Computers and Electrical Engineering 102 (2022) 108224

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Automatic illness prediction system through speech☆

Husam Ali Abdulmohsin [a,*], Belal Al-Khateeb [b], Samer Sami Hasan [a], Rinky Dwivedi [c]

[a] *Computer Science Department, College of Science, University of Baghdad, Baghdad, Iraq*
[b] *College of Computer Science and information Technology, University of Anbar, Anbar, Iraq*
[c] *Maharaja Surajmal Institute of Technology, New Delhi, India*

A R T I C L E I N F O

Keywords:
Medical speech transcription and intent dataset
Automatic disease prediction
Mel frequency Ceptstral coefficient
Spectral centroid variability
Forward-backward filter
Neural network

A B S T R A C T

Due to the COVID-19 epidemic and the curfew caused by it, many people have sought to find an ADPS on the internet in the last few years. This hints to a new age of medical treatment, all the more so if the number of internet users continues to expand. As a result, automatic illness prediction online applications have attracted the interest of a large number of researchers worldwide. This work aims to develop and implement an automated illness prediction system based on speech. The system will be designed to forecast the sort of ailment a patient is suffering from based on his voice, but this was not feasible during the trial, therefore the diseases were divided into three categories (painful, light pain and psychological pain), and then the diagnose process were implemented accordingly. The medical dataset named "speech, transcription, and intent" served as the baseline for this study. The smoothness, MFCC, and SCV properties were used in this work, which demonstrated their high representation to human being medical situations. The noise reduction forward-backward filter was used to eliminate noise from wave files captured online in order to account for the high level of noise seen in the deployed dataset. For this study, a hybrid feature selection method was created and built that combined the output of a genetic algorithm (GA) with the inputs of a NN algorithm. Classification was performed using SVM, neural network, and GMM. The greatest results obtained were 94.55% illness classification accuracy in terms of SVM. The results showed that diagnosing illness through speech is a difficult process, especially when diagnosing each type of illness separately, but when grouping the different illness types into groups, depending on the amount of pain and the psychological situation of the patient, the results were much higher.

1. Introduction

Researchers and companies have been attracted to the automatic disease prediction system (ADPS) in recent years because it enables rapid access to medical treatment, avoids traffic and long-distance travel, and is especially beneficial to senior citizens, regardless of the new era the world is entering with the spread of COVID-19 and the periodic curfews implemented by many countries

Abbreviations: ADPS, Automated Disease Prediction System; CPU, Central Processing Unit; GA, Genetic Algorithm; GB, Giga Byte; GMM, Gaussian Mixture Model; MFCC, Mel Frequency Cepstral Co-efficient; NN, Neural Network; RAM, Random Access Memory; RSM, Response Service Methodology; SCV, Spectral Centroid Variability; SVM, Support Vector Machine.

☆ This work is implemented in University of Baghdad, College of science, Computer science department labs, Baghdad, Iraq, not funded.

* Corresponding author.
E-mail address: husam.a@sc.uobaghdad.edu.iq (H.A. Abdulmohsin).

https://doi.org/10.1016/j.compeleceng.2022.108224
Received 30 January 2022; Received in revised form 29 June 2022; Accepted 8 July 2022
Available online 21 July 2022
0045-7906/

to prevent the virus from spreading.

It is critical to understand the human speech production process in order to deal with ADPSs professionally. Phonetics is the study of the sounds produced by human speech. Speech is produced by pushing air from the lungs to the larynx (respiration), where the vocal cords may be open to allow for air flow or may vibrate, producing sound (phonation). The articulators in the mouth and nose, who are responsible for articulation, will affect the lungs' airflow.

Acoustic and articulatory phonetics are two perspectives of spoken sounds. Acoustic phonetics is a subfield of linguistics, but it is also a branch of physics, focusing on the acoustic and physical aspects of sound waves. Articulatory phonetics is the method through which human bodies are employed to produce spoken sounds [1]. Three processes are required to generate speech. To begin, an energy source is required. Anything that makes sound requires an energy source, and in the case of human speaking sounds, the energy source is the air coming from the lungs. The second mechanism is the sound source, which is the vocal cords of the larynx (or vocal folds). The articulators are the third mechanism, and they are used to filter or shape the sound. The oral cavity (mouth space), the nasal cavity (space within and behind the nose), the jaws, teeth, lips, and, most importantly, the tongues are all articulators [2].

Acoustic voiced sounds are produced when air from the lungs passes through the vocal cords, causing the vocal cords to vibrate and produce periodic signals. Due to the unique physical qualities of each human being's vocal cords, distinct frequencies are produced during voice production, which reflects the unique attributes of each human being's acoustics [3]. The epiglottis is the second component of the human speech production system. Each individual's epiglottis has a unique bulk and flexibility [4,5]. This also results in the addition of new elements to speech, regardless of the tong position or the mouth cavity, which both result in various pronunciations for the same words when spoken by different persons. The varied properties of the epiglottis also vary across genders, with female formants being more frequent than male formants and the spectrum of voiced sounds often decreasing in loudness with increasing frequency. All of these acoustic phenomena are a result of speech production. As a result, it is feasible to identify gender-specific elements in acoustic speech signals [6].

According to the source of creation, speech characteristics can be classified as acoustic or articulation features. If a human being is weary or in pain, his articulators will reflect his medical condition, and his sound will be created differently, since speaking requires a great deal of energy, and the first thing that will be impacted in his voice is his fatigue or discomfort [7–9]. The amount of verbal and gestural communication generated is influenced by the severity of acute pain. When pain intensity rises, so does spontaneous verbal communication [10]. Furthermore, research demonstrates that when pain is more acute, not only the use of nonverbal pain behaviors, but also the usage of co-speech gestures, rises in the realm of visible physical communication [11]. These gestures are rich in semantic content and offer crucial information about a sufferer's pain experience. In contrast to pain behaviors, demonstrating that when pain is more acute, people use multimodal communication resources to give greater information about their suffering. [12–14]. This is regarded evidence that any human health condition may be translated through his voice, but the most critical aspect is to select the most pertinent characteristics that will translate the sickness, infection, discomfort, or any other health issue.

