Building an Arabic Dialectal Diagnostic Dataset for Healthcare

Jinane Mounsef
Department of Electrical Engineering and Computing Sciences
Rochester Institute of Technology
Dubai, United Arab Emirates

Maheen Hasib
School of Mathematical and Computer Sciences
Heriot Watt University
Dubai, United Arab Emirates

Ali Raza
Department of Electrical Engineering and Computing Sciences
Rochester Institute of Technology
Dubai, United Arab Emirates

Abstract—**Accurate diagnosis of patient conditions becomes challenging for medical practitioners in urban metropolitan cities. A variety of languages and spoken dialects impedes the diagnosis achieved through the exploratory journey a medical practitioner and patient go through. Natural language processing has been used in well-known applications, such as Google Translate, as a solution to reduce language barriers. Languages typically encountered in these applications provide the most commonly known, used or standardized dialect. The Arabic language can benefit from the common dialect, which is available in such applications. However, given the diversity of dialects in Arabic in the healthcare domain, there is a risk associated with incorrect interpretation of a dialect, which can impact the diagnosis or treatment of patients. Arabic language dialect corpuses published in recent research work can be applied to rule-based natural language applications. Our study aims to develop an approach to support medical practitioners by ensuring that the diagnosis is not impeded based on the misinterpretation of patient responses. Our initial approach reported in this work adopts the methods used by practitioners in the diagnosis carried out within the scope of the Emirati and Egyptian Arabic dialects. In this paper, we develop and provide a public Arabic Dialect Dataset (ADD), which is a corpus of audio samples related to healthcare. In order to train machine learning models, the dataset development is designed with multi-class labelling. Our work indicates that there is a clear risk of bias in datasets, which may come about when a large number of classes do not have enough training samples. Our crowd sourcing solution presented in this work may be an approach to overcome the sourcing of audio samples. Models trained with this dataset may be used to support the diagnosis made by medical practitioners.**

Keywords*—*Dialectal Arabic (DA); healthcare diagnosis; natural language processing (NLP); multi-class labeling; crowd sourcing

I. Introduction

Research into healthcare practices often highlight language barriers as a major hurdle that impedes a seamless doctor patient interaction and hinders the possibility of positive diagnostic outcomes [1], [2], [3]. In emergency rooms, where timing is critical and information exchange between the patient and the doctor can save lives, the language barrier is seen as an obstacle that must be overcome. It is important to understand the role that language plays, as there has been an increase in culturally and linguistically diverse populations among patients and practitioners [2]. As a result of this ongoing global migration and globalization, many societies all over the world have become multicultural, with many migrants who do not speak the official language of the host country. In the United Arab Emirates (UAE) in particular, the local citizens make up only about 11% of the population, with the rest being from various parts of the world, mostly from Southeast Asian countries and various Arab countries [4]. With this diversity in culture and population, it causes a communication gap in the healthcare service and creates challenges in providing quality individual and holistic healthcare [5].

A large number of published studies [6], [7] address the language barrier that exists between the doctor and the patient, resulting in a poor understanding of diagnosis, relevant investigations, and medication instructions [7]. Surveys, such as that conducted by [8], have shown that language barrier can also lead to poor comprehension and compliance with recommendations for follow-up and treatment. This increases the likelihood of the occurrence of adverse medical events, thereby lowering patient satisfaction. Hence, communication barriers must be considered as part of any strategy aimed at improving patient safety and risk management in healthcare organizations [9].

Attempts to solve this communication problem have been reported in [10], [11], [12], [13] by using Google Translate (GT) or an interpreter. However, when language is a barrier, GT remains the most easily accessible and a free initial mode of communication between the doctor and the patient [10]. A good medical interpreter needs to be able to accurately translate the complex medical terminology. Evidence in [14] suggests that when limited English proficiency patients have access to skilled professional interpreters or bilingual physicians, they have better communication, patient satisfaction, and outcomes, as well as fewer interpretation errors.

The limitations related to healthcare and diagnosis have been reported in [14], [15], [16]. GT has only 57.7% accuracy and should not be relied on for critical medical communications [10]. In [12], it is mentioned that those patients who require interpreters but do not receive them, have a poor self-reported understanding of their diagnosis and treatment. Ad hoc interpreters misinterpret or omit up to half of all physicians' questions and are more likely to make clinically significant errors. The three most commonly used forms of interpreters are ad hoc, professional, and telephone interpreters. For ad hoc interpreters, people commonly use friends or family members of the patient or staff in the work setting, such as housekeepers, secretaries, or medical personnel, who are untrained in interpreting. It is increasingly recognized that the use of untrained or ad hoc interpreters can lead to inaccurate communication and ethical breaches.

Language barrier leads to a communication breakdown between the patient and healthcare providers [6], [17]. Hence, it is critical for the doctor and the patient to communicate effectively, which is why it is advantageous if the doctor is fluent in the patient's language [18], [19]. However, doctors may not always be able to communicate in the patients' native language, which creates hurdles, as shown in [20].

Early health communication linguistic research [21] focused mainly on small datasets, such as language samples gathered from face-to-face clinical interactions or research interviews [22]. The main drawback of this was that the findings presented were based on limited datasets and hence, researchers have started to use corpus linguistic methods [23]. A corpus is a collection of authentic text or audio organized into datasets [24].

The Arabic language is classified into three forms: classical, modern standard and dialectal. Classical Arabic (CA) is the language used in both ancient history literary and religious texts. Modern Standard Arabic (MSA) is the "official" Arabic used in books, magazines, media, legal documents, etc. Dialectal Arabic (DA) is the informal Arabic that native speakers use to communicate in their daily lives. Research in [25], [26], [27], [28] focused on the development of an Arabic language corpus, such as Egyptian, Tunisian and Iraqi dialects.

Over the last decade, Arabic and its dialects have made growth in the field of Natural Language Processing (NLP) [29]. Dialects vary and mutate over the large geographical span of the Arab peninsula with popularity borne by the Egyptian dialect in a majority of countries in the Middle East and North African regions. The UAE population distribution provides the motivation for the work reported in this paper. Of the 40% Arab population in the UAE, Emirati and non-Emirati Arabic groups are almost in equal proportion. This allows us to focus our research efforts on the Emirati and Egyptian dialects based on the fact that the Egyptian dialect is the most spoken in the non-Emirati Arab community. Although research in [30] focused on developing a pediatric assistant that supported Arabic dialects, particularly Egyptian dialects, there is a lack of research for both Emirati and Egyptian dialects in the UAE healthcare sector.

With the growth of AI applications, researchers are looking for ways to use NLP (a strand of AI) to solve real-world problems related to language. So, a dataset that can enhance the functionality of NLP-based applications is highly desirable for researchers in this field and is a main contribution for science. Researchers working on language translation tools and chatbots will be very interested in such tools when focusing on the development of solutions for the healthcare industry. These datasets can be used to develop machine learning models to detect and classify linguistics.

