Applied Intelligence (2023) 53:29845–29862
https://doi.org/10.1007/s10489-023-04992-9

Arabic text detection: a survey of recent progress challenges and opportunities

Abdullah Y. Muaad[1,2] · **Shaina Raza**[3] · **Usman Naseem**[4,5] · **Hanumanthappa J. Jayappa Davanagere**[1]

Accepted: 29 August 2023 / Published online: 4 November 2023

Abstract
The Arabic language plays a crucial role in the world after becoming the sixth official language of the United Nations (UN). In the last ten years, there has been a rising growth in the number of Arabic texts, which requires algorithmic to be more effective and efficient to represent Arabic Text (AT), detecting patterns, and classifying text into the right class. Many algorithms are available for English text, but it is not the same for Arabic because of the complexity of morphology and diversity of the Arabic dialects. This study provides a survey of research in the field of Arabic Text Detection (ATD) published from 2017 to 2023. In addition, it has been conducted in a two-fold manner. Firstly, we survey based on eleven topics related to ATD. Secondly, we survey based on three stages of ATD namely pre-processing, representation, and detection. We explore all available datasets and open sources related to AT. It is revealed through the reviewed research that there are many topics of still interest to address. Furthermore, based on our observation deep-based methods yield better results only because they comprehend both the context and semantics of the language. However, they are also slower than traditional representations. Thus, hybrid models seem to be a promising way forward. Finally, we highlight new directions and discuss the open challenges and opportunities which assist researchers in identifying future work.

Keywords Arabic language · Natural language processing · Text detection · Text representation · Text pre-processing

1 Introduction

Arabic Text (AT) is rapidly expanding and is fast becoming one of the top ten languages in the world. Furthermore, AT as unstructured data introduces new challenges, which is a greater problem than preserving the quality of the organized text. Text Detection (TD) is the task of understating the pattern of text and classifying it in such a way that helps humans make a decision in many real-life situations. Researchers and developers have recently placed a great emphasis on TD to tackle various difficulties that individuals or groups face in daily life, such as sentiment analysis [1, 2], the prediction of human behavior [3], hate speech detection [4], gender detection, misogyny detection [5], sarcasm detection [6], fake news detection [7, 8], dialect detection [9] and so on.

Pre-processing plays an important role to make the AT ready for further processing, not only for text classification tasks but also for clustering, anomaly detection, and other tasks too [10–12]. There are different techniques to prepare Arabic text such as tokenization, stop word removal, lemmatization, and stemming [13]. The representation of text is also a very important task to enhance the performance of TD. Many representation models such as character level [14], sentence and word level [15], have been introduced and used in the theory of information retrieval. Different feature extraction techniques are also used for

✉ Abdullah Y. Muaad
abdullahmuaad9@gmail.com

✉ Shaina Raza
shaina.raza@utoronto.ca

✉ Hanumanthappa J. Jayappa Davanagere
hanumsbe@gmail.com

Usman Naseem
usman.naseem@sydney.edu.au

[1] Department of Studies in Computer Science, University of Mysore, Manasagangothri, Mysore 570006, India

[2] Sana'a Community College, 5695 Sana'a, Yemen

[3] University of Toronto, Toronto, Canada

[4] School of Computer Science, University of Sydney, Sydney, Australia

[5] College of Science and Engineering, James Cook University, Douglas, Australia

Table 1 The existing surveys related to Arabic text detection

Year	Ref	Pre-processing	Features Extraction	Detection/ Classification	Comparative Qualitative Analysis	Experimental Analysis	Quantitative Analysis	Covering Different Topic	Topic
2021	[23]	✓		✓		✓	✓		Fake News
2021	[24]		✓	✓		✓	✓		Sarcasm
2021	[25]	✓	✓	✓	✓	✓	✓		Dialect
2021	[26]			✓	✓		✓		Fake Information
2023	[27]		✓	✓			✓	✓	Different language
2020	[28]	✓	✓	✓		✓			Fake News
2022	[29]		✓	✓	✓			✓	Many Topics
2023	[30]		✓	✓	✓		✓	✓	General
2023	[31]	✓	✓	✓	✓		✓		Sarcasm
	Our study	✓	✓	✓	✓	✓	✓	✓	Coverage of Many Topics

extracting features such as TF-IDF [15], and fast text [16], among many others. Furthermore, dimensionality reduction techniques are also used to decrease the dimensions of textual data such as Chi-square [17, 18].

In recent years, there has been a significant increase in the volume of AT on social media [19, 20]. As a result, there is a growing need to conduct comprehensive research on the techniques and tools employed in state-of-the-art Arabic Text Detection (ATD). Many factors drove us to do this study, including the expanding number of internet users, the scarcity of academics focused on ATD, and a lack of tools and applications. Additionally, there is a potential to apply new Deep Learning (DL) based models that can effectively process a large quantity of AT.

The primary objective of this survey on ATD is to provide a comprehensive overview of the current state-of-the-art techniques, methodologies, challenges, and future directions in the field of ATD. The survey aims to achieve the following specific objectives:

- Review and evaluate the existing literature on ATD techniques, including both ML and DL-based approaches.
- Identify and explore emerging research trends and topics in the field of ATD.
- Assess the performance and effectiveness of feature extraction methods used in ATD, considering various evaluation metrics.
- Explore the availability and characteristics of publicly available datasets for training and evaluating ATD models.
- Identify the key challenges and limitations in current ATD approaches and propose directions for future research.

1.1 Searching strategy

The timeline for this survey is from 2017 to 2023. We start our survey with queries, such as 'Arabic text detection survey' or 'Arabic text detection systematic review', "Arabic dialect identification", "Arabic Irony identification", "Arabic Plagiarism identification", "Arabic crisis identification", "Arabic hate speech identification", Arabic misogyny identification", "Arabic Sports-fanaticism formalism detection", "Arabic real or fake news identification", "Arabic COVID-19 misinformation identification", "Arabic sarcasm identification" or "Arabic offensive identification". We chose different Bibliographies such as Springer Link, Association for Computing Machinery (ACM) Digital Library, IEEE Explore the digital library, and Elsevier for finding the existing works. To expand this work, we have taken other portals to find the journals, conferences, and workshops from Google Scholar too. We browse the conference proceedings of these journals to find more papers related to ATD due to the lack of studies on ATD.

Our research aims to fill a gap in the existing literature on Arabic Text Detection (ATD) surveys. As shown in Table 1, the previous surveys are mostly focused on a specific topic such as fake news detection, and sarcasm, in Arabic, but there is a lack of comprehensive surveys that cover a wide range of ATD topics. Our research makes several key contributions to the field of ATD. First, we compare existing ATD surveys, providing insights into their strengths and limitations. Second, we present a systematic taxonomy for ATD, encompassing eleven topics and three stages of ATD. Additionally, we conduct a quantitative analysis of ATD models, examining their timelines, main categories, and sub-categories. We also offer a qualitative analysis of proposed techniques used

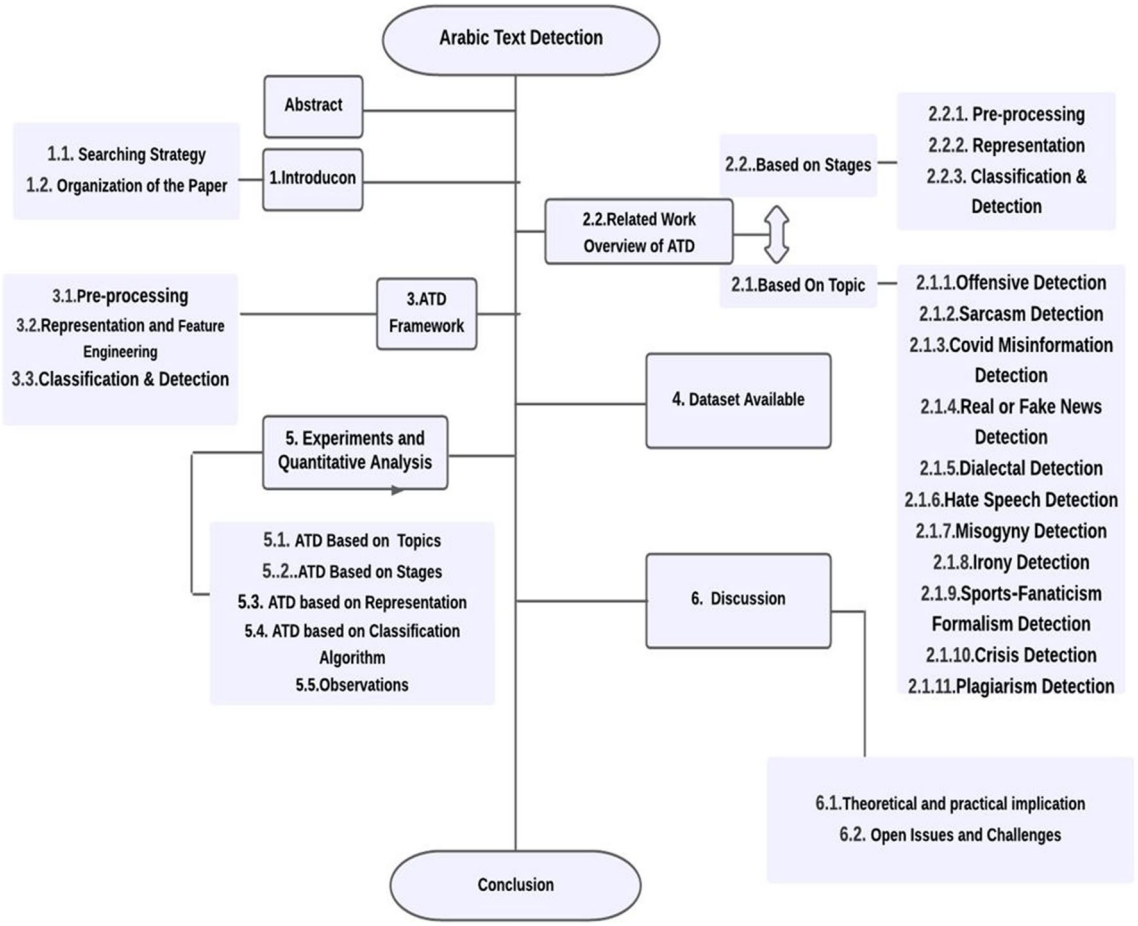

Fig. 1 Mind map diagram of ATD Arabic text detection survey

in the various stages of ATD. We provide a comprehensive list of available tools and datasets for ATD research. Lastly, we outline the challenges and future research directions in ATD. There many works for other languages which address only one stage such as representation and-classification [21, 22].

