

## Sentiment Analysis Methods for Arabic Content on Social Media: A Systematic Review

Reem K AlMotairi<sup>1</sup> , Mohammed Hadwan<sup>1,2\*</sup> 

<sup>1</sup> Department of Information Technology, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia

<sup>2</sup> Department of Computer Science, College of Applied Sciences, Taiz University, Taiz 6803, Yemen

Corresponding Author Email: [M.hadwan@qu.edu.sa](mailto:M.hadwan@qu.edu.sa)

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.290138>

### ABSTRACT

**Received:** 8 June 2023

**Revised:** 16 October 2023

**Accepted:** 4 December 2023

**Available online:** 27 February 2024

#### Keywords:

*Arabic sentiment analysis, systematic review, social media, opinion analysis, deep learning and machine learning*

The topics of sentiment analysis either written or auditory texts are among important research areas in artificial intelligence (AI). The researchers in Natural Processing language (NLP) concerned more about the development of sentiment analysis methods and applications. For the available literature, the English Language has most research studies for the sentiment analysis. Arabic language is a membership in an entirely distinct language family than English explains why it needs the researchers to explore it from the scratch. Consequently, languages other than Arabic are more closely connected to one another than Arabic is. The Indo-European languages are utterly alien to the grammar, syntax, pronunciation, and lexicon. As a result, Arabic sentiment analysis has recently attracted the attention of the scientific researchers. This is due to the huge number of ideas and thoughts that are posted daily by social media users around the world. Manual processing of such huge data to obtain valuable information is an impossible task. The aim of this systematic review is to present a comprehensive review the major contributions in the field of Arabic sentiment analysis (ASA). The previous studies were primarily focused on dealing with certain sentiment analysis tasks, according to a comprehensive analysis of the accessible literature. The approaches found in the literature for ASA is classified into three main groups: (i) supervised, (ii) unsupervised, and (iii) hybrid. The literature's primary points include the fact that sentiment analysis in Arabic is difficult due to the language's complexity and wide variety of local dialects. These research findings, while intriguing, were not all in agreement. This difference is mostly attributable to the method chosen, the job being examined, besides the peculiarities and nuances of the Arabic diversity being studied. The evaluation of the literature revealed that, Naïve bayes (NB), K-Nearest Neighbour (KNN) and Support Vector Machine (SVM) are among the most popular classifiers applied to ASA. The lack of trusted Arabic data sets to allow the researchers examine the proposed ASA methods is among the main issues not yet solved. Therefore, this research can help the researchers to get updated about the literature related to Arabic sentiment analysis datasets and existed methods and techniques.

## 1. INTRODUCTION

Social network have rapidly expanded in recent years; therefore, social networks (SNs) have taken the globe by storm [1]. TikTok, Facebook, Twitter, and others are examples of popular SNs. These social media networks have proliferated throughout the world attracting a significant number of internet users to actively participate in global communication and collaboration.

SNs platforms are becoming more and more crucial than ever for the dissemination of ideas about everything [2, 3]. On SNs, members can utilize a variety of social data formats to share and express their thoughts and experiences.

These forms include textual data (like reviews, tweets, and comments), visual data (like liked and shared photos), and multimedia data (like sounds and movies) [3-5]. Sentiment analysis (SA) seeks to identify a group of people's views about a certain issue on one or more social media platforms [5]. For

decision-makers, corporate executives, and others, understanding public opinions and concerns voiced on these many platforms is a vital issue. According to the study of Duwairi and Qarqaz [6], SA is an automatic process that quickly ascertains the sentiment of sizable volumes of text or voice. It's important to note that SA goes beyond hashtag counts that many social media sites provide. This is due to the latter analyses post contents to create an overall mood rather than only counting hashtags users use.

Sentiment analysis (SA) is the science of studying and analysing how people respond to and accept an item (such as blogs, products, books, films, etc.). This is by utilizing computational power to analyse the text and algorithmic methods to identify whether the user's comments are positive, negative, or neutral. Furthermore, to analyse the emotions (e.g., angry, sad, happy) [6-9]. SA is often referred to as "Opinion Mining" (OM), which is a branch of Natural Language Processing (NLP). The OM objective is to extract

the emotions, sentiments, or more general opinions contained in a human-written text [10]. Therefore, SA has been a trend research subject in the realm of NLP because of its crucial role in comprehending community opinions and delivering meaningful opinion-based judgments [9].

Arabic language is among the languages widely utilized on SNs [3]. Around 422 million people speak Arabic natively or in one of its many dialects. However, little SA research has been conducted in Arabic compared to other major languages. One of the languages most widely utilized on social media is Arabic [3]. But there is a dearth of research on SA for the Arabic language, with most earlier studies concentrating on English. Language-wise, Arabic differs from English in several ways. The two languages have different grammatical and structural elements. Arabic has a rich and complicated morphology, as well as several dialects in addition to its standard form, which makes Arabic text analysis more difficult [11]. As a result, accurate representation of Arabic text is essential since its performance is highly dependent on it [7, 12]. The vocabulary's textual and phonetic organization makes the Arabic language distinct. Each word has a hidden or overt meaning, which may be either positive or negative. Arabic sentence analysis (ASA) is very important that's why this research is to review and analyze the existing research papers to see what issues exist that researchers can further explore.

