International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(15s) 2026 pp. 70-87
https://doi.org/10.70917/ijcisim-2026-4349
Article

Arabic Sentiment Analysis (ASA): A Comprehensive Survey of Approaches, Challenges, and Future Directions

Ola Adnan Altiti[1]

[1]Engineering and Artificial Intelligence Department, Al-Balqa Applied University, Jordan
olatiti@bau.edu.jo

Abstract: Arabic sentiment analysis (ASA) has received increasing attention due to the rich availability of Arabic content across online platforms. However, the Arabic language presents several issues including rich morphology, dialectal diversity, and complex grammatical structures. This survey introduced a comprehensive systematic review of ASA research from 2018 to 2025, analyzing over 70 peer-reviewed studies. We examine three main methodological approaches:(1) traditional machine learning approaches, (2) deep learning techniques, and (3) transformer-based architectures including AraBERT, MARBERT, and CAMeLBERT and emerging aspect-based sentiment analysis (ABSA) research. Additionally, we provide a novel comparative analysis of a set of publicly available datasets and NLP tools. We identify four critical challenges: data scarcity, code-switching, dialectal generalization and evaluation standardization. Finally, we propose future directions including cross-lingual transfer learning, multimodal sentiment analysis, domain-specific ASA applications. The survey provides researchers with a structured roadmap for improving Arabic sentiment analysis systems.

Keywords: Arabic language, machine learning, deep learning, transformer-based architectures, Arabic Sentiment Analysis (ASA)

1. Introduction

Recently, Arabic Sentiment Analysis (ASA) has become a crucial research area in the Natural Language Processing (NLP) field due to the exponential growth of Arabic data on social networking sites and news articles [1][2][3]. Nevertheless, the Arabic language poses challenges because of its rich inflectional system, orthographic variations, and the diverse regional dialects [4][5][6]. All these linguistic characteristics make sentiment analysis in Arabic far more complicated compared to other languages such as English [7] [8]. Therefore, there is an increasing demand for sophisticated sentiment analysis systems specifically for the Arabic language. Over the past few years, researchers have progressively developed systems to perform sentiment analysis on texts, moving from lexicon-based approaches, which relied on manually constructed sentiment dictionaries, through to machine learning models and further to deep learning methods using convolutional and recurrent architectures [2][67].

More recently, transformer-based architectures such as MARBERT, AraBERT and multilingual BERT have shown incredible performance by capturing contextual and semantic features across dialects [9][11]. In parallel, Aspect-Based Sentiment Analysis (ABSA) has developed into a distinct subfield, enabling a more nuanced evaluation of user opinions on aspects or features [11][14].

Researchers have paid attention to developing sentiment analysis systems for Arabic. However, ASA is still a challenging task as it requires complex features of the Arabic language such as rich morphology [12][34]. Moreover, the presence of multiple dialects complicates the task of extracting sentiment [13][17]. Moreover, due to the rapid development of methodologies, the abundance of newly created resources and the emergence of domain-specific sentiment applications (from healthcare to politics), there is an urgent need for a comprehensive survey that systematically reviews the existing approaches, benchmarks, tools and challenges. The difference between this survey and earlier reviews is that:

- offering the first comprehensive analysis integrating traditional, deep learning, and transformer approaches within a unified framework
- providing an up-to-date comparative evaluation of recent transformer models (AraBERT, MARBERT, CAMeLBERT, QARiB, AraELECTRA) on standard benchmarks
- Systematically reviewing ABSA research with quantitative performance comparisons.
- Identifying emerging trends including multimodal sentiment analysis and cross-lingual transfer learning
- Providing a comprehensive review of datasets and tools with accessibility information

The remainder of this survey is organized as follows: section II provides a review of Arabic language characteristics and ASA, section III provides a literature review of the ASA approaches, section IV summarizes the recent surveys regarding ASA and provides a comprehensive comparison, section V shows the available datasets and tools, section VI discusses challenges as well as future directions, and finally, the section VII concludes the work.

2. background

Nowadays, Arabic Sentiment Analysis (ASA) has received serious attention in the research community owing to the fast growth of Arabic content on online platforms and social media sites. Therefore, understanding the features of the Arabic language as well as the methods and techniques used in ASA is an essential step for developing effective systems. This section illustrates Arabic language Characteristics and highlights the ASA research.

A. Arabic Language Characteristics

Arabic is one of the most widely spoken languages in the world, ranking 4th in the world with more than 362 million Native speakers [14]. It is the official language of more than 22 countries. Arabic holds importance in cultural, religious and political domains on the Arab world. It is used in digital content, online platforms, and social media and continues to grow rapidly. Moreover, it is among the six official languages of the United Nations, which makes it more significant worldwide. Furthermore, the Arabic alphabet is composed of 28 fundamental letters. In addition to these letters, the orthographic system employs a set of diacritical marks (tashkeel) to indicate short vowels (a, i, and u) [3].

Modern Standard Arabic (MSA) is offered as the formal variant in media, literature and education, while several dialects dominate informal communication especially in digital spaces such as Levantine, Gulf, and Egyptian [15]. Thus, Arabic can be divided into various forms [3][17][18]:

- Classical Arabic (CA): The formal language used in religious and classical literature. It is rarely used on social media.
- Modern Standard Arabic (MSA): The formal language of books, news and the press; widely used in NLP research.
- Dialectal Arabic (DA): Regional spoken varieties (Egyptian, Levantine, Gulf, Maghrebi, etc.) that differ from MSA in terms of syntax, vocabulary and phonology. TABLE I. illustrates how the same word can differ in various dialects of Arabic.

Table 1. Examples of the same word in different Arabic dialects.

Word	Maghrebi	Levantine	Egyptian	Gulf
سعيد جدًا (very happy)	فرحان بزاف	مبسوط كتير	فرحان أوي	مستانس وايد
ممتاز(Excellent)	زوين	منيح	جامد	زين
سيئ (bad)	خايب	سيء	وحش	شين

Fig. 1. presents the classification of Arabic language varieties, showing the division into Modern Standard Arabic (MSA), Classical Arabic (CA), and Dialectal Arabic (DA). The figure also provides examples of Arabic dialects, including Egyptian, Levantine, and Gulf. TABLE II. shows the same sentence in different Arabic dialects, which illustrates the linguistic variety that presents challenges for Arabic Natural Language Processing (NLP) and sentiment analysis tasks.

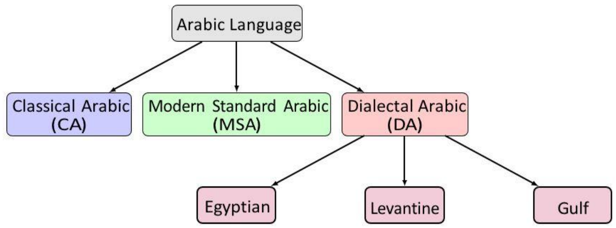

Fig 1. Classification of Arabic language varieties

Table II. Examples of the same sentence in different Arabic varieties

Arabic Form	Sentence Example
Classical Arabic (CA)	أريد أن أذهب إلى المدرسة
Modern Standard Arabic (MSA)	أريد الذهاب إلى المدرسة
Egyptian Dialect (EGY)	عايزة اروح المدرسة
Levantine Dialect (LEV)	بدي روح المدرسة
Gulf Dialect (GLF)	أبغى اروح المدرسة
Maghrebi Dialect (MAG)	بغيت نمشي المدرسة

B. Arabic Sentiment Analysis (ASA)

Sentiment Analysis (SA) is a subfield of NLP that enables the extraction of opinions, emotions, and attitudes from textual data [2][3][8][20]. The goal of SA is to detect the sentiment that exists in a segment of text, whether it is positive, negative or neutral [18]. Furthermore, SA can be studied at different granularity levels, each addressing specific aspects of opinion mining [14][21][22][23]:

- The document level: the entire text is performed as a single unit of opinion, where the task is to classify the sentiment of the whole document, whether it's positive, negative, or neutral [2][21]. While effective for longer reviews, such as articles, this level often leads to mixed sentiments toward different aspects within the same text.
- The sentence level:analyze the individual sentences, allowing for more precise sentiment identification, though it still struggles with sentences expressing ambiguous sentiments [12].
- Aspect-based Sentiment Analysis (ABSA): sentiments are identified for the features or aspects mentioned in the text [11][14]. This is useful in domains such as service reviews,

where users may give opinions on several aspects. Fig. 2. illustrates the three main levels of sentiment analysis: document-level, sentence-level, and aspect-level classification. At the document level, a single label is assigned to the entire text. At the sentence level, each sentence is classified independently according to its sentiment. At the aspect level, sentiment is identified to aspects or objects in the text (e.g., price or service), enabling fine-grained polarity detection even when multiple sentiments coexist in the same.

