International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V73I3P116

Volume 73 Issue 3, 212-229, March 2025

Review Article

Enhancing Automatic Recognition of Isolated Arabic Speech Using Artificial Intelligence Techniques: A Systematic Review

Mithal Khaleel Ismael[1], Goh Chin Hock[1], Hazem Noori Abdulrazzak[2]

[1]*Institute of Power Engineering, Universiti Tenaga Nasional, Kajang 43000, Malaysia.*
[2]*Computer Communication Engineering Department, Al-Rafidain University College, Baghdad, Iraq.*

[2]*Corresponding Author : hazem.n@ruc.edu.iq*

Received: 17 December 2024 Revised: 11 February 2025 Accepted: 14 February 2025 Published: 28 March 2025

Abstract *- Automatic Speech Recognition (ASR) and spoken language systems have indeed made remarkable strides, fueled by advances in artificial intelligence, deep learning, and computational power. Modern ASR systems are now more accurate, robust, and capable of handling a variety of applications, from voice assistants to real-time transcription services. This review discusses the developments in isolated Arabic speech recognition using different AI methodologies. It emphasizes fundamental techniques, such as deep learning and machine learning algorithms, and then evaluates their effectiveness in enhancing recognition accuracy. This paper highlights the inherent obstacles of the Arabic language, including dialectal differences and phonetic complexities. It also examines the importance of feature extraction and model training in improving performance. The methodologies used in speech detection and processing, identifying gaps and correlations between known patterns, and presenting recent patterns are illustrated in this paper. A systematic review of the selected studies was conducted to identify and select relevant papers. The evaluation indicates that despite significant progress, additional research is needed to overcome current limitations and enhance practical application.*

Keywords *- Dialect recognition, MSA, Speech recognition, Artificial Intelligence, Machine Learning, ASR.*

1. Introduction

In recent years, spoken communication has increasingly replaced written communication as the primary means of exchanging information and building social connections. Spoken language, both in human-to-human and human-machine interactions, is preferred due to its immediacy and natural flow [1]. The quest to develop systems that can engage fluently in spoken dialogue has captivated scientists and engineers over the past century, with machine designs aiming to emulate human behavior. Homer is recognized as an early pioneer in acoustic and electronic engineering [2], while Bell Labs developed the first electrical speech synthesizer in the 1930s. During World War II, secure voice transmission became a focus of technological advancement [3]. Researchers in speech recognition have long faced the challenge of developing comprehensive systems from the ground up. Recently, experiments with neural networks have been conducted to interpret human actions through audiovisual inputs, and audio recognition systems have found applications in speech-enabled devices, home automation, machine translation, and medical and educational systems [4]. Voice modeling is the foundational step in voice recognition [5].

It connects acoustic information with linguistic data, and most of the calculations in acoustic modeling are dedicated to feature extraction and statistical representation, significantly impacting the recognition process [6]. Features extracted from speech are transformed into statistical representations that model the distribution of specific sounds. This modeling establishes a link between the extracted features and the structures of linguistic units [7]. Various feature extraction methods have been documented, including those based on human auditory perception and speech production mechanisms [8]. These techniques address challenges in speaker-independent speech recognition, enhancing the extraction of features for robust voice models [9, 10]. Technological advancements in recent years have expanded the role of dialect recognition and speech classification using deep learning for human-machine interfaces, machine control, and other applications [11]. The choice of classification method is equally important, as several studies have examined different algorithms and approaches to dialect classification [12]. Techniques in the field include Deep Neural Networks (DNNs), Hidden Markov Models (HMMs), discriminative training, Artificial Neural Networks (ANNs), and sequence-to-sequence audio modeling. For many years, Gaussian

Mixture Models (GMMs) and HMMs have dominated voice recognition systems [13-16], with HMMs handling time-based variability and sequential data structures, while GMMs provide local classifications [17]. However, GMMs are statistically limited in capturing non-linear interactions, such as those between phonemic features and human sound inputs. They are sensitive to training-test mismatches, especially when noise is introduced. Traditional approaches tackle these challenges using engineering techniques, including noise reduction, speech enhancement, and tailored input features [18]. Unlike raw speech, traditional systems rely on spectral features that require transformation from raw waveforms. Commonly employed features include Mel-Frequency Cepstral Coefficients (MFCCs) and Perceptual Linear Predictive Coefficients (PLPs) [19]. N-gram models are frequently used for probabilistic word prediction [20]. Speech recognition technology stands out as one of AI's most promising applications for smart home systems, enabling spoken communication with computers [21].

For typical residents, speech recognition is a complementary tool, while for elderly or disabled residents, it provides essential support that facilitates interaction with their environment, enhancing safety, comfort, and efficiency [22]. In homes where disabled individuals reside, particularly those with visual impairments, voice recognition is expected to facilitate integration and independence [23, 24]. The authors demonstrated a voice-controlled smart home system using pre-programmed commands to control different home areas [25]. Prior research in smart homes has produced valuable insights, particularly in energy saving [26], communication [27], and services for disabled individuals [28, 29]. Despite numerous studies on speech recognition in AI, systematic reviews on Arabic dialects and speech processing for smart devices remain scarce. This study bridges the gap by focusing on the automatic recognition of isolated Arabic speech using machine learning and deep learning. A Systematic Literature Review (SLR) was conducted, analyzing research from Arab and international sources between 2018 and 2024. This study seeks to answer the following questions:

Key research questions:

- RQ1: What are the most common algorithms and techniques used in isolated speech recognition?
- RQ2: What are the features extracted in this study?
- RQ3: How does improving the quality of the input audio data contribute to the accuracy of isolated speech recognition systems?
- RQ4: What are the techniques used to collect audio recordings?
- RQ5: What are the technical challenges in isolated speech recognition, especially regarding recognition accuracy in different environments?
- RQ6: What are the criteria used to evaluate the performance of isolated speech recognition systems?

The list of the abbreviations which is used in this paper is shown in Table 1.

Table 1. Abbreviations

Technology	Description
MSA	Modern Standard Arabic
ASR	Automatic Speech Recognition
DNNs	Deep Neural Networks
HMMs	Hidden Markov Models
GMMs	Gaussian Mixture Models
ANNs	Artificial Neural Networks
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
KNN	K-Nearest Neighbors
SVM	Support Vector Machine
GB	Gaussian Backend
RNN	Recurrent Neural Network
NN	Neural Network
AI	Artificial Intelligence
MFCCs	Mel-Frequency Cepstral Coefficients
PLPs	Perceptual Linear Predictive Coefficients
LRE	Language Recognition Evaluation
FB	Forward-Backward
DTW	Digital Wavelet Transform
BP	Back Propagation
GRU	Gated Recurrent Unit
PCA	Principal Component Analysis
FFBPNN	Feed Forward Back Propagation Neural Networks
LPC	Linear Predictive Coding
GFCC	Gammatone Frequency Cepstral Coefficient
PNCC	Linear Prediction Cepstral Coefficients
Mod GDF	Modified Group Delay Function
VQ	Vector Quantization
BiLSTM	Bidirectional Long Short-Term Memory
LBG	Linde–Buzo–Gray
VQLBG	Vector Quantization of Linde-Buzo-Gray

The paper is structured as follows: Section 2 covers related works, while Section 3 details the methodology and strategy. Section 4 presents the results of selected studies, and Section 5 discusses key findings, research gaps, challenges, and data analysis insights. Finally, Section 6 provides the conclusion.

2. Related Work

In this research, various surveys were conducted, examining topics related to speech models, classification methods, Arabic dialect recognition, and their applications in smart devices. The study emphasizes research papers that utilize deep learning and machine learning for classifying isolated Arabic speech, focusing particularly on the automatic recognition of Arabic dialects. Additionally, an overview is provided on implementing Deep Neural Networks (DNNs) comprising multiple hidden layers. Recent techniques employed in training these classes have received positive feedback in studies [30, 31]. This research also evaluates the potential of these advanced techniques as alternatives to traditional Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs) for acoustic modeling in speech recognition [32]. Findings indicate that DNNs, trained with novel methodologies and extensive hidden layers, outperform GMMs and HMMs in several aspects related to speech recognition criteria, with significant growth in feature quality observed in some instances [33, 34].The paper reviews recent speech classification and voice recognition advancements across various languages and dialects. It highlights methods and technologies in speech classification using cutting-edge artificial intelligence techniques and their applications across diverse fields.

The authors [35] discuss advancements in language extraction modeling, acoustic modeling, and speech understanding, introducing more sophisticated spectrograms using MFCC-DNN models compared to traditional GMM-HMM methods. This study underscores the need to refine DNN architectures to enhance acoustic measurement capabilities. Before the rise of deep learning, GMM-HMM models were commonly utilized in audio modeling for training and recognition with MFCC features [36]. However, the advent of AI has popularized deep learning in voice classification, primarily through DNNs with a SoftMax classifier following a multi-layer network structure [37]. While each method has distinct advantages, the simplicity of classifiers and the complexity of DNN features often surpass the high complexity of GMM-HMM models. The limitations of MFCCs in identifying and differentiating basic speech units are noted [38].