Other factors that increase the challenge of health detection in general are the following:

- Certain psychological situations among different individuals, under certain circumstances, can affect the patient's judgment of his health situation. Sickness can be faked.
- To discern health automatically from vocal expressions, we must delete superfluous and dangerous information such as organ noise and human reactions that might result in abnormal voices.
- This work has made several contributions, that overcome the limitations of the state-of-art researches, which are listed below:
- Gender and speaker independency, a new features selection method enabled the system to function independently with respect to gender and speaker, and this is a huge contribution to health detection.
- Accuracy compared to literature survey; high classification accuracy results are achieved through the recognition strength of the features extracted from the feature extraction method deployed. Notably, when applied to the medical speech, transcription, and intent dataset, the method achieves 94.55% classification accuracy, and this is the highest accuracy result gained compared to the state-of-art works.
- Generality, the same system setup was used to detect health for all languages, genders and ages represented in the dataset, and no tweaking was done for each category of people. This demonstrates that the characteristics derived in this work are highly capable of detecting illness and that they function across languages, recording frequencies, genders, and ages and are speaker independent.
- Age and language independent, all the results given in this work were gained by recognizing all limits of ages, all languages, and all genders in the datasets utilized without removing any category from the dataset or merging multiple categories into one group to avoid misclassification. This was a good challenge because all three datasets contained conflicting emotions, including happy and anger, neutral and calm, and disgust and surprise, which are difficult to distinguish.

The related works will be described in Section 2 of this article. The statistics and features of the benchmark datasets used in this research, as well as the frame-work for the proposed technique, will be discussed in depth in Section 3. Section 4 will present and debate the work's highest categorization findings. Section 5 will summarize the findings from this effort.

2. Related Work

Numerous studies have established the presence of distinct acoustic and physiological characteristics in human voices that can be utilized to identify patient symptoms, but they have not yet achieved the needed classification accuracy. Several scholars have investigated ADPSs using a variety of various methodologies and processes.

One of the processes upon which ADPS's rely is on clinical reports to identify the patient's highest expected condition [15]. Another mechanism by which healthcare ADPSs operate is through the question-and-answer approach, which relies on user text input (guided by users). The algorithm will then provide a list of the most likely illnesses [16–19]. Numerous researchers have applied machine learning techniques to the diagnosis of cardiac disease [20]. Another technique is to utilize natural language processing and deep learning to extract text that may be used to identify certain symptoms for the purpose of illness prediction [21–23] and several additional studies. According to a 2014 study conducted by Maree Johson, nearly 20,000 state-of-art papers in the field of voice to text ADPS were published in 2014. These works utilized computational linguistics, natural language processing, human language technology, or text mining [24]. According to the literature study undertaken for this work, the suggested work belongs to the field of ADPS since it does not include linguistics, natural language processing, or speech conversion for the purpose of text mining. This paper developed an ADPS that estimates the degree of pain a patient is experiencing based just on his speech, without requiring any additional language processing. Numerous difficulties were encountered, one of which was the psychological state of the human person and its influence on voice. When a person is psychologically unstable, his vowels are stated differently than when he is stable [25,26]. The other difficulty was dealing with the similarity of children's voices between the ages of three and seven years; hence, children were excluded from the studies used to validate the approach described in this study, and since children rely on adults to speak out for them. Noise and distortion were a challenge in many of the recordings in the used dataset, as they will be in real life. The suggested effort has several limits that must be established, such as its interaction with youngsters and its lack of familiarity with other languages. We were unable to identify the voice signal for each condition in this work, but we were able to differentiate between diseases that have psychological effects and those that do not, as well as diagnose diseases based on the intensity of pain, strong and mild pain.

3. Dataset Selection

The dataset used in this study is version 1 of the medical voice, transcription, and intent dataset. The collection is comprised of spoken descriptions in (wav) audio format that detail the patient's medical problems. Each vocal description is accompanied with a transcription in the (csv) text format, and then each is tagged with a specific type of illness. This is a single-language dataset comprised of North American English slang. It enumerates 25 symptoms, all of which are back pain, acne, blurry vision, body feeling weakened,

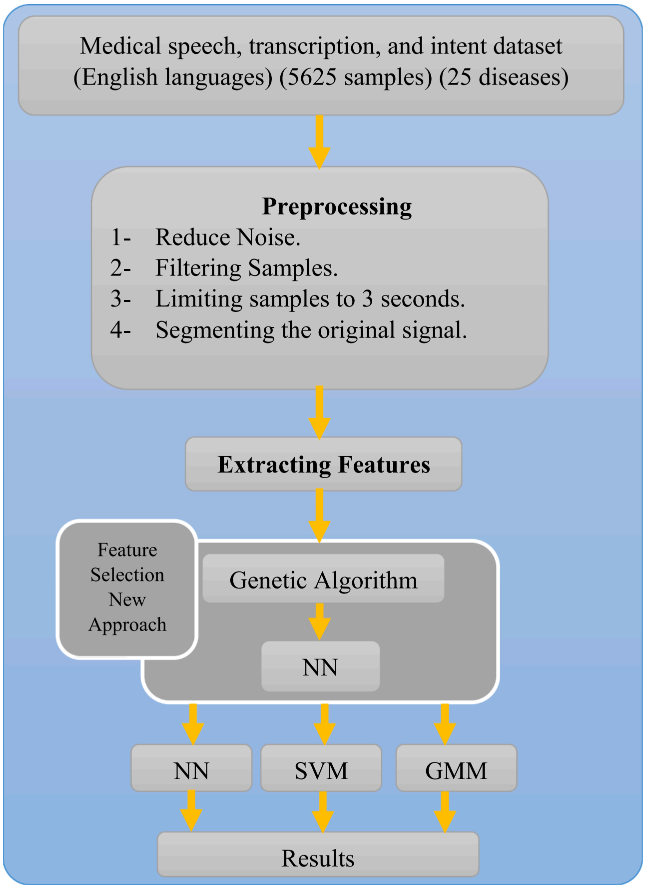

Fig. 1. The block diagram of the method proposed in this work.