The development of Arabic sentiment analysis systems is essential due to the widespread availability of digital content. ASA aims to automatically detect opinions, attitudes, and emotions from Arabic texts [21]. However, Arabic language is complicated with many different dialects, making the

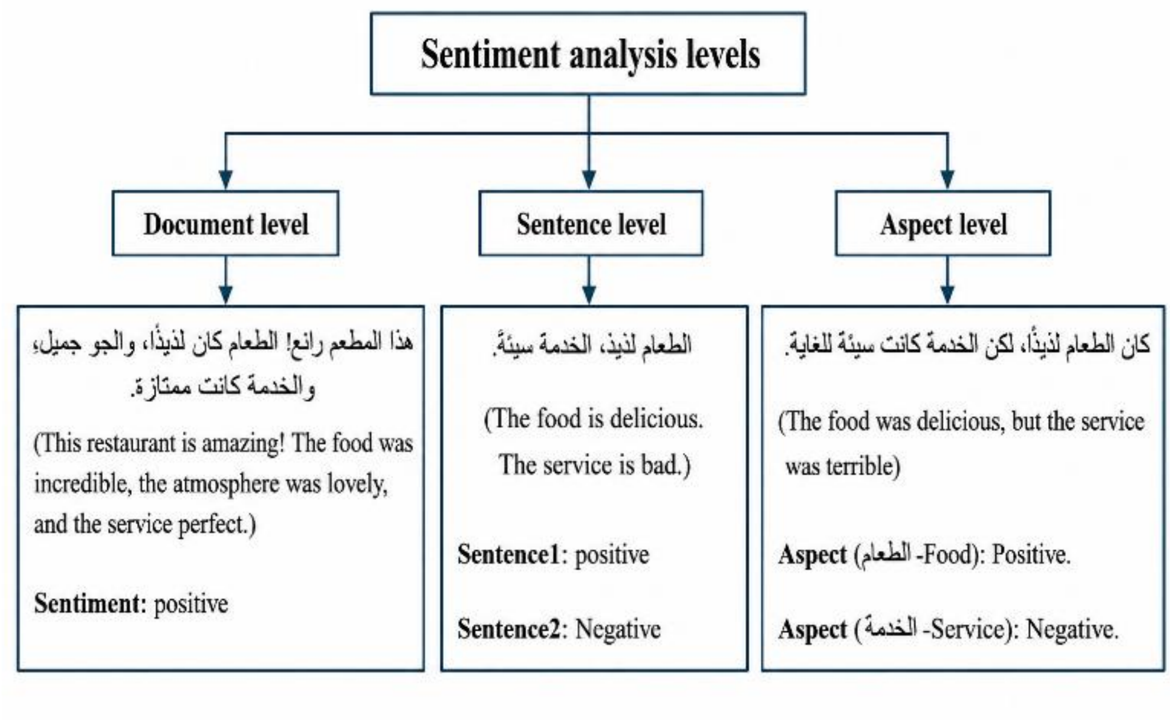

Fig. 2. Examples of sentiment analysis levels in Arabic.

development of sentiment analysis systems difficult. Hence, several ASA gained significant interest because of the advances in the NLP domain. Early approaches utilized traditional machine learning models [20] such as Naive Base (NB) [26], and Support Vector Machines (SVM) [27] with a set of pre-processing steps and feature engineering [28]. Furthermore, deep learning classifiers, including Convolutional Neural Networks (CNNs) [41] and Long Short-Term Memory (LSTM) networks [23], improved the ASA task [20]. Later, transformer-based models such as AraBERT [9] and MARBERT [10] enhanced the performance of such systems. Recent models, such as QARiB [52] and AraELECTRA [16], have applied advanced pretraining strategies to further enhance classification performance. Pre-processing steps such as diacritics removal, normalization and tokenization remain critical for robust ASA. More details regarding the approaches provided in section III.

To significantly improve sentiment analysis accuracy and robustness, we should shed light on Arabic linguistic understanding in ASA research. Furthermore, combining information on morphology, orthography, and dialectal variation [16]. Therefore, deep insights into the Arabic language remain a cornerstone for the development of advanced ASA systems [67].

Fig. 3. illustrates the process of ASA, from data collection to sentiment output. Several steps included in the process are as follows:

Data Collection: The first step is to collect Arabic text from several sources such as forum comments, news articles or social media. The diversity and quality of the data collected are very important for the success of the analysis, which ideally covers both Modern Standard Arabic and different dialects [20] [23].

Text Pre-processing: After data collection, pre-processing steps are applied to clean and normalize the text for analysis. This includes removing diacritics, punctuation, and stop-words, as well as handling special symbols and hashtags. These steps reduce noise in the data and enhance the performance of models.

Feature Extraction: the text is converted into numeric representations using techniques such as TF-IDF, Bag of Words (BoW) and Word Embeddings (e.g. Word2Vec). Feature extraction encodes linguistic and contextual information for the successful classification of sentiment.

Sentiment Classification: After the generation of text features, the classification of sentiment is done using the models of machine learning or deep learning. The most popular models are Support Vector Machines (SVM), LSTM and transformer-based models such as AraBERT and BERT. These models are trained in classifying the sentiment of Arabic text into positive, negative and neutral classes.

Output: The last stage contains the sentiment predictions for each segment of Arabic text. These outputs are valuable for many applications, such as social media opinion mining and user feedback analysis.

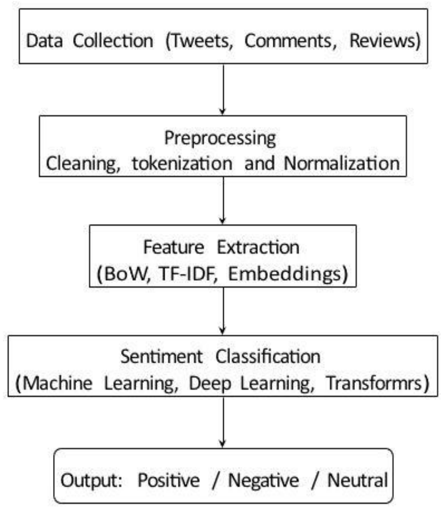

Fig. 3. The process of Arabic Sentiment Analysis (ASA) from input text to sentiment classification

3. LITERATURE REVIEW

Nowadays, the technology revolution has made it possible for digital content to grow exponentially on social media, online forums and review websites, which have enabled opinion-rich content (e.g., status updates and comments) to become

increasingly abundant.[4]. The rapid expansion of Sentiment Analysis (SA) research has been driven by the dissemination of user-generated text and the need to efficiently process and interpret large volumes of such data. Since the early 2000s, SA has been one of the most prominent areas within Natural Language Processing (NLP). The earliest major academic work in the English language dedicated to sentiment analysis (SA) is attributed to [18]. English is getting the most attention because of the extensive availability of large-scale, high-quality datasets [19][3]. In contrast, ASA has progressed more slowly, primarily due to the shortage of diverse datasets encompassing Modern Standard Arabic (MSA) and its various dialects. Nevertheless, the fast growth of Arabic content on digital sites has led to a significant interest in ASA. Hence, a vast number of studies have considered various approaches for sentiment classification including traditional machine learning methods, deep learning techniques and transformer-based architecture. The timeline diagram in Fig. 4. Overviews the evolution of ASA techniques, starting from lexicon-based approaches, through machine learning, deep learning and transformer architectures. The inclusion of Aspect-Based Sentiment Analysis (ABSA) as a dedicated stage reflects the field's evolution from general polarity classification to fine-grained opinion mining. This development shows how methodological development has progressively corresponded with the linguistic complexity of Arabic, enabling more accurate sentiment interpretation. This section provides a review of different approaches used to handle ASA.

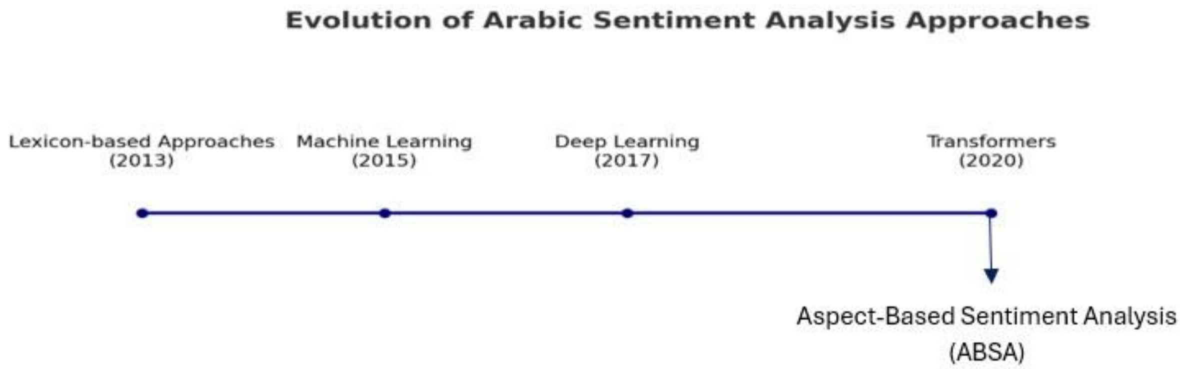

Fig. 4. Evolution of Arabic Sentiment Analysis approaches, showing the transition from lexicon-based approaches to transformers, and the emergence of Aspect-Based Sentiment Analysis (ABSA).