1.2 Organization of the paper

The survey has eight sections as we mention here: Sect. 1 begins with an introduction. Then, in Sect. 2, we review the related work. The proposed approach for ATD describes in Sect. 3. Then, Sect. 4 lists the available datasets, and Sect. 5 explores quantitative analysis with implications. Section 6 discusses the experimental analysis including different topics such as the theoretical and practical implications of this study, challenges, and open issues. Finally, Sect. 7 shows the conclusion. For greater clarity of this work, Fig. 1 uses a mind map graphic to show the survey's flow.

2 Related work overview of ATD

ATD is still in its developing stage and it is limited to a few notable works [32]. So, we aim to study and survey the existing work on different topics related to ATD. In sub-Sect. 2.1, we study related work on ATD based on different topics. After that in Sect. 2.2, we study and explore existing work based on different stages.

2.1 Based on topic

In this section, Table 2 summarizes the existing papers related to different topics related to ATD. There are many works available for other languages such as English [33], and Italian [34] with different topics but compare to Arabic the available works are still very less. Our observation based on Table 2 proves this point and enthusiasm us to do this survey and mentions different open areas and also mentions many challenges for future work. Figure 2 illustrates a distribution of ATD studies in this Survey.

Table 2 A summary of the existing papers related to different topics for Arabic text detection

Topic in Arabic text	الموضوع في النص العربي	Number of papers	Ref. and Year	Observation
Offensive Detection	الكشف الهجومي	8	[35–41]	All topics regarding Arabic are still required for research
Sarcasm Detection	كشف السخرية	7	[32, 42–47]	
Covid Misinformation and romurs Detection	كشف المعلومات المضللة لفيروس كوفيد	7	[48–54]	
Real or fake news Detection	كشف أخبار حقيقية أو مزيفة	7	[7, 23, 28, 55–57]	
Dialectal Detection	كشف اللهجات	6	[9, 18, 58–61]	
Hate speech Detection	كشف الكلام الذي يحض على الكراهية	4	[37, 39, 62]	
Misogyny Detection	كشف كراهية النساء	4	[19, 20, 63, 64]	
Irony Detection	كشف المفارقة	2	[65, 66]	
Sports-fanaticism formalism Detection	الرياضة - التشكلية والكشف عن التعصب	1	[67]	
Crisis Detection	كشف الأزمات	2	[68, 69]	
Plagiarism Detection	كشف السرقة الأدبية	2	[70, 71]	

Fig. 2 ATD based on topics distribution studied in this survey

2.1.1 Offensive detection

Offensive Detection refers to the identification of objectionable language in text-based conversations on digital platforms, the goal is to address harmful practices such as hate speech, misogyny, and racism [36]. Typically, an offensive detection system is built using a training set tagged with a list of offensive words. Mubarak et al., presented a new model for Arabic hate speech detection via Arabic Twittersphere. They considered different feature representations for offensive language detection such as character-level, n-grams with pre-trained embedding (Mazajak) [72].

Otiefy et al. introduced a new system for the identification of Arabic offensive language identification. They used a Convolution Neural Network (CNN), highway network, Bi-LSTM, and attention layers. One of the best models that they found was a linear support vector machine (LSVM) and they use a combination of word n-grams and characters. The macro-F1 was equal to 86.9% on the golden dataset [73].

Husain explored ensemble machine learning (ML) models for offensive Arabic language detection [38]. Additionally, Alsafari et al., investigated the framework for hate and offensive speech based on context [39], comparing five word-embedding models. Their study concluded that skip-gram models produced more effective representations than others.

In another work, Husain & Uzuner proposed transfer learning across several Arabic offensive detection datasets using the AraBERT model [41, 45]. The results reported that Arabic monolingual BERT models performed better than BERT multilingual models. While various methods have tackled these challenges, the Arabic language still poses several difficulties, indicating promising avenues for future research.

2.1.2 Sarcasm detection

Sarcasm is a type of sentiment when people use positive words in the text to express their bad [74]. Sarcasm presents a major challenge in sentiment analysis due to its ambiguity. Abu-Farha & Magdy., created a dataset called ArSarcasm, consisting of 10,547 tweets. They trained BiLSTM model for the sarcasm detection task, achieving an F1-score of 0.46 [42]. They later created a new dataset namely ArSarcasm-v2 [43], and shared high-level descriptions of top-performing teams in their shared task, achieving a sarcasm detection F1-score of 0.6225.

Hengle et al., designed a hybrid model to combine Mazajak [75] with sentence representations based on sentence level obtained from AraBERT [32]. They got 0.62 in terms of the F1-sarcastic score and 0.715 in terms of F-PN score (macro average of the F-score of the positive and negative classes) for the sarcasm and sentiment detection tasks, respectively. There are various models for other languages that have used the same idea [76]. Abuzayed & Al-Khalifa., applied seven

BERT- models and they used data augmentation to identify the sentiment of a tweet. Finally, their model detects if a tweet is sarcasm or not. Their goal is to solve the imbalanced data problem. For both tasks, their MARBERT and BERT-based models with data augmentation outperformed all other models [44].

Wadhawan,. presented a model for the detection task in the EACL WANLP-2021 Shared Task 2 [45]. They have used two transfer learning models called AraELECTRA, and AraBERT. Lichouri et al., presented a simple but intuitive detection system based on the investigation of several preprocessing steps and their combinations [46], presenting a comparison between LSVC and BiLSTM classifiers. Overall, sarcasm detection, particularly in the Arabic language, remains an intriguing area for future research due to the existing challenges yet to be fully addressed.

Mahdaouy et al. introduced an end-to-end deep MTL model that allows knowledge interaction between tasks [47]. Although different methods have addressed these challenges, the Arabic language still presents numerous difficulties. Overall, sarcasm detection, particularly in the Arabic language, remains an intriguing area for future research due to the existing challenges yet to be fully addressed.

2.1.3 COVID-19 misinformation detection

Misinformation Detection spread information without verification that it is accurate. COVID-19 misinformation detection. Raza., is one of the critical topics at this time which has a lot of research work as we will explore in this section [77]. Haouari et al., introduced the ArCOV19-Rumors Twitter dataset for the misinformation detection task. They created an Arabic Twitter dataset called ArCOV-19 [52]. Hadj Ameur & Aliane., presented AraCOVID19-MFH model [50]. They used a multi-label fake news and hate speech detection dataset for the detection task. Hasanah et al., created a dataset containing 12,000 verified Twitter accounts in Surabaya. Naelah o. Bahurmuz et al., proposed a new detection model for rumors in Arabic text. They used two transformers-based models, AraBERT and MARBERT [51]. Many researchers have created a new dataset related to information about ArCOV19 such as Mohamed Seghir et al. [50] and [52]. The misinformation detection task is still really interesting.

2.1.4 Real / fake news detection

Real / Fake News defined is as the task of classifying news as real or fake. It is a very important topic in ATD. Alkhair et al., introduced Arab corpus for the task of fake news analysis. They present several exploratory analyses to retrieve some useful knowledge [55]. Saadany et al., introduced a novel model for classifying Arabic text and they used ML for detecting whether an Arabic news article is true or satirical. They studied to identify the linguistic properties of Arabic fake news with satirical content. They built ML models to detect fake news with an accuracy of 98.6% [28]. Antoun et al., state-proposed a new model for addressing three important challenges in automated text detection: fake news detection, domain identification, and bot identification in tweets. Their experiments showed the superiority of advances in language models that can provide a deep understanding of the language [56]. Al-Yahya et al., presented neural network and transformer-based language models used for Arabic fake news detection [23]. AraBERTv02 outperformed all compared models in terms of generalization. This task is very important because many people are using this for different purposes so we are here more interested to address the multitasking problem as future work.

2.1.5 Dialectal detection

Dialectal identification is the task of the ability to distinguish between members of the same language family. Dialectal identification/detection is a hot topic in the Arabic language. Many researchers have contributed to this topic. Several researchers have made significant contributions to this field. Mubarak et al., focused on building a large Arabic offensive tweet dataset. They utilized a dataset to determine and detect text dialects. Their results with F1 were 83.2% based on SOTA dataset [35]. Abdul-Mageed, Zhang, et al., introduced the Arabic Dialect Identification model which has been done in NADI 2021 [59]. El Mekki, El Mahdaouy, Essefar, et al., aimed to detect dialect and standard language identification for many Arabic dialects[61]. El Mekki, El Mahdaouy, Berrada, et al., proposed a model for cross-dialect sentiment analysis detection [60]. In the last, dialectal this time is a really hot area regarding ATD due to the nature of the Arabic language has many dialects and many of them don't handle until this time such as the Yemeni dialect.

2.1.6 Hate speech detection

Hate speech is sharing your opinions or feelings on social media in a negative aspect that may harm others. There are many reasons for hate speech, including religious, sectarian, ethnic, regional, and partisan. However, we see that partisan discourse and political interests in most cases use other reasons to achieve the desires of a person, group, or party. So, it is an important topic in the Arabic language. Many researchers have contributed to this topic as we will explore here. Alshalan et al., analyzed and studied hate speech using Twitter data. They used DL and topic modeling to achieve this task [48]. Alshalan & Al-Khalifa., aimed to investigate CNN and RNN to detect hate speech. They evaluated BERT on the task of hate speech detection for Arabic text. Their experimental results on the dataset using the CNN model give the best performance with an F1-score of 0.79

and AUROC of 0.89. They compared four different models: BERT, CNN, CNN + GRU, and, Gated Recurrent Unit (GRU) [37, 62] proposed model for detection tasks based on hate speech and offensive language. The experiments show that using a multi-task learning setting was extremely useful due to the high correlation between the two tasks. Faris et al., proposed a smart deep-learning approach for the automatic detection of cyber hate speech [78]. Husain., investigated the impact of pre-processing on text classification, and hate speech classification for Arabic text [40]. They prove significant impacts of pre-processing, on hate speech classification. Alshalan & Al-Khalifa, proposed a model to detect hate speech using CNN and RNN [62].