cough, ear ache, emotional pain, feeling cold, feeling dizzy, foot ache, hair falling out, difficulty breathing, head ache, heart hurts, infected wound, injury from spots, internal pain, joint pain, knee pain, muscle pain, neck pain, open wound, shoulder pain, skin issue, and stomach ache. Each symptom has a unique number of recordings, which required us to disregard some of the recordings for certain symptoms in order to get an equal number of recordings for each symptom, which is 225 records for each symptom. This work has a total of 5625 recordings [27]. We found a lot of distortion and noise in some of the wave recordings throughout our study, which we attribute to the low quality of the microphones used to record those samples and the loud setting in which they were recorded. This necessitated the use of a forward-backward digital filter, dubbed the zero-phase filter, in all recordings. The justification for adopting zero-phase filtering is because it retains filtered features exactly where they appeared in the unfiltered signal, which has no effect on classification performance. These filters conduct zero-phase digital filtering in both directions, forward and backward. After filtering the data in the forward direction, the filter sequence will be reversed, re-filtering the signal [28].

As we have addressed most of the main concepts of health detection systems, we will now explain our framework proposed in this research in detail.

4. Framework

The suggested technique is depicted in Fig. 1 using a block diagram. The dataset and the suggested method's major phases will be illustrated in depth in the following four sections.

4.1. Preprocessing Phase

Numerous processes were applied during this stage. To begin, there is the noise reduction step. The noise reduction forward-backward filter was used to eliminate noise from wave files recorded online in order to account for the significant level of noise observed in several samples from the medical speech, transcription, and intent dataset utilized in this study. Second, the process of sampling and filtering. Each of the 25 illnesses included in the collection had at least 225 samples, and several had significantly more. Thus, we picked 225 samples for each illness and eliminated samples with excessive noise based on our listening to all samples, which was a lengthy procedure. Trimming was the third procedure. The majority of the recordings were made under various time constraints. As a result, all samples were shortened to three seconds. Segmentation was the fourth procedure. All samples were segmented at a ratio of (0.05%) to the original signal, and the overlap ratio used in this study was (0,025%), which results in a (50%) overlap, and the total number of segments created will be (2n-1), where n is the number of original segments.

4.2. Feature Extraction

A feature is a quantifiable quality derived from the substance under observation [29]. The most critical component of feature extraction is identifying the aspects that are most relevant to the issue statement. In this work, the smoothness, mel frequency cepstral coefficient (MFCC) with 12 degrees, and spectral centroid variability (SCV) features were extracted. These features are strongly related to human beings experiencing pain, and were chosen based on experiments conducted on 15 different types of speech features. The mean and standard deviation in multiple degrees were calculated for each feature, therefore, each feature generated 8 features, representing the (mean, and seven different degrees for the standard deviation. As a results, 112 features were generated from the feature extraction phase, all were utilized through experiment (8 smoothness features, 96 MFCC features, and 8 SCV features).

If smoothness is defined as the transaction of speech via air, then smooth speech results in a delayed transaction, whereas rough speech results in a faster transaction [30]. Smoothness was determined in two domains: the temporal domain and the spectral domain. Smoothness is determined using Eq. (1) and (2) [31].

$$GV_t = \frac{1}{P}\sqrt{\sum_{j=1}^{P}(var_t(j))^2} \tag{1}$$

$$GV_s = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(var_s(i))^2} \tag{2}$$

Where var_t and var_s denote the temporal and spectral domain variances of the spectral feature, P denotes the feature's dimension, and N denotes the feature's time domain length

4.3. Feature Selection

The evolutionary feature selection method (genetic method) was used in this study was to filter feature groups before giving them to the neural Network (NN), in order to choose the best feature groups according to a given criterion, and then to pass the best selected feature group to the NN.

The idea of the genetic algorithm was to weight each feature group (each two features) with a fitness function, that shows the classification power of that certain group, and the weight will be used to combine the groups to form new ones. Weighted groups will be checked, and merged, until the criterion is fulfilled, at which point the best group will be handed to the NN. The neural network is used for classification or as an integrated feature selection technique, and it is also utilized in data mining and machine learning [32,33]. Three considerations motivated the employment of a filter ranking algorithm in this work. To begin, filter out irrelevant variables. Second, to take advantage of the variable selection criteria based on the order of the variable ranking approaches. Finally, and maybe most importantly, their simplicity and success as reported by online apps. Each variable is scored using a ranking criterion, and then a threshold is determined through experimentation and utilized to exclude variables that fall below that threshold [29].

Filter feature selection techniques are those that are applied before to classification; hence, ranking methods are considered filter methods. The fundamental premise behind feature selection methods is to pick unique features that convey important information about various classes in the dataset by leveraging a feature's basic attribute. This quality is referred to as feature relevance, and it quantifies a feature's ability to categorize distinct classes [29,34,35].

4.4. Neural Network

In this study, the support vector machine (SVM), backpropagation NN, and Gaussian mixture model (GMM) classifiers were used to predict the patient's discomfort. The structure of the neural network was two hidden neural network layers, the first layer with 80 hidden nodes, and the second layer with 40 hidden nodes. The input were 25 diseases and the output were 3 groups of diseases. The fitness function of the NN, was a NN that evaluated the classification accuracy of each sub group. Three kernel functions used with the SVM, were the polynomial kernel of degree, calculated through Eq. (3):

$$\mathrm{p},\ \left(x_i,\ x_j\right)^{\mathrm{p}} \tag{3}$$

The radial basis function kernel, calculated through Eq. (4):

$$\mathrm{e}^{-\mathrm{a}||x_i\ -\ x_j||2},\ \text{with a} > 0 \tag{4}$$

and the sigmoid function, calculated through Eq. (5):

$$\tanh\ \left(a.\ x_i\ .\ x_j + c\right) \tag{5}$$

The radial kernel function achieved the highest accuracies.