A. Lexicon-Based Approaches

The earliest research in ASA relied on manually crafted or semi-automatically expanded sentiment lexicons, where words were manually or semi-automatically annotated with polarity scores [24]. A notable example is AraSenTi proposed by [15], which provided one of the largest sentiment lexicons for Arabic, covering both MSA and dialects. AraSenTi introduced a large-scale Arabic sentiment lexicon for Twitter data. The authors automatically developed the lexicon and evaluated its effectiveness through both intrinsic (in-domain) and extrinsic (cross-domain) tasks. Additionally, the authors in [24] created CNSAMSASAT, a sentiment analysis tool made especially for MSA and colloquial Arabic, by creating polarity lexicons with positive, negative, and neutral classifications using manually annotated comments from social media. Other studies built a multi-lingual lexicon, for example [7], addressed the scarcity of sentiment lexicons by building high-quality resources for 136 major languages. The study integrated multiple linguistic sources into a large-scale knowledge graph and propagated sentiment from seed words to build lexicons for each language. In [25], they provided a hybrid framework for sentiment analysis in Modern MSA and Egyptian dialect, combining textual content with non-verbal signals such as emojis. Plutchik's Wheel of Emotions has been employed to build an emoji-based lexicon, which was used as a source for capturing non-verbal emotional cues. Experimental results showed that this approach yielded an accuracy of approximately 90%. However, while lexicon-based systems were central to the adaptation of SA for Arabic, they suffered from problems with morphology and were not efficient in handling orthographic and dialectal variations, which are prevalent in social media platforms [14].

B. Machine Learning (ML) Approaches

With the advent of supervised Machine Learning (ML), the SA field has shifted from lexicon-driven approaches to data-driven approaches. Early works used traditional classifiers such as Naive Bayes (NB) [26], Support Vector Machines (SVM) [27], Decision Tree (DT), and Logistic Regression (LR), which became widely used in this domain due to their robustness in handling high-dimensional text representations [67]. These classifiers were primarily applied on features, including Bag-of-Words (BoW) representations that capture the distribution of word frequencies, Part-of-Speech (POS) tags to encode syntactic information, and Term Frequency-Inverse Document Frequency (TF-IDF) to weigh the importance of the word across documents. Additionally, [22] approached ASA from a machine learning perspective using NB, SVM, and K-Nearest Neighbour (KNN) on a dataset of comments and tweets generated with the RapidMiner tool. Experimental results imply that SVM achieved the highest precision, whereas KNN (K=10) yielded the highest recall. In [28], they introduced a feature-based sentence-level approach incorporating Arabic idioms, contextual multipliers, and syntactic features, achieving over 95% accuracy with SVM classifiers on MSA and Egyptian dialectal tweets. The same authors also presented MIKA, a manually annotated corpus for MSA and colloquial Arabic sentiment analysis. In [29], they presented an Arabic Jordanian Twitter corpus classified into positive and negative sentiment, covering both MSA and Jordanian dialect content. It evaluated multiple supervised machine learning approaches under various weighting techniques, stemming methods, and N-gram configurations. Experimental findings showed that the SVM using a TF-IDF representation with stemming and bi-gram features, achieved high performance compared to the NB classifier's best scenario. In [30], they performed Naive Bayes (NB), Decision Tree (DT), SVM on Modern Arabic Corpus that contains Arabic Twitter. Term Frequency-Inverse Document Frequency (TFIDF), Arabic stemming, normalization, stop-word removal has been performed. The results have shown that DT has performed better than the other techniques by getting 78% of f-measure. Similarly, in [31], they applied several machine learning classifiers, including NB, SVM, DT, Stochastic Gradient Descent (SGD), and LR, The models have been used on ASTD Egyptian Dialect tweets

and RES Multi-Dialect restaurant reviews. Moreover, TF-IDF has been used for feature representation with a set of pre-processing steps including removing stop-words, normalization, tokenization and lemmatization. The evaluation results showed that Bernoulli Naive Bayes (B-NB) was the best performing classifier with an accuracy of 82% on the ASTD dataset containing Egyptian Dialect tweets, whereas the SVM classifier yielded the best accuracy of 87.7% on the RES dataset comprising multi-dialect reviews. In [32], they applied a set of machine learning classifiers, including: SVM, LR, and K-Nearest Neighbour (KNN). The models were run to the Arabic-Review (ARev) database with specific features such as Bags of words (BOW). Additionally, several pre-processing steps were performed: Tokenization, normalization, deleting non-Arabic letters, numbers, usernames, and external links. The results showed that LR and SVM were the best performing classifiers with an accuracy of 93% and 92%, respectively. Moreover, in [33], they examined the effectiveness of different feature extraction and machine learning methods for ASA, with a set of pre-processing techniques. These techniques were: stop words removal, deleting repetitions, punctuation, hashtags, mentions, elimination of non-Arabic characters, deletion of diacritics and tashkeel, stemming and tokenization. Ar-Twitter and ASTC were utilized to evaluate the methodology. The study compared TF-IDF with unigrams and bigrams and Bag of Words (BoW) as well as five machine learning techniques: Random Forest (RF), DT, LR, SVM and XGBoost. The results indicated that RF with unigram achieved the highest accuracy equals to 87.7%, on the ASTC dataset, while SVM obtained 90.3% accuracy on the Ar-Twitter dataset. In [13], they employed topic modeling to construct a topic-based sentiment dataset for Arabic texts. Also, balancing strategies such as under-sampling, over-sampling, and imbalanced learning were conducted to examine their effect on classification performance. The research also investigated the impact of dataset size on machine learning models. Experimental results demonstrated that SVM achieved the highest accuracy, with F1-scores of 0.97% for sentiment analysis and 96% for topic classification in Arabic tweets. TABLE III. summarizes a set of studies that applied traditional machine learning models for ASA. In the reviewed studies, several techniques have been applied to enhance ASA, primarily relying on classifiers such as NB, SVM, LR, and DT, which were trained on handcrafted features including BoW, n-grams, POS tags, and TF-IDF representations.

As shown in the table, TF-IDF is the most widely adopted technique for feature representation. Obviously, studies demonstrated the effectiveness of SVM and NB in handling high-dimensional sparse feature spaces, achieving relatively high accuracy compared to a baseline lexicon-based method. Therefore, ASA using machine learning yields accuracy between 79% and 96% across a range of datasets and domains. Furthermore, the pre-processing stage plays a significant role in the performance and typically involves cleaning the data from special characters, hashtags, URLs, normalization and removing stop words. Although these methods achieved reasonable performance, their dependence on surface-level features limited their ability to capture the semantic and contextual features of the Arabic language, which required extensive feature engineering when performed to dialectal Arabic or noisy online text.

