applied sciences

MDPI

Review

Arabic Automatic Speech Recognition: A Systematic Literature Review

Amira Dhouib *, Achraf Othman *, Oussama El Ghoul, Mohamed Koutheair Khribi and Aisha Al Sinani

Mada Center, Doha P.O. Box 24230, Qatar

* Correspondence: adhouib@mada.org.qa (A.D.); aothman@mada.org.qa (A.O.)

Abstract: Automatic Speech Recognition (ASR), also known as Speech-To-Text (STT) or computer speech recognition, has been an active field of research recently. This study aims to chart this field by performing a Systematic Literature Review (SLR) to give insight into the ASR studies proposed, especially for the Arabic language. The purpose is to highlight the trends of research about Arabic ASR and guide researchers with the most significant studies published over ten years from 2011 to 2021. This SLR attempts to tackle seven specific research questions related to the toolkits used for developing and evaluating Arabic ASR, the supported type of the Arabic language, the used feature extraction/classification techniques, the type of speech recognition, the performance of Arabic ASR, the existing gaps facing researchers, along with some future research. Across five databases, 38 studies met our defined inclusion criteria. Our results showed different open-source toolkits to support Arabic speech recognition. The most prominent ones were KALDI, HTK, then CMU Sphinx toolkits. A total of 89.47% of the retained studies cover modern standard Arabic, whereas 26.32% of them were dedicated to different dialects of Arabic. MFCC and HMM were presented as the most used feature extraction and classification techniques, respectively: 63% of the papers were based on MFCC and 21% were based on HMM. The review also shows that the performance of Arabic ASR systems depends mainly on different criteria related to the availability of resources, the techniques used for acoustic modeling, and the used datasets.

Keywords: Arabic language processing; automatic speech recognition; Arabic Speech-To-Text; systematic literature review

check for updates

Citation: Dhouib, A.; Othman, A.; El Ghoul, O.; Khribi, M.K.; Al Sinani, A. Arabic Automatic Speech Recognition: A Systematic Literature Review. *Appl. Sci.* **2022**, *12*, 8898. https://doi.org/10.3390/app12178898

Academic Editor: Lijiang Chen

Received: 2 August 2022
Accepted: 31 August 2022
Published: 5 September 2022

1. Introduction

Automatic Speech Recognition (ASR) represents a particular case of digital signal processing, which comprises statistics, phonetics, linguistics, and machine learning. It can be defined as a technology by which the spoken words are converted into textual representation, using software to recognize human voice and speech [1,2]. Recently, automatic speech recognition systems have become the subject of increasing interest for diverse speech/language researchers and academics. This interest is reflected in their emergence in various areas, such as health, education, dictation, and robotics [3]. With the rapid progress of technologies, ASR systems are adopted by various applications due to their functionality and ease of use. For example, they are applied in dictation software, which can be a constructive PC tool for accessibility benefits. Another application is using voice control commands and search with mobile devices. They can also be associated with speech translation from a source to a target language. The ASR systems have been considered for a long time as a valuable and helpful input technique for a range of categories of disabilities since it is based on speech as an input technique alternating traditional manual techniques via keyboards and mouses [4]. However, automatic speech recognition is considered a challenging task in the signal processing field as it requires several layers of processing to reach a high level of accuracy and lower Word Error Rate (WER).

The Arabic language is considered one of the official languages in twenty-two countries situated in the Middle East, Africa, and the Gulf. It is ranked as the fifth most extensively used language worldwide [5] and is used by more than 422 million native and non-native people [6,7]. According to [8], the Arabic language can be classified according to three primary types, namely:

- Classical Arabic represents the most formal and standard form of Arabic as it is mainly used in the Holy Quran and the religious instructions of Islam [1];
- Modern Standard Arabic (MSA) represents the current formal linguistic standard of the Arabic language. It is generally used in written communication and media, and is taught in educational institutions [1];
- Dialectal Arabic (DA), also called colloquial Arabic, is a variation of the same language specific to countries or social groups used in everyday life. Various dialects of Arabic exist, and, sometimes, more than one DA can be used within a country [1].

Based on the study of Elnagar et al. [6], the DA can be categorized according to the following varieties:

1. North Africa, which includes the Tunisian, Algerian, Moroccan, Mauritanian, and Libyan dialects;
2. Gulf dialect, which includes Qatari, Kuwaiti, Saudi, Omani, Bahraini, and Emirati dialects;
3. Nile Basin, including the Egyptian and Sudanese dialects;
4. Yemeni dialect;
5. Iraqi dialect;
6. Levantine dialect is often used in Syria, Lebanon, Palestine, and western Jordan [9].

Compared to the existing research on ASR for the English language, the ASR for the Arabic language received little attention due to its consideration as a limited resource language [10]. The main challenges of the Arabic language remain specific to the existence of enormous dialects with various pronunciations, the morphological complexity, and the difficulty of acquiring a diacritized transcription of the speech corpora, which is not very commonly open-source, etc. [11]. Toward building a robust Arabic ASR system, it is highly recommended and more accurate to use extensive speech collections. According to [12], an extensive vocabulary means that the dataset contains approximately 20k to 60k words.

Several overviews and survey studies have been published to review various aspects of Arabic speech recognition. In 2018, the authors of [13] published a literature survey paper that discusses the Arabic ASR. The survey shows that few freely available continuous speech corpora exist. It also shows a need to compile large corpora. In another study, Algihab et al. [14] review the available studies on Arabic speech recognition along with the available services and toolkits for the development of Arabic speech recognition systems. The focus was on Arabic ASR using deep learning. Seventeen papers were reviewed and presented according to the recognized entity and learning techniques. A more recent study by Abdelhamid et al. [8] presents Arabic speech recognition systems from the end-to-end methodology perspective. The study focuses on two types of the Arabic language, namely MSA and dialectal Arabic. It presents the end-to-end Arabic speech recognition systems proposed between 2017 and 2019. It also presents the available API Services and toolkits essential for building end-to-end models. Another work that reports on the reviews of ASR systems for isolated Arabic words was proposed in 2021 by Shareef and Irhayim [15]. The authors focused on ASR systems based on artificial intelligence techniques and summarized 16 studies according to four criteria. These include speech recognition types, classification techniques, feature extraction techniques, and accuracy rates.

This paper is a follow-up of studies conducted about automatic speech recognition studies proposed for the Arabic language, where the need for broader research on this topic was recognized. The goal is to conduct a Systematic Literature Review (SLR) of Arabic automatic speech recognition to guide researchers by providing them with the most significant studies published recently. To the best of our knowledge, this is the first systematic review that presents the landscape of Arabic ASR studies. Our goal is to highlight the progress made in the Arabic ASR field over ten years, starting from 2011 till 2021. This systematic literature review will also guide speech and language researchers and academics to define the significant research gaps in the field and to open perspectives for future research.

The remaining paper is organized into five sections. Section 2 presents a brief background of Arabic ASR. Section 3 describes the adopted research method in this systematic literature review. The formulated primary and secondary research questions are answered in Section 4. The conclusions are presented in the last section.

2. Background

Automatic speech recognition concerns the automated conversion of speech or audio waves into texts exploitable by a machine through analyzing and processing speech signals using different techniques such as Convolutional Neural Network (CNN) [16] or deep learning [17]. The design of an ASR architecture system depends on various components and tasks like preprocessing, noise detection, speech classification, and feature extraction. Figure 1 presents a generic architecture used in the development of ASR systems. Three main modules can be identified in a traditional speech recognition system [4]. The first one corresponds to speech pre-processing, which aims to remove undesirable noises from the speech signal and identify speech activity [18]. The second module concerns feature extraction, in which essential data are extracted from a speech. The third module refers to the classification, which aims to find the parameter set from memory.

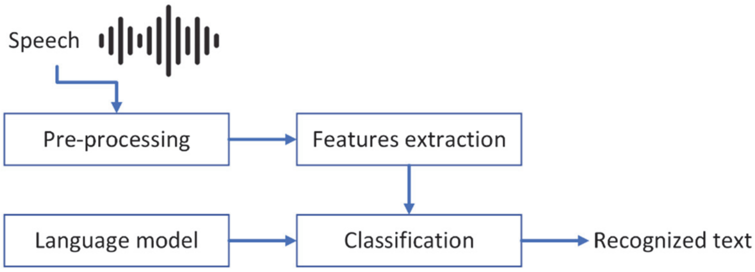


Figure 1. Automatic speech recognition architecture.

The critical challenge in developing highly accurate Arabic ASR systems is selecting feature extraction and classification techniques [19]. Mel Frequency Cepstral Coefficient (MFCC) and Perceptual Linear Predictive (PLP) are the most common techniques used for feature extraction. In the systematic literature review presented by Nassif et al. [20], for instance, 69.5% of the retained papers used the MFCC technique to extract features from speech. A wide range of techniques can also be used for classification. Examples of these techniques are Artificial Neutral Network (ANN), Hidden Markov Model (HMM), and Dynamic Time Warping (DTW).

The development of Arabic speech recognition systems has increased in the past decades thanks to the availability of different open sources and toolkits for building and assessing ASR. These toolkits are Hidden Markov Model Toolkit (HTK), Carnegie Mellon University (CMU) Sphinx engine, and KALDI Speech Recognition Toolkit. Speech recognition can be subdivided into four types [15], namely:

- Isolated word recognition in which speakers pause momentarily between every spoken word;
- Continuous speech recognition allows speakers to speak almost naturally, with little or no breaks between words. The systems related to the second type are more complex than isolated word recognition and need large volumes of data to achieve excellent recognition rates;
- Connected words allow a minimal pause between the isolated utterances to be used together;
- Spontaneous speech remains normal-sounding and not conversational speech.

Each category of speech recognition can be further categorized according to two sub-categories, namely:

1. Speaker-dependent, in which the system is based only on the speech of a specific speaker for which the system is trained [21];
2. Speaker-independent means speech recognition systems can be found in any speaker's speech [21].

As presented earlier, this paper aims to review and analyze the existing studies on automatic speech recognition for the Arabic language. Seven fundamental research questions are tackled to provide insight, for instance, into the used toolkits for Arabic ASR and the applied feature extraction and classification techniques. The following section presents these questions and details the adopted search method in this SLR.

3. Method

The systematic literature review was based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol [22]. First, the research questions are formulated, followed by the search strategy. Next, the inclusion and exclusion criteria are presented. Finally, the quality assessment and data extraction process are stated.

Figure 2 illustrates the PRISMA flow chart showing a report of the obtained outcomes in each phase for the current systematic literature review of Arabic Speech to Text.

