

# Artificial Intelligence Model to Detect and Classify Arabic Dialects

Iman S. Alansari

Computer Science Department, College of Computer Science and Engineering, Taibah University, Madinah, Saudi Arabia  
Email: iansari@taibahu.edu.sa

**How to cite this paper:** Alansari, I.S. (2023) Artificial Intelligence Model to Detect and Classify Arabic Dialects. *Journal of Software Engineering and Applications*, 16, 287-300.  
<https://doi.org/10.4236/jsea.2023.167015>

**Received:** June 28, 2023

**Accepted:** July 28, 2023

**Published:** July 31, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

The Arabic Dialect (AD) detection method involves analyzing the matching sound wave for various characteristics that identify the speaker's dialect. Among these features are accent, intonation, stress, vowel length, vowel type, and other acoustic characteristics. Data from different speakers of different dialects is usually used in training machine learning algorithms. Based on this data, an algorithm is created to accurately identify the speaker's dialect. Arabic dialects can be detected and classified using several models and techniques available in literature. Various models have been proposed from different perspectives. Therefore, this paper discussed different studies about AD for building an understanding of conceptual deep learning model to detect and classify Arabic dialects. The model captured the semantic, syntactic, and phonological characteristics of these dialects using Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The proposed model consists of six stages: Natural Language Processing (NLP) stage, feature engineering techniques, neural networks, language models, optimization techniques, and evaluation techniques. Each stage of the proposed model has several techniques that can be used to detect and classify AD. The accuracy and capability of the proposed model will be performed in the future work.

## Keywords

AI, Arabic Dialect, CNN, RNN

## 1. Introduction

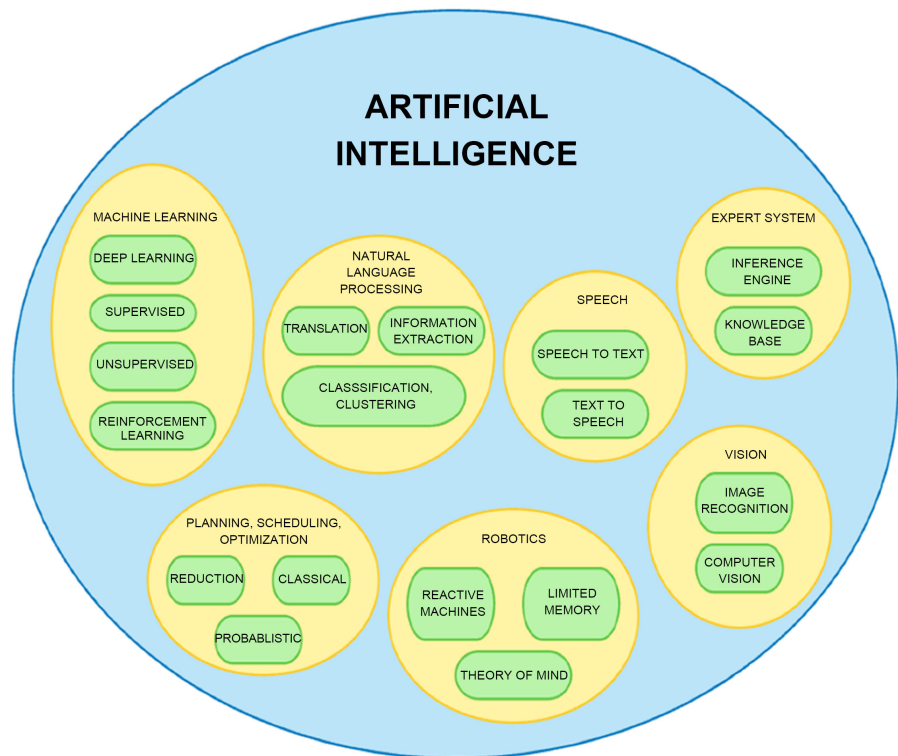
There are many varieties of Arabic languages spoken around the world, including Modern Standard Arabic (MSA), which is one of the most spoken languages in the world. There are several ways in which MSA is used in a formal commu-

nication as well as written Arabic language, including culture, media, and education in the Arabic world. MSA is, however, not the native language of the Arabic people all over the world since Arabic has a wide variety of dialects depending on the region in which it is spoken in the country. Additionally, Arabic is a religion language that includes a wide variety of religious discourses, ranging from historical and literature texts [1] [2] [3].

The Arab world uses email more than any other communication method, including diary entries, dialogues, and social media platforms like Twitter. Dialectal Arabic is usually used in these communications. Media, education, and government use Modern Standard Arabic (MSA), the most commonly used Arabic form [4]. Arabic dialects are mostly used in informal settings for everyday communication, such as during interactions with friends and family [4]. Several verbal dialects of Arabic, along with other terms native to the region, are also associated with this language. The Arabic dialect used by Arab nations serves as a means of interacting with various dialects of the language spoken by Arab nations, while also serving as a shared language among them [5]. There are many Arabic dialects in the Arabic-speaking world that have become native languages of speakers. AD is one of these dialects. Despite many semantic, syntactic, morphological, and lexical similarities that Arabic dialects share with Modern Standard Arabic varieties (MSA), they exhibit significant differences across almost all language subsystems.