Table III. Summary of Machine Learning Models for Arabic Sentiment Analysis

Ref	Models	Features	processing	Dataset	Best Result
[22]	NB, SVM, KNN	TF-IDF, n-grams	Tokenization, Stemming, removing Stop words	in-house dataset of tweets.	Acc= 75.25 %
[28]	SVM	lexicon features, POS	remove redundant words	MSA and Egyptian dialectal tweets.	Acc=95%
[15]	SVM	TF-IDF, lexicon	stop-words removal, normalization, tokenization	ASTD (Egyptian tweets)	Acc. 72%
[30]	SVM, NB, DT	TF-IDF	Stemming, normalization, and removing stop words	Twitter data (MSA)	F1=78 %
[29]	SVM, NB	TF-IDF, N-grams	Stemming	MSA and Arabic Jordanian dialect corpus	Acc=88.72%
[35]	RF, SVM	Word embeddings	Normalization, diacritics removal	ArSentD-LEV	Acc=87%

Ref	Models	Features	processing	Dataset	Best Result
[36]	SVM, NB, DT	-	-	Arabic data from different websites (twitter, product reviews)	Acc = 96%
[37]	LR, KNN, DT, SVM, NB	TF-IDF, Count Vector-izer	Removal of punctuation, URLs, usernames, hashtags, stop-words, Normalization	ASTC twitter data	Acc=93%
[6]	Random Forest	Bag-of-Words, sentiment lexicons	Cleaning, stop-word removal, tokenization	Ar-Twitter, Multi-domain tweets	Acc= 87.7%
[39]	Ensemble (RF+SVM+DT+KNN+NB+LR)	TF-IDF, embeddings	Normalization, stemming	the trip advisor (TA) dataset -booking of hotels for Arabic	Acc=91%
[40]	Ensemble learning: KNN, SVM, DT, and RF	TF-IDF	Hashtag removal, punctuation and number removal, diacritic removal, removal of duplicate tweets, URL removal, stop word removal, normalization	Twitter reviews (coffee products);	Acc= 95.95 %
[33]	DT, LR, SVM, RF, XGBoost	BoW, TF-IDF, unigrams, bigrams	Removing stop words, emojis, hashtags, mentions,punctuation marks.	ASTC, Ar-Twitter	Acc=90.3%

C. Deep Learning Approaches (DL)

The advent of neural networks led to a shift toward models capable of automatically learning semantic and syntactic representations. Hence, deep learning models, such as Convolutional Neural Networks (CNN) [41] and Long Short-Term Memory (LSTM) networks [23], gained popularity, becoming widely used and enhancing the performance of ASA tasks. These models were able to overcome some limitations of classical approaches, particularly in recognizing patterns in noisy or short texts. CNNs and RNNs, as demonstrated in [42], showed strong improvements over ML by capturing contextual features in text. In [43], they proposed a system for ASA by applying word embeddings built from a 3.4-billion-word corpus and a CNN. The model exceeds the performance of existing methods on 4 out of 5 balanced and unbalanced datasets. Initializing word vectors with pre-trained vectors (CBOW58) significantly improves accuracy. Similarly, in [44], they utilized the CNNs and LSTM networks for ASA by addressing the morphological and orthographic complexities of the language. It compared sentiment analysis at different levels, including character-level, word-level, and Ch5gram-level representations. The methods were used on the Arabic Health Services (AHS) dataset. The proposed approach achieved accuracy of 94.24% for the Main-AHS subset and 95.68% for the Sub-AHS subset. Additionally, [45] proposed a new framework for ASA aimed to using deep learning and several pre-processing techniques. Two models of LSTM and CNN networks were applied and evaluated, with the addition of pre-processing steps such as stemming, tokenization, normalization and removal of stop words. A key contribution was the development of the Moroccan Sentiment Analysis Corpus (MSAC), a rich and publicly available dataset. Experimental results demonstrated that deep learning outperformed machine learning approaches such as SVM and NB. In [46], they employed deep learning models that combined a CNN and bidirectional long short-term memory (Bi-LSTM) for feature extraction, combined with a linear SVM classifier, to process the embedded vectors produced by the Bi-LSTM and CNN. The proposed method was evaluated on a set of publicly available datasets, and the results showed that it achieved high performance compared to the baseline CNN and SVM algorithms across all datasets. In [47], they applied CNN), LSTM, and Hybrid CNN-LSTM on LARB and HARD datasets. They explored different word embeddings (Word2Vec, fastText). Additionally, a set of pre-processing steps was performed, including tokenization, cleaning text by removing punctuation and diacritics, and stemming. The results showed that CNN with Word2Vec on HARD achieved 94.68%. In[12], an ensemble model composed of a CNN and an LSTM is used for predicting the sentiment of Arabic tweets. The model achieves an F1-score of 64.46% on the ASTD.

However, despite the successes of using deep learning approaches in the ASA task, the models required large, annotated datasets, and their performance may drop in resource-scarce dialects such as Moroccan [45].

D. Transformers and Pre-trained Models

The emergence of transformer-based architectures was the most significant advancement in ASA. These models exceeded the constraints of lexicon-based, machine learning, and deep learning approaches. Moreover, these models consistently outperform deep learning baselines on various benchmark datasets, such as ASTD, ArSAS, and SemEval, according to several studies [48][11]. Language models based on transformers, such as BERT [49] and its Arabic variant AraBERT [9], have proven their ability to capture deep contextual representations and have demonstrated superior performance in sentiment analysis tasks [49][10]. The introduction of AraBERT [9] which is the first large-scale Arabic pre-trained language model, provided superior performance in sentiment analysis benchmarks and marked a significant milestone, demonstrating strong performance across multiple Arabic datasets. Subsequently, MARBERT and ARBERT [10] further developed this work by leveraging large-scale social media corpora and dialectal data. Additionally, CAMeLBERT [51] focused mainly on multi-dialect corpora, offering enhanced handling of diverse Arabic varieties. Recently, QARiB [52] and AraELECTRA [16] have further pushed the performance by integrating advanced pretraining strategies and larger datasets. AraBERT and MARBERT have been widely deployed due to their ability to understand deep contextual representations across MSA and dialects [10], which incorporated dialectal diversity from massive social media corpora. In [48], they compared the performance of different deep learning models (LSTM, CNN, GRU and hybrids) and shallow learning classifiers alongside transformer-based approaches such as AraBERT. Using large book review and multi-dialect Arabic hotel datasets. The results showed that deep learning outperforms shallow techniques, whereas transformer models outperform deep learning models. Similarly, [53] provided a comprehensive evaluation of transformer models, highlighting the robustness of AraBERT v2.0, ArBERT, and MARBERT in aspect-based sentiment analysis. Comparative studies have further demonstrated the strength of these models in NLP tasks, such as Arabic text classification [54], confirming their flexibility. Moreover, efforts tend to employ ensemble techniques such as voting ensemble using AraBERT embeddings [56] enhanced with AraBERT, achieving the highest accuracy and F1 scores.

TABLE IV. compares the latest and most prominent pre-trained language models commonly used in Arabic sentiment analysis tasks. Obviously, the reported results demonstrate that transformer-based models consistently perform well on different benchmark datasets demonstrating their effectiveness in understanding the linguistic features of Arabic. Among the models, MARBERT performs best with F1-scores ranging from 82% to 88% due to its pre-training on a large collection of Arabic tweets and social media content. Additionally,

AraBERT provides competitive results with F1-scores from 79% to 86% on many benchmark datasets, demonstrating the effectiveness of general-purpose Arabic language model. Similarly, CAMeLBERT and QARiB have robust performance, especially on datasets with multiple dialects. Therefore, the results show that including dialectal coverage in pre-training contributes positively to sentiment classification performance. In contrast, ARBERT reports a slightly lower result than MARBERT, however, it still achieves good classification accuracy on Modern Standard Arabic datasets. AraELECTRA also performs competitively, suggesting that lightweight transformer architectures can yield effective sentiment representations while being computationally efficient.

In general, the comparison results show that recent transformer-based models have achieved great progress on the ASA task. However, the variations in the performance on datasets suggest that the effectiveness of models is influenced by dialectal coverage, dataset size, and annotation quality. Hence, the selection of an appropriate pre-trained model should also consider the linguistic and domain characteristics of the application.

Tbale IV. Comparison of recent Transformer-based architecture applied to Arabic Sentiment Analysis

Model	Dataset(s)	Performance
AraBERT [9]	ASTD, ArSenTD-Lev, LABR	F1: 79–86%
MARBERT [10]	Twitter (large-scale), AraEval	F1: 82–88%
ARBERT [10]	PADIC, ASTD	Acc: ~83%
CAMeLBERT [51]	Multi-dialect corpora	F1: 80–85%
QARiB [52]	Arabic news + tweets	Acc: 84–87%
AraELECTRA [16]	Twitter, ArSAS	F1: ~85%

1) Aspect-Based Sentiment Analysis (ABSA)

Transformer-based architecture has played a major role in progressing Arabic Aspect-Based Sentiment Analysis (ABSA). ABSA seeks to identify sentiment about certain entities or aspects in a piece of text. Recently, ABSA has attracted more attention in the Arabic NLP research, since it can capture fine-grained sentiment at the aspect level rather than at the document or sentence level. Traditional approaches to ABSA depend on rule-based or lexicon-driven methods, but they often cannot cope with the morphology and diversity of Arabic dialects. However, with the advent of deep learning and neural architectures like CNNs and LSTMs; aspect extraction has improved.