Despite the recent expansion of publicly accessible Arabic information on social networks and the ongoing improvement of Arabic NLP tools, ASA research continues to encounter many difficulties. The majority of these difficulties are connected to the varieties of nature of Arabic as a language. The main Arabic language varieties: (i) the formal Modern Standard Arabic (MSA), that is utilized in the holy Quran, and (ii) the colloquial or Dialectal Arabic (DA), this blends numerous distinct dialects [4].

The researchers have highlighted the most difficult problems experienced during ASA:

- Complex morphology: because Arabic is a semitic language. It uses a representation of root-and-pattern, in which a specific set of consonants is used as the "root" and other words are created by vowels added (a, o, and i) or short vowels to it.
- Lack of resources: despite the wealth of Arabic information available online, sentiment datasets and lexicons are few.
- Negation and sarcasm: in Arabic, the word for "not" that expresses negation is a particular term. Negation should be correctly identified and managed since it can change a sentence's meaning and produce an entirely other polarity.
- Arabizi usage: Arabizi is a recently developed Arabic dialect written with Roman script letters and Arabic numerals.

- Variations across dialects: each dialect has its own vocabulary, syntactic and grammatical norms, as well as unique idioms. However, even though all tongues are descended from MSA thus share vocabulary, words in common or phrases across two-fold dialects may have vastly dissimilar meanings.

This research is arranged as follows: The methodology is introduced in Section 2. Results are shown in Section 3. To pinpoint study gaps and draw conclusions for future studies, Section 4 includes an analysis and discussion of the existed research work. Final the conclusion is presented in Section 5.

## 2. REVIEW Methodology

A systematic literature review (SLR) is used in this research as a method for locating and analysing the SA in Arabic language. Using methodical, precise, and strict criteria, a systematic review is conducted with the intention of not just summarizing the most recent research on the subject but also including some level of analytical commentary. In the subsections that follow, the specifics are provided. Figure 1 displays the followed methodology.

### 2.1 Research questions

The existed published research on sentiment analysis helped us to identify our research question as follows: RQ1. What is the research's present state? Who are the authors? When is it published?

- RQ2. What are the ASA approaches that work the best?
- RQ3. What are the examined studies' biggest flaws and restrictions?
- RQ4. What avenues should ASA research take in the future?

### 2.2 Search strategy

The search terms and data sources are provided to follow the determination of the research questions. The study topics were first examined to determine the search keywords. The following key-phrases were identified and used: "Arabic", "Arabic text", "Social media" and "sentiment analysis". Additionally, all conceivable combinations of these phrases, such as "sentiment analysis and social media," were searched for using the boolean operators "OR" and "AND". The second step was to do a search on the most common four scholastic databases: (i) Springer, (ii) IEEE Explore, (iii) ScienceDirect/Elsevier, and (iv) Google Scholar. These databases were selected because they are including the most of research available in the computer science. Using the selected search criteria, we searched all indexed publications' titles, abstracts, and keywords from 2014 through 2022.

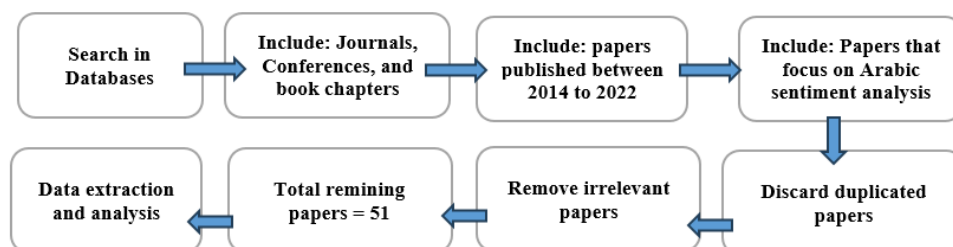


Figure 1. The research methodology

### 3. PRIMARY STUDIES AND DISCUSSION

The study of SA, examines how people feel about various entities, including goods, organizations, services, people, issues, events, themes, and their qualities. It is a significant and effective topic of computer science research [13]. At this stage, it is vital to understand the open ideas, thoughts, and questions offered, which is why SA is gaining huge interest [14].

SA is the practice of finding and retrieving certain views in a text using text analytics, computational linguistics, and NLP techniques. Its objective is to categorize a specified text into sentiment polarity, i.e., figure out whether the presented viewpoint is positive, negative, or neutral [15]. This part offerings and examines the key findings of the systematic review about the usage of SNs by ASA.