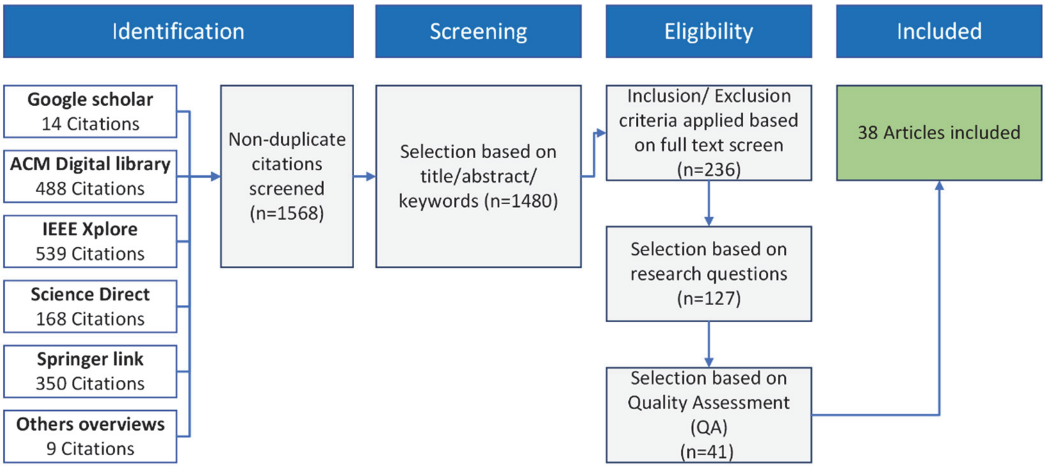

Identification
Screening
Eligibility
Included
Google scholar
14 Citations
ACM Digital library
488 Citations
IEEE Xplore
539 Citations
Science Direct
168 Citations
Springer link
350 Citations
Others overviews
9 Citations
Non-duplicate citations screened (n=1568)
Selection based on title/abstract/ keywords (n=1480)
Inclusion/ Exclusion criteria applied based on full text screen (n=236)
Selection based on research questions (n=127)
Selection based on Quality Assessment (QA) (n=41)
38 Articles included

Figure 2. PRISMA flow chart of the study selection.

3.1. Research Questions (RQ)

The first step of this SLR consists of defining the research questions and Secondary Research Questions (SRQ). As presented earlier, the purpose is to review the Arabic ASR studies conducted between 2011 and 2021. A total of seven RQs and three SRQs were defined to carry out a detailed review of the field. The RQs and SRQs related to the purposes are as follows:

RQ 1. What is the bibliographic information of the existing studies?

- SRQ 1.1. What are the most active countries?
- SRQ 1.2. How has the number of studies evolved across the years?
- SRQ 1.3. What are the types of venues (i.e., journals, conferences, or workshops) used by the authors of studies?

RQ 2. What is the considered variant of Arabic in speech recognition studies?
RQ 3. What are the toolkits most often used in the Arabic speech recognition field?
RQ 4. Which datasets were most often used, and types of Arabic speech recognition were identified in these datasets?
RQ 5. What are the used feature extraction and classification techniques for Arabic speech recognition studies?
RQ 6. What are the current gaps and future research in the Arabic ASR field?

RQ 7. What is the performance of Arabic recognition systems in terms of accuracy rate or WER?

3.2. Search Strategy

3.2.1. Search Strings

We started our SLR by defining the main keywords used in the related research studies and the research questions. To ensure a more comprehensive search, alternate synonyms, acronyms, and spelling variations of words were included for the different keywords. In the following, we divided the keywords into four categories. The Boolean operator OR was used by combining the keywords in each category. Then, we used the Boolean operator AND incorporated the keywords across the categories. Table 1 presents the defined categories along with the keywords.

Table 1. Search terms.

Category 1	Category 2	Category 3	Category 4
Arabic Arabic language Multilingual	Automat [1] Computer	Speech recogni [2] Speech trans [3] Speech to text Voice to text Voice recogniti [2] SRT [4] ASR STT	System Technology Tool

[1] Automated or automatic. [2] Recognition or recognizer. [3] Translator or transformation. [4] Speech recognition technology.

The search string induced based on the keywords in each category is as follows:

- C1: "Arabic" OR "Arabic Language" OR "Multilingual";
- C2: "Automat*" OR Computer";
- C3: "Speech recogni*" OR "Speech trans*" OR "Speech to text" OR "Voice to text" OR "Voice recogni*" OR "SRT" OR "ASR" OR "STT";
- C4: "System" OR "Tool" OR "Technology"

The resulting string can be formulated as (C1) AND (C2) AND (C3) AND (C4). The results of the search were then imported to Mendeley Reference Management Software.

3.2.2. Electronic Databases

Five electronic databases were used to collect data. These include ACM Digital Library, ScienceDirect, IEEE Xplore, SpringerLink, and Google Scholar. Table 2 presents the used electronic databases along with their links. In this study, the search was performed for published journal papers, conferences, and workshop proceedings.

Table 2. List of used online electronic databases.

Database Source	Link
Google Scholar	https://scholar.google.com/ (accessed on 1 August 2022)
ACM Digital Library	https://dl.acm.org (accessed on 1 August 2022)
IEEE Xplore	https://ieeexplore.ieee.org/Xplore/home.jsp (accessed on 1 August 2022)
Science Direct	https://www.sciencedirect.com (accessed on 1 August 2022)
Springer Link	https://link.springer.com/ (accessed on 1 August 2022)

Table 3 illustrates the procedure for conducting queries in each electronic database, along with some notes.

Table 3. Online electronic databases.

Databases	Query String	Notes
Google Scholar	allintitle: ("Arabic" OR "Multilingual") AND ("Speech recognition" OR "Speech recognizer" OR "Speech transformation" OR "Speech to text" OR "voice to text" OR "Voice recognition") AND ("system" OR "technology")	Custom range: 2011–2021 The number of characters per query is limited, so the shorter query was applied.
ACM Digital Library	("Arabic" OR "Arabic Language" OR "Multilingual") AND ("Automatic" OR "Automated" OR "Computer") AND ("Speech recognition" OR "Speech recognizer" OR "Speech transformation" OR "Speech to text" OR "voice to text" OR "Voice recognition" OR "Voice recognizer" OR "SRT" OR "ASR" OR "STT") AND ("system" OR "tool" OR "technology")	Filter year: 2011–2021 Document Type: research articles (295) and journals (193).
IEEE Xplore	("Arabic" OR "Arabic Language" OR "Multilingual") AND ("Automat*" OR "Computer") AND ("Speech recogni*" OR "Speech trans*" OR "Speech to text" OR "voice to text" OR "Voice recogni*" OR "SRT" OR "ASR" OR "STT") AND ("system" OR "tool" OR "technology")	Filter year: 2011–2021 Document type: Conferences (473) and Journals (66). IEEE Xplore recommends using short expressions. It does not work correctly with many disjunction terms.
Science Direct	("Arabic" OR "Multilingual") AND ("Speech recogni" OR "Speech trans" OR "Speech to text" OR "voice to text" OR "Voice recogni") AND ("system" OR "technology")	Filter by article types: "Research articles (168)." Filter by years: 2011–2021 The number of Boolean connectors is limited to 8 in the search. For this reason, we have used the shorter string The wildcard '*' is not supported
Springer Link	("Arabic" OR "Arabic Language" OR "Multilingual") AND ("Automat" OR "Computer") AND ("Speech recogni" OR "Speech trans" OR "Speech to text" OR "voice to text" OR "Voice recogni" OR "SRT" OR "ASR" OR "STT") AND ("system" OR "tool" OR "technology")	Two filters were applied to the result: Language is English Years: 2011–2021 Articles (189), conference papers (161)

3.3. Study Selection

A total of 1559 articles were retrieved from the search string presented in Table 3. We added 9 papers from other overviews and surveys. Out of 1568 papers, 88 were duplicates, and consequently, they were removed. A total of 1480 papers were retained after this step. Then, the abstract, keywords, and title of all the papers were checked by one author referred to as a reviewer. A total of 236 articles were retrieved after this step. In the following, a set of inclusion and exclusion criteria were applied to decide which research papers to review. Table 4 illustrates the adopted inclusion and exclusion criteria. By applying these criteria, the number of research papers was reduced further down to 127.

Table 4. Inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
Papers undertaking Arabic speech recognition.	Papers venue is not journals, conferences, or workshops.
Papers focused only on spoken words in Arabic.	Papers undertaking spoken Arabic digit recognition.
Papers directly answer one or more of the RQ.	Papers not being available.
Published papers are between 2011 and 2021.	Studies in duplicity.
Papers are written in English	Theoretical papers.

Next, two reviewers were involved in checking anonymously if the papers addressed one or more of the research questions presented previously in Section 3.1. In this way, the papers unable to cover the research questions related to this SLR would not be retained. In this step, the candidate papers were imported to a Web application called Rayyan [23], using the RIS format. This Web application supports the collaboration of the authors of systematic literature reviews by voting on papers based on the S/RQs criteria.

Three voting options can be used in the Rayyan application, namely: "exclude", "include", and "maybe".

- The papers with two "include" votes or one "maybe" and one "include" vote were retained.

- The papers that received two “exclude” votes or one “maybe” and one “exclude” vote were eliminated from the dataset.
- The papers that received two “maybe” votes or one “include” and one “exclude” vote were resolved through discussion. In those cases, a deciding vote on including or excluding the paper are made by the third reviewer.

Finally, the Quality Assessment (QA) of 41 candidate papers for inclusion was performed. The quality assessment of papers is described in the next section.

3.4. Quality Assessment

During this step, the quality assessment of each candidate’s paper for inclusion was carried out. The goal was to assess the quality and relevance of the papers’ contents. In this SLR paper, the quality assessment procedure was based on the study of [24]. The following quality assessment questions were proposed accordingly:

- QA 1. Are the research goals/aims clearly explained?
- QA 2. Does the technique/methodology in the research clearly describe?
- QA 3. Are all study questions answered?
- QA 4. Are the research findings reported?
- QA 5. Do the results of the study add to the speech recognition field?
- QA 6. Are the limitations of the current study adequately addressed?

Three-point scales were used to answer each QA question, namely “yes”, which was given 1 point, “partially”, which was given 0.5 points, and “no”, which was given 0 points. If the paper answers the QA question, it receives 1 point.

It scores 0.5 points if it partially addresses the QA question. In the case of a paper that did not address the QA question, it receives 0. The quality assessment of the research papers was performed by evaluating their quality against the QA questions. The answers to the QA questions of all the papers are presented in Table 5. The total score was calculated for each research study. A threshold was defined, such as if the total score was equal to or greater than three, then the study was included. In the case when the research study was less than three, then it was excluded. In this study, three papers were excluded. These papers are highlighted in blue color as we can see from Table 5. At the end of this process, the final number of studies to be retained was 38.