ABSA is introduced in several studies for application. In [58], they introduced a benchmark dataset for Arabic ABSA, and deep neural networks were used to determine strong baselines. Later, in [11], they extended the SemEval ABSA tasks to cover Arabic reviews. More recent studies have applied transformer-based models, such that AraBERT and MARBERT have been fine-tuned for extraction with promising results [60][65]. Additionally, CAMeLBERT and QARiB have shown improvements when performed for ABSA tasks, particularly in handling dialectal variations [51][66]. In [61], they proposed a hybrid transformer model integrating zero-shot learning (ZSL) techniques, which enabled efficient ABSA in domains with limited annotated data, such as Arabic game reviews. Their results showed superior performance. Similarly, [62] introduced an unsupervised integrated framework combining ABSA with text summarization for traffic services using transformers, demonstrating the adaptability of transformer-based models in real-world smart city applications. The authors in [63] employed AraBERT and MARBERT in the healthcare domain for aspect-based sentiment analysis, achieving significant improvements over earlier approaches. TABLE V. presents a comparative review of recent Arabic ABSA studies, focusing on datasets, methods, and performance results. Together, these contributions implied that ABSA in Arabic is headed towards more robust and generalizable systems, empowered by transformer-based architectures.

Table V. Recent studies on (ABSA) for Arabic.

Study	Methodology	Dataset	Performance
[58]	Lexicon + ML features	Ara-ABSA Arabic hotel reviews	Accuracy: 72%
[11]	SemEval ABSA extension	SemEval Arabic task	F1: 68%
[60]	AraBERT fine-tuning	Arabic Tourism Reviews	F1: 97%
[51]	CAMeLBERT	Multi-dialect corpora	F1: 81%
[51]	MARBERT-based	Twitter-ABSA (dialectal)	Acc: 83%
[53]	QARiB	Large-scale Arabic data (News)	Acc: 85%
[62]	Unsupervised Transformers	Traffic service texts	F1: 79%
[61]	Hybrid Transformers	Arabic game reviews	Acc: 87%

4. SURVEY STUDIES ON ARABIC SENTIMENT ANALYSIS (ASA)

A set of surveys has provided a comprehensive overview of the progress in Arabic Sentiment Analysis (ASA) and summarize the developments, techniques and challenges regarding this domain. The earlier surveys focused mainly on lexicon-based and machine learning methods. For instance, [50] reviewed the earliest sentiment analysis studies on Arabic tweets, emphasizing pre-processing challenges such as stemming and handling diacritics. Similarly, [65] reviewed Arabic NLP tasks in general and devoted a section to ASA,

highlighting the need for public datasets. In [20], they summarized the different ASA approaches, highlighting the shift from traditional models such as SVM and NB to deep learning models like LSTM and CNN. In [2], they provided an extensive survey on machine and deep learning and compared the performance of CNN and RNN on various datasets.

Recently, surveys have moved towards transformer-based models and cross-lingual applications, such as in [8], where a systematic review of the literature focuses on the essential pre-processing steps, such as stemming, removal of stop words, and normalization. The study demonstrated that pre-processing and effective feature extraction significantly improve ASA performance. Additionally, [66] provided a broad review of ASA approaches, emphasizing the role of transformer-based architectures such as BERT and AraBERT, as well as the challenges posed by dialectal variations and the scarcity of large-scale datasets. TABLE VI. highlights the evolution of ASA surveys over the period (2018-2025). Earlier surveys [2][20][65][50] primarily focused on traditional machine learning approaches and linguistic pre-processing challenges. However, recent surveys [5][14][64][55] expanded their coverage to include deep learning models and transformer-based architectures. Furthermore, most existing surveys [66] either focus on specific methodological aspects or provide descriptive summaries of the literature without offering a comprehensive comparison of datasets, tools, sentiment analysis levels, and emerging research directions. While previous surveys have significantly contributed to structuring the evolution of ASA, most of them have been focused on specific aspects of the field, such as traditional machine learning techniques, deep learning architectures, pre-processing strategies, or transformer-based models. Furthermore, many surveys provided descriptive summaries of existing studies without offering a comprehensive comparison across datasets, tools, and sentiment analysis levels. In addition, the rapid advancement of transformer-based language models and the growing interest in Aspect-Based Sentiment Analysis (ABSA) have introduced new challenges and opportunities that are not fully covered in earlier reviews.

TABLE VI. summarizes the evolution of the major methodological paradigms employed in Arabic Sentiment Analysis (ASA), highlighting their representative techniques, strengths, limitations, and reported performance ranges. Early studies primarily relied on lexicon-based approaches, which offered interpretable sentiment classification without the need for large, annotated datasets. However, their effectiveness was often constrained by limited dialectal coverage and the substantial effort required to construct and maintain sentiment lexicons. Subsequently, traditional machine learning methods, including Support Vector Machines (SVM), Naïve Bayes (NB), and Random Forest (RF), achieved improved classification performance by leveraging handcrafted features, although their success remained highly dependent on feature engineering and preprocessing quality. The emergence of deep learning approaches marked a significant advancement in ASA by enabling automatic feature extraction and improved contextual representation. Models such as CNNs, LSTMs, and BiLSTMs demonstrated superior performance across various benchmark datasets, particularly in handling complex linguistic patterns. More recently, transformer-based architectures, including AraBERT and MARBERT, have established the current state of the art, achieving the highest reported performance levels due to their ability to capture rich contextual and semantic information. Nevertheless, these models require substantial computational resources and large-scale training data. Overall, the comparison illustrates a clear progression from manually engineered approaches toward increasingly sophisticated contextual models, reflecting the broader evolution of natural language processing techniques in the Arabic sentiment analysis domain.

Therefore, this is not the first survey focused on ASA, but to the best of our knowledge, this is the first survey that provides a more comprehensive and up-to-date synthesis of the ASA literature by integrating traditional approaches, deep learning methods, transformer-based architectures, datasets, tools, linguistic challenges, and future research directions within a single framework. The pie chart in Fig. 5. shows the distribution of ASA approaches between 2018 and 2025, highlighting the evolution of methodologies in the field. We performed a systematic literature review according to PRISMA guidelines. Our search strategy involved Scopus, IEEE Xplore, ACL Anthology, and Google Scholar using queries: ("Arabic" OR "Arabic language") AND ("sentiment analysis" OR "opinion mining" OR "aspect-based sentiment" OR Arabic Sentiment Analysis). After full-text review, we retrieved over 70 papers from 2018 to 2025 that met our inclusion criteria.

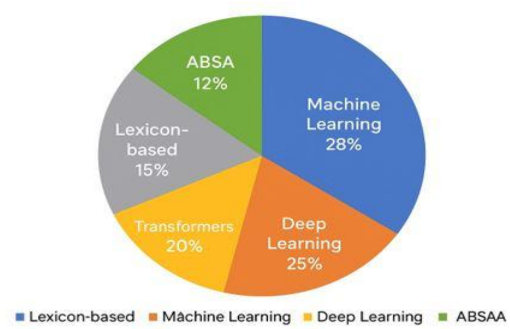

Fig. 5. Distribution of Arabic Sentiment Analysis approaches

Table VI. Comparative Analysis of ASA Approaches

Approach	Year Range	Key Methods	Strengths	Weaknesses	Performance
Lexicon-based	2015-2018	AraSenTi, SentiWS	No training needed, interpretable	Dialect coverage, manual effort	Acc: 60-75%
ML	2014-2020	SVM, NB, RF	Interpretable, lower computational cost	Feature engineering, limited context	Acc: 75-90%

Approach	Year Range	Key Methods	Strengths	Weaknesses	Performance
DL	2016-2022	CNN, LSTM, BiLSTM	Automatic feature learning, context	Requires large data, black box	Acc: 80-95%
Transformers	2020-2026	AraBERT, MARBERT	State-of-the-art, contextualized	Computational cost, data requirements	Acc: 85-90%+

As shown, machine learning (ML) methods are still the most widely used, comprising 28% of studies. Deep learning (DL) approaches are closely behind with 25% showing the effect of CNNs and LSTMs to handle complex linguistic structures of Arabic. Transformers account for around 20%, showing the rapid transition to contextualized models like AraBERT that greatly enhance performance in sentiment analysis problems. Further, traditional lexicon-based approaches are still present in around 15% of the studies and Aspect-Based Sentiment Analysis (ABSA) has emerged as a specialized subfield representing 12% of recent contributions. We can observe a clear evolution from the rule-based methods to the deep contextualized models in this distribution. ABSA is gaining importance for fine-grained sentiment understanding.

5. Datasets and Tools

The availability of annotated datasets, along with supporting frameworks and tools, is crucial for the progress of Arabic Sentiment Analysis (ASA) for pre-processing, feature extraction, and model development. Nevertheless, Arabic resources are relatively limited in terms of size and diversity when compared to English and other languages.

Table VII. SUMMARY OF RECENT SURVEYS ON ASA.

