University of Zawia Journal of Engineering Sciences and Technology (UZJEST)
Volume 2, Issue 2, (2024), pp 227-245, ISSN 3006-2764

A Systematic Review of Sentiment Analysis Techniques, Challenges and Future Directions for Arabic Dialects

Hani Glgam[1,2]

[1] Department of Information Technology, College of Science and Technology, Jado, Libya

[2] Department of Information Technology, Libyan Academy for Postgraduate Studies, Tripoli, Libya

Corresponding author email: haniglgam83@gmail.com

Received: 14-11-2024 | Accepted: 13-12-2024 | Available online: 15-12-2024 | DOI:10.26629/uzjest.2024.21

ABSTRACT

At present time, the internet and social media are crucial for content sharing and expressing opinions on topics. These opinions can be analysed to assess product quality, identify issues, or improve products, but manual analysis of thousands of comments can be time-consuming. Therefore, Sentiment analysis is a method that uses natural language processing and computational linguistics to identify and extract subjective information from texts, determining the emotional state expressed, whether positive, negative, or neutral. A systematic literature review (SLR) on sentiment analysis of Arabic dialects (SAAD) is presented in this study. The main causes of diversity among these dialects are variations in syntax, lexicon, and grammar, which makes it challenging for scholars to classify DA polarity. This study has determined every stage that significantly affects the machine learning model used for dialect sentiment analysis, including text pre-processing, text annotation, feature extraction, and the approaches used. Additionally, the study has identified the issues and unresolved problems with sentiment analysis of the Arabic dialect (SAAD), which should be the main focus of future studies.

Keywords: Sentiment Analysis, Arabic Dialects, Machine Learning, Dataset Pre-processing, challenges

How to cite this article:
Glgam, H. A Systematic Review of Sentiment Analysis Techniques, Challenges and Future Directions for Arabic Dialects. Univ Zawia J Eng Sci Technol. 2024;2:227-245.

مراجعة منهجية لتقنيات تحليل المشاعر والتحديات والتوجهات المستقبلية للهجات العربية

هاني-أحمد جلغام

قسم تكنولوجيا المعلومات، كلية العلوم والتقنية، جادو، ليبيا

ملخـــــــص البحــــــث

في الوقت الحاضر، يعد الإنترنت ووسائل التواصل الاجتماعي أمرًا بالغ الأهمية لمشاركة المحتوى والتعبير عن الآراء حول الموضوعات. يمكن تحليل هذه الآراء لتقييم جودة المنتج أو تحديد المشكلات أو تحسين المنتجات، لكن التحليل اليدوي لآلاف التعليقات قد يستغرق وقتًا طويلاً. ولذلك، فإن تحليل المشاعر هو أسلوب يستخدم معالجة اللغات الطبيعية

واللغويات الحاسوبية لتحديد واستخراج المعلومات الذاتية من النصوص، وتحديد الحالة العاطفية المعبر عنها، سواء كانت إيجابية أو سلبية أو محايدة. في هذه الدراسة تم تقديم مراجعة منهجية للأدبيات حول المنهجيات المتبعة لتحليل المشاعر في اللهجات العربية. علاوة على ذلك، الأسباب الرئيسية للتنوع بين هذه اللهجات هي الاختلافات في تركيب الجملة، والمعجم، والقواعد، مما يجعل من الصعب على البحاث تصنيف قطبية اللهجات العربية. حددت هذه الدراسة كل المراحل التي تؤثر بشكل كبير على نموذج التعلم الآلي المستخدم لتحليل مشاعر اللهجات، بما في ذلك المعالجة المسبقة للنص، والتعليق التوضيحي للنص أي" تصنيف النص"، واستخراج الميزات، والأساليب المستخدمة في تحليل المشاعر باستخدام تعلم الالة. بالإضافة إلى ذلك، حددت الدراسة القضايا والمشكلات التي لم يتم حلها في تحليل المشاعر للهجات العربية في الدراسات السابقة، والتي ينبغي أن تكون المحور الرئيسي للدراسات المستقبلية.

الكلمات المفتاحية: تحليل الآراء، اللهجات العربية، تعلم الالة، معالجة مجموعة البيانات، التحديات

1. Introduction

Special attention has been paid to sentiment analysis in the domains of production, marketing, and advertising. In fact, the rise of social media sites like Facebook, Instagram, LinkedIn, and Twitter has enabled people to voice their thoughts and emotions about a wide range of topics, goods, concepts, or services, and so on. Thus, one of the NLP field's most active study fields is automatic sentiment analysis [1]. Sentiment analysis (SA), sometimes referred to as opinion mining, is a crucial area of natural language processing (NLP) that establishes whether a text's representation of emotions is positive, negative, or neutral sentiment in multiple languages. [2]. Every language has various dialects, which are frequently distinguished by the speakers' geographic locales. One of the primary languages of the world is Arabic. Languages have different dialects based on geographical locations, with Arabic having classical Arabic (CA), the language of Muslims' Holy Quran, formal (Modern Standard Arabic - MSA) and informal (Dialectal Arabic - DA) variants. Arabic dialects include Gulf (GLF), Iraqi (IRQ), Levantine (LEV), Egyptian (EGY), and North African (NOR) dialects [3].

There are still some differences between DA and MSA at the phonological, morphological, lexical, and syntactic levels, even though DA and modern Arabic have a similar vocabulary. Different dialects may deduce distinct syntactic meanings from the same word. For instance, the meaning of "طرّش" varies among the Gulf countries. In many Gulf countries, it means 'send a message or a text', in Lebanon, it might mean "paint something in white", in Morocco, it can mean "slap someone in the face", and in Yemen, it can mean "to vomit". another example stated in MSA, is the phrase "How are you?" takes on distinct forms in each of the dialects (Egyptian: أزيك , Levantine: كيفك, Gulf: شلونك, North African: شنو حالك , Iraqi: اشونك, Yemeni: كيوفك)[3]. Furthermore, the authors in [4] covered the problem of negation, which was introduced by means of the word (ما , mA) and the suffix (ش) with the sukun on it (« sh ») as in(ماكليتش) which means " I did not eat". Thus, pre-processing procedures are essential for any Arabic dialect's application since, as stated by the same study, DA has multiple scripts (Roman and Arabic) and is being utilized on social media for written correspondence.

Three main methods have been developed for sentiment analysis, which are categorized as lexicon-based, machine learning, and hybrid. While the lexicon-based approach uses systems that do sentiment analysis based on a set of human-crafted dictionaries, the machine learning method uses machine learning techniques to categorize sentiment based on a large annotated training data set. In the final strategy, the systems integrate machine learning and lexicon-based techniques into a single technique[5]. The process of applying sentiment analysis involves several steps, beginning with text annotation, text pre-processing, text extraction, and classification models. Priorities were set on examining the SA research conducted within DA. The writers of [6] set out to look into the procedures and approaches used for Sentiment Analysis for Arabic Dialects SAAD classification. Further, the authors of [7] described the difficulties that sentiment analysis systems face; the most frequent ones are sarcasm, spam, polarity fuzziness, poor review quality, and so on.

In contrast to the latter, this literature examines the most recent research on the whole pipeline used in the SAAD workflow by offering an analysis of each evaluated study's performance as well as an insight into all of its experiments. With regard to the state of various pre-processing techniques, lexicon-based, machine learning, and deep learning approaches as well as transformer-based models in current research, this literature review aims to give an overview of important research methodologies and suggestions.

This paper's reminder is arranged as follows. The methodology of systematic review is provided in Section 2. Related works are shown in Section 3. Datasets region, source, and size are then summarized in detail in section 4. After that, sections 5 and 6 present various pre-processing and feature extraction techniques. Then, section 7 presents findings, discussion and, challenges. Section 8, finally provides a conclusion and some directions for future research.