5. Hardware and Software Platform

All tests were conducted in MATLAB R2019a on a computer equipped with an Intel Core i7-8750H Central Processing Unit (CPU) running at 2.2 GHz and 32 GB of Random Access Memory (RAM) (RAM). The MATLAB instructions are executed on the Windows 10 platform.

6. System Setting and Tuning

All of the findings described will be derived using the same system parameters and will not be changed for each sickness category contained in the dataset. The objective was to design a system capable of addressing a wide variety of languages and accents using the same parameters in order to achieve generalizability. The variables that were fixed during the system's setup were tweaked until the highest classification accuracy was achieved, but then they were fixed. Those most important variables are the following:

- **The length of the window dividing the signal** specifies the length of the segment that will be extracted from the original signal in order to extract the desired characteristics. The window length was calculated through Eq. (6):

$$\text{Window}\ _{Length} = 0.05\ \text{X Frequency} \tag{6}$$

- **The length of the step** sets the overlap between the current and previous windows in order to capture the majority of the possible divisions of the original signal. The step length was calculated through Eq. (7):

$$\text{Step}\ _{Size} = 0.025\ \text{X Frequency} \tag{7}$$

- **The standard deviation degree** sets the number of degrees needs for the standard deviation to achieve the highest accuracy.

There are other variables that were fixed, but we have discussed the most important ones.

7. Experimental Results

Many experiments were conducted through our work, two of the most important experiments will be discussed that achieved the highest results.

7.1. Experiment number 1 (Exp1)

The purpose of this experiment was to anticipate the disease represented in each wave recording and to classify each recording according to one of the 25 diseases represented in the dataset. The confusion matrix of the best accuracy classification results obtained using the three classifiers SVM, NN, and GMM is shown in Figs. 2, 3 and 4. As shown in the three confusion matrices, the classification accuracies produced by the SVM, NN, and GMM were 50.8 %, 48.7 %, and 31.5 %, respectively.

The three classifiers used produced poor results that are unacceptable in any ADPS. Through analyzing the results some conclusions were gained, such as:

1 The classification accuracy of each individual disease lye near the mean of the overall classification accuracy, which shows the difficulty of gaining high classification accuracy from this approach.
2 It has been noticed that some diseases have been misclassified with a certain group of diseases and not with all diseases which led to the idea of clustering the diseases into three different groups. For example, in Fig. 2, column (1), ear ache, foot ache, head ache, Infected wound, injury from spots, internal pain, joint pain, knee pain, muscle pain, neck pain, open wound, shoulder pain, and stomach ache wave samples have been misclassified with back pain by (9, 14, 11, 8, 7, 9, 11, 8, 9, 12, 23, 10, and 7) wave samples. Whereas, the all-other diseases were not misclassified classified with back pain in that high number of wave samples, the misclassifications with other diseases were (1, 1, 0, 0, 1, 1, 0, 1, 4, 2, and 1), which shows that a certain number of diseases share the same effect on speech, different than other diseases.
3 The highest classification accuracy gained from the SVM classifier was 70.2% with the heart hurts disease, which make sense, because any pain in the heart affects the breath of the human being, and as results, will affect his speech.
4 There are some characteristics that are shared among similar recordings of the same condition, as seen by the findings of Exp1. Regardless of the modest findings obtained, the 50.8 % classification accuracy indicates that there are some common characteristics among the recordings, but these characteristics need to be increased.
5 We discovered that illnesses are classified into three groups based on the misclassification distribution of samples in the confusion matrices. Each illness within the same group is categorized incorrectly as another disease within the same category. After analyzing the disorders, we discovered that they may be classified into three categories based on the level of discomfort they produce.
6 There are certain diseases that affect the human voice, primarily due to the amount of pain they cause. These diseases include (painful diseases, back pain, internal pain, joint pain, knee pain, muscle pain, neck pain, open wound, shoulder pain, stomach ache, injury from spots, infected wound, head ache, and ear ache. Foot discomfort) (which will be referred to as Group 1 in Exp2), and these disorders have an effect on the phonetics of articulation. There are several disorders that impact the acoustic phonetics of