Table 5. Quality assessment of the candidate papers for inclusion and their characteristics.

Ref	Publication Title	Publication Year	Type of Venue	QA questions						Sum
				QA1	QA2	QA3	QA4	QA5	QA6	
[25]	A complete KALDI Recipe for building Arabic speech recognition systems	2014	Workshop	1	1	1	1	1	0	5
[26]	A comparative study of Arabic speech recognition system in noisy environments	2021	Journal	1	1	1	1	1	0	5
[27]	Effect of characteristics of speakers on MSA ASR performance	2013	Conference	1	1	1	1	0.5	0.5	5
[28]	An accurate HSMM-based system for Arabic phonemes recognition	2017	Conference	0.5	0.5	0.5	1	0.5	0	3
[29]	Active learning for accent adaptation in Automatic Speech Recognition	2012	Workshop	1	0	0.5	0.5	0.5	0	2.5
[30]	AALTO system for the 2017 Arabic multi-genre broadcast challenge	2017	Workshop	0.5	0.5	1	0.5	0.5	0	3
[31]	Arabic corpus Implementation: Application to Speech Recognition	2018	Conference	1	1	0.5	0.5	1	0	4
[32]	Arabic isolated word recognition system using hybrid feature extraction techniques and neural network	2017	Journal	1	1	1	1	1	0.5	5.5
[33]	Arabic Speech Recognition System Based on MFCC and HMMs	2020	Journal	1	0.5	1	0.5	0.5	0	3.5
[34]	Automatic speech recognition system for Tunisian dialect	2018	Journal	1	1	1	0.5	1	1	5.5
[35]	Arabic Speech Recognition by End-to-End, Modular Systems, and Human	2021	Journal	1	1	1	1	1	0.5	5.5
[4]	Constructing accurate and robust HMM/GMM models for an Arabic speech recognition system	2017	Journal	1	1	1	1	1	1	6
[36]	Development of the Arabic Loria Automatic Speech Recognition system (ALASR) and its evaluation of Algerian dialect	2017	Journal	1	1	1	0.5	1	0.5	5
[37]	Development of the MIT ASR system for the 2016 Arabic multi-genre broadcast challenge	2016	Workshop	1	0.5	1	1	0.5	0	4
[38]	Diacritics Effect on Arabic Speech Recognition	2019	Journal	1	1	1	1	1	0	5
[39]	Hybrid continuous speech recognition systems by HMM, MLP, and SVM: a comparative study	2014	Journal	1	1	1	1	1	0	5
[40]	Graphical Models for the Recognition of Arabic continuous speech based Triphones modeling	2015	Conference	1	0.5	1	1	1	0	4.5
[41]	Hybrid Arabic Speech Recognition System Using FFT, Fuzzy Logic, and Neural Network	2016	Journal	1	0.5	1	0.5	1	1	5
[42]	Investigating the impact of phonetic cross language modeling on Arabic and English speech recognition	2014	Conference	0.5	0	1	0.5	0.5	0	2.5
[43]	Lexicon Free Arabic Speech Recognition Recipe	2016	Conference	1	1	1	1	1	0	5
[44]	Arabic Speech Recognition Using MFCC Feature Extraction and ANN Classification	2017	Conference	1	0.5	1	0.5	0.5	0	3.5
[45]	Robust Front-End based on MVA processing for Arabic Speech Recognition	2017	Conference	0.5	0.5	1	0.5	0.5	0	3
[46]	Selection and combination of hypotheses for dialectal speech recognition	2016	Conference	1	0	0.5	0.5	0.5	0	2.5
[47]	Self-Supervised Speech Enhancement for Arabic Speech Recognition in Real-World Environments	2021	Journal	1	0	1	1	1	0	4
[1]	TAMEEM V1.0: speakers and text independent Arabic automatic continuous speech recognizer	2017	Journal	1	1	1	1	1	0.5	5.5
[48]	Multi-Dialect Arabic Speech Recognition	2020	Conference	1	0.5	1	0.5	1	0	4
[49]	Dynamic Time Warping Inside a Genetic Algorithm for Automatic Speech Recognition	2019	Conference	1	1	1	1	1	0	5
[50]	Control Interface of an Automatic Continuous Speech Recognition System in Standard Arabic Language	2020	Conference	1	1	1	0	0.5	0	3.5
[51]	The Effect of Diacritization on Arabic Speech Recognition	2017	Conference	1	0.5	1	1	1	0	4.5
[52]	Toward enhanced Arabic speech recognition using part of speech tagging	2011	Journal	1	0.5	1	1	1	0.5	5
[53]	Tunisian Dialectal End-to-end Speech Recognition based on Deep Speech	2021	Conference	1	1	1	0.5	1	0	4.5
[54]	The impact of phonological rules on Arabic speech recognition	2017	Journal	1	1	1	1	1	0.5	5.5

Table 5. *Cont.*

Ref	Publication Title	Publication Year	Type of Venue	QA questions						Sum
				QA1	QA2	QA3	QA4	QA5	QA6	
[55]	Arabic speech Recognition using end-to-end deep learning	2021	Journal	1	1	1	1	1	0	5
[56]	Convolutional Neural Network for Arabic Speech Recognition	2021	Journal	1	1	1	0.5	1	0	4.5
[57]	Automatic speech recognition of Arabic multi-genre broadcast media	2017	Conference	1	1	1	1	1	0.5	5.5
[58]	Bidirectional deep architecture for Arabic speech recognition	2019	Journal	1	1	1	1	1	0	5
[11]	WERD: using social text spelling variants for evaluating dialectal speech recognition	2017	Conference	1	0.5	1	1	1	0.5	5
[59]	LIUM ASR systems for the 2016 multi-genre Broadcast Arabic Challenge	2016	Workshop	1	1	0.5	1	0.5	0	4
[19]	Arabic Isolated Word Speaker Dependent Recognition System	2016	Journal	1	1	1	1	1	0	5
[60]	Speech Recognition of Isolated Arabic words via using Wavelet Transformation and Fuzzy Neural Network	2016	Journal	1	0.5	1	0.5	0.5	0	3.5
[61]	Development of a TV Broadcasts Speech Recognition System for Qatari Arabic	2014	Conference	1	1	1	1	1	0.5	5.5

4. Results and Discussion

The 38 papers related to Arabic speech recognition constitute the dataset that will be analyzed. Based on the research questions in Section 3.1, each study is analyzed using the following criteria: feature techniques, type of Arabic language, toolkit, and speech recognition. Table 6 presents the retained papers classified according to these criteria. The important legends for Table 6 are as follows:

MLLT: Maximum Likelihood Linear Transform.
PLP: Perceptual Linear Predictive.
MLP: Multi-Layer Perceptron.
GMM: Gaussian Mixture Model.
PCA: Principal Component Analysis.
RASTA-PLP: Relative Spectral-Perceptual Linear prediction.
fMMLR: Feature Space Maximum Likelihood Linear Regression.
bMMI: Boosted Maximum Mutual Information.
SGMM: Subspace Gaussian Mixture Model.
DNN: Deep Neural Networks.
HSMM: Hidden Semi-Markov Model.
MPE: Minimum Phone Error.
TDNN: Time-Delay Neural Network.
LSTM: Long Short-Term Memory.
BLSTM: Bidirectional Long Short-Term Memory.
FFBPNN: Feed-Forward Back Propagation Neural Network.
SVM: Support Vector Machine.
DBN: Dynamic Bayesian Networks.
BDRNN: Bidirectional Recurrent Neural Network.
CTC: Connectionist Temporal Classification.
GLSTM: Guiding Long-Short Term Memory.
CHMM: Coupled Hidden Markov Model.
RNN: Recurrent Neural Network.
LPC: Linear Predictive Coding.
MFCC: Mel Frequency Cepstral Coefficient.
PCA: Principal Component Analysis.
ANN: Artificial Neural Network.
SAT-GMM: Speaker Adaptive Training Gaussian Mixture Model-Gaussian Mixture Model.
FFT: Fast Fourier Transform.
MFSC: Log Frequency Spectral Coefficients.
GFCC: Gammatone-Frequency Cepstral Coefficients.
WLPCCS: Weighted Linear Prediction Cepstral Coefficients.
MAP: Maximum a Posterior Adaptation.
DTW: Dynamic Time Warping.
GMMD: Gaussian Mixture Model Derived.

Table 6. Summary of the retained papers and their features.

Ref	Feature Techniques		Type of Arabic Language	Toolkits	Type of Speech Recognition		Datasets
	Extraction Techniques	Classification Techniques			Mode	Speaker Dependency	
[25]	MFCC+LDA+MLLT	GMM, GMM-fMLLR, GMM-MPE, GMM-bMMI, SGMM-fMLLR, SGMM-bMMI, DNN, DNN-MPE	MSA	MADA, KALDI	Continuous speech	Speaker independent	A mix of 127 h of Broadcast Conversations (BC) and 76 h of Broadcast Report (BR).
[26]	MFCC	GMM-HMM, DNN-HMM	MSA	CMU Sphinx, KALDI	Isolated words	Speaker independent	A free Arabic Speech Corpus for Isolated Words dataset [62].
[27]	MFCC	HMM	MSA	HTK	continuous speech	Speaker independent	The Algerian Arabic Speech Database (ALGASD) [63].
[28]	MFCC	HSMM, GMM	MSA	HTK	Isolated words	N/A	Classical Arabic letters sound from Holy Quran.
[30]	LDA	TDNN, TDNN-LSTM, TDNN-BLSTM	MSA, Egyptian dialect	KALDI	Continuous words	Speaker independent	The 3rd Multi-Genre Broadcast challenge (MGB-3): an enormous audio corpus of primarily MSA speech and 5 h of Egyptian data [64].
[31]	MFCC, PLP, LPC	HMM	Tunisian dialect	HTK	Continuous and connected words	N/A	21 recordings of 5 words of vocabulary.
[32]	MFCC, PLP, PCA, RASTA-PLP	FFBPNN	MSA	N/A	Isolated words	Speaker dependent	11 modern standard Arabic words.
[33]	MFCC	HMM	MSA	MATLAB	Isolated words	N/A	24 Arabic words Consonant-Vowel [33].
[34]	PLP, LDA, MLLT, fMLLR	GMM-HMM	Tunisian dialect	KALDI	Continuous words	Speaker independent	The Tunisian Arabic Railway Interaction Corpus (TARIC). It contains information requests in the Tunisian dialect about railway services.
[35]	MFCC	TDNN-LSTM	MSA, Egyptian, Gulf, Levantine, North African	KALDI	Continuous words	Speaker independent	Arabic Multi-Genre Broadcast corpus: MGB0F [1] [65], MGB3 [64], and MGB5 [66].
[4]	MFCC	GMM-HMM	MSA	HTK, MATLAB	Isolated words	N/A	A speech database containing more than 8 h of speech recorded from the Holy Quran concerning Tajweed rules.
[36]	MFCC	DNN-HMM	MSA extended to the Algerian dialect	KALDI	Isolated words	Speaker independent	A corpus of Algerian dialect sentences extracted from the Parallel Arabic DIalect Corpus (PADIC) [67].