This study proposes combining both approaches by selecting the output from the final DNN layer as the final feature transformation, which is then employed to train and recognize the GMM-HMM model, thereby balancing feature complexity and model accuracy [39]. In voice recognition, accents of non-native speakers present a significant challenge due to the presence of phonemes that are uncommon in standard pronunciation. Dialect classification can enhance recognition systems by identifying the speaker's ethnicity, allowing the system to switch to settings optimized for that specific dialect. The authors [40] employed dialect detection through Dual Neural Networks (DNN and RNN), hypothesizing that dialectal differences stem from phonetic and pronunciation characteristics. This study proposes combining long-term and short-term feature training, with the DNN focusing on long-term feature prediction for dialects, while RNNs handle short-term features. Results from both networks are integrated using a probabilistic fusion algorithm [41]. However, the proposed model encounters misclassification issues among geographically close languages, such as between Hindi and Telugu or Japanese, Korean, and Chinese [42].This research aims to identify a new dataset's most effective feature extraction model. Mel-Frequency Cepstral Coefficients (MFCC) are identified as a widely used feature extraction method due to their efficiency in generating minimal data while retaining essential information through short-time spectral analysis.

Therefore, this study employs MFCC as the feature extraction method [43]. The dataset comprises 1023 .wav files containing samples from six language classes: Arabic, English, French, Korean, Mandarin, and Spanish [44]. Following feature extraction, the KNN method is utilized to analyze MFCC-derived data vectors and evaluate KNN accuracy, recall, precision, and F1 scores for accent determination. In a prior study [45], speech emotion recognition was explored using GMM and KNN, concluding that KNN is a straightforward classification method with extensive applications. The datasets used were the NIST LRE dataset for language classification and the Mozilla Common Voice dataset for accent analysis.

A vector/I-vector-based representation with a Gaussian Backend (GB) was employed [46], achieving an accuracy of 89.4% in 30 seconds on the LRE Development dataset with the MMSE s-vector configuration [47]. Furthermore, as the use of smart home devices increases, intelligent interaction in this context can provide users with greater comfort and convenience. Voice recognition powered by artificial intelligence allows users to control their home environment entirely through audio commands. Researchers increasingly use Artificial Neural Networks (ANN) to categorize user inputs, facilitating realistic dialogues and enabling device management via voice or text commands [48, 49].

3. Methodology and Strategy

3.1. Design

This study has a mixed methodological review (qualitative, quantitative, and mixed) and includes the study on the techniques used in automatic recognition of isolated Arabic speech, voice and dialect recognition, voice command applications, and speech in smart devices [50] and adherence to the Preferred Standards for Reporting Requirements for Meta-Analyses and Systematic Reviews (PRISMA 2020) [51, 52]. The following steps include (1) eligibility criteria, (2) information sources, (3) search terms, (4) study selection, (5) data collection and synthesis process, and (6) critical recommendation.

3.2. Eligibility Criteria

Studies are eligible for inclusion if they: Provided empirical research and conceptual evidence directly related to the research topic and advanced methods and protocols used by researchers from 2018 to 2024 and published in scientific journals. The main reason for selecting studies is the frequent use of dialects in Arab countries. This has prompted the need to research methods for speech and audio recognition and classification in AI to facilitate future researchers to use dialects in other applications. Technologies play an important role in AI; however, access to these technologies remains a major challenge for communities' limited resources. Studies written in languages other than English were ignored. We also excluded traditional reviews, studies unrelated to the research topic, and papers not readily available in full text.

3.3. Information Sources

The search methodology was to obtain as many relevant strategies as possible, and different procedures were applied to as diverse studies as possible from sources [53]. A systematic and comprehensive search expansion was conducted across six online databases: (Web of Science, Research Gate, IEEE Xplore, SpringerLink, Science Direct, and the Digital Library of the Association for Computing Machinery (ACM)). Search engines (Google Scholar). Also, the American Academy of Sciences (ACADEMIA). Experts were contacted to gather opinions on the search and identify additional or unpublished studies. Finally, Google Scholar was utilized to review relevant studies' reference lists, helping uncover additional hidden research. It is worth noting that each database searched from 2018 to 2024 is closely related to research studies that refer to research on Arabic dialects, isolated speech recognition, and modern methods of phoneme and speech classification. Therefore, searching through these databases is necessary to display the latest research.

3.4. Search Terms

The importance of the research process led to the optimization of keywords. First, keywords and concepts were compiled from the acquired studies and compared with the research objectives and questions. Second, synonyms and relevant characteristics were developed.

The identified keywords were then tested across different databases and optimized accordingly. Table 2 summarizes the final list of the collected search terms. Then, logical operators were associated with different sets of keywords and designed as follows:

1- Arabic dialect recognition and Arabic accent recognition. 2- Voice recognition or isolated speech or voice command techniques or isolated speech techniques. 3- Artificial intelligence applications in isolated speech recognition or artificial intelligence applications in voice command recognition. 4- Speech or voice classification methods and methods for arabic dialect recognition.

Table 2. Categories and keywords in the research process

Category	Keywords
Dialect or accent	Arabic dialect recognition, Arabic accent recognition
Speech or voice recognition	voice recognition or isolated speech or voice command techniques or isolated speech techniques
Applications in artificial intelligence	artificial intelligence applications in isolated speech recognition or artificial intelligence applications in voice command recognition
Techniques used for classification	speech classification methods and voice or sound classification methods and methods for Arabic dialect recognition

3.5. Study Selection

The procedure of selecting the study for analysis and evaluation and determining its relevance to the research is based on the objectives of the systematic review. This study illustrates the stages that the search method went through from 2018-2024. In the first stage (S1), records are identified from different information sources (academic journals, search engines, Google and reference lists). Once all the records are collected, the first filter is applied in the second stage (S2), eliminating duplicates. Mendeley is used to remove duplicate entries and efficiently manage all the records. Once all duplicates are removed, the records are screened based on “title, keywords and abstract” during this third stage (S3), i.e. studies that do not meet the eligibility criteria are excluded. Also, during this stage, we look at “full-text” studies. Screening and reviewing the studies were within the scope of this systematic review and relevant. This stage (S4) aims to fully examine the research for all studies that achieved full eligibility for the research methodology, as the selected studies were verified in terms of the research gap, the research problem, and the results extracted from the research, thus obtaining 32 articles that met all the requirements of the research methodology.

3.6. Data Collection and Synthesis

All embedded articles were reviewed and verified. Mendeley was used to collect essential publication data, including date, title, authors, publisher, DOI, summary URL, keywords, pages, size, and issues. The articles were summarized based on their raw ratings and stored as Microsoft Word and Excel files for easy filtering. Each article was thoroughly read in its entirety. Through the process of generating the recommended classification, we identified some highlights and notes regarding the articles being scanned and the continuous classification of all articles; moreover, the scientific technique was used to classify articles with a lot of comments and identify the groups [54]. Depending on each author's style, comments were categorized as important and unimportant. Then, the primary results were characterized, described, tabulated, and concluded. The results are included in the supplementary material.

3.7. Critical Appraisal

The authors of [55] evaluated the quality of the selected studies to prevent misinterpretation and bias. The criteria from the Quality Assurance and Assessment Tools for Systematic Review of Contradictory Data were followed in this process. Table 3. shows the checklist used to assess each study individually. Six assessment criteria form the basis of this checklist:

- (CQ1) Did the abstract contain the idea of the research, the objectives, the method of work, and the results clearly?
- (CQ2) Was previous research reviewed for the research work, the results summarized, and the research gap identified, with a clear explanation of the research problem and its objectives in the research introduction?
- (CQ3) Were the techniques used in the research clearly explained and detailed?
- (CQ4) Were the strategies used in collecting data explained in detail?
- (CQ5) Was the data analysis described accurately?
- (CQ6) Were the results clear and described in detail?
- (CQ7) Suggestions for future studies and recommendations for future research work provided.

The selected studies were classified into three quality categories: Good (G), Average (A), and Poor (P). This classification was based on predefined methodological criteria to ensure an unbiased assessment. The primary objective was to evaluate and categorize each study according to its quality, ensuring consistency and reliability. A structured approach was adopted to conduct a sensitive intervention, analyzing each study's strengths and weaknesses. The classification helped identify methodological gaps and inconsistencies in the literature.

Additionally, it provided a framework for understanding the significance of the results obtained from the reviewed studies. The evaluation process aimed to maintain a high standard of systematic review by reducing misinterpretation and bias. Critical assessments were conducted through a structured consensus approach among reviewers to ensure accuracy.

The results were then examined for relevance, impact, and methodological soundness. Finally, the findings contributed to a more comprehensive understanding of the field and highlighted areas for further research.