back pain	75	1	0	2	0	11	1	1	0	14	0	2	7	1	8	8	12	13	14	12	12	13	9	1	13	32.6
acne	1	119	12	23	0	0	2	3	1	1	16	1	0	0	0	0	0	1	1	1	1	0	1	11	1	60.7
blurry vision	1	21	127	21	1	1	1	1	1	1	17	2	0	1	0	3	1	1	0	0	1	0	0	14	1	58.5
body feels weak	0	24	19	126	1	0	0	1	0	0	13	1	1	0	0	0	0	1	1	1	1	0	1	14	1	61.2
cough	0	1	1	0	129	4	16	12	9	1	0	12	0	8	0	0	1	1	1	0	1	1	0	1	1	64.5
* ear ache	9	0	1	1	1	99	0	0	1	8	0	0	8	1	7	12	11	9	11	14	9	10	10	1	12	42.1
emotional pain	1	0	0	0	22	0	131	11	13	1	1	11	1	10	0	1	1	0	0	0	1	1	0	0	2	63.0
feeling cold	1	1	0	1	15	0	17	144	11	0	0	13	0	9	0	0	0	1	0	1	0	1	1	0	3	65.8
* feeling dizzy	0	1	0	2	16	1	23	13	135	0	1	11	1	9	1	0	0	0	0	0	1	2	0	1	1	61.6
foot ache	14	1	1	0	0	12	0	1	0	108	0	0	10	0	12	9	11	13	13	10	11	14	9	1	17	40.4
hair falling out	1	18	21	19	1	0	0	1	1	1	157	0	3	0	1	0	1	1	1	0	1	0	0	27	1	61.3
heard to breath	4	0	2	2	15	0	16	15	14	0	1	150	1	17	1	1	0	1	2	0	0	0	1	0	0	61.7
*head ache	11	2	3	0	1	14	0	1	2	12	0	1	105	1	8	6	9	8	7	11	5	34	13	2	14	38.9
heart hurts	2	0	0	0	17	0	13	12	15	0	2	13	1	158	1	0	1	0	2	0	0	1	1	1	3	65.0
*Infected wound	8	1	1	0	1	8	0	0	1	7	0	0	13	1	116	10	7	7	7	8	6	4	7	2	7	52.3
injury from spots	7	1	0	1	0	9	0	0	2	7	0	1	11	0	7	105	18	7	12	7	5	5	8	1	6	47.7
internal pain	9	2	0	0	1	8	0	1	2	7	1	0	7	1	8	8	76	10	8	8	8	6	9	1	8	40.2
joint pain	11	1	0	1	0	9	1	1	3	7	0	1	8	1	7	8	12	78	18	7	9	4	7	1	5	39.0
knee pain	8	1	1	0	0	8	0	0	3	9	1	1	8	1	8	9	11	13	72	8	9	6	32	0	6	33.5
muscle pain	9	2	1	0	1	8	1	0	3	7	0	1	7	1	7	8	6	7	11	97	9	7	22	0	7	43.7
neck pain	12	0	1	0	1	7	0	1	2	7	0	1	9	1	8	8	11	11	11	12	111	7	9	0	7	46.8
open wound	23	1	1	0	0	7	1	1	2	8	0	1	9	2	9	7	14	16	12	10	9	96	9	0	7	39.2
shoulder pain	10	1	0	1	1	8	1	1	3	9	0	1	8	0	6	12	14	16	13	12	8	9	62	1	6	30.5
skin issue	1	25	32	24	1	0	0	1	1	1	12	1	0	2	0	2	0	1	0	0	0	0	1	144	0	57.8
*stomach ache	7	1	1	1	0	11	1	3	0	9	3	0	7	0	10	8	8	9	8	6	7	4	13	1	96	44.9
	33.3	52.9	56.4	56.0	57.3	44.0	58.2	64.0	60.0	48.0	69.8	66.7	46.7	70.2	51.6	46.7	33.8	34.7	32.0	43.1	49.3	42.7	27.6	64.0	42.7	50.1

Fig. 2. The Confusion Matrix of the best classification accuracy gained from predicting 25 diseases in Exp 1 using SVM.

back pain	71	0	2	0	0	9	2	0	2	9	1	2	7	0	6	10	19	8	11	8	10	11	10	2	8	34.1
acne	1	86	31	11	0	0	2	3	2	2	23	0	2	0	2	2	2	1	0	1	2	2	1	11	1	45.7
blurry vision	0	23	56	34	0	1	3	3	3	3	25	0	3	0	2	3	1	0	0	3	2	0	1	13	0	31.3
body feels weak	1	25	33	86	0	1	0	3	1	1	12	0	2	0	2	4	2	0	0	1	2	2	1	18	1	43.4
cough	4	2	0	0	112	3	11	15	17	0	2	20	2	12	4	2	1	0	0	0	2	1	1	2	0	52.6
* ear ache	9	2	1	0	3	71	0	2	0	12	1	3	8	3	7	11	14	11	15	11	13	12	11	4	8	30.6
emotional pain	0	2	1	0	25	2	167	13	19	0	2	21	2	9	2	1	1	0	0	0	3	0	1	2	1	60.9
feeling cold	0	3	1	0	24	3	8	139	16	0	1	19	2	9	3	0	0	0	0	1	0	0	1	2	1	59.7
* feeling dizzy	0	4	2	1	18	2	9	13	112	2	0	23	0	12	1	4	2	1	1	1	3	1	0	0	0	52.8
foot ache	8	4	3	2	1	21	0	0	1	58	0	5	7	0	8	11	16	12	11	9	9	10	14	0	10	26.4
hair falling out	0	17	33	32	2	1	0	2	0	2	82	0	4	1	2	2	0	0	0	2	2	2	0	17	1	40.2
heard to breath	1	2	4	3	17	2	11	12	18	3	2	95	0	9	1	1	0	1	1	0	2	0	0	0	1	51.1
*head ache	9	3	2	4	2	23	0	3	1	11	3	1	121	1	7	13	13	9	13	16	8	14	9	2	2	41.7
heart hurts	1	3	2	4	14	1	11	11	17	3	4	26	2	159	2	1	0	1	1	2	0	0	0	0	0	60.0
*Infected wound	8	2	3	9	2	8	0	0	0	8	3	2	7	1	111	15	8	8	7	8	8	9	8	2	8	45.3
injury from spots	11	3	1	1	1	7	0	0	2	9	4	0	8	0	7	66	7	12	7	9	9	9	12	0	9	34.0
internal pain	13	2	2	2	1	8	0	0	1	7	3	0	6	0	7	10	61	11	7	7	12	6	14	0	10	32.1
joint pain	13	4	2	1	1	9	0	0	2	18	4	0	0	0	7	10	31	107	8	7	9	9	12	0	7	41.0
knee pain	10	3	1	0	0	7	0	0	1	12	3	0	8	1	8	9	8	10	105	7	8	6	11	1	9	46.1
muscle pain	12	2	0	0	1	8	0	0	1	11	2	0	7	2	7	11	6	5	8	106	9	17	14	0	8	44.7
neck pain	10	3	0	0	0	9	0	1	1	22	0	1	6	1	8	9	9	8	8	6	77	9	12	1	9	36.7
open wound	18	2	0	0	0	8	0	1	2	12	0	0	7	0	7	9	7	4	8	5	14	83	12	0	8	40.1
shoulder pain	14	2	0	0	0	8	0	0	3	17	0	1	6	0	6	8	7	8	8	5	11	14	70	1	8	35.5
skin issue	0	24	45	34	1	0	0	2	2	2	45	3	2	2	1	2	1	3	0	1	2	0	0	145	1	45.6
*stomach ache	11	2	0	1	0	13	1	2	1	3	3	3	6	3	7	11	9	5	6	9	8	8	10	2	114	47.9
	31.6	38.2	24.9	38.2	49.8	31.6	74.2	61.8	49.8	25.8	36.4	42.2	53.8	70.7	49.3	29.3	27.1	47.6	46.7	47.1	34.2	36.9	31.1	64.4	50.7	43.7

Fig. 3. The Confusion Matrix of the best classification accuracy gained from predicting 25 diseases in Exp 1 using NN.

human speech, which may be classified into two categories. To begin, psychological ailments such as acne, blurred eyesight, physical weakness, hair loss, and skin problems (which will be called Group 2 in Exp2). Second, disorders associated with frequency such as (cough, emotional discomfort, feeling chilly, difficulty breathing, heart ache, and dizziness) (which will be called Group 3 in Exp2).