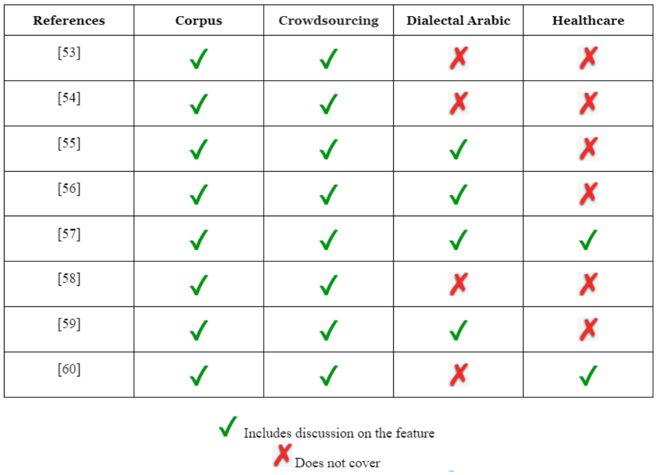
In this work, the approach used by medical practitioners in the UAE were analyzed following some interviews with doctors representing the Dubai Health Authority (DHA). A database of questions was developed in the English Language and a platform was developed to allow users to record a translated version of the typical response to a question in their Arabic dialect (Emirati or Egyptian). These responses were labeled and stored as an audio dataset in a raw mp3 format for researchers to use. Recordings with poor audio quality were manually removed after an inspection. The development of AI-based tools can improve the diagnostic outcomes for patients and lead to a better care. Issues that are diagnosed early allow for treatments, which can lead to correct prognosis and timely interventions.

The remainder of this paper is organized as follows: Section II provides the related work in the field of Arabic NLP (ANLP); Section III discusses the DA challenges in healthcare; Section IV presents our approach and methodology for building a dataset to train a Machine Learning (ML) algorithm; Section V deals with the ML approach to which the dataset can be applied. Finally, Section VI provides the limitations of the work in addition to the social and ethical considerations. We conclude this paper with future directions of research that can be supported through this dataset and its enhancement.

II. Literature Review

A considerable amount of literature has been published on the growth of the DA, which has been recognized as the primary language for informal communication [31]. This growth has been aligned with the emerging trends in online social media platforms (e.g. Twitter, Snapchat, Facebook etc.). The online social media factor attributes to the varsity of DA, which has led to a wider use in offline social interaction. As highlighted in [32], the forms of DA differ depending on the geographic distribution and the socio-economic conditions of the speakers. Each form of DA is considered a divergence from the formal variety, the MSA. This divergence is evidenced in the challenges presented by [33], highlighting Arabic Internet users adopting a mixture of MSA and DA, thereby increasing the challenges related to text processing for machine translation through NLP.

The recent advances in computational power and processing of large data arrays have led to the enhancement of NLP as a technique to empower machines in gaining a better understanding of the human language [34]. Deep learning (DL), a consequence of the rise in computational power, has increased the opportunity for more tasks to be carried out by NLP. The Arabic NLP community (ANLP) [35] has yet to grasp the advantages offered by NLP with DL, despite the growth of DA aligned with social media platforms mentioned earlier. However, recent work published by [36] shows a new focus on ML and DL, based on lexicon and corpus approaches. Islam et al. [37] discuss the importance of NLP in the healthcare industry with researchers finding ways to improve diagnostics. Furthermore, automation for initial and informal diagnosis has been envisaged through the use of NLP to provide an on-demand self-service for patients [38]. The communication between a doctor and a patient usually involves personal questions, which reveal intimate information necessary for an accurate diagnosis [19]. These challenges are amplified in the Arabic culture and the framework of respect and politeness around the spoken dialogue [39]. To cater for these challenges, techniques used by practitioners have included the use of an interpreter [40]. However, the framework of politeness mentioned earlier also reduces the success of using an interpreter in the Arabic culture. Another



References	Corpus	Crowdsourcing	Dialectal Arabic	Healthcare
[53]	✓	✓	✗	✗
[54]	✓	✓	✗	✗
[55]	✓	✓	✓	✗
[56]	✓	✓	✓	✗
[57]	✓	✓	✓	✓
[58]	✓	✓	✗	✗
[59]	✓	✓	✓	✗
[60]	✓	✓	✗	✓

✓ Includes discussion on the feature
✗ Does not cover

Fig. 1. List of Published Papers using Crowd Sourcing.

alternative approach used by doctors is known as code switching [41], where the bilingual capability of the practitioner can be used to combine English words and terms (of important medical consequence) and words in DA, which can provide an atmosphere of comfort and respect for the patient in the cultural setting. Although code switching has been reported to be a positive and useful approach, the bilingual capability is not ubiquitous, given the divergence of DA. Hence, approaches, which are supported by artificial intelligence (AI), play a role in bridging the gap. This gap is wider, where the growth in healthcare tourism [42] is supported by the deployment of practitioners accustomed to MSA or a different strand of DA [43].

Approaches in sentiment analysis, which apply NLP as reported in [44], have been used in healthcare based on text computation extracted from web sources, such as blogs and social media, and with no methodical approach used for the attainment. Again, the politeness of the Arabic culture plays a significant role in the level of accuracy that may be gained through these approaches, as there is a limited directness. Furthermore, online sources will also suffer from a lack of contextual information, which can be used to train models used in AI. Researches [31], [44], [45], [46] were developed using a corpus-based approach for sentiment analysis in healthcare. The techniques that have been used are Lexicon. There are several challenges mentioned in [45], [46] facing the sentiment analysis and evaluation process, especially in the medical domain. There are health related blogs and forums where people discuss their health issues, symptoms, diseases, medication, etc. [47]. These challenges become obstacles in analyzing the accurate meaning of sentiments and detecting the suitable sentiment polarity. The approaches used to overcome these challenges are discussed in [48].

ML implementations of audio analytics have gained popularity in applications, such as Alexa, Siri, Google Assistant and Google Home. The capabilities offered are founded on models, which can extract information from audio signals [49]. Audio processing capabilities supported by open-source codes and public big data have led to an accelerated development of applications, such as Shazam, which provides an audio similarity search capability [50]. Applications of audio analytics can range from customer sentiment analysis in recordings taken from call centers to media content monitoring for parental control. In the domain of healthcare, several ML-based approaches have been proposed to consider clinical decisions supported using text-free data [51]. In recent years, clinical speech and audio processing have offered new opportunities, such as automatic transcription of patient conversations, synthesis of clinical notes, as well as identification of disorders related to speech [52]. Chatbots, such as Babylon Health, have been developed for diagnosis and prevention of diseases. Such chatbots use speech recognition technologies to compare symptoms reported by the user with a database of diseases to recommend a course of actions based on the identified disease in addition to the patient history. In the special case of machine translation, there is currently limited evidence in the literature of the use of healthcare-oriented ML-based translators in clinical practice [53]. Instead, physicians commonly use either professional interpreters, who are usually costly, or digital translating services, such as GT, which are not tailored to properly function with a medical lexicon.

There are different techniques for processing the audio files for ML. The approaches used in the literature are audio spectrograms in [54] and Mel Frequency Cepstral Coefficients (MFCC) in [55]. In [56], the input of fixed length video segments of 10 seconds was converted to an audio spectrogram and fed into a Convolutional Neural Network (CNN) based on Alexnet [57] and VGG-16 [58] architectures. In [55], the use of audio spectrograms to feed into a CNN with the use of MFCCs was discussed. The authors also used a second approach, where the audio spectrogram was used as an input, using a modified VGG-16 architecture based on transfer learning [59]. Results in [55] showed that the accuracy using MFCC was higher than the spectrogram-based approach.