An artificial intelligence system refers to a mechanism or process that simulates human intelligence processes using machines, specifically computers [6]. Several processes contribute to the learning process, such as learning (the acquisition of information and rules for using it), reasoning (application of the rules to reach approximate or definite conclusions), and self-correction [7] [8]. AI is used in many different applications, such as expert systems, speech recognition, and machine vision, as shown in **Figure 1**. There are several fields in which artificial intelligence is used, such as medicine, finance, and engineering. This technology has numerous applications, such as enhancing the efficiency of tasks and processes, as well as creating more accurate and reliable results. In addition to automating manual processes, it can also be used to improve customer service and improve customer relationships. The implementation of artificial intelligence has facilitated the automation of repetitive and tedious tasks in businesses, thus leading to significant cost and time savings. Additionally, AI can be used to make predictions about future events and identify patterns in the data for the purpose of future predictions [9] [10]. Nevertheless, AI has been used in software engineering for bug classification [11] and source code summarization. Bug classification helps identify and classify bug reports, so developers can prioritize and address critical reports first. Source code summarization is the task of creating a concise and accurate summary of code, which can be used as a reference by developers [12].

Thus, the purpose of this research is to build a conceptual deep learning model that can be used to develop a method for determining and classifying dialects



**Figure 1.** Artificial intelligence techniques based on evolutionary taxonomy [13].

of Arabic. Therefore, this proposed model identified and classified Arabic dialects according to their characteristics. This model is expected to accurately predict Arabic dialects based on text in the dataset.

This study proposed a conceptual deep-learning model that can be used to detect and classify Arabic dialects. This model combined Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to capture the semantic, syntactic, and phonetic features of dialects. CNNs can be used to extract semantic and syntactic features, such as word and phrase structures, from dialects. RNNs can be used to extract phonetic features, such as intonation, pronunciation, and stress, from dialects. Subsequently, the model can be trained using a dataset of dialectal audio recordings.

The proposed model exhibits more accurate and efficient way to detect and classify Arab dialects than traditional approaches. However, the scope of this study is limited to the development of the conceptual deep learning model, whereas the experimental data will be discussed in future work.

This paper is structured as follows: Section 2 presents related works, and Section 3 discusses the methodology. The results and discussion are presented in Section 4. The conclusions and future work of this study are discussed in Section 5.

## 2. Related Works

Several security assessment models and framework have been proposed in the

literature. For example, the authors in [14] investigated ANSI/API Security Risk Assessment methods for developing proactive security policies that respect threats. The researchers in [15] proposed a security framework which consists of four stages: risk assessment, implementation of security risk controls, monitoring and review. The researchers in [16] introduced a security risk management model for Saudi organizations. Several factors were proposed in this model such the security policies and process, the size and culture of the organizations. The authors in [17] proposed a model for communicating and implementing information security policies. They examined present information security policy progress and methods from secondary sources. This was to gain a deeper understanding of the processes critical to the development and life cycle of information security policies. As part of the proposed model, the various steps involved in developing, implementing, and evaluating an effective information security policy are described.

There is growing interest in identifying Arabic dialects as a new research area in the fields of Machine Learning (ML) and Natural Language Processing (NLP). Dialectal Arabic text (DA) has become an important means of informal communication online because of the substantial use of dialectal Arabic text. In this new context, the emergence of a new form of textual data online has been inseparably mixed with the standard MSA. This research work is mainly focused on constructing dialectal corpora [18], in addition to differentiating between MSA and the varieties of DA in terms of language classification.

The authors in [18] have described a general approach for recognizing DA text in Modern Standard Arabic (MSA) text in a small sample size of 1.6K sentences, which was used as an example of DA text recognition. As a result of the project, they produced annotated text at the sentence and word level.

It was reported by [19] that the percentage of DA in MSA could be measured. In addition, further research on DA has been greatly enriched by the construction of the public Arabic Online Commentary (AOC) dataset, which was developed by [20]. The dataset was analyzed based on data collected from online Arabic forums. The authors used crowdsourcing to classify it into the following six classifications: MSA, Egyptian, Gulf, Iraq, Levantine, and Maghrebi, as well as three additional labels.

The authors in [21] utilized a small subset of the AOC to identify whether a given text was either MSA or DA (the Egyptian dialect, *i.e.*, EGP) using a supervised SVM classifier to determine the status of the text. The accuracy of the SVM classifier was 85.5%.

Furthermore, in [22], researchers used a dataset consists of five Arabic dialects, namely Levantine, Gulf, Egyptian, Maghrebi, and Iraqi, and achieved a results of 81% for accuracy.

Similar subset of AOC was employed in another study [23] using more features, and the results were able to significantly boost the performance of the classifier while producing a binary output (*i.e.*, MSA or EGP).

