

Review

A Short Review of Arabic Aspect-Based Sentiment Analysis: Methods, Challenges and Future Directions

Hamza Youseef ¹, Luis Gonzaga Baca Ruiz ^{2,*} , David Criado Ramón ¹ 
and María del Carmen Pegalajar Jiménez ¹ 

¹ Department of Computer Science and Artificial Intelligence, University of Granada, 18014 Granada, Spain; hamzehyousef@correo.ugr.es (H.Y.); dcriado@ugr.es (D.C.R.); mcarmen@decsai.ugr.es (M.d.C.P.J.)

² Department of Software Engineering, University of Granada, 18014 Granada, Spain

* Correspondence: bacaruiz@ugr.es

Abstract

The need for Arabic Aspect-Based Sentiment Analysis (ABSA) has grown steadily alongside the expansion of digital content, while the linguistic complexity of Modern Standard Arabic and its diverse dialects introduces significant challenges. However, progress in the field remains constrained by methodological fragmentation, inconsistent task definitions, heterogeneous datasets, and non-standardized evaluation practices. Based on a systematic analysis of 57 studies, this work presents an analytical and interpretive review that moves beyond performance-oriented surveys to examine the methodological foundations of Arabic ABSA research. The review follows a rigorous and transparent study selection process and applies a structured analytical framework to analyze task formulations, dataset characteristics, modeling approaches and evaluation strategies. Our findings reveal persistent challenges, including ambiguous aspect definitions, insufficiently documented annotation protocols, structural annotation biases, and limited robustness across domains and dialects. A heavy reliance on Transformer-based architectures and new Arabic foundation models can create an illusion of progress. Researchers often evaluate these models on small and homogeneous datasets. Consequently, strong in-domain performance obscures limited cross-domain and cross-dialectal generalizability. This study concludes by outlining actionable research directions, emphasizing clearer task standardization, more rigorous annotation guidelines, unified evaluation, and broader dialectal coverage to enhance reproducibility and scalability in Arabic ABSA systems.

Keywords: Arabic aspect-based sentiment analysis; Arabic natural language processing; sentiment analysis; transformer-based models; Arabic dialects



Academic Editor: Gianni D'Angelo

Received: 5 March 2026

Revised: 13 April 2026

Accepted: 16 April 2026

Published: 19 April 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Arabic has undergone a profound digital transformation in recent years, emerging as one of the most widely used languages online and generating a rapidly expanding volume of user-produced content across social and interactive platforms [1]. This expansion has amplified the need for computational methods capable of handling the full linguistic and sociolinguistic complexity of Arabic, which differs fundamentally from high-resource languages in both structural properties and patterns of digital use. A central characteristic of Arabic is its dual linguistic system: Modern Standard Arabic (MSA) as the formal written variety, and a continuum of dialects that dominate everyday communication. These

dialects exhibit extensive variation in phonology, morphology, and lexicon and lack standardized orthographic conventions, greatly complicating their treatment in natural language processing (NLP) pipelines [2].

The challenges of dialectal processing are magnified by the prevalence of mixed-code writing, Arabizi (Arabic text written using Latin characters and numerals), and fluid alternation between MSA and spoken varieties, especially in online contexts. Extensive analyses demonstrate that dialects such as Algerian Arabic employ hybrid lexical systems, frequent code-switching with French and Tamazight, and inconsistent orthographic practices, all of which reduce the effectiveness of MSA-centric NLP tools and require dialect-specific resources for reliable processing [3]. Moreover, Arabic online writing operates within a sociolinguistic environment shaped by diglossia, where dialectal forms (being more emotionally expressive, accessible, and cognitively natural) have increasingly become the default medium of digital interaction. This shift has blurred traditional boundaries between spoken and written Arabic and further diversified the linguistic landscape in which computational systems must operate [4].

In addition to dialectal variation, Arabic exhibits extensive reliance on figurative and metaphorical expressions that encode sentiment implicitly. Empirical studies show that state-of-the-art sentiment classifiers experience marked performance degradation on metaphor-rich inputs, reflecting their inability to capture deep semantic associations not directly observable in surface lexical polarity [5]. These limitations are especially consequential for ABSA, where models must identify implicit aspects, infer contextual polarity shifts, and interpret sentiment embedded in culturally specific figurative structures.

The broader Arabic NLP literature consistently identifies structural constraints that hinder methodological progress: limited availability of large, high-quality annotated datasets; inconsistent annotation guidelines; divergent task formulations; and the absence of standardized evaluation benchmarks [6]. As a result, most Arabic ABSA studies rely on small, domain-restricted datasets (often hotel reviews) and their findings are therefore difficult to compare or generalize. Comparative analyses of Arabic sentiment classification further demonstrate that model performance varies sharply across datasets and preprocessing configurations, with classical machine learning models frequently outperforming deep architectures due to data scarcity and domain dependence [7].

This stands in contrast to other data-driven domains, where advanced models achieve robust and generalizable performance when supported by large-scale, long-term datasets and guided by domain-specific constraints [8,9]. These observations underscore the need for methodological scrutiny in Arabic ABSA beyond numerical accuracy.

Compounding these issues, recent work on multilingual transformers shows that Arabic is underrepresented in large pretraining corpora, leading to unstable internal language representations and cross-lingual interference that undermines the reliability of ABSA predictions [10]. Taken together, these structural, linguistic, and methodological constraints position Arabic ABSA as a fragmented research area where empirical results lack comparability and deeper theoretical foundations remain insufficiently developed [11].

Within this context, recent surveys such as [1,6] provide illustrative overviews of Arabic NLP and deep neural network approaches and cover ABSA. However, their metric-focused perspectives leave important conceptual and methodological aspects insufficiently addressed.

Therefore, key elements like precise task definitions, systematic dataset construction, and rigorous annotation practices remain underexplored. As a result, this oversight restricts the ability to capture the linguistic complexity of Arabic, and it simultaneously hinders reliable and reproducible ABSA research. To address these gaps, the present study examines task formulations, dataset development, annotation strategies, and evaluation

practices across the Arabic ABSA literature. Finally, we establish a framework to highlight methodological strengths and weaknesses, and we thereby provide actionable directions to foster linguistically grounded research.

The structure of this paper is as follows: Section 2 presents the linguistic background and theoretical foundations relevant to Arabic ABSA. Section 3 describes the methodological approach employed in collecting and evaluating the literature. Section 4 synthesizes current research across tasks, datasets, annotation strategies, and evaluation practices. Section 5 provides a critical analysis of persisting challenges and structural limitations. Section 6 concludes with a summary of key insights and directions for future work.

2. Background

Arabic presents a uniquely challenging landscape for NLP, shaped by its complex morphology, orthographic ambiguity, rich system of clitics, and the coexistence of MSA with a wide spectrum of regional dialects. The morpho-syntactic structure of Arabic relies on a nonconcatenative root-pattern system that generates extensive lexical variation and sparsity, complicating fundamental tasks such as tokenization, segmentation, and part-of-speech tagging. The presence of proclitics and enclitics further increases ambiguity by producing long agglutinated forms that require careful morphological decomposition. These linguistic characteristics generate substantial difficulties for downstream sentiment analysis (SA) and especially ABSA, which depends on stable lexical and syntactic cues for identifying aspect terms and modeling polarity.

Dialectal variation compounds these structural challenges. Research on dialectal Arabic NLP demonstrates that dialects diverge from MSA (and from one another) at phonological, morphological, and lexical levels, often introducing irregular orthographic practices, code-mixed forms, and borrowed vocabulary that are not represented in standard resources [12]. The lack of normalized spelling conventions leads to multiple surface realizations of the same lexical item, inflating the out-of-vocabulary (OOV) rate and degrading the performance of tokenization and segmentation pipelines. Parallel corpus studies further show that dialects differ systematically in morphophonological patterns, cliticization behavior, and lexical choice, which undermines cross-dialect generalization and reduces the portability of NLP models across regions [13]. These divergences directly affect ABSA, where aspect cues and sentiment expressions are highly sensitive to lexical and morphological variation.

Early approaches to Arabic sentiment analysis relied on surface lexical cues or hand-crafted features, but their performance was hindered by limited domain coverage, weak generalization, and sensitivity to dataset noise. Subsequent work highlighted the OOV problem as a major obstacle, particularly in social media contexts characterized by inconsistent spelling, dialect variation, and creative orthographic forms. Embedding-expansion techniques have been proposed to mitigate sparsity by generating morphological and orthographic variants, enriching representation spaces and improving the robustness of classifiers across dialectal forms and informal text [14]. Complementary research on embedding quality has shown that lemma-based representations outperform word-level embeddings due to reduced morphological sparsity and improved semantic coherence, yielding higher stability and better downstream sentiment classification results [15]. These findings underscore the importance of morphological normalization and corpus diversity for producing reliable vector representations.

Transformer-based models have dramatically altered the landscape of Arabic SA, enabling context-aware encoding and large-scale pretraining on dialectal and social media corpora. Benchmarks across a wide range of pretrained Arabic models (such as AraBERT, MARBERT, ARBERT, QARiB and Arabic ALBERT) demonstrate that perfor-

mance is tightly coupled with the linguistic composition of the pretraining data. Models pretrained on dialect-rich corpora consistently outperform those trained primarily on MSA, highlighting the centrality of dialectal coverage for robust sentiment modeling [16–19]. However, even with transformers, sentiment and sarcasm detection remain sensitive to segmentation choices, morphological irregularities, and the heterogeneity of online language. Unified toolkits based on transformers have shown substantial improvements across multiple social media tasks, revealing that deep contextual encoders can compensate for resource scarcity and annotation inconsistencies but only when supported by sufficient dialect diversity and adequate preprocessing pipelines [20].

Beyond token-level modeling, research has increasingly recognized that ABSA requires discourse-level understanding. Reviews often contain multiple sentences whose sentiments interact, and sentence-level models struggle with ambiguity when aspects appear in contexts that depend on earlier or later discourse. Hierarchical architectures that embed sentences and aspects within multi-sentence structures have shown marked improvements by incorporating global coherence and long-distance dependencies [21]. Earlier neural systems for ABSA demonstrated that even simple convolution-based classifiers can perform competitively across languages when aspect information is incorporated into token representations. Nevertheless, they exhibit limitations in handling conflicting sentiments across multiple aspects within a sentence [22].

To better understand the impact of tokenization on Arabic NLP, extensive comparative studies have analyzed character-, word-, subword-, and morphology-aware tokenization strategies. Results demonstrate that tokenization is a task-dependent variable and that no single approach is universally optimal. Word tokenization performs best in low-resource sentiment classification tasks, while morphology-based tokenizers improve performance in structurally complex domains. Character-level tokenization is effective for highly phonological tasks such as poem-meter classification. These findings highlight the necessity of aligning tokenization choices with task characteristics and linguistic properties [23].