#### 3.1 Arabic sentiment analysis approaches

It is noticed that ASA approaches fall into three categories: (i) supervised learning-based, (ii) lexicon-based, and (iii) hybrid-based. We summarized the selected papers in tables based on the proposed categories; each of them is covered in its own subsection as follows.

##### 3.1.1 Supervised learning-based approaches

Several authors were used the supervised machine learning approach to analyse sentiments. In the literature, numerous methods have been proposed to approach the task of SA.

For example, in the study of AlSalman [16] the method that used to enhances the corpus-based ASA approach is the machine learning. The Naïve Bayes (NB) classifier was used with a 4-gram tokenizer, stemming, word frequency, and inverse document frequency (TF-IDF). The dataset contained about 2,000 Arabic tweets classified into two groups (negative and positive). The proposed NB classifier has an accuracy rate of 87.5%.

Abuuznien et al. [17] mainly concentrated on obtaining and examining the feeds of Sudan's SNs on the services of ridesharing. To assess the usefulness of four classifiers of machine-learning, they applied them to a 2,116 tweets dataset. The classifiers were NB, support vector machine (SVM), Logistic Regression, and K-nearest Neighbours (KNN). The SVM obtained the maximum accuracy level of 95%.

El-Masri et al. [7] employed two machine learning models on 8000 Arabic tweets. SVM and NB classifier were the methods used. Although the SVM classifier performed well, the NB results were excellent. NB had the largest improvement, with a 70% rate of accuracy.

Al-Tamimi et al. [8] used 5,986 Arabic comments that annotated manually of YouTube. They used the comments from the most popular Arabic-language videos to demonstrate the efficiency of proposed method. They used the SVM-RBF, Bernoulli NB, and KNN classifiers to carry out and compare several of supervised classification trials. Using their imbalanced 2-class (positive and negative) normalized dataset with connected and unconnected comments, SVM-RBF produced the greatest f-measure of 88.8%. They noticed that, particularly when utilizing bigger datasets. The pre-processing and normalization stages invariably enhanced the classification results.

SA of Arabic text has been considered by Duwairi and Qarqaz [6]. Were researchers used crowdsourcing. A dataset comprising 2591 tweets and comments was gathered and annotated. The polarity of a particular comment was determined using the NB, SVM, and KNN classifiers. The

information was divided into training and testing sets using a cross-validation of 10-fold. SVM was able to reach the best precision, which is 75.25%. The KNN (K = 10) model had the best recall, which is equivalent to 69.04%.

A detailed summary of previous research work is shown in Table 1.

**Table 1.** Previous research work summary related to ASA

Ref.	Algorithm/Classifiers	Dataset	Evaluation
[6]	NBs, SVM and KNN classifiers	2591 tweets/comm ents	Best: SVM acc=75.25
[8]	SVM, KNN and NP classifiers	8,053 comments	Best: SVM F1=88.8% Best: NB
[7]	NBss and SVM	8000 tweets.	NB=70% SVM=34%. Best: SVM SVM=95% KNN=71% NB=80% Logistic Regression=81%
[17]	NB, SVM, Logistic Regression and KNN	2116 tweets	acc=87.5%
[16]	DMNBTF-IDF	2000 Arabic tweets	Best: KNN KNN=78.46% DT=59.92% NB=54.78% SVM=55.38%
[18]	DTree, SVM, KNN, and NB	8000 user reviews	Best: SVM acc=75.25%
[6]	NBs, SVM and KNN-	2591 tweets/c omments	

##### 3.1.2 Approaches of Deep learning (DL)

DL models have become more common in NLP during the last several years [9, 19]. Artificial neural networks (ANNs) with several undetectable layers among the input and output layer make-up the Deep Neural Networks (DNNs).

Among the most well-known deep learning models are Generative Adversarial Networks (GANs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs).

RNNs is one of the Long Short-Term Memory Networks (LSTMs) which also covered in the study of Mulki et al. [9]. There are several studies on SA using various techniques and languages. ASA studies based on the SLR defined criteria is summarized in this section.

Heikal et al. [20] suggested an ensemble model based on DL that brought together two classifiers (i) CNN and (ii) LSTM models. Arabic tweet sentiments were predicted using the suggested methodology. They made use of the 10,000 tweets from the ASTD dataset. Which are divided into 4 groups (positive, negative, neutral, and objective). The LSTM models obtained the greatest accuracy, which was 64.75% with a loss rate of 0.2, while all other variables stayed constant. 65.05% was the ensemble model accuracy, while the accuracy of the CNN model was 64.30% in the regular setup. The CNN model employed a fully connected layer with 100 of a batch size, and this resulted in the best accuracy of 64.30%.

A book dataset of Arabic reviews recognized as LABR, or Large-Scale Arabic Book Reviews, was analysed using DL by Al-Bayati et al. [10]. 16,448 reviews made up the dataset, and either positive or negative results were anticipated. This article tested various LSTM output sizes with various batch sizes using an LSTM neural network. When the batch size was 256

and the LSTM output was 50, the best accuracy results were achieved at about 82%.