2. Methodology

Systematic literature review (SLR) is a systematic method used to identify, collect, select, and analyse primary studies on a specific subject, adhering to PRISMA guidelines [8]. It involves formulating research questions, conducting literature searches, screening data, and analysing the findings.

2.1. Research questions

The performance of SAAD is influenced by numerous aspects. These elements include datasets and sentiment analysis methods, pre-processing methods, feature extraction, and more. In light of these elements, the study questions were formulated as follows:

- What approaches to sentiment analysis are used in SAAD?
- What are the datasets used to perform sentiment analysis for Arabic dialects?
- What kinds of preprocessing methods does SAAD employ?
- What feature extraction methods does SAAD use?
- Which SAAD methods yield the best results in terms of accuracy performance?
- What are SAAD's next research challenges and directions?

2.2. Search strategy

The literature search process followed the PRISMA criteria, which involves three stages: identification, screening, and inclusion, as illustrated in Figure 1. The reviewer searched for journal articles and conference proceedings from 2018 to 2024 using just Google Scholar because of limited access to digital databases such as Springer Link, IEEE/IET Electronic Library, Elsevier Science Direct, etc. The following keywords are employed in the search process: challenges, techniques, Arabic dialects, and sentiment analysis. To perform the queries, research expressions join the keywords using the logical operator AND. Only titles, abstracts, and keywords are included in the search index.

2.3. Inclusion and exclusion criteria

The screening phase involves using criteria for inclusion and exclusion to determine if primary studies meet specific requirements. A paper is included if it meets all inclusion criteria, while it is excluded if it meets at least one excluding requirement as depicted in Table 1. Relevant publications are searched

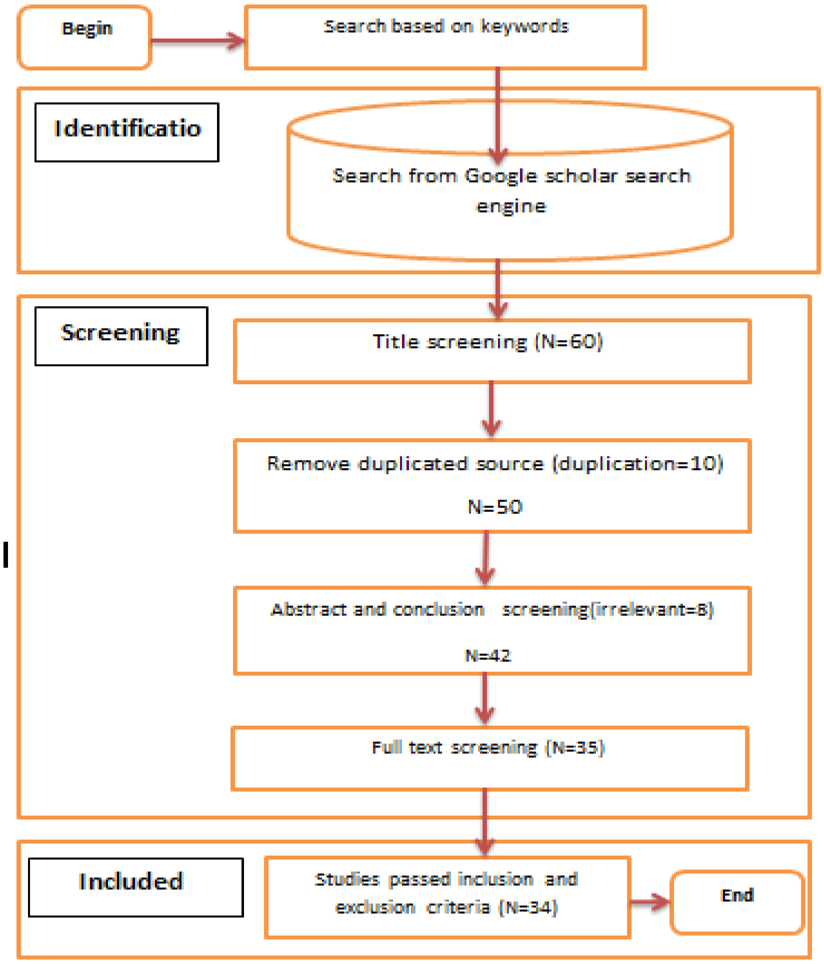

Figure 1. The flow diagram of the systematic literature review [8]

by scanning the study's title and abstract, and if unclear, the entire content is read. At the inclusion stage, 34 papers were included for more in-depth analysis.

Table 1. Inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criteria
• The article is published in journals Or conferences. • The paper is written in English by Arabic native speakers. • The research focuses on study questions of the systematic literature review SLR. • The paper was published between 2018 and 2024.	• Duplicated papers are publications older than 7 years. • The article does not focus on sentiment analysis of Arabic dialects SAAD. • The study deals with Arabic dialect identification but it is not about sentiment analysis classification. • Study focuses on SA of other languages such as English, Urdu, etc…

2.4. Analysis of Data

The inclusion phase involves reviewing summary data from primary studies, including author, publisher, and publication years as well as data focusing on SLR research topics, including approaches, pre-processing techniques, public datasets, and current problems.

3. Related work of Sentiment Analysis Methods

Various methodologies and tools have been developed for the classification and analysis of Modern Standard Arabic (MSA). Recently, there has been a growing focus on sentiment analysis of different Arabic dialects. Researchers have explored several challenges associated with sentiment analysis of Arabic dialects (SAAD), including morphological complexity, language dependency, stop-word removal, stemming, negation detection, and annotation, among others. The literature identifies five approaches to this field: lexicon-based, machine learning, deep learning, hybrid and transformer-based approaches as illustrated in Figure 2.

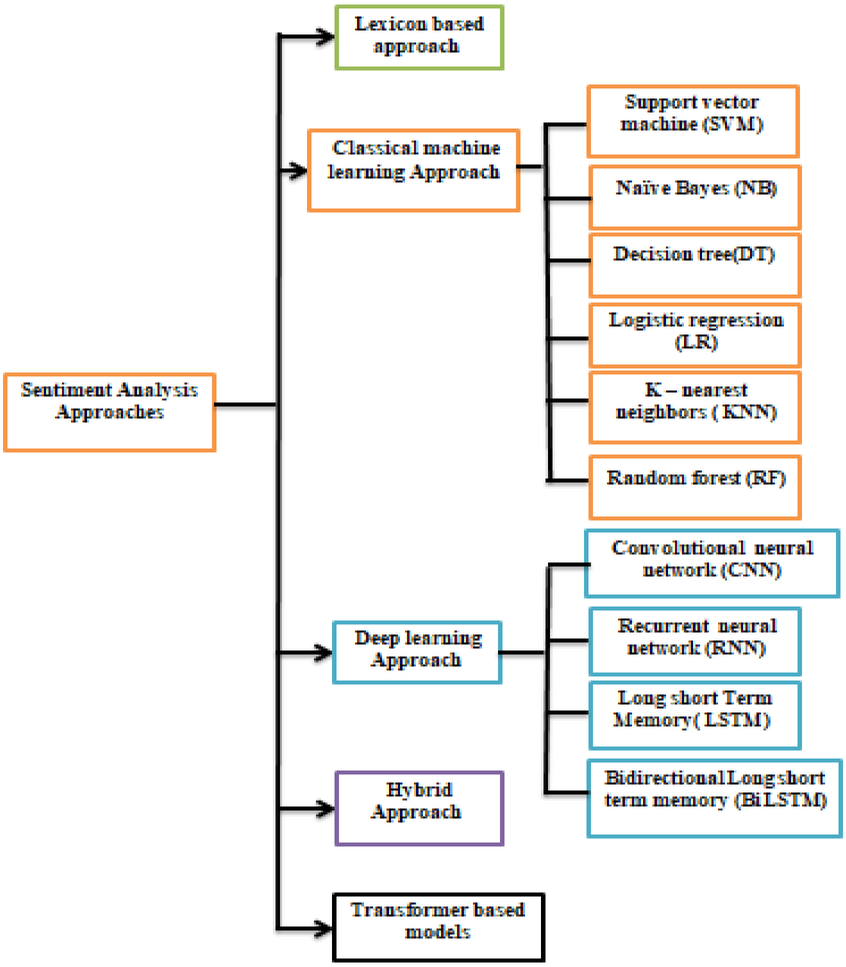

Figure 2. Sentiment analysis approaches [9]