Table 6. *Cont.*

Ref	Feature Techniques		Type of Arabic Language	Toolkits	Type of Speech Recognition		Datasets
	Extraction Techniques	Classification Techniques			Mode	Speaker Dependency	
[37]	LDA, MFCC, MLLT, fMLLR	GMM, DNN, CNN, TDNN, H-LSTM, GLSTM	MSA	KALDI, CNTK, SRILM	Continuous speech	N/A	A 1.200-h speech corpus was made available for the 2016 Arabic Multi-genre Broadcast (MGB) Challenge [2].
[38]	MFCC+LDA+MLLT	GMM-SI, GMM SAT, GMM MPE, GMM MMI, SGMM, SGMM-bMMI, DNN, DNN-MPE	MSA	KALDI, SRILM	Isolated words	Speaker independent	The total length of the corpus is 23 h of 4754 sentences with 193 speakers.
[39]	MFCC	HMM, MLP-HMM, SVM-HMM	MSA	HTK	Continuous speech	N/A	ARABIC_DB for large vocabulary continuous Speech recognition [68]
[40]	MFCC	HMM, Hybrid SVM-HMM DBN, Hybrid SVM-DBN	MSA	HTK	Isolated words	N/A	ARABIC_DB [68].
[41]	FFT	Fuzzy logic, Neural network	MSA	MATLAB	Isolated words	N/A	2 Arabic words. (مرحبا, شكراً)
[43]	Filter bank, Neutral network	GMM-HMM, HMM-GMM-Tandem, BDRNN, CTC	MSA	HTK	Continuous speech	N/A	8-h Aljazeera broadcast, which was collected and transcribed by Qatar Computing Research Institute (QCRI) using an advanced transcription system [69].
[44]	MFCC	ANN	MSA	HTK	Isolated words	N/A	3 Arabic letters of sa (س), tsa (ث), sya (ش).
[45]	MFCC, RASTA-PLP, PNCC	HMM	MSA	HTK	Isolated words	Speaker dependent	4 isolated Arabic words.
[47]	DAE	HMM	MSA	CMU Sphinx	Isolated words	Speaker independent	A free Arabic mobile parallel multi-dialect speech corpus containing 15492 utterances from 12 speakers [70].
[1]	MFCC	CHMM	MSA	CMU Sphinx	Continuous recognition	Speaker independent	The corpus contains recordings of 415 Arabic sentences.
[48]	MFCC	CNN, RNN	MSA, Egyptian, Gulf	CMU Sphinx	Isolated words	N/A	A corpus covering multiple dialects and different accents.
[49]	MFCC, FFT, Filter Bank	DTW, Genetic algorithm	MSA	MATLAB	Isolated words	Speaker independent	Corpus A is composed of 30 words. Corpus B is recorded in a natural environment containing 33 words.
[50]	N/A	HMM	MSA	HTK	Continuous speech	Speaker independent	The ALGerian Arabic Speech Database (ALGASD) [63].
[51]	N/A	N/A	MSA	CMU Sphinx	Continuous speech	N/A	Broadcast news from As-Sabah TV.
[52]	N/A	N/A	MSA	CMU Sphinx	Continuous speech	N/A	A 5.4-h speech corpus of MSA.
[53]	MFCC	DNN, RNN	Dialectal Tunisian	KenLM	Continuous words	N/A	Tunisian dialect paired speech collection" TunSpeech" consisting of 11 h.
[54]	N/A	HMM	MSA	CMU Sphinx	Continuous speech	N/A	Broadcast news using MSA.

Table 6. *Cont.*

Ref	Feature Techniques		Type of Arabic Language	Toolkits	Type of Speech Recognition		Datasets
	Extraction Techniques	Classification Techniques			Mode	Speaker Dependency	
[55]	MFCC, FBANK	CNN-LSTM	MSA	KALDI	Isolated words	Speaker independent	51 thousand words that required seven hours of recording via a single young male speaker
[56]	MFSC, GFCC	CNN	MSA	Librosa, Spafe, Keras, Tensor-flow	Isolated Words	Speaker independent	9992 utterances of 20 words spoken by 50 native male Arabic speakers.
[57]	MFCC, LDA, MLLT, fMLLR	DNN, TDNN, LSTM, BLSTM	MSA, Egyptian, Gulf, Levantine, North African	CNTK, KALDI, SRILM	Continuous speech	Speaker independent	An Egyptian broadcast data.
[58]	Mel Frequency, Filter Bank	LSTM, MLP	MSA	N/A	Isolated words	Speaker independent	Arabic TV commands and digits.
[11]	MFCC	TDNN, LSTMs, BiLSTMs	Egyptian dialect	KALDI	Continuous words	N/A	2 h of Egyptian Arabic A broadcast news speech data.
[59]	(PLP, BN, GMMD) + (i-vectors, fMLLR, MAP)	DNN, SAT-GMM, TDNN	MSA	MADAMIRA, KALDI	Continuous speech	Speaker independent	1128 h of Arabic broadcast speech.
[19]	MFCC, LPC	DTW, GMM	MSA	Praat, MATLAB	Isolated word	Speaker dependent	40 Arabic words.
[60]	LPC, LPCS, WLPCCS	Neural Network, Neuro-Fuzzy Network	MSA	MATLAB	Isolated words	Speaker independent	10 isolated Arabic words.
[61]	fMLLR, LDA, MLLT	GMM-HMM	MSA, Qatari dialect	KALDI	Continuous speech	Speaker independent	Different TV series and talk show programs.

[1] The abbreviation that refers to the Multi-Genre Broadcast. [2] http://www.mgb-challenge.org/arabic.html (accessed on 1 August 2022).

4.1. RQ1: What Is the Bibliographic Information of the Existing Studies?

4.1.1. SRQ1.1. What Are the Most Active Countries?

Table 7 presents the number of papers related to MSA and DA contributed by each country of the first author. The analysis illustrates that the first authors of seven papers are from Tunisia, with four studies dedicated to MSA and three to the Tunisian dialect. Algeria is second with five research papers, while Qatar is third with four papers. Two countries, namely Morocco and Kuwait, had three studies each. The research focus in these two countries was specifically on MSA. France, Jordan, and Malaysia had two research studies each. Lastly, eleven countries (i.e., Libya, Palestine, Saudi Arabia, Finland, Spain, Indonesia, USA, UK, Yemen, Egypt, Iraq) had one study each. Based on our analysis, it is noticed that most of the studies related to DA focus on the Egyptian dialect (five out of ten studies explicitly dedicated to dialectal Arabic). It can be noted that the Egyptian dialect is considered the first-ranked DA in terms of the number of speakers in the Arab world.

Table 7. The most active countries according to the retained studies. Some studies support more than one type of Arabic. Then, the same study was repeated, which increased the number of studies related to MSA/DA.

Countries of the First Author	Studies Related to MSA	Studies Related to Dialect Arabic	Total Number of Studies
Morocco	3	0	3
Qatar	4	3 (Qatari, Egyptian, Gulf/Levantine, /North African/Egyptian)	4
Finland	1	1 (Egyptian)	1
Palestine	1	0	1
Tunisia	4	3 (Tunisian)	7
Libya	1	0	1
Saudi Arabia	1	0	1
France	2	1 (Algerian)	2
USA	1	1 (Egyptian/Gulf, Levantine/North African)	1
Kuwait	3	0	3
Jordan	2	0	2
Spain	1	0	1
Indonesia	1	0	1
Algeria	5	0	5
UK	1	1 (Egyptian/Gulf)	1
Yemen	1	0	1
Egypt	1	0	1
Iraq	1	0	1

4.1.2. SRQ1.2. How Has the Number of Papers Evolved?

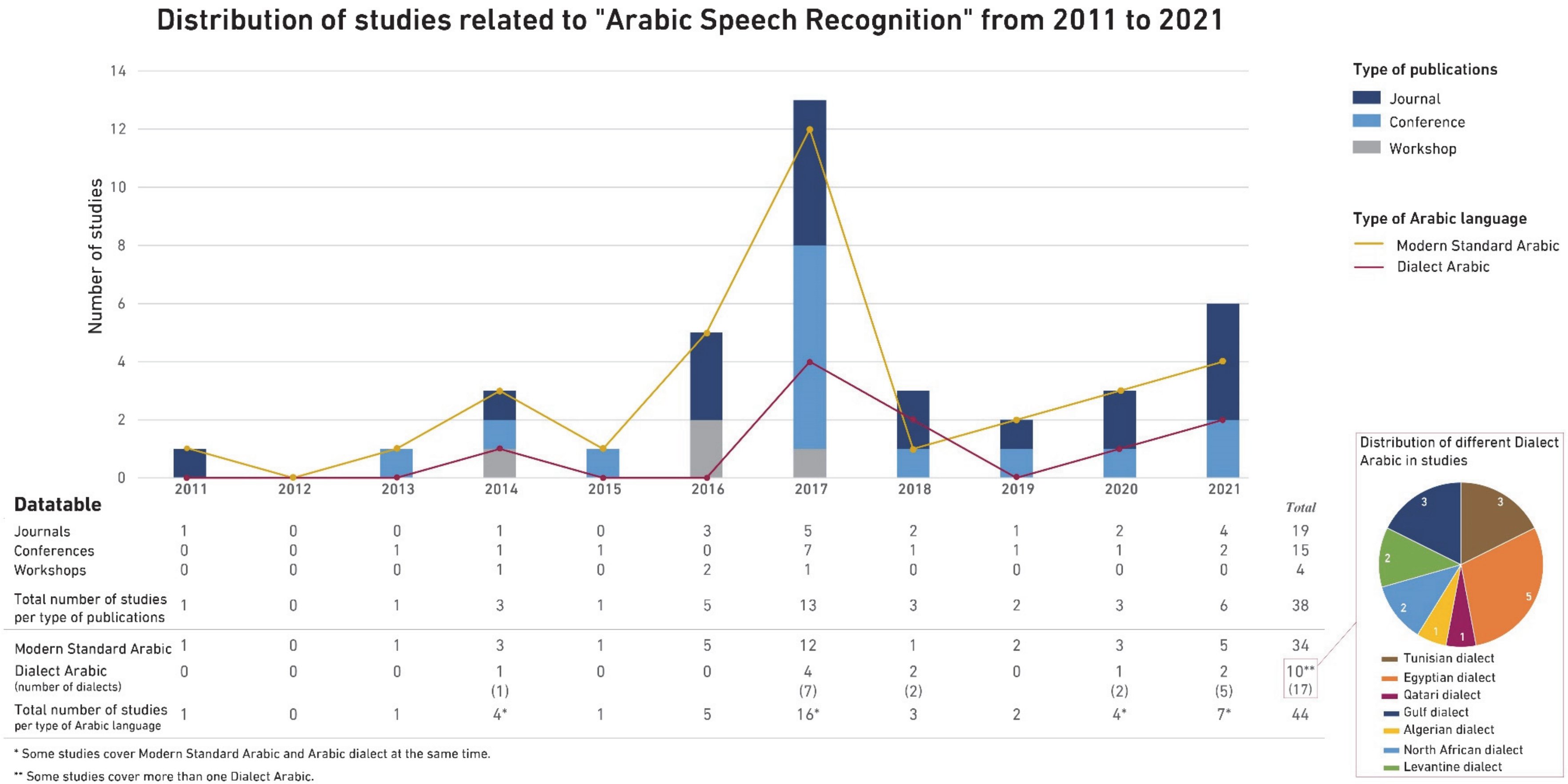
Figure 3 illustrates the distribution of Arabic speech recognition studies per publication years, types, and variants of Arabic language and dialects. The vertical columns present the number of published research papers per year. As we can see from the figure, the number of research studies on Arabic speech recognition systems has increased since 2016.