2.1.7 Misogyny detection

Misogyny is hatred of, contempt for, or prejudice against women. One of the most important topics which increase in many countries via social media platforms such as Facebook etc. Mulki & Ghanem, created a dataset for the misogynistic Arabic language (LeT-Mi) [63]. They employ Multi-Task Learning (MTL). They applied SOTA systems with ML. The result in terms of accuracy was equal to 88. Frenda et al., presented a new model for the detection of misogyny using Spanish and English texts created on Twitter [19]. Muaad et al., have proposed and implemented a new technique for Arabic misogynistic detection using ML and DL [20]. Finally, they mentioned AraBERT m. Few researchers have been working on this topic as we will mention above so this really interesting area at this time.

2.1.8 Irony detection

The irony is a psychological feeling that consists of blind animosity (hatred without cause). There are some works based on this topic. Ghanem et al., proposed the first multilingual model for three languages (French, English, and Arabic). Their accuracy is equal to 80.5 ON OSACT DATASET [65]. Abdel-Salam., proposed a new model for irony detection based on Multi-headed-LSTM-CNN-GRU [66]. Irony detection is one of the most important topics to solve the problem of hatred, but unfortunately, there is not enough work until this time to reduce and discover this phenomenon.

2.1.9 Sports-fanaticism detection

Sports-fanaticism formalism is a psychological feeling that consists of blind animosity (hatred without cause). Alqmase et al., proposed a model that aims to classify Arabic text as anti-fanatic and fanatic [67]. So, they built a classification model for sentimental analysis. Sports-fanaticism formalism detection is one of the most important topics to solve the problem of hatred, but unfortunately, there is not enough work until this time to reduce and discover this phenomenon.

2.1.10 Crisis detection

Crisis detection is one of the most important topics, but unfortunately, there is not enough work until this time to reduce and discover this phenomenon. Alharbi et al., created a corpus for four high-risk floods that occurred in 2018 [68]. Alaa Alharbi et al., created a dataset for a crisis detection task. The dataset called Kawarith which have a multi-dialect Arabic Twitter for crisis events [79]. In general, these topics are still hot areas for research, and many challenges have been accomplished by different methods in other languages, for the Arabic language is still interesting for future work.

2.1.11 Plagiarism detection

Plagiarism is when someone uses another person's words or ideas and passes them off as their own. Plagiarism detection is one of the most important topics to solve the problem of hatred, but unfortunately, there is not enough work until this time to reduce and discover this phenomenon. Suleiman et al. proposed a word2vec model to detect plagiarism. The authors used OSAC corpus for training the word2vec model [70]. Ghanem et al., proposed a hybrid Arabic plagiarism detection model. called (HYPLAG). HYPLAG deals with all different types of plagiarism [71].

Based on our research this topic is one of the most important topics in academic studies but unfortunately only very less work.

2.2 Based on stages

There are many topics still not covered by TD techniques, especially in the Arabic language, which are still open for future research. One example detection of people's behavior about any phenomenon such as COVID-19. ATD has three main steps as we will see in Sect. 3 proposed methodology for ATD. Based on these three steps we explored every stage of a typical machine-learning pipeline and find the challenges that are still open for future work.

2.2.1 Pre-processing

Pre-processing is the first step in the ATD model. It is very important to prepare Arabic text for further

Table 3 A summary of the latest studies related to a different topic in Arabic text detection in terms of pre-processing

Ref	Objective	Pre-processing methods	Dataset	Evaluation metrics
[52]	Created a dataset for misinformation detection called ArCOV19-Rumors	propagation networks	Create new Dataset	NA
[40]	This investigated the impact of the pre-processing for hate speech	Pre-processing	Create Corpora and Tools	F1 scores of 89% and 95%, respectively,
[43]	Aim to create a new ArSarcasm-v2 dataset to encourage researchers to work on Arabic sarcasm,	Pre-processing	create dataset	NA
[49]	This study aimed to draw a comparison of the public's reaction to Twitter among the countries of West Asia (a.k.a the Middle East) and North Africa t	Pre-processing	Data set for multi-language	NA
[35]	In this paper, they focus on building a large Arabic offensive tweet dataset	using SOTA techniques	Create new dataset	NA
[80]	Designed a new method to prepare and analyze the Arabic text	Normalization	Create new dataset	None
[81]	Designed a new model to detect stopword	Stop word	Collect data from different resources such as Facebook and tweets	F-measure Metric
[82]	Designed a model for stemming tasks	Stemming and light stemming	OSAC	Accuracy
[83]	Aimed to use regular expressions to design a new morphological model for Arabic text	Morphological Model	Some Surat from the Holy Quran	False-positive and False-negative rate

processing although this is a very important step, we don't find so much work here because most of these processes will be included in other steps, especially with a new model such as AraBERT. Table 3 covers some of this work.

2.2.2 Representation

Representation is the second step in ATD which consider a very important step that affects the performance of the model. In this step, we will explore many different techniques starting with the traditional method bag of words until the last transfer technique for example BERT. In this step, we explore different works in the following Table 4.

2.2.3 Classification and detection

It is the step that the model starts to learn the pattern and use any approaches that help to understand the meaning after proper representation. In this section, we will explore some models such as ML and DL classification techniques as follows in Table 5.

3 ATD framework

The proposed methodology of any ATD model generally follows stages as we explore in Fig. 3. In this section, we quickly explore these stages to give a simple idea for ATD and to make this survey more efficient. Figure 3, describes all steps for ATD. We will explain every step bravely as follow:

3.1 Pre-processing

Pre-processing is a crucial step for the ATD system which makes text ready for further processing. There are many methods to prepare text as mentioned by [10]. Many steps for pre-processing and we will explain some of them as follow:

Segmentation and Tokenization: Segmentation of text is the first method to cut down the text to words or sentences or characters based on the requirement of the study purpose. A form of a segmentation text document into sentences, words, or characters is called a token. Tokenization is the process of encoding text to a number based on the representation method [10]. This is the initial phase

Table 4 a summary of the latest studies related to a different topic in Arabic text detection in terms of representation

Ref	Objective	Representation methods	Dataset	Evaluation metrics
[36]	Aimed to build effective offensive tweet detection	n-gram for representation	CD stance	88.6 precision
[65]	Aimed to propose a model for an irony detection system based on multicultural and multilingual (French, English, and Arabic)	Multilingual word representation	Arabic dataset (Ar = 11, 225 tweets)	Precision-Recall -F1 -Accuracy = 80.5 ON OSACT dataset
[28]	This study aims to identify the linguistic properties of Arabic fake news	CNN with a pre-trained word	a new dataset (3185 articles) FAKE AND the 'BBC Arabic, CNN-Arabic, and Al-Jazeera news	Accuracy 98.59
[56]	This paper proposed a model for fake news detection	XLNET	The dataset, provided by QICC	NA
[41]	This paper aimed for applying a model for Arabic offensive detection datasets	AraBERT	Aljazeera.net Deleted and many other datasets	Precision-Recall F1 Accuracy = 94 ON OSACT DATASET
[44]	This paper aimed to propose a model for a sentiment of a tweet	BERT	The ArSarcasm-v2 dataset	F1-sarcastic with data augumentation0.86 but without is 0.68
[45]	This paper aimed to propose a strategy to detect sarcasm and tackle the task in two steps	BERT	The ArSarcasm-v2 dataset	F1- sarcastic with data augumentation0.86 but without is 0.69
[23]	presented a comprehensive comparative study of fake news detection	BERT	The ArSarcasm-v2 dataset	F1- sarcastic with data augumentation0.86 but without is 0.70
[50]	Created a dataset called "AraCOVID19-MFH" for multi-label	BERT	The ArSarcasm-v2 dataset	F1- sarcastic with data augumentation0.86 but without is 0.71
[52]	created ArCOV-19, the first Arabic Twitter dataset about COVID-19	BERT	The ArSarcasm-v2 dataset	F1-sarcastic with data augumentation0.86
[63]	Created Levantine Twitter dataset for Misogynistic language	BERT—LSTM	Misogynistic Dataset for Arabic	Accuracy Precision, Recall, Macro, and F1 = 88
[20]	Aimed to Detection Misogyny from Arabic Twitter for Arabic text	ML and BERT	Let-Mi: An Arabic Levantine Twitter Dataset	Precision, Recall, Macro-F1, and accuracy
[59]	Aimed to detect dialect identification	fine-tuned multi-lingual BERT Base	Twitter APIto crawl data from 100 provinces	Precision, Recall, Macro-F1, and accuracy
[47]	The proposed deep-learning model	Bidirectional Encoder Representation from Transformers (BERT)	ArSarcasm-v2 dataset	Precision, Recall, Macro-F1, and accuracy
[60]	proposed a new unsupervised domain adaptation method for Arabic dialect and sentiment analysis	BERT	Finetuning deep pre-trained language	Improvement rate by 20.8% using zero-shot
[61]	Aimed to detect dialects of Arabic languages	MARBERT	NADI shared task's datasets	Macro-F1 and accuracy

in pre-processing, and we can use regular expressions to identify characters, words, and sentences. Then make data ready for the next step as ML algorithms require.

Stop word removal is the act of eliminating frequently occurring terms from the text that has no significant meaning. Before moving on to the representation steps, we can eliminate words, digits, and characters. Eliminating them has the advantage of reducing dimensionality and accurately defining the subject. There are some available lists for these tasks such as Arabic-stop-word-list.[1] We can use this library from Python,[2] PyArabic[3], or using rand to generate Arabic text.[4]

[1] https://github.com/yalhag1/Arabic-stop-word-list.
[2] https://pypi.org/project/Arabic-Stopwords/.
[3] https://pypi.org/project/PyArabic/.
[4] https://tahadz.wordpress.com/2020/08/10/arrand/.