3.1 Lexicon based Approach

While many research studies have focused on developing lexicons for English language texts suitable for the NLP domain, only a small number of research papers have examined the creation of lexicons for Arabic language texts, either in Modern Standard Arabic or Arabic dialects [6].

In [5], the authors developed a lexicon-based sentiment analysis system for the Libyan dialect, using adjective-adverb combinations to classify tweets into seven categories. They used a publicly accessible Twitter corpus and rigorously tested the system on 5000 annotated tweets, achieving an F-score of 82.19%, demonstrating its effectiveness in sentiment classification.

Another study by Almosawi and Mahmood [10], used a lexicon-based approach for sentiment analysis of student feedback on lecturer performance, achieving 98% accuracy. The approach outperformed machine learning algorithms like Naive Bayes, Support Vector Machines, and K-Nearest Neighbors. The researchers used an electronic questionnaire to gather a dataset of 2,217 terms in modern Arabic and the southern Iraqi dialect for a dictionary of viewpoints in higher education.

In [11], The study introduced the Lexicon-based Sentiment Analysis on Arabic Texts (LSAnArTe) framework, specifically designed to analyse customer perceptions of coffee products in Arabic on social media. This model utilizes a self-constructed lexicon with approximately 10,769 sentences, enhancing the accuracy of sentiment classification. The LSAnArTe framework outperformed the Amazon Comprehend tool, achieving an accuracy score of 93.79% compared to Amazon's 51.90%.This emphasizes the effectiveness of LSAnArTe in sentiment analysis for Arabic texts .

It can be seen that Lexical-based methods can be an effective approach because they take into account the meaning of words used in the dialect, as well as the use of negation and other linguistic elements that can influence the sentiment of a comment. However, lexical-based approaches are still limited in their coverage of dialect-specific vocabulary and are unable to account for the highly inflected nature of dialect Arabic. In other words, each Arabic dialect has unique vocabulary and expressions, which can lead to gaps in the lexicon. A lexicon developed for one dialect may not include terms used in another, limiting its effectiveness and also the problem of domain-specific meaning, different dialects may have specific terms related to particular domains (e.g., cultural references, local slang) that are not captured in a general lexicon. This can lead to inaccuracies in sentiment classification when analysing texts from diverse contexts which leads to poor sentiment analysis results. In summary, while lexicon-based approaches offer valuable insights and feature capture in sentiment analysis of Arabic dialects, they also face challenges related to coverage, flexibility, and the effort required for effective implementation.

3.2 Machine learning based Approach

When compared to English, Arabic language analysis is much more complicated. Arabic social media users often communicate in dialect, an unstructured, grammatical slang language that makes it difficult to understand them. According to recent research, several studies have looked at the application of traditional machine learning algorithms for Arabic sentiment analysis. This is a summary of related work in this field, organized to fit in the related work part:

The study in [2], explores the impact of sampling techniques and classification algorithms on an imbalanced Arabic dataset, revealing that fear and stress were the primary factors influencing people's perceptions of COVID-19. The research used over-sampling and under-sampling techniques on machine learning and ensemble algorithms, finding resampling-based approaches effective in addressing the classification problem with an F1-score value of 0.99.

In [12], the study proposed a sentiment analysis system for analysing customer opinions of Libya's major telecommunication companies, Libyana, Almadar Aljadid, and Libya Telecom and Technology, using Twitter data. Five machine learning models were applied, but most were over fit due to class imbalance. Pre-processing and cleaning steps were performed to improve performance. The SVM model was most effective at predicting Libyan telecom firm customer opinions with an accuracy of 80.67%, followed by the NB model for Almadar Aljadid with an accuracy of 81.19%. The decision tree model performed best in Libya Telecom and Technology , with a 75% success rate..

Another study introduced by Inas M. Milad and Abdulrazag M. Atomi proposed a SVM model to classify extracted features into positive, neutral, or negative polarities using N-gram and Arabic

WordNet tools. The model was evaluated using the SYR dataset, consisting of 2000 Arabic tweets about Syrian wars. Experimental tests showed a total accuracy of 0.78 [13].

According to the paper written by Ramadan A. Alfared and Hanan M. Aljarm three machine learning algorithms using Libyan tweets for sentiment analysis. The researchers found 89.1% agreement in sentiment annotations, with the decision tree algorithm outperforming the other two. The Decision Tree, Support Vector Machine, and Naive Bayes models achieved 72%, 69%, and 65% accuracy, respectively. [14].

In [15], researchers implemented several machine learning algorithms to carry out sentiment classification on both balanced and unbalanced dataset. The created Emirati dataset was subjected to preprocessing and TF-IDF features extraction techniques in order to enhance its classification performance and get it ready for the sentiment analysis experiment. The best result was achieved 80.80% , and this was obtained when the ensemble model implemented on unbalanced dataset for sentiment classification.

The paper in [16], proposes a weakly supervised categorization system called "q8SentiLabeler" to replace human annotators in sentiment analysis in Kuwait. A dataset of over 16.6k posts was created, eliminating bias. The performance of the dataset was tested using deep-learning language models and conventional machine-learning classifiers. ML models used to test the dataset were LR, SVM, M-NB, and Bagging which have achieved an accuracy rate of 0.91, 1.00, 0.78, and 0.99 respectively.

In [17], the researchers created a comprehensive Arabic sentiment lexicon and used it to accurately label Facebook comments from Jordanian users regarding telecom services, achieving 98% accuracy. The labeled dataset was then used to train supervised machine learning models. The results of the classification were 97.8, 96.8, and 95.6% for Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), and Naïve Bayes (NB) classifiers, respectively.

Hussein et al. [18] used Six distinct machine learning algorithms in the suggested model which are decision tree, random forest, stochastic gradient descent, logistic regression, Naïve Bayes, and support vector machines. They tested the model on two datasets: RES Multi-Dialect restaurant feedback and ASTD Egyptian-Dialect tweets. The SVM classifier achieved the highest accuracy of 87.7%, while the Bernoulli Naive Bayes classifier achieved the best results of 82%.

Habberrih and Abuzaraida published a paper investigated the impact of Stemming and Stop-words removal techniques on machine learning classifiers, Support Vector Machine (SVM) and Logistic Regression (LR), in detecting sentiment from Libyan dialect poetry. The results show that Stop-words removal may negatively affect classifier performance, but SVM outperformed LR in both experiments, achieving an accuracy of 71.63% , while LR achieved 70.92% in the second experiment [19].