Following the conclusion of our trials, we observed that several disorders had been clustered together. Each category of illnesses is misclassified within itself and not in relation to other groupings. When we classified those diseases into categories, as indicated in Table 1, we discovered that some behaviors were shared by diseases within the same category. Certain disorders are accompanied by pain, while others produce no discomfort but impair the human being's emotional state, and yet others have a direct effect on the vocal folds, affecting the frequency of speech. As a result, Experiment 2 was conducted.

7.2. Experiment number 2 (Exp2)

The purpose of this experiment is to classify illnesses into three categories. Where each set of disorders has a similar pattern of behavior or effect on the human body or mind and speech. Following the implementation of this experiment, some conclusions were reached, such as:

1- High classification accuracy was achieved using all three classifiers used in this study, although the SVM classifier achieved the highest classification accuracy, as seen by the confusion matrix in Fig. 5.
2- Figs. 6 and 7 illustrate the classification results of the NN and GMM classifiers, respectively.
3- Figs. 5, 6, and 7, still show that there is a partial misclassification between the three prevented groups of disease, and this can be justified by the poor representation of the illness in the wave recordings, or there might be some intersection in the features extracted from the three different groups of diseases, that caused this misclassification.
4- The least classification accuracy was gained through the GMM, through classifying group 2, that was misclassified with group 1 via 164 wave samples.
5- No uniform misclassification was found in Exp 2, all three groups are misclassified with each other. the only uniform found was, group 1, achieved the highest classification accuracy with respect to all three classifiers, 96.83%, 94.35%, and 94.92%, with SVM, NN, and GMM, respectively.

back pain	34	2	3	2	1	11	1	1	2	12	0	2	19	1	8	9	15	16	14	12	14	11	9	0	9	16.3
acne	2	56	34	24	2	2	1	2	1	2	18	1	2	2	1	1	2	0	1	0	1	1	1	22	1	31.1
blurry vision	0	30	54	33	1	2	0	1	2	3	21	1	2	1	1	2	3	0	1	0	1	1	0	21	1	29.7
body feels weak	3	31	27	70	1	1	1	2	1	2	21	0	3	1	0	1	2	0	1	0	1	1	0	19	1	36.8
cough	4	2	3	3	68	4	20	18	20	2	1	28	0	11	0	1	2	1	1	1	1	1	0	0	0	35.4
* ear ache	13	3	4	5	2	56	3	2	1	12	2	2	19	1	9	8	11	9	10	10	12	8	12	0	7	25.3
emotional pain	0	3	1	3	23	2	88	15	28	2	0	29	0	17	0	1	1	1	0	1	3	0	2	0	0	40.0
feeling cold	4	2	2	2	20	3	23	78	29	2	1	21	0	18	0	0	1	0	1	0	0	1	2	1	0	37.0
* feeling dizzy	2	0	2	2	33	2	23	27	59	0	3	27	1	14	0	0	0	0	0	13	0	0	0	0	1	28.2
foot ache	9	0	2	2	1	22	1	3	2	58	0	3	16	2	11	12	12	16	15	9	15	16	15	1	10	22.9
hair falling out	1	32	24	24	2	3	1	1	2	0	73	1	1	3	0	2	1	0	1	1	1	1	0	32	2	34.9
heard to breath	1	2	4	4	22	3	27	24	33	0	1	67	1	20	1	0	1	0	0	0	2	1	0	1	0	31.2
*head ache	12	4	2	2	4	21	2	3	1	11	31	0	54	2	15	10	11	6	12	8	8	11	7	1	12	21.6
heart hurts	0	4	4	4	21	4	24	21	31	1	1	29	0	111	0	0	1	0	0	2	1	1	0	1	0	42.5
*infected wound	10	2	2	2	3	8	1	2	3	14	19	1	18	0	101	10	13	8	8	9	9	10	9	0	7	37.5
injury from spots	14	3	3	1	2	12	1	2	2	15	0	1	12	2	8	66	13	18	9	11	8	17	9	1	8	27.7
internal pain	17	4	4	3	4	11	1	3	1	11	0	0	13	3	8	20	56	15	11	12	7	11	8	1	7	24.2
joint pain	15	2	3	2	5	11	0	2	1	10	0	1	11	2	9	10	14	68	12	11	9	15	9	0	7	29.7
knee pain	13	1	4	2	3	7	1	4	1	9	0	0	10	4	8	15	11	11	65	10	11	16	9	2	7	29.0
muscle pain	14	3	2	2	2	6	0	2	1	8	1	1	9	2	9	14	15	11	14	71	9	14	13	4	7	30.3
neck pain	17	2	2	0	1	7	1	4	1	13	1	2	9	5	8	13	11	9	14	9	70	13	10	1	8	30.3
open wound	14	0	2	0	1	8	1	2	0	11	1	3	9	1	8	8	10	9	13	16	11	50	10	2	9	25.1
shoulder pain	16	0	3	0	0	8	0	3	3	11	0	0	8	0	9	11	10	14	12	10	13	11	88	1	9	36.7
skin issue	1	35	30	32	1	4	4	1	0	1	29	1	0	1	0	0	0	0	1	1	1	0	3	112	0	43.4
*stomach ache	9	2	4	1	2	7	2	2	0	15	1	4	8	1	11	11	9	13	9	8	17	14	9	2	112	41.0
	15.1	24.9	24.0	31.1	30.2	24.9	39.1	34.7	26.2	25.8	32.4	29.8	24.0	49.3	44.9	29.3	24.9	30.2	28.9	31.6	31.1	22.2	39.1	49.8	49.8	31.7

Fig. 4. The Confusion Matrix of the best classification accuracy gained from predicting 25 diseases in Exp 1 using GMM.