Finally, the broader challenges identified across dialectal NLP—resource scarcity, annotation inconsistency, orthographic instability, and weak cross-dialect transfer—remain directly relevant to the current limitations of Arabic ABSA research. Surveys of dialectal Arabic NLP highlight persistent fragmentation in datasets, uneven dialect representation, and inadequate methodological grounding for cross-dialect modeling [24]. Together, these linguistic, computational, and methodological constraints explain why Arabic ABSA continues to exhibit high variance across datasets, unstable evaluation practices, and limited generalizability. Addressing these foundations is essential for developing robust, interpretable, and linguistically grounded ABSA systems.

3. Methodology

This review adopts the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework to ensure a structured, transparent, and reproducible methodology. The review process consisted of three main phases: identification, screening, and eligibility assessment.

The database searches were conducted between January and December 2025. During the screening and eligibility phases, the approximately 150 retrieved records were reviewed. A total of 93 records from this initial search were excluded from the primary review corpus because they were either irrelevant to Arabic ABSA, written in Arabizi, lacked clear task specifications, or provided insufficient methodological details. This selection process resulted in a final corpus of 57 studies for detailed quantitative and qualitative

synthesis. We would like to note that additional background literature and foundational surveys cited in this manuscript were sourced outside this search.

For the data extraction and coding phase, we developed a standardized protocol to capture key variables from the selected studies. Specifically, we extracted data concerning task formulations (single and multi-task), dataset characteristics (e.g., domain focus, linguistic coverage), modeling approaches (classical machine learning, deep learning, Transformer-based architectures), and evaluation metrics. We independently extracted the information and subsequently cross-checked the coded variables to resolve any discrepancies and mitigate potential bias. The complete study selection process is visually summarized in the PRISMA flow diagram provided in the Figure 1.

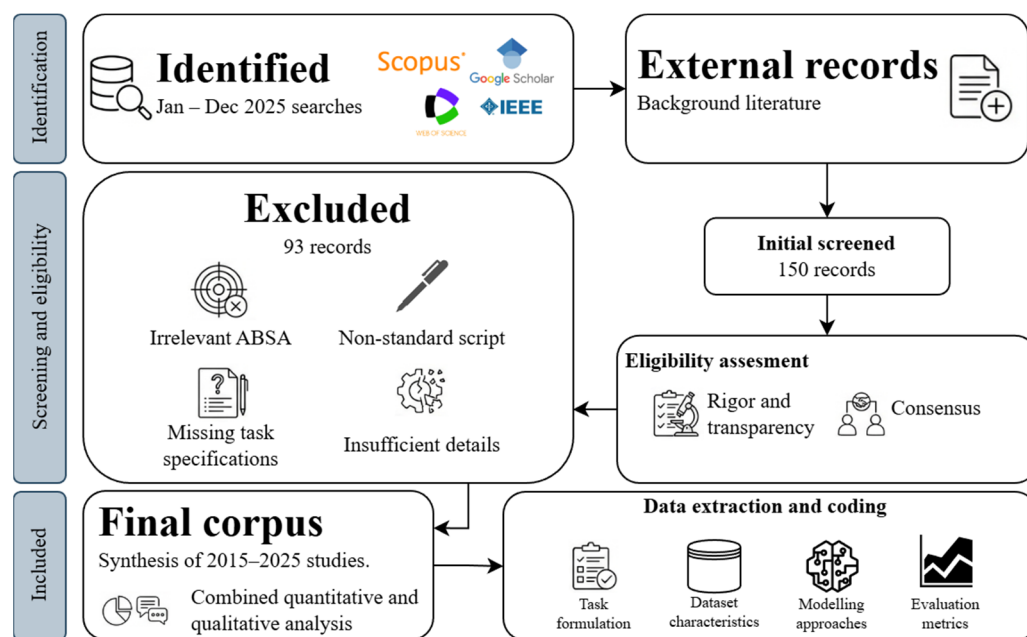


Figure 1. PRISMA diagram of the research methodology followed.

The conducted search across scholarly databases includes Scopus, Web of Science, IEEE Xplore, ACM Digital Library, and Google Scholar. To capture key aspects of ABSA, we crafted specific Boolean queries. These queries featured terms like “Arabic sentiment analysis”, “aspect extraction” and “aspect-level polarity classification”. This initial search retrieved approximately 150 records.

Next, we examined titles and abstracts to filter out studies irrelevant to Arabic ABSA. For instance, we excluded works focused on general sentiment analysis without aspect-level formulations. Furthermore, we removed studies written in Arabizi. Authors write Arabizi with Latin characters and numerals, and this format does not conform to the Arabic script targeted in our review. The exclusion of such studies from the core review corpus was a deliberate decision. Although Arabizi is highly prevalent in informal online communication, processing it introduces a fundamentally different set of natural language processing challenges. Most notably, handling Arabizi typically requires an intermediate transliteration step to convert Latin characters and numerals back into Arabic script before standard morphological analyzers or Arabic-specific pretrained language models can be effectively applied. Moreover, the extreme lack of standardized orthography in Arabizi diverges significantly from the primary focus of this review. Addressing the unique pipelines required for Arabizi would require a dedicated review.

Subsequently, we performed an eligibility assessment. We reviewed full-text articles to evaluate methodological transparency and empirical rigor. Therefore, we retained studies that explicitly defined ABSA tasks, detailed dataset construction, and described anno-

tation and evaluation procedures. On the other hand, we excluded articles without clear task specifications or sufficient methodological details. Throughout this stage, all authors cross-checked the inclusion and exclusion decisions. This collaborative step ensured consensus and reduced selection bias.

Finally, we established the definitive corpus. After we applied these procedures, we selected 57 studies published between 2015 and 2025 for detailed analysis. Readers can find a comprehensive summary of these works in Table A2 in Appendix A. This table provides key attributes and publication details.

4. Analytical Overview

Following the study selection and coding procedures described in Section 3, this section provides an analytical overview of the reviewed Arabic ABSA literature. Rather than listing individual studies or reporting isolated performance results, we focus on synthesizing recurring methodological patterns identified across the selected corpus. To support this analysis, Table 1 presents a structured mapping of the reviewed studies according to the core analytical dimensions examined in this section. A summarized representation is provided in Table 1, while the complete coding details are available in Table A1.

Table 1. Summary of modeling approaches and datasets.

Method Category	N	Approach	Dataset
ML	9	ML models (SVM, CRF, J48) on T1–T4 handle few ABSA tasks each, as full coverage needs richer features or advanced architectures.	Mainly benchmark datasets (SemEval-2016) with domain-specific review corpora; minimal cross-domain evaluation.
DL	28	Trend: Arabic ABSA evolved from RNNs (Bi-LSTM/BiGRU with embeddings) to Transformer-based models (AraBERT, MARBERT, QARiB) with fine-tuning and attention, reflecting state-of-the-art modeling.	Datasets cover hotels, healthcare, tourism, social media, news, and education, with hotel reviews being most frequent; several include multi-domain or augmented resources for Arabic ABSA.
Lexicon-based	5	Studies focused on polarity, frequency, and linguistic rules, using heuristic, lexicon, and rule-based techniques to capture Arabic sentiment patterns.	All datasets are in MSA, covering book reviews, hotel/product reviews, and news or social media posts.
Hybrid	15	Hybrid approaches combine ML, DL, Transformers, CRF, GCN, ontologies, and lexicon/rule-based techniques, often integrating pretrained embeddings, attention mechanisms, and ensemble strategies to enhance Arabic ABSA performance.	Datasets cover hotel reviews, book reviews, game reviews, social media posts, and customer feedback, with the largest focus on hotel reviews from SemEval-2016 and HAAD.

The overview is structured around five core analytical dimensions: task formulations, dataset characteristics, annotation practices, modeling approaches, and evaluation protocols. These dimensions reflect the principal stages at which methodological decisions are made and where variation across studies is most pronounced. Examining them together enables a comparative perspective that reveals dominant research patterns as well as implicit assumptions underlying empirical implementations. For instance, the modeling choices documented in Table 1 clearly reflect a significant shift toward transformer-based architectures, but this trend often occurs without a corresponding diversification in dataset domains, a pattern that has significant implications for the field's progress.

By abstracting across individual studies, this section provides a coherent characterization of how Arabic ABSA has been operationalized in practice. The resulting synthe-

sis establishes an empirical foundation for the critical discussion in Section 5, where the implications of these methodological patterns are examined in relation to reproducibility, robustness, and linguistic adequacy.

4.1. Task Formulations

Across the reviewed literature, Arabic ABSA is predominantly framed as a set of inter-related subtasks rather than a single end-to-end problem. Researchers commonly decompose ABSA into components such as aspect-level polarity classification, aspect or opinion target extraction, and aspect category detection, reflecting a modular approach to opinion analysis. This decomposition allows for more focused modeling and evaluation at each stage, but it also introduces variability in task definitions, annotation schemes, and performance metrics, making cross-study comparisons challenging.

Figure 2 illustrates this conceptual decomposition, highlighting how Arabic ABSA is commonly operationalized through distinct but interrelated task formulations. Rather than depicting a sequential processing pipeline, the figure emphasizes the analytical separation of subtasks that characterizes much of the existing literature.

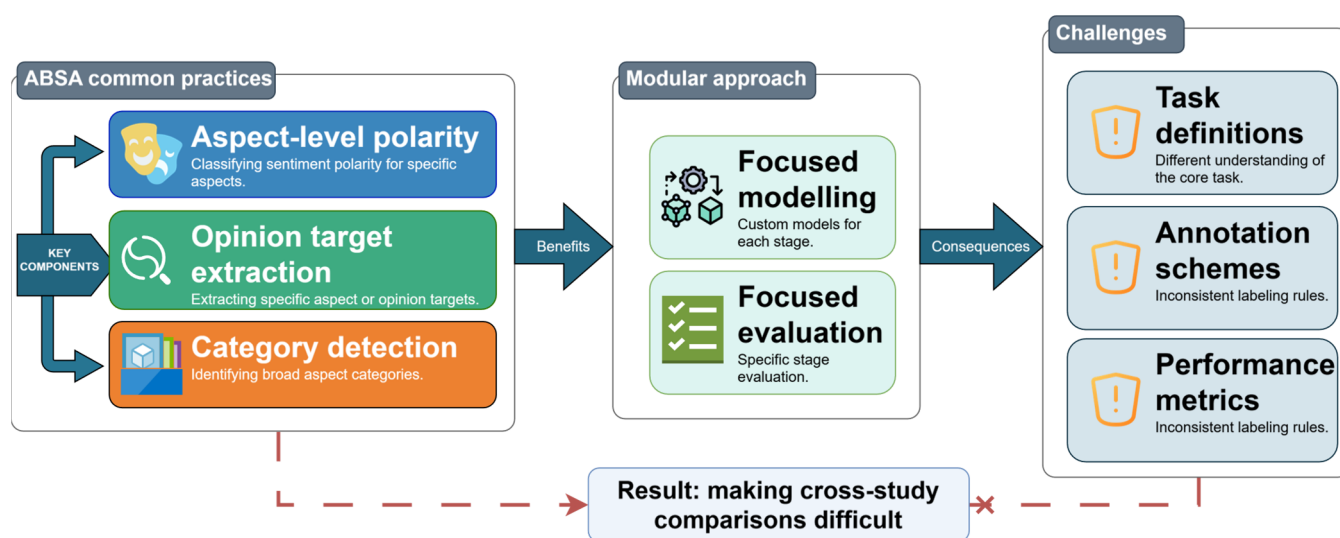


Figure 2. Conceptual decomposition of Arabic ABSA into commonly adopted task formulations.