Table 3. Checklist items used for critical appraisal

Ref	Year	CQ1	CQ2	CQ3	CQ4	CQ5	CQ6	CQ7
[56]	2019	G	A	G	A	A	G	P
[57]	2019	A	A	G	G	A	G	A
[58]	2020	A	A	G	G	A	P	P
[59]	2022	A	G	G	A	G	G	P
[60]	2019	G	A	G	G	G	G	P
[61]	2019	G	G	G	A	G	G	P
[62]	2018	A	G	G	G	G	G	G
[63]	2020	G	A	G	G	A	A	A
[64]	2018	G	A	G	G	G	A	P
[65]	2018	G	A	A	G	A	A	A
[66]	2019	G	A	G	G	G	G	G
[67]	2024	G	G	G	G	G	G	P
[68]	2020	A	A	A	G	A	P	P
[69]	2020	G	A	A	G	A	G	P
[70]	2021	G	G	A	G	A	G	P
[71]	2024	G	G	G	A	G	G	G
[72]	2019	G	G	G	G	A	G	P
[73]	2021	G	G	G	G	G	G	A
[74]	2021	G	G	G	G	A	G	P
[75]	2022	G	A	G	G	A	G	G
[76]	2021	G	A	G	G	A	G	A
[77]	2021	G	G	A	G	G	G	P
[78]	2018	G	A	A	G	G	G	G
[79]	2018	A	G	G	G	A	A	A
[80]	2020	G	A	A	G	A	G	P
[81]	2021	A	G	A	G	G	G	P
[82]	2024	G	G	A	G	A	G	G
[83]	2023	G	G	G	G	P	G	P

[84]	2019	G	G	A	A	A	G	G
[85]	2020	G	G	A	G	A	A	A
[86]	2020	G	G	G	G	A	G	A
[87]	2018	G	A	G	G	G	A	P

4. Summary of Quantitative Data

Firstly, this section provides a detailed summary of the quantitative data from the selected studies, including essential features such as the references, main focus, number of records analyzed, modeling techniques used, feature extraction methods, and the data sources. Additionally, it reports the accuracy metrics, publication location, journal names, and study evaluation, which are systematically organized in Table 4 for easy comparison.

Secondly, Figure 1 visually presents the results of the studies and the selection process, which follows the PRISMA 2020 guidelines to ensure transparency and reproducibility. In the third and final section, a categorization of the studies is provided, grouping them according to relevant research themes and outcomes. This includes critically evaluating each study's methodology, reliability, and contribution to the field. All results in this section are closely aligned with the research questions.

Table 4. Quantitative summary of studies

Ref	Main Focus	# Records	Model	Features	Data source	Accuracy	Future Studies
[56]	Moroccan dialect	20 Speakers	HMM	MFCCs	Audio Recording	90%	-------
[57]	MSR	88 Spoken digit	LSTM/ GRU	MFCC/FB	The laboratory of automatic and signals, University of Badji-Mokhtar - Annaba, Algeria. TV.	96%	The researcher directed the evaluation of the systems with a more realistic dataset for noisy environments**.**
[58]	MSA	50 Speakers	HMM	MFCC	King Faisal University	65.72%.	The researcher hopes to explore many deeper DAE configurations and the number of neurons in each layer in the future.
[59]	MSR	50 male Arabic	KNN–DTW	MFCC	Microphone	99%	-------
[60]	MSA	50 Different persons	BP, ANN	LPC	Commercial Microphone	94.5%	-------
[61]	MSA	7 male speakers, utter 340	HMMs	MFCC	Recorded by native male Saudi speakers	92.70%	Focus on HMM parameters, acoustics and particle filter elements. You can also perform other experiments as a hybrid system, like ANN with HMM.
[62]	MSA	20 Speakers	ANN	Rasta PLP, ΔRasta-PLP, PCA, FFBPNN	A Sennheiser dynamic microphone	0.68%	Automatic speech recognition and its applications are focused on real-time, such as robotics.
[63]	MSA	40 Speakers	KNN	MFCCs	The recording app on a mobile phone	98.1250%	Create speech recognition apps that accommodate the Arabic language.
[64]	Arabic Tunisian language	HCopy command	HMM	MFCC, PLP, and LPC	HTK platform	98.64%	-------
[65]	MSA	88 speakers	LSTM or GRU	MFCC	Laboratory of automatic and signals, University of Badji-Mokhtar - Annaba, Algeria	98.77%	Assessing this type of system with noisy (More realistically) speech signals.

[66]	MSA	1GP Speakers	DNN-HMM	MFCC	The recording microphone	93%	The aim is to take advantage of continuous speech or verbal data to modify techniques and apply them more widely.
[67]	MSA	2538 speakers Recorded by Android Application	CNN	MFCC	Recorded by Android Application	84%	The objective is to create the most precise and dependable Arabic voice recognition systems within the domain of the IoT.
[68]	MSA	1730 Recordings	HMM	MFCC	One single computer and one single microphone	93%	Establishing a global network utilising HMM to identify isolated Arabic words.
[69]	MSA	4500 Word	CNN, ANN	MFCC	Microphone to input voice	85%	The microcontroller must utilise pulse width modulation (PWM) to provide a changing DC voltage in response to spoken commands.
[70]	MSA	16PP Recorded	G Naïve Bayesian	MFCC	A laptop microphone	98 %	-------
[71]	MSA	1G speakers	Feed Forward Neural Networks and Keras-based NN	MFCCs, MFFCs	Speakers at Hashemite University in Zarqa, Jordan, by Microphone	93.09%.	Dimensionality reduction and extracting speech features using convolutional neural networks to improve Arabic speech recognition on a large dataset.
[72]	Different dialects **:** Yemen, Egypt, Sudan, Iraq, and Saudi Arabia	104 Native Arabic speakers	LSTM	MFCCs	Apps like WhatsApp	94%	-------
[73]	MSA	50 Native-Arabic speakers	GMM-HMM and DNN-HMM	MFCC	Recordings of single utterances	97.1%	Kaldi and CMU Sphinx tools for Hindi, Arabic, Spanish, and English in noisy situations compared.
[74]	MSA	50 Native male Arabic speakers	CNN	GFCC	Department of Management Information Systems, King Faisal University by microphone	99.77 %	-------
[75]	MSA	50 Speakers	CNN, LSTM	MFCC	Systems Department of King Faisal University	98,42%	Develop an end-to-end Arabic automatic speech recognition system using larger databases and the KALDI toolkit or other methods.
[76]	Various dialects	12000 Arabic spoken	CNN	MFCCs	Microphone to Input voice	91.625 %	-------
[77]	MSA	6 Speakers of the Arabic alphabet were recorded	CNN	MFCC	Microphone to input voice	92.86%	The study investigates the technique applied to adult vocalisations of various letters in the Arabic alphabet that exhibit nearly identical pronunciations.

[78]	Tunisian dialect	1P Voluntary speakers	FFBP, ANN	PLP, MFCC	Dynamic microphone	98.5%	Extend our database to all Arabic dialects, use advanced algorithms to account for speech non-linearity, and improve Arabic speech recognition and dialect classification.
[79]	MSA	890 Speech samples	KNN	MFCC	Recorded using COOLEDIT	83.3%	-------
[80]	MSA, Yemen dialect	50 speakers	SVM	PNCC, ModGDF, MFCC	Recorded by microphone from students	93-97%	-------
[81]	MSA, dialect (Levantine, Gulf, and Egyptian)	52 speakers dialects, 12 speakers MSA	HMMs	FCN, MFCC	The home, in a moving car, in a public place and a quiet place	97%	An over-complete DAE model provides fresh insights into unsupervised learning; thus, we plan to study its possibilities in future research.
[82]	MSA, Medina Dialect	70364 Tokens	HMM	MFCC	Recording by microphone R	%92.09	Future studies may use deep learning with large corpus data. A graphical user interface program can improve user experience.
[83]	MSA	107 Speakers	CNN	__	Apps (WhatsApp, Facebook Messenger)	89.61%.	-------
[84]	MSA	11 Arabic spoken	SVM	WAT-MFCC)	Recording by microphone	100%	This approach should be tested online or offline on datasets such as TIMIT. It should be implemented on FPGA and Arduino hardware architectures to track performance improvements in real-time.
[85]	MSA	119 Speakers	CNN	MFCC	The audio was recorded with a microphone in a quiet environment.	98.5%	A sequential data classifier like long short-term memory may use convolutional neural network-extracted properties.
[86]	MSA	100 Arab Moroccan speakers	VQ and GMM	MFCC	Recording by microphone	97.92%	To obtain better results, this method is compared with other methods to improve the model with other types of noise as the data size increases.
[87]	Malwi, Majhi and Doabi dialects	100 speakers	HMM	MFCC	These are desktop-mounted microphones and mobile phone	84.04%	The system comprises ongoing speech recognition and a substantial amount of data. Utilising many dialects.

4.1. Methodology

In this section of the research methodology, the elements of the coloring report (PRISMA 2020) were applied to the review of selected literature from 2018 to 2024, *w*here a detailed study was conducted on all the studies, which numbered 1850 studies from the databases (Web of Science, Research Gate, IEEE Xplore, Springer Link, Science Direct, the ACM, Google Scholar, and ACADEMIA*). A*fter deletion and removal, according to the methods of selecting the methodology, which were explained in the section on selecting studies and the results extracted from this report, until they reached the studies that meet all the quality criteria, which numbered 32 studies and are shown in Figure 1 in the (PRISMA 2020) flow chart.

