Genetic Algorithms Application to Recognize the Arabic Back Obstruent Sounds in Continuous Speech

Aissiou Mohammad
Electronics Department, Senior lecturers at National Higher School of Advanced Technologies, Algeria
aissoummaissoumm@yahoo.fr

Abstract: *Our work concerns the application of Genetic Algorithms (GAs) to recognize the eight Arabic back obstruents regardless of their positions in syllables. Speech recognition was been cast as a pattern classification problem where we would like to classify an input acoustic signal into one of all possible phonemes. Since GAs are widely used population-based evolutionary search algorithms, they were applied at the Acoustic-Phonetic Decoding (APD) step of the Automatic Speech Recognition (ASR) domain. The Modern Standard Arabic (MSA) is characterized by the presence of glottal, pharyngeal and velar sounds called back consonants. They belong to the class of obstruents and are distinguished from other sounds by their place of articulation, which is defined as a set of anatomical locations ranging from the palate to the glottis. These consonants are the subject of this study because of the difficulties encountered to categorize and classify them correctly since their rear articulation points make them difficult to recognize. The used GA is characterized by parametric matrices based-evaluation function during which intervenes by operating modifications of the vectors parameters representing the phonemes and by selecting the best ones. Hence, we performed a chronological classification of matrices representative of speech segments based on adapted genetic modeling. The research focuses on both the empirical adjustment of GA parameters and the judicious choice of genetic reproduction operators on which the reliability of the genetic recognition algorithm and the overcoming of the premature convergence phenomenon largely depend. The experimental results demonstrate that the proposed methods achieves interesting performances compared to other conventional speech recognition ones*.

Keywords: *Speech processing, acoustic-phonetic decoding, automatic speech recognition, genetic algorithms searching, arabic back obstruents*.

Received April 10, 2025; accepted October 28, 2025
https://doi.org/10.34028/iajit/23/3/13

1. Introduction

Speech recognition, known as Automatic Speech Recognition (ASR) is the ability of a machine to identify phonemes, words and phrases in spoken language and convert them to a machine-readable format. The main purpose of ASR systems is to recognize natural languages that are spoken by human beings. However, a more recent description defines ASR as the designing and action of systems in mapping acoustic signals to a string of speech units [35, 64]. Speech recognition is the task that will be addressed in this work.

The recognition of continuous speech is an attempt to transcribe naturally spoken utterances. The fluency of speech in natural speech imposes co-articulation between adjacent phonemes and words in a phrase. This leads to neglecting some phonemes, especially between words, which makes the recognition process very difficult to achieve. An ASR system is composed of two important stages; feature extraction and pattern matching.

A typical speech recognition system is developed with major components that include acoustic front-end, acoustic model, lexicon, language model and decoder. Acoustic front-end takes care of converting the speech signal into appropriate features, which provides useful information for recognition.

The functionality of automatic speech recognition system can be described as an extraction of a number of speech parameters from the acoustic speech signal for each word or sub-word unit. The word or phoneme patterns are stored and later when a word or sub-word unit is to be recognized its pattern is compared to the stored patterns and the word that gives the best match is selected. This technique is generally, referred to as pattern recognition.

Acoustic front-end involves signal processing and feature extraction. In speech recognition, the main goal of the feature extraction step is to compute a parsimonious sequence of feature vectors providing a compact representation of the given input signal.

Acoustic modeling of speech typically refers to the process of establishing statistical representations for the feature vector sequences computed from the speech waveform.

ASR methodologies are broadly classified into three approaches, namely, acoustic-phonetic approach pattern 6matching recognition approach and artificial intelligence approach.

The intention of this paper is the application of Genetic Algorithms (GAs) at the acoustic phonetic decoding level of ASR to recognize Modern Standard Arabic (MSA) back obstruent sounds in continuous speech regardless of their position and in syllables.

In the recent years, metaheuristic algorithms are used to solve real-life complex problems arising from different fields such as economics, engineering, politics, management, and engineering [3]. Among the metaheuristic algorithms, GAs is a well-known algorithm, which is inspired from biological evolution process. GA mimics the Darwinian theory of survival of fittest in nature [4].

Genetic algorithms are useful to search a space for solutions in multidirectional way from large spaces and poorly defined one. The main advantage of using GAs is their ability to expand the search space for the best parameters by applying the principle of maintenance and manipulation of a large population [4, 5].