By categorizing Arabic tweets into 5 categories-"none, religion, racism, sexism, or general hatred" Al-Hassan and Al-Dossari [21] sought to identify hate speech in these tweets. The four DL models that were compared by the authors were LTSM, CNN + LTSM, GRU, and CNN + GRU. On a dataset of 11,000 tweets, these classifiers were used. Based on the obtained results, the CNN+LTSM achieved the best results with a success rate of 72%.

Cheng and Tsai [22] suggested a DL-based framework to deal with slang and other unique social languages. They used three deep learning models to analyse 40,000 Arabic tweets on various subjects. The LSTM, BiLSTM, and GRU techniques were used. The accuracy rate of LSTM, which had the biggest improvement, was 88.83%.

Ombabi et al. [3] suggested a DL model for ASA that expertly combined a CNN plan with 2-LSTM layers. The input layer of this plan is maintained by FastText word embedding. A multi-domain corpus demonstrated excellent performance in terms of precision, recall, F1-Score, and accuracy, scoring 89.10%, 92.14%, 92.44%, and 90.75%, respectively, based on the conducted trials. In the study of Mohammed and Kora [23], the authors presented a 40,000 tagged corpus of Arabic tweets covering a variety of themes. For ASA, they proposed three DL models: (i) CNN, (ii) LSTM, and (iii) RCNN. the researchers tested the effectiveness of the mentioned models on the suggested corpus via word embedding. According to the findings, LSTM surpassed CNN and RCNN with 81.3%, accuracy compared to 75.72% and 78.46%, respectively. The accuracy of the LSTM has increased to 88.05% after using a corpus-based data augmentation approach.

Hamed et al. [24] introduced an ASA corpus with 60K comments that they obtained from Facebook and classified as positive and negative. They put out a DL method that enables businesses and organizations to gauge how well their services are received by customers by examining those customers' Facebook comments.

With the use of word embedding, they provided three DL models for ASA: (i) convolutional neural networks (CNN), (ii) recurrent convolutional neural networks (RCNN), and (iii) long-term memory (LSTM). When the model's accuracy was evaluated, it is been clear that it had an accuracy average of 82.1%, that was superior to both CNN and RCNN. The accuracy of the LSTM is increased by 8.7% when that data is applied to the body.

In Alharbi et al. [12] study, the authors proposed a DL-based model based on SA to predict how strongly people feel about things. Higher-level representations are learned using two kinds of RNNs: (i) the Gated Recurrent Unit (GRU) and (ii) the LSTM. Then, three diverse classification techniques were used to create the result to reduce the data dependence issue and improve the model's resilience. based on the experimental findings, the model was accurate between 81.11 and 94.32% in all the datasets that were chosen. In addition, the model decreased the error rate of relative classification by 26% in comparison to leading-edge models. A summary of earlier studies employing the ML and DL technique is presented in Table 2 and Table 3.

### 3.1.3 Hybrid approach

The combined or hybrid approach makes use of both lexicon and machine-learning-based techniques. This technique is established in the relevant literature and is often

recognised to achieve better than the lexicon-based method and the machine-learning method by themselves [14]. The scores of lexical are often utilized as characteristics to feed into the classifier. Regarding lexicons, the study tested dialectical or informal Arabic (DA), modern standard Arabic (MSA), and MSA and DA together. SVM and NB were the most often used machine learning classifiers [15]. Yet other strategies, like entropy and KNN, were also used. A summary of earlier studies employing the hybrid technique is shown in Table 4.

**Table 2.** ML and DL ASA research works summary

Ref.	Algorithm/Classifiers	Dataset	Evaluation
[4]	CNN, LSTM and RCNN	40k Arabic tweets	Best: LSTM LSTM= 81.3%. CNN = 75.72% RCNN= 78.46%
[3]	Deep CNN-LSTM	15.100 reviews	acc=90.75%
[20]	Arabic-SA	10,000 tweets	CNN =64.30 LSTM=64.75
[10]	CNN and LSTM	LABR (16448) tweets	acc=82%
[24]	LSTM	60k Arabic comments	Best: LSTM LSTM= 82.1% CNN = 76.96% RCNN= 79.26%
[12]	CNN, LSTM and RCNN	1159K reviews and tweets	acc=94.32%
[21]	LSTM and GRU	11,000 tweets	Best: CNN+LSTM CNN+LSTM= 72% LSTM=71% GRU=69% CNN+GRU=71%
[22]	CNN+ LSTM, LSTM, GRU and CNN+GRU	40,000 tweets	Best: LSTM LSTM=88.83% BiLSTM=87.17% GRU=64.92%