Table 5 A summary of the latest studies related to the different topics in Arabic text detection in terms of detection and classification

Ref	Objective	Detection and classification methods	Dataset	Evaluation metrics
[68]	Aimed proposed model of crisis detection	(ML) and (DNN) classifiers	gold standard Arabic Twitter corpus	Accuracy 94.08
[55]	Created Arab corpus fake news	SVM, D, and MNB	Create dataset	95,35
[48]	Proposed model for detection of hate speech	CNN	Create AR COVID	NA
[62]	Aimed to detect hate speech in Arabic tweets	CNN, GRU, CNN + GRU, and BERT	GHSD dataset in Saudi dialect	F1-score of 0.79
[37]	Presented a model for offensive language and hate speech	CNN-BiLSTM,	the SemEval 2020 Arabic offensive language dataset	NA
[72]	provided an overview of the offensive language detection	FastText, CNN + RNN, and BERT	NA	NA
[73]	proposed hybrid model based CNN, highway network, Bi-LSTM, that enhanced the predictive	CNN, highway network, Bi-LSTM,	Golden dataset under CodaLab username "yasserotiefy"	Macro-F1 86.9% on
[42]	In this paper, they presented ArSarcasm, an Arabic sarcasm detection dataset	BiLSTM	The dataset contains 10,547 tweets	F1-score of 0.46,
[38]	Investigated model to detect offensive Arabic using a single and an ensemble learning	SVM, LR, and DT	(OSACT)dataset	F1 score of 88%,
[39]	investigated how the word-embedding affects deep neural networks for hate and offensive speech detection task	CNN, GRU, BILSTM, and hybrid CNN + BILSTM	Arabic hate speech multi-class dataset	87.22 F-M for binary classification
[78]	The objective of this paper was to propose a smart deep-learning approach for the automatic detection of cyber hate speech	LSTM and CNN	Hence, a dataset is collected from Twitter	Accuracy, precision, recall, and F1 measure
[27]	Aimed to propose a model for sarcasm detection tasks	CNN, BLSTM, and AraBERT language model	sarcasm shared-task 2021 follows	CNN, BLSTM, and AraBERT language model
[32]	Aimed to propose a hybrid model for sarcasm and sentiment detection tasks	CNN and BILASTM	the ArSarcasm v2 dataset	NA
[66]	Proposed model for the sarcasm detection task	Multi-headed-LSTM-CNN-GRU and also MARBERT	ArSarcasm-v2 dataset (Abu Farha	NA
[67]	Proposed model to classify text into fanatic and anti-fanatic	LR	Built lexicon and corpus	Accuracy 91%
[53]	Collected 12,000 sample data. Then proposed a model to classify COVID-19 into seven classes	W2V, fastText and CNN, RNN, LSTM	collects 12,000 datasets	Accuracy is 97.3% and 99.4%
[46]	presented a simple but intuitive detection system	LSVC) and BiLSTM	The arSarcasm-v2 dataset	Accuracy score of 57.87%

Word filtering: In this step, users can use different methods such as the rule-based algorithm and lexicon to filter some words and delete them or replace them with some additional terms. The filtering methods could be useful for spam tweets [13].

Emojis description: there are many different emoticons such as happy etc. can be used to express sentiments. So, in addition, it is important to capture this useful information. The data need to prepare for further processing.

Cleaning: it is the process of deleting unwanted words or symbols such as punctuations, additional whitespaces, diacritics, and non-Arabic characters. Cleaning text is one method to decrease the number of a word which help to decrease dimension and improve accuracy.

Normalization: it is the process of transforming the alphabet into another. The normalization steps are as follows: Different forms of ("آ" ,"إ" ,"أ ") were replaced by" ي" ," ا " is replaced by "ى" and "ة " is replaced by"ه ".

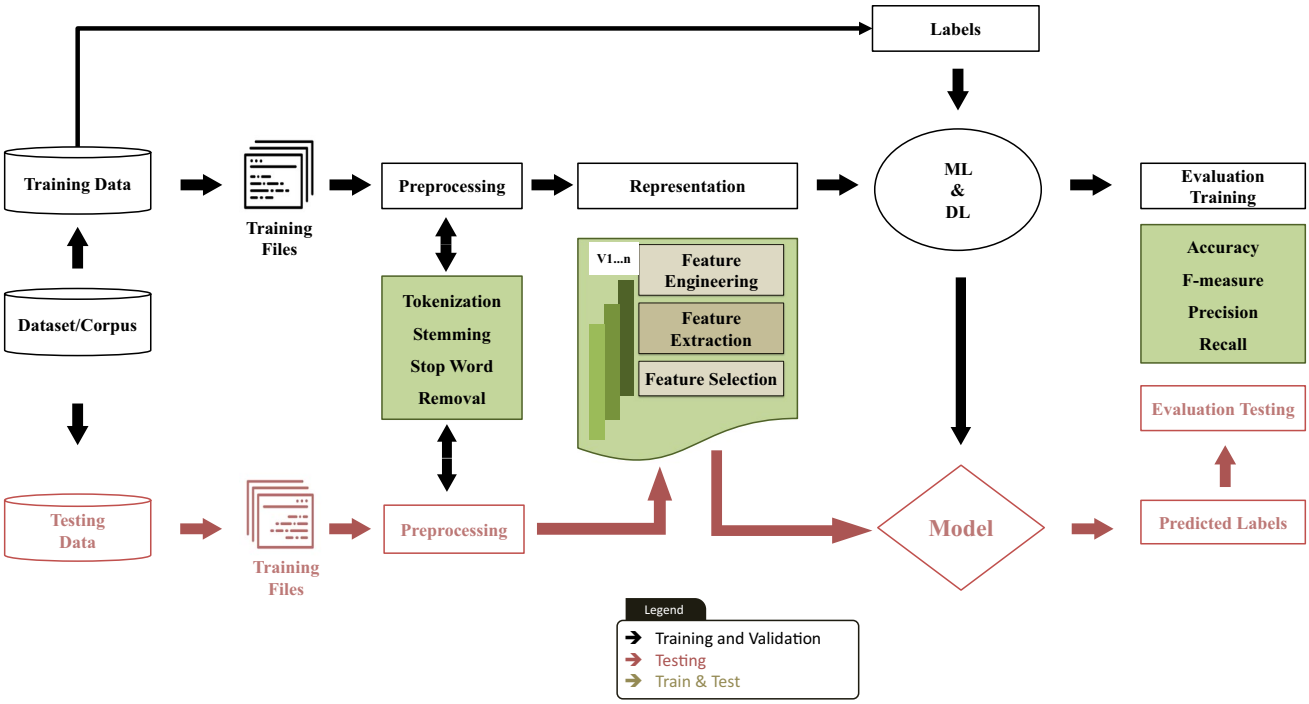

Fig. 3 Generic architecture of automatic Arabic text detection system

Lemmatization and stemming: It is one of the pre-processing techniques which cut down a word to its base. Lemmatization used lexical knowledge to transform a word into its base. Lemmatization is slower than stemming because stemming doesn't use lexical knowledge. Finally, lemmatization depends on the dictionary whereas stemming depends on a regular expression. Many other tools have been used for Arabic text for pre-processing in CAMEL-lab[5] such as MADAMIRA,[6] MADAR,[7] Computational Approaches to Modeling Language,[8] Gumar Corpus,[9] and ADIDA.[10]

3.2 Representation

Representation of text is an essential step in allowing the machine to understand, identify, analyze, and classify text data in the same way that people do. It is the process of converting unstructured data into structured which became machine-readable numeric data. Different techniques have been used for this goal. These methods can be divided into two categories: statistical ML methods such as a bag of words and DL methods such as Word2Vector (W2V) which are all referred for feature engineering techniques. At the same time, the text is a sequence of characters or a sequence of words, or even a sequence of sentences, and based on the problem we can decide the right level is called a representation of text. We summarize different representation methods and grasp the merits and demerits as shown in Sect. 5.5 Table 7. Using traditional representation techniques to represent text will not handle the semantic meaning and syntactic for this reason, we need to work with a new feature extraction model such as word embedding and AraBERT. AraBERT is a new technique that works based on transfer learning and the attention concept is all you need [84]. AraBERT model was trained with 2 tasks to encourage bifacial prediction and sentence-level understanding. It is trained on 2 unsupervised tasks and

[5] https://github.com/CAMeL-Lab.

[6] https://camel.abudhabi.nyu.edu/madamira/.

[7] https://sites.google.com/nyu.edu/madar/.

[8] https://nyuad.nyu.edu/en/research/faculty-labs-and-projects/computational-approaches-to-modeling-language-lab.html.

[9] https://camel.abudhabi.nyu.edu/gumar/?page=search&term=%D8%AA%D8%B1%D8%A7.

[10] https://adida.abudhabi.nyu.edu/#.

Table 6 Dataset available for Arabic text detection in different topics

Ref	Website	Dataset Name	No of classes	No of words	No of Documents	Remark
[85]	https://sourceforge.net/projects/ar-text-mining/files/Arabic-Corpora/	CNN	6	2,241,348	5070	OSAC
[85]	https://sourceforge.net/projects/ar-text-mining/files/Arabic-Corpora/	BBC	7	1,860,786	4,763	OSAC
[85]	https://sourceforge.net/projects/ar-text-mining/files/Arabic-Corpora/	OSAc	10	18,183,511	22,429	OSAC
[43]	https://github.com/iabufarha/ArSarcasm	ArSarcasm dataset	2	NA	10,547	Hate speech
[42]	https://github.com/iabufarha/ArSarcasm-v2	ArSarcasm-v2 dataset	2	NA	15,548	Sarcasm
[63]	https://github.com/bilalghanem/let-mi	An Arabic Levantine Twitter Dataset for Misogynistic Language	2&8	NA	7865& 6550	Misogynistic
[52]	https://gitlab.com/bigirqu/ArCOV-19/-/tree/master/	ArCOV-19	-	-	-	COV-19
[62]	https://github.com/raghadsh/Arabic-Hate-speech	dataset (GHSD)	3	-	9,316	Hate Speech
[54]	https://github.com/mohaddad/COVID-FAKES	Bilingual (Arabic/English) COVID-19 Twitter dataset	2	-	-	English -Arabic
[86]	https://github.com/nuhaalbadi/Arabic_hatespeech	Religious Hate Speech in the Arabic	-	2	7136	Arabic
[87]	https://github.com/UBC-NLP/marbert	Coverage Topics§	-	-	-	Arabic
[88]	https://github.com/HKUST-KnowComp/MLMA_hate_speech	MLMA_hate speech	-	-	-	Multilingual
[89]	https://alt.qcri.org/resources/ArCovidVac.zip	ArCovidVac	2 & 3 &10	-	10000	COVID-19 with three different goals Informative, Stance, and Fine-grained categorization

used for many tasks in NLP. We have compared different techniques of representation based on the experiment as follows in the Table 4.