It can be concluded that, Machine learning algorithms, particularly Support Vector Machines (SVM) and Naive Bayes (NB), have shown effective performance in sentiment analysis of Arabic dialects, achieving high accuracy rates in various studies and also machine learning techniques do not need a large volume of data compared to deep learning methods . Furthermore, Machine learning techniques are adaptable and improve over time as more data becomes available, which is crucial given the diversity of Arabic dialects, making sentiment analysis faster and more efficient, especially in Arabic dialects, due to their ability to automate large volumes of data from social media. On the other hand, The morphological richness of Arabic dialects poses challenges for ML models, as they may struggle to capture the nuances and variations in word forms. Moreover, The performance of ML models heavily relies on the quality of sentiment resources. Poor-quality resources can lead to inaccurate sentiment classification.

3.3 Deep Learning Based Approach

Deep learning (DL) is a machine learning branch that learns embedded and abstract representations from raw data with minimal human intervention. It has shown excellent performance in sentiment analysis of English text. However, only a few recent works have explored deep learning models for SAAD.

Study conducted by [1] proposed a deep learning approach for sentiment analysis of Tunisian dialect comments on supermarket Facebook pages. It evaluates CNN, LSTM, and Bi-LSTM models for sentence-level and aspect-level sentiment analysis. The LSTM and Bi-LSTM models outperformed the others, with an F-Measure of 87% for sentence-level sentiment analysis.

In the same context, a study by [20] used a deep learning model to analyze sentiment in Arabic tweets during the COVID-19 outbreak in Saudi Arabia. Data was collected from Jeddah, Dammam, and Riyadh, and sentiments were classified using convolutional neural networks (CNN) and bi-directional long short memory (BiLSTM) techniques. The results showed a 92.80% accuracy rate for CNN and 91.99% for BiLSTM.

In [21], the authors assessed their created ASA models against other models that employ deep learning and machine learning methods. The enhanced three models, CNN-Model, LSTM-Mode2, and CNN+LSTM-Model3, have accuracy of 96.83%, 94.74%, and 96.91%, respectively. The models performed much better than previous efforts.

The paper in [22], presented three variants of Deep Learning (DL) based on multiplicative Long Short Term Memory (mLSTM), Long Short Term Memory (LSTM), and Gated Recurrent Unit (GRU) using AraVec. The experimental study was validated using three Arabic language corpora (TEAD, ATSAD, and ASTD) and two learning modes (Hold out and 10-folds cross validation). The mLSTM deep learning model achieved the best results in the TEAD, ATSAD, and ASTD dataset, with an accuracy of 99.75% ,97.52and 96.54 respectively.

Alsokkar et al. [23] used Support Vector Machine (SVM) and BiLSTM algorithms on Kaggle to classify notes as positive, negative, or neutral. They used Natural Language Processing (NLP) and pre-processing tools to analyse three thousand notes from telecommunication companies in Jordanian dialect. SVM achieved a 66% accuracy rate, while BiLSTM achieved a 99.21% accuracy rate. Previous studies have shown SVM produces better accuracy with larger dataset sizes.

In [3], the study presented the development of SGRU, SBi-GRU, and an ensemble approach using multiple models (SGRU, SBi-GRU, and AraBERT) to generate the best-suited model for the Arabic language. They also introduced automatic sentiment refinement (ASR) for discarding stop words. AraVec was found to be the most effective Arabic embedding compared to Fasttext and ArabicNews. The ensemble method, which combined the strengths of different classifiers, achieved the best performance, with a 90% accuracy surpassing that of other singular models (SGRU, SBi-GRU, and AraBERT).

It is clearly seen that deep learning models, such as CNN and Bi LSTM ensemble, mLSTM model in conjunction with AraVec, have shown impressive results in sentiment analysis, achieving accuracies up to 99.75% in some studies. This indicates their effectiveness in capturing complex patterns in data. Further, deep learning reduces manual feature extraction in Arabic dialects, allowing for the automatic learning of intricate and varied linguistic features. This technique also handles the complexities of Arabic dialects, including variations in grammar and vocabulary, which traditional machine-learning methods struggle with. However, deep learning models require substantial computational resources and large datasets for training, which can be a limitation given the scarcity of annotated datasets for Arabic dialects. Furthermore, deep learning models are prone to overfitting, especially when trained on small datasets. This can lead to poor generalization of unseen data.

3.4 Hybrid-based Approach

The goal of hybrid approaches is to integrate the most advantageous aspects of both ML and DL techniques. Despite the hybrid approach appearing to have high performance in SAAD, few studies have been published on this method.

Abubaera and Jiddah developed a hybrid CNN-GRU model to analyze Arabic sentiment using a dataset of 13,782 customer reviews from a Libyana telecommunication company. The model uses data augmentation to increase dataset size and balance, enabling more effective deep learning training. Two experiments showed that data augmentation significantly improved model performance, with the first achieving 82.19% accuracy and the second at 91.54%, demonstrating its effectiveness in addressing under-sampled positive data samples [24].

Mahboub et al. [25] developed a machine learning model for identifying Arabic Libyan dialects, achieving accuracy rates of 72%, 76%, and 77%. They used SVM, LR, and NB with the TF-IDF feature representation technique, building a lexicon of Arabic Libyan dialects from a relational database. The model transformed pre-processed Libyan dialect text into Modern Standard Arabic (MSA) form, enhancing overall performance. The model is a hybrid approach, combining linguistic resources (lexicon) and statistical learning methods (classifiers) to improve the identification process.

In [26], researchers proposed hybrid models to improve single deep learning models, achieving higher accuracy by layering multiple ensembles. They successfully predicted Arabic sentiment using CNN, LSTM, GRU, BiGRU, BiLSTM, CNN-BiGRU, CNN-GRU, CNN-LSTM, and CNN-biLSTM. The model successfully interpreted feelings conveyed in Arabic, surpassing standard forms of deep learning by a margin of 0.9112. The procedure starts with extracting Arabert model features and training nine deep-learning models.

The paper in [27] proposed a hybrid machine and deep learning model using LSTM, BiLSTM, and logistic regression, incorporating an attention mechanism. Tests on a dataset of 34,905 Twitter-generated Arabic language comments on Twitter, specifically Egyptian, Gulf, Jordanian, and Yemeni. showed the model outperformed long-short-term memory, bidirectional long-short-term memory, and logistic regression models in classifying dialects. The model used different word representations, with an accuracy rating of 83.31%, 81.31%, and 81.22%, respectively.

In summary, hybrid CNN-GRU model shown promise in enhancing sentiment analysis for Arabic dialects. The model leverages the strengths of Convolutional Neural Networks (CNNs) in capturing local patterns and features from the text data, while Gated Recurrent Units (GRUs) are effective in capturing sequential dependencies. This combination can lead to improved performance in understanding context and sentiment in text. On the other hand, it presents challenges related to complexity, over fitting, and interpretability that need to be addressed for practical applications. Furthermore, with the use of several deep learning models, there is a potential risk of over fitting, especially if the datasets are not sufficiently large or diverse. Over fitting can result in a model that performs well on training data but poorly on unseen data. Additionally, the model combines long short-term memory (LSTM) and bidirectional LSTM (BiLSTM) architectures, which are effective in capturing long-range dependencies and contextual information. The bidirectional nature of BiLSTMs enables the model to consider both past and future contexts when making predictions, which is crucial for accurately classifying dialects where the context of neighboring words plays a significant role.

3.5 Transformer Based Models

Most of the current methods for sentiment analysis in the Arabic dialects depend on traditional machine learning and deep learning methods. Comparatively to other languages like English, there is less research on the use of BERT models.