Table 1
The groups of diseases generated after analyzing the results of Exp1.

No.	Acoustic phonetic feature related disease Frequency Related	Psychological Related	Articulator phonetics feature related disease Painful diseases
1	cough	acne	back pain
2	emotional pain	blurry vision	internal pain
3	feeling cold	body feels weak	joint pain
4	heard to breath	hair falling out	knee pain
5	heart hurts	skin issue	muscle pain
6	feeling dizzy		neck pain
7			open wound
8			shoulder pain
9			stomach ache
10			injury from spots
11			infected wound
12			head ache
13			ear ache
14			Foot ache

Group1	3050	50	69	96.24%
Group 2	42	1056	26	93.95%
Group 3	58	19	1255	94.22%
	96.83%	93.87%	92.96%	94.55%

Fig. 5. The Confusion Matrix of the best classification accuracy gained from predicting 3 groups of diseases using SVM.

Group1	2972	113	72	94.14%
Group 2	91	961	35	88.41%
Group 3	89	51	1243	89.88%
	94.35%	85.42%	92.07%	90.62%

Fig. 6. The Confusion Matrix of the best classification accuracy gained from predicting 3 groups of diseases using NN.

Group1	2990	164	146	90.61%
Group 2	76	904	39	88.71%
Group 3	84	57	1167	89.22%
	94.92%	80.36%	86.44%	87.24%

Fig. 7. The Confusion Matrix of the best classification accuracy gained from predicting 3 groups of diseases using GMM.

6- Group 2 have achieved the lowest classification accuracies with NN and GMM, it might be justified by its poor wave samples, but as long as the same groups have achieved the second place in classification with SVM, so that justification is not totally true, but the right justification is, the three group of diseases, still have common features, that are leading to this misclassification.

The limitation of the proposed work, was the time needed by the genetic algorithm to select the best group of features. Adding to this limitation, through experiment, it wasn't applicable to diagnose the 25 diseases represented in the dataset, therefore, it was a must to group identical diseases in three different groups. Those groups were fixed through try and error during experiment. Diseases that are related to acoustic phonetic feature were determined, and diseases related to articulator phonetics feature were also determined. Then the first was divided to frequency related diseases and psychological related diseases.

8. Conclusion

Diagnosing illnesses via speech is a difficult process that was previously unattainable, but with the advent of machine learning, everything is conceivable. Our objective was to diagnose illnesses directly from human speech using a machine learning approach, however we were unable to get the desired results, despite the 50.1 percent classification accuracy achieved using the SVM classifier, but on the other hand, high accuracy result of (94.55%) was gained when divided the diseases into three related groups.

This investigation revealed that illnesses may be classified into several categories, including severe and light discomfort, psychological or physical, emotional or non-emotional, and others. Each of the above groups exhibits similar behavior when it comes to communication. This reflection leaves distinct characteristics associated with each form of sickness.

The advantage of the proposed work is the capability of illness diagnoses through speech, which is considered a new approach in this field, and is extremely required in certain applications were speech is the only source of information available to diagnose the illness of a patient.

Through experiment, the proposed work faced a couple of limitation. First, the time, the needed by the genetic algorithm was high. Second, the failure of the proposed work to diagnose each of the 25 diseases represented in the dataset utilized, which means the failure is extracting features that can diagnose each of the 25 diseases separately.

Future Work

The findings of Exp1 can be increased by using more feature types or machine learning methods. Although the genetic technique utilized in Exp2 produced positive findings, it was a time-consuming procedure. Thus, developing a more efficient feature selection approach can accelerate the ADPS.

As the challenges and the extreme number of random factors of this field have been discussed in the introduction section, the response service methodology (RSM) program can be used to predict the perfect setting to the system proposed, by suggesting the experiments with the highest probability of achieving the best classification results.

Searching for features that are more disease related, that can diagnose each disease separately, will be more applicable, because when patients use ADPS's, they usually are looking for exclusive answers, not broad answers.

The input of this work is only speech, but adding other data such as human hand or arm movement as an input, and using both types of data as combination, and predicting the disease accordingly, might increase the classification accuracy, since human bodies translate pain through moving parts of it, such as moving their arms, hand, shoulders, toughing their foreheads or legs.

Declaration of Competing Interest

None.

Data availability

The authors are unable or have chosen not to specify which data has been used.