3.3 Detection and classification model

The classification task is crucial for text classification due to the characteristics of text as widespread forms of sequence data. There are many algorithms have been proposed. Some widely used algorithms for detection and classification include Naive Bayes, Support Vector Machines (SVM), decision tree-based algorithms like Random Forest and Gradient Boosting, as well as deep learning techniques such as Recurrent Neural Networks (RNNs) and CNNs. Transfer learning with pre-trained language models like BERT has also shown promising results. The choice of the model depends on numerous factors such as the nature of the text data and specific task requirements.

4 Dataset available

There are several datasets available for ATD however, they are quite limited in comparison to English and most of them are very small and it is not efficient for many real-world scenarios. Some of these datasets mention with their links as follows in Table 6, and some other dataset available on the internet has been mentioned in [60], and also Masader.[11]

5 Experiments and quantitative analysis

This section provides a quantitative analysis of the reviewed research papers focusing on ATD across various topics and stages. We delve into the quantitative aspects of the existing work, considering a total of 48 articles as part of our study.

[11] https://arbml.github.io/masader/.

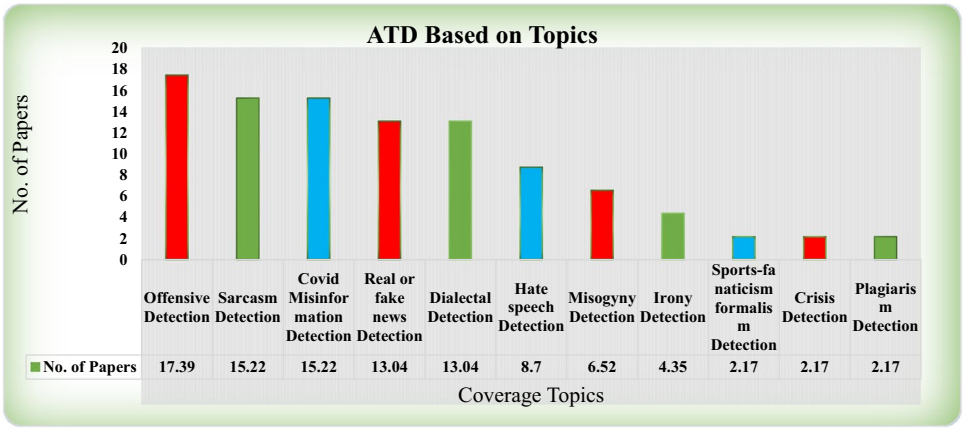

Fig. 4 Distributions of ATD papers based on covering topics

Our analysis is centered on the stages and topics relevant to ATD. By conducting these analyses, we aim to address the following questions:

- How many research papers have been published on each topic and stage of ATD across the timeline from 2017 to 2021?
- Which stages of ATD models have received the most and least attention in research?
- How is the distribution of papers for each topic/stage based on the methods employed?
- What are the key challenges and limitations that ATD still faces, requiring further research?

5.1 ATD based on topics

The total number of reviewed research papers related to the topic was 46, which is divided into 16 topics. However, Fig. 4, shows the distribution of published papers among the topics starting with Offensive which is equal to 17.39%, and ending with plagiarism detection which is equal to 2.17. It can be observed that the offensive topic has obtained the highest percentage because some available dataset is there. Whereas topics such as prediction of human behavior, and exploring halal tourism tweets detection have not any publications. Finally, by observing Fig. 4, we can prove that there are so many topics still open for future work. At the same time not only these topics but there are others available.

5.2 ATD based on stages

The total number of reviewed research papers related to the stages of ATD is equal to 48 articles. However, Fig. 5, shows the distribution of published papers among the Pre-processing, Representation, and detection, with 29.17%, 35.42%, and 35.42%, respectively.

It can be observed that the Representation and detection category has obtained the highest percentage 35.42% because many researchers will include pre-processing in their model. Finally, we can observe and conclude that focusing on steps is not important as topics because analyzing data based on a topic will not discover challenges for future work. The most important here is that this work opens our minds to make sure that there is gab in finding suitable representation techniques and detection algorithms too.

5.3 ATD based on representation

There are 48 journal and conference papers available for the representation stage. At the end of this paper, several characteristics related to ATD have been studied. We compare them based on different characteristics as we explore them in Table 7.

5.4 ATD based on classification algorithm

Table 8 explores our comparison of the classification algorithms based on understanding some merits and demerits of these algorithms. Based on our research which we discussed in sub-Sect. 1.1 and sub-Sect. 2.2.3. We understand that any algorithm has its own merits and demerits. Table 8, the characteristic refers to the merits and demerits algorithm.

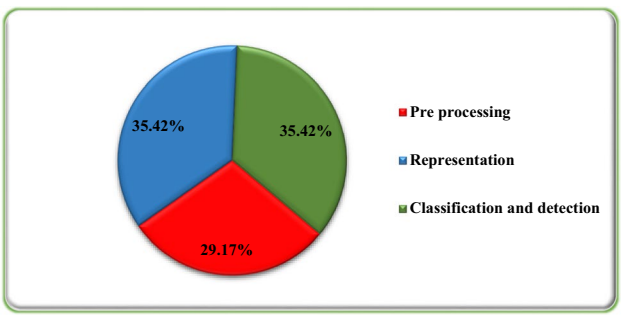

Fig. 5 Distributions of ATD papers based on pre-processing, representation and classification

Table 7 Comparing ML & DL representation models evaluated by semantics, syntactical, context, and out of vocabulary Y: Yes, and N: No

Representation Model	ML/DL Models	Semantics	Syntactical	Context	Out of Vocabulary
One Hot encoding	Classical Models	N	N	N	N
BoW		N	N	N	N
TF		N	N	N	N
TF-IDF		N	N	N	N
W2V[12]	DL Models	Y	Y	N	N
Glo2Vec[13]		Y	Y	N	N
AraBERT[14]		Y	Y	Y	N
FasText[15]		Y	Y	N	Y

[12] https://github.com/bakrianoo/aravec

[13] https://nlp.stanford.edu/projects/glove/

[14] https://github.com/aub-mind/arabert

[15] https://fasttext.cc/docs/en/crawl-vectors.html

Table 8 Comparing ML & DL classification models evaluated by passive aggressive classifier PAC, Logistic Regression LR, K-Nearest Neighbor KNN, Linear Support Vector Machine (LSVC), Decision Tree DT, Random Forest RF, AraBERT

Classification Algorithm Characteristics	PCA	LR	KNN	LSVC	DT	RF	AraBERT
Easy to implement	Y	Y	Y	Y	Y	N	N
Robust and easy to understand	Y	Y	Y	Y	Y	Y	N
Requires less amount of data	Y	N	Y	Y	Y	N	N
Effective in higher dimension	Y	N	N	Y	N	Y	Y
Find an efficient architecture difficult	Y	N	N	N	N	N	Y
Is it a fast algorithm	Y	Y	Y	Y	Y	N	N
Less pre-processing required	Y	N	Y	Y	Y	Y	Y
Is it black-box	N	N	N	N	N	N	Y
Fails in case of non-linear problems	Y	N	N	N	N	N	N
Parallel processing capability	Y	N	N	N	N	N	Y

5.5 Observations

According to the quantitative analysis results in Sects. 5.1, and 5.2, also to qualitative analysis in Tables 3, 4, and 5, we explore how the existing methods distribute based on different topics and stages related to ATD. Using coverage topic and stage gives us good visualization to understand how the existing work is handling ATD for different tasks. From a theoretical standpoint, ATD is critical in perception, planning, reasoning, and decision-making, not just for governments or organizations, but also for individuals. We carefully select and optimize the study of this topic to be capable of handling many real scenarios. This is because most of the increase in users writing every day and their writing affects a lot of people such as rumors and fake news etc. In addition, ATD is one of the most important hot topics for different domains. Thus, we are highly enthusiastic to mention these implications of ATD which exactly will help people in many real-life scenarios such as health, education, and economics, and making them safe from fake news, rumors, etc. Based on our observation we might consider ATD in comparison to other languages to be a promising subject in which researchers can work on several topics such as offensive, sarcasm, COVID-19 misinformation, real or fake news, dialectal, hate speech, misogyny, irony, Sports-fanaticism formalism, crisis, plagiarism, psychopathic personality trait, exploring halal tourism tweets detection, gender detection and prediction of human behavior.

6 Discussion

This section presents and discusses the survey findings. In each of the sections that follow. (6.1, 6.2), we present and discuss our findings in conjunction with responding to the questions and also find challenges and open questions for the future.

6.1 Theoretical and practical implication

ATD has become one of the hot topics for researchers working on Arabic processing systems. With the strongly

increased data and users in the Arabic language in general, it has, as a result, become difficult to understand, classify, monitor, and track AT is a difficult task. At the same time, lack of resources and tools for the Arabic language. In this work, we survey the results of a review article of existing work. Specifically, we considered 11 topics that were presented in Sect. 2.1 from journals and conferences in the last 5 years. Our study proves that all the topics we have studied are still interesting for future research. And also, it turned out that all of them can help in different real-life scenarios. So, it has become one of the hot topics for researchers working on Arabic processing systems. There are many new ideas that ATD can support, such as fine-grained analysis, cross-domain transfer, multi-modal integration, monitoring rumors, detecting hate speech, and ethical considerations; all these domains can utilize ATD technique so this survey has a lot of important to detect and use the proper the right way for each situation.