**Table 3.** Summary of ML and DL ASA research works

Ref.	Algorithm/Classifiers	Dataset	Evaluation
[19]	NB, Logistic Regression, SVM and LSTM	32,000 tweets	Best: LSTM CVM = 63% LSTM=70%
[25]	NB, CNN-LSTM	58k tweets	Best: CNN-LSTM CNN-LSTM= 98% NB=87.6%
[26]	NB, DNN	11000 reviews from Google Play	NB=87.30% DNN=95.87%
[1]	MLP, BNB and SVM	17.000comments from Facebook	Best: MLP prec=78% recall=78%
[11]	LR, SVM, M-NB and Bagging	16.6k tweets	acc=93%

### 3.1.4 Approaches of a Lexicon-based

Lexicon-based approaches do not need labelled data or a training phase to create the sentiment classifier. The emotion scores of each word in a phrase or document that makes up that sentence or document are calculated using the lexicon. A sentiment lexicon includes a collection of subjective words and phrases with a positive or negative score. This indicates

the polarity and intensity of the feeling associated with each word or phrase. Sentiment lexicons may be created manually or mechanically, for broad or specialized purposes. The Straightforward Sum (SFS) approach and the Double Polarity (DP) method are two weighing methods that are utilized to decide the sentiment weight or score of each element in the Lexicon [9].

**Table 4.** Hybrid ASA summary of previous research works

Ref.	Hybrid features	Algorithm/classifier	Dataset	Evaluation
[27]	N-grams, polarity of token polarity of tweet and Arabic Senti-Lex	SVM NB	1093 Positive tweets 978 Negative tweets	acc=96%
[28]	Bing Lui, AFINN, and MPQA lexicons	RF bagging SVM LR NB	51k reviews	SVM: acc=94.3
[29]	N-grams sentence-level syntactic score from Arabic Senti-Lex	SVM, NB, LLR, KNN	8860 positive negative reviews	F1=97%
[30]	Stemmed word uni-grams and bi-grams, starts with Link, ends With Link and num Of Pos, num Of Neg, length and segments.	SVM, CNB MNBU	28,132 tweets	acc= 85.03%

According to the study by Duwairi et al. [30], when working with MSA data, there is a big chance of discovering a stem in the emotion lexicon than there is of discovering the original word. This has been looked at using a manually created MSA sentiment lexicon, as described in Section 4.1. To examine the proposed algorithm, 4,400 positive and negative tweets were manually gathered and analysed. Stopwords were eliminated from the data during pre-processing, but negations were left in. The MSA Khoja stemmer8 was used to stem the input data. Experiments were done with and without stemming to investigate the effect. The input tweets sentiment score was computed using the SFS technique with a switch negation strategy. The findings showed that stemming enhanced sentiment classification ability for such MSA data, with accuracy increasing from 23% to 46% and the F1-score rising from 31.3% to 55.51%.

Al-Ghaith [31] introduced the Saudi-Dialect Sentiment Lexicon (SaudiSentiPlus). Which includes 7139 terms used to analyse the sentiment of tweets in Saudi-dialect. Additionally, provide a novel approach based on the presentation of two lexicon-based methods to handle the suffixes and prefixes of lexicon words. This technique significantly improves the functionality or accuracy of the (SaudiSentiPlus) lexicon. Every step included a measurement of the F-Score, recall, precision, and accuracy. The authors concentrated on Saudi dialect hashtags while creating their testing sample from Twitter (971 thousand tweets from 162 hashtags) were used. The dataset's tweets were divided into three groups (positive, negative, and neutral), as shown in the assessment section. The findings demonstrate that SaudiSentiPlus' accuracy with the two lexicon-based method accomplished 81%.

Assiri et al. [32] presented a polarity weighting technique called WLBA. This technique gives weights to the polarity

terms by learning from the data itself to derive the sentiment score. This technique analyses and counts the frequency of a pair of (polarity and non-polarity) words co-occurring while considering the context of the polarity terms. The polarity word is later assigned a weight based on the number of connections it has with the non-polarity term across the entire corpus. Utilizing methods based on corpuses and dictionaries, a Saudi lexicon was created (see Section 4.1). WLBA performed poorly matched to SFS and DP for used datasets after applying the model. This is due to the ignorance of the intricate structural and specifications of lexical of the Saudi corpus. Though, WLBA surpassed other approaches with 81% accuracy of, compared to 72% and 43% earned by SFS and DP methods, respectively. Additionally, the accuracy attained for the Egyptian dataset was 76%, compared to the accuracy scores of 71% and 68% obtained using the SFS and DP methods, respectively.

#### 4. CHALLENGES FOR THE ANALYSIS OF ARABIC SENTIMENTS

Even Nevertheless, public Arabic material on social media has increased recently there are still difficulties with ASA research because of the difficult structure of Arabic language itself. Conducting sentiment analysis research presents several obstacles; we will focus this section just on those challenges of which remain unsolved. This provides opportunities for future academics to address these issues, discover solutions, and improve the accuracy of ASA.