Ref.	Domain	Limits
[20] 2018	ML, lexicon, hybrid approaches	ABSA
[2] 2019	Lexicon-based approach, Corpus-based approach	Limited transformers
[65] 2019	Lexicon-based approach, Corpus-based approach	Limited ABSA/tools/datasets
[8] 2020	ML, Lexicon-based Approach, hybrid approaches	Limited ABSA
[5] 2021	ABSA only	Limited ML &DL approaches
[64] 2022	Arabic sentiment analysis using RNNs	Limited Machine learning, Deep learning approaches and ABSA
[55] 2023	comprehensive review of the Arabic aspect-based sentiment analysis	Limited ML & DL
[17] 2024	Machine learning Approach, Lexicon-based Approach, hybrid approaches)	Limited ABSA
[66] 2025	Contemporary ASA	Limited tools/datasets
[64] 2025	Arabic sentiment analysis using RNNs	Limited ML&DL approaches &ABSA

As a result, many benchmark datasets and NLP tools have been developed for Arabic, which considerably facilitate the progress of this task. Below, the most widely used datasets and tools have been summarized:

1) Publicly Available Datasets

Several benchmark datasets have been introduced to address the lack of annotated corpora for sentiment analysis in Arabic. These datasets vary in domain (e.g., Twitter, reviews, news), size, and annotation methodology. TABLE VIII. shows a summary of major datasets related to ASA.

- **LABR (Large Arabic Book Reviews)** [24]: A corpus of more than 60,000 book reviews gathered from Goodreads, annotated with positive, negative, neutral. LABR is one of the largest available Arabic sentiment datasets.
- **Arabic Hotel Reviews (AHR)** [68]: is one of the widely used corpora for sentiment analysis in the hospitality domain. It consists of 20,000 hotel reviews written in Modern Standard Arabic and dialectal Arabic, collected from online platforms.
- **AraSenTi-Tweet** [69]: Includes more than 17,000 Saudi Arabian tweets, labelled for binary and multi-class sentiment classification. It addresses dialectal Arabic challenges.
- **SemEval Arabic Subtasks** [70]: Provided as a part of the SemEval competitions. The datasets cover different sentiment analysis challenges, including aspect-based sentiment analysis and emotion detection.
- **ASTD (Arabic Sentiment Tweets Dataset)** [71]: Contains over 10,000 manually labeled tweets classified into positive, negative, neutral. It focuses on Egyptian dialect and MSA.

Table VIII. SUMMARY OF COMMON ARABIC SENTIMENT ANALYSIS DATASETS

Dataset	Domain	Size	Language Variety	Sentiment labels
LABR	Book reviews	63K	MSA	5-star rating scale
AHR	Hotel reviews	20K	MSA	Positive, Negative, Neutral
ASTD	Twitter	10K	MSA, Egyptian Dialect	positive, negative, neutral,
AraSenTiTweet	Twitter	12K	Saudi Dialect	Positive, Negative, Neutral
SemEval Subtasks	Twitter	6k	MSA, Dialects	Positive, Negative, Neutral

2) Available Tools and Frameworks

In addition to datasets, several tools have been developed to support Arabic NLP tasks including sentiment analysis. These tools provide functionalities such as tokenization, morphological analysis, and part-of-speech tagging, which are critical pre-processing steps for ASA. Some of the widely used tools include:

- CAMeL Tools [72]: An open-source Python toolkit that provides a suite of NLP tools for Arabic, including morphological analysis, disambiguation, lemmatization, and named entity recognition. CAMeL is particularly useful in pre-processing ASA datasets.
- AraBERT [9]: A pre-trained transformer model based on BERT architecture, specifically trained on large-scale Arabic corpora. AraBERT has demonstrated state-of-the-art performance in various ASA tasks and is widely used by researchers.
- Farasa [73]: fast and accurate Arabic NLP tools for POS tagging, segmentation and diacritization. Farasa is particularly efficient in handling MSA texts.
- MADAMIRA [74]: A system integrating morphological analysis and disambiguation, useful for pre-processing in Arabic NLP applications.
- Mazajak [57]: is an Arabic sentiment analysis tool that utilizes deep learning models to classify the sentiment of Arabic text. It is widely used for analyzing opinions expressed in social media and online content.
- Scikit-learn and TensorFlow: general-purpose machine learning libraries which can be used to develop and evaluate ASA models, mainly in implementing classifiers and deep learning models.
- PyTorch: is an open-source deep learning framework developed by Meta AI. It provides a flexible and efficient platform for building, training, and deploying neural networks. It has become one of the most widely used frameworks for developing modern sentiment analysis models.

TABLE IX. summarizes the most widely used tools and frameworks in Arabic Sentiment Analysis (ASA). These tools provide the essential infrastructure for building efficient Arabic sentiment analysis systems. Arabic-specific NLP toolkits, such as MADAMIRA, Farasa, AraNLP [76], and CAMeL, are primarily used for preprocessing, including tokenization, stemming, lemmatization, morphological analysis, and dialect identification. General-purpose deep learning frameworks, such as PyTorch and Hugging Face Transformersare extensively used for developing, training, and fine-tuning modern deep learning and transformer-based models.

Table IX. Common tools and frameworks used in Arabic sentiment analysis

Tool	Category	Main Application	Availability
MADAMIRA	Arabic NLP Toolkit	Morphological analysis, POS tagging, and lemmatization	Licensed
Farasa	Arabic NLP Toolkit	Text segmentation, stemming, and preprocessing	Open source
PyTorch	Deep Learning Framework	Development and training of deep learning and transformer-based sentiment analysis models	Open source
AraNLP	Arabic NLP Toolkit	Arabic text preprocessing, normalization, tokenization, and stemming	Open source
Hugging Face Transformers	NLP Library	Fine-tuning and deployment of transformer-based language models	Open source
CAMeL Tools	Arabic NLP Toolkit	Tokenization, morphology, NER, dialect identification, and preprocessing	Open source

6. Challenges and FUTURE directions

A. Challenges in ASA

Arabic sentiment analysis continues to face challenges despite significant advances in natural language processing. A major issue is the dialectal diversity of Arabic, which makes it difficult for models trained on Modern Standard Arabic (MSA) to effectively generalize across regional dialects [15][75]. Despite the popularity of transformer-based approaches, researchers still observe a drop in accuracy for input in dialectal forms. [9].

This challenge is attributed to the linguistic and resource-related complexities of Arabic. The language's rich morphology generates extensive lexical variation [34], causing sparsity in data representation. The scarcity of large-scale balanced sentiment datasets further complicates this problem, although recent corpora like ASTD and LABR have made notable contributions [24][35]. Moreover, issues such as code-switching between Arabic and English, as well as the pervasive use of informal text, introduce significant noise that challenges even transformer-based models [42][50]. Furthermore, the absence of standardized benchmarks prevents fair comparison across models and slows down progress in general. [43].

Therefore, there is a high demand for building larger dialect-inclusive resources, developing unified evaluation standards, and adapting advanced models like AraBERT and MARBERT to better handle informal, code-switched content. Recent studies emphasize the potential of pre-trained transformer models fine-tuned on dialectal data, demonstrating improved robustness in handling real-world social media text [9][64][90]. Only by strategically addressing these challenges can Arabic sentiment analysis evolve into an impactful research domain.

B. Future Directions

The advancement of Arabic Sentiment Analysis (ASA) will heavily depend on emerging deep learning architectures and enhancing domain adaptability. While transformer-based approaches such as these have demonstrated significant improvements, limitations related to dialectal variation, domain coverage, and lack availability of resources still hinder broader applicability [51][65]. Researchers are now exploring Aspect-Based Sentiment Analysis (ABSA) for fine-grained opinion mining, cross-lingual transfer learning to exploit multilingual corpora, and multimodal sentiment analysis that integrates textual, visual, and audio signals for richer context [53][61][62]. Furthermore, they emphasize the importance of incorporating knowledge graphs and unsupervised domain adaptation to compensate for the shortage of annotated Arabic datasets [66]. Therefore, future research in ASA should combine these directions, focusing on ABSA. An integrated approach will not only improve performance in various domains but will also guarantee socially responsible and transparent applications of ASA [53][61].

7. Conclusion

Arabic Sentiment Analysis (ASA) has witnessed substantial progress, driven by the growing availability of computational resources, data and evolving machine learning classifiers.

While traditional methods such as Support Vector Machines and Naive Bayes have laid the groundwork for ASA, the rise of deep learning and transformer-based models has significantly enhanced the ability to handle the linguistic features of the Arabic language. Despite this progress, several issues remain, including dialectal diversity, scarcity of annotated datasets and code-switching. Solving these issues requires the development of comprehensive, diverse corpora, and the adoption of techniques such as cross-lingual transfer learning and multimodal analysis. This survey provides a comprehensive review of Arabic sentiment analysis research from 2018 to 2025, analyzing over 70 reviewed studies across three methodological approaches: machine learning, deep learning, and transformer-based approaches. Key findings include:

1. Transformer-based models (AraBERT, MARBERT) consistently achieve state-of-the-art performance (F1: 79-88%) across standard benchmarks, with MARBERT showing superior dialectal coverage.
2. Data scarcity remains a major challenge compared with resource-rich languages such as English.
3. Aspect-based sentiment analysis has emerged as the most active research direction, with recent transformer models achieving 81-87% accuracy.
4. Critical gaps persist in dialectal generalization, evaluation standardization, and handling of code-switching and informal text.