We conclude this experimental analysis and discussion section by proving the importance of this survey and summarizing the key insights gained from the survey responses. Emphasize the significance of the findings and their implications for advancing ATD research and applications.

6.2 Open issues and challenges

The research on the problem of ATD is still in the nascent stage as we mention above so there are many challenges still open for research. As we divided this survey into two sections in the related work section. We explore some of these challenges in this section and for more details about these challenges we can refer [90–93]. In this section, we explore some of them as follows:

- Lack of studies for ATD compared to English languages.
- Finding roots of Arabized words for example (programs) (برامج.) and (computer)(13] (كمبيوتر[.
- Handling the negation of Arabic sentences, especially with various dialects [94].
- Arabic dialects with different meaning is another challenge.
- The problem of synonyms, and broken plural forms is widespread which makes it difficult to recognize and understand the meaning.
- Pre-processing still has many challenges for ATD tasks such as normalization, stemming and lemmatization.
- Finding effective context representation models for a specific task are still required.
- Out of vocabulary to make the model able to understand any word that has not been seen before.
- Designing proper classification methods for new concepts such as multi-label and multi-task detection.
- Many other issues in this part such as orthographic, ambiguity, dialectal variation, and stemming [90, 91].
- Lack of availability of benchmark datasets and lack of lexicons.
- proposing augmentation data techniques to solve the imbalance of classes could get better performance, especially with minor classes.
- Creating new datasets for different detection and classification tasks.
- We plan for mitigating biases and study fairness in ATD for different topics [95].

Finally, by comparing Arabic to another language such as English there are many topics still interest such as psychopathic personality trait detection, exploring halal tourism tweets detection, gender detection, identification, prediction of human behavior, political platforms, and many more still not handled [77].In addition, we plan to study in deep how to investigate Arabic text understanding. Furthermore, we also plan to create and apply our own transformation models, which work based on attention is all you need.

7 Conclusion

This survey so far is limited to papers published related to ATD for 2017 to 2023 only. So, the study included a complete survey for ATD across a variety of topics, as well as three stages, namely pre-processing, representation, and detection. Our analysis indicates that the existing work is still very less and more challenging and can be addressed in the future. Furthermore, it turned out that most of the topics that we have explored are still not addressed (interest for the future). Therefore, this survey calls for more rigor and better research practices with respect to different topics in ATD. In addition, this study opened up new avenues of challenges and future directions such as multi-models, multi-language models, and difficulties regarding the nature of the Arabic language (dialectal, morphology, and stemming). Finally, based on our understanding, this study is helpful for the research community in finding gaps and challenges related to the ATC system in different domains such as healthcare, economics, business, education, etc. We make sure our survey has a new idea by addressing different topics. Furthermore, this work is a review article and theory-driven so it is a summarization of previous works with some demonstrations.

Acknowledgements The first author is grateful to the Department of Studies in Computer Science and the chairman of the Department, University of Mysore, Manasagangothri, Mysore-570006, Karnataka, India. for all the facilities offered for this research. Besides, the first author wishes to acknowledge Sana'a Community College, Sana'a 5695, Yemen, for their support and all the facilities provided for this

research work. A Special thanks and gratitude to Wadeea R.Naji for all help and support in this work.

Data Availability NA.

Declarations

Conflict of interest There are no conflicts of interest associated with publishing this paper.

References

1. Catelli R, Fujita H, De Pietro G, Esposito M (2022) Deceptive reviews and sentiment polarity: Effective link by exploiting BERT. Expert Syst Appl 209:118290. https://doi.org/10.1016/J.ESWA.2022.118290
2. Catelli R et al (2022) Cross lingual transfer learning for sentiment analysis of Italian TripAdvisor reviews. Expert Syst Appl 209:118246. https://doi.org/10.1016/J.ESWA.2022.118246
3. Davahli MR et al (2020) Identification and prediction of human behavior through mining of unstructured textual data. Symmetry (Basel) 12(11):1–23. https://doi.org/10.3390/sym12111902
4. Rajput G, Punn NS, Sonbhadra SK, Agarwal S (2021) Hate speech detection using static BERT embeddings. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 13147:67–77. https://doi.org/10.1007/978-3-030-93620-4_6/COVER
5. Samghabadi NS, Patwa P, Srinivas PYKL, Mukherjee P, Das A, Solorio T (2020) Aggression and misogyny detection using BERT: a multi-task approach. In: Proceedings of the Second workshop on trolling, aggression and cyberbullying. European Language Resources Association (ELRA), Marseille, pp 126–131. https://aclanthology.org/2020.trac-1.20
6. Yaghoobian H, Arabnia HR, Rasheed K (2021) Sarcasm detection: a comparative study. Available http://arxiv.org/abs/2107.02276
7. Mahlous AR, Al-Laith A (2021) Fake news detection in Arabic tweets during the COVID-19 pandemic. Int J Adv Comput Sci Appl 12(6):778–788. https://doi.org/10.14569/IJACSA.2021.0120691
8. Raza S, Ding C (2022) Fake news detection based on news content and social contexts: a transformer-based approach. Int J Data Sci Anal. https://doi.org/10.1007/s41060-021-00302-z
9. Abo MEM et al (2021) A multi-criteria approach for arabic dialect sentiment analysis for online reviews: Exploiting optimal machine learning algorithm selection. Sustainability 13(18):10018. https://doi.org/10.3390/su131810018
10. Naseem U, Razzak I, Eklund PW (2020) A survey of pre-processing techniques to improve short-text quality: a case study on hate speech detection on twitter. Multimed Tools Appl. https://doi.org/10.1007/s11042-020-10082-6
11. Murshed BAH, Mallappa S, Ghaleb OAM, Al-ariki HDE (2021) *Efficient twitter data cleansing model for data analysis of the pandemic tweets*, vol. 348, no. March. Springer International Publishing. https://doi.org/10.1007/978-3-030-67716-9_7
12. Modhaffer M, Sivaramakrishna CV (2017) Prepositional verbs in Arabic: A corpus-based study. Lang India 17(10):154. Available https://www.researchgate.net/publication/320677565_Prepositional_Verbs_in_Arabic_A_Corpus-based_Study
13. Alhaj YA, Xiang J, Zhao D, Al-Qaness MAA, AbdElaziz M, Dahou A (2019) A study of the effects of stemming strategies on Arabic document classification. IEEE Access 7:32664–32671. https://doi.org/10.1109/ACCESS.2019.2903331
14. Muaad AY, Jayappa H, Al-antari MA, Lee S (2021) ArCAR: A novel deep learning computer-aided recognition for character-level Arabic text representation and recognition. Algorithms 14(7):216. https://doi.org/10.3390/a14070216
15. Alhaj YA, Al-qaness MAA, Dahou A, Abd Elaziz M, Zhao D, Xiang J (2020) Effects of light stemming on feature extraction and selection for Arabic documents classification, vol 874. In: Studies in computational intelligence, pp 59–79. https://doi.org/10.1007/978-3-030-34614-0_4
16. Joulin A, Grave E, Bojanowski P, Mikolov T (2017) Bag of tricks for efficient text classification. 15th Conf Eur Chapter Assoc Comput Linguist EACL 2017 - Proc Conf 2:427–431. https://doi.org/10.18653/v1/e17-2068
17. Alhaj YA, Xiang J, Zhao D, Al-Qaness MAA, Abd Elaziz M, Dahou A (2019) A study of the effects of stemming strategies on arabic document classification. IEEE Access 7:32664–32671. https://doi.org/10.1109/ACCESS.2019.2903331
18. Alali M, MohdSharef N, Azmi Murad MA, Hamdan H, Husin NA (2019) Narrow convolutional neural network for Arabic dialects polarity classification. IEEE Access 7:96272–96283. https://doi.org/10.1109/ACCESS.2019.2929208
19. Frenda S, Ghanem B, Montes-y-Gómez M (2018) Exploration of misogyny in Spanish and english tweets. CEUR Workshop Proc 2150:260–267
20. Muaad Y, Hanumanthappa J, Al-antari MA, V BBJ, Chola C (2021) "AI-based Misogyny Detection from Arabic Levantine Twitter Tweets," Proc 1st Online Conf Algorithms, MDPI Basel, Switzerland, pp. 4–11 https://doi.org/10.3390/IOCA2021-10880,pp.4-11,2021
21. Rauniyar K, Shiwakoti S, Poudel S, Thapa S, Naseem U, Nasim M (2023) Breaking barriers: exploring the diagnostic potential of speech narratives in Hindi for alzheimer's disease. In: Proceedings of the 5th clinical natural language processing workshop, pp 24–30. https://doi.org/10.18653/v1/2023.clinicalnlp-1.4
22. Naseem U, Razzak I, Khan SK, Prasad M (2021) "A comprehensive survey onword representation models: From classical to state-of-the-artword representation language models",. ACM Trans Asian Low-Resource Lang Inf Process 20(5):1–35. https://doi.org/10.1145/3434237
23. Al-Yahya M, Al-Khalifa H, Al-Baity H, Alsaeed D, Essam A (2021) "Arabic fake news detection: Comparative study of neural networks and transformer-based approaches," Complexity 2021. https://doi.org/10.1155/2021/5516945
24. Farha IA, Magdy W (2021) A comparative study of effective approaches for Arabic sentiment analysis. Inf Process Manag 58(2):102438. https://doi.org/10.1016/j.ipm.2020.102438
25. Elnagar A, Yagi SM, Nassif AB, Shahin I, Salloum SA (2021) Systematic literature review of dialectal Arabic: Identification and detection. IEEE Access 9:31010–31042. https://doi.org/10.1109/ACCESS.2021.3059504
26. Seyam A, Nassif AB, Talib MA, Nasir Q, Nassif B (2021) Deep learning models to detect online false information: a systematic literature review. In: ArabWIC 2021: The 7th annual international conference on Arab women in computing in conjunction with the 2nd forum of women in research, pp 1–5. https://doi.org/10.1145/3485557.3485580
27. Jahan MS, Oussalah M (2023) A systematic review of hate speech automatic detection using natural language processing. Neurocomputing 546:126232. https://doi.org/10.1016/J.NEUCOM.2023.126232
28. Saadany H, Orašan C, Mohamed E (2020) "Fake or Real? A Study of Arabic Satirical Fake News." Online, pp. 70–80. Accessed: May 29, 2023. [Online]. Available: https://aclanthology.org/2020.rdsm-1.7
29. Alhumoud SO, Al Wazrah AA (2022) Arabic sentiment analysis using recurrent neural networks: a review. 55(1). Springer Netherlands. https://doi.org/10.1007/s10462-021-09989-9
30. Wahdan A, Al-Emran M, Shaalan K (2023) "A systematic review of Arabic text classification: areas, applications, and