The authors in [24] produced better results for their binary classifier when

compared to n-grams when they used lexical and morphological features rather than just n-grams in their study on Twitter data.

According to [25], a number of different classifiers based on bigrams using Naive Bayes have been proposed to characterize 18 different Arabic dialects with an overall F1-measure of approximately 78% have been achieved.

Similar work was proposed in [26] to classify four dialects: Levantine, Egyptian, Saudi, and Iraqi. It consists of two parts, each containing different types of data. In the first case, there is a small corpus of manually annotated instances, and in the second case, there is a large collection of words crawled from the Web using wordmarks.

The authors in [27] used deep learning techniques such as word2vec in order to be able to identify the equivalent words from different dialects of the Arab world. In this research, a set of Arabic dialect corpora was merged and extended, followed by the application of deep learning techniques to achieve the best results with dialectal word synonyms based on the existing Arabic dialect corpora.

In [28], the authors investigated an end-to-end model to improve Dialectal Arabic Speech Recognition (DASR) based on deep learning to improve the accuracy of the speech recognition process. This approach is based on a hybrid model composed of a Convolutional Neural Network (CNN) in conjunction with Long Short-Term Memory (LSTM), which is referred to as (CNN-LSTM).

The authors in [29] introduced a novel transcribed corpora for Yamani Arabic, Jordanian Arabic, and multi-dialect Arabic. It was also found that several baseline deep neural models were developed for end-to-end recognition of Arabic dialects using sequence-to-sequence deep neural networks.

The authors of [30] examined different approaches for identifying dialects in Arabic broadcast speech. Using a speech recognition system, phonetic and lexical features were obtained, as well as bottleneck features, using an i-vector framework. Their study examined both generative and discriminative classifiers, and a multi-class Support Vector Machine (SVM) was used to combine both features.

An algorithm based on Recurrent Neural Networks (RNN) was developed in [31] to classify dialects based on four dialect regions, namely the Maghreb, Levantine, Gulf (along with Iraqi), and Nile. In this study, a dataset was retrieved from the MADAR corpus, consisting of more than 110,000 sentences.

According to [32], researchers trained a dialect classification model using the AOC dataset. The authors used six deep learning methods and found that bidirectional RNNs classified Arabic text into MSA, Egyptian, Gulf, and Levantine texts with a total of 82.45% accuracy.

A novel Arabic resource with dialect annotations was described by [33]. The findings resulted in a large monolingual dataset called the Arabic Online Commentary Dataset, which contains an extensive collection of dialectal Arabic content.

Researchers in [34] classified Arabic Facebook posts using two approaches. First, syntactic features were used to express opinions using patterns common to

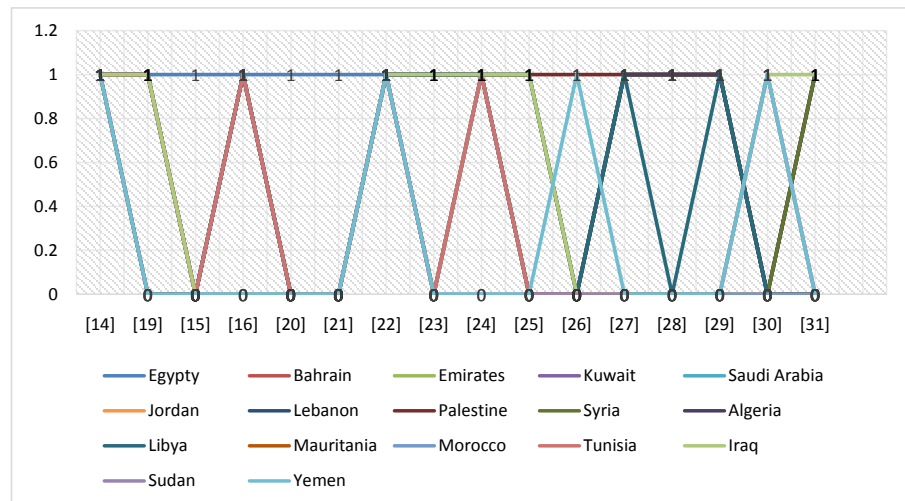
different Arabic dialects. This pattern was highly accurate even when tested against a new corpus. The focus of this approach was on informal Arabic texts, which has never been studied before.

An Arabic Dialect Dataset (ADD), which consists of a corpus of audio samples related to healthcare issues, was developed and made publicly available by [35] [36] [37]. The datasets developed for this project have multiclass labelling. This study revealed that the datasets were subject to bias as there may be lack of training samples for several classes. **Table 1** summarizes the existing studies on AD.