Bourahouat et al. [28] have used pre-trained Arabic BERT models that are pre-trained in the Arabic language, namely AraBERT, QARIB, ALBERT, AraELECTRA, and CAMeLBERT for sentiment analysis in Moroccan Arabic, integrating them with deep learning and machine learning algorithms like SVM and CNN. The approach achieves impressive accuracy, reaching 96% when using the QARIB model even on imbalanced data. The research highlights the superior results of BERT-based models compared to CNN or SVM, contributing to the advancement of sentiment analysis in Moroccan Arabic dialects.

In the study introduced in [29], attempted to identify sentiments from Arabic commentary. Researchers employed a total of seven algorithms for this: two from deep learning (ARABERT and neuron network) and five from machine learning (logistic regression, decision tree, Bayesian, SVM, and XGBoost). The TF_IDF used to extract the features for the first six algorithms, and BERT for ARABERT served as our foundations. Overall, the three models LR, NN, and Arabert archived the best outcomes, with respective rates of 82%, 80%, and 88%.

In [30], AraXLNet pre-trained language model was build which is the first Arabic language model based on XLNet, and showcasing its application in Arabic sentiment analysis to enhance the prediction accuracy of such tasks. The model was tested by AraSenTi, ASTD, SemEval, and AJGT datasets. The suggested model AraXLNet with Farasa segmenter, attained F1-Score results of 95%, 93%, and 86%, and 88% respectively. In contrast, AraBERT performed worse on the same datasets, obtaining 84.65%, 92.13%, 85.05%, and 91.94% respectively.

It can be said that results in the reviewed studies showed the use of Arabert, both as classifier and embedding model, consistently achieved better results than traditional ML and DL algorithms such as SVM and CNN. This superior outcome can be attributed to many factors, including the extensive language knowledge obtained by Transformers over training on varied language modelling objectives. However, Arabert may not perform optimally on tasks or domains that significantly differ from the data it was trained on, requiring additional training on another Arabic dialects such as Libyan dialect and more fine-tuning or adaptations.

4. Datasets Region, Source, and Size

The study examines the differences in Arabic dialects due to slang from neighbouring countries. Researchers used datasets from various geographical areas, primarily containing MSA and Arabic dialects. Table 2 shows the limited number of research studies suggesting datasets containing Arabic dialects, including North African, Egyptian, Levantine, Iraqi, and Gulf countries [1,2,3,5,13,15,16,19,20,21,22,24,26,28,29,30,31,32,33]. Language experts and professors have gathered datasets from social media sites such as Twitter, Facebook, and several platforms and then categorized them into various polarities. Because they are usually unstructured, these data need to be pre-processed before continuing.

Despite significant progress in Arabic natural language processing, there is a lack of complete and high-quality annotated datasets, especially for dialectal Arabic, which hinders the development and training of reliable machine learning models [32]. Furthermore, compared to English datasets, Arabic datasets have a small size, especially regarding Libyan dialect datasets. Deep learning approaches and models based on transformers need large datasets to be trained. Thus, it will not achieve promising results.

Therefore, large annotated datasets must be created to make valuable progress in the sentiment Analysis field.

Table 2. datasets designed for Arabic dialects sentiment analysis

Ref.	Dialect	Size	Source	Domain
[1]	Tunisian dialect	44k	Facebook	Field of supermarkets
[2]	MSA, DA	15,779k	Twitter	COVID-19
[5]	Libyan dialects	5000K	Twitter	freely available Libyan dialect
[28]	5 national dialects, Moroccan dialect	MSDA= 50,000K MAC= 18,000k	Twitter	______
[31]	Libyan dialect	2938k	Twitter	News, sport, technology, love, religion
[24]	Libyan dialects	13,782k	Twitter	Telecom companies
[13]	DA, MSA	2000k	Twitter	wars in Syria
[15]	MSA, Emiratis dialect, DA	ESAAD= 70,000k	Instagram platform	__________
[20]	Saudi Arabia dialect	157.214k	Twitter	Covid-19
[21]	Saudi Arabia dialect	27.294k	Twitter	Communication companies
[29]	MSA, Moroccan Dialect	10,254k	Facebook	Morocco's parliamentary elections
[22]	Multi Arabic dialects, Egyptians dialect	TEAD= 5,615,943 ATSAD= 58,751 ASTD= 10,006	Twitter	COVID-19
[26]	MSA,DA	HARD= 490587 BRAD= 510,600	Booking GoodRead.com websites	Hotel reviews, books reviews
[32]	MSA, DA	67,000k	online platforms	ordering food
[33]	Saudi Arabia	4604k	Trip Advisor website	121 Saudi hotels
[16]	Kuwait dialect	16,676k	instagram	Justice, humanity, religious, feminism
[30]	Large MSA,DA corpus	OpenSubtitles= 60 Million sentence HARD= 93,700K LABR= 63,000 k BRAD= 510,600k	Different platforms.	Pertaining AraXLNet on Hotel Arabic Reviews, book reviews.
[3]	Gulf dialect	56,674k	Twitter	news, sports, and politics
[19]	Misurata sub-dialect	22,762K	---------------	poems

5. Dataset Pre-Processing

Pre-processing Arabic dialects on social media is challenging due to numerous dialect regions, spelling errors, extra letters, diacritical marks, and elongations, which often involve various processes.

5.1 Tokenization

This is the first preprocessing stage, in which the text is divided into words (tokens) and punctuated or white spaced between them. The outcome of this operation is a group of words. Tokenization was applied in this experiment with the NLTK library [15].

5.2 Cleaning

Arabic dialect text requires multiple cleaning techniques to improve Natural Language Processing (NLP) comprehension. Common methods include removing non-Arabic characters, punctuation, repeated letters, lengthening, diacritics, URLs, usernames, emoticons, and hash tags. However, removing non-Arabic characters may result in data loss due to code switched and Arabize words. According to [22], the distracting and unnecessary material is eliminated from the tweets. For instance, on Arabic characters, numbers, and single Arabic characters and unique symbols like ($% & / _ - …) Punctuation like (. : "" ; ') and usernames the tweet's image and URL are deleted.

5.3 Stop words removal

This natural language processing (NLP) technique identifies words in a text that are of little to no meaning and eliminates them as possible noise in the data. The NLTK Arabic language stop words package was used in this study for preprocessing [24]. It is preferable to eliminate all excluded words in the Arabic language, such as: ·أنتم، متى، (أما، من), هذه), as well as several Libyan dialect excluded words (توا، ليش، عليش) [12].

5.4 Normalization

Normalization involves transforming a word or a letter into its standard form. This involved replacing Arabic letters (أ إ آ) with (ا) , replacing (ة) with (ه), replacing (ئ) with (ي), and replacing (ؤ) with (و). Elongated words were also returned to their original form for instance the word " حلووووو" was transformed to its original form " حلو "[15].

5.5 Steaming

It is the process of taking words down to their most basic, uninflected forms. Light and root stemming can help with this [19]. It is mentioned in [1] Light rooting seeks to preserve the word's base while eliminating all prefixes and suffixes. For instance, by deleting the suffix مغازتنا mgAztnA / [our supermarket] becomes مغازة [supermarket].

5.6 Negation detection

Since negation detection has the power to reverse the polarity of a feeling or attitude, it may be seen as a significant hurdle while treating DA. According to [4], the authors covered the problem of negation, which was introduced by means of the word (ما , mA) and the suffix (ش) with the sukun on it (« sh ») as in(ماكليتش) which means " I did not eat". The negation issue was addressed in a few studies, including [17,10,19] which included negation words as features in lexicon-based approaches. In contrast, the ML-based studies used the bi-gram as a feature extractor because it depends on the existence of a prefix and a suffix in the same word as in the Libyan and Moroccan dialect. By using the prefix " ما " and the suffix "ش', we can express the negation of the word "كليت" and the resulting word is "ماكليتش", which means "I did not eat." Furthermore, this issue has only been handled utilizing a lexicon-based strategy that makes use of a set of rules specific to the dialect under consideration. As a result, negation and its range in the sentence in DA are still problems for machine learning systems to handle.