The approach proposed in this paper supports speech recognition (SR) under text and speech processing as a branch of NLP. Advances in SR through big data and DL have yielded significant gains through innovative approaches, as found in recent academic work [60], [61]. Limited work on the use of an audio corpus is referenced in [54], where voice recognition based on whole sentences without speech segmentation is done. To improve and innovate in the field of voice recognition for healthcare diagnosis in the DA, this corpus provides the big data component, which is labelled and can be used for supervised learning. The labelling approach of the corpus can support the development of voice recognition with DL for improvement in the healthcare diagnostic outcomes. The approach used to create the corpus is similar to Amazon Mechanical Turk. Other approaches, which use the crowd sourcing and labelling mechanism, are mentioned in Fig. 1.

III. Healthcare Challenges with DA

With the discovery of oil in the UAE in the late 1950s, a huge economic and social transformation began resulting in an increase of the need for foreign labor. Since then, the UAE has seen a huge number of expatriates trickling in to gain from the profitable economic bustle including trade, real estate, construction, and healthcare, thereby outnumbering the national population. While the nationals whose spoken language is Arabic only amount to 11.48% of the entire

TABLE I. SAME MEANING OF DIFFERENT HEALTHCARE WORDS IN EGYPTIAN AND EMIRATI DIALECTS

Egyptian Dialect	Emirati Dialect	Meaning
وجع	عوار	pain
غمَمان	لوعة	nausea
أرّجع	بزوع	I vomit
بتوجعني	يعورني	It hurts

TABLE II. EXAMPLES OF HEALTHCARE STATEMENTS IN DIFFERENT DAS WITH THEIR TRANSLATIONS IN ENGLISH

Questions/ Answers	Translation by GT	Correct Translation
زوري واجعني (Egyptian dialect)	visit and see me	my throat hurts
انا دايخ (Egyptian dialect)	I am deaf	I am dizzy
انا عيان (Egyptian dialect)	I am an eye	I am sick
شو رح يصير لو ما سويت هاي العملية؟ (Syrian dialect)	What would happen if you did not make this process happen?	What will happen if I do not get the surgery?
حاسة حالي تعبانة (Jordanian dialect)	My current sense is tired	I am tired
فيني لوعة (Emirati dialect)	Vinny is crazy	I feel nauseated

population in 2022, the expatriates total around 88.52% of the population, mainly coming from India, Pakistan, Bangladesh, Philippines, Iran, Egypt, Nepal, Sri Lanka and China [4]. As most of these migrants are non-Arabic speakers, the common communication language used in the different public and private service sectors in the UAE is English. The Arabic language has also been a barrier to Arabic speakers across the country, as the Arab migrants use different dialects or DAs, which diverge from the MSA, and are often classified regionally as Egyptian, North African, Levantine, Gulf, and Yemeni or sub-regionally as Tunisian, Algerian, Moroccan, Lebanese, Syrian, Jordanian, Saudi, Kuwaiti and Qatari [62]. Moreover, many of the Arabic speakers cannot use the English language to communicate efficiently with non-Arabic speakers. This suggests that the language in the UAE is a contributing factor to misinformation in several critical service providers including the healthcare sector.

In the UAE, the two mostly spoken Arabic dialects are Emirati (11.48%) and Egyptian (4.23%) [4]. While the majority of the Arabic speaking residents use these two dialects to communicate on a daily basis [63], hospitals are staffed mostly with physicians who speak exclusively English at workplace. Therefore, these physicians are frequently obliged to deal with Arabic speaking patients who have no language in common. Moreover, the diversity of the Arabic language through its several DAs poses another serious challenge that needs to be addressed [62]. Table I shows an example of several Arabic healthcare terms expressed differently in Emirati and Egyptian dialects, but having the same meaning. A situation of this kind, with barriers in language and medical understanding, creates serious problems for quality, security and equitability of medical care. Rosse et al. [64] defined a language barrier as “a communication barrier resulting from the parties concerned speaking different languages” and supported previous claims that these barriers pose a significant threat to quality of care and patient safety.

Therefore, the language barrier poses challenges for achieving high levels of satisfaction among physicians and patients, providing high-quality healthcare service, and maintaining patient safety. To address these challenges, several solutions are available today, but they all have their drawbacks. There has been recently a rising demand of English-Arabic translating services for medical, anatomy and healthcare lexicons in Arabic cosmopolitan countries, which usually attract multilingual/multidialectal expatriates [65]. Knowing that professional interpreters are very expensive and are not always available for some dialects, the framework of politeness in the Arab culture also reduces the success of using an interpreter for DA speakers. On the other hand, GT, increasingly often used when no other alternatives exist, is known to be unreliable for medical communication [66] and human interference is greatly needed to produce accurate and effective translation. For instance, an Egyptian patient reporting a pain in the leg, ''''رجلي وجعاني , is interpreted as “My feet are hungry” instead of “My leg hurts”. Another example shows the inadequacy of GT, which fails to translate two different statements having the same meaning into the same expression. The two Egyptian dialect statements ''''انا مصدع and ''''انا عندي صداع ,which both report a headache, are construed as “I’m cracked” and “I have a headache”, respectively. Table II, which illustrates some examples of translation from several DAs to English, shows the inaccurate translation of GT when used in the healthcare sector.

All these factors lead us to explore and suggest a better solution to the language dilemma that physicians face in providing high-quality and safe care to DA-speaking patients with limited English proficiency in the UAE. Our work proposes a new approach of efficiently addressing this problem by using the crowd sourcing and labelling mechanism for the Emirati and Egyptian dialects in the medical diagnosis context. To our knowledge, no other work has followed this approach to address the challenge of English-Arabic interpretation in the medical field, specifically related to the emergency diagnosis questionnaire.

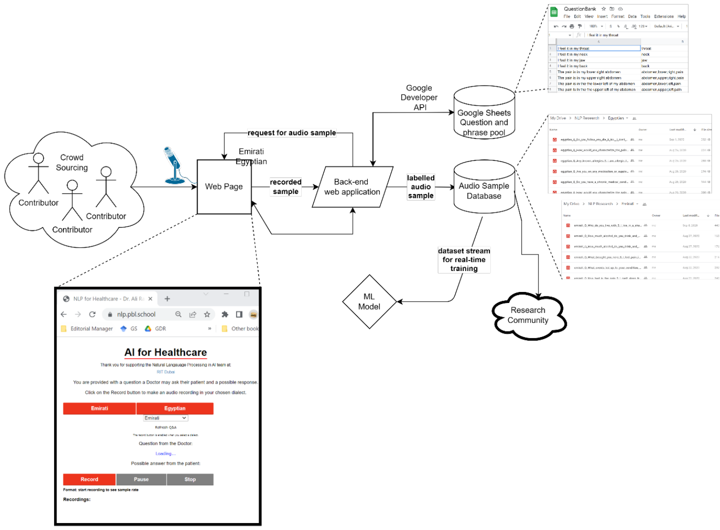
IV. APPROACH AND METHODOLOGY

The crowd sourcing approach used in this work is termed as “crowd labeling”. In this approach, an automated labeling implementation is developed, which allows for audio recordings to be labeled at the time they are recorded and uploaded by contributors.