Aspect-level polarity classification occupies a central position and is frequently treated as the primary objective, even when combined with other subtasks. Extraction-oriented formulations remain common but are typically evaluated independently, while fully end-to-end task definitions that jointly identify aspects and assign polarity are comparatively rare [25]. Collectively, these patterns underscore the prevalence of task decomposition as a defining methodological choice in Arabic ABSA research.

4.2. Datasets

Datasets constitute a foundational component of Arabic ABSA, as they directly shape task formulation, modeling choices, and evaluation outcomes. According to the reviewed literature, dataset construction practices exhibit a high degree of concentration and methodological homogeneity, revealing structural constraints that limit the generalizability and robustness of empirical findings.

A dominant characteristic of Arabic ABSA datasets is their narrow domain focus. Most datasets are derived from consumer review platforms, particularly hotel and restaurant reviews [26]. While such domains provide relatively explicit aspect mentions, they may introduce domain bias and limit transferability to other genres. This imbalance is vi-

sually summarized in Figure 2, where the domain dimension highlights the concentration of existing datasets within a small set of recurring application contexts.

Beyond domain concentration, Figure 3 illustrates the limited coverage of Arabic linguistic variety in current ABSA datasets. Most resources are MSA-dominant or restricted to a single dialectal variety, implicitly assuming linguistic homogeneity. In contrast, datasets designed to support multi-dialect or cross-variety evaluation occupy a marginal position in the figure, reflecting their scarcity in the literature.

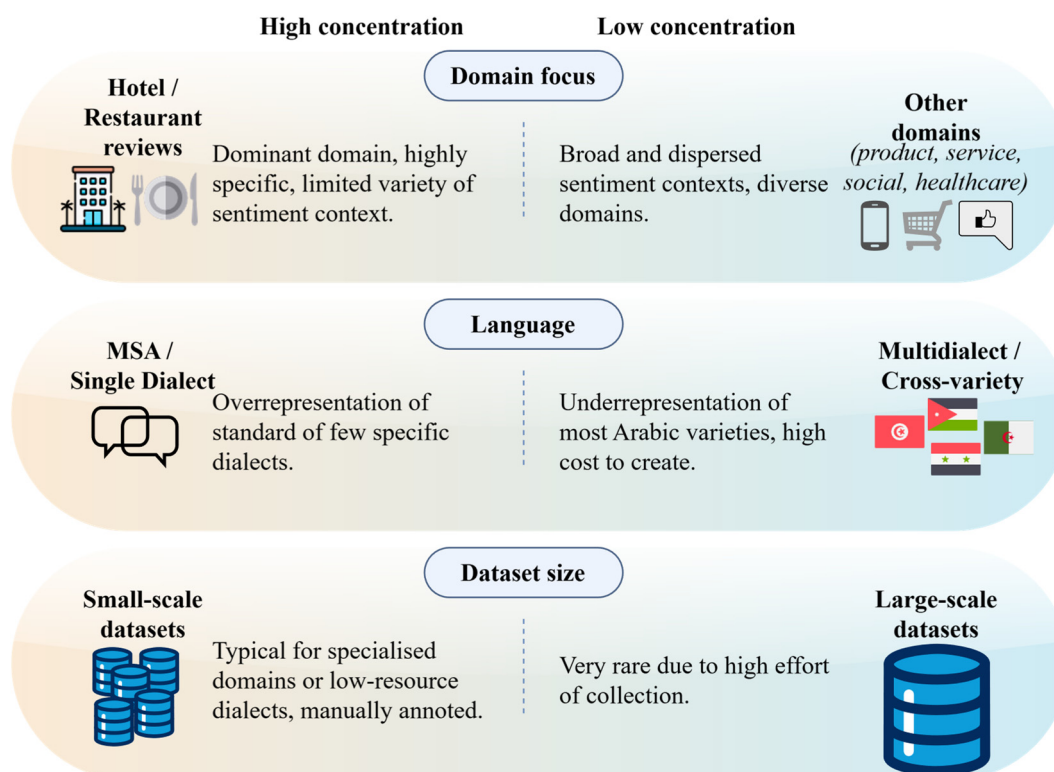


Figure 3. Structural concentration of datasets in Arabic ABSA.

Dataset scale represents a third constraining dimension. As depicted in Figure 3, small-scale datasets dominate the current landscape, reinforcing the interaction between limited data size, narrow domain coverage, and restricted linguistic diversity. Together, these visual patterns emphasize that many empirical results in Arabic ABSA are conditioned by structural properties of the data rather than by modeling choices alone.

4.3. Annotation Practices

Now, turning to the conceptual abstraction, Figure 3 depicts the main sources of annotation variability observed in Arabic ABSA research. Annotation practices play a decisive role in shaping the empirical foundations of Arabic ABSA, as they determine how linguistic phenomena are transformed into computationally tractable representations [27]. Rather than functioning as a neutral preprocessing step, annotation introduces a series of methodological decisions that directly influence model behavior, evaluation validity, and the comparability of results across studies. The figure places the annotated dataset at the center, emphasizing that the final labeled resource is the outcome of interacting design choices rather than an objective ground truth [27].

One major source of variability arises from ambiguity in task definition. As illustrated in Figure 4, studies differ in how they conceptualize core units such as aspects, aspect terms, and aspect categories. Decisions regarding whether to annotate only explicit mentions or

to allow implicit aspects, as well as whether annotation is performed at the term or category level, lead to structurally different datasets [28], even when similar source texts are used.

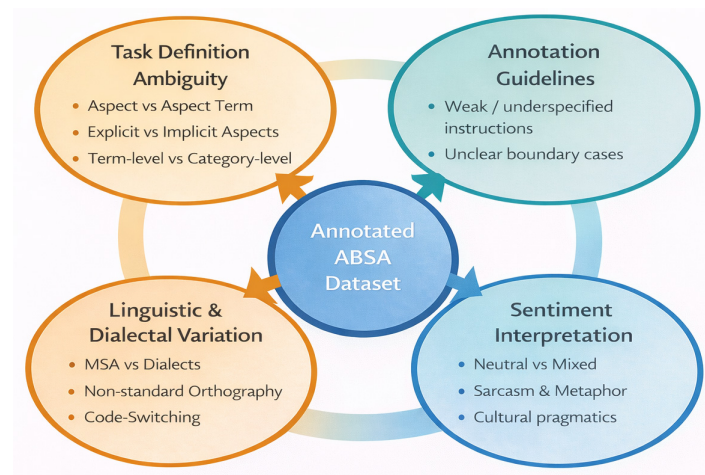


Figure 4. Sources of annotation variability in Arabic ABSA.

Annotation guidelines constitute a second influential factor. Figure 4 highlights how underspecified or weakly documented instructions introduce inconsistencies in annotator decisions, particularly in borderline cases. Limited reporting on annotator training, disagreement resolution, and quality control further obscures the logic underlying the resulting labels, reducing reproducibility across datasets [29].

Linguistic and dialectal variation represents a third dimension of annotation variability. As shown in Figure 3, annotators are often required to interpret non-standard orthography, dialectal expressions and code-switched content. In the absence of dialect-aware annotation guidelines, such variation is frequently handled implicitly, introducing latent biases that favor MSA-like constructions and marginalize dialect-specific sentiment expressions [30,31].

Finally, it underscores the role of subjective sentiment interpretation. Differences in how annotators distinguish neutral from mixed sentiment or interpret sarcasm, metaphor, and culturally embedded expressions [32] introduce additional layers of inconsistency. These interpretive differences propagate directly into model training and evaluation, contributing to performance variance that cannot be attributed solely to modeling choices.

Taken together, the interactions depicted in Figure 3 demonstrate that annotation variability in Arabic ABSA is cumulative and systemic [29]. Without clearer task definitions, explicitly documented guidelines, and linguistically informed annotation strategies, advances in modeling architectures risk being grounded in datasets whose internal consistency and comparability remain limited.

4.4. Modeling Approaches

Modeling approaches in Arabic ABSA have evolved alongside advancements in natural language processing, but they remain strongly conditioned by upstream methodological factors, particularly dataset construction and annotation design. Early approaches and baseline evaluations in Arabic ABSA heavily relied on traditional feature engineering and classical machine learning classifiers, such as Support Vector Machine (SVM), Naive Bayes and Random Forests, sometimes supplemented by sentiment lexicons [33–37]. While computationally efficient, researchers quickly identified the need for more complex architectures to capture the rich morphological context of Arabic. Neural architectures subsequently enabled contextual encoding and joint learning of related subtasks, provid-

ing greater modeling flexibility. But empirical gains were often inconsistent and sensitive to dataset size and annotation stability.

The adoption of transformer-based language models, such as AraBERT and MARBERT, represents the most outstanding shift in the field, offering contextualized representations that partially address morphological sparsity and lexical variation. When pre-trained on dialect-rich or social media corpora, these models improve aspect-level task performance; nevertheless, gains remain uneven across datasets and domains, reflecting sensitivity to pretraining-finetuning alignment, annotation quality, and task formulation. Many studies employ transformers as direct replacements for earlier architectures without revisiting underlying task definitions or annotation assumptions, potentially amplifying existing dataset biases rather than mitigating them.

Overall, progress in Arabic ABSA modeling is tightly coupled with upstream methodological choices. Architectural advances alone yield diminishing returns when applied to narrowly scoped datasets with unstable annotation foundations. Sustainable improvements therefore require a principled alignment between modeling strategies, dataset design, and annotation practices to ensure reproducibility, robustness, and meaningful generalization [11].

4.5. Evaluation Practices

Evaluation practices in Arabic ABSA constitute a major source of fragmentation and limited comparability across studies. Although most research adopts widely used evaluation metrics, inconsistencies in experimental design, reporting conventions, and evaluation scope significantly constrain the interpretability and reproducibility of empirical findings.