This survey contributes a unified framework for understanding ASA evolution, a comprehensive review of resources, and a prioritized research agenda. We recommend that future research prioritize: (1) development of dialect annotated corpora, (2) transfer learning approaches, (3) standardized evaluation benchmarks, and (4) applications in under-explored domains such as healthcare and politics. Limitations of this survey include the rapid pace of transformer model development, which may affect the timeliness of our comparative analysis, and potential selection bias in our literature search. We have made our search methodology and inclusion criteria explicitly available to allow replication and future updates. The continued advancement of ASA requires collaborative efforts between computational linguists, domain experts, and Arabic-speaking communities to ensure models are both linguistically accurate and culturally appropriate.

Reference

1. H. ElSahar and S. R. El-Beltagy, “Building large Arabic multi-domain resources for sentiment analysis,” in Proc. Int. Conf. Intelligent Text Processing and Computational Linguistics, Springer, 2015, pp. 23–34.
2. M. Al-Ayyoub, M. Al-Smadi, and Y. Jararweh, “A comprehensive survey of Arabic sentiment analysis: Approaches, resources, and applications,” ACM Comput. Surveys, vol. 53, no. 4, pp. 1–45, 2020, doi:10.1145/3395031.
3. H. Rahab, M. Djoudi, and A. Zitouni, “Sentiment analysis of Arabic documents: Main challenges and recent advances,” in NLP for Global and Local Business, 2021, pp. 307–331.
4. A. Shoukry and A. Rafea, “Sentence-level Arabic sentiment analysis,” in Proc. Int. Conf. Collaboration Technologies and Systems (CTS), IEEE, 2012, pp. 546–550.
5. R. Obiedat et al., “Arabic aspect-based sentiment analysis: A systematic literature review,” IEEE Access, vol. 9, pp. 152628–152645, 2021.
6. N. Al-Twairesh, H. Al-Khalifa, A. Alsalman, and Y. Al-Ohali, “Sentiment analysis of Arabic tweets: Feature engineering and a hybrid approach,” arXiv preprint arXiv:1805.08533, 2018.
7. Y. Chen and S. Skiena, “Building sentiment lexicons for all major languages,” in Proc. ACL, 2014, pp. 383–389.
8. A. Ghallab, A. Mohsen, and Y. Ali, “Arabic sentiment analysis: A systematic literature review,” Appl. Comput. Intell. Soft Comput., 2020, Art. ID 7403128.
9. W. Antoun, F. Baly, and H. Hajj, “AraBERT: Transformer-based model for Arabic language understanding,” arXiv:2003.00104, 2020.
10. M. Abdul-Mageed and A. Elmadany, “ARBERT & MARBERT: Deep bidirectional transformers for Arabic,” in Proc. ACL, 2021.
11. M. Pontiki et al., “Semeval-2016 task 5: Aspect-based sentiment analysis,” in Proc. 10th Int. Workshop Semantic Evaluation (SemEval-2016), ACL, 2016, pp. 19–30, doi:10.18653/v1/S16-1002.
12. M. Heikal, M. Torki, and N. El-Makky, "Sentiment analysis of Arabic tweets using deep learning," Procedia Computer Science, vol. 142, pp. 114–122, 2018, doi: 10.1016/j.procs.2018.10.466.
13. Alderazi et al., “Generative artificial intelligence in topic-sentiment classification for Arabic text: A comparative study with possible future directions,” PeerJ Comput. Sci., vol. 10, p. e2081, 2024.
14. Alqusair, M. Taileb, and H. Al-Barhamtoshy, “A survey of deep learning techniques for Arabic aspect-based sentiment analysis,” IEEE Access, 2025.

15. Badaro, R. Baly, H. Hajj, N. Habash, and W. El-Hajj, “A large-scale Arabic sentiment lexicon for Arabic opinion mining,” in Proc. EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP), 2014, pp. 165–173.
16. W. Antoun, F. Baly, and H. Hajj, “AraELECTRA: Pre-training text discriminators for Arabic language understanding,” in Proc. Arabic Natural Language Processing Workshop, 2021.
17. A. A. Aladeemy et al., “Advancements and challenges in Arabic sentiment analysis: A decade of methodologies, applications, and resource development,” Heliyon, vol. 10, no. 21, 2024.
18. A. Pang and L. Lee, “Opinion mining and sentiment analysis,” Found. Trends Inf. Retr., vol. 2, no. 1–2, pp. 1–135, 2008.
19. Y. Eid, H. Zayed, and W. Medhat, “A-MASA: Arabic multi-domain aspect-based sentiment analysis datasets,” Procedia Comput. Sci., vol. 244, pp. 202–211, 2024.
20. M. Alrefai, H. Faris, and I. Aljarah, “Sentiment analysis for Arabic language: A brief survey of approaches and techniques,” arXiv:1809.02782, 2018.
21. A. El-Halees, "Arabic opinion mining using combined classification approach," in Proceedings of The International Arab Conference on Information Technology (ACIT), Zarqa, Jordan, Dec. 2011, pp. 264–271.
22. R. M. Duwairi and I. Qarqaz, “Arabic sentiment analysis using supervised classification,” in Proc. Int. Conf. Future Internet of Things and Cloud, IEEE, 2014, pp. 579–583.
23. S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.
24. M. Aly and A. F. Atiya, “LABR: Large-scale Arabic book reviews dataset,” in Proc. EMNLP, 2013, pp. 1172–1182.
25. N. El-Naggar, Y. El-Sonbaty, and M. Abou El-Nasr, “Sentiment analysis of MSA and Egyptian dialectal Arabic tweets,” in Proc. Computing Conf., IEEE, 2017, pp. 880–887.
26. P. Langley, W. Iba, and K. Thompson, “An empirical comparison of Bayesian classifiers,” in Proc. AAAI, 1992, pp. 223–228.
27. Cortes and V. Vapnik, “Support-vector networks,” Mach. Learn., vol. 20, pp. 273–297, 1995.
28. S. Ibrahim, S. M. Abdou, and M. Gheith, “Sentiment analysis for Modern Standard Arabic and colloquial Arabic,” arXiv:1505.03105, 2015.
29. K. M. Alomari, H. M. ElSherif, and K. Shaalan, “Arabic tweets sentiment analysis using machine learning,” in Int. Conf. Ind. Eng. Appl. Artif. Intell., Springer, 2017, pp. 602–610.
30. M. M. Altawaier and S. Tiun, “Comparison of machine learning approaches on Arabic Twitter sentiment analysis,” Int. J. Adv. Sci. Eng. Inf. Technol., vol. 6, no. 6, pp. 1067–1073, 2016.
31. A. H. Hussein, I. F. Moawad, and R. M. Badry, "Arabic sentiment analysis for multi-dialect text using machine learning techniques," International Journal of Advanced Computer Science and Applications, vol. 12, no. 12, pp. 693–700, 2021, doi: 10.14569/IJACSA.2021.0121286.
32. Fouadi, H. El Moubtahij, H. Lamtougui, and A. Yahyaouy, “Sentiment analysis of Arabic comments using machine learning and deep learning models,” Indian J. Comput. Sci. Eng., vol. 13, no. 3, pp. 598–606, 2022.
33. A. A. Aladeemy et al., “Machine learning algorithms for predicting and analyzing Arabic sentiment,” SN Comput. Sci., vol. 5, no. 8, pp. 1–10, 2024.
34. M. Al-Haj and D. Hussein, “Sentiment analysis of Arabic tweets: Feature engineering and a hybrid approach,” in Proc. Int. Conf. Computational Linguistics and Intelligent Text Processing, Springer, 2016, pp. 307–320.
35. G. Badaro, F. Baly, H. Hajj, N. Habash, and W. El-Hajj, “ArSentD-LEV: A multi-topic corpus for Arabic sentiment analysis of Levantine tweets,” Procedia Comput. Sci., vol. 142, pp. 66–73, 2018.
36. M. A. Algburi, A. Mustapha, and S. A. Mostafa, “Comparative analysis for Arabic sentiment classification,” in Proc. Int. Conf. Appl. Comput. Support Ind. Innov. Technol., Springer, 2019, pp. 271–285.
37. N. K. Bolbol and A. Y. Maghari, “Sentiment analysis of Arabic tweets using supervised machine learning,” in Proc. ICPET, IEEE, 2020, pp. 89–93.
38. M. Saidi, K. Shahi, N. Zare, and H. Hassanpoor, “A hybrid approach for sentiment analysis of Arabic tweets,” in Proc. 10th Int. Conf. Signal Processing and Intelligent Systems (ICSPIS), IEEE, 2024, pp. 105–109.
39. N. Hicham, S. Karim, and N. Habbat, “Customer sentiment analysis for Arabic social media using a novel ensemble machine learning approach,” Int. J. Electr. Comput. Eng., vol. 13, no. 4, pp. 4504–4515, 2023.
40. O. Alsemaree, A. S. Alam, S. S. Gill, and S. Uhlig, “Sentiment analysis of Arabic social media texts: A machine learning approach to deciphering customer perceptions,” Heliyon, vol. 10, no. 9, 2024.
41. Y. Kim, “Convolutional neural networks for sentence classification,” in Proc. EMNLP, 2014, pp. 1746–1751.
42. A. Al Sallab, H. Hajj, R. Baly, K. Shaban, and W. El-Hajj, “Deep learning models for sentiment analysis in Arabic,” in Proc. 2nd Workshop on Arabic Natural Language Processing, 2015.
43. Dahou, S. Xiong, J. Zhou, M. H. Haddoud, and P. Duan, “Word embeddings and convolutional neural networks for Arabic sentiment classification,” in Proc. COLING 2016, 2016, pp. 2418–2427.
44. M. Alayba, V. Palade, M. England, and R. Iqbal, “A combined CNN and LSTM model for Arabic sentiment analysis,” in Int. Cross-Domain Conf. Machine Learning and Knowledge Extraction, Springer, 2018, pp. 179–191.