4.1.3. SRQ1.3. What Are the Types of Venues Used by the Authors of the Studies?

The total number of retained studies is 38. As we can see from Figure 3, a total of 19 studies were published in journals, 15 studies were presented at conferences, and 4 studies were presented in workshops.

4.2. RQ2. What Is the Considered Variant of Arabic in Speech Recognition Studies?

In this section, we describe the types or variants of Arabic adopted by the reviewed studies. As we can see from Figure 3, the recognition of MSA has piqued the interest of several researchers. Among the 38 reviewed studies, 34 of them cover MSA (89.47% of the retained studies), whereas 10 of the retained studies were dedicated to different dialects of Arabic (26.32% of the reviewed studies). Note that some studies can be dedicated to more than one variant of Arabic. The studies dedicated for MSA include, for example, [4,25–28,30,32,33,35,37–40,43]. Some of these studies cover MSA along with other dialectal Arabic, such as the Algerian dialect [36] (1 study out of 38), the Qatari dialect [61] (1 study out of 38), and the Egyptian dialect (4 studies out of 38) [30]. It can also be observed that some Arabic dialects were not supported by retained studies such as the Iraqi and Yemeni dialects.



Datatable

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	*Total*
Journals	1	0	0	1	0	3	5	2	1	2	4	19
Conferences	0	0	1	1	1	0	7	1	1	1	2	15
Workshops	0	0	0	1	0	2	1	0	0	0	0	4
Total number of studies per type of publications	1	0	1	3	1	5	13	3	2	3	6	38
Modern Standard Arabic	1	0	1	3	1	5	12	1	2	3	5	34
Dialect Arabic (number of dialects)	0	0	0	1 (1)	0	0	4 (7)	2 (2)	0	1 (2)	2 (5)	10** (17)
Total number of studies per type of Arabic language	1	0	1	4*	1	5	16*	3	2	4*	7*	44

* Some studies cover Modern Standard Arabic and Arabic dialect at the same time.

** Some studies cover more than one Dialect Arabic.

Figure 3. Distribution of Arabic speech recognition studies per publication years, types, and variants of Arabic language and dialects. Some studies support more than one variant of Arabic and can focus on multi-dialects of Arabic. Accordingly, the same study has been repeated, which increases the total number of studies per type of Arabic language.

4.3. RQ 3. What Are the Toolkits Most Often Used in the Arabic Speech Recognition Field?

Across the retained studies, researchers have used different toolkits to support implementing and evaluating Arabic speech recognition systems. Two studies have not mentioned the toolkit used [32,58]. Most of the retained studies have either used KALDI or CMU Sphinx, or HTK toolkits. The study of [1] used the CMU Sphinx toolkit to evaluate the speech corpus. The author claimed that there are many technical differences between CMU sphinx and HTK tools. These include (1) CMU Sphinx has more advanced features compared to HTK; (2) HTK toolkit is more user-friendly than CMU Sphinx; (3) CMU Sphinx is often better than HTK, mainly in terms of accuracy rate.

According to Abushariah [1], the CMU Sphinx toolkit is more suitable to be used, particularly for speaker-independent, extensive vocabulary, and continuous speech recognition systems. The CMU sphinx along with HTK toolkits are suitable to be used for training acoustic models due to their abilities to implement speaker-independent, large vocabulary, continuous speech recognition systems in any language [10,71,72]. However, it is noticed based on Table 6 that the CMU Sphinx and HTK tools were used to support isolated words in a few studies such as [26] and [40], respectively.

Other different toolkits such as SRILM language modeling were used in [37]. The previous study was based on the KALDI speech recognition toolkit to build and evaluate acoustic models and CNTK to train acoustic models. The KALDI toolkit was used as well in many other research studies such as [25,34–36,61]. Most of these studies were proposed to support dialectal Arabic. Some other research efforts such as [19,33,41,49,60] used MATLAB, a closed-source software. All these speech recognizers were dedicated mainly to MSA, particularly to the MSA isolated words category. Few of the retained studies have used other toolkits such as MADA [25] and KenLM [53].

Figure 4 illustrates the frequency of use of each toolkit according to the variant of Arabic.

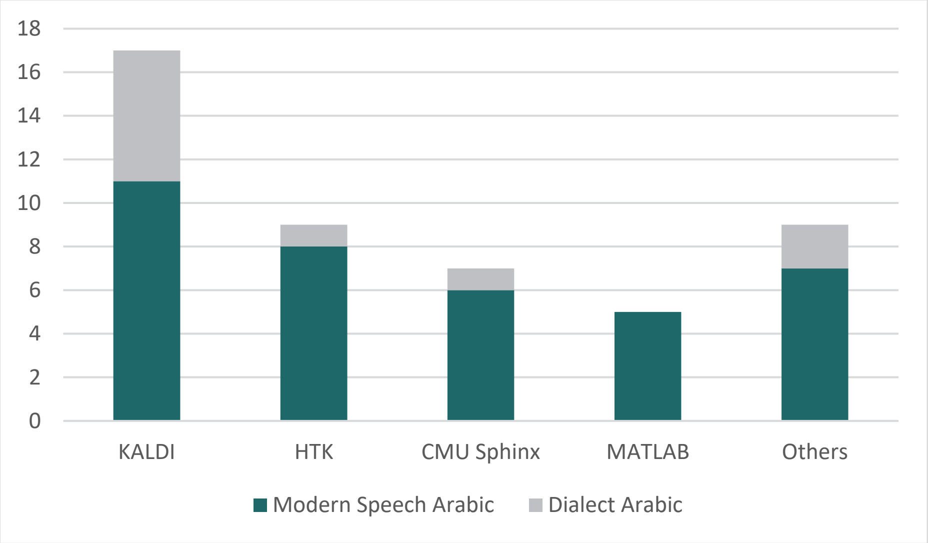


Figure 4. Frequency of use of toolkits according to the used type of Arabic. Some studies were based on more than one toolkit. Then, the same study was repeated, which increased the total number of studies.

4.4. RQ 4. Which Datasets Were Most Often Used, and What Types of Arabic Speech Recognition Were Identified in These Datasets?

As we can see from Table 6, most of the datasets were used only once. Examples of these datasets are TunSpeech and TARIC. A limited number of studies were adopted on the same dataset. For instance, two studies, namely [27,50], were based on the Algerian Arabic speech database [63]. The Arabic multi-genre broadcast datasets were also adopted by three studies, namely [30,35,37].

Among the retained studies, two main modes of speech recognition have been supported, namely isolated words and continuous words. The following subsections describe the existing studies according to each of these modes.

4.4.1. Isolated Words

Within the retained studies, 19 were dedicated to the isolated words category. For this category, studies including [19,26,32,33,40,41,45,47–49,58,60] were directed towards recognizing MSA. Some other studies for the isolated words category have contributed to MSA along with dialect Arabic such as [48], which focused on two variants of dialect Arabic (i.e., Egyptian and Gulf).

4.4.2. Continuous Words

Based on the data extracted from the 38 retained studies, all the research contributions for continuous words focused on recognizing speech from broadcast news or broadcasted reports, or conversations. These studies include, for example, [11,25,35,37,51,54,57,59,61] as shown in Table 6. Furthermore, the research contributions toward continuous Arabic words for speech-independent are more than what we have seen for speech-dependent, which makes sense as most of the continuous speech systems are dedicated to recognizing broadcast news/conversions. Among these studies, we can cite [25,59].

4.5. RQ 5. What Are the Used Feature Extraction and Classification Techniques for Arabic Speech Recognition Studies?

Results show that a wide range of feature techniques was used in Arabic speech recognition systems (see Table 6). These techniques can be classified into two categories: (1) feature extraction and (2) classification techniques. The following sub-sections present the retained studies according to these two categories.

4.5.1. Feature Extraction Techniques

The analysis of the retained studies shows that the MFCC acoustic feature extraction technique was the most used. A total of 24 studies out of 38 were based on the MFCC technique, which constitutes 63% of the reviewed papers. Other studies were based on alternate speech feature extraction methods, for instance, Linear prediction coefficient (LPC) (8% of the reviewed studies). Table 8 illustrates the commonly used feature extraction techniques, the percentage of their use, and the studies that were based on these techniques. As shown in this Table, 10% of the studies were based on the PLP technique, and 18% were based on the LDA technique. Some researchers have adopted a combination between MFCC and other feature extraction methods, such as [19,32], and they achieved better accuracy results than others.

Table 8. The commonly used extraction techniques. Some studies are based on more than one technique. Then, the same study was repeated, which increased the total number of studies.