Aspect-level sentiment classification is commonly evaluated using standard metrics such as precision, recall and F1-score, occasionally supplemented by accuracy. However, these metrics are often applied under heterogeneous conditions, including different data splits, preprocessing pipelines, and label mappings that are not always fully documented. As a result, reported performance differences frequently reflect experimental configuration choices rather than substantive methodological improvements.

A recurring limitation lies in the reliance on aggregate performance scores. Most studies report overall metrics averaged across all aspects and sentiment classes, without providing fine-grained analysis by aspect category, sentiment polarity, domain, or dialect. This practice obscures systematic weaknesses, particularly in minority classes, implicit aspects, and dialect-specific sentiment expressions. Consequently, models may appear robust despite exhibiting substantial failure modes under more challenging linguistic or contextual conditions.

Evaluation practices are further constrained by limited cross-dataset and cross-domain testing. Many studies evaluate models exclusively on a single dataset, often using random train–test splits drawn from the same domain and annotation scheme. Such evaluations tend to overestimate generalization performance and provide little insight into a model’s ability to transfer across domains, dialects, or alternative annotation conventions. When cross-dataset evaluation is conducted, it frequently reveals sharp performance degradation, underscoring the dataset-dependent nature of reported results.

Additional inconsistency arises from the evaluation of heterogeneous task formulations. Aspect extraction, aspect category detection, and aspect-level polarity classification are sometimes assessed using incompatible metrics or loosely defined success criteria. In joint or multi-task settings, evaluation protocols are rarely standardized, and trade-offs between subtasks are not systematically analyzed, further complicating meaningful comparison across studies.

Reproducibility is also hindered by incomplete reporting of experimental details. Information regarding preprocessing steps, hyperparameter tuning, random seed control, and evaluation scripts is often omitted or insufficiently specified. Even when identical datasets are employed, these omissions make it difficult to replicate reported results or to assess whether observed improvements are statistically meaningful.

Overall, evaluation practices in Arabic ABSA remain largely performance-driven and insufficiently standardized. The predominant focus on headline scores, single-dataset evaluation, and underspecified experimental setups limits the validity and comparability of empirical claims. Without evaluation protocols that explicitly account for domain variation, dialectal diversity, and annotation heterogeneity, reported results risk overstating methodological progress. Strengthening evaluation practices is therefore essential for establishing reliable benchmarks and for enabling cumulative and reproducible advancement in Arabic ABSA research [14].

5. Critical Synthesis and Open Challenges

This section provides a synthesis of the methodological patterns identified across the reviewed Arabic ABSA literature. Moving beyond descriptive analysis, it integrates insights from task formulation, dataset design, annotation practices, modeling approaches, and evaluation protocols to examine their combined influence on research outcomes.

Rather than revisiting individual findings, the analysis focuses on explaining the persistence of recurring methodological patterns and their implications for robustness, comparability, and generalization. By emphasizing structural interdependencies and shared methodological constraints, the following subsections identify the principal challenges shaping current research trajectories and outline directions toward more coherent and sustainable development in Arabic ABSA research.

5.1. Illusion of Progress

A recurring theme across contemporary Arabic ABSA research is the perception of steady methodological progress, primarily reflected through incremental improvements in reported performance metrics [38–40]. At first glance, this trajectory suggests a rapidly advancing field, particularly following the widespread adoption of deep neural networks and transformer-based architectures [1]. However, closer examination of the empirical conditions under which these improvements are obtained reveals a more nuanced reality.

Much of the reported progress in Arabic ABSA is achieved within tightly constrained experimental settings. Models are typically trained and evaluated on small, domain-specific datasets drawn from relatively homogeneous sources such as hotels, restaurants, or product reviews [41–43]. Evaluation protocols frequently rely on random train–test splits within the same dataset, ensuring substantial lexical, topical, and stylistic overlap between training and evaluation data [44]. Under such conditions, even modest architectural refinements or hyperparameter adjustments can yield measurable performance gains, creating an appearance of sustained advancement.

This dynamic gives rise to an illusion of progress in which improvements primarily reflect increased compatibility between models and narrowly scoped datasets rather than genuine advances in aspect-level sentiment understanding. The effect is further reinforced by the sequential adoption of increasingly complex architectures, which are often evaluated against baselines sharing the same dataset biases and annotation assumptions. Consequently, performance gains are frequently interpreted as evidence of architectural superiority, even when they largely stem from representational advantages within a fixed and limited data regime. When models are exposed to more challenging conditions (such as cross-domain settings, alternative annotation schemes or dialectally diverse in-

puts), their performance often degrades substantially, although such evaluations remain relatively uncommon.

Importantly, this dynamic does not imply that recent modeling advances lack value. Rather, it highlights a misalignment between what is measured and what constitutes meaningful progress for Arabic ABSA. Improvements in in-domain accuracy do not necessarily translate into robustness, generalization or practical applicability across the diverse linguistic and contextual landscape of Arabic. Without evaluation frameworks that explicitly test these dimensions, reported gains risk overstating the maturity of the field.

Taken together, the evidence suggests that much of the perceived progress in Arabic ABSA is fragile and context-dependent. Addressing this illusion requires shifting evaluative emphasis away from isolated performance improvements toward more demanding experimental designs that foreground generalization, linguistic diversity, and methodological transparency. Only under such conditions can progress be assessed in a manner that reflects substantive advances rather than incremental optimization within constrained settings.

5.2. Annotation as a Hidden Bottleneck

Among the structural factors constraining progress in Arabic ABSA, annotation practices constitute one of the most influential yet least explicitly addressed bottlenecks [45]. While annotation is often treated as a technical prerequisite for model training, the analysis of the reviewed corpus reveals that annotation design decisions exert a profound and lasting impact on dataset composition, model behavior and evaluation outcomes.

A primary source of instability lies in the absence of standardized task definitions at the annotation level. Across studies, core concepts such as aspect, aspect term, opinion target, and aspect category are operationalized in markedly different ways [46,47]. Some datasets restrict annotation to explicitly mentioned aspect terms, while others allow implicit or inferred aspects without clearly articulated criteria [48]. These definitional discrepancies result in datasets that encode fundamentally different assumptions, even when they are described under the same ABSA task label.

Annotation guidelines are frequently underspecified or insufficiently documented [49]. Many studies provide limited information regarding annotator training, decision rules for ambiguous cases, or procedures for resolving disagreement. As a consequence, it is often unclear whether annotated labels reflect consistent linguistic judgments or ad hoc interpretations shaped by individual annotator intuition. This opacity limits the interpretability of the resulting datasets and undermines confidence in downstream experimental results.

The challenges of annotation are further amplified by the linguistic characteristics of Arabic. Dialectal variation, non-standard orthography, code-switching, and culturally embedded sentiment expressions introduce ambiguity that cannot be resolved without explicit, linguistically informed guidelines. However, the reviewed studies rarely describe how such phenomena are handled during annotation. In practice, this often leads to implicit normalization strategies or unacknowledged biases that privilege MSA-like constructions, marginalizing dialect-specific sentiment cues.

These annotation-related inconsistencies propagate directly into the modeling and evaluation stages. Models trained on datasets with unstable or weakly specified annotation schemes may achieve high performance under matched evaluation conditions, yet their behavior becomes difficult to interpret or compare across studies. Differences in reported results may thus reflect variation in annotation logic rather than genuine differences in modeling effectiveness.

Crucially, annotation bottlenecks are rarely foregrounded in discussions of experimental limitations. Inter-annotator agreement scores, when reported, are typically pre-

sented as validation indicators without deeper analysis of disagreement patterns or sources of ambiguity [50]. This practice obscures the extent to which annotation uncertainty shapes empirical outcomes and contributes to inflated perceptions of model reliability.

Taken together, these observations indicate that annotation practices represent a hidden but central constraint on Arabic ABSA research. Without clearer task definitions, transparent documentation of annotation procedures, and explicit consideration of linguistic and dialectal variability, improvements in modeling architecture risk being evaluated against unstable ground truths. Addressing annotation as a first-class methodological concern is therefore essential for enabling meaningful comparison, reproducibility, and cumulative progress.

5.3. Dataset Bias and Limited Linguistic Coverage

Dataset design constitutes a second structural axis through which limitations in Arabic ABSA research are reinforced. Across the reviewed studies, datasets consistently exhibit a narrow scope along multiple dimensions, including domain coverage, linguistic variety and scale [51]. These design characteristics extend beyond practical constraints, strongly influencing what models learn and how their performance is interpreted [52].

Domain bias is particularly pronounced. A large proportion of Arabic ABSA datasets originate from a limited set of recurring application domains, most notably hotel, restaurant and product reviews [53]. For instance, 30 out of 57 studies focus on the hotel domain, highlighting the dominance of this single domain. Although these domains provide relatively explicit sentiment cues, they exhibit restricted lexical and syntactic diversity. Consequently, models trained on these datasets tend to internalize domain-specific patterns, achieving strong in-domain performance while remaining less robust when applied to other genres such as social media discourse, political commentary, or informal user-generated content where sentiment is often implicit or context-dependent [54].

Linguistic coverage further constrains dataset representativeness. Despite the central role of dialectal Arabic in online communication, many datasets remain dominated by MSA or a single dialectal variety. Among the reviewed studies, 36 focus on MSA, 18 employ mixed linguistic settings and only 3 explicitly address dialectal Arabic, reflecting a critical underrepresentation of real-world varieties. Multi-dialect and cross-variety resources are comparatively rare, and explicit reporting of linguistic composition is frequently limited. This imbalance encourages models to implicitly assume linguistic homogeneity [55], reducing their ability to generalize across the Arabic dialect continuum, where variation functions as a core linguistic property rather than peripheral noise.

Dataset scale compounds these challenges, as many Arabic ABSA datasets are relatively small, limiting exposure to diverse sentiment expressions and increasing sensitivity to annotation noise and sampling artifacts. Under these conditions, models may achieve high performance by exploiting recurring lexical patterns rather than developing robust aspect-sentiment representations. This tendency is further reinforced by the repeated reliance on a limited set of benchmark datasets. For instance, nearly 50% of the reviewed studies focus on the hotel domain (see Table A2 for a comprehensive breakdown). Such a constrained data regime promotes incremental optimization while limiting exploration of broader linguistic and contextual variability. Moreover, these data-related limitations are rarely framed as central methodological challenges and are often acknowledged only briefly in discussions of experimental limitations. Nevertheless, their cumulative impact is substantial: dataset bias shapes model behavior, constrains evaluation outcomes and may create a misleading impression of progress. Without deliberate efforts to diversify domains, expand linguistic coverage, and improve dataset documentation practices, Ara-

bic ABSA research risks remaining confined to a narrow empirical space that inadequately reflects real-world language use.

