pour la reconnaissance automatique de la parole en langue arabe. ThÃ¨se de doctorat, Universite de Sfax.

33. **Mdhaffar, S., Bougares, F., Esteve, Y., Hadrich-Belguith, L. (2017).** Sentiment analysis of tunisian dialects: Linguistic ressources and experiments. Third Arabic natural language processing workshop (WANLP), pp. 55–61.

34. **Messaoudi, A., Cheikhrouhou, A., Haddad, H., Ferchichi, N., BenHajhmida, M., Korched, A., Naski, M., Ghriss, F., Kerkeni, A. (2022).** Tunbert: Pretrained contextualized text representation for tunisian dialect. International Conference on Intelligent Systems and Pattern Recognition, Springer, pp. 278–290.

35. **Mohammed, A., Kora, R. (2019).** Deep learning approaches for arabic sentiment analysis. Social Network Analysis and Mining, Vol. 9, No. 1, pp. 52.

36. **Mulki, H., Haddad, H., Babaoğlu, I. (2017).** Modern trends in arabic sentiment analysis: A survey. Traitement Automatique des Langues, Vol. 58, No. 3, pp. 15–39.

37. **Nasr, L. I., Masmoudi, A., Belguith, L. H. (2023).** Natural tunisian speech preprocessing for features extraction. 2023 IEEE/ACIS 23rd

International Conference on Computer and Information Science (ICIS), IEEE, pp. 73–78.

38. **Obeid, O., Zalmout, N., Khalifa, S., Taji, D., Oudah, M., Alhafni, B., Inoue, G., Eryani, F., Erdmann, A., Habash, N. (2020).** Camel tools: An open source python toolkit for arabic natural language processing. Proceedings of the twelfth language resources and evaluation conference, pp. 7022–7032.

39. **Rakhlin, A. (2016).** Convolutional neural networks for sentence classification. GitHub, Vol. 6, pp. 25.

40. **Rodríguez-Ibánez, M., Casánez-Ventura, A., Castejón-Mateos, F., Cuenca-Jiménez, P.-M. (2023).** A review on sentiment analysis from social media platforms. Expert Systems with Applications, Vol. 223, pp. 119862.

41. **Sabbeh, S. F., Fasihuddin, H. A. (2023).** A comparative analysis of word embedding and deep learning for arabic sentiment classification. Electronics, Vol. 12, No. 6, pp. 1425.

42. **Salloum, S. A., Al-Emran, M., Monem, A. A., Shaalan, K. (2018).** Using text mining techniques for extracting information from research articles. Intelligent natural language processing: Trends and Applications, pp. 373–397.

43. **Sayadi, K., Liwicki, M., Ingold, R., Bui, M. (2016).** Tunisian dialect and modern standard arabic dataset for sentiment analysis: Tunisian election context. Second international conference on Arabic computational linguistics, ACLING, pp. 35–53.

44. **Tan, K. L., Lee, C. P., Lim, K. M. (2023).** A survey of sentiment analysis: Approaches, datasets, and future research. Applied Sciences, Vol. 13, No. 7, pp. 4550.

45. **Teach Mideast (2024).** Arabizi the Arabic chat language changing the way young people write speak. https://teachmideast.org/articles/arabizi-the-arabic- chat-language-changing-the-way-young-people-write-speak/. [Online].

46. **Tian, H., Gao, C., Xiao, X., Liu, H., He, B., Wu, H., Wang, H., Wu, F. (2020).** SKEP: Sentiment knowledge enhanced pre-training for sentiment analysis. arXiv preprint arXiv:2005.05635.

47. **Wikipedia (2022).** Arabic chat alphabet. en.wikipedia.org/wiki/Arabic_chat_alphabet. [Online].

48. **Zhang, L., Wang, S., Liu, B. (2018).** Deep learning for sentiment analysis: A survey. Wiley interdisciplinary reviews: data mining and knowledge discovery, Vol. 8, No. 4, pp. e1253.

Article received on 30/04/2025; accepted on 31/08/2025.
**Corresponding author is Abir Masmoudi.*