Used Techniques	Percentage	References
fMLLR	13%	[34,37,38,57,59,61]
MFCC	63%	[1,4,11,19,25–28,31–33,35–40,44,45,48,49,53,55,57]
LDA	18%	[25,30,34,37,38,57,61]
LPC	8%	[19,31,60]
PLP	10%	[31,32,34,59]

4.5.2. Feature Classification Techniques

Table 9 shows the most common feature classification techniques used in the retained papers. Different studies adopt the HMM as a feature classification technique, which constituted 21% of the reviewed studies. Other research efforts have been based on a hybrid of the Hidden Markov Model (HMM) and Gaussian Mixture Model (GMM), especially for continuous speech recognition. In this case, the HMM was used to identify the temporal variability of speech while the Gaussian Mixture Model was applied to define how HMM states fit the frame acoustic input. As shown in Table 9, 13% of the retained studies were based on the combination of GMM-HMM.

Table 9. The commonly used classification techniques.

Used Techniques	Percentage	References
HMM	21%	[27,31,33,45,47,50,54]
GMM-HMM	13%	[4,26,34,43,61]

4.6. RQ 6. What Are the Current Gaps and Future Research in the Arabic ASR Field?

Despite the successful applications of Arabic speech recognition in different areas, there are still many gaps and limitations. One of the main limitations addressed in the studies focusing on DA is the minimal number of large datasets for dialect speech. It is known in the speech recognition community that preparing large training datasets for DA acoustic modeling is too tricky compared to MSA. It has been noticed that many of the retained studies were based on a small corpus. Using common datasets can be a cost-effective and possible way to gradually enhance the research area since the results can be compared and enhanced. Few speech datasets are publicly available for Arabic dialects, such as Spoken Tunisian Arabic Corpus (STAC) (It consists of five transcribed hours obtained from different Tunisian TV radios and channels TV) [73], MGB-3 (It emphasizes dialectal Arabic using a multi-genre collection of Egyptian YouTube videos), etc.

The adoption of a manual diacritized corpus and the lack of diacritized text remain significant limitations in many studies [38,51]. One adopted solution for the latter limitation consists of using grapheme units instead of the phoneme, which are the natural units of speech. Additionally, some software, such as Apptek, has been used for automatic diacritization of an Arabic corpus [10].

Aside from the limitations mentioned above, it should be noted that most of the selected studies focus on the non-diacritized Arabic scripts instead of a diacritized version. Only 6 studies out of 38 have been focused on diacritized Arabic speech [38,48,51,52,54,55]. A possible explanation is that the diacritized version of Arabic scripts may decrease the accuracy compared with the non-diacritized version, which prompts researchers to focus especially on non-diacritized scripts.

In terms of the techniques used for implementing speech recognition models, it was observed that the use of deep learning techniques (e.g., neural networks, recurrent neural networks, deep neural networks, etc.) is effective in increasing the accuracy of ASR. However, limited studies have been based on these techniques compared to the ones using alternative feature techniques. It is essential to highlight that using deep learning techniques requires many corpora to train a model. As already presented, a very limited common large corpus exists, which explains and is in line with the restricted number of studies using deep learning techniques. More practice-led research is needed to build more common large datasets for both MSA and dialectal Arabic.

Furthermore, many studies have claimed that the pronunciation variation phenomenon represents an additional challenge to ASR systems. A possible solution consists of adopting data extracted from pronunciation dictionaries to create rules, which enable the generation of pronunciation variants. Another possible solution is to estimate variants using speech data. The prominent approach that can be adopted for modeling pronunciation is the decision tree. The study of [74], for instance, was based on this approach to avoid over-generation pronunciation and pronunciation variant generation. According to Loots and Niesler [75], a decision tree is a practical approach to producing pronunciation variants by generalizing the pronunciation.

4.7. RQ 7. What Is the Performance of Arabic Speech Recognition Systems?

The performance of ASR is typically defined in terms of accuracy and speed. Accuracy is usually rated with the word error rate, whereas speed is estimated with the real-time factor. In this section, the analysis of studies is presented based on RQ4 and RQ5. Table 10 summarizes the performance of each of the retained Arabic ASR. It can be noted that excellent performances on the ASR proposed, for instance, by [41,60], have been achieved using deep learning techniques.

Table 10. Performance of the retained Arabic speech recognition systems.

References	Performance
[25]	The obtained WER scored 15.81% on BR and 32.21% on a broadcast conversation.
[26]	The best WER for clean data was 96.2%. It was obtained with 256 mixtures per state. For the noisy data test, the best WER was 49.2% average for SNR levels under babble noise obtained with 256 mixtures.
[27]	The WER was 9%.
[28]	The overall system accuracy was 98%, and it was enhanced by around 1% by implementing HSMM instead of standard HMM to reach 99%.
[30]	The obtained WER was 13.2% on the MGB2 test. The obtained WER was 37.5% on MGB3.
[31]	The obtained average accuracy rate was 91.56% by using MFCC. The obtained average accuracy rate was 95.34% by using PLP. The obtained average accuracy rate was 86.15% by using LPC.
[32]	The WER reached 0.26% when using a combination of RASTA-PLP, PCA, and FFBPNN techniques.
[33]	The obtained accuracy was 92.92%. The obtained WER was 7.08%.
[34]	The obtained WER was 22.6% on the test set.
[35]	The ASR achieved 12.5% WER on MGB2. It achieved 27.5% WER on MGB3 and 33.8% WER on MGB5.
[4]	The recognition rate reached 95.5% for system 1. The recognition rate reached 94% for system 2. The recognition rate reached 97% for system 3.
[36]	The WER was 14.02% for MSA. The WER was 89% for the Algerian dialect.
[37]	The overall WER was 18.3%.
[38]	The WER scored 4.68%. Adding diacritics increased WER by 0.59.
[39]	The lowest average of WER was 11.42% for SVM/HMM, 11.92% for MLP/HMM, and 13.42% for HMM standards.
[40]	The use of HMM led to a recognition rate of 74.18%. The hybridization of MLP with HMM led to a recognition rate of 77.74%. The combination of SVM with HMM led to a recognition rate of 78.06%. The hybridization of SVM with DBN realized the best performance, which was 87.6%.
[41]	The achieved accuracy was 98%.
[43]	The obtained WER ranged between 3.9% and 55%.
[44]	The average accuracy is 92.42%, with recognition accuracy of each letter (sa (س), sya (ش), tsa (ث)) at 92.38%, 93.26%, and 91.63%, respectively.
[45]	The obtained accuracy ranged between 97.26% and 98.95%.
[47]	The WER ranged between 30.17% and 34.65%.
[1]	The average WER was 2.22% for speakers independent of the text-dependent data set. The achieved average WER was 7.82% for speakers independent with text-independent data set.
[48]	The system has achieved an overall WER of 14%.
[49]	The best recognition rate given by the system is 79% for multi-speaker recognition and 65% for independent speaker recognition.
[50]	The performance of the ACSRS setup gave the region of Algiers a recognition rate of 97.74% for words and 94.67% for sentences.

Table 10. *Cont.*

References	Performance
[51]	The WER scored 76.4% for the non-diacritized text system and 63.8% for the diacritized text-based system.
[52]	The obtained accuracy was 90.18%.
[53]	The best performance achieved was 24.4% of WER.
[54]	The experimental results show that the non-diacritized text system scored 81.2%, while the diacritized text-based system scored 69.1%.
[55]	The achieved result is 31.10% WER using the standard CTC-attention method. For CNN-LSTM with the attention method, the best result is obtained from this model: 28.48% as WER.
[56]	Accuracy when using GFCC with CNN was 99.77%. The maximum accuracy obtained when using GFCC with CNN was 99.77%.
[57]	The best WER obtained on MGB-3 using a 4-g re-scoring strategy is 42.25% for a BLSTM system, compared to 65.44% for a DNN system.
[58]	For TV commands, accuracy is over 95% for all models.
[11]	The WER is much higher on the dialectal Arabic dataset, ranging from 40% to 50%. The WER for the proposed ASR system using all five references achieved a score of 25.3%.
[59]	The final system was a combination of five systems where the result obtained succeeded the best single LIUM ASR system with a 9% of WER reduction and succeeded the baseline MGB system that the organizers provided with a 43% WER reduction.
[19]	The system accuracy reached 94.56%.
[60]	The recognition rate of trained data reached 97.8%. The recognition rate of non-trained data reached 81.1%.
[61]	The proposed ASR achieved a 28.9% relative reduction in WER.

In terms of the used datasets, it was found that the larger the speech corpus size used to train the recognizer, the better the accuracy and the lower the WER. In [38], for example, the authors claim that the WER continuously decreases as the corpus size increases. As presented in Section 4.6, the speech recognition system using the non-diacritics dataset can achieve better performance than the non-diacritics version.

In [38], Abed et al. examined the effect of diacritization on Arabic ASR. The authors used diacritized and non-diacritized versions and checked how diacritics could impact the word error rate. In all their results (except a few models with few corpora), the diacritics increased WER for the used models. Additionally, in [51], the experimental results show that the word error rate scored 76.4% for the non-diacritized text system, while it scored 63.8% for the diacritized text-based system. It can also be noticed that better accuracy is achieved with speaker-dependent systems compared to speaker-independent systems since the former is adapted to an individual user. According to [38], speaker-independent systems might struggle when a new user uses the ASR system. In [1], for instance, Abushariah conducted two experiments with and without an adaptation to the speakers using different sentences. In their results, the obtained average WER was 7.64% for speaker-dependent, whereas the average WER for speaker-independent was 7.82%.

To summarize, the overall performance of ASR systems depends significantly on different factors, mainly the used datasets, the techniques for acoustic modeling, and the type of speech recognition. Accordingly, building a precise acoustic model using large datasets can be considered the key to suitable recognition performance.

5. Conclusions

This paper aims to compile the existing scientific knowledge about Arabic ASR studies published between 2011 and 2021. For that, a systematic review of the literature is conducted. A total of 38 conferences, workshops, and articles papers were reviewed from five academic databases: Google Scholar, IEEE Xplore, Science Direct, ACM Digital Library, and Springer Link. Our results and discussion revolve around seven fundamental research questions. The purposes were to provide insight into the used toolkits for implementing and evaluating Arabic ASR, the variants of the Arabic language, the used feature extraction and classification techniques, the performance of Arabic ASR systems, the type of speech recognition, the existing gaps, and future research in the Arabic ASR field.