future directions," Soft Comput 6. https://doi.org/10.1007/s00500-023-08384-6
31. Rahma A, Azab SS, Mohammed A (2023) A comprehensive survey on Arabic sarcasm detection: approaches, challenges and future trends. IEEE Access 11:18261–18280. https://doi.org/10.1109/ACCESS.2023.3247427
32. Hengle A, Kshirsagar A, Desai S, Marathe M (2021) Combining context-free and contextualized representations for Arabic sarcasm detection and sentiment identification, pp 357–363. https://aclanthology.org/2021.wanlp-1.46. Accessed 29 May 2023
33. Nadeem MI et al (2022) SHO-CNN: A metaheuristic optimization of a convolutional neural network for multi-label news classification. Electronics 12(1):113. https://doi.org/10.3390/electronics12010113
34. Catelli R et al (2023) A new Italian Cultural Heritage data set: detecting fake reviews with BERT and ELECTRA leveraging the sentiment. IEEE Access 11(May):52214–52225. https://doi.org/10.1109/ACCESS.2023.3277490
35. Mubarak H, Rashed A, Darwish K, Samih Y, Abdelali A (2021) Arabic offensive language on Twitter: analysis and experiments, pp 126–135. https://aclanthology.org/2021.wanlp-1.13 . Accessed 29 May 2023
36. Mubarak H, Darwish K (2019) Arabic offensive language classification on twitter. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 11864(April 2020):269–276. https://doi.org/10.1007/978-3-030-34971-4_18
37. Abu Farha I, Magdy W (2020) Multitask learning for {A}rabic offensive language and hate-speech detection. In: Proceedings of the 4th Workshop on open-source Arabic corpora and processing tools, with a shared task on offensive language detection, pp 86–90. Available https://www.aclweb.org/anthology/2020.osact-1.14
38. Husain F (2020) Arabic offensive language detection using machine learning and ensemble machine learning approaches. ArXiv. http://arxiv.org/abs/2005.08946. Accessed 29 May 2023
39. Alsafari S, Sadaoui S, Mouhoub M (2020) Hate and offensive speech detection on Arabic social media. Online Soc Netw Media 19:100096. https://doi.org/10.1016/J.OSNEM.2020.100096
40. Husain F (2020) OSACT4 shared task on offensive language detection: intensive preprocessing-based approach. European language resource association, Marseille, pp 53–60
41. Husain F, Uzuner O (2021) Transfer learning approach for Arabic offensive language detection system BERT-based model. arXiv. Available https://api.semanticscholar.org/CorpusID:231879804
42. Abu-Farha I, Magdy W (2020) From Arabic sentiment analysis to sarcasm detection: The ArSarcasm dataset. In: Proceedings of the 4th Workshop on open-source arabic corpora and processing tools, with a shared task on offensive language detection. European Language Resource Association, Marseille, pp 32–39. https://aclanthology.org/2020.osact-1.5
43. Abu Farha I, Zaghouani W, Magdy W (2021) Overview of the WANLP 2021 shared task on sarcasm and sentiment detection in Arabic. In: Proceedings of the sixth Arabic natural language processing workshop, pp 296–305. Available https://www.aclweb.org/anthology/2021.wanlp-1.36
44. Abuzayed A, Al-Khalifa H (2021) Sarcasm and sentiment detection in Arabic tweets using BERT-based models and data augmentation. In: Proceedings of the sixth Arabic natural language processing workshop, pp 312–317. Available https://api.semanticscholar.org/CorpusID:233365108
45. Wadhawan A (2021) AraBERT and Farasa segmentation based approach for sarcasm and sentiment detection in Arabic tweets, pp 395–400. https://aclanthology.org/2021.wanlp-1.53. Accessed 29 May 2023
46. Lichouri M, Abbas M, Benaziz B, Zitouni A, Lounnas K (2021) Preprocessing solutions for detection of sarcasm and sentiment for {A}rabic. In: Proceedings of the sixth Arabic natural language processing workshop, pp 376–380. Available https://www.aclweb.org/anthology/2021.wanlp-1.49
47. El Mahdaouy A, El Mekki A, Essefar K, El Mamoun N, Berrada I, Khoumsi A (2021) Deep multi-task model for sarcasm detection and sentiment analysis in Arabic language, pp 334–339. https://aclanthology.org/2021.wanlp-1.42. Accessed 29 May 2023
48. Alshalan R, Al-Khalifa H, Alsaeed D, Al-Baity H, Alshalan S (2020) Detection of hate speech in COVID-19-related tweets in the Arab Region: Deep learning and topic modeling approach. J Med Internet Res 22(12):e22609. https://doi.org/10.2196/22609
49. Jafarian H et al (2021) "Topic discovery on farsi, English, French, and Arabic tweets related to COVID-19 using text mining techniques," 26–33. https://doi.org/10.3233/shti210084
50. HadjAmeur MS, Aliane H (2021) AraCOVID19-MFH: Arabic COVID-19 multi-label fake news & hate speech detection dataset. Procedia Comput Sci 189(May):232–241. https://doi.org/10.1016/j.procs.2021.05.086
51. Bahurmuz NO, Amoudi GA, Baothman FA, Jamal AT, Alghamdi HS, Alhothali AM (2022) Arabic rumor detection using contextual deep bidirectional language modeling. IEEE Access 10(November):114907–114918. https://doi.org/10.1109/ACCESS.2022.3217522
52. Haouari F, Hasanain M, Suwaileh R, Elsayed T (2021) "ArCOV-19: The first Arabic COVID-19 twitter dataset with propagation networks." 82–91. Accessed: May 29, 2023. [Online]. Available: https://aclanthology.org/2021.wanlp-1.9
53. Hasanah NA, Suciati N, Purwitasari D (2021) Identifying degree-of-concern on covid-19 topics with text classification of twitters. Regist J Ilm Teknol Sist Inf 7(1):50–62. https://doi.org/10.26594/register.v7i1.2234
54. Elhadad MK, Li KF, Gebali F (2021) "COVID-19-FAKES: A twitter (Arabic/English) dataset for detecting misleading information on COVID-19," 256–268. https://doi.org/10.1007/978-3-030-57796-4_25
55. Alkhair M, Meftouh K, Smaïli K, Othman N (2019) "An Arabic corpus of fake news: Collection, analysis and classification," 292–302. https://doi.org/10.1007/978-3-030-32959-4_21
56. Antoun W, Baly F, Achour R, Hussein A, Hajj H (2020) "State of the art models for fake news detection tasks," 2020 IEEE Int Conf Informatics, IoT, Enabling Technol ICIoT 519–524. https://doi.org/10.1109/ICIoT48696.2020.9089487
57. Shishah W (2022) JointBert for detecting Arabic fake news. IEEE Access 10(June):71951–71960. https://doi.org/10.1109/ACCESS.2022.3185083
58. Alruily M (2020) Issues of dialectal Saudi twitter corpus. Int Arab J Inf Technol 17(3):367–374. https://doi.org/10.34028/iajit/17/3/10
59. Abdul-Mageed M, Zhang C, Elmadany A, Bouamor H, Habash N (2021) NADI 2021: The second nuanced Arabic dialect identification shared task. In: Proceedings of the sixth Arabic natural language processing workshop. association for computational linguistics, Kyiv, pp 244–259. http://arxiv.org/abs/2103.08466
60. El Mekki A, El Mahdaouy A, Berrada I, Khoumsi A (2021) "Domain adaptation for Arabic cross-domain and cross-dialect sentiment analysis from contextualizedword embedding," 2824–2837. https://doi.org/10.18653/v1/2021.naacl-main.226
61. El Mekki A, El Mahdaouy A, Essefar K, El Mamoun N, Berrada I, Khoumsi A (2021) BERT-based multi-task model for country and province Level MSA and dialectal Arabic identification. In: Proceedings of the Sixth Arabic Natural Language Processing Workshop, pp 271–275. Available https://www.aclweb.org/anthology/2021.wanlp-1.31
62. Alshalan R, Al-Khalifa H (2020) A deep learning approach for automatic hate speech detection in the saudi twittersphere. Appl Sci 10(23):1–16. https://doi.org/10.3390/app10238614