As a part of this study, existing studies about AD have been reviewed using different datasets such as Egypt, Gulf, Moghrabi, Levantine, Iraqi, Yemeni, and Sudani datasets as shown in **Figure 2**. It can be concluded that Arabic dialects have many differences in terms of pronunciation, lexical terms, syntactic structure, and morphological features. Therefore, there is a significant lexical difference between Arabic dialects. In this region, the local language has a significant influence on lexical variation. Turkish, Kurdish, and Persian are the native languages of Levant and North Africa, respectively. In syntactic structures, words are ordered differently, and sentences are used differently [38] [39]. The dialects of the Arabian Peninsula tend to use a conservative word order. A further difference between dialects is morphological, with Levantine and North African dialects being more conservative than those of the Arabian Peninsula [40] [41] [42]. Moreover, some studies have examined automatic dialect identification, which involves classifying text into one of various dialects of Arabic using Decision Trees (DT) and Support Vector Machines (SVM). In addition, Arabic dialect language models have also been developed.

**Table 1.** Summarizes the existing studies on AD.

Ref	Year	Arabic Dialect Dataset																					
		Egypt																Iraq		Sudan		Yemen	
		Gulf		Levantine				Moghrabi				Iraq		Sudan		Yemen							
		Bahrain	Emirates	Kuwait	Oman	Qatar	Saudi Arabia	Jordan	Lebanon	Palestine	Syria	Algeria	Libya	Mauritania	Morocco	Tunisia							
[18]	2008	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑			
[20]	2011	☑	☑	☑	☑	☑	☑	☑	☒	☑	☒	☒	☑	☑	☑	☑	☑	☑	☑	☒	☒		
[21]	2013	☑	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒		
[22]	2014	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☒	☒		
[23]	2014	☑	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒		
[24]	2016	☑	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒		
[25]	2014	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑		
[26]	2016	☑	☒	☒	☒	☒	☒	☑	☑	☑	☑	☒	☒	☒	☒	☒	☒	☑	☒	☒	☒		
[27]	2022	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☒	☒		
[28]	2022	☑	☒	☒	☒	☒	☒	☑	☑	☑	☑	☒	☒	☒	☒	☒	☒	☑	☒	☒	☒		
[29]	2023	☒	☒	☒	☒	☒	☒	☒	☑	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☑		
[30]	2015	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒		
[31]	2018	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒		
[32]	2018	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒		
[33]	2014	☒	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☒	☒	☒	☒	☒	☒	☒	☒		
[34]	2012	☑	☒	☒	☒	☒	☒	☒	☒	☑	☑	☑	☒	☒	☒	☒	☒	☑	☒	☒	☒		



**Figure 2.** Summaries of the existing studies on AD.

### 3. Methodology

This study aims to develop an integrated scalable deep learning model align with the design science methodology [43] [44]. This model can be employed to effectively identify and categories Arabic dialects. A conceptual model was developed through an in-depth review of the existing literature on deep learning, Arabic dialects, and design science research with the aim to create a conceptual model that can be used in the future. For this purpose, a model has been developed to capture the characteristics and distinction of the linguistic features of Arabic dialects in a system. It involves six stages of deep learning as shown in **Figure 3**.

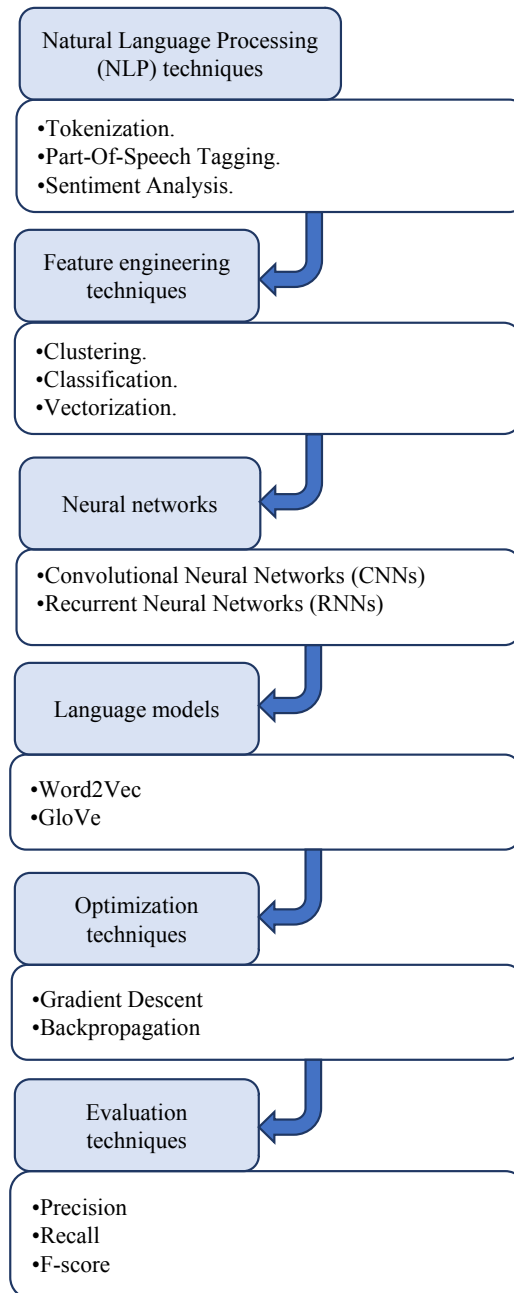
**1) Natural Language Processing (NLP) Methods:** include part-of-speech tagging, tokenization, and sentiment analysis:

a) Tokenization: tokenization of a text stream is the process of breaking it down into meaningful symbols, words, or phrases, called tokens, to digest the data. For example, an Arabic sentence like “ذهب الطالب الى المدرسة” can be tokenized into “ذهب”, “الطالب”, “الى”, “المدرسة”.