On the other hand, It can be said that negation detection can be improved by using QARiB as word embedding with one of deep learning models such as mLSTM and BiLSTM enriched by hand-crafted rules and lexicon-based features, which leads to an accuracy improvement of the model. As mentioned in [28], QARiB is a BERT model specifically designed for dialectal Arabic, trained on an extensive dataset of 420 million tweets and 180 million sentences of text, enabling it to effectively understand and handle the nuances of dialects in its language representations. Fine-tuning it effectively requires a reasonable amount of labelled data to learn the specific task and negation detection in Arabic dialects. If the dataset is small, the model might over fit, so data augmentation techniques will be needed to increase the size of the dataset.

5.7 Imbalanced Data

Class imbalance in datasets is a critical issue that affects the performance of these methods. The problem of data set imbalance occurs when different classes have significantly different sizes in the data set; the majority classes constitute a large part of the data, while the minority classes have a small amount of data. When collecting data for SA, we may experience data imbalance; if the nature of this data is ignored, ML algorithms will be biased towards the majority class, resulting in misclassification of the minority class relative to the majority class. To correct class imbalance, many solutions have been developed, which can be at the data or algorithm level [2].

The study in [28], stated that the available data presents the problem of data imbalance, which has a significant impact on the developed model. The problem of data imbalance can be solved either by collecting more data, removing redundancy, strengthening the training dataset, or performing data augmentation. The lack of resources and data is compounded when it comes to Arabic dialects, such as Moroccan dialect. For the latter, few useful open access resources are available.

Al-Khazaleh et al. in [2] discussed data-level sampling techniques to address imbalanced datasets, including over-sampling and under-sampling methods. Over-sampling techniques, such as SMOTE and Borderline-SMOTE, generate synthetic samples for the minority class to improve its representation in the dataset. Under-sampling techniques remove samples from the majority class to achieve balance, but care must be taken to avoid losing valuable information. The results show that resampling-based approaches can significantly improve classifier performance on imbalanced datasets, with over-sampling techniques yielding better results than under-sampling methods. The RF ensemble classifier using the oversampled SMOTE, BSMOTE or ADASYN data shows good efficiency with an F1 value of 0.99.

In [34], OpenAI's ChatGPT was used as a zero-short learning model to create human-like annotation tools for sentiment tasks. The feature vector was then used to detect and identify sentiment in sentences using BERT, BiLSTM, and a SoftMax function. Oversampling methods were used to to address the imbalanced data dilemma and visualize text contributions for customer reviewers. The method performed better than the latest methods on tested datasets, with an average score of 98.9%, F1-measure of 97.7%, and an AUC of 91.90% when combined with pre-trained models.

Overall, it can be said that sampling techniques help to adjust the class distribution in a dataset, making it more balanced. This is crucial because imbalanced datasets can lead to biased machine learning models that favor the majority class. Further, methods like SMOTE (Synthetic Minority Over-sampling Technique) create synthetic examples of the minority class. This increases the representation of the minority class, allowing the model to learn better from these examples and improve its predictive performance

5.8 Annotation

Sentiment annotation is the process of labeling an opinion or emotion within a text with positive, negative, or neutral polarity. This is crucial for sentiment analysis, especially in datasets extracted from social networks. It can be done manually, semi-automatically, or automatically. Manual annotation is the most precise method, but it's time-consuming and costly. The majority of reviewed studies have employed manual annotation techniques. When manual annotation is done by non-experts, it can occasionally be subjective because it depends on the sentences as understood by humans. This task remains challenging to complete in the absence of explicit dialect rules, which is why automated and semi-automatic annotation systems were applied in just two studies [16,17].

In the study introduced in [34], ChatGPT was employed as a text annotation tool to evaluate sentiment analysis (SA) on customer reviews from financial institutions. It captures human sentiment effectively compared to traditional lexical-based models (LBM) that rely on predefined sentiment lexicons. Furthermore, the annotated data from ChatGPT was integrated into a model that includes BERT and BiLSTM layers. This combination helps in extracting feature vectors and determining the sentiment orientation of the dataset. The effectiveness of ChatGPT in annotating sentiment was evaluated against other sentiment analysis techniques, demonstrating its superior performance in generating an accurate sentiment annotation.

6. Feature Extraction

The study aims to analyse feature extraction techniques in SAAD publications, focusing on their application over time and frequency in specific studies, while also addressing the unsolved task of encoding contextual and semantic information in DA.

According to [27] The goal of the feature extraction techniques is to convert the textual input into numerical representations. The model is tested using each of the three main word representation algorithms, which are TF-IDF, Word2Vec, and GloVe. In the same context, [19] N-grams are text-based features that break up text into n-groups, with three main varieties: unigrams, bigrams, and trigrams. Further, Bag of Words (BOW) is the easiest to extract and provides good data coverage. However, bigrams and trigrams can identify patterns of negation or sentiment expression, allowing for a more in-depth examination of the text's sentiment.

Reviewed studies show various techniques, including TF-IDF, N-grams, BOW, and CountVectorizer, are used to extract text insights. TF-IDF is the most commonly used technique, followed by lexicon-based features and N-grams. However, deep learning techniques heavily rely on word embedding, and new word embedding methods like AraBERT, Word2Vec, GloVe, and FastText have been discovered and demonstrated their significant performance. Further, Word2vec is the most commonly used feature extractor, followed by FastText, as per most reviewed studies.

7. Findings and Discussion

The main objective of this section is to present and discuss the findings of the conducted SLR and also to summarize and analyze the key processes in the SAAD pipeline, as well as to focus on the limitations in the studies reviewed. Table 3 illustrates the best results of reviewed studies that have been chosen for each approach to be discussed.

Based on the reviewed studies, the lexicon-based approach was applied in three research studies, with the most promising result being achieved on an Iraqi dataset of 802 feedback on lecturer performance, achieving an accuracy score of 98% [10]. Conversely, Libyan dataset of

5000 tweets was analysed using a lexicon-based method, resulting in an f1-score of 82.19% on various labels [5]. This disparity may be caused by imbalanced distribution of classes in the dataset. It can be concluded that Lexicon-based approach is superior due to their ability to capture additional elements like negation, but they are laborious and manual in data gathering due to the lack of a common system for multi-dialect datasets.

Table 3. The best results of each approach

Ref.	Approach	Classifier	Feature extraction	Dialect	Max performance
[10]	Lexicon based	Lexicon	Lexicon features	Iraqi dialect	98% Accuracy
[16]	Machine learning	LR,**SVM**, M-NB, Bagging	TF-IDF	Kuwait dialect	0.91,1.0, 0.78,0.99 Accuracy
[22]	Deep learning	mLSTM	AraVec	Multi dialect, Egyptian dialect	99.75%, 97.52%, 96.54% accuracy
[21]		Ensemble CNN+ LSTM	------------	Saudi Arabia	96.91% Accuracy
[24]	Hybrid	CNN-GRU + data augmentation	--------------	Libyan dialect	91.54% Accuracy
[28]	Transformer based models	QARIB pretrained model	QARIB	5 national dialects, Moroccan dialect	95% accuracy
[30]		AraXlNet	--------------	Large MSA, DA corpus	95%,86%,88%, 93% Accuracy

This approach is also known as domain reliance, as the same term can evoke different feelings in different contexts. "For instance, 'كبير جدا' (very big) is positive if it is about accommodation. However, it is negative if it is about address"[9, p411].