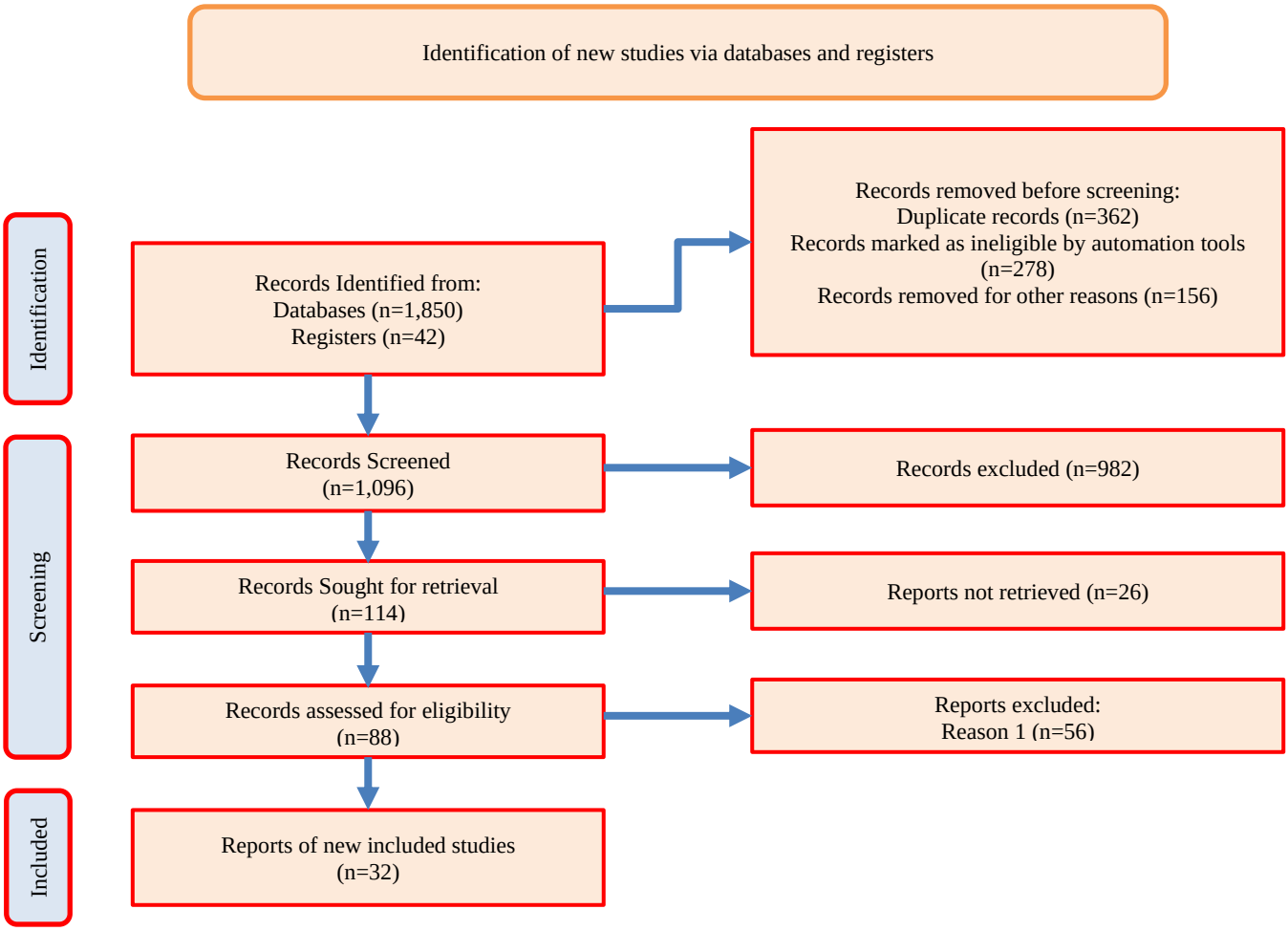

Fig. 1 (PRISMA 2020) Flow chart for the screening and selection process of the selected studies

4.2. Study Trends

Figure 2 shows the annual distribution of the selected studies. The number of studies has increased dramatically in recent years, indicating that the study of discrete speech recognition and its applications are gradually attracting the attention of academics and researchers. Most of the selected studies were conducted between 2018 and 2024.

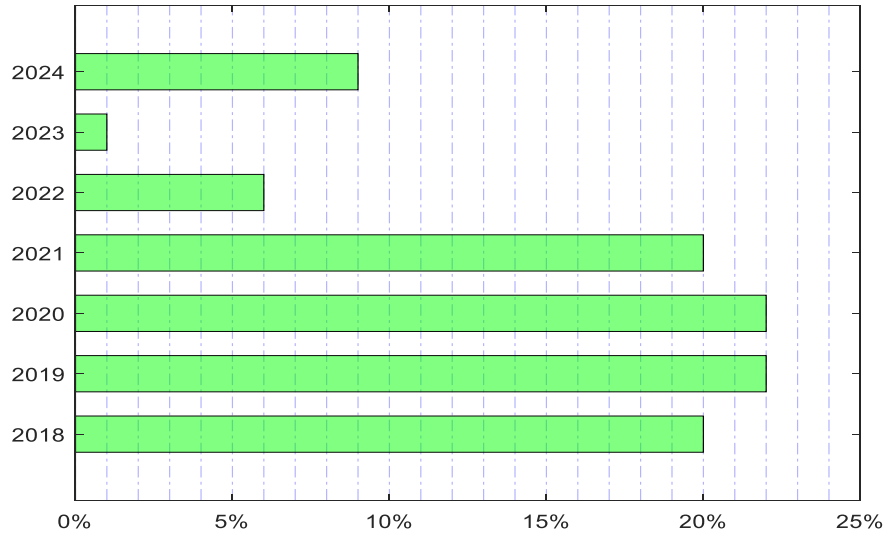

Fig. 2 Distribution of studies per year

4.3. Studies Published

To accommodate the 32 selected studies, the publications were categorized into two groups. Of these, 19 studies were published in journals, while 13 studies were presented at conferences. Each study was evaluated according to specific study requirements, such as research objectives, methodology, and outcomes. The classification helped identify patterns in the type of publication venue and its potential impact on the study quality. Journal publications were assessed for their peer-review process, impact factor, and relevance to the research field, while conference papers were evaluated for their contribution to emerging trends and innovations. Table 5 provides a detailed summary of the publications in the selected journals, listing relevant information such as the journal name, volume, issue, and publication date. This breakdown allows for a clearer understanding of the distribution of studies across different platforms. It also highlights the importance of journal articles in establishing the foundational research, while conference papers often provide insights into the latest developments in the field.

Table 5. Summary of the published studies

Ref	Journal	Conference	Ref	Journal	Conference
[56]		√	[72]		√
[57]	√		[73]	√	
[58]		√	[74]	√	
[59]	√		[75]		√
[60]	√		[76]	√	
[61]	√		[77]	√	
[62]	√		[78]	√	
[63]		√	[79]	√	
[64]		√	[80]	√	
[65]		√	[81]	√	
[66]	√		[82]	√	
[67]	√		[83]	√	
[68]		√	[84]	√	
[69]		√	[85]	√	
[70]		√	[86]	√	
[71]	√		[87]	√	

4.4. Classification of Studies

In Table 6, the diversity of studies in speech recognition research is presented. Most studies, accounting for 75%, focused on the classical Arabic language, while only 25% addressed dialects, speech classification, and their applications. This indicates a significant gap in research on Arabic dialects, highlighting the need for more studies in this area due to the diversity of dialects across regions.

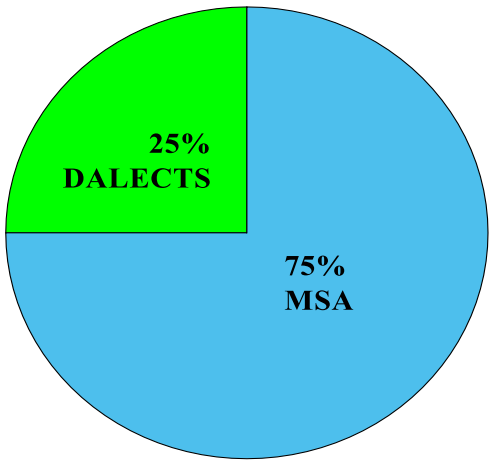

Fig. 3 Percentage of speech classification

Table 6. Classification of Studies in Speech Recognition

Studies	MSA	Dialects
[57-63, 65-71, 73-75, 77, 79, 83-86]	MSA	-----
[56]		Moroccan dialect
[64]		Arabic Tunisian language
[72]		Dialects (Yemen, Egypt, Sudan ,Iraq, KSA)
[76]		various dialects
[78]		Tunisian dialect
[80]	MSA	Yemen dialect
[81]	MSA	Dialect (Levantine, Gulf, and Egyptian)
[82]	MSA	MEDINA DIALECT
[87]		Malwi, Majhi and Dubai dialects
SUM	26	6

The study emphasizes the importance of expanding research in recognizing and classifying Arabic dialects to enhance the effectiveness of speech recognition systems. Figure 3 illustrates the percentages of studies dedicated to recognizing and classifying the classical Arabic language compared to Arabic dialects, further underlining the disparity between the two focus areas.

5. Results Derived from the Methodological Questions

5.1. RQ1: What are the Most Common Algorithms and Techniques Used in Isolated Speech Recognition

Various algorithms have been used in isolated speech recognition studies, with different techniques showing varying use rates, as shown in Figure 4. Analyzing the use of these algorithms provides a clear view of the common patterns and approaches in the field. Through the analysis of techniques used in the reviewed studies, it is clear that HMM and CNN are the most popular techniques, each representing 25% of the methods. This popularity is attributed to their robustness and ability to effectively model and analyze speech data, particularly in isolated speech recognition tasks. In contrast, traditional algorithms like GMM and Naïve Bayes have seen a noticeable decline in usage, signalling a shift away from classical machine learning techniques. This shift reflects the broader trend toward deep learning methods, especially LSTM networks and CNNs, which are well-suited to handle the complexities of speech data due to their capacity to learn temporal dependencies and spatial features, respectively. Furthermore, emerging techniques combining DNN with HMM have shown promise in improving recognition accuracy by capturing temporal dynamics and contextual information. Other advanced techniques, such as SVM, ANN, and KNN, are also rising. These methods leverage deep machine learning architectures to enhance performance, often by improving feature extraction and classification accuracy. The increasing trend of utilizing deep learning and hybrid models underscores the growing emphasis on performance optimization in speech recognition. This diversity of techniques reflects the dynamic nature of the field and the ongoing development of more efficient, accurate systems capable of handling the complexities of isolated speech recognition.