b) Part-of-speech Tagging: the part-of-speech tagging is used to determine which word in a text belongs to a particular part of speech. For example, an Arabic sentence like “ذهب الطالب الى المدرسة” can be tagged as follows: “ذهب” (verb), “الطالب” (noun), “الى” (preposition), “المدرسة” (noun).

c) Sentiment Analysis: this process is called sentiment analysis, and it entails identifying and extracting subjective information from texts using a set of statistical algorithms. For example, an Arabic sentence such as “ذهب الطالب الى المدرسة” can be given a sentiment score of “positive”, since it expresses a positive action (going to school).

**2) Feature Engineering Techniques:** this method applies feature engineering techniques to extract semantic information from Arabic dialects for better understanding. Examples of this technique are clustering, classification, and vectorization.



**Figure 3.** Conceptual deep learning model that can be used to detect and classify Arabic dialects.

a) Clustering: A clustering technique involves dividing data points into groups based on their similarity using features generated through feature engineering. It can be applied to categorize similar dialects into clusters based on their similarities based on the vocabulary, syntax, and pronunciation used in each dialect of Arabic.

b) Classification: Classification techniques are used to classify data points based on their characteristics when feature engineering is used to assign labels to them. For example, the origins or relationships among other dialects can be se-



parated by classifying Arabic dialects based on the region where they originated.

c) Vectorization: Features are converted into numerical vectors by vectorizing text using features. In this case, machine learning models can be trained using inputs created from Arabic sentences by vectorizing them and converting them into numerical vectors via vectorization.

### 3) Recognizing and Choosing Security Risk Assessment Models

a) Convolutional Neural Networks (CNNs): To classify Arabic dialects, CNNs extract several features from textual data. Using a CNN, the presence of phonemes, syllables, and other linguistic features in text data can be detected to classify dialects.

b) Recurrent Neural Networks (RNNs): In machine learning, RNN analyze the context of words within sentences. In addition to detecting sentence structures, RNNs can also detect syntactic rules to classify them based on dialects. Therefore, RNNs may be able to detect gender agreement. Besides, computer algorithms can be applied for dialect classification.

c) The language models: Word2Vec and GloVe are two programs available in the market that represent dialect-specific words and phrases in Arabic dialects. Natural language processing (NLP) uses popular language models such as Word2Vec and GloVe. Examples of such tasks include machine translation, sentiment analysis, and text classification. These models are designed to capture the complex structure and dialects of the Arabic language. Several examples have been provided to illustrate this point.

d) Word2Vec: Using a shallow neural network, Word2Vec embeds words in a vector format. Embeddings for words can be created using several Arabic dialects. There are three main dialects of Arabic: Standard Arabic, Egyptian Arabic, and Gulf Arabic. From this model, the parameters are devised.

e) GloVe: As part of the global log-bilinear regression model, GloVe represents words as vectors rather than as words embedded in matrices. Besides Maghrebi Arabic, Levantine Arabic, and Iraqi Arabic, this model can also be used to create language embeddings for a wide variety of Arabic dialects.

**4) Optimization Techniques** as a rule, there are two main categories of Arabic dialect models used for training:

a) Gradient Descent: To minimize the error between the predicted output and the targeted output, the parameters of the model are adjusted so that the model's predicted output is as close as possible to the targeted output. Gradient descent can be stochastic, mini-batch, or momentum, with the former being the most common type.

b) Backpropagation: The model weights are updated by propagating errors from the output to the input layer. Examples of these include dropout mode, batch normalization, and reverse-mode autodiff.

**5) Evaluation Techniques:** There are many metrics that can be used to evaluate the performance of a model, including precision, recall, and F-score.

**6) Precision:** In math, precision is a measure of the proportion of predictions that are correct compared to all predictions within a set. By identifying words

native to a dialect, the precision can be used to measure the accuracy of a model for classifying spoken Arabic dialects.

**7) Recall:** Compared with all possible correct predictions, recall measures the proportion of correct predictions made by the model. For instance, a recall measurement could be used to evaluate a model for classifying spoken Arabic dialects based on its ability to identify words from the dialect being evaluated, which are also used in other dialects.

a) **F-Score:** F-Score measures the balance between precision and recall. An F-Score can be used to evaluate the accuracy of a model in identifying not only native Arabic words but also words from that dialect, which are also used in other dialects when analyzing spoken Arabic dialects.