GA plays a good role in selecting the fittest parameter set in feature selection, phonetic decoding of ASR. In ASR, GA improves the accuracy by using clustering-based fitness function for choosing the elementary population [6].

It is a population based search algorithm for solutions. The new populations are produced by iterative use of genetic operators on individuals present in the population. The chromosome representation, selection, crossover, mutation, and fitness function computation are the key elements of GA [4, 7].

GA dynamically changes the search process through the probabilities of crossover and mutation and reached to optimal solution. GA can modify the encoded genes, evaluate multiple individuals and produce multiple optimal solutions. Hence, GA has better global search capability [9].

In this research, GAs were used to assign objects that are acoustic segments to well-defined classes, by designing a single-objective supervised genetic classification model. It is conceived to overcome the MSA phonetic complexities. To achieve this goal, we relied on the process of adjusting GA parameters and the choice of genetic representation, evaluation function and genetic operators.

During the feature extraction phase, the relevant speech parameters were extracted. They constitute the parametric matrices coefficients. These matrices are reduced and relevant representations of both back obstruent consonants to be recognized and the speech continuum phonemes. We represent the acoustic segments of speech by the Mel-Frequency Cepstral Coefficients (MFCC).

The remainder of this paper is organized as follows:

Section 2 is dedicated to work already done in this area. Section 3 is an overview of the MSA phonetic system while section 4 is dedicated to Arabic speech processing with feature representation and acoustic modelling of MSA back consonants in continuous speech continuum. In section 5 distance classifiers were studied. In section 6, we detailed GA modelling and highlighted GA tuning for Arabic ASR. Various types of crossover, mutation, selection, and encoding techniques are discussed. The choose of these reproduction operators and their rates were discussed based on series of experiments. Sections 4 and 6 presents the methodology used to conduct this work. Section 6 in part and section 7 in its entirety are dedicated to the presentation of the research results and their interpretations.

2. Related Work

Many serious efforts have been done to develop Arabic speech recognition systems. Numerous collaborative studies have been carried out to explore this area. [11].

Several overviews and survey studies have been published to review various aspects of Arabic speech recognition. [12, 13]. These surveys show that few freely available continuous speech corpora exist and the available services and toolkits for the development of Arabic speech recognition systems.

Several works were presented according to the recognized entity (phonemes, syllables, words, and speaker), the nature of Arabic language (classical, modern standard, dialectal), the speech representations (discrete, continuous, spontaneous), the feature selection technics, the traditional and artificial classifiers and learning techniques. These studies present Arabic speech recognition systems from the end-to-end methodology perspective. The study focuses on two types of the Arabic language, namely MSA and dialectal Arabic [13, 14].

2.1. Basic Classification Models for ASR

There are different classification approaches; acoustic-phonetic approach, pattern recognition approach and artificial intelligence approach [112].

Most conventional speech recognition systems, including those for Arabic speech recognition are generally based on statistical approaches such as Hidden Markov Models (HMMs) that perform the modeling of the sequential structure of speech signals [15, 130]. The HMMs are mainly based on Gaussian Mixture Models (GMMs) that are considered the best statistical representation of the spectral distributions of speech waveforms. GMM involves modeling the acoustic features of speech using a combination of Gaussian distributions. The GMM is widely used in ASR systems for its ability to approximate complex probability distributions effectively [16, 17].

A hybrid algorithm based on combination of HMM and learning vector. Learning Vector Quantization (LVQ) method showed an important contribution in producing highly discriminative reference vectors for classifying static patterns.

Support Vector Machines (SVMs) is one of the popular techniques for pattern recognition and is considered to be the state-of-the-art tool for linear and nonlinear classification. Recently, SVMs has been adopted to perform the task of speech recognition. SVMs can be implemented independently or as a hybrid model [18]. However, the two major problems with SVMs for ASR application are the inability to handle variable inputs, and the high computational cost in classifying more than two classes at once.

Dynamic Time Warping (DTW) has been found useful in many other disciplines, including data mining, gesture recognition, robotics. Dynamic Time Warping is an algorithm for measuring similarity between two sequences, which may vary in time or speed [8].