45. Oussous, F.-Z. Benjelloun, A. A. Lahcen, and S. Belfkih, “ASA: A framework for Arabic sentiment analysis,” J. Inf. Sci., vol. 46, no. 4, pp. 544–559, 2020.
46. O. Alharbi, “A deep learning approach combining CNN and Bi-LSTM with SVM classifier for Arabic sentiment analysis,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 6, 2021.
47. N. Elhassan et al., “Arabic sentiment analysis based on word embeddings and deep learning,” Computers, vol. 12, no. 6, p. 126, 2023.
48. A. B. Nassif, A. M. Darya, and A. Elnagar, “Empirical evaluation of shallow and deep learning classifiers for Arabic sentiment analysis,” ACM Trans. Asian Low-Resour. Lang. Inf. Process., vol. 21, no. 1, pp. 1–25, 2021.
49. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” arXiv:1810.04805, 2018.
50. Assiri, A. Emam, and H. Aldossari, “Arabic sentiment analysis: A survey,” Int. J. Adv. Comput. Sci. Appl., vol. 6, no. 12, 2015.
51. G. Inoue, B. Alhafni, N. Baimukan, H. Bouamor, and N. Habash, “The interplay of variant, size, and task type in Arabic pre-trained language models,” in Proc. 6th Arabic Natural Language Processing Workshop, 2021, pp. 92–104.
52. A. Abdelali, S. Hassan, H. Mubarak, K. Darwish, and Y. Samih, “Pre-training BERT on Arabic tweets: Practical considerations,” arXiv:2102.10684, 2021.
53. M. Almasaud and H. Al-Baity, “On the robustness of Arabic aspect-based sentiment analysis: A comprehensive exploration of transformer-based models,” J. King Saud Univ. Comput. Inf. Sci., 2024, doi:10.1016/j.jksuci.2024.102352.
54. Aliah, D. Berezkin, and F. Kozlov, “Enhancing Arabic text classification: The impact of dataset variety on BERT models,” in Proc. IEEE Int. Conf. Data Science and Advanced Analytics (DSAA), 2025, pp. 1–8, doi:10.1109/DSAA58524.2025.10971045.
55. R. Bensoltane and T. Zaki, “Aspect-based sentiment analysis: An overview in the use of Arabic language,” Artif. Intell. Rev., vol. 56, no. 3, pp. 2325–2363, 2023.
56. D. Ghoul, J. Patrix, G. Lejeune, and J. Verny, “A combined AraBERT and voting ensemble classifier model for Arabic sentiment analysis,” Nat. Lang. Process. J., vol. 8, p. 100100, 2024.
57. Abu Farha and W. Magdy, “Mazajak 2.0: Arabic sentiment and emotion analysis on social media,” Inf. Process. Manag., vol. 58, no. 2, p. 102438, 2021, doi:10.1016/j.ipm.2020.102438.
58. M. Al-Smadi, M. Al-Ayyoub, Y. Jararweh, and O. Qawasmeh, “Enhancing aspect-based sentiment analysis of Arabic hotels’ reviews using morphological, syntactic and semantic features,” Inf. Process. Manag., vol. 56, no. 2, pp. 308–319, 2019.
59. S. Zaid, A. H. Alharbi, and H. Samra, “Multi-aspect sentiment classification of Arabic tourism reviews using BERT and classical machine learning,” Data, vol. 10, no. 11, p. 168, 2025.
60. M. Hammad, N. AbuEnnab, and M. Al-Refai, “An aspect-based sentiment analysis model for Arabic game reviews based on hybrid transformer models,” Neural Comput. Appl., pp. 1–23, 2025.
61. A. Alotaibi and F. Nadeem, “An unsupervised integrated framework for Arabic aspect-based sentiment analysis and abstractive text summarization of traffic services using transformer models,” Smart Cities, vol. 8, no. 2, p. 62, 2025.
62. AlNasser and A. AlMuhaideb, “Listening to patients: Advanced Arabic aspect-based sentiment analysis using transformer models,” Big Data Cogn. Comput., vol. 8, no. 11, p. 156, 2024, doi:10.3390/bdcc8110156.
63. S. O. Alhumoud and A. A. Al Wazrah, “Arabic sentiment analysis using recurrent neural networks: A review,” Artif. Intell. Rev., vol. 55, no. 1, pp. 707–748, 2022.
64. Guellil, F. Azouaou, and M. Mendoza, “Arabic sentiment analysis: Studies, resources, and tools,” Soc. Netw. Anal. Min., vol. 9, no. 1, p. 56, 2019.
65. Z. Shi and R. Agrawal, “A comprehensive survey of contemporary Arabic sentiment analysis: Methods, challenges, and future directions,” in Findings Assoc. Comput. Linguistics: NAACL 2025, 2025, pp. 3760–3772.
66. A. Alsayat and N. Elmitwally, “A comprehensive study for Arabic sentiment analysis (challenges and applications),” Egypt. Inform. J., vol. 21, no. 1, pp. 7–12, 2020.
67. Elnagar, Y. S. Khalifa, and A. Einea, “Hotel Arabic-reviews dataset construction for sentiment analysis applications,” in Intelligent Natural Language Processing: Trends and Applications, Cham, Switzerland: Springer Int. Publ., 2017, pp. 35–52.
68. N. Al-Twairesh, H. Al-Khalifa, A. Al-Salman, and Y. Al-Ohali, “AraSenTiTweet: A corpus for Arabic sentiment analysis of Saudi tweets,” Procedia Comput. Sci., vol. 117, pp. 63–72, 2017.
69. S. Rosenthal, N. Farra, and P. Nakov, “SemEval-2017 task 4: Sentiment analysis in Twitter,” arXiv:1912.00741, 2019.
70. M. Nabil, M. Aly, and A. Atiya, “ASTD: Arabic sentiment tweets dataset,” in Proc. EMNLP, 2015.
71. O. Obeid, W. Zaghouani, M. El-Haj, H. Bouamor, and N. Habash, “CAMeL Tools: An open-source Python toolkit for Arabic natural language processing,” in Proc. 12th Language Resources and Evaluation Conf. (LREC), 2020, pp. 7022–7032.

72. A. Abdelali, K. Darwish, N. Durrani, and H. Mubarak, “Farasa: A fast and furious segmenter for Arabic,” in Proc. NAACL-HLT Demonstrations, 2016, pp. 11–16.
73. Pasha et al., “MADAMIRA: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic,” in Proc. LREC, vol. 14, 2014, pp. 1094–1101.
74. M. Abdul-Mageed and M. Diab, “AWATIF: A multi-genre corpus for Modern Standard Arabic subjectivity and sentiment analysis,” in Proc. LREC, 2012.
75. M. Althobaiti, U. Kruschwitz, and M. Poesio, "AraNLP: A Java-based library for the processing of Arabic text," in Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014), Reykjavik, Iceland, May 2014, pp. 4134–4138.