## 4. Results and Discussion

This study reviewed existing studies on AD involving different datasets, including Egypt, Gulf, Moghrabi, Levantine, Iraqi, Yemeni, and Sudanese. Artificial intelligence has been used to detect and classify Arabic dialects using conceptual and supervised learning. Rule-oriented systems, Support Vector Machines, and n-grams have been employed in the past to detect and categorize Arabic dialects. Despite these advantages, there are some limitations, such as the need to manually create features and pre-process the data prior to analysis. To overcome these limitations, this paper introduces a conceptual deep learning approach based on deep learning concepts to detect and classify Arabic dialects using six stages as follows Natural Language Processing (NLP), feature engineering techniques, neural networks, language models, optimization techniques, and evaluation techniques. To prepare the data for further analysis, the model used a variety of techniques during the NLP stage. The tokenization process involves tokenizing, stemming, removing stop words, and tagging words based on its parts of speech. The Feature Engineering techniques are used to extract and numerically represent the characteristic features from the data. For analysis purposes, these data must be collected. Neural networks are used to identify patterns and make predictions based on the accessible data. This stage uses several neural network architectures to classify data based on dialects, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), prior to the Recurrent Neural Network stage. Moreover, data context understanding, and expression can also be improved by using language models. Particularly, this model uses regularization and dropout techniques to improve its performance. Furthermore, the model will be evaluated using different evaluation techniques as well as cross-validation. The effectiveness of the model can be evaluated by testing it using unseen data. The model can be refined and tuned to further improve its performances.

## 5. Conclusion

There are several factors contained within each acoustic wave that are used to

identify the dialect of a speaker in AD identifiers. As a result, in this study, we have developed a model for detecting and classifying ADs. As a result of combining convolutional neural networks (CNNs) and recurrent neural networks (RNNs), a model that captures the semantic, syntactic, and phonological characteristics of extensive dialects is developed. In the proposed model, there are six stages such as Natural Language Processing (NLP), Feature Engineering techniques, neural networks, language models, optimization techniques, and evaluation techniques. There are several techniques that can be applied at each stage of the process to detect and categorize AD in this study. In future work, it will be possible to verify the accuracy of the proposed model by conducting experiments to verify it. Moreover, to improve the accuracy of the model, further work will include an investigation of alternative datasets and hyperparameter tuning as methods to improve the model's accuracy. In addition, the research endeavor seeks to find ways to improve the efficacy of the model, as well as integrating genetic and medical imaging data into the model to enhance its efficacy.

### Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

### References

- [1] Wei, G. (2022) Research on Internet Text Sentiment Classification Based on BERT and CNN-BiGRU. 2022 11th International Conference on Communications, Circuits and Systems (ICCCAS), Singapore, 13-15 May 2022, 285-289. <https://doi.org/10.1109/ICCCAS55266.2022.9824526>
- [2] Omran, T.M., Sharef, B.T., Grosan, C. and Li, Y. (2023) Transfer Learning and Sentiment Analysis of Bahraini Dialects Sequential Text Data Using Multilingual Deep Learning Approach. *Data & Knowledge Engineering*, **143**, Article ID: 102106. <https://doi.org/10.1016/j.datak.2022.102106>
- [3] Qureshi, K.N., et al. (2022) A Blockchain-Based Efficient, Secure and Anonymous Conditional Privacy-Preserving and Authentication Scheme for the Internet of Vehicles. *Applied Sciences*, **12**, Article No. 476. <https://doi.org/10.3390/app12010476>
- [4] Gobert, M. (2023) Helping Gulf Arab Learners Negotiate the Linguistic Challenges Posed by English as a Medium of Instruction. In: Wyatt, M. and El Gamal, G., Eds., *English as a Medium of Instruction on the Arabian Peninsula*, Routledge, London. <https://doi.org/10.4324/9781003183594-12>
- [5] Gravano, A. (2009) Turn-Taking and Affirmative Cue Words in Task-Oriented Dialogue. Columbia University, New York.
- [6] Altowayti, W.A.H., et al. (2022) The Role of Conventional Methods and Artificial Intelligence in the Wastewater Treatment: A Comprehensive Review. *Processes*, **10**, Article No. 1832. <https://doi.org/10.3390/pr10091832>
- [7] Rasool, M., Ismail, N.A., Al-Dhaqm, A., Yafooz, W.M.S. and Alsaeedi, A. (2023) A Novel Approach for Classifying Brain Tumours Combining a SqueezeNet Model with SVM and Fine-Tuning. *Electronics*, **12**, Article No. 149. <https://doi.org/10.3390/electronics12010149>
- [8] Mohammed, M.Q., et al. (2022) Review of Learning-Based Robotic Manipulation in