Nine ML papers have been reviewed, with most having lower accuracy rates (65%-80%). This may be caused by multiple classes and unbalanced datasets. On the other hand, a study on Kuwaiti dialect sentiment analysis using LR, SVM, M-NB, and Bagging achieved perfect performance rates 0.91, 1.00, 0.78, and 0.99 respectively [16]. These perfect performance scores imply a possibility of over-fitting. In summary, the evaluated papers use various machine learning algorithms, with SVM and NB being the most widely used. SVM is used in sentiment analysis of Arabic dialects in SAAD works, as it can discriminate across hyper planes, maximize class borders, and outperform most existing algorithms.

Regarding DL approach, six papers have been reviewed. Table 4 shows the optimal outcomes are achieved by the study [22] used the mLSTM model in conjunction with AraVec to deliver 99.75%, 97.52%, and 96.54% accuracy, respectively, on Twitter datasets TEAD, ATSAD, and ASTD with three labels (positive, negative, and neutral). Furthermore, 96.91% accuracy rate was attained by the CNN and Bi LSTM ensemble when 27.294k tweets from Saudi Arabia were gathered via Twitter [21]. This is due to RNN's ability to capture the long-term context of words inside sentences or documents. Further, CNN architectures may also be helpful for handling shorter messages, like tweets. It can be concluded that deep learning has shown promising performance in sentiment analysis for Arabic dialects, but it is still infrequently used compared to machine learning approaches. This is because DLs methods require significant processing power and data to produce accurate generalization models.

In the case of using Hybrid approach four studies were reviewed, the best result were obtained by combining CNN and GRU to evaluate the dataset which contains 13,782 customer reviews on the Libyana telecommunication company. Two experiments were conducted, the first experiment, using pre-processed data without augmentation, achieved an accuracy of 82.19%. The second experiment, incorporating data augmentation, significantly improved accuracy to 91.54%, demonstrating the effectiveness of data augmentation in enhancing sample size and addressing under-sampled positive data samples [24]. In summary, the hybrid CNN-GRU model has shown potential in improving sentiment analysis for Arabic dialects by combining the strengths of Convolutional Neural Networks (CNNs) in capturing local patterns and features, and Gated Recurrent Units (GRUs) in capturing sequential dependencies.

Lastly, regarding transformer based models three studies were reviewed. The best outcomes obtained in two studies [28,30]. The first study utilized QARIB pre-trained model for sentiment analysis in Moroccan Arabic dialect and achieved an accuracy of 95%. The second study used AraXlNet model which was tested by AraSenTi, ASTD, SemEval, and AJGT datasets. The suggested model, with Farasa segmenter, attained F1-Score results of 95%, 93%, 86%, and 88%. In contrast, AraBERT performed worse on the same datasets, obtaining 84.65%, 92.13%,85.05%, and 91.94% respectively. It can be said that Studies show that AraBERT and AraXlNet classifiers, consistently outperforms traditional ML and DL algorithms like SVM and CNN. This is due to Transformers' extensive language knowledge and training on diverse modeling objectives. However, it may not perform optimally on tasks or domains that differ significantly from the trained data, requiring additional training and adaptations.

Overall, the study examines SAAD accuracy from four perspectives: dataset characteristics, pre-processing techniques, feature extraction, and classification approaches. It reveals that DL models improve with more data provided. Cleaning noisy and inconsistent data enhances sentiment classification. Normalization techniques can deal with inconsistent data and improve stemming performance. Stop words, which convey little information or cause unnecessary noise, must be deleted. Since stop words have been removed manually in most research, the list of stop words that are utilized varies heavily depending on domain and area. However, based on [19], suggested that eliminating stop words in Libyan poetry may not always be beneficial or require further research, as it may lead to the loss of text meaning. The findings suggest that machine learning classifiers may suffer when this strategy is applied.

Feature extraction is the most important pre-processing step since it enables models to extract contextual and semantic information Thus, reviewed studies revealed that TF-IDF is the most commonly used technique, followed by lexicon-based features and N-grams. However, deep learning approach heavily rely on word embedding, and techniques like AraBERT, Word2Vec, GloVe, and FastText are widely used for feature extraction. Deep Learning (DL) is the most commonly used ML approach for extracting characteristics, especially with large datasets. Using DL in SAAD can tackle problems that ML approaches like SVM and NB cannot handle, such as negation, which requires a lexicon-based approach. Therefore, a hybrid approach may provide the best results.

In the area of sentiment analysis of Arabic dialects, there are still many gaps that remain unsolved. These ramifications can be observed from different perspectives:

- **Morphological Complexity:** Arabic dialects exhibit significant morphological variations, making it difficult to analyze sentiment accurately. The complexity arises from the rich inflectional forms of words, which can alter meaning significantly.
- **Resource Scarcity:** there is a notable lack of resources, such as annotated datasets and sentiment lexicons specifically tailored for Arabic dialects. This scarcity hampers the development of effective sentiment analysis models.
- **Domain Dependency:** sentiment expressions may differ across various domains, making it challenging to create models that generalize well across different contexts.
- **Negation:** the presence of negation in dialects can significantly alter sentiment meaning, complicating the analysis process.
- **Automatic Annotation:** this means labeling data automatically, which can be tricky. For example, if a sentence expresses happiness, it needs to be marked correctly so the computer understands it.
- **Stop Words Automatic Detection:** stop words are common words like (ليش , علاش ,توا , من، أما، متى، أنتم, هذه,), that don't add much meaning. Removing them helps focus on the important words in a sentence.

8. Conclusions and Future Directions

This SLR's primary goal is to examine the most recent Arabic sentiment analysis research using DA, 35 relative research articles published between 2018 and 2024 were examined after numerous works were filtered through a quality assessment pipeline. The analysis focused on preprocessing techniques like normalization, stemming, stop word removal, cleaning, negation detection, and annotation in order to explore approaches and methods used for SAAD. To offer some good practices, the study were able to look at feature extraction techniques in terms of ML and DL algorithms. Additionally, the goal was to evaluate which data representation was most appropriate for SAAD in terms of size, area, and source. Additionally, the results were evaluated using a variety of machine learning techniques, including probabilistic, non-parametric, and parametric algorithms. Based on these evaluations, the study has demonstrated that most studies showed a substantial correlation between performance and sentiment resource quality.

SLR focused on recent literature contributions related to SAAD; these findings have the following implications for the direction of future researches:

- Examine the dialectical content that is readily available on social media in order to create complete Arabic resources that take into account dialects and Arabize.
- To address domain and dataset dependencies, automatically remove stop words from a huge corpus of Arabic dialects.
- Build a large word embedding using a large corpus of DA.
- Explore self-supervised techniques for automatic annotation.
- Enlarge DA dictionaries to improve lexicon-based approach.
- Build hybrid model using the combination of deep learning techniques and lexicon-based features such as negation features gave the best results.
- Future studies could explore cross-linguistic models using transfer learning to adapt data from one language to languages with fewer resources, like Arabic dialects, enhancing sentiment analysis precision and efficacy.

- Studying deep learning structures like transformers and capsule networks could enhance Arabic sentiment analysis by capturing complex patterns and improving existing methods.

REFERENCES

[1] Masmoudi, A., Hamdi, J., & Belguith, L. H.. Deep learning for sentiment analysis of Tunisian dialect. *Computación y Sistemas*, 2021, 25(1), 129-148.

[2] Al-Khazaleh, M. J., Alian, M., & Jaradat, M. A. Sentiment analysis of imbalanced Arabic data using sampling techniques and classification algorithms. *Bulletin of Electrical Engineering and Informatics*, 2024, 13(1), 607-618.