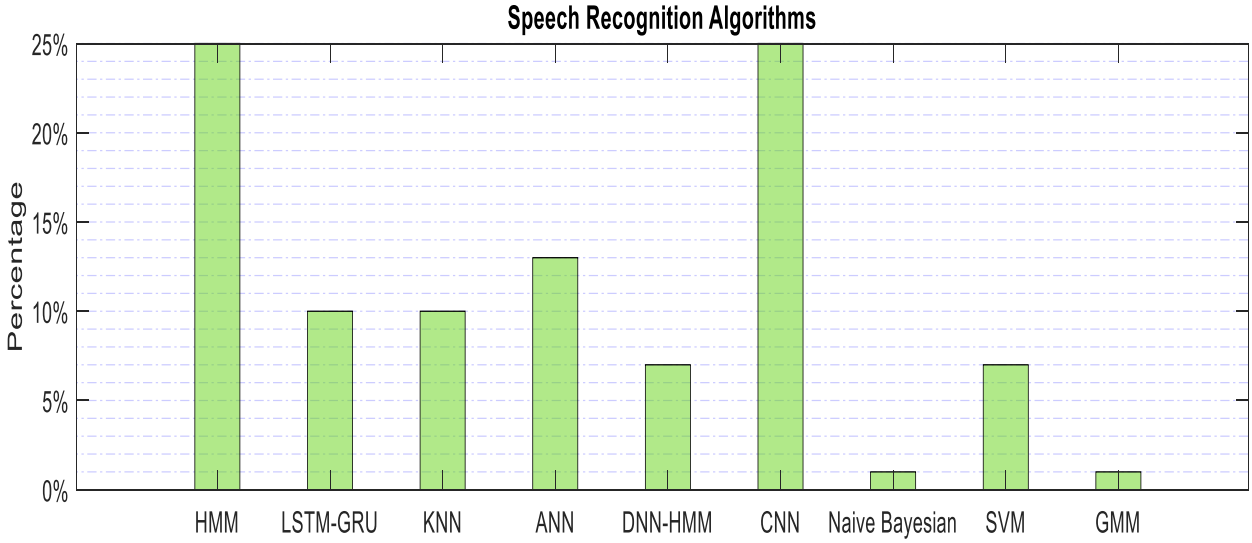

Fig. 4 Techniques used in studies for classifying speech

5.2. RQ2: What are the Features Extracted in this Study?

The isolated Arabic speech recognition methodology mainly uses machine learning and deep learning techniques. Among the features used, the Mel Frequency Coefficients (MFCC) feature was used in 88% of the reviewed studies, which represents the highest percentage. This widespread adoption of the Mel Frequency Coefficients feature can be attributed to its effectiveness in extracting speech features with high accuracy. In contrast, other features, including Perceptual Linear Prediction (PLP), Principal Component Analysis (PCA), Fully Connected Networks (FCN), Power Normalized Mel Frequency Coefficients (PNCC), Mel Frequency Coefficients (GFCC), and Linear Prediction (LPC), together accounted for only 11% of the studies and only one study did not mention the feature used in the research.

5.3. RQ3: How Does Improving the Quality of the Input Audio Data Contribute to the Accuracy of Isolated Speech Recognition Systems?

This study contains a wide range of speech and linguistic data, primarily in Modern Standard Arabic and several Arabic dialects, including Moroccan, Yemeni, Tunisian, and others. Records vary regarding speaker demographics, recording characteristics, and linguistic element concentration. A large part of this methodology covers Modern Standard Arabic (MSA) speech, with the dataset of recordings ranging from 12 to 2,538 speakers across the selected studies. The number of

studies for Arabic dialects is Moroccan Arabic (20 speakers), Tunisian Arabic (10 volunteer speakers), Yemeni Arabic (50 speakers), and others such as Levantine and Gulf dialects. Some records also include datasets with speakers from diverse dialects (e.g., Yemen, Egypt, Sudan, Iraq, and Saudi Arabia), totaling 104 native speakers.

5.4. RQ4: What are the Techniques Used to Collect Audio Recordings?

Collecting audio recordings is vital in building audio databases that support research and development in various fields, such as speech recognition and natural language processing. Figure 5 provides information about the different techniques used to collect audio recordings and the number of samples collected using each technique.

Microphones were the primary means of collecting audio recordings, as they were used in 19 cases, indicating their use as a primary option for obtaining pure audio quality. For using recording applications on smartphones, such as popular applications (WhatsApp and Facebook Messenger), to record audio to collect five audio samples. This technology represents a practical and easy solution for collecting audio recordings in diverse environments. The HTK (Hidden Markov Model Toolkit) platform recorded a single sample. It is a specialized tool used in processing and analyzing audio and creating speech recognition models. Signals and Automation Laboratory, Badji Mokhtar University - Annaba, Algeria. Two samples were recorded in this laboratory, reflecting the use of specialized scientific environments to achieve high-quality recordings. Hashemite University, Zarqa, Jordan. One sample was recorded using a microphone at this university, indicating the diversity of audio data regarding dialects or environments. As for King Faisal University, three samples were recorded, highlighting the contribution of this institution in collecting diverse data. The COOLEDIT software was used to record a single sample. It is an effective tool for editing and enhancing audio, providing high recording quality and advanced editing features.

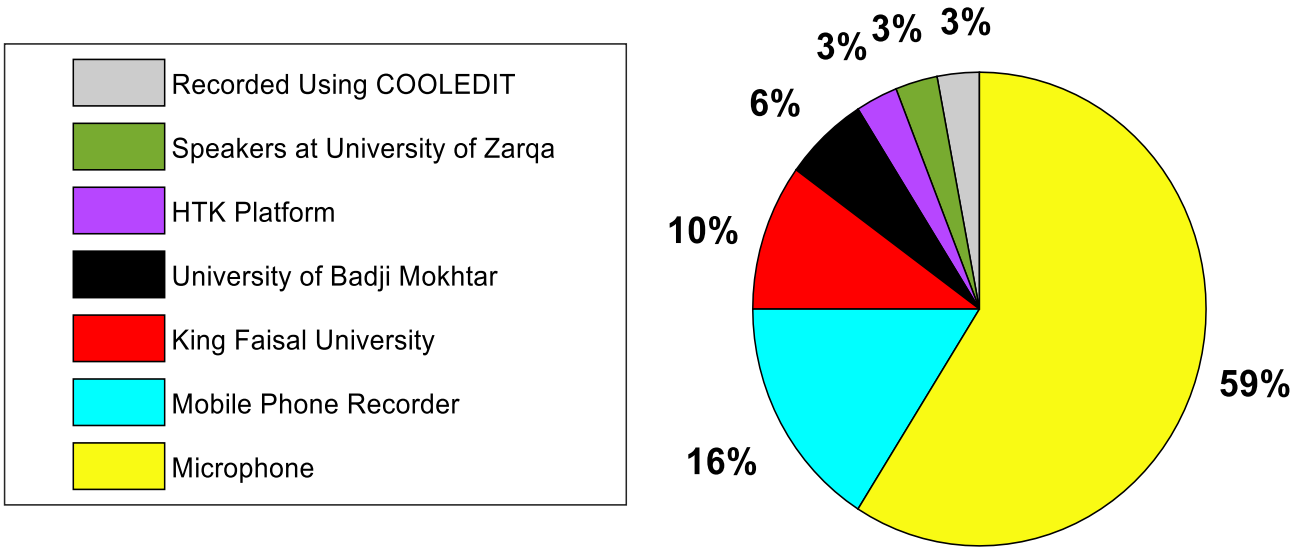

Fig. 5 Techniques used in audio recordings

5.5. RQ5: What are the Technical Challenges in Isolated Speech Recognition, Especially Regarding Recognition Accuracy in Different Environments?

Although many studies have addressed the automatic recognition of isolated Arabic speech and achieved good results, some challenges and problems face automatic speech recognition, including the recognition of Arabic dialects. Given the diversity of this type of language, the challenge facing the various Arabic dialects is the dataset to accurately recognize the dialects. It was noted that most studies use a small dataset, and therefore, searching for a large dataset to train the model is very difficult due to the complex phonetic and morphological richness [63]. One of the main difficulties is the variation in pronunciation across different regions, which may lead to inconsistencies in speech recognition systems. In addition, the lack of a reference for vowels in the Arabic text often complicates the accurate transcription of the spoken language [78]. The second challenge is automatic speech recognition in the presence of a noisy environment in the background. The system's performance in recognizing and classifying speech deteriorates in the presence of noise [59]. Although many studies have been conducted to solve this problem, they have not achieved the desired results using artificial intelligence algorithms [66]. The last challenge is the use of a microphone to record audio data. Using a microphone type may lead to poor speech recognition quality if it is average or poor quality, or the inability to separate noise is a problem in model training [80].