Artificial Neural Networks (ANN) are great classifiers, and they produce the best results for pattern recognition problems. They are used for their capability to learn and organize according to the dataset provided at the training stage. They work exceptionally well with unknown data and can classify unknown data effectively. There are various types of ANN.

Convolutional Neural Networks (CNNs) are widely used in Arabic speech recognition to capture local patterns and features in speech data. Using CNN facilitates feature learning and classification, enhancing Arabic ASR performance [19, 20].

Recurrent Neural Networks (RNNs) represent a neural network design frequently employed for handling sequential information, like speech signals [21, 22]. In the context of Arabic ASR, an RNN can be utilized to convert the input speech signal into a sequence of phonemess, which can then be further processed to recognize the spoken words [24, 25].

Deep Neural Networks (DNNs) are used to improve the automatic recognition speech by combining the convolutional CNNs and Long Short-Term Memory (LSTM) neural networks [26, 27].

2.2. Genetic Algorithms Applied for ASR

GAs were applied both to optimize basic models for ASR and as searching algorithm for solutions at feature selection or at the evaluation level. However, there are few works in the literature using GAs as classifier for ASR to classify and recognize speech sounds and especially very few works dealing with the automatic recognition of MSA back consonants [28, 29].

As an optimization algorithm in hybrid approach, GAs were used intensively in many works to enhance the ASR [32, 33].

2.3. Hybrid Classifiers

Speech recognition hybrid systems combine two core approaches: traditional methods like HMM-based methods, SVMs, ANNs or evolutionary algorithms like GAs and tabu search algorithm [32, 33].

As classifiers, hybrids models such as HMM-ANN, HMM-SVM, HMM-GMM, HMM-DNN, HMM-AGs, ANN-AGs and end-to-end deep learning models have been emerged as a breakthrough in improving the performance of ASR. They have better performance than the classical ones alone. Hybrids methods enhance the speech recognition rate independently to feature extraction technics “Table 1”.

Table 1. Comparison of classifiers performances for ASR.

Speech recognition approaches		**Range of average rate of ASR**	**References**
Classical approaches	HMM	[81%, 92,92%]	[34, 35, 36 , 37, 38, 39]
	SVM	[76,15%, 86,50%]	[40, 41, 42, 43, 44, 45]
	DTW	[68%, 81.75%]	[46, 47, 48, 49, 50, 51, 52]
Neuronal approaches CNN/ DNN/ ANN		[85,15%, 95,26%]	[53,54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 20,,21 24, 25, 26, 143]
Mata-heuristics/ Evolutionary algorithms Tabu Search		[89,85% , 98, 15%]	[32, 33, 67 , 69, 70]
Hybrid approaches	HMM/ANN	[94,19%, 97%]	[71, 72, 73, 31, 74, 127]
	HMM/GMM /SVM/LVQANN	[81%, 96,47%]	[72, 73, 74, 18, 75]
	HMM/AG	[93,32 %, 98,48%]	[76, 77, 78, 79, , 31]
	ANN//AG DNN/AG	[93%, 98,35%]	[29, 30, 80, 81]

2.4. Comparison of Classifiers Performance

Table 1 shows that the best classifiers for the ASR are the hybrid models and the evolutionary ones under the same conditions. Among them, HMM/ANN and ANN/AG are the best models in term of recognition average rate, followed by the evolutionary algorithms such as AGs. The neuronal based methods show too interesting scores [82].

Hybrid speech classifiers allowed for significant gain in performance with respect to standard models in a variety of situations. In addition, hybrid systems are easily implemented since they require much less parameters than ASR systems based on standard classifiers [83, 84].

3. Overview of Arabic Phonetic System

The study and classification of the sounds of speech is called phonetics. The phonetic system of the Arabic language contains 34 phonemes composed of 6 vowels and 28 consonants. They can be classified as sonorants or obstruents. Speech sounds involving turbulent noise are referred to in generative phonology as obstruents called turbulent sounds, a natural class subsuming stops, affricates and fricatives. Obstruent sounds are produced with a vocal tract configuration sufficient to increase the air pressure inside the mouth significantly over that of the ambient air. Sonorants subsume vowels, glides, liquids, nasals. The originality of Arabic phonetics is based on emphatic, pharyngeal, and laryngeal consonants, and huru:f l mad, because they give a special value to the language. Arabic has a relatively large number of speech sounds whose primary or secondary articulation lies in the pharynx. Among these sounds, there are the pharyngeal and the emphatic or pharyngealized class [85, 86].