A. Architecture Overview

Fig. 2 provides an overview of the application architecture using HTML, CSS, JavaScript, PHP and Google Developer APIs. Visitors to the webpage can see a landing page, which offers them the option to record their phrase in the Emirati or Egyptian dialect. The selection prompts a back-end process on the web server to get a random question and associated phrase from a pool, which is hosted in a Google sheets document. This is done using Google developer APIs.

The contributor records the phrase and uploads it to the audio sample database, which is hosted in Google drive. This repository is available for public access for other researchers. The audio files are labelled automatically. When a file is



	A	B
1	I feel it in my throat	throat
2	I feel it in my neck	neck
3	I feel it in my jaw	jaw
4	I feel it in my back	back
5	The pain is in my lower right abdomen	abdomen,lower,right,pain
6	The pain is in my upper right abdomen	abdomen,upper,right,pain
7	The pain is in the the lower left of my abdomen	abdomen,lower,left,pain
8	The pain is in the the upper left of my abdomen	abdomen,upper,left,pain

NLP for Healthcare - Dr. Ali Ra

nlp.pbl.school

AI for Healthcare

Thank you for supporting the Natural Language Processing in AI team at:
RIT Dubai

You are provided with a question a Doctor may ask their patient and a possible response.

Click on the Record button to make an audio recording in your chosen dialect.

Emirati | Egyptian

Emirati

Refresh Q&A

The record button is enabled when you select a dialect.

Question from the Doctor:

Loading...

Possible answer from the patient:

Record | Pause | Stop

Format: start recording to see sample rate

Recordings:

Fig. 2. High-Level Implementation Architecture.

uploaded, the following attributes are included in the file name: `<dialect>_Q_<possible question from Doctor>_S_<possible answer from patient>_L_<predefined labels>.mp3`

An example from the repository is provided here: `emirati_Q_What_brought_you_here_S_I_feel_pain_in_my_leg_L_leg_pain.mp3`

The audio samples are recorded with a 48 MHz sampling rate and stored in the mp3 format.

B. Application Overview

Fig. 3 provides an overview of the implementation with a summary of the steps:

1) User visits the webpage (https://nlp.pbl.school/)
2) User reads the welcome note.
3) User selects the dialect they wish to contribute in.
4) Once the question-and-answer set is generated, the user records the audio.
5) The user reviews the recording quality and uploads it to the Google drive storage location. The user can proceed to step 7 to continue with contributions or proceed to step 8 and exit the application by closing their browser page.
6) If the audio sample is not clear, the user can decide to re-record it and return to step 5.
7) The user can choose to launch a new contribution and restart at step 3.
8) The user can choose to exit the application at any time.

C. Access to Repository

The repository of the collected audio files can be accessed from Google drive using the link below. As new recordings are published, this drive is automatically updated. Researchers have unlimited access to the ADD dataset:

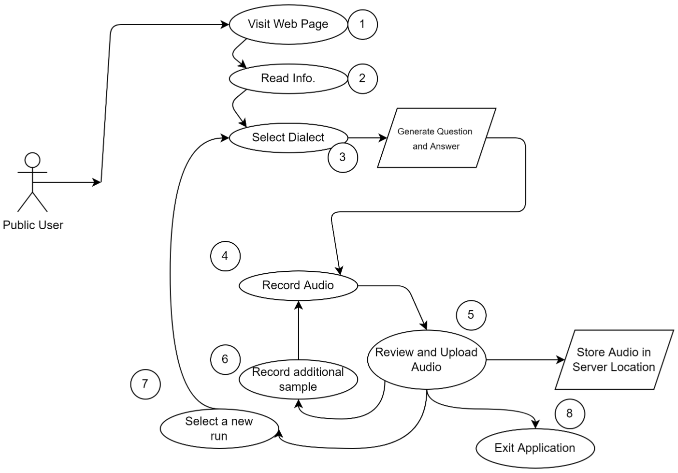


Fig. 3. Crowd Labeling Application Flow Diagram.

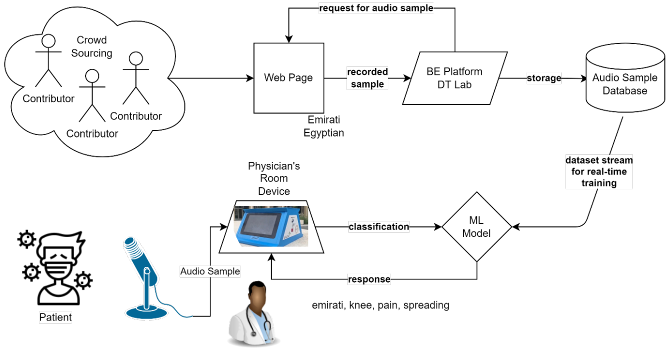


Fig. 4. Implementation of ML Pipeline for Patient-Doctor Diagnostic Device.

https://drive.google.com/drive/folders/160HE3q_FcqNJMyC4M5Hx1e2SffKhDeW8?usp=sharing

Access to this repository can also be automated using Google drive APIs by researchers who intend to create a continuous ML pipeline. With this approach, each time a new file is uploaded, the ML pipeline implemented can fetch the file and update the ML algorithm. The current implementation of this is shown in Fig. 4. The implementation consists of a two-monitor device based on the Raspberry Pi 4, and placed in the room of a physician. The doctor can select a question, which is played in the dialect of the patient. The response from the patient is recorded and uploaded to the ML model for classification. The response from the ML model provides the physician with the phrase from the database that best matches the input from the patient.

Preliminary results on the accuracy of the classification have been produced, as described in Section V. We intend to publish detailed results in the future when the size of the corpus reaches a minimum threshold.

V. ML Proposed Applications

We address in this section the need for capacity development in this area by providing some conceptual ML methods

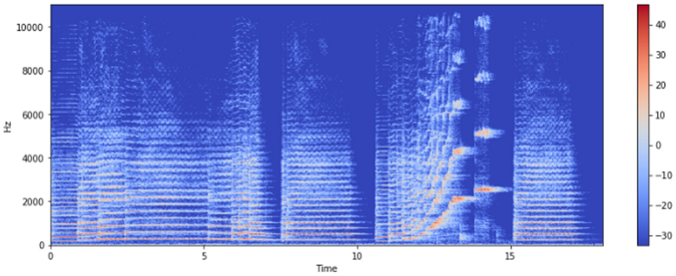


Fig. 5. Spectrogram of an Audio Clip with a Spectral Range of 0 kHz to 10 kHz.

for researchers to developing predictive algorithms for several healthcare speech recognition-related applications. Examples of such applications include speech tagging, dialect classification, diagnosis multi-labelling and language translation, using our freely available public ADD dataset for Emirati and Egyptian dialect lexicons.

Libraries, such as Librosa, provide useful tools for audio signal processing, using a Fourier Transform to convert an audio signal into a frequency domain from the time domain. Librosa supports the display, processing, and analysis of key spectral features, such as spectral-flux, spectral centroids and fundamental frequency components, which are essential features for ML-based spectral analysis [49]. Fig. 5 shows an implementation of Librosa's spectral display capabilities in which spectral images of our collected audio samples are generated. The image is normalized to a specific color profile and a frequency profile from 0 to 4 kHz, which aligns to the human speech audio range.