Our findings illustrate that this is still an emerging research area where the number of studies has increased over the years. Many studies focus on MSA, whereas a relatively limited number of papers concentrate on dialectal Arabic. Some of these papers were dedicated to more than one variant of Arabic, and the reviewed studies did not support different dialects of the Arabic language. It would be interesting to focus on DA and use common datasets for dialect speech to enhance this research area gradually. Many toolkits were used to build and assess Arabic ASR. The most prominent ones were KALDI, HTK, and CMU Sphinx open-source toolkits. Concerning the used feature extraction techniques, MFCC was the most used one, followed by LDA, then PLP, and LPC. The results show also that HMM was the most adopted classification technique in the reviewed studies. Different limitations have also been addressed in the reviewed studies. The pronunciation variation phenomenon and the low availability of common large diacritized text for the Arabic language can be considered significant challenges that might limit research in this field. It would be interesting then to focus on the non-diacritized Arabic scripts and to develop more large and common datasets for both MSA and dialectal Arabic.

Author Contributions: Conceptualization, A.D. and A.O.; methodology, A.D. and A.O.; software, A.D., A.O. and O.E.G.; validation, A.D., A.O. and O.E.G.; formal analysis, A.D.; investigation, A.D.; resources, A.D.; data curation, A.D.; writing—original draft preparation, A.D.; writing—review and editing, A.D., A.O., O.E.G., M.K.K. and A.A.S.; visualization, A.O.; supervision, A.O.; project administration, A.O.; funding acquisition, A.O. and A.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This article was funded by the Mada Innovation Program (MIP'2021–2022) at Mada Center, Qatar.

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

References

1. Abushariah, M.A.M. TAMEEM V1.0: Speakers and Text Independent Arabic Automatic Continuous Speech Recognizer. *Int. J. Speech Technol.* **2017**, *20*, 261–280. [CrossRef]
2. Sen, S.; Dutta, A.; Dey, N. *Audio Processing and Speech Recognition: Concepts, Techniques and Research Overviews*; SpringerBriefs in Applied Sciences and Technology; Springer: Singapore, 2019; ISBN 9789811360978.
3. Jaber, H.Q.; Abdulbaqi, H.A. Real Time Arabic Speech Recognition Based on Convolution Neural Network. *J. Inf. Optim. Sci.* **2021**, *42*, 1657–1663. [CrossRef]
4. Khelifa, M.O.M.; Elhadj, Y.M.; Abdellah, Y.; Belkasmi, M. Constructing Accurate and Robust HMM/GMM Models for an Arabic Speech Recognition System. *Int. J. Speech Technol.* **2017**, *20*, 937–949. [CrossRef]
5. Al-Anzi, F.S.; AbuZeina, D. Synopsis on Arabic Speech Recognition. *Ain Shams Eng. J.* **2021**, *13*, 101534. [CrossRef]
6. Elnagar, A.; Yagi, S.M.; Nassif, A.B.; Shahin, I.; Salloum, S.A. Systematic Literature Review of Dialectal Arabic: Identification and Detection. *IEEE Access* **2021**, *9*, 31010–31042. [CrossRef]
7. Mubarak, H.; Darwish, K. Using Twitter to Collect a Multi-Dialectal Corpus of Arabic. In Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP), Doha, Qatar, 25 October 2014; Association for Computational Linguistics: Stroudsburg, PA, USA, 2014; pp. 1–7.
8. Abdelhamid, A.; Alsayadi, H.; Hegazy, I.; Fayed, Z. End-to-End Arabic Speech Recognition: A Review. In Proceedings of the 19th Conference of Language Engineering (ESOLEC'19), Alexandria, Egypt, 28 September 2020; pp. 26–30.
9. Abuata, B.; Al-Omari, A. A Rule-Based Stemmer for Arabic Gulf Dialect. *J. King Saud Univ. Comput. Inf. Sci.* **2015**, *27*, 104–112. [CrossRef]
10. Abushariah, M.; Ainon, R.; Zainuddin, R.; Elshafei, M.; Khalifa, O. Arabic Speaker-Independent Continuous Automatic Speech Recognition Based on a Phonetically Rich and Balanced Speech Corpus. *Int. Arab. J. Inf. Technol.* **2012**, *9*, 84–93.
11. Ali, A.; Nakov, P.; Bell, P.; Renals, S. WERD: Using Social Text Spelling Variants for Evaluating Dialectal Speech Recognition. In Proceedings of the 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Okinawa, Japan, 16–20 December 2017; pp. 141–148.
12. Jurafsky, D. *Speech & Language Processing*; Pearson Education: Delhi, India, 2000.
13. Al-Anzi, F.; AbuZeina, D. Literature Survey of Arabic Speech Recognition. In Proceedings of the 2018 International Conference on Computing Sciences and Engineering (ICCSE), Kuwait, Kuwait, 11–13 March 2018; pp. 1–6.

14. Algihab, W.; Alawwad, N.; Aldawish, A.; AlHumoud, S. Arabic Speech Recognition with Deep Learning: A Review. In *Social Computing and Social Media. Design, Human Behavior and Analytics, Proceedings of the International Conference on Human-Computer Interaction, Orlando, FL, USA, 26–31 July 2019*; Meiselwitz, G., Ed.; Springer International Publishing: Cham, Switzerland, 2019; pp. 15–31.
15. Shareef, S.R.; Irhayim, Y.F. A Review: Isolated Arabic Words Recognition Using Artificial Intelligent Techniques. *J. Phys. Conf. Ser.* **2021**, *1897*, 012026. [CrossRef]
16. Sitaula, C.; He, J.; Priyadarshi, A.; Tracy, M.; Kavehei, O.; Hinder, M.; Withana, A.; McEwan, A.; Marzbanrad, F. Neonatal Bowel Sound Detection Using Convolutional Neural Network and Laplace Hidden Semi-Markov Model. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2022**, *30*, 1853–1864. [CrossRef]
17. Subramanian, A.S.; Weng, C.; Watanabe, S.; Yu, M.; Yu, D. Deep Learning Based Multi-Source Localization with Source Splitting and Its Effectiveness in Multi-Talker Speech Recognition. *Comput. Speech Lang.* **2022**, *75*, 101360. [CrossRef]
18. Labied, M.; Belangour, A.; Banane, M.; Erraissi, A. An Overview of Automatic Speech Recognition Preprocessing Techniques. In Proceedings of the 2022 International Conference on Decision Aid Sciences and Applications (DASA), Chiangrai, Thailand, 23–25 March 2022; pp. 804–809.
19. Kourd, A.; Kourd, K. Arabic Isolated Word Speaker Dependent Recognition System. *Br. J. Math. Comput. Sci.* **2016**, *14*, 1–15. [CrossRef]
20. Nassif, A.B.; Shahin, I.; Attili, I.; Azzeh, M.; Shaalan, K. Speech Recognition Using Deep Neural Networks: A Systematic Review. *IEEE Access* **2019**, *7*, 19143–19165. [CrossRef]
21. Bhardwaj, V.; Ben Othman, M.T.; Kukreja, V.; Belkhier, Y.; Bajaj, M.; Goud, B.S.; Ur Rehman, A.; Shafiq, M.; Hamam, H. Automatic Speech Recognition (ASR) Systems for Children_ A Systematic Literature Review. *Appl. Sci.* **2022**, *12*, 4419. [CrossRef]
22. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; for the PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *BMJ* **2009**, *339*, b2535. [CrossRef] [PubMed]
23. Rayyan Systems Inc. Available online: https://www.rayyan.ai/ (accessed on 1 August 2022).
24. Kitchenham, B.; Stuart, C. Guidelines for Performing Systematic Literature Reviews in Software Engineering, Version 2.3. EBSE Technical Report. EBSE-2007-01. Available online: http://www.elsevier.com/framework_products/promis_misc/525444systematicreviewsguide.pdf (accessed on 1 August 2022).
25. Ali, A.; Zhang, Y.; Cardinal, P.; Dahak, N.; Vogel, S.; Glass, J. A Complete KALDI Recipe for Building Arabic Speech Recognition Systems. In Proceedings of the 2014 IEEE Spoken Language Technology Workshop (SLT), South Lake Tahoe, NV, USA, 7–10 December 2014; pp. 525–529.
26. Ouisaadane, A.; Safi, S. A Comparative Study for Arabic Speech Recognition System in Noisy Environments. *Int. J. Speech Technol.* **2021**, *24*, 761–770. [CrossRef]
27. Droua-Hamdani, G.; Sellouani, S.-A.; Boudraa, M. Effect of Characteristics of Speakers on MSA ASR Performance. In Proceedings of the 2013 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA), Sharjah, United Arab Emirates, 12–14 February 2013.
28. Khelifa, M.O.M.; Belkasmi, M.; Abdellah, Y.; ElHadj, Y.O.M. An Accurate HSMM-Based System for Arabic Phonemes Recognition. In Proceedings of the 2017 Ninth International Conference on Advanced Computational Intelligence (ICACI), Doha, Qatar, 4–6 February 2017; pp. 211–216.
29. Nallasamy, U.; Metze, F.; Schultz, T. Active Learning for Accent Adaptation in Automatic Speech Recognition. In Proceedings of the 2012 IEEE Spoken Language Technology Workshop (SLT), Miami, FL, USA, 2–5 December 2012; pp. 360–365.
30. Smit, P.; Gangireddy, S.R.; Enarvi, S.; Virpioja, S.; Kurimo, M. Aalto System for the 2017 Arabic Multi-Genre Broadcast Challenge. In Proceedings of the 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Okinawa, Japan, 16–20 December 2017; pp. 338–345.
31. Helali, W.; Hajaiej, Z.; Cherif, A. Arabic Corpus Implementation: Application to Speech Recognition. In Proceedings of the 2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET), Hammamet, Tunisia, 22–25 March 2018; pp. 50–53.
32. Boussaid, L.; Hassine, M. Arabic Isolated Word Recognition System Using Hybrid Feature Extraction Techniques and Neural Network. *Int. J. Speech Technol.* **2018**, *21*, 29–37. [CrossRef]
33. Elharati, H.A.; Alshaari, M.; Këpuska, V.Z. Arabic Speech Recognition System Based on MFCC and HMMs. *J. Comput. Commun.* **2020**, *8*, 28–34. [CrossRef]
34. Masmoudi, A.; Bougares, F.; Ellouze, M.; Estève, Y.; Belguith, L. Automatic Speech Recognition System for Tunisian Dialect. *Lang. Res. Eval.* **2018**, *52*, 249–267. [CrossRef]
35. Hussein, A.; Watanabe, S.; Ali, A. Arabic Speech Recognition by End-to-End, Modular Systems and Human. *Comput. Speech Lang.* **2021**, *71*, 101272. [CrossRef]
36. Menacer, M.A.; Mella, O.; Fohr, D.; Jouvet, D.; Langlois, D.; Smaïli, K. Development of the Arabic Loria Automatic Speech Recognition System (ALASR) and Its Evaluation for Algerian Dialect. *Procedia Comput. Sci.* **2017**, *117*, 81–88. [CrossRef]
37. AlHanai, T.; Hsu, W.-N.; Glass, J. Development of the MIT ASR System for the 2016 Arabic Multi-Genre Broadcast Challenge. In Proceedings of the 2016 IEEE Spoken Language Technology Workshop (SLT), San Diego, CA, USA, 13–16 December 2016; pp. 299–304.
38. Abed, S.; Alshayeji, M.; Sultan, S. Diacritics Effect on Arabic Speech Recognition. *Arab. J. Sci. Eng.* **2019**, *44*, 9043–9056. [CrossRef]