##### 4.1 Extracting dialectal Arabic features

Arabic text is written from right to left, unlike Latin, and is differentiated by the lack of upper- or lower-case letters. Only 28 letters that make up Arabic language alphabet, 25 of which are consonants and just 3 are vowels [33]. Arabic script also uses markings of diacritical as short-vowels in addition to these vocal parts. The bulk of texts lack diacritical markings, causes challenges to the lexical ambiguity for computer techniques [34]. Informal dialectal Arabic (DA) is notoriously unstructured and hard to standardize, making it tough to analyse. It differs from MSA phonologically, morphologically, and syntactically, which makes morphological analysers and POS taggers exceedingly difficult to use. Different dialects of MSA may use different negations and stop words. Building lexicons for many dialects is extremely challenging since ideas have no similar lexical options in unlike dialects [15, 34].

##### 4.2 Use of Arabizi

Arabizi, or the Latin letters use to show Arabic words, is a currently trend on SNs. It can be challenging to tell if a term written in Latin letters is Arabic or English when posted by Arabic social media users since they frequently code-switch between Arabic and English in writing [15, 35]. This issue has not yet been focussed in the literature on ASA.

##### 4.3 Named Entity Recognition (NER)

In Arabic, positive adjectives are seen in significant numbers of Arabic names. Therefore, names in Arabic that are formed from Arabic adjectives, such as, should be identified since the classifier could mistake them for attitudes if they are

not [15, 35]. For example, Jamila (جميلة), which means beautiful in Arabic [36], the name “سعيد” links to the adjective “سعيد” which means “happy” [36]. Additionally, not capitalization to Arabic proper nouns like they are in Latin languages, makes it more difficult to identify things. To analyse Arabic texts and discriminate between entity names and emotion words, a named entity recognition system is essential [36].

#### 4.4 Handling compound phrases and idioms

Arabic writing on social media frequently uses idioms and compound phrases [15]. Arabic speakers frequently express their ideas using well-known idioms and complex words. The issue here is these idioms and phrases include hidden views that are difficult to be picked up by a sentiment tool if they are not present in the training dataset or the sentiment lexicons being used [36].

#### 4.5 Sentiment lexicons shortage

Even though numerous academics have created sentiment lexicons for the Arabic, they didn't make it accessible to the public [36]. AbdulMajeed and Diab [35] have recently suggested an effort to create an extensive multi-genre, multi-dialect Arabic sentiment lexicon. Thus, it supports only the Levantine and Egyptian dialects, and it has not yet been completely integrated into SSA tasks.

#### 4.6 Lack of corpora and datasets

The 5<sup>th</sup> widely used language in the world is Arabic; however, there aren't numerous Arabic reviews or online resources. There are fewer datasets available to do sentiment analysis on than there are for the English language. Because sentiment analysis accuracy based on the volume of data, it is challenging to compare performance between languages [15]. Any SSA system's accuracy is dependent on the accessibility of sizable, annotated corpora, that are currently a limited resource for Arabic [35].

### 5. CONCLUSIONS

In this research, ASA was investigated and critically analysed. The key ASA studies found in the literature was gathered and reviewed. It is clearly that, the NLP research community has become more interested in sentiment analysis for Arabic. According to existed research, it is obvious that three types of techniques are employed for the sentiment classification task which are: (i) lexicon-based, (ii) corpus-based, and (iii) hybrid. Machine learning algorithms are the major focus of corpus-based and hybrid techniques. The most popular algorithms for Arabic are NB, SVM, and KNN. As compared to English sentiment analysis, deep learning has not yet been extensively studied. Researchers have exposed that; the sentiment sources quality has a significant influence on the performance. Because of their dialectical character, Arabic sentiment resources currently available are undeserving for SNs analysis. Most of the researchers concluded that, the results for other types of data are encouraging. To be effective, though, several issues must be resolved. The researchers suggested looking at the dialectical content on social-media in more details to create complete Arabic resources that consider both Arabizi and dialects. In addition, the researchers

suggested to moving away from straightforward word-level SA toward concept-based and extra investigating word embedding to manage the Arabic language complexity.

Instead, it is found that, aspect-based SA, opinion-holder extraction, spam identification, and issues with domain dependency have not well studied while examining works that have caught the interest of scholars. To improve the literature in the SA field, it is believed that this systematic review will give academics a thorough overview of ASA in terms of resources, methods, and unresolved difficulties.

### ACKNOWLEDGMENT

The author(s) gratefully acknowledge Qassim University, represented by the Deanship of Scientific Research, on the financial support for this research under the number (COC-2022-1-3-J- 31192) during the academic year 1444 AH / 2022 AD.