3.1. Arabic Back Obstruent Sounds

The MSA is characterized by the presence of back glottal, pharyngeal and velar consonants, which are distinguished from other consonants by their place of articulation, defined as a set of anatomical locations ranging from the palate to the glottis. Pharyngeal, glottal, and uvular consonants are called back consonants relative to their places of articulation these sounds are extremely abundant in Arabic [98].

Despite the long history of research on the acoustic characteristics of back consonants, current state-of-the-art ASR systems are still incapable of performing accurate fine phoneme distinctions for these classes of sounds. One of the main reasons for this is the dynamic, short, speaker and context dependent nature of these sounds. The information that exists in the literature is neither sufficient nor consistent enough to be integrated in an ASR system [88, 89].

The International Phonetic Alphabet (IPA) transcription was chosen for its simplicity "Table 2".

Table 2. Transcription of MSA back obstruent consonants to IPA [90].

MSA back consonants (in Arabic)	MSA back consonants (API)
[ء]	[?]
[ع]	[ɛ]
[غ]	[γ]
[ه]	[h]
[ح]	[ħ]
[خ]	[X]
[ق]	[q]
[ك]	[k]

Arabic back consonants are divided into plosive and fricative sounds according to the manner of articulation. Plosive sounds differ from each other according to their place of articulation and the same is true for constrictives. The eight back obstruent consonants, which are the subject of the study, are illustrated in Table 3.

Table 3. Arabic back obstruents from an articulatory.

	Place of articulation of MSA back obstruents			
Articulatory mode	**Post palatal**	**Velar uvular**	**pharyngeal**	**Laryngeal or glottal**
Occlusive or plosive	[k]	[q]		[?]
Constrictive or fricative		[X]	[ħ] voiced	[h] voiced
		[γ] voiced	[ɛ] voiced	

Plosives are also called oral stops. Oral stops can be either voiced or voiceless. Nasal stops are usually voiced. A plosive consonant completely cuts off the airflow through the mouth. This means that there is no airflow at all for the duration of the sound. In a Fricative consonant, the articulators involved in the constriction approach get close enough to each other to create a turbulent airstream [23].

Pharyngeal phonemes use the pharyngeal cavity in the vocal tract as place of articulation. Articulating empathic phonemes involves the rising and retracting of the back of the tongue [87, 92].

3.2. Databases for Arabic Speech Recognition

Databases are an essential part of speech recognition since classification process relies on the labeled data. Quality of the data affects the success of the recognition process. Incomplete, low-quality, or faulty data may lead to incorrect predictions; hence, data should be carefully designed and collected.

The main challenge lies in the limited availability of suitable datasets for Arabic ASR. Due to this scarcity, individual researchers have had to create their own speech corpora, but unfortunately, these datasets have not been publicly shared. As of now, our knowledge indicates the existence of seven available corpora for Arabic ASR systems, which are composed of complete Arabic speech datasets.

For this work, the speech database is the Algerian Arabic Speech Database (ALGASD) was used, a Modern Standard Arabic (MSA) speech corpus composed of utterances pronounced by 300 Algerian native speakers selected from eleven regions of Algeria. One of the objectives of this corpus design is to be representative of the regional accents of MSA spoken in Algeria. Useful information related to the speakers, such as gender, age, and education level, is provided [93].

The experiment was conducted using a medium-sized corpus of one hundred naturally spoken sentences and words in continuous MSA language. This corpus is spoken by several Algerian male speakers in a low-noise environment. It contains more than one hundred well-balanced back obstruent consonants, present in all three possible positions: initial, medial, and final.

4. Arabic Speech Processing

Whether recognition is global or analytical, the speech recognition process begins with acoustic pre-processing of the speech signal, the goal of which is to reduce the flow of information and eliminate redundancies present within it [93].

4.1. Speech Signal Pre-Processing

The pre-processing of the speech signal is mainly performed to make the speech signal ready for the feature extraction phase of ASR. The speech signal is pre-processed in these ordered steps. In the first step, Voice Activity Detection (VAD) is executed to keep only the voice parts of a speech signal. Then, noise is reduced [96, 97].