39. Zarrouk, E.; Ben Ayed, Y.; Gargouri, F. Hybrid Continuous Speech Recognition Systems by HMM, MLP and SVM: A Comparative Study. *Int. J. Speech Technol.* **2014**, *17*, 223–233. [CrossRef]
40. Zarrouk, E.; Benayed, Y.; Gargouri, F. Graphical Models for the Recognition of Arabic Continuous Speech Based Triphones Modeling. In Proceedings of the 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Takamatsu, Japan, 1–3 June 2015.
41. Hamdan, S.; Shaout, A. Hybrid Arabic Speech Recognition System Using FFT, Fuzzy Logic and Neural Network. *IRACST Int. J. Comput. Sci. Inf. Technol. Secur.* **2016**, *6*, 4–10.
42. Alotaibi, Y.A.; Meftah, A.H.; Selouani, S.-A. Investigating the Impact of Phonetic Cross Language Modeling on Arabic and English Speech Recognition. In Proceedings of the 2014 9th International Symposium on Communication Systems, Networks Digital Sign (CSNDSP), Manchester, UK, 23–25 July 2014; pp. 585–590.
43. Ahmed, A.; Hifny, Y.; Shaalan, K.; Toral, S. Lexicon Free Arabic Speech Recognition Recipe. In Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016, Cairo, Egypt, 24–26 November 2016; Hassanien, A.E., Shaalan, K., Gaber, T., Azar, A.T., Tolba, M.F., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 147–159.
44. Wahyuni, E.S. Arabic Speech Recognition Using MFCC Feature Extraction and ANN Classification. In Proceedings of the 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, 1–2 November 2017; pp. 22–25.
45. Techini, E.; Sakka, Z.; Bouhlel, M. Robust Front-End Based on MVA and HEQ Post-Processing for Arabic Speech Recognition Using Hidden Markov Model Toolkit (HTK). In Proceedings of the 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), Hammamet, Tunisia, 30 October–3 November 2017; pp. 815–820.
46. Soto, V.; Siohan, O.; Elfeky, M.; Moreno, P. Selection and Combination of Hypotheses for Dialectal Speech Recognition. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 5845–5849.
47. Dendani, B.; Bahi, H.; Sari, T. Self-Supervised Speech Enhancement for Arabic Speech Recognition in Real-World Environments. *Trait. Signal.* **2021**, *38*, 349–358. [CrossRef]
48. Ali, A.R. Multi-Dialect Arabic Speech Recognition. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020.
49. Maouche, F.; Benmohammed, M. Dynamic Time Warping Inside a Genetic Algorithm for Automatic Speech Recognition. In Proceedings of the International Symposium on Modelling and Implementation of Complex Systems, Laghouat, Algeria, 16–18 December 2018; Chikhi, S., Amine, A., Chaoui, A., Saidouni, D.E., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 180–192.
50. Zaidi, B.F.; Boudraa, M.; Selouani, S.-A.; Sidi Yakoub, M.; Hamdani, G. Control Interface of an Automatic Continuous Speech Recognition System in Standard Arabic Language. In Proceedings of the 2020 SAI Intelligent Systems Conference, London, UK, 3–4 September 2020; Arai, K., Kapoor, S., Bhatia, R., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 295–303.
51. Al-Anzi, F.S.; AbuZeina, D. The Effect of Diacritization on Arabic Speech Recogntion. In Proceedings of the 2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), Aqaba, Jordan, 11–13 October 2017.
52. AbuZeina, D.; Al-Khatib, W.; Elshafei, M.; Al-Muhtaseb, H. Toward Enhanced Arabic Speech Recognition Using Part of Speech Tagging. *Int. J. Speech Technol.* **2011**, *14*, 419–426. [CrossRef]
53. Messaoudi, A.; Haddad, H.; Fourati, C.; Hmida, M.B.; Elhaj Mabrouk, A.B.; Graiet, M. Tunisian Dialectal End-to-End Speech Recognition Based on DeepSpeech. *Procedia Comput. Sci.* **2021**, *189*, 183–190. [CrossRef]
54. Al-Anzi, F.S.; AbuZeina, D. The Impact of Phonological Rules on Arabic Speech Recognition. *Int. J. Speech Technol.* **2017**, *20*, 715–723. [CrossRef]
55. Alsayadi, H.A.; Abdelhamid, A.A.; Hegazy, I.; Fayed, Z.T. Arabic Speech Recognition Using End-to-end Deep Learning. *IFT Signal Process.* **2021**, *15*, 521–534. [CrossRef]
56. Abdelmaksoud, E.; Hassen, A.; Hassan, N.; Hesham, M. Convolutional Neural Network for Arabic Speech Recognition. *Egypt. J. Lang. Eng.* **2021**, *8*, 27–38. [CrossRef]
57. Najafian, M.; Hsu, W.-N.; Ali, A.; Glass, J. Automatic Speech Recognition of Arabic Multi-Genre Broadcast Media. In Proceedings of the 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Okinawa, Japan, 16–20 December 2017; pp. 353–359.
58. Zerari, N.; Abdelhamid, S.; Bouzgou, H.; Raymond, C. Bidirectional Deep Architecture for Arabic Speech Recognition. *Open Comput. Sci.* **2019**, *9*, 92–102. [CrossRef]
59. Tomashenko, N.; Vythelingum, K.; Rousseau, A.; Estève, Y. LIUM ASR Systems for the 2016 Multi-Genre Broadcast Arabic Challenge. In Proceedings of the 2016 IEEE Spoken Language Technology Workshop (SLT), San Diego, CA, USA, 13–16 December 2016; pp. 285–291.
60. Al-Irhayim, D.Y.F.; Hussein, M.K. Speech Recognition of Isolated Arabic Words via Using Wavelet Transformation and Fuzzy Neural Network. *Comput. Eng. Intel. Syst.* **2016**, *7*, 21–31.
61. Elmahdy, M.; Hasegawa-Johnson, M.; Mustafawi, E. Development of a TV Broadcasts Speech Recognition System for Qatari Arabic. *LREC* **2014**, *14*, 3057–3061.

62. Alalshekmubarak, A.; Smith, L.S. On Improving the Classification Capability of Reservoir Computing for Arabic Speech Recognition. In Proceedings of the International Conference on Artificial Neural Networks, Hamburg, Germany, 15–19 September 2014; Stefan, W., Cornelius, W., Włodzisław, D., Timo, H., Petia, K.-H., Sven, M., Günther, P., Alessandro, E.P.V., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 225–232.
63. Droua-Hamdani, G.; Selouani, S.A.; Boudraa, M. Algerian Arabic Speech Database (ALGASD): Corpus Design and Automatic Speech Recognition Application. *Arab. J. Sci. Eng.* **2010**, *35*, 157–166.
64. Ali, A.; Vogel, S.; Renals, S. Speech Recognition Challenge in the Wild: Arabic MGB-3. In Proceedings of the 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Okinawa, Japan, 16–20 December 2017; pp. 316–322.
65. Ali, A.; Bell, P.; Glass, J.; Messaoui, Y.; Mubarak, H.; Renals, S.; Zhang, Y. The MGB-2 Challenge: Arabic Multi-Dialect Broadcast Media Recognition. In Proceedings of the 2016 IEEE Spoken Language Technology Workshop (SLT), San Diego, CA, USA, 13–16 December 2016.
66. Ali, A.; Shon, S.; Samih, Y.; Mubarak, H.; Abdelali, A.; Glass, J.; Renals, S.; Choukri, K. The MGB-5 Challenge: Recognition and Dialect Identification of Dialectal Arabic Speech. In Proceedings of the 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Singapore, 14–18 December 2019; pp. 1026–1033.
67. Meftouh, K.; Harrat, S.; Jamoussi, S.; Abbas, M.; Smaili, K. Machine Translation Experiments on PADIC: A Parallel Arabic DIalect Corpus. In Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation, Shanghai, China, 30 October–1 November 2015; pp. 26–34.
68. Al-Diri, B.; Sharieh, A.; Hudaib, T. *Database for Arabic Speech Recognition ARABIC_D*; Paper or Report (Technical Report); University of Jordan: Amman, Jordan, 2002.
69. Khurana, S.; Ali, A. QCRI Advanced Transcription System (QATS) for the Arabic Multi-Dialect Broadcast Media Recognition: MGB-2 Challenge. In Proceedings of the 2016 IEEE Spoken Language Technology Workshop (SLT), San Diego, CA, USA, 13–16 December 2016; pp. 292–298.
70. Almeman, K. The Building and Evaluation of a Mobile Parallel Multi-Dialect Speech Corpus for Arabic. *Procedia Comput. Sci.* **2018**, *142*, 166–173. [CrossRef]
71. Kacur, J.; Rozinaj, G. Practical Issues of Building Robust HMM Models Using HTK and SPHINX Systems. In *Speech Recognition*; Mihelic, F., Zibert, J., Eds.; InTech: Rijeka, Croatia, 2008; pp. 171–192. ISBN 978-953-7619-29-9.
72. Novak, J.R.; Dixon, P.R.; Furui, S. An Empirical Comparison of the Tˆ3, Juicer, HDecode and Sphinx3 Decoders. In Proceedings of the Eleventh Annual Conference of the International Speech Communication Association, Chiba, Japan, 26–30 September 2010; pp. 1890–1893.
73. Zribi, I.; Ellouze, M.; Hadrich Belguith, L.; Blache, P. Spoken Tunisian Arabic Corpus "STAC": Transcription and Annotation. *Res. Comput. Sci.* **2015**, *90*, 123–135. [CrossRef]
74. Ahmed, B.H.A.; Ghabayen, A.S. Arabic Automatic Speech Recognition Enhancement. In Proceedings of the 2017 Palestinian International Conference on Information and Communication Technology (PICICT), Gaza, Palestine, 8–9 May 2017; pp. 98–102.
75. Loots, L.; Niesler, T. Automatic Conversion between Pronunciations of Different English Accents. *Speech Commun.* **2011**, *53*, 75–84. [CrossRef]