In a typical classification application, the images can be processed and numerically represented with the pixel color information being the input for each neuron in an Artificial Neural Network (ANN) or Convolutional Neural Network (CNN) model, such as AlexNet or GoogLeNet, as proposed by Boddapati et al. [67]. The training of the model will allow for classification of dialects and diagnosis types. This is summarized by the process pipeline shown in Fig. 6. In this section, we develop and test the proposed design process on the ADD dataset, as one possible approach of using the dataset for Arabic/English translation of the emergency diagnosis patient answers.

A. Audio Processing

A total of 301 recordings in the Egyptian dialect and 138 recordings in the Emirati dialect are collected as part of the ADD after manually removing the faulty audio samples from 12% of the Egyptian dialect recordings and 22% of the Emirati dialect recordings. The errors include wrong translation, blank recordings, and duplications. The maximum duration for an audio sample is 10 seconds, the minimum is 1 second, and the average is between 2 and 3 seconds.

Next, we consider the recorded raw audio samples, where a sample of an Egyptian dialect speech waveform is shown in Fig. 7. The waveform only shows the change of the signal's amplitude over time without giving any insight about the different frequency components pertaining to the recorded audio signal. Therefore, we convert the audio waveform to a spectrogram, which is a 2D image representing sequences of spectra with time along one axis, frequency along the other, and brightness or color indicating the strength of a frequency component at each time frame. This representation can thus be used with CNNs that are usually applied on images and can be applied directly to sound in this case. Moreover, learning more about the signal's frequencies gives us a better understanding of the recorded audio signal and allows us to filter out any unwanted disturbance, as distortion and noise can be visualized in a spectrogram.

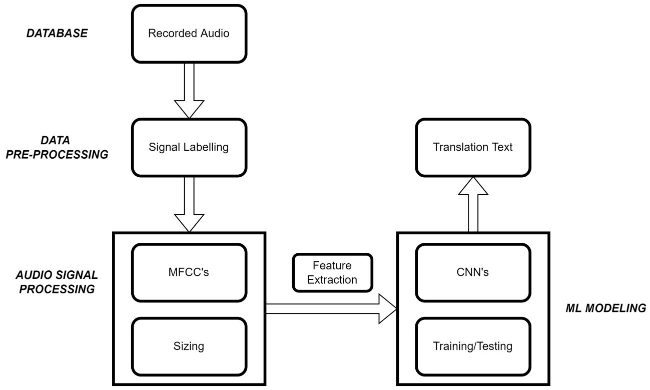


Fig. 6. Proposed Classification Pipeline.

B. Machine Learning Models

For speech classification, Mel Frequency Cepstral Coefficients (MFCCs) are commonly used for their classification and identification effectiveness, as they can describe concisely the shape of the spectral envelope expressed on a Mel-scale. However, MFCCs are also known as being "lossy" representations, which are preferably ruled out when working with a high-quality sound. Therefore, the spectrograms can be used directly as 2D images to feed pre-trained CNNs. In the case of a limited size dataset, transfer learning is used to relax the hypothesis that the training data must be large, independent and identically distributed with the test data. This motivates many work [68], [69], [70], [71], [72], [73] to use transfer learning in the presence of insufficient training data for speech and language classification. A network, which is pre-trained on a large-sized dataset, such as the ImageNet [74], will keep its structure and connection parameters, when used by a network-based deep transfer learning, to compute intermediate image representations for smaller-sized datasets. Therefore, a network, such as a CNN, is trained on the ImageNet to learn image representations that can be efficiently transferred to other visual recognition tasks with limited amount of training data. The front-layers of the network are operated as a feature extractor, where the extracted features are classified using ML classifiers, such as a support vector machine (SVM).

C. Preliminary Results

We propose a similar machine learning model for the spectrograms classification to identify the English translation of the corresponding Arabic audio sample. Fig. 8 shows the different steps followed in the described model. We perform

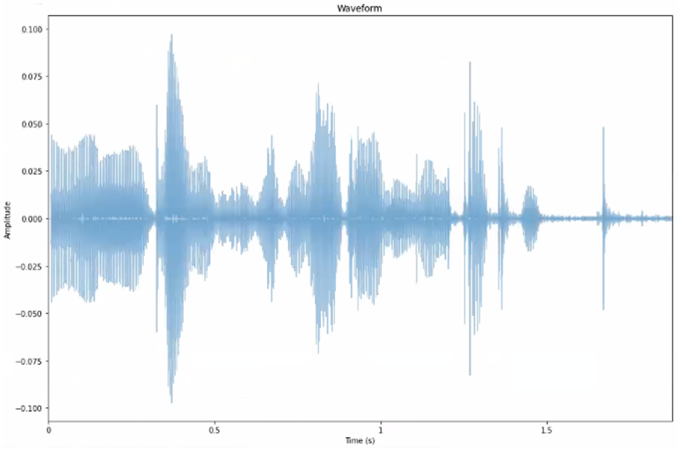


Fig. 7. Audio Waveform Sample of an Egyptian Dialect Recorded Answer.

identification experiments with a closed-set experimental protocol, where our model should predict the label to which the input image belongs. For experimentation, we use four classes of the dataset labeled as: "I smoke", "I do not smoke", "I live with my family", and "I live alone". The training set consists of 80% of the spectrogram images, while the testing set includes the remaining 20% of the images. Since the ADD dataset is relatively small, it is easy for the CNN model to over-fit and not generalize well on the testing data. To alleviate this problem, we augment the dataset with a factor of five by editing the pitch of the audio recordings to four different values. Other augmentation techniques are possible to use, such as time stretching or adding noise. We use the augmented training set of the resulting spectrograms on a VGG16 network. We extract features from different layers of the same network to explore the different classification accuracies. We perform an exhaustive search on the layers and report results with the layer that gives the highest accuracy. We find that the best performance corresponds to the last convolutional layer of VGG16. We use the extracted features to train a one-against-one multi-class linear SVM. The achieved recognition accuracy is of 78%.

Although the result might not look promising enough, it shows that the proposed model performs decently well for a small-sized dataset and proves to utilize the ADD in one of many ways to implement Emirati and Egyptian dialect sentences classification into their respective English translated expressions.

VI. Limitations, Social and Ethical Considerations

The main limitations of the approach presented in this work were the residual response bias of participants, the short period of time to conduct the research and failure to fully explore all the possible responses for both dialects.

The current work relies on translation crowdsourcing, which is particularly known for its innovative approach, combining the concept of crowdsourcing and that of natural and non-professional translation. The translation project that formed the basis of this work comprised of untrained translator participants who were native speakers of the Arabic dialect with no previous experience of translation. Even though their contribution to the crowdsourcing platform, as native speakers, undoubtedly gave a high credibility to the correctness of the translated responses, there was still an obvious response bias in the research given that they were translator novices, with many participants translating the response differently based on their knowledge and background. The work shows that these limitations in volunteer translation suggest that one might bridge the gap between trained and untrained translators through collaboration and mutual training in this kind of projects as in [75].