5.4. Evaluation Inflation and Reproducibility Challenges

Evaluation practices play a decisive role in sustaining methodological fragility within Arabic ABSA. The reviewed literature reveals a strong emphasis on headline performance metrics, typically reported as aggregate precision, recall, or F1-scores. While such metrics provide a convenient summary of model behavior, their widespread use without complementary diagnostic analysis significantly limits interpretability.

A central concern is the predominance of single-dataset, in-domain evaluation. Models are commonly assessed using random splits drawn from the same dataset on which they are trained, guaranteeing substantial overlap in lexical, topical, and stylistic properties [39,54,56]. Under these conditions, evaluation metrics often overestimate robustness and obscure the extent to which performance depends on dataset-specific characteristics. Cross-domain and cross-dialect evaluation, which would provide a more stringent test of generalization, remains the exception rather than the norm.

Evaluation inflation is further intensified by limited granularity in reporting. Most studies present overall performance scores without disaggregating results by aspect type, sentiment polarity, domain, or linguistic variety [51]. This practice hides systematic weaknesses, particularly in minority sentiment classes, implicit aspects, and dialect-specific expressions. Consequently, models that perform well on dominant patterns may appear broadly effective despite failing in precisely the cases that are most challenging and practically relevant.

Reproducibility concerns compound these issues. Experimental details such as pre-processing steps, hyperparameter selection, random seed control, and evaluation scripts are often insufficiently specified [53]. Even when identical datasets are used, these omissions make it difficult to replicate reported results or to assess whether observed improvements are statistically meaningful. The lack of standardized benchmarks and evaluation protocols further undermines cumulative comparison across studies.

Taken as a whole, these evaluation practices contribute to an inflated perception of progress while weakening the evidentiary basis of empirical claims. Without more rigorous, transparent and generalization-oriented evaluation frameworks, it remains difficult to distinguish genuine methodological advances from performance gains driven by experimental convenience.

5.5. Methodological Coherence

The challenges spanning annotation, dataset design, modeling, and evaluation are not isolated shortcomings but mutually reinforcing aspects of a fragmented Arabic ABSA research landscape. Addressing them demands a shift from piecemeal optimizations toward methodological coherence across the entire research pipeline.

At the task formulation level, the reviewed studies reveal notable fragmentation: 21 out of 57 studies (37%) focus on single tasks, 29 studies (51%) implement partial multi-task setups covering two or three subtasks, and 6 studies (11%) pursue fully integrated end-to-end ABSA pipelines. Such distribution highlights the underutilization of comprehensive approaches that unify multiple ABSA components, emphasizing the need for clarity and consistency in task definitions and assumptions.

Annotation practices should be treated as a central methodological concern, supported by transparent guidelines, linguistically informed decision rules, and explicit reporting of uncertainty and disagreement. Dataset design should prioritize diversity in both domain and linguistic coverage: 36 studies (63%) focus exclusively on MSA, 18 stud-

ies (32%) employ mixed linguistic settings, and 3 studies (5%) address dialectal Arabic, reflecting a critical underrepresentation of real-world varieties. Thorough reporting of data sources, composition, and limitations is essential to improve reproducibility and interpretability. Evaluation protocols should move beyond in-domain performance to test robustness, transferability, and sensitivity to linguistic variation. Cross-dataset and cross-dialect evaluation should be integral components of empirical validation rather than optional extensions.

It is crucial to mention that methodological coherence does not imply a rigid standardization or the imposition of a single “correct” approach. Instead, it calls for shared design principles that make assumptions explicit, results interpretable and comparisons meaningful. By aligning modeling innovation with principled data, annotation, and evaluation practices, Arabic ABSA research can move beyond incremental performance gains toward sustainable, reproducible, and linguistically grounded progress.

6. Conclusions and Future Directions

This paper has presented a comprehensive analytical and interpretative review of Arabic ABSA, grounded in a structured synthesis of task formulations, dataset design choices, annotation practices, modeling approaches and evaluation protocols. By moving beyond performance-oriented comparisons, this study has sought to uncover the structural and methodological forces that shape empirical outcomes and condition perceived progress in the field.

Our analysis demonstrates that many of the limitations observed in contemporary ABSA research are not primarily attributable to modeling capacity but rather to upstream methodological design. Fragmented task definitions, narrowly scoped and linguistically homogeneous datasets, inconsistent annotations practices, and evaluation protocols centered on in-domain performance collectively constrain robustness, generalization, and comparability. Within this context, recent advances in neural and transformer-based architectures, while valuable, often operate within restricted empirical regimes that amplify apparent gains without resolving deeper foundational challenges.

A central insight of this review is that progress in Arabic ABSA cannot be sustained through architectural innovation alone. Without coherent alignment between task design, data construction, annotation logic, and evaluation strategy, improvements remain context-dependent and difficult to interpret. The persistent reuse of limited benchmarks, coupled with underreported methodological variability, further reinforces research silos and impedes cumulative advancement. Looking forward, future research in Arabic ABSA should prioritize methodological rigor alongside modeling innovation. This includes the development of clearly specified and comparable task formulations, transparent and linguistically informed annotation frameworks, and datasets that reflect the full diversity of Arabic language use across domains and dialects. Evaluation practices should be redesigned to foreground robustness, transferability, and reproducibility, with greater emphasis on cross-domain and cross-dialect testing. More broadly, advancing Arabic ABSA requires a shift in research culture, from optimizing headline scores toward building reusable, interpretable, and linguistically grounded systems. By adopting shared methodological principles and making design assumptions explicit, the field can move beyond incremental performance gains and toward sustainable progress with practical relevance for real-world Arabic language understanding applications.

Author Contributions: Conceptualization, M.d.C.P.J. and L.G.B.R.; methodology, H.Y.; software, H.Y.; validation, H.Y., D.C.R. and L.G.B.R.; formal analysis, M.d.C.P.J. and L.G.B.R.; investigation, H.Y.; resources, H.Y.; data curation, H.Y. and L.G.B.R.; writing—original draft preparation, H.Y. and L.G.B.R.; writing—review and editing, H.Y., D.C.R., M.d.C.P.J. and L.G.B.R.; visualization, H.Y. and

L.G.B.R.; supervision, D.C.R., M.d.C.P.J. and L.G.B.R.; project administration, M.d.C.P.J. and L.G.B.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were create or analyzed in this study. Data sharing is not applicable to this study.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ABSA	Aspect-Based Sentiment Analysis
BERT	Bidirectional Encoder Representations from Transformers
DL	Deep Learning
LSTM	Long Short-Term Memory
ML	Machine Learning
MSA	Modern Standard Arabic
NLP	Natural Language Processing
OOV	Out-of-vocabulary
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analysis
SA	Sentiment Analysis
SVM	Support Vector Machine

Appendix A

Table A1. Definition of ABSA task types.

Task	Task Name	Description
T1	AE (Aspect Extraction)	Identifying aspect terms in the text.
T2	ALC (Aspect-Level Classification)	Determining the sentiment (Positive/Negative/Neutral) for a given aspect.
T3	ATE (Aspect Term Extraction)	Extracting specific terms and their corresponding categories.
T4	E2E (End-to-End ABSA)	Jointly extracting aspects and predicting their sentiments simultaneously.

Table A2. Summary of reviewed ABSA studies categorized by task type, methodology, dataset domain and reported results.

Ref	Task Type	Method Category	Dataset Source and Domain	Result
[57]	T2	Hybrid: CNN/LSTM + SVM/NB	MSA: Book Reviews	LSTM > Deep > Classical ML (LABR)
[33]	T1–T4	Lexicon: Enhanced Lexicon	MSA: News Posts	F1: 0.424
[34]	T1, T2	ML: SVM + AraVec	Saudi Dialect: Airline Tweets	F1: 0.79 (AD)
[58]	T1–T4	Transformer: FAST-LCF-ATEPC + AraBERT/MARBERT/QARib	MSA: Edu Service Surveys	Best Aspect F1: 0.58 Polarity Acc: 0.84 Cat F1: 0.98
[59]	T2	Hybrid: AraBERT + LSTM-GRU + Attention	MSA: Hotel Reviews	Acc: 94.71% (SOTA) Neutral F1: 0.18
[26]	T2, T3	ML: MNB/SVM/LR/DT/RF	MSA + Dialectal: M-App Reviews	Best Acc: 64.42%
[60]	T3	Hybrid: SVM/LR/RF + CNN	MSA: Hotel Reviews	ARBERT F1: 0.71 (Improvement > 17%)
[61]	T3, T4	Hybrid: CNN + LR/SVM/RF	Saudi Dialect: Telecom Tweets	CNN + Emb Acc: 81% CNN + POS Acc: 75% NER-City P: 89%
[30]	T3	Transformer: AraBERT + Token Augmentation	MSA + Dialectal: Hotel Reviews	AraBERT F1: 0.663 Aug F1: 0.661

Table A2. Cont.