- Cluttered Environments. *Sensors*, **22**, Article No. 7938.  
<https://doi.org/10.3390/s22207938>
- [9] Kong, L. (2013) An Improved Information-Security Risk Assessment Algorithm for a Hybrid Model. *International Journal of Advanced Computer Technology*, **5**.
- [10] Mohammed, M.Q., *et al.* (2021) Deep Reinforcement Learning-Based Robotic Grasping in Clutter and Occlusion. *Sustainability*, **13**, Article No. 13686.  
<https://doi.org/10.3390/su132413686>
- [11] Nagwani, N.K. and Suri, J.S. (2023) An Artificial Intelligence Framework on Software Bug Triaging, Technological Evolution, and Future Challenges: A Review. *International Journal of Information Management Data Insights*, **3**, Article ID: 100153.  
<https://doi.org/10.1016/j.jjimei.2022.100153>
- [12] Gao, S., *et al.* (2023) Code Structure-Guided Transformer for Source Code Summarization. *ACM Transactions on Software Engineering and Methodology*, **32**, 1-32.  
<https://doi.org/10.1145/3522674>
- [13] Gupta, C., Johri, I., Srinivasan, K., Hu, Y.-C., Qaisar, S.M. and Huang, K.-Y. (2022) A Systematic Review on Machine Learning and Deep Learning Models for Electronic Information Security in Mobile Networks. *Sensors*, **22**, Article No. 2017.  
<https://doi.org/10.3390/s22052017>
- [14] Moore, D.A. (2013) Security Risk Assessment Methodology for the Petroleum and Petrochemical Industries. *Journal of Loss Prevention in the Process Industries*, **26**, 1685-1689. <https://doi.org/10.1016/j.jlp.2013.10.012>
- [15] Hassan, M., Saeedi, K., Almagwashi, H. and Alarifi, S. (2023) Information Security Risk Awareness Survey of Non-governmental Organization in Saudi Arabia. In: Visvizi, A., Troisi, O. and Grimaldi, M., Eds., *Research and Innovation Forum 2022. RIIFORUM 2022. Springer Proceedings in Complexity*, Springer, Cham, 39-71.  
[https://doi.org/10.1007/978-3-031-19560-0\\_4](https://doi.org/10.1007/978-3-031-19560-0_4)
- [16] Alshareef, N.M.N. (2022) Information Security Risk Management (ISRM) Model for Saudi Arabian Organisations. Curtin University, Perth.
- [17] Tuyikeze, T. and Flowerday, S. (2014) Information Security Policy Development and Implementation: A Content Analysis Approach. *8th International Symposium on Human Aspects of Information Security and Assurance*, Plymouth, 8-9 July 2014, 11-20.
- [18] Habash, N., Rambow, O., Diab, M. and Kanjawi-Faraj, R. (2008) Guidelines for Annotation of Arabic Dialectness. *Proceedings of the LREC Workshop on HLT & NLP within the Arabic World*, Marrakech, 49-53.
- [19] Diab, M., Habash, N., Rambow, O., Altantawy, M. and Benajiba, Y. (2010) COLABA: Arabic Dialect Annotation and Processing. *LREC Workshop on Semitic Language Processing*, Malta, 17-23 May 2010, 66-74.
- [20] Zaidan, O. and Callison-Burch, C. (2011) The Arabic Online Commentary Dataset: An Annotated Dataset of Informal Arabic with High Dialectal Content. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. Human Language Technologies*, Portland, 19-24 June 2011, 37-41.
- [21] Elfardy, H. and Diab, M. (2013) Sentence Level Dialect Identification in Arabic. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Sofia, 4-9 August 2013, 456-461.
- [22] Cotterell, R. and Callison-Burch, C. (2014) A Multi-Dialect, Multi-Genre Corpus of Informal Written Arabic. *9th International Conference on Language Resources and Evaluation*, Reykjavik, 26-31 May 2014, 241-245.
- [23] Tillmann, C., Mansour, S. and Al-Onaizan, Y. (2014) Improved Sentence-Level