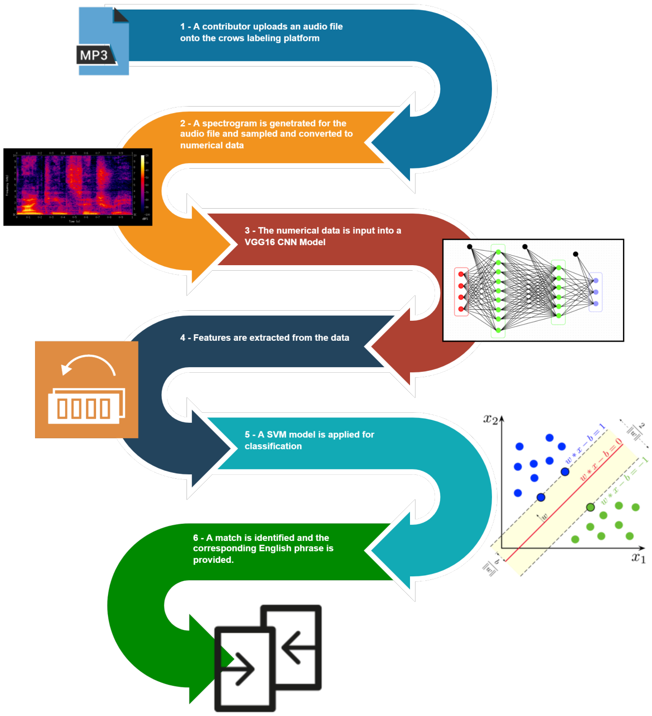

MP3
1 - A contributor uploads an audio file onto the crows labeling platform
2 - A spectrogram is genetrated for the audio file and sampled and converted to numerical data
3 - The numerical data is input into a VGG16 CNN Model
4 - Features are extracted from the data
5 - A SVM model is applied for classification
6 - A match is identified and the corresponding English phrase is provided.
x_2
x_1
$w * x - b = 1$
$w * x - b = 0$
$w * x - b = -1$
$\frac{2}{\|w\|}$
$\frac{b}{\|w\|}$
w

Fig. 8. One Proposed Design of an ML Scheme Application.

To accommodate building the dataset and making it public to the community, only a short period of time was available for the research. The crowdsourcing platform and initial recordings were achieved in nine months. Participants were selected from the close network of school, family, and friends who spoke the Emirati and Egyptian dialects. Time restrictions limited the size of the dataset, which includes a total of 301 Egyptian and 138 Emirati dialects recordings. This resulted in an imbalanced dataset on multiple levels. The Egyptian dialect recordings are almost double the size of the Emirati ones, knowing that the acquaintances of the crowdsourcing developers consisted mainly of Egyptians. On the other hand, by looking at the participants gender figures, 95% of the Emirati participants are male, while 63% of the Egyptian participants are female. Finally, most of the participants (80%) are from the young generation, thereby minimizing the contribution of children, middle age and elder samples of the Emirati and Egyptian populations. This systemic inequity based on age and gender disparities has received a lot of attention recently in voice biometrics [76] and reveals that age and gender subgroups have a significant different voice characteristic. In fact, non-negligible gender disparities exist in speaker identification accuracy and show that the average accuracy can be significantly higher for female speakers than males due to mainly voice inherent characteristic difference [76]. The results in [77] also indicate a significant age effect on the voice acoustic parameters (fricative spectral center of gravity, spectral skewness, and speaking STSD) revealing that certain speech and voice features change with age.

Even though the research was deemed successful by curating the first medical diagnosis dataset for the Emirati and Egyptian dialects, it did fail to completely explore all the possible responses to the emergency diagnosis questionnaire, as not all the questions were presented to the participants in the crowdsourcing platform, and thus many responses missed being recorded in the Arabic dialects, at least in either one of them. In hindsight, participants should have been provided with a wide palette of different responses for them to translate before they were able to stop recording.

In order to protect the rights of the research participants and also to maintain scientific integrity, it was important to take ethical and social considerations into account [78]. The participants were therefore given the option of voluntary involvement, which allowed them to withdraw at any point without any obligations. It was also made clear to them that there were no negative consequences or repercussions to their refusal to participate. The identities of the research participants were kept anonymous, as no personal identifiable data were asked, such as their names, email addresses, phone numbers, photos and videos. The research used data pseudonymization, where the audio recordings of the participants were given artificial identifiers, such as sample001, sample002 etc.

VII. Conclusion

In this work, the implementation of a crowd labeling approach to develop an audio corpus is proposed. An application of the audio corpus to train an ML algorithm, which interfaces with a device in a physician's room is provided. Although results on the accuracy of the classification of this application are preliminary and incomplete due to the small size of the dataset, the implementation approach can be replicated by researchers by splitting the corpus into training and testing sets to evaluate different ML techniques or test a specific ML pipeline. The crowd labeling approach developed here can be extended to other applications, where data is collected from public contributors. Special care has been taken to ensure that no personal identifiable information about the contributors is collected or stored. The future proposed work will collect more audio samples and include additional Arabic dialects. A dashboard will be supporting the dataset to provide statistics on the collected audio samples.

Acknowledgment

The authors would like to thank Hana Abbas, Anan Hefny, Haneen Al Attar, Mohamed Dorrah, Amr Mohamed, Arjun Nair, Habiba Elshahat, and Sherif Reda for helping with this work.

References

[1] K. Gerrish, R. Chau, A. Sobowale, and E. Birks, "Bridging the Language Barrier: the Use of Interpreters in Primary Care Nursing," *Health & Social Care in the Community,* 2004, 12(5), 407-413.

[2] R. F.I. Meuter, C. Gallois, N. S. Segalowitz, A. G. Ryder, and J. Hocking, "Overcoming Language Barriers in Healthcare: a Protocol for Investigating Safe and Effective Communication when Patients or Clinicians use a Second Language," *BMC Health Services Research,* 2015, 15(1), 1-5.

[3] P. A. Ali and R. Watson, "Language Barriers and their Impact on Provision of Care to Patients with Limited English Proficiency: Nurses' Perspectives," *Journal of Clinical Nursing,* 2018, 27(5-6), e1152-e1160.

[4] GMI. United Arab Emirates Population Statistics 2022, 2022, https://www.globalmediainsight.com/blog/uae-population-statistics/. Accessed June 5, 2022.

[5] E. Hadziabdic and H. Katarina, "Working with Interpreters: Practical Advice for use of an Interpreter in Healthcare," *International Journal of Evidence-Based Healthcare,* 2013, 11(1), 69-76.

[6] H. S. Al-Neyadi, A. Salam, and M. Malik, "Measuring patient's satisfaction of healthcare services in the UAE hospitals: Using SERVQUAL." *International Journal of Healthcare Management,* 2018, 11(2), 96-105.

[7] A. Saqer and Q. Alaa, "Language Miscommunication in the Healthcare Sector: A Case Report." *Journal of Patient Safety and Quality Improvement,* 2019, 7(1), 33-35.

[8] C. L. Timmins, "The impact of language barriers on the health care of Latinos in the United States: a review of the literature and guidelines for practice," *Journal of midwifery and women's health,* 2002, 47(2), 80-96.