5.6. RQ6: What are the Criteria Used to Evaluate the Performance of Isolated Speech Recognition Systems?

Automated recognition of isolated Arabic speech is one of the areas that has received a great deal of attention recently, as the application of this field in artificial intelligence has greatly and rapidly helped in various areas of life. Therefore, most research has focused on improving the use and application of this type of challenge by enhancing the small data set by artificially augmenting the data as well as integrating some algorithms in deep learning to obtain better results and extract features as well as evaluating the results through accuracy, as shown in Table 7.

Table 7. Performance evaluation

References	Year	Evaluation of Studies
[56]	2019	This research aims to recognize voices using Moroccan dialects and identify the unknown speaker by extracting the unknown phonetic features and comparing them with the previously stored phonetic features.
[57]	2019	The results obtained through the delta-delta features show their effectiveness for classifying speech signals and training the model using the data in the first and second sections, where the accuracy reached 96% compared to other techniques using the FBs and MFCC features.
[58]	2020	The research aims to improve the communication between the client and the server through mobile networks to reduce the impact of speech recognition performance degradation caused by speech codecs. Using a deep self-encoding algorithm to enhance the degraded speech by 2.711 CODEC and transport.
[59]	2022	In this research, the KNN algorithm was used to recognize isolated Arabic speech, as the proposed algorithm proved its ability to recognize speech in noisy environments.
[60]	2019	This research dealt with recognizing voice speech commands using different features, such as neural networks and fuzzy logic behavior, and it was concluded that neural networks are better in accuracy and precision, but they require high configuration and the longest processing time.
[61]	2019	This research deals with recognising speech commands using different features of neural networks and fuzzy logic behavior. It concluded that neural networks are better in accuracy and precision, but they need high configuration and longer processing time.
[62]	2018	This paper discusses hybrid techniques for feature extraction, and the researcher points out that using these techniques, PLP and Rasta-PLP techniques with the "Trainscg" algorithm gave better results with an error rate of 0.68%.
[63]	2020	This paper presents a comparison between MFCC feature extraction using the KNN method in terms of training time using a database where it was the best with the three speech recognition techniques in terms of recognition accuracy for extracting the temporal dispersion feature of waves using Long Short-Term Memory (LSTM) classifier.
[64]	2018	In this paper, the speech recognition model under the HTK platform, with different features such as MFCC, PLP, and LPC, is used. Thus, the effectiveness of MFCC is proven by experimental results, which mainly depend on the datasets used for this task.
[65]	2018	In this paper, the results show that the MFCC features (presented to the classification system) are effective in speech recognition using the BiLSTM technique, and the results obtained are good.
[66]	2019	This paper presented good results using spoken Arabic numerals, especially when using meaningful parameters and features to expand its scope of use in AI applications.
[67]	2024	The researcher used a large vocabulary of 148 commands, and the purpose of this research was to enhance voice recognition and support for smart devices.
[68]	2020	This work focused on providing a new Arabic isolated word database by using the new concept of phonemes using the G8 standard consonants that improve speech recognition accuracy.
[69]	2020	This paper presented a dataset using four Arabic voice commands and used the SGDM algorithm implemented using MATLAB codes, which achieved good results for application in the ATmega328 project.
[70]	2021	The aim of this paper: The researcher found the use of the naive Bayesian algorithm to be simple, efficient in classification, and perform well.
[71]	2024	The proposed model found an accurate way to recognize speech, determine the gender of the speaker, and understand it in a noisy environment, and thus the extracted results were good.
[72]	2019	This work achieved good results using the LSTM technique but took a long time to train. Also, the results extracted using MFCC features were good but caused confusion when evaluating the results between the numbers.
[73]	2021	Evaluation The results obtained from the comparison between the hybrid model DNNHMM with the

		Kaldi toolkit and the GMM-HMM model with the Pocketsphnix toolkit We conclude that the first model gave better results in speech recognition in noise.
[74]	2021	CNNs have many advantages in acoustic modeling. One of these advantages is solving the problem of small data using data augmentation with the GFCC feature extraction technique, which achieved good results of 99.77%.
[75]	2022	This paper aims to recognize discrete speech using the CNN+LSTM technique, which improved the model's accuracy by 98.42%.
[76]	2021	This system proposed a CNN technique using MFCC features to extract features using a large dataset of 12,000 samples of Arabic numerals and commands in different dialects.
[77]	2021	In this study, the stuffing method was proposed for Arabic letter classification, and its performance was compared with the spectral, MFCC, and mel-spectrogram techniques. It was found that the stuffing and spectral techniques have superior performance over mel-spectrogram and MFCC for Arabic phoneme recognition.
[78]	2018	In this research, the researcher proposed a speech recognition system based on hybrid feature recognition techniques to extract FFBPNN for classification. For the purpose of validating the work, experiments have shown that PLP was used first, followed by ΔPLP, and then VQLBG was used second. Provides an average test performance of 98.54%. The LBG algorithm is also better than the PCA algorithm in terms of performance.
[79]	2018	This paper aimed to create an automatic device Technique for detecting pronunciation errors and applying it to A Speaker's speech. Use MFCC, using functions to calculate the Melcepstrum of the speech signal. KNN technology is used to obtain high accuracy in speech recognition.
[80]	2020	In this paper, three features are extracted. The algorithms, namely MFCC, PNCC, and ModGDF, instead of SVM, were used for the classification process. The results showed PNCC was more efficient, PNCC got 93-97%. The accuracy rate, ModGDF, was 9P%, and MFCC was 88%.
[81]	2021	The proposed approach is an effective alternative to implementing robust speech recognition, where a deep autoencoder is used to reduce noise and produce the final enhanced speech signal to be recognized. Meanwhile, the two-step model is unsupervised. Achieves significant improvement in speech quality (PESQ) and clarity (STOI) of about 0.835 and 0.06, respectively.
[82]	2024	In this research, specific audio applications were used to process and annotate embedded audio files. The total number of audio file integrations recorded was 15, performed across 21 experiments. The Medina dialect ASR system exploited hidden Markov models (HMM).
[83]	2023	In this article, three well-performing CNN architectures are identified. The accuracy of AlexNet, ResNet, and GoogLeNet was respectively 86.19%, 83.46%, and 89.61%. The results demonstrated the superiority of GoogLeNet, emphasizing the potential of CNN architectures for modeling low-resource acoustic features in languages such as Arabic.
[84]	2019	The research aims to compare the WAT-MFCC and SVM systems for isolated speech classification in a noisy environment, and the results were better than those of MLP and HMM techniques.
[85]	2020	We conclude from this work that CNN can recognize isolated speech with an accuracy of 98.5%.
[86]	2020	In this research, the researcher compared the results extracted from comparing the following techniques using clean and noisy environments: MFCC+GMM at 95.39%, MFCC+VQ at 90.04% and the best results were obtained from the combination of MFCC+GMM+VQ at 97.92%.
[87]	2018	The study aimed to recognize speech using isolated words in Malay, Majeh, and Dua dialects, and the results varied in accuracy at the speech recognition level due to individual characteristics and dialect differences.

6. Conclusion

This paper presents an in-depth study of the automatic recognition of isolated Arabic speech, speech command recognition, and the applications used in this domain. The review aims to identify the most significant literary methodologies and key studies published between 2018 and 2024. Critical research from prestigious journals on Scopus and conferences across eight leading databases-Web of Science, IEEE Xplore, ResearchGate, Springer Link, ScienceDirect, ACM, Google Scholar, and ACADEMIA- was thoroughly examined.

The paper begins with an introduction, followed by a detailed description of the study's strategies and the formulation of key research questions. The methodology section elaborates on the studies reviewed, identifying the strengths and weaknesses of each, and highlights the challenges and gaps specific to Arabic speech recognition, especially with respect to Arabic dialects. The paper also discusses future trends and suggests possible solutions to the problems encountered in the existing studies. The findings emphasize the need for continued research to address these issues, enabling future advancements in Arabic speech

recognition systems, particularly within artificial intelligence applications. The analysis of research methodologies revealed several important insights:

- Increasing the data size for training and testing enhances performance, with larger datasets yielding better results in shorter timeframes.
- A diverse dataset representing various Arabic dialects from different regions is essential for achieving comprehensive and accurate recognition.
- Adding noise to speech samples has been shown to reduce the error rate, highlighting the importance of noise robustness.
- Incorporating new features and methods for feature extraction can significantly improve recognition outcomes.
- Voice command recognition systems are pivotal for the automatic control of smart devices, further demonstrating the applicability of speech recognition in real-world settings.

These findings underscore the need for ongoing innovation and research to enhance the performance and quality of isolated Arabic speech recognition systems, particularly in the context of Arabic dialects and their applications in emerging artificial intelligence technologies.