References

[1] Zhang Z. Mechanics of human voice production and control. J Acoust Soc Am 2016;140(4):2614–35.
[2] Carr P. English phonetics and phonology: An introduction. John Wiley & Sons; 2019. p. 382.
[3] Anderson C. Essentials of Linguistics. Pressbooks by McMaster University; 2018.
[4] Rabiner LR, Schafer RW. Digital processing of speech signal. Pearson; 1978. p. 528.
[5] Deller JR, Proakis JG, Hansen JH. Discrete time processing of speech signals. Prentice Hall PTR; 1993.
[6] Harrington J, Cassidy S. The acoustic theory of speech production, in Techniques in Speech Acoustics. Springer; 1999. p. 29–56.
[7] Rose P. Forensic Speaker Identification, 1. London: Taylor Francis; 2002. p. 380.
[8] Kreiman J, Papçun G. Voice discrimination by two listener populations. J Acoust Soc Am 1985;77(S1). S9–S9.
[9] EUiott JR. Auditory and F-pattern variations in Australian okay: a forensic investigation. Acoust Aust 2001;(29):41.
[10] Vickers ER, Cousins MJ, Woodhouse AJ. Pain description and severity of chronic orofacial pain conditions. Aust Dent J 1998;43(6):403–9.
[11] Labus JS, Keefe FJ, Jensen MP. Self-reports of pain intensity and direct observations of pain behavior: when are they correlated? Elsevier Sci B.V. 2003;102 (1–2):109–24.
[12] Sullivan MJ, Pascal T, André S, Richard C, John K, William S. The influence of communication goals and physical demands on different dimensions of pain behavior. Elsevier Sci B.V. 2006;125(3):270–7.
[13] Rowbotham S, Holler J, Lloyd D, Wearden A. How do we communicate about pain? A systematic analysis of the semantic contribution of co-speech gestures in pain-focused conversations. J Nonverbal Behav 2012;36(1):1–21.
[14] Rowbotham S, Holler J, Lloyd D, Wearden A. Handling pain: The semantic interplay of speech and co-speech hand gestures in the description of pain sensations. Speech Commun 2014;57:244–56.
[15] Wang X, Chused A, Elhadad N, Friedman C, Markatou M. Automated knowledge acquisition from clinical narrative reports. In: AMIA Annual Symposium Proceedings. American Medical Informatics Association; 2008.
[16] Hossain M, Laskar M, Rahman T. Automated disease prediction system (ADPS): a user input-based reliable architecture for disease prediction. Int J Comput Appl 2015.
[17] Dragu N, Elkhoury F, Miyazaki T, Morelli RA, di Tada N. Ontology-based text mining for predicting disease outbreaks. In: Twenty-Third International FLAIRS Conference; 2010.
[18] Maude J. Patients could provide initial differential. Br J Gen Pract 2021.
[19] Shepperd S, Charnock D, Gann B. Helping patients access high quality health information. Br Med J (Clin Res Ed) 1999;319(7212):764–6.
[20] Das R, Turkoglu I, Sengur A. Effective diagnosis of heart disease through neural networks ensembles. Expert Syst Appl 2009;36(4):7675–80.
[21] Kumar R, Pradhan S, Rebaka T, Prakash J. Disease Prediction from Speech Using Natural Language Processing and Deep Learning Method. In: Congress on Intelligent Systems. Springer; 2020.
[22] Dreisbach C, Koleck TA, Bourne PE, Bakken S. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. Int J Med Informatics 2019;125:37–46.
[23] Gangavarapu T, Krishnan GS, Kamath S, Jeganathan J. FarSight: long-term disease prediction using unstructured clinical nursing notes. IEEE Trans Emerg Topics Comput 2020;9(3):1151–69.
[24] Johnson M, Lapkin S, Long V, Sanchez P, Suominen H, Basilakis J, et al. A systematic review of speech recognition technology in health care. BMC Med Inform 2014;14(1):1–14.
[25] Shaqra FA, Duwairi R, Al-Ayyoub M. Recognizing emotion from speech based on age and gender using hierarchical models. Procedia Comput Sci 2019;151: 37–44.
[26] Abdulmohsin HA, Wahab HA, Hossen AMJA. Speech Emotion Recognition Survey. J Mech Conti. Math Sci 2020;15(9):24.
[27] Mooney, P., medical speech, transcription, intent, 2018, 1.
[28] Gustafsson F. Determining the initial states in forward-backward filtering. IEEE Trans Signal Process 1996;44(4):988–92.
[29] Chandrashekar G, Sahin F. A survey on feature selection methods. Comput Electr Eng 2014;40(1):16–28.
[30] Fant G. Acoustic theory of speech production. Walter de Gruyter; 1970.
[31] Nghia PT, Van Tao N, Huong PTM, Diep NTB, Hien PTT. A Measure of Smoothness in Synthesized Speech. REV Journal on Electronics Communications in Mathematical Physics 2016;6(1-2):35–9.
[32] Yoon H, Park C-S, Kim JS, Baek J-G. Algorithm learning based neural network integrating feature selection and classification. Expert Syst Appl 2013;40(1): 231–41.
[33] Ledesma S, Cerda G, Avina G, Hernandez D, Torres M. Feature selection using artificial neural networks. In: Mexican International Conference on Artificial Intelligence. Springer; 2008. p. 351–9.
[34] Abdulmohsin HA, Wahab HBA, Hossen AMJA. A New Hybrid Feature Selection Method Using T-test and Fitness Function. CMC-COMPUTERS MATERIALS CONTINUA 2021;68(3):3997–4016.
[35] Kohavi R, John GH. Wrappers for feature subset selection. Artif Intell 1997;97(1–2):273–324.

Asst. Prof. Dr. Husam Ali Abdulmohsin (M'41) was born in Baghdad, Iraq in 1979. This author became a Member (M) of IEEE in 2020. He received the B.S. degree in 2000 and M.S. degrees in 2003, both in computer science from the University of Al-Nahrain, Baghdad, Iraq. Fulfilled his Ph.D degree in the computer science department, Technology University, Baghdad, Iraq.

Prof. Dr. Belal Al-Khateeb received the B.Sc. (honors) (first class) degree in computer science from Al-Nahrain University, Baghdad, IRAQ, in 2000, and the M.Sc. degree in computer science from Al-Nahrain University, Baghdad, IRAQ, in 2003, and the Ph.D. degree from the School of Computer Science, University of Nottingham, Nottingham, U.K., in 2011.

Asst. Prof. Dr. Samer Sami hasan was born in Baghdad, Iraq in 1978. This author became a Member (M) of IEEE in 2014. He received the B.S. degree in 2000 and M.S. degrees in 2003, both in computer science from the University of Al-Nahrain, Baghdad, Iraq. Fulfilled his Ph.D degree in the computer science department, UKM University, SELANGOR, MALAYSIA.

Dr Rinky Dwivedi has completed her B.Tech in Computer Science and Engineering from Guru Gobind Singh Indraprastha University, Delhi in 2004 and M.E. in Computer Technology and Application from Delhi College of Engineering, Delhi in 2008. She has received her Doctorate in 2016 from Delhi Technological University, New Delhi. Dr. Rinky has over 19 years of experience in Academics.