[9] T. Loney, T. C. Aw, D. G. Handysides, R. Ali, I. Blair, and M. Grivna, "An analysis of the health status of the United Arab Emirates: the 'Big 4' public health issues," *Global Health Action,* 2013, 6(1), 20100.

[10] S. Patil and P. Davies, "Use of Google Translate in Medical Communication: Evaluation of Accuracy," *British Medical Journal,* 2014, 349.

[11] E. Hadziabdic and H. Katarina Hjelm, "Arabic-Speaking Migrants' Experiences of the use of Interpreters in Healthcare: a Qualitative Explorative Study," *International Journal for Equity in Health,* 2014, 13(1), 1-12.

[12] E. Hadziabdic, "The use of Interpreter in Healthcare: Perspectives of Individuals, Healthcare Staff and Families," Linnaeus University Press, 2011.

[13] E. C. Khoong, E. Steinbrook, C. Brown, and A. Fernandez, "Assessing the use of Google Translate for Spanish and Chinese Translations of Emergency Department Discharge Instructions," *JAMA Internal Medicine,* 2019, 179(4), 580-582.

[14] G. Flores, "The Impact of Medical Interpreter Services on the Quality of Healthcare: a Systematic Review," *Medical Care Research and Review,* 2005, 62(3), 255-299.

[15] S. Giordano, "Overview of the Advantages and Disadvantages of Professional and Child Interpreters for Limited English Proficiency Patients in General Health Care Situations," *Journal of Radiology Nursing,* 2007, 26(4), 126-131.

[16] J. A. Rodriguez, A. Fossa, R. Mishuris, and B. Herrick, "Bridging the Language Gap in Patient Portals: An Evaluation of Google Translate," *Journal of General Internal Medicine,* 2020, 1-3.

[17] M. Alhamami, "Language barriers in multilingual Saudi hospitals: Causes, consequences, and solutions," *International Journal of Multilingualism,"* 2020, 1-13.

[18] J. F. Ha and L. Nancy, "Doctor-patient communication: a review," *Ochsner Journal,* 2010, 10(1), 38-43.

[19] L. ML. Ong, J. CJM. De Haes, A. M. Hoos, and F. B. Lammes, "Doctor-patient communication: a review of the literature," *Social Science and Medicine,* 1995, 40(7), 903-918.

[20] P. A. Ali and J. Stacy, "Speaking my patient's language: bilingual nurses' perspective about provision of language concordant care to patients with limited English proficiency," *Journal of advanced nursing,* 2017, 73(2), 421-432.

[21] P. Crawford, B. Brown, and K. Harvey, "Corpus linguistics and evidence-based health communication," *The Routledge handbook of language and health communication,* 2014, 75-90.

[22] S. Adolphs, ”Applying corpus linguistics in a health care context,” *Journal of applied linguistics,* 2004, 1(1).

[23] D. Knight, ”A multi-modal corpus approach to the analysis of backchanneling behaviour,” 2009, Dissertation, University of Nottingham.

[24] Defined.AI. The Challenge of Building Corpus for NLP, 2020, Librarieshttps://www.defined.ai/blog/the-challenge-of-building-corpus-for-nlp-libraries/. Accessed on May 22, 2022

[25] A. Elnagar, ”Systematic literature review of dialectal Arabic: identification and detection,” *IEEE Access,* 2021, 9, 31010-31042.

[26] K. Almeman and M. Lee, “Automatic building of Arabic multi dialect text corpora by bootstrapping dialect words,” in Proceedings of 1st International Conference on Communications, Signal Processing, and their Applications. (ICCSPA), Feb. 2013, 1–6.

[27] R. Boujelbane, M. E. Khemekhem, S. BenAyed, and L. H. Belguith, “Building bilingual lexicon to create Dialect Tunisian corpora and adapt language model,” in Proceedings of 2nd Workshop Hybrid Approaches Translation, 2013, 88–93.

[28] R. Al-Sabbagh and R. Girju, “YADAC: Yet another dialectal Arabic corpus,” in Proceedings of LREC, 2012, 2882–2889.

[29] I. Guellil, ”Arabic natural language processing: An overview,” *Journal of King Saud University-Computer and Information Sciences,* 2021, 33(5), 497-507.

[30] T. Wael, ”Intelligent Arabic-Based Healthcare Assistant,” in Proceedings of 3rd Novel Intelligent and Leading Emerging Sciences Conference (NILES), October 2021.

[31] H. Bouamor, N. Habash, M. Salameh, W. Zaghouani, O. Rambow, D. Abdulrahim, O. Obeid, S. Khalifa, F. Eryani, and A. Erdmann, ”The Madar Arabic Dialect Corpus and Lexicon,” in *Proc. Eleventh International Conference on Language Resources and Evaluation,* Myazaki, Japan, 2018.

[32] M. Embarki and M. Ennaji, *Modern Trends in Arabic Dialectology,* Red Sea Press, 2011.

[33] A. Farghaly and K. Shaalan, ”Arabic Natural Language Processing: Challenges and Solutions,” *ACM Transactions on Asian Language Information Processing,* 2009, 8(4), 1-22.

[34] A. Torfi, R.A. Shirvani, Y. Keneshloo, N. Tavvaf, and E. Fox, ”A Natural Language Processing Advancements by Deep Learning: A Survey,” arXiv preprint arXiv:2003.01200, 2020.

[35] M. Al-Ayyoub, A. Nuseir, K. Alsmearat, Y. Jararweh, and B. Gupta, ”Deep Learning for Arabic NLP: A Survey,” *Journal of Computational Science,* 2018, 26, 522-531.

[36] M. A. Omari and M. Al-Hajj, ”Classifiers for Arabic NLP: Survey,” *International Journal of Computational Complexity and Intelligent Algorithms,* 2020, 1(3), 231-258.

[37] Md. S. Islam, Md. M. Hasan, X. Wang, Xiaoyi, and H. D. Germack, ”A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining,” *Healthcare,* 2018, 6(2), 54.

[38] R. Ooms and M. Spruit, ”Self-Service Data Science in Healthcare with Automated Machine Learning,” *Applied Sciences,* 2020, 10(9), 2992.

[39] A. Y. Samarah, ”Politeness in Arabic Culture,” *Theory and Practice in Language Studies,* 2015, 5(10), 2005-2016.

[40] E. Hadziabdic and K. Hjelm, ”Working with Interpreters: Practical Advice for use of an Interpreter in Healthcare,” *International Journal of Evidence-Based Healthcare,* 2013, 11(1), 69-76.

[41] M. Alhamami, ”Switching of Language Varieties in Saudi Multilingual Hospitals: Insiders’ Experiences,” *Journal of Multilingual and Multicultural Development,* 2020, 41(2), 175-189.

[42] P. Ram, ”Management of Healthcare in the Gulf Cooperation Council (GCC) Countries with Special Reference to Saudi Arabia,” *International Journal of Academic Research in Business and Social Sciences,* 2014, 4(12), 24.

[43] T. Khoja, S. Rawaf, W. Qidwai, D. Rawaf, K. Nanji, and A. Hamad, ”Healthcare in Gulf Cooperation Council Countries: a Review of Challenges and Opportunities,” *Cureus,* 2017, 9(8).