63. Mulki H, Ghanem B (2021) Let-Mi: an Arabic Levantine Twitter dataset for Misogynistic language. Association for computational linguistics. Note, pp 154–163. https://aclanthology.org/2021.wanlp-1.16. Accessed 29 May 2023
64. Alduailaj AM, Belghith A (2023) Detecting Arabic cyberbullying tweets using machine learning. Mach Learn Knowl Extr 5(1):29–42. https://doi.org/10.3390/make5010003
65. Ghanem B, Karoui J, Benamara F, Rosso P, Moriceau V (2020) "Irony detection in a multilingual context," Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 12036:141–149. https://doi.org/10.1007/978-3-030-45442-5_18
66. Abdel-Salam R (2021) WANLP 2021 shared-task: towards irony and sentiment detection in Arabic tweets using multi-headed-LSTM-CNN-GRU and MaRBERT. In: Proceedings of the Sixth Arabic Natural Language Processing Workshop, pp 306–311. Available: https://www.aclweb.org/anthology/2021.wanlp-1.37
67. Alqmase M, Al H, Habib M (2021) Sports-fanaticism formalism for sentiment analysis in Arabic text. Soc Netw Anal Min 11:1–24. https://doi.org/10.1007/s13278-021-00757-9
68. Alharbi A, Lee M {2019} Crisis detection from Arabic tweets. In: Proceedings of the 3rd workshop on Arabic corpus linguistics, pp 72–79. Available https://www.aclweb.org/anthology/W19-5609
69. Alharbi A, Lee M (2021) Kawarith: an Arabic Twitter corpus for crisis events, pp 42–52 2021. https://aclanthology.org/2021.wanlp-1.5. Accessed 30 May 2023
70. Suleiman D, Awajan A, Al-Madi N (2017) "Deep learning based technique for plagiarism detection in Arabic texts," Proc - 2017 IntConf New Trends Comput Sci ICTCS 2017 2018:216–222. https://doi.org/10.1109/ICTCS.2017.42
71. Ghanem B, Arafeh L, Rosso P, Sánchez-Vega F (2018) HYPLAG: Hybrid arabic text plagiarism detection system. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 10859:315–323. https://doi.org/10.1007/978-3-319-91947-8_33/COVER
72. Mubarak H, Darwish K, Magdy W, Elsayed T, Al-Khalifa H (2020) Overview of OSACT4 Arabic offensive language detection shared task. In: Proceedings of the 4th workshop on open-source Arabic corpora and processing tools, with a shared task on offensive language detection. European Language Resource Association, Marseille, pp 48–52. https://www.aclweb.org/anthology/2020.osact-1.7
73. Otiefy Y, Abdelmalek A, El Hosary I (2020) "WOLI at SemEval-2020 Task 12: Arabic offensive language identification on different twitter datasets," 14th Int Work Semant Eval SemEval 2020 - co-located 28th Int Conf Comput Linguist COLING 2020 Proc 2237–2243. https://doi.org/10.18653/V1/2020.SEMEVAL-1.298
74. Faraj D, Faraj D, Abdullah M (2021) SarcasmDet at sarcasm detection task 2021 in Arabic using AraBERT pretrained model. In: Proceedings of the sixth Arabic natural language processing workshop, pp 345–350. https://aclanthology.org/2021.wanlp-1.44
75. Abu Farha I, Magdy W (2019) "Mazajak: An online Arabic sentiment analyser," 192–198. https://doi.org/10.18653/v1/w19-4621
76. Catelli R, Casola V, De Pietro G, Fujita H, Esposito M (2021) Combining contextualized word representation and sub-document level analysis through Bi-LSTM+CRF architecture for clinical de-identification. Knowledge-Based Syst 213:106649. https://doi.org/10.1016/J.KNOSYS.2020.106649
77. Raza S, (2021) "Automatic fake news detection in political platforms - A transformer-based approach," 68–78. https://doi.org/10.18653/v1/2021.case-1.10
78. Faris H, Aljarah I, Habib M, Castillo PA (2020) "Hate Speech Detection usingWord Embedding and Deep Learning in the Arabic Language Context," ICPRAM 2020 - Proc 9th Int Conf Pattern Recognit Appl Methods (March):453–460. https://doi.org/10.5220/0008954004530460
79. Lin Z et al (2021) Medical visual question answering: A survey. http://arxiv.org/abs/2111.10056
80. Hegazi MO, Al-Dossari Y, Al-Yahy A, Al-Sumari A, Hilal A (2021) Preprocessing Arabic text on social media. Heliyon 7(2):e06191. https://doi.org/10.1016/j.heliyon.2021.e06191
81. Namly D et al (2019) A bi-technical analysis for arabic stop-words detection. Compusoft 8(5):3126–3134
82. Saad M (2010) The impact of text preprocessing and term weighting on Arabic text classification. THESIS, p 112. https://api.semanticscholar.org/CorpusID:208123315
83. Ben Othman MT, Al-Hagery MA, El Hashemi YM (2020) Arabic text processing model: verbs roots and conjugation automation. IEEE Access 8:103913–103923. https://doi.org/10.1109/ACCESS.2020.2999259
84. Vaswani A et al (2017) Attention is all you need. Adv Neural Inf Process Syst 30. http://arxiv.org/abs/1706.03762. Accessed 29 May 2023
85. Saad M, Ashour W (2010)"OSAC: Open Source Arabic Corpora," 6th Int Conf Electr Comput Syst (EECS'10), Nov 25–26, 2010, Lefke, Cyprus (November):118–123. https://doi.org/10.13140/2.1.4664.9288
86. Albadi N, Kurdi M, Mishra S (2018) "Are they our brothers? analysis and detection of religious hate speech in the Arabic Twittersphere," Proc 2018 IEEE/ACM Int Conf Adv Soc Netw Anal Mining, ASONAM 2018 69–76. https://doi.org/10.1109/ASONAM.2018.8508247
87. Abdul-Mageed M, Elmadany AR, Nagoudi EMB (2021) "ARBERT & MARBERT: Deep bidirectional transformers for Arabic," ACL-IJCNLP 2021 - 59th Annu Meet AssocComput Linguist 11th Int Jt Conf Nat Lang Process Proc Conf (ii):7088–7105. https://doi.org/10.18653/v1/2021.acl-long.551
88. Ousidhoum N, Lin Z, Zhang H, Song Y, Yeung DY (2020) "Multilingual and multi-aspect hate speech analysis," EMNLP-IJCNLP 2019 - 2019 Conf Empir Methods Nat Lang Process 9th Int Jt Conf Nat Lang Process Proc Conf 4675–4684. https://doi.org/10.18653/v1/d19-1474
89. Mubarak H, Hassan S, Chowdhury SA, Alam F (2022) ArCovidVac: analyzing Arabic tweets about COVID-19 vaccination. arXiv no i. http://arxiv.org/abs/2201.06496. Accessed 1 May 2023
90. Habash NY (2010) Introduction to Arabic natural language processing. 3(1). https://doi.org/10.2200/S00277ED1V01Y201008HLT010
91. Farghaly A (2009) Arabic natural language processing: Challenges and solutions. ACMTrans Asian Lang Inf Process 8(May):1–10. https://doi.org/10.1145/1644879.1644881.http
92. Shaalan K, Siddiqui S, Alkhatib M, Abdel Monem A (2018) Challenges in Arabic natural language processing. In: Computational linguistics, speech and image processing for arabic language, world scientific, pp 59–83. https://doi.org/10.1142/9789813229396_0003
93. Albalawi RM, Jamal AT, Khadidos AO, Alhothali AM (2023) Multimodal Arabic rumors detection. IEEE Access 11(January):9716–9730. https://doi.org/10.1109/ACCESS.2023.3240373
94. Guru DS, Ali M, Suhil M (2018) "A Novel Term Weighting Scheme and an Approach for Classification of Agricultural Arabic Text Complaints," 2nd IEEE Int Work Arab DerivScr Anal Recognition ASAR 24–28. https://doi.org/10.1109/ASAR.2018.8480317
95. Raza S, Pour PO, Bashir SR (2023) Fairness in machine learning meets with equity in healthcare, pp 1–8. Available: http://arxiv.org/abs/2305.07041

Abdullah Yahya Mohammed Muaad received his B.SC degree in computer science from Hail University in 2008 and M.S.C from University of Mysore, Mysuru, Karnataka, India, in 2018. He submitted his Ph.D thesis to Department of Studies in Computer Science, University of Mysore. He received 5 awards during his study. He has authored and co-authored more than 30 journal and conference papers in well-reputed international journals. Also, he worked as a general registrar and lecturer in Sanaa Community College Yemen from 2010 to 2014. Also, he is a reviewer in many SCI journals. His main areas of interest and current research interests are natural language processing, text representation, image processing, cyber security, machine and deep learning. His recent AI-based publications have earned a lot of attention from researchers as well as international journal editorials.

Shaina Raza is a Responsible AI Scientist at the Vector Institute of Artificial Intelligence in Toronto, Canada. She works in the intersections of technology, ethics, and society. In the past, she was awarded the esteemed Canadian Institutes of Health Research (CIHR) Health System Impact (HSI) Fellowship, where she worked on health equity and introduced Fairness in Machine Learning (ML) in the health domain. Holding a Ph.D. in Computer Science, she specializes in Recommender Systems and Fairness in ML. Dr. Raza's diverse research interests cover a wide range of topics, from Information Systems and Fairness in LMs (LM) to Public Health Equity and Ethical AI. Over the years, she has been involved in several projects focusing on ML and LM. Her work particularly emphasizes fairness, debiasing, and their real-world applications in diverse sectors including news, public health, and finance.

Usman Naseem is a Lecturer at the College of the School of Science and Engineering, James Cook University, Australia. His research interest lies in Natural Language Processing (NLP) and Computational Social Science, focusing on understanding human communication in social contexts and developing socially aware language technologies. Prior to persuading his research, he worked in leading ICT companies for over 10 years. He publishes and serves as Program Committee, including the Area Chair for several top-tier venues, including ACL, EMNLP, NAACL, COLING, WSDM, Webconf, and AAAi. He also delivered several invited talks and a tutorial on recommender systems at Webconf 2023. His work in NLP has attracted attention from the World Health Organization and earned him the Nepean Blue Mountains Local Health District Board Chair's Quality Award in 2021 and the prestigious IEEE Best Transactions Paper Award in 2022.

Hanumanthappa. J is a Professor in the Department of Studies in Computer Science, at University of Mysore. He completed his Doctorate in Computer Science from Mangalore University in 2014. He is a reviewer and program committee member for many journals. He has vast experience of more than 20 years in teaching. He has published more than 100 papers in international SCI (Peer- Reviewed) Journals and Conferences, and 02 publications in books/book- chapters. He has also Published 02 Patents on “Root Cause Analysis, Threat Interpretation, And Network Survivability Prediction Device for Heterogeneous Networks(Patent Application No:202141000707, Advanced Social IoT for real-time monitoring of Diabetics Patient in Health Care System(Patent Application No:202141051287) on 07/01/2021 and 09/11/2021. His research interests include Performance Issues of Protocols, Wireless Networking, data mining, and machine learning.