References

[1] Vlado Delić et al., "Speech Technology Progress Based on New Machine Learning Paradigm," *Computational Intelligence and Neuroscience*, vol. 2019, no. 1, pp. 1-19, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[2] Siddique Latif et al., "Deep Representation Learning in Speech Processing: Challenges, Recent Advances, and Future Trends," *Arxiv*, pp. 1-25, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Yogesh Kumar, and Navdeep Singh, "A Comprehensive View of Automatic Speech Recognition System: A Systematic Literature Review," *International Conference on Automation, Computational and Technology Management*, London, UK, pp. 168-173, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[4] Anjuli Kannan et al., "Large-Scale Multilingual Speech Recognition with a Streaming End-To-End Model," *arXiv Preprint*, pp. 2130-2134, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[5] Chung-Cheng Chiu et al., "State-Of-The-Art Speech Recognition with Sequence-To-Sequence Models," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Calgary, AB, Canada, pp. 1-5, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[6] Jing-Xuan Zhang et al., "Sequence-To-Sequence Acoustic Modeling for Voice Conversion," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 3, pp. 631-644, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[7] Garima Sharma, Kartikeyan Umapathy, and Sridhar Krishnan, "Trends in Audio Signal Feature Extraction Methods," *Applied Acoustics*, vol. 158, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Parashar Dhakal et al., "A Near Real-Time Automatic Speaker Recognition Architecture for Voice-Based User Interface," *Machine Learning and Knowledge Extraction*, vol. 1, no. 1, pp. 504-520, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[9] Motaz Hamza, Touraj Khodadadi, and Sellappan Palaniappan, "A Novel Automatic Voice Recognition System Based on Text-Independent in A Noisy Environment," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 4, pp. 3643-3650, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Suvarnsing Bhable, Ashish Lahase, and Santosh Maher, "Automatic Speech Recognition (ASR) of Isolated Words in Hindi Low-Resource Language," *International Journal for Research in Applied Science & Engineering Technology*, vol. 9, no. 2, pp. 260-265, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Ankita Wadhawan, and Parteek Kumar, "Deep Learning-Based Sign Language Recognition System for Static Signs," *Neural Computing and Applications*, vol. 33, no. 12, pp. 7957-7968, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Imane Guellil et al., "Arabic Natural Language Processing: An Overview," *Journal of King Saud University-Computer and Information Sciences*, vol. 33, no. 5, pp. 497-507, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Samira Hazmoune et al., "A New Hybrid Framework Based on Hidden Markov Models and K-Nearest Neighbors for Speech Recognition," *International Journal of Speech Technology*, vol. 21, no. 3, pp. 689-704, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[14] Abdelmadjid Benmachiche, and Amina Makhlouf, "Optimization of Hidden Markov Model with Gaussian Mixture Densities for Arabic Speech Recognition," *WSEAS Transactions on Signal Processing*, vol. 15, pp. 85-94, 2019. [Google Scholar] [Publisher Link]

[15] Vishal Passricha, and Rajesh Kumar Aggarwal, "End-To-End Acoustic Modeling Using Convolutional Neural Networks," *Intelligent Speech Signal Processing*, pp. 5-37, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[16] Marvin Coto-Jiménez, "Improving Post-Filtering of Artificial Speech Using Pre-Trained LSTM Neural Networks," *Biomimetics*, vol. 4, no. 2, pp. 1-17, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[17] Brahim Fares Zaidi et al., "Deep Neural Network Architectures for Dysarthric Speech Analysis and Recognition," *Neural Computing and Applications*, vol. 33, no. 15, pp. 9089-9108, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[18] Kateřina Žmolíková et al., "Speakerbeam: Speaker Aware Neural Network for Target Speaker Extraction in Speech Mixtures," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 4, pp. 800-814, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] Mutian Zhuang, "*New Features for Speech Processing Standard Pronunciation Classification*," Master Thesis, Northern Illinois University, 2021. [Google Scholar]

[20] Jinchuan Tian et al., "Improving Mandarin End-To-End Speech Recognition with Word N-Gram Language Model," *IEEE Signal Processing Letters*, vol. 29, no. 8, pp. 812-816, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] T. Serrenho, and P. Bertoldi, "*Smart Home and Appliances: State of The Art*," European Commision, Technical Report, pp. 1-59, 2019. [Google Scholar]

[22] Chao-Han Huck Yang et al., "Decentralizing Feature Extraction with Quantum Convolutional Neural Network for Automatic Speech Recognition," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Toronto, ON, Canada, pp. 6523-6527, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[23] Roshna Muhamad M. Amin, and Miran H.M. Baban, "Implementation of Controlling Home Appliances Via Secure Short Message Service Technology," *Tikrit Journal of Engineering Science*, vol. 28, no. 1, pp. 21-30, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[24] Pete Warden, "Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition," *Arxiv*, pp. 1-11, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[25] Dragos Mocrii, Yuxiang Chen, and Petr Musilek, "Iot-Based Smart Homes: A Review of System Architecture, Software, Communications, Privacy and Security," *Internet of Things*, vol. 1-2, pp. 81-98, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[26] Marcin Woźniak, and Dawid Połap, "Intelligent Home Systems for Ubiquitous User Support by Using Neural Networks and Rule-Based Approach," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2651-2658, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[27] Yu Xiao, and Maria Watson, "Guidance on Conducting a Systematic Literature Review," *Journal of Planning Education and Research*, vol. 39, no. 1, pp. 93-112, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[28] Debajyoti Pal et al., "Analyzing the Elderly Users' Adoption of Smart-Home Services," *IEEE Access*, vol. 6, pp. 51238-51252, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[29] Berrak Sisman et al., "An Overview of Voice Conversion and Its Challenges: From Statistical Modeling to Deep Learning," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 132-157, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[30] Mahmood S. Mahmood, and Najla B. Al Dabagh, "Improving IOT Security Using Lightweight Based Deep Learning Protection Model," *Tikrit Journal of Engineering Science*, vol. 30, no. 1, pp. 119-129, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[31] Ali Bou Nassif et al., "Speech Recognition Using Deep Neural Networks: A Systematic Review," *IEEE Access*, vol. 7, pp. 19143-19165, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[32] Mahmoud A. Elawi, Noori H. Noori, and Auday T.S. Al-Bayati, "Comparison Between the Relations of HpGe Detector Efficiency Curve and Background 'Spectrum Shape'," *Tikrit Journal of Pure Science*, vol. 26, no. 1, pp. 96-100, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[33] Haoye Lu, Haolong Zhang, and Amit Nayak, "A Deep Neural Network for Audio Classification with A Classifier Attention Mechanism," *Arxiv*, pp. 1-15, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[34] Yishuang Ning et al., "Review of Deep Learning Based Speech Synthesis," *Applied Sciences*, vol. 9, no. 19, pp. 1-16, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[35] Salima Harrat, Karima Meftouh, and Kamel Smaïli, "Maghrebi Arabic Dialect Processing: An Overview," *Journal of International Science and General Applications*, vol. 1, no. 1, pp. 1-8, 2018. [Google Scholar] [Publisher Link]

[36] Zhaojuan Song, "English Speech Recognition Based on Deep Learning with Multiple Features," *Computing*, vol. 102, no. 3, pp. 663-682, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[37] Mine Büşra Gelen, and Alparslan Serhat Demir, "Selection of Information Technology Personnel for an Enterprise in the Process of Industry 4.0 with the MultiMoora Method," *Sakarya University Journal of Science*, vol. 23, no. 43328, pp. 663-675, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[38] Apeksha Shewalkar, Deepika Nyavanandi, and Simone A. Ludwig, "Performance Evaluation of Deep Neural Networks Applied to Speech Recognition: RNN, LSTM and GRU," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 9, no. 4, pp. 235-245, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[39] Mohammad Ayache et al., "Speech Command Recognition Using Deep Learning," *2021 Sixth International Conference on Advances in Biomedical Engineering*, Werdanyeh, Lebanon, pp. 24-29, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[40] Anand Umashankar, "*Large Scale Speech Recognition with Deep Learning*," Master Thesis, Aalto University, pp. 1-83, 2021. [Google Scholar] [Publisher Link]

[41] Samer Alsammarraie, and Nazar K. Hussein, "A New Hybrid Grasshopper Optimization-Backpropagation for Feedforward Neural Network Training," *Tikrit Journal of Pure Science*, vol. 25, no. 1, pp. 118-127, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[42] Tanvira Ismail, "A Survey of Language and Dialect Identification Systems," *Adalya Journal*, vol. 9, no. 1, pp. 682-690, 2020. [Google Scholar] [Publisher Link]

[43] Aalaa Ahmed Mohammed, and Yusra Faisal Al-Irhayim, "Speaker Age and Gender Estimation Based on Deep Learning Bidirectional Long-Short Term Memory (BiLSTM)," *Tikrit Journal of Pure Science*, vol. 26, no. 4, pp. 76-84, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[44] Dwi Sari Widyowaty, Andi Sunyoto, and Hanif Al Fatta, "Accent Recognition Using Mel-Frequency Cepstral Coefficients and Convolutional Neural Network," *Proceedings of the International Conference on Innovation in Science and Technology*, pp. 43-46, 2021. [Google Scholar] [Publisher Link]

[45] Mohammed Jawad Al Dujaili, Abbas Ebrahimi-Moghadam, and Ahmed Fatlawi, "Speech Emotion Recognition Based on SVM and KNN Classifications Fusion," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 2, pp. 1259-1264, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[46] Thibault Viglino, Petr Motlicek, and Milos Cernak, "End-To-End Accented Speech Recognition," *Interspeech*, pp. 2140-2144, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[47] Shreyas Ramoji, and Sriram Ganapathy, "Supervised I-Vector Modeling for Language and Accent Recognition," *Computer Speech & Language*, vol. 60, pp. 1-39, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[48] Samantha Wray, "Classification of Closely Related Sub-Dialects of Arabic Using Support-Vector Machines," *Proceedings of the 11th International Conference on Language Resources and Evaluation*, pp. 3671-3674, 2018. [Google Scholar] [Publisher Link]