Ref	Task Type	Method Category	Dataset Source and Domain	Result
[52]	T1–T3	Hybrid: POS + Rules + Weights	Mixed: Social Media Reviews	F1 (P/N/Neu): 93.9%/88.4%/100% Acc: 92.15%
[31]	T1	Lexicon: POS + AWN + Rules	MSA: Hotel and Product Reviews	F1 (Hotels/Products): 69.56%/68.29%
[62]	T1, T2	DL: BiGRU + MUSE Embeddings	MSA: Hotel Reviews	BiGRU + MUSE AE: 93%/92.8% PC: 90.9%/91.4%
[49]	T2, T3	Transformer: BERTopic + CAMELBERT/AraBART/mT5	MSA + Dialectal: Traffic Services Tweets	ACD F1: 92.8% APC F1: 99.2% E2E: 92% AraBART R-L: 0.79
[63]	T1–T3	ML: SVM (Top-1000 Unigrams, Multi-label)	MSA: Hotel Reviews	F1: 33.7 Acc: 74.8 (Average Over Subgroups)
[43]	T1	Hybrid: AraBERT + Flair + BiGRU-CRF	MSA: Hotel Reviews	BF-BiLSTM-CRF AE F1: 79.7%
[44]	T1, T2	ML: CRF + J48 (POS/NER/ng/BOW)	MSA: 2014 Israel-Gaza News	T1 F1: 0.38/0.76/0.82 T2 Acc: 0.61/0.86
[64]	T3, T4	Lexicon: Frequency/Polarity/Weighted Lexicons	MSA: Book Reviews	T3 F1: +15.2% T4 Acc: 71%
[65]	T1, T2	Hybrid: Ontology + LSTM/BiGRU	Tunisian Dialect: Multi-Domain Comments	F1: 94% (Phones) 65% vs. 48% (CNN + W2V)
[66]	T1, T2	Hybrid: AraBERT + CRF	MSA + Dialectal: Hotel Reviews	T1 F1: 97.8 T2 F1: 96.2 T3 F1: 95.2 T2 Acc: 98.3
[50]	T1–T4	Hybrid: Dual BERT Models	MSA + Arabizi: Game Reviews	T1: 91.6% T2: 91% T3: 79.1% T4: 88.2%
[41]	T2	Hybrid: AraBERT + LCF + Dep-GCN	MSA: Hotel Reviews	CDM Acc: 92.8% F1: 81.6%
[46]	T3, T4	ML: SVM (Unigrams, Multi-label)	MSA + Dialectal: Laptop Reviews	T1 F1: 0.315 T2 Acc: 0.732
[67]	T2	Transformer: BERT-base + Linear Softmax	MSA + Dialectal: Hotel and News Reviews	Acc: Hotels: 89.51% News: 85.73% HAAD: 73.23%
[68]	T2	DL: IA-LSTM L/T/R Enc + Interactive Attention	MSA + Dialectal: Hotel and News Reviews	Acc: Hotels: 83.10% News: 80.82%
[35]	T1, T2	Transformer: Encoder–Decoder BERT (Generation and Analysis)	MSA: Hotel Reviews	BERT2BERT F1: 77.3% (HAAD) 96.0% (Hotels)
[42]	T2, T3	DL: C-IndyLSTM ACD; MBRA + Bi-IndyLSTM + MH-Attn ASC	MSA: Hotel Reviews	C-IndyLSTM ACD F1: 58.05% MBRA ASC Acc: 87.31% (Beats RNN/CNN)
[55]	T1, T3	DL: AraBERT FT (BERT + BiGRU + CRF ATE; S/P-BERT + BERT + CNN ACD)	MSA: Facebook Posts	BERT + BiGRU + CRF F1: 88.1% (ATE) P-BERT F1: 84.1% (ACD)
[47]	T1, T4	DL: BGRU-CNN-CRF (T3); IAN-BGRU + Char CNN + Word Embeddings + BiGRU + CRF	MSA: Hotel Reviews	OTE F1: 70.7% ALSC Acc: 84% (Beats RNN/CNN)
[36]	T1, T4	Lexicon: Ontology-based	MSA: HAAD	Improvement: +19% (Ext) +55% (Pol) vs. Base
[37]	T3, T4	Transformer: MARBERT/AraBERT/ArabicBERT + GPT-4 FS (QA/NLI)	MSA + Saudi Dialect: Patient Reviews	Joint: 92.1% QARib 2-stage: 82.4%/78.4% GPT-4: 88.9%/89.3%
[69]	T1–T4	Hybrid: GPT-4o-mini + Human + FT	MSA + Saudi Dialect: Hospital Reviews	AraBERTv2 Acc: 96.3% (Aspect) 94.6% (Sentiment)
[70]	T1–T4	DL: CNN + Word2Vec	MSA + Saudi Dialect: Telecom Tweets	F1: 0.81
[71]	T1–T4	Transformer: Multi-task AraBERT + BiLSTM/GRU + Sent-pair BERT	MSA: Hotel Reviews	ATE F1: 80.3% ACD: 68.2% APC: 89.7% ACPC: 89.4% (Beats Single-task)
[72]	T2, T4	Transformer: Seq2Seq Dialect-to-MSA + BERT (QA/NLI)	MSA + Dialectal: Hotel and Book Reviews	T2 Acc: 74.9%/84.7% (HAAD/SemEval) T4: 76.9% (HAAD (SOTA Polarity))
[48]	T1, T2	Transformer: Arabic BERT + BIO + CRF	MSA: Book Reviews	T1/T2 F1: 47.6% Joint Acc: 95.2% (CRF Improved)
[73]	T2, T3	DL: BiLSTM-CRF + AB-LSTM-PC + Attention	MSA: Hotel Reviews	T1/T2 F1: 70% T3 Acc: 82.6%
[74]	T1, T2	ML: SVM/J48/BayesNet/NB/IBK + Rich Features	MSA: Hotel Reviews	T1 F1: 93.4 T2 F1: 89.8 T3 Acc: 95.4
[75]	T1, T2	Hybrid: SVM (Rich Features) + 5-layer RNN	MSA: Hotel Reviews	SVM T1/T2/T3: 93.4/89.8/95.4 RNN: 48%/48%/87% (4–5 × Faster)
[45]	T1–T3	Hybrid: Transformers + Bi-LSTM-CRF/LSTM/SVM	MSA + Dialectal: Food-Delivery Tweets	T1: 77% (Giga/AraBERT) T2: 82% (AraBERT) T3: 81% (GigaBERT) (Beats BiLSTM-CRF/LSTM/SVM)
[76]	T1, T3	Hybrid: ML (AdaBoost/J48/SMO/CRF) + Rule-based (Nooj/Lexical Rules)	MSA: Hotel Reviews	Pos F1: 97.49% Neg F1: 96% Acc: 97% (ML Extraction + Rule Polarity Correction)
[28]	T1, T2	Transformer: AraBERT/MARBERT/QARib + Token	MSA + Dialectal: Multi-Domain	PROD F1: 65.5% TELC: 69.5% (Synthetic Data: 56.6% to 69.5%)
[54]	T3	Transformer: FT AraBERT (Multi-label)	MSA: Hotel Reviews	AraBERT + DWLF/INS F1: 67.3%
[77]	T1–T3	Transformer: FT MARBERT (Sentiment and Aspect)	MSA + Dialectal: SPL Dataset	Sentiment Acc: 93.5% F1: 93% ATE F1: 91.8% 71.9% (Negative SPL Posts)

Table A2. Cont.

Ref	Task Type	Method Category	Dataset Source and Domain	Result
[32]	T1, T2	Transformer: AraBERT/MARBERT FT + SA ABSA	MSA: Hotel Reviews	AraBERT + Cosine-Attn + CRF Best: Joint F1: 72.92% ATE F1: 80.36%
[78]	T2	Transformer: BERT + BiGRU + Attention + CRF (BIO)	MSA + Saudi Dialect: Tourism Reviews	F1: 0.97/0.92 Classical: 0.75–0.91
[38]	T2, T3	Transformer: FT AraBERTv2/ArBERT/MARBERT/mBERT	MSA + Dialectal: Restaurant Reviews	AraBERT v2.0 F1: 95.75% (ACD) 92.83% (ASC)
[39]	T3	Transformer: Few-shot + GPT-4o FT	MSA: Gaza and Book Reviews	F1: 95% (Gaza) 32% (Book)
[79]	T3	Hybrid: AraBERT + ML (SVM/LR/DT, BR/CC)	MSA: Hotel Reviews	TF-IDF + N-grams (CC) F1: 62% (Augmented)
[80]	T2	DL: XLNet + BiLSTM + Topics (LDA/NMF)	MSA + Dialectal: COVID-19 Tweets	XLNet-NMF Acc: 0.942
[56]	T3	Transformer: BERT + TCN + BiGRU	MSA: Gaza 2014 News	BERT-TCN-BiGRU F1: 88.11%
[81]	T3	DL: BiGRU + Multi-embeddings	MSA: Hotel Reviews	BiGRU + BERT Vectors F1: 65.5%
[82]	T2	ML: SVM/RF/NB/KNN + TF-IDF and N-grams (POS)	Saudi Dialect: 2 Datasets	RF: 86.3% (Electronics) SVM: 86.5% (Clothes)
[83]	T3	Transformer: AraBERT (Semi-supervised)	MSA: Hotel Reviews	Teacher–Student Ensemble F1: 0.682
[84]	T1, T2	Transformer: AraBERTv02 + LCF-ATEPC + MHSA (MTL)	MSA: Hotel Reviews	AR-LCF-ATEPC-Fusion ATE F1: 78.56% APC Acc: 93.18%
[40]	T1	ML: CRF + Linguistic Features	MSA: Arabic News	CRF Template 14 F1: 95%
[85]	T1	ML: AdaBoost + Lexical/Semantic/Syntactic Features	MSA: SemEval-2016 Hotel Reviews	AdaBoost F1: 96.9%
[86]	T4	Transformer: BERT + BiGRU + CRF (Polarity-aware BIO)	MSA: Hotel Reviews	Joint F1: 72.92% ATE F1: 80.36%

References

- Alqusair, D.; Taileb, M.; Al Barhamtoshy, H.M. A survey of deep learning techniques for arabic aspect-based sentiment analysis. *IEEE Access* **2025**, *13*, 25350–25368. [\[CrossRef\]](#)
- Shoufan, A.; Alameri, S. Natural Language Processing for Dialectal Arabic: A Survey. In Proceedings of the Second Workshop on Arabic Natural Language Processing, Beijing, China, 26–31 July 2015; pp. 36–48. Available online: <https://aclanthology.org/W15-3205.pdf> (accessed on 1 June 2025).
- Moudjari, L.; Akli-Astouati, K.; Benamara, F. An algerian corpus and an annotation platform for opinion and emotion analysis. In Proceedings of the Twelfth Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; pp. 1202–1210. Available online: <https://aclanthology.org/2020.lrec-1.151.pdf> (accessed on 1 June 2025).
- Arak, M.H. Digital transformation and its impact on arabic language evolution in social media. *ICRRD Qual. Index Res. J.* **2024**, *5*, 174–186. [\[CrossRef\]](#)
- Alsiyat, I.; Piao, S.S. Metaphorical expressions in automatic arabic sentiment analysis. In Proceedings of the Twelfth Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; pp. 4911–4916. Available online: <https://aclanthology.org/2020.lrec-1.604.pdf> (accessed on 1 June 2025).
- Bensoltane, R.; Zaki, T. Aspect-based sentiment analysis: An overview in the use of arabic language. *Artif. Intell. Rev.* **2023**, *56*, 2325–2363. [\[CrossRef\]](#)
- Oueslati, O.; Cambria, E.; HajHmida, M.B.; Ounelli, H. A review of sentiment analysis research in arabic language. *Future Gener. Comput. Syst.* **2020**, *112*, 408–430. [\[CrossRef\]](#)
- Li, R.; Prozzi, J.A.; Hong, F. Quantification of post-rainfall moisture content in pavement unbound layers using long-term pavement performance data. *Transp. Res. Rec.* **2025**, *2680*, 03611981251380276. [\[CrossRef\]](#)
- Zhan, W.; Chen, Y.; Liu, Q.; Li, J.; Sacchi, M.D.; Zhuang, M.; Liu, Q.H. Simultaneous prediction of petrophysical properties and formation layered thickness from acoustic logging data using a modular cascading residual neural network (mcarnn) with physical constraints. *J. Appl. Geophys.* **2024**, *224*, 105362. [\[CrossRef\]](#)
- Wu, C.; Ma, B.; Liu, Y.; Zhang, Z.; Deng, N.; Li, Y.; Chen, B.; Zhang, Y.; Xue, Y.; Plank, B. M-absa: A multilingual dataset for aspect-based sentiment analysis. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, Suzhou, China, 4–9 November 2025; pp. 2530–2557. [\[CrossRef\]](#)
- Shi, Z.; Agrawal, R. A comprehensive survey of contemporary arabic sentiment analysis: Methods, challenges, and future directions. *Find. Assoc. Comput. Linguist. NAACL* **2025**, *2025*, 3760–3772. [\[CrossRef\]](#)
- Alqahtani, S.; Mishra, A.; Diab, M. A multitask learning approach for diacritic restoration. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 8238–8247. [\[CrossRef\]](#)