[44] L. Abualigah, H. E. Alfar, M. Shehab, A. Hussein, and M. A. Alhareth, ”Sentiment Analysis in Healthcare: a Brief Review,” *Recent Advances in NLP: The Case of Arabic Language,* 2020, 129-141.

[45] D.M.E.D.M Hussein, ”A Survey on Sentiment Analysis Challenges,” *Journal of King Saud University - Engineering Sciences,* 2018, 30(4), 330-338.

[46] K. Denecke and Y. Deng, ”Sentiment Analysis in Medical Settings: New Opportunities and Challenges,” *Artificial Intelligence in Medicine,* 2015, 64(1), 17-27.

[47] C. L. Ventola, ”Social Media and Health Care Professionals: Benefits, Risks, and Best Practices,” *Pharmacy and Therapeutics*, 2014, 39(7), 491-499.

[48] A. Alnawas, ”The Corpus Based Approach to Sentiment Analysis in Modern Standard Arabic and Arabic Dialects: A Literature Review,” *Politeknik Dergisi*, 2018, 21(2), 461-470.

[49] G. Mendels, “How to Apply Machine Learning and Deep Learning Methods to Audio Analysis,” Medium, Towards Data Science, November 18, 2019, https://towardsdatascience.com/how-to-apply-machine-learning-and-deep-learning-methods-to-audio-analysis-615e286fcbbc. Accessed February 8, 2021.

[50] T. Cooper, “How Shazam Works,” Medium, 29 January, 2018, medium.com/@treycoopermusic/how-shazam-works-d97135fb4582. Accessed February 8, 2021.

[51] J.A. Reyes-Ortiz, B.A. Gonzalez-Beltran, and L. Gallardo-Lopez, ”Clinical Decision Support Systems: a Survey of NLP-Based Approaches from Unstructured Data,” in *26th International Workshop on Database and Expert Systems Applications (DEXA),* Valencia, Spain, September, 2015.

[52] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, ”Secure and Robust Machine Learning for Healthcare: A Survey,” in arXiv, 2020.

[53] M. Nunez, ”Medical Translation and Artificial Intelligence,” *SimulTrans*, 2020. https://www.simultrans.com/blog/medical-translation-artificial-intelligence. Accessed February 24, 2021.

[54] J. TeCho, ”A Corpus-Based Approach for Keyword Identification using Supervised Learning Techniques,” in *Proc. 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology,* Krabi, Thailand, 2008.

[55] V. Bansal, G. Pahwa, and N. Kannan, ”Cough Classification for COVID-19 Based on Audio Mfcc Features using Convolutional Neural Networks,” in *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON),* India, 2020.

[56] S. Arjun, A. Biswas, A. Gandhi, S. Patil, and O. Deshmukh, ”LIVELINET: A Multimodal Deep Recurrent Neural Network to Predict Liveliness in Educational Videos,” *International Educational Data Mining Society,* 2016.

[57] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ”Imagenet Classification with Deep Convolutional Neural Networks,” *Advances in neural information processing systems*, 2012, 25, 1097-1105.

[58] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[59] S. J. Pan and Q. Yang, ”A Survey on Transfer Learning,” *IEEE Transactions on Knowledge & Data Engineering,* 2009, 22(10), 1345-1359.

[60] P. T. Chen, C. L. Lin, and W. N. Wu, ”Big Data Management in Healthcare: Adoption Challenges and Implications,” *International Journal of Information Management,* 2020, 53, 102078.

[61] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, ”Speech Recognition Using Deep Neural Networks: A Systematic Review,” *IEEE Access,* 2019, 7, 19143-19165, doi: 10.1109/ACCESS.2019

[62] N. Haeri, ”Form and Ideology: Arabic Sociolinguistics and Beyond,” *Annual Review of Anthropology,* 200, 29, 61-87.

[63] V. Sergon. Expatica. Introduction to Arabic: the Language of the UAE, March 11, 2021, https://www.expatica.com/ae/education/language-learning/introduction-to-arabic-the-language-of-the-united-arab-emirates-71422. Accessed February 11, 2021.

[64] F.V. Rosse, M.D. Bruijne, J. Suurmond, M. EssinkBot, and C. Wagner, ”Language Barriers and Patient Safety Risks in Hospital Care. A mixed Methods Study,” *International Journal of Nursing Studies,* 2016, 54, 45–53.

[65] S. Fares, F. B Irfan, R. F. Corder, M. A. AlMarzouqi, A. H. AlZaabi, M. M. Idrees, and M. Abbo, ”Emergency Medicine in the United Arab Emirates,” *International Journal of Emergency Medicine*, 2014, 7(4).

[66] S. Patil and P. D. Sumant, ”Use of Google Translate in Medical Communication: Evaluation of Accuracy,” *The BMJ,* 2016, 349(7392).

[67] V. Boddapatia, A. Petefb, and J.L.L. Rasmussonb, ”Classifying environmental sounds using image recognition networks,” *Procedia Computer Science,* 2017, 112, 2048–2056.

[68] S. Seo and S.B. Cho,”Offensive Sentence Classification using Character-Level CNN and Transfer Learning with Fake Sentences,” in *International Conference on Neural Information Processing,* Guangzhou, China, November, 2017.

[69] D. Wang and T.F. Zheng, ”Transfer Learning for Speech and Language Processing,” in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA),* Hong Kong Polytechnic University, Hong Kong, December, 2015.

[70] B. Sertolli, R. Zhao, B.W. Schuller, and N. Cummins, ”Representation transfer learning from deep end-to-end speech recognition networks for the classification of health states from speech,” *Computer Speech & Language*, 2021, 68, 101204.

[71] J.C. Vásquez-Correa, C.D. Rios-Urrego, T. Arias-Vergara, M. Schuster, J. Rusz, E. Nöth, and J.R. Orozco-Arroyave, ”Transfer Learning Helps to Improve the Accuracy to Classify Patients with Different Speech Disorders in Different Languages,” *Pattern Recognition Letters*, In press, 2021.

[72] Y. Chen, B. Gao, L. Jiang, K. Yin, J. Gu and W.L. Woo, and Wai Lok, ”Transfer learning for wearable long-term social speech evaluations,” *IEEE Access*, 2018, 6, 61305-61316.

[73] K. Feng and T. Chaspari, ”Low-resource language identification from speech using transfer learning,” in *29th International Workshop on Machine Learning for Signal Processing (MLSP),* Pittsburg, PA, US, October, 2019.

[74] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, and L. Fei-Fei, ”Imagenet: A large-scale hierarchical image database,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Florida, US, June, 2009.

[75] A. Senn, ”How did participants experience volunteering for a translation crowdsourcing project?,” PhD dissertation, University of Geneva, 2021.

[76] X. Chen, L. Zhengxiong, S. Srirangaraj, and X. Wenyao, ”Exploring racial and gender disparities in voice biometrics,” *Scientific Reports,* 2022, 12(1), 1-12.

[77] S. Taylor, C. Dromey, S.L. Nissen, K. Tanner, D. Eggett, and K. Corbin-Lewis, ”Age-related changes in speech and voice: spectral and cepstral measures,” *Journal of Speech, Language, and Hearing Research,*, 2020, 63(3), 647-660.

[78] E. West, ”Ethics and integrity in nursing research,” *Handbook of Research Ethics and Scientific Integrity,* 2020, 1051-1069.