- Arabic Dialect Classification. *Proceedings of the 1st Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects*, Dublin, 23 August 2014, 110-119. <https://doi.org/10.3115/v1/W14-5313>
- [24] Nasr, M., Ateia, M. and Hassan, K. (2016) Artificial Intelligence for Greywater Treatment Using Electrocoagulation Process. *Separation Science and Technology*, **51**, 96-105. <https://doi.org/10.1080/01496395.2015.1062399>
- [25] Sadat, F., Kazemi, F. and Farzindar, A. (2014) Automatic Identification of Arabic Language Varieties and Dialects in Social Media. *Proceedings of the 2nd Workshop on Natural Language Processing for Social Media (SocialNLP)*, Dublin, 24 August 2014, 22-27. <https://doi.org/10.3115/v1/W14-5904>
- [26] Durandin, O.V., Hilal, N.R. and Strebkov, D.Y. (2016) Automatic Arabic Dialect Classification. *Computational Linguistics and Intellectual Technologies: Proceedings of the Annual International Conference "Dialogue 2016"*, Moscow, 1-4 June 2016, 1-13.
- [27] Ramadan, H., Alqahtani, M. and Algoson, A. (2022) Identifying Equivalent Words from Different Arabic Dialects Using Deep Learning Techniques. *2022 20th International Conference on Language Engineering (ESOLEC)*, Cairo, 12-13 October 2022, 124-128. <https://doi.org/10.1109/ESOLEC54569.2022.10009555>
- [28] Alsayadi, H.A., Al-Hagree, S., Alqasemi, F.A. and Abdelhamid, A.A. (2022) Dialectal Arabic Speech Recognition using CNN-LSTM Based on End-to-End Deep Learning. *2022 2nd International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, Ibb, 25-26 October 2022, 1-8. <https://doi.org/10.1109/eSmarTA56775.2022.9935427>
- [29] Nasr, S., Duwairi, R. and Quwaider, M. (2023) End-to-End Speech Recognition for Arabic Dialects. *Arabian Journal for Science and Engineering*. <https://doi.org/10.1007/s13369-023-07670-7>
- [30] Ali, A., *et al.* (2015) Automatic Dialect Detection in Arabic Broadcast Speech. *Interspeech 2016*, San Francisco, 8-12 September 2016, 2934-2938. <https://doi.org/10.21437/Interspeech.2016-1297>
- [31] Alzu'bi, D. and Duwairi, R. (2021) Detecting Regional Arabic Dialect Based on Recurrent Neural Network. *2021 12th International Conference on Information and Communication Systems (ICICS)*, Valencia, 24-26 May 2021, 90-93. <https://doi.org/10.1109/ICICS52457.2021.9464605>
- [32] Elaraby, M. and Abdul-Mageed, M. (2018) Deep Models for Arabic Dialect Identification on Benchmarked Data. *Proceedings of the 5th Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2018)*, Santa Fe, 20 August 2018, 263-274.
- [33] Zaidan, O.F. and Callison-Burch, C. (2014) Arabic Dialect Identification. *Computational Linguistics*, **40**, 171-202. [https://doi.org/10.1162/COLI\\_a\\_00169](https://doi.org/10.1162/COLI_a_00169)
- [34] Itani, M.M., Zantout, R.N., Hamandi, L. and Elkabani, I. (2012) Classifying Sentiment in Arabic Social Networks: Naive Search versus Naive Bayes. *2012 2nd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA)*, Beirut, 12-15 December 2012, 192-197. <https://doi.org/10.1109/ICTEA.2012.6462864>
- [35] Yahya, A.E., Gharbi, A., Yafooz, W.M.S. and Al-Dhaqm, A. (2023) A Novel Hybrid Deep Learning Model for Detecting and Classifying Non-Functional Requirements of Mobile Apps Issues. *Electronics*, **12**, Article No. 1258. <https://doi.org/10.3390/electronics12051258>
- [36] Mounsef, J., Hasib, M. and Raza, A. (2022) Building an Arabic Dialectal Diagnostic

- Dataset for Healthcare. *International Journal of Advanced Computer Science and Applications*, **13**, 859-868. <https://doi.org/10.14569/IJACSA.2022.01307100>
- [37] Al-Dhaqm, A., Abd Razak, S., Ikuesan, R.A., Kebande, V. R. and Siddique, K. (2020) A Review of Mobile Forensic Investigation Process Models. *IEEE Access*, **8**, 173359-173375. <https://doi.org/10.1109/ACCESS.2020.3014615>
- [38] Slunečková, L. (2018) ESP Students and the Mysteries of English Word Order. In: Jančaříková, R., Ed., *Interpretation of Meaning across Discourses*, Masarykova Univerzita Nakladatelství, 109-120.
- [39] Ngadi, M., Al-Dhaqm, R. and Mohammed, A. (2012) Detection and Prevention of Malicious Activities on RDBMS Relational Database Management Systems. *International Journal of Scientific & Engineering Research*, **3**, 1-10.
- [40] Zu'bi, A. (2023) Some Linguistic Features of the Dialect of Acre and Their Possible Explanation by the History of the City. *Journal of Semitic Studies*, Article ID: Fgac029. <https://doi.org/10.1093/jss/fgac029>
- [41] Saleh, M.A., Othman, S.H., Al-Dhaqm, A. and Al-Khasawneh, M.A. (2021) Common Investigation Process Model for Internet of Things Forensics. 2021 *2nd International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, Cameron Highlands, 15-17 June 2021, 84-89. <https://doi.org/10.1109/ICSCEE50312.2021.9498045>
- [42] Al-Dhaqm, A., Razak, S., Siddique, K., Ikuesan, R.A. and Kebande, V.R. (2020) Towards the Development of an Integrated Incident Response Model for Database Forensic Investigation Field. *IEEE Access*, **8**, 145018-145032. <https://doi.org/10.1109/ACCESS.2020.3008696>
- [43] Al-Dhaqm, A., *et al.* (2017) CDBFIP: Common Database Forensic Investigation Processes for Internet of Things. *IEEE Access*, **5**, 24401-24416. <https://doi.org/10.1109/ACCESS.2017.2762693>
- [44] Al-Dhaqm, A., *et al.* (2020) Categorization and Organization of Database Forensic Investigation Processes. *IEEE Access*, **8**, 112846-112858. <https://doi.org/10.1109/ACCESS.2020.3000747>