13. Barhoumi, A.; Camelin, N.; Aloulou, C.; Estève, Y.; Belguith, L.H. Toward qualitative evaluation of embeddings for arabic sentiment analysis. In Proceedings of the Twelfth Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; pp. 4955–4963.
14. Alyami, S.; Jamal, A.; Alhothali, A. Domain-adaptive pre-training for arabic aspect-based sentiment analysis: A comparative study of domain adaptation and fine-tuning strategies. *arXiv* **2025**, arXiv:2509.16788. [\[CrossRef\]](#)
15. Nazir, A.; Rao, Y.; Wu, L.; Sun, L. Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. *IEEE Trans. Affect. Comput.* **2020**, *13*, 845–863. [\[CrossRef\]](#)
16. Farha, I.A.; Magdy, W. Benchmarking transformer-based language models for arabic sentiment and sarcasm detection. In Proceedings of the Sixth Arabic Natural Language Processing Workshop, Kiev, Ukraine, 19 April 2021; pp. 21–31. Available online: <https://aclanthology.org/2021.wanlp-1.3/> (accessed on 1 July 2025).
17. Abdul-Mageed, M.; Zhang, C.; Hashemi, A.; Nagoudi, E.-M.-B. Aranet: A deep learning toolkit for arabic social media. In Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, Marseille, France, 11–16 May 2020; pp. 16–23. Available online: <https://aclanthology.org/2020.osact-1.3/> (accessed on 1 July 2025).
18. Ruder, S.; Ghaffari, P.; Breslin, J.G. A hierarchical model of reviews for aspect-based sentiment analysis. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–4 November 2016; pp. 999–1005. [\[CrossRef\]](#)
19. Ruder, S.; Ghaffari, P.; Breslin, J.G. Insight-1 at semeval-2016 task 5: Deep learning for multilingual aspect-based sentiment analysis. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), San Diego, CA, USA, 16–17 June 2016; pp. 330–336. [\[CrossRef\]](#)
20. Hsu, T.-W.; Chen, C.-C.; Huang, H.-H.; Chen, H.-H. Semantics-preserved data augmentation for aspect-based sentiment analysis. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Punta Cana, Dominican Republic, 7–11 November 2021; pp. 4417–4422. [\[CrossRef\]](#)
21. Qin, H.; Tian, Y.; Xia, F.; Song, Y. Complementary learning of aspect terms for aspect-based sentiment analysis. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, Marseille, France, 20–25 June 2022; pp. 7029–7039. Available online: <https://aclanthology.org/2022.lrec-1.760/> (accessed on 1 July 2025).
22. Qin, H.; Chen, G.; Tian, Y.; Song, Y. Improving federated learning for aspect-based sentiment analysis via topic memories. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Punta Cana, Dominican Republic, 7–11 November 2021; pp. 3942–3954. [\[CrossRef\]](#)
23. Alyafeai, Z.; Al-shaibani, M.S.; Ghaleb, M.; Ahmad, I. Evaluating various tokenizers for arabic text classification. *Neural Process. Lett.* **2023**, *55*, 2911–2933. [\[CrossRef\]](#)
24. Alyami, S.; Alhothali, A.; Jamal, A. Systematic literature review of arabic aspect-based sentiment analysis. *J. King Saud Univ. Comput. Inf. Sci.* **2022**, *34*, 6524–6551. [\[CrossRef\]](#)
25. Youssef, L.; Elhoussaine, Z. AraABSAMD: A Novel Arabic Dataset for Aspect-Based Sentiment Analysis in the Moroccan Education Domain. *Array* **2026**, *30*, 100782. [\[CrossRef\]](#)
26. Voskergian, D. Amar_absa: Arabic mobile app reviews dataset targeting aspect-based sentiment analysis tasks. In Proceedings of the 2022 Innovations in Intelligent Systems and Applications Conference (ASYU), Antalya, Turkey, 7–9 September 2022. [\[CrossRef\]](#)
27. Elwy, I.; Sabty, C. Takween: A comprehensive arabic collection and annotation tool. *Procedia Comput. Sci.* **2024**, *244*, 186–193. [\[CrossRef\]](#)
28. Eid, Y.; Zayed, H.; Medhat, W. A-masa: Arabic multi-domain aspect-based sentiment analysis datasets. *Procedia Comput. Sci.* **2024**, *244*, 202–211. [\[CrossRef\]](#)
29. Zaghouani, W.; Mohit, B.; Habash, N.; Obeid, O.; Tomeh, N.; Rozovskaya, A.; Farra, N.; Alkuhlani, S.; Oflazer, K. Large scale arabic error annotation: Guidelines and framework. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland, 26–31 May 2014; pp. 2362–2369.
30. Almasre, M.A. Enhance the aspect category detection in arabic language using arabert and text augmentation. In Proceedings of the 2022 Fifth National Conference of Saudi Computers Colleges (NCCC), Jeddah, Saudi Arabia, 17–18 December 2022; pp. 1–6. [\[CrossRef\]](#)
31. Mataoui, M.; Bendali Hacine, T.E.; Tellache, I.; Bakhtouchi, A.; Zelmatti, O. A new syntax-based aspect detection approach for sentiment analysis in arabic reviews. In Proceedings of the 2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP), Algiers, Algeria, 25–26 April 2018; pp. 1–6. [\[CrossRef\]](#)
32. Zaid, S.; Alharbi, A.H.; Samra, H. Multi-aspect sentiment classification of arabic tourism reviews using bert and classical machine learning. *Data* **2025**, *10*, 168. [\[CrossRef\]](#)

33. Al-Sarhan, H.; Al-So'ud, M.; Al-Smadi, M.; Al-Ayyoub, M.; Jararweh, Y. Framework for affective news analysis of arabic news: 2014 gaza attacks case study. In *Proceedings of the 7th International Conference on Information and Communication Systems (ICICS)*, Irbid, Jordan, 5–7 April 2016; pp. 200–205. [\[CrossRef\]](#)
34. Ashi, M.M.; Siddiqui, M.A.; Nadeem, F. Pre-trained Word Embeddings for Arabic Aspect-Based Sentiment Analysis of Airline Tweets. In *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2018*; Hassanien, A., Tolba, M., Shaalan, K., Azar, A., Eds.; *Advances in Intelligent Systems and Computing*; Springer: Berlin/Heidelberg, Germany, 2019; Volume 845. [\[CrossRef\]](#)
35. Chouikhi, H.; Jarray, F.; Alsuhailani, M. A sequence-to-sequence neural network for joint aspect term extraction and aspect term sentiment classification tasks. In *Proceedings of the 18th International Conference on Web Information Systems and Technologies (WEBIST)*, Valletta, Malta, 25–27 October 2022; pp. 115–122. [\[CrossRef\]](#)
36. Behdenna, S.; Barigou, F.; Belalem, G. An ontology-based approach to enhance explicit aspect extraction in standard arabic reviews. *Int. J. Comput. Digit. Syst.* **2022**, *11*, 245–256. [\[CrossRef\]](#) [\[PubMed\]](#)
37. AlNasser, S.; AlMuhaideb, S. Listening to patients: Advanced arabic aspect-based sentiment analysis using transformer models towards better healthcare. *Big Data Cogn. Comput.* **2024**, *8*, 156. [\[CrossRef\]](#)
38. Youssef, L.; Elhoussaine, Z.; Soufiane, N.; Noureddine, M. Enhancing arabic aspect category detection using large language models (llms). *Results Eng.* **2025**, *26*, 105049. [\[CrossRef\]](#)
39. Ameer, A.; Hamdi, S. Arabic aspect category detection for hotel reviews based on data augmentation and classifier chains. In *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing (SAC '23)*, Tallinn, Estonia, 27–31 March 2023; pp. 838–841. [\[CrossRef\]](#)
40. Trigui, S.; Boujelben, I.; Jamoussi, S.; Ben Ayed, Y. Adal system: Aspect detection for arabic language. *Adv. Intell. Syst. Comput.* **2021**, *1179*, 31–40. [\[CrossRef\]](#)
41. Bensoltane, R.; Zaki, T. Knowledge-enhanced graph convolutional networks for arabic aspect sentiment classification. *Soc. Netw. Anal. Min.* **2024**, *14*, 6. [\[CrossRef\]](#)
42. Al-Dabet, S.; Tedmori, S.; Al-Smadi, M. Enhancing arabic aspect-based sentiment analysis using deep learning models. *Comput. Speech Lang.* **2021**, *69*, 101224. [\[CrossRef\]](#)
43. Fadel, A.S.; Saleh, M.E.; Abulnaja, O.A. Arabic aspect extraction based on stacked contextualized embedding with deep learning. *IEEE Access* **2022**, *10*, 30247–30259. [\[CrossRef\]](#)
44. Al-Smadi, M.; Al-Ayyoub, M.; Al-Sarhan, H.; Jararweh, Y. Using aspect-based sentiment analysis to evaluate arabic news affect on readers. In *Proceedings of the 2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing (UCC)*, Limassol, Cyprus, 7–10 December 2015; pp. 436–441. [\[CrossRef\]](#)
45. Al-Jarrah, I.; Mustafa, A.M.; Najadat, H. Aspect-based sentiment analysis for arabic food delivery reviews. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2023**, *22*, 220. [\[CrossRef\]](#)
46. Al-Ayyoub, M.; Gigieh, A.; Al-Qwaqenah, A.; Al-Kabi, M.N.; Talafhah, B.; Alsmadi, I. Aspect-based sentiment analysis of arabic laptop reviews. In *Proceedings of the 2017 International Arab Conference on Information Technology (ACIT)*, Yasmine Hammamet, Tunisia, 22–24 December 2017.
47. Abdelgwad, M.M.; Soliman, T.H.A.; Taloba, A.I.; Farghaly, M.F. Arabic aspect based sentiment analysis using bidirectional gru based models. *J. King Saud Univ. Comput. Inf. Sci.* **2022**, *34*, 6652–6662. [\[CrossRef\]](#)
48. Chouikhi, H.; Alsuhailani, M.; Jarray, F. Bert-based joint model for aspect term extraction and aspect polarity detection in arabic text. *Electronics* **2023**, *12*, 515. [\[CrossRef\]](#)
49. Alotaibi, A.; Nadeem, F. An unsupervised integrated framework for arabic aspect-based sentiment analysis and abstractive text summarization of traffic services using transformer models. *Smart Cities* **2025**, *8*, 62. [\[CrossRef\]](#)
50. Hammad, M.; AbuEnnab, N.; Al-Refai, M. An aspect-based sentiment analysis model for arabic game reviews based on hybrid transformers models. *Neural Comput. Appl.* **2025**, *37*, 10309–10331. [\[CrossRef\]](#)
51. Al Katat, S.; Zaki, C.; Hazimeh, H.; El Bitar, I.; Angarita, R.; Trojman, L. Natural language processing for arabic sentiment analysis: A systematic literature review. *IEEE Trans. Big Data* **2024**, *10*, 576–599. [\[CrossRef\]](#)
52. Abd-Elhamid, L.; Elzanfaly, D.; Eldin, A.S. Feature-based sentiment analysis in online arabic reviews. In *Proceedings of the 2016 11th International Conference on Computer Engineering & Systems (ICCES)*, Cairo, Egypt, 20–21 December 2016; pp. 240–245. [\[CrossRef\]](#)
53. Chebolu, S.U.S.; Derronnecourt, F.; Lipka, N.; Solorio, T. A review of datasets for aspect-based sentiment analysis. In *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, Bali, Indonesia, 1–4 November 2023; pp. 570–585. [\[CrossRef\]](#)
54. Ameer, A.; Hamdi, S.; Yahia, S.B. Multi-label learning for aspect category detection of arabic hotel reviews using arabert. In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023)*, Lisbon, Portugal, 22–24 February 2023; Volume 2, pp. 241–250. [\[CrossRef\]](#)