[49] Radek Martinek et al., "Voice Communication in Noisy Environments in A Smart House Using Hybrid LMS+ICA Algorithm," *Sensors*, vol. 20, no. 21, pp. 1-24, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[50] Hannah Snyder, "Literature Review as A Research Methodology: An Overview and Guidelines," *Journal of Business Research*, vol. 104, pp. 333-339, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[51] Matthew J. Page et al., "Updating Guidance for Reporting Systematic Reviews: Development of the PRISMA 2020 Statement," *Journal of Clinical Epidemiology*, vol. 134, pp. 103-112, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[52] Catrin Sohrabi et al., "PRISMA 2020 Statement: What's New and the Importance of Reporting Guidelines," *International Journal of Surgery*, vol. 88, pp. 39-42, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[53] Xin Huang et al., "Synthesizing Qualitative Research in Software Engineering: A Critical Review," *Proceedings of the 40th International Conference on Software Engineering*, Gothenburg Sweden, pp. 1207-1218, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[54] Hayrol Azril Mohamed Shaffril, Samsul Farid Samsuddin, and Asnarulkhadi Abu Samah, "The ABC of Systematic Literature Review: The Basic Methodological Guidance for Beginners," *Quality and Quantity*, vol. 55, no. 4, pp. 1319-1346, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[55] Quan Nha Hong, and Pierre Pluye, "A Conceptual Framework for Critical Appraisal in Systematic Mixed Studies Reviews," *Journal of Mixed Methods Research*, vol. 13, no. 4, pp. 446-460, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[56] Bezoui Mouaz, Beni Hssane Abderrahim, and Elmoutaouakkil Abdelmajid, "Speech Recognition of Moroccan Dialect using Hidden Markov Models," *Procedia Computer Science*, vol. 151, pp. 985-991, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[57] Naima Zerari et al., "Bidirectional Deep Architecture for Arabic Speech Recognition," *Open Computer Science*, vol. 9, no. 1, pp. 92-102, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[58] Bilal Dendani, Halima Bahi, and Toufik Sari, "Speech Enhancement Based on Deep AutoEncoder for Remote Arabic Speech Recognition," *Image and Signal Processing: 9th International Conference*, Marrakesh, Morocco, pp. 221-229, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[59] Fawaz S. Al-Anzi, "Improved Noise-Resilient Isolated Words Speech Recognition Using Piecewise Differentiation," *Fractals*, vol. 30, no. 8, pp. 1-12, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[60] Lubna Eljawad et al., "Arabic Voice Recognition Using Fuzzy Logic and Neural Network," *International Journal of Applied Engineering Research*, vol. 14, no. 3, pp. 651-662, 2019. [Google Scholar] [Publisher Link]

[61] Yousef A. Alotaibi et al., "A Canonicalization of Distinctive Phonetic Features to Improve Arabic Speech Recognition," *Acta Acustica United with Acustica*, vol. 105, no. 6, pp. 1269-1277, 2019. [Google Scholar]

[62] Lotfi Boussaid, and Mohamed Hassine, "Arabic Isolated Word Recognition System Using Hybrid Feature Extraction Techniques and Neural Network," *International Journal of Speech Technology*, vol. 21, pp. 29-37, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[63] Lina Tarek Benamer, and Osama A.S. Alkishriwo, "Database for Arabic Speech Commands Recognition," *3rd Conference for Engineering Sciences and Technology*, 2020. [Google Scholar] [Publisher Link]

[64] Wafa Helali, Zied Hajaiej, and Adnane Cherif, "Arabic Corpus Implementation: Application to Speech Recognition," *2018 International Conference on Advanced Systems and Electric Technologies*, Hammamet, Tunisia, pp. 50-53, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[65] Naima Zerari et al., "Bi-Directional Recurrent End-To-End Neural Network Classifier for Spoken Arab Digit Recognition," *2018 2nd International Conference on Natural Language and Speech Processing*, Algiers, Algeria, pp. 1-6, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[66] Abdelkader Guerid, and Amrane Houacine, "Recognition of Isolated Digits Using DNN-HMM and Harmonic Noise Model," *IET Signal Processing*, vol. 13, no. 2, pp. 207-214, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[67] Nourredine Oukas et al., "ArabAlg: A New Dataset for Arabic Speech Commands Recognition for Machine Learning Purposes," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 989-1005, 2024. [Google Scholar]

[68] Ahmed Boumehdi, and Abdellah Yousfi, "Construction of a Database for Speech Recognition of Isolated Arabic Words," *Proceedings of the 13th International Conference on Intelligent Systems: Theories and Applications*, New York, United States, pp. 1-4, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[69] Hani S. Hassan, S. Jammila Harbi, and Maisa'a Abid Ali Kodher, "RETRACTED: Arabic Command Based Human Computer Interaction," *International Conference for Modern Applications of Information and Communication Technology*, Baghdad, IRAQ, vol. 1530, no. 1, pp. 1-12, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[70] Shaker K. Ali, and Zahraa M. Mahdi, "Arabic Voice System to Help Illiterate or Blind for Using Computer," *Journal of Physics: Conference Series: International Conference of Modern Applications on Information and Communication Technology*, Babylon-Hilla City, Iraq, vol. 1804, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[71] Amjad Rehman Khan, "Automatic Gender Authentication from Arabic Speech Using Hybrid Learning," *Journal of Advances in Information Technology*, vol. 15, no. 4, pp. 532-543, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[72] Abdulaziz Saleh Mahfoudh Ba Wazir, and Joon Huang Chuah, "Spoken Arabic Digits Recognition Using Deep Learning," *2019 IEEE International Conference on Automatic Control and Intelligent Systems*, Selangor, Malaysia, pp. 339-344, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[73] Abdelkbir Ouisaadane, and Said Safi, "A Comparative Study for Arabic Speech Recognition System in Noisy Environments," *International Journal of Speech Technology*, vol. 24, pp. 761-770, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[74] Engy Ragaei Abdelmaksoud et al., "Convolutional Neural Network for Arabic Speech Recognition," *Egyptian Journal of Language Engineering*, vol. 8, no. 1, pp. 27-38, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[75] Rafik Amari et al., "Deep Convolutional Neural Network for Arabic Speech Recognition," *14th International Conference on Computational Collective Intelligence*, Hammamet, Tunisia, pp. 120-134, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[76] Hiba Qasim Jaber, and Huda Abdulaali Abdulbaqi, "Real Time Arabic Speech Recognition Based on Convolution Neural Network," *Journal of Information and Optimization Sciences*, vol. 42, no. 7, pp. 1657-1663, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[77] Asroni Asroni et al., "Arabic Speech Classification Method Based on Padding and Deep Learning Neural Network," *Baghdad Science Journal*, vol. 18, no. 2, pp. 925-936, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[78] Mohamed Hassine, Lotfi Boussaid, and Hassani Massaoud, "Tunisian Dialect Recognition Based on Hybrid Techniques," *International Arab Journal of Information Technology*, vol. 15, no. 1, pp. 58-65, 2018. [Google Scholar] [Publisher Link]

[79] Moner N.M. Arafa et al., "A Dataset for Speech Recognition to Support Arabic Phoneme Pronunciation," *International Journal of Image, Graphics and Signal Processing*, vol. 10, no. 4, pp. 31-38, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[80] Abdulmalik A. Alasadi et al., "Efficient Feature Extraction Algorithms to Develop an Arabic Speech Recognition System," *Engineering, Technology & Applied Science Research*, vol. 10, no. 2, pp. 5547-5553, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[81] Bilal Dendani, Halima Bahi, and Toufik Sari, "Self-Supervised Speech Enhancement for Arabic Speech Recognition in Real-World Environments," *Signal Processing*, vol. 38, no. 2, pp. 349-358, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[82] Haneen Bahjat Khalafallah, Mohamed Abdel Fattah, and Ruqayya Abdulrahman, "Speech Corpus for Medina Dialect," *Journal of King Saud University-Computer and Information Sciences*, vol. 36, no. 2, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[83] Zoubir Talai, Nada Kherici, and Halima Bahi, "Comparative Study of CNN Structures for Arabic Speech Recognition," *Information Systems Engineering*, vol. 28, no. 2, pp. 327- 333, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[84] Mohamed Walid et al., "Real-Time Implementation of Isolated-Word Speech Recognition System on Raspberry Pi 3 Using WAT-MFCC," *International Journal of Computer Science and Network Security*, vol. 19, no. 3, pp. 42-49, 2019. [Google Scholar] [Publisher Link]

[85] Jihad Anwar Qadir, Abdulbasit K. Al-Talabani, and Hiwa A. Aziz, "Isolated Spoken Word Recognition Using One-Dimensional Convolutional Neural Network," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 20, no. 4, pp. 272-277, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[86] Abdelkbir Ouisaadane, Said Safi, and Miloud Frikel, "Arabic Digits Speech Recognition and Speaker Identification in Noisy Environment Using a Hybrid Model of VQ and GMM," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 4, pp. 2193-2204, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[87] Shipra J. Arora, "Dialectal Variations of Isolated Word Recognition," *INFOCOMP Journal of Computer Science*, vol. 17, no. 1, pp. 38-44, 2018. [Google Scholar] [Publisher Link]