[3] Al Wazrah, A., & Alhumoud, S. Sentiment analysis using stacked gated recurrent unit for Arabic tweets. *IEEE Access*, 2021, 9, 137176-137187.

[4] Nassr, Z., Sael, N., & Benabbou, F. Preprocessing Arabic dialect for sentiment mining: State of art. The International Archives of the Photogrammetry. *Remote Sensing and Spatial Information Sciences*,2020, 44, 323-330.

[5] Alhammi, H. A., & Haddar, K. Building a Libyan dialect lexicon-based sentiment analysis system using semantic orientation of adjective-adverb combinations. *International Journal of Computer Theory and Engineering*, 2020, 12(6), 145-150.

[6] Al Shamsi, A. A., & Abdallah, S. Text mining techniques for sentiment analysis of Arabic dialects: Literature review. *Adv. Sci. Technol. Eng. Syst. J*, 2021, 6(1), 1012-1023.

[7] Oueslati, O., Cambria, E., HajHmida, M. B., & Ounelli, H. A review of sentiment analysis research in Arabic language. *Future Generation Computer Systems*, 2020, 112, 408-430.

[8] Sarkis-Onofre, R., Catalá-López, F., Aromataris, E., & Lockwood, C. How to properly use the PRISMA Statement. *Systematic Reviews*, 2021, 10, 1-3.

[9] Oueslati, O., Cambria, E., HajHmida, M. B., & Ounelli, H. A review of sentiment analysis research in Arabic language. *Future Generation Computer Systems*, 2020, 112, 408-430.

[10] Almosawi, M. M., & Mahmood, S. A. Lexicon-based approach for sentiment analysis to student feedback. *Webology*. (ISSN: 2022, 1735-188X), 19(1).

[11] Alsemaree, O., Alam, A. S., Gill, S. S., & Uhlig, S. An analysis of customer perception using lexicon-based sentiment analysis of Arabic Texts framework. *Heliyon*, 2024, 10(11).

[12] Omar, A., Essgaer, M., & Ahmed, K. M. Using machine learning model to predict Libyan telecom company customer satisfaction. *International Conference on Engineering & MIS (ICEMIS)*. 2022, July, (pp. 1-6). IEEE.

[13] Milad, I. M., & Atomi, A. M. Sentiment analysis for Arabic sentences using nlp and classification techniques. *Al Academia Journal for Basic and Applied Sciences* (AJBAS), 2022, 4, 1–17.

[14] Alfared, R. A., & Aljarm, H. M. Sentiment analysis of libyan tweets using machine learning algorithms. *Malaysian Journal of Industrial Technology* (MJIT), 2024, 8, 1–12.

[15] A. Al Shamsi, A., & Abdallah, S. Sentiment analysis of Emirati dialect. *Big Data and Cognitive Computing*, 2022, 6(2), 57.

[16] Husain, F., Alostad, H., & Omar, H. Bridging the Kuwaiti Dialect Gap in Natural Language Processing. *IEEE*, 2024, Access.

[17] Nahar, K. M., Jaradat, A., Atoum, M. S., & Ibrahim, F. Sentiment analysis and classification of arab jordanian Facebook comments for Jordanian telecom companies using lexicon-based approach and machine learning. *Jordanian Journal of Computers and Information Technology*, 2020, 6(3).

[18] Hussein, A. H., Moawad, I. F., & Badry, R. M. Arabic sentiment analysis for multi-dialect text using machine learning techniques. *International Journal of Advanced Computer Science and Applications*, 2021, 12(12), 693-700.

[19] Habberrih, A., & Abuzaraida, M. A. Sentiment Analysis of Libyan Dialect Using Machine Learning with Stemming and Stop-words Removal. *In 5TH International conference on communication engineering and computer science* (CIC-COCOS'24), 2024, pp. 259-264.

[20] Alqarni, Arwa & Rahman, Atta. Arabic Tweets-Based Sentiment Analysis to Investigate the Impact of COVID-19 in KSA: A Deep Learning Approach. Big *Data and Cognitive Computing*. 2023, 7. 10.3390/bdcc7010016.

[21] Almutairi, S., & Alotaibi, F. Arabic Sentiment Analysis Evaluation of Saudi Arabia's Telecommunication Using Several Deep Learning Algorithms. *Journal of Electrical Systems*, 2024, 20(3s), 1321-1331.

[22] Abdelminaam, D. S., Neggaz, N., Gomaa, I. A. E., Ismail, F. H., & Elsawy, A. A. Arabic dialects: An efficient framework for Arabic dialects opinion mining on twitter using optimized deep neural networks. *Ieee Access*, 2021, 9, 97079-97099.

[23] Alsokkar, A., Otair, M.A., Alfar, H.E., Nasereddin, A.Y., Aldiabat, K., & Abualigah, L.M. Sentiment analysis for Arabic call center notes using machine learning techniques. *Journal of Autonomous*

Intelligence. 2024,7(3).
[24] Abubaera, M. M., & Jiddah, S. M. Natural Language Processing and Sentiment Analysis for Libyan Arabic Language Dataset. *International Journal of All Research Education and Scientific Methods* (IJARESM), 2024, Volume 12, Issue 6, 2753-2761.
[25] Abdelhadi Mahboub, Mohamed & Krishna, T.Gopi & Rao, Pyla. An identification model used for Arabic Libyan dialects based on machine learning approach. *International Journal of Advanced Research*. 2024, 12. 889-902.
[26] Hicham, N., Nassera, H., & Karim, S. Enhancing arabic e-commerce review sentiment analysis using a hybrid deep learning model and fasttext word embedding. *EAI Endorsed Transactions on Internet of Thing*s, 2024, 10.
[27] Yafooz, W. M. Enhancing Arabic Dialect Detection on Social Media: A Hybrid Model with an Attention Mechanism. *Information*, 2024, 15(6), 316.
[28] Bourahouat, G., Abourezq, M., & Daoudi, N. Improvement of Moroccan Dialect Sentiment Analysis Using Arabic BERT-Based Models. *Journal of Computer Science*. 2023, 20(2):157-167
[29] Ait Elouli, A., Cherrat, E.M., Ouahi, H., & Bekkar, A. Sentiment Analysis from Texts Written in Standard Arabic and Moroccan Dialect Based on Deep Learning Approaches. *International Journal of Computing and Digital Systems*. 2024, 16(1),447-458.
[30] Alduailej, A., & Alothaim, A. AraXLNet: pre-trained language model for sentiment analysis of Arabic. *Journal of Big Data*, 2022, 9.
[31] Alfared, Ramadan & Ly, Ramadan & Alhammi, Husien. A Topic-based Twitter Sentiment Analysis Training Dataset for Libyan Dialect. *Conference: International Conference on Technical Sciences (ICST2019)*. 2022.
[32] Elshafy, M.F., Aly, D.T., & Gheith, P.M. Analyse the Enhancement of Sentiment Analysis in Arabic by doing a Comparative Study of Several Machine Learning Techniques. *International Journal for Research in Applied Science and Engineering Technology*. 2024. 12(IV), 2007-2027.
[33] Alosaimi, S., Alharthi, M.A., Alghamdi, K.K., Alsubait, T., & Alqurashi, T. Sentiment Analysis of Arabic Reviews for Saudi Hotels Using Unsupervised Machine Learning. *Journal of Computer Science*, 2020, 16, 1258-1267.
[34] Mathebula, M., Modupe, A., & Marivate, V. ChatGPT as a Text Annotation Tool to Evaluate Sentiment Analysis on South African Financial Institutions. *IEEE Access*, 2024, 12, 144017-144043.