55. Bensoltane, R.; Zaki, T. Towards arabic aspect-based sentiment analysis: A transfer learning-based approach. *Soc. Netw. Anal. Min.* **2022**, *12*, 77. [\[CrossRef\]](#)
56. Bensoltane, R.; Zaki, T. Comparing word embedding models for arabic aspect category detection using a deep learning-based approach. *E3S Web Conf.* **2021**, *297*, 01072. [\[CrossRef\]](#)
57. Alassaf, M.; Qamar, A.M. Aspect-based sentiment analysis of arabic tweets in the education sector using a hybrid feature selection method. In Proceedings of the 2020 14th International Conference on Innovations in Information Technology (IIT), Al Ain, United Arab Emirates, 17–18 November 2020; pp. 202–207. [\[CrossRef\]](#)
58. Alshaikh, K.A.; Almatrafi, O.A.; Abushark, Y.B. Bert-based model for aspect-based sentiment analysis for analysing arabic open-ended survey responses: A case study. *IEEE Access* **2024**, *12*, 2288–2302. [\[CrossRef\]](#)
59. Alghamdi, A.; Taileb, M.; Almani, N. Stacked lstm-gru with double attention model for arabic aspect-based sentiment analysis. In Proceedings of the 2025 2nd International Conference on Advanced Innovations in Smart Cities (ICAISC), Jeddah, Saudi Arabia, 9–11 February 2025. [\[CrossRef\]](#)
60. Youssef, L.; ELhoussaine, Z. Arabic aspect category detection using traditional neural networks and arbert. In Proceedings of the 2024 11th International Conference on Wireless Networks and Mobile Communications (WINCOM), Casablanca, Morocco, 23–25 July 2024. [\[CrossRef\]](#)
61. AlShammari, N.; AlMansour, A. Aspect-based sentiment analysis and location detection for arabic language tweets. *Appl. Comput. Syst.* **2022**, *27*, 119–127. [\[CrossRef\]](#)
62. Al-Smadi, M.; Hammad, M.M.; Al-Zboon, S.A.; Al-Tawalbeh, S.; Cambria, E. Gated recurrent unit with multilingual universal sentence encoder for arabic aspect-based sentiment analysis. *Knowl.-Based Syst.* **2023**, *261*, 107540. [\[CrossRef\]](#)
63. Al-Smadi, M.; Qwasmeh, O.; Talafha, B.; Al-Ayyoub, M.; Jararweh, Y.; Benkhelifa, E. An enhanced framework for aspect-based sentiment analysis of hotels' reviews: Arabic reviews case study. In Proceedings of the 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST), Barcelona, Spain, 5–7 December 2016; pp. 106–111. [\[CrossRef\]](#)
64. Obaidat, I.; Mohawesh, R.; Al-Ayyoub, M.; Al-Smadi, M.; Jararweh, Y. Enhancing the determination of aspect categories and their polarities in arabic reviews using lexicon-based approaches. In Proceedings of the 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), Amman, Jordan, 3–5 November 2015; pp. 1–6. [\[CrossRef\]](#)
65. Belguith, M.; Aloulou, C.; Gargouri, B. Aspect level sentiment analysis based on deep learning and ontologies. *SN Comput. Sci.* **2024**, *5*, 58. [\[CrossRef\]](#)
66. Shafiq, G.M.; Hamza, T.; Alrahmawy, M.F.; El-Deeb, R. Enhancing arabic aspect-based sentiment analysis using end-to-end model. *IEEE Access* **2023**, *11*, 142060–142071. [\[CrossRef\]](#)
67. Abdelgwad, M.M.; Soliman, T.H.A.; Taloba, A.I. Arabic aspect sentiment polarity classification using bert. *J. Big Data* **2022**, *9*, 115. [\[CrossRef\]](#)
68. Balla, H.A.M.N.; Salvador, M.L.; Delany, S.J. Interactive learning approach for arabic target-based sentiment analysis. In Proceedings of the 13th International Conference on Recent Advances in Natural Language Processing (RANLP 2021), Online, 1–3 September 2021; pp. 116–125.
69. Alamoudi, E.; Solaiman, E. Ehsan: Leveraging chatgpt in a hybrid framework for arabic aspect-based sentiment analysis in healthcare. *arXiv* **2025**. [\[CrossRef\]](#)
70. Alshammari, N.F.; AlMansour, A.A. Aspect-based sentiment analysis for arabic content in social media. In Proceedings of the 2020 2nd International Conference on Electrical, Communication and Computer Engineering (ICECCE), Istanbul, Turkey, 14–15 April 2020; pp. 1–6. [\[CrossRef\]](#)
71. Fadel, A.; Saleh, M.; Salama, R.; Abulnaja, O. Mtl-arabert: An enhanced multi-task learning model for arabic aspect-based sentiment analysis. *Computers* **2024**, *13*, 98. [\[CrossRef\]](#)
72. Chennafi, M.E.; Bedlaoui, H.; Dahou, A.; Al-qaness, M.A.A. Arabic aspect-based sentiment classification using seq2seq dialect normalization and transformers. *Knowledge* **2022**, *2*, 388–401. [\[CrossRef\]](#)
73. Al-Smadi, M.; Talafha, B.; Al-Ayyoub, M.; Jararweh, Y. Using long short-term memory deep neural networks for aspect-based sentiment analysis of arabic reviews. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 2163–2175. [\[CrossRef\]](#)
74. Al-Smadi, M.; Al-Ayyoub, M.; Jararweh, Y.; Qawasmeh, O. Enhancing aspect-based sentiment analysis of arabic hotels' reviews using morphological, syntactic and semantic features. *Inf. Process. Manag.* **2019**, *56*, 308–319. [\[CrossRef\]](#)
75. Al-Smadi, M.; Qawasmeh, O.; Al-Ayyoub, M.; Jararweh, Y.; Gupta, B. Deep recurrent neural network vs. Support vector machine for aspect-based sentiment analysis of arabic hotels' reviews. *J. Comput. Sci.* **2018**, *27*, 386–393. [\[CrossRef\]](#)
76. Trugui, S.; Boujelben, I.; Jamoussi, S.; Ayed, Y.B. A hybrid method for arabic aspect-based sentiment analysis. *Int. J. Hybrid Intell. Syst.* **2020**, *16*, 99–110. [\[CrossRef\]](#)
77. Alraimi, D.A.Q.; Paputungan, I.V. Decoding fan and societal sentiment: Absa of the saudi pro league's recent evolution. *J. Sains Nalar Appl. Teknol. Inf.* **2025**, *4*, 44–53. [\[CrossRef\]](#)
78. AlMasaud, A.; Al-Baity, H.H. On the robustness of arabic aspect-based sentiment analysis: A comprehensive exploration of transformer-based models. *J. King Saud Univ. Comput. Inf. Sci.* **2024**, *36*, 102264. [\[CrossRef\]](#)

79. Sweidan, A.H.; El-Bendary, N.; Elhariri, E. Autoregressive feature extraction with topic modeling for aspect-based sentiment analysis of arabic as a low-resource language. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2024**, *23*, 23. [[CrossRef](#)]
80. Bensoltane, R.; Zaki, T. Combining bert with tcn-bigru for enhancing arabic aspect category detection. *J. Intell. Fuzzy Syst.* **2023**, *44*, 4123–4136. [[CrossRef](#)]
81. Alrefae, R.; Alqahmi, R.; Alduraibi, M.; Almatrafi, S.; Alayed, A. Enhancing customer experience through arabic aspect-based sentiment analysis of saudi reviews. *Int. J. Adv. Comput. Sci. Appl.* **2024**, *15*, 387–394. [[CrossRef](#)]
82. Almasri, M.; Al-Malki, N.; Alotaibi, R. A semi supervised approach to arabic aspect category detection using bert and teacher-student model. *PeerJ Comput. Sci.* **2023**, *9*, e1425. [[CrossRef](#)]
83. Fadel, A.S.; Abulnaja, O.A.; Saleh, M.E. Multi-task learning model with data augmentation for arabic aspect-based sentiment analysis. *Comput. Mater. Contin.* **2023**, *75*, 4419–4444. [[CrossRef](#)]
84. Touati, I.; Ellouze, M. Opinion target extraction from arabic news articles using shallow features. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing (SAC '18), Pau, France, 9–13 April 2018; pp. 838–841. [[CrossRef](#)]
85. Al-Smadi, M.; Qawasmeh, O.; Al-Ayyoub, M.; Jararweh, Y.; Gupta, B. Human annotated arabic dataset of book reviews for aspect based sentiment analysis. In Proceedings of the 2015 3rd International Conference on Future Internet of Things and Cloud (FiCloud), Rome, Italy, 24–26 August 2015; pp. 726–730. [[CrossRef](#)]
86. Bensoltane, R.; Zaki, T. Neural multi-task learning for end-to-end arabic aspect-based sentiment analysis. *Comput. Speech Lang.* **2025**, *89*, 101683. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.