### REFERENCES

- [1] Mdhaffar, S., Bougares, F., Esteve, Y., Hadrich-Belguith, L. (2017). Sentiment analysis of Tunisian dialects: Linguistic resources and experiments. In Third Arabic Natural Language Processing Workshop (WANLP), pp. 55-61.
- [2] El-Beltagy, S.R., Khalil, T., Halaby, A., Hammad, M. (2018). Combining lexical features and a supervised learning approach for Arabic sentiment analysis. *Computational Linguistics and Intelligent Text Processing*, 9624: 307-319. [https://doi.org/10.1007/978-3-319-75487-1\\_24](https://doi.org/10.1007/978-3-319-75487-1_24)
- [3] Ombabi, A.H., Ouarda, W., Alimi, A.M. (2020). Deep learning CNN-LSTM framework for Arabic sentiment analysis using textual information shared in social networks. *Social Network Analysis and Mining*, 10: 1-13. <https://doi.org/10.1007/s13278-020-00668-1>
- [4] Alsayat, A., Elmitwally, N. (2020). A comprehensive study for Arabic sentiment analysis (challenges and applications). *Egyptian Informatics Journal*, 21(1): 7-12. <https://doi.org/10.1016/j.eij.2019.06.001>
- [5] Zahidi, Y., El Younoussi, Y., Al-Amrani, Y. (2021). A powerful comparison of deep learning frameworks for Arabic sentiment analysis. *International Journal of Electrical & Computer Engineering*, 11(1): 745-752. <http://doi.org/10.11591/ijece.v11i1.pp745-752>
- [6] Duwairi, R.M., Qarqaz, I. (2014). Arabic sentiment analysis using supervised classification. In 2014 International Conference on Future Internet of Things and Cloud, Barcelona, Spain, pp. 579-583. <https://doi.org/10.1109/FiCloud.2014.100>
- [7] El-Masri, M., Altrabsheh, N., Mansour, H., Ramsay, A. (2017). A web-based tool for Arabic sentiment analysis. *Procedia Computer Science*, 117: 38-45. <https://doi.org/10.1016/j.procs.2017.10.092>
- [8] Al-Tamimi, A.K., Shatnawi, A., Bani-Issa, E. (2017). Arabic sentiment analysis of YouTube comments. In 2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), Aqaba, Jordan, pp. 1-6. <https://doi.org/10.1109/AEECT.2017.8257766>
- [9] Mulki, H., Haddad, H., Babaoğlu, I. (2017). Modern

- trends in Arabic sentiment analysis: A survey. *Traitement Automatique des Langues*, 58(3): 15-39.
- [10] Al-Bayati, A.Q., Al-Araji, A.S., Ameen, S.H. (2020). Arabic sentiment analysis (ASA) using deep learning approach. *Journal of Engineering*, 26(6): 85-93. <https://doi.org/10.31026/j.eng.2020.06.07>
- [11] Husain, F., Al-Ostad, H., Omar, H. (2022). A weak supervised transfer learning approach for sentiment analysis to the Kuwaiti dialect. In *Proceedings of the The Seventh Arabic Natural Language Processing Workshop (WANLP)*, Abu Dhabi, United Arab Emirates, pp. 161-173. <https://doi.org/10.18653/v1/2022.wanlp-1.15>
- [12] Alharbi, A., Kalkatawi, M., Taileb, M. (2021). Arabic sentiment analysis using deep learning and ensemble methods. *Arabian Journal for Science and Engineering*, 46: 8913-8923. <https://doi.org/10.1007/s13369-021-05475-0>
- [13] Guellil, I., Azouaou, F., Mendoza, M. (2019). Arabic sentiment analysis: Studies, resources, and tools. *Social Network Analysis and Mining*, 9: 1-17. <https://doi.org/10.1007/s13278-019-0602-x>
- [14] Alhumoud, S.O., Al Wazrah, A.A. (2022). Arabic sentiment analysis using recurrent neural networks: A review. *Artificial Intelligence Review*, 55(1): 707-748. <https://doi.org/10.1007/s10462-021-09989-9>
- [15] El-Masri, M., Altrabsheh, N., Mansour, H. (2017). Successes and challenges of Arabic sentiment analysis research: A literature review. *Social Network Analysis and Mining*, 7: 1-22. <https://doi.org/10.1007/s13278-017-0474-x>
- [16] AlSalman, H. (2020). An improved approach for sentiment analysis of Arabic tweets in twitter social media. In *2020 3rd International Conference on Computer Applications & Information Security (ICCAIS)*, Riyadh, Saudi Arabia, pp. 1-4. <https://doi.org/10.1109/ICCAIS48893.2020.9096850>
- [17] Abuuznien, S., Abdelmohsin, Z., Abdu, E., Amin, I. (2021). Sentiment analysis for Sudanese Arabic dialect using comparative supervised learning approach. In *2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*, Khartoum, Sudan, pp. 1-6. <https://doi.org/10.1109/ICCCEEE49695.2021.9429560>
- [18] Hadwan, M., Al-Hagery, M., Al-Sarem, M., Saeed, F. (2022). Arabic sentiment analysis of users' opinions of governmental mobile applications. *Computers, Materials and Continua*, 72(3): 4675-4689.
- [19] Alshutayri, A., Alamoudi, H., Alshehri, B., Aldahri, E., Alsaleh, I., Aljojo, N., Alghoson, A. (2022). Evaluating sentiment analysis for Arabic Tweets using machine learning and deep learning. *Romanian Journal of Information Technology & Automatic Control/Revista Română de Informatică și Automatică*, 32(4): 7-18. <https://doi.org/10.33436/v32i4y202201>
- [20] Heikal, M., Torki, M., El-Makky, N. (2018). Sentiment analysis of Arabic tweets using deep learning. *Procedia Computer Science*, 142: 114-122. <https://doi.org/10.1016/j.procs.2018.10.466>
- [21] Al-Hassan, A., Al-Dossari, H. (2022). Detection of hate speech in Arabic tweets using deep learning. *Multimedia Systems*, 28(6): 1963-1974. <https://doi.org/10.1007/s00530-020-00742-w>
- [22] Cheng, L.C., Tsai, S.L. (2019). Deep learning for automated sentiment analysis of social media. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 1001-1004. <https://doi.org/10.1145/3341161.3344821>
- [23] Mohammed, A., Kora, R. (2019). Deep learning approaches for Arabic sentiment analysis. *Social Network Analysis and Mining*, 9: 1-12. <https://doi.org/10.1007/s13278-019-0596-4>
- [24] Hamed, S., Ezzat, M., Hefny, H. (2022). Augmented Deep Learning Model for Social Network Sentiment Analysis. *International Journal of Intelligent Systems and Applications in Engineering*, 10(4): 246-255.
- [25] Suleiman, D., Odeh, A., Al-Sayyed, R. (2022). Arabic sentiment analysis using Naïve Bayes and CNN-LSTM. *Informatica*, 46(6). <https://doi.org/10.31449/inf.v46i6.4199>
- [26] Sari, S., Kalender, M. (2021). Sentiment analysis and opinion mining using deep learning for the reviews on Google play. *Innovations in Smart Cities Applications*, 183: 126-137. [https://doi.org/10.1007/978-3-030-66840-2\\_10](https://doi.org/10.1007/978-3-030-66840-2_10)
- [27] Mustafa, H.H., Mohamed, A., Elzanfaly, D.S. (2017). An enhanced approach for Arabic sentiment analysis. *International Journal of Artificial Intelligence and Applications (IJAIA)*, 8(5): 1-14. <https://doi.org/10.5121/ijaia.2017.8501>
- [28] Hadwan, M., Al-Sarem, M., Saeed, F., Al-Hagery, M.A. (2022). An improved sentiment classification approach for measuring user satisfaction toward governmental services' mobile apps using machine learning methods with feature engineering and SMOTE technique. *Applied Sciences*, 12(11): 5547. <https://doi.org/10.3390/app12115547>
- [29] Al-Moslmi, T., Albared, M., Al-Shabi, A., Omar, N., Abdullah, S. (2018). Arabic senti-lexicon: Constructing publicly available language resources for Arabic sentiment analysis. *Journal of Information Science*, 44(3): 345-362. <https://doi.org/10.1177/0165551516683908>
- [30] Duwairi, R.M., Ahmed, N.A., Al-Rifai, S.Y. (2015). Detecting sentiment embedded in Arabic social media—a lexicon-based approach. *Journal of Intelligent & Fuzzy Systems*, 29(1): 107-117. <https://doi.org/10.3233/IFS-151574>
- [31] Al-Ghaith, W. (2019). Developing lexicon-based algorithms and sentiment lexicon for sentiment analysis of Saudi Dialect Tweets. *International Journal of Advanced Computer Science and Applications*, 10(11): 83-88.
- [32] Assiri, A., Emam, A., Al-Dossari, H. (2018). Towards enhancement of a lexicon-based approach for Saudi dialect sentiment analysis. *Journal of Information Science*, 44(2): 184-202. <https://doi.org/10.1177/0165551516688143>
- [33] Alqurashi, T. (2023). Arabic sentiment analysis for Twitter data: A systematic literature review. *Engineering, Technology & Applied Science Research*, 13(2): 10292-10300. <https://doi.org/10.48084/etasr.5662>
- [34] Boudad, N., Faizi, R., Thami, R.O.H., Chiheb, R. (2018). Sentiment analysis in Arabic: A review of the literature. *Ain Shams Engineering Journal*, 9(4): 2479-2490. <https://doi.org/10.1016/j.asej.2017.04.007>
- [35] Abdul-Mageed, M., Diab, M., Kübler, S. (2014). SAMAR: Subjectivity and sentiment analysis for Arabic social media. *Computer Speech & Language*, 28(1): 20-

37. <https://doi.org/10.1016/j.csl.2013.03.001>  
[36] Al-Twairesh, N., Al-Khalifa, H., Al-Salman, A. (2014). Subjectivity and sentiment analysis of Arabic: trends and challenges. In 2014 IEEE/ACS 11th International

Conference on Computer Systems and Applications (AICCSA), Doha, Qatar, pp. 148-155. <https://doi.org/10.1109/AICCSA.2014.7073192>