In the second step, the pre-emphasis is applied to flatten the magnitude spectrum and balance the high and low-frequency components of the speech signal. On the other hand, pre-emphasis filtering is used for reducing the high dynamic range of speech waveform and enhancing the signal-to-noise ratio. After that, the framing of the whole signal is applied. The framing step fills this gap by splitting the continuous speech signal into a series of blocks called frames of 20-40ms length to allow block processing of the speech signal [99, 100].

In our case, the speech signal, after its acquisition by microphone, is filtered and sampled. A band-pass filter with cut-off frequencies 100Hz and 4.8 KHz was used to filter speech signal before processing. The sampling rate was set to 10 KHz with 16-bit resolution for all recorded speech tokens.

The last step of speech pre-processing is the application of a windowing function and a normalization of the vocal tract to improve the signal spectrum of the speech signal. A 256-point Hamming window was used to select the data to be analyzed [101].

4.2. Feature Extraction and Acoustic Modelling

Acoustic feature extraction is a vital step in Arabic ASR, as it involves converting the raw speech signal into a set of relevant and compact features that effectively represent the linguistic content of the speech [104].

The speech signal has a non-stationary nature, the properties of a signal are not static over time which makes its analysis a very difficult task, however, in short time intervals, the speech signal is considered as a stationary signal and easy to process. Hence, the use of short-term analysis methods in continuous speech processing is justified since the speech continuum is framing into a certain number of stationary short segments. [101, 103]

The time-domain waveform of a speech signal carries all of the auditory information. However, to extract statistically significant information from the waveform, methods must be capable of condensing the information of every segment into a limited number of parameters, or characteristics, while keeping the signal's discriminating power [102].

Distinctive phonetic features have an important role in Arabic speech phoneme recognition. In a given language, they are extrapolated from acoustic features using different methods [102].

These features should characterize each segment so that other segments with similar features may be grouped by comparing their characteristics [51]. Various approaches for extracting features speech signals have been developed based on extensive research in mathematics, acoustics, and speech technology.

MFC analysis is the most popular feature extraction technique for speech recognition [103]. Its popularity stems from the fact that it tries to mimic the human hearing system]. MFCC features have been employed to keep the phonetically crucial aspects of the speech signal [93]. By converting traditional frequency to Mel scale, MFCC considers human perception for sensitivity at appropriate frequencies. The results in coefficients indicate features in the speech signal, such as power, pitch, and vocal tract shape [104, 117].

Cepstral analysis technic was used to extract vocal tract coefficients from the Arabic speech signal. The temporal variations in the speech signal are represented by the first and second derivatives of the twelve MFCCs and energy. In addition, the Voicing Onset Time (VOT) is appended to MFCCs features to properly characterize the back obstruents [104].

These discriminating parameters, as an acoustic matrix form, constitute the set of reference data for the used AG. Each back obstruent consonant denoted x is represented by a reference parameter matrix denoted Mrx. It is obtained by the concatenation of a certain number denoted m_x of the representative short-term parametric vectors denoted Vrx_{40k}. The speech continuum is segmented into a number of acoustic segments lasting 10 ms and are represented by short-term parametric vectors composed of 40 distinctive coefficients.

$$Vrx_{40k} = (\lambda_{1k}, \lambda_{2k}, \ldots, \lambda_{40k}) \tag{1}$$

The specific order reference parameter matrices denoted q_{40mx}, which represent the MSA back obstruent consonant, acoustically, are obtained during the learning phase.

$$Mrx = \begin{pmatrix} \lambda_{11} \cdots & \lambda_{21} & \cdots \lambda_{i1} & \cdots\cdots & \lambda_{401} \\ \cdots\cdots\cdots & & & & \\ \lambda_{1k} \cdots & \lambda_{2k} & \cdots \lambda_{ik} & \cdots\cdots & \lambda_{40k} \\ \cdots\cdots\cdots & & & & \\ \lambda_{1m} \cdots & \lambda_{2m} & \cdots \lambda_{im} & \cdots\cdots & \lambda_{40m_x} \end{pmatrix} \tag{2}$$

$$q_{40mx} = 40 \times m_x \tag{3}$$

To recognize a back obstruent sound in continuous speech, we segment the vocal continuum into a set of short-term acoustic vectors and we create a parametric matrix denoted *Msx* by successively concatenating mx short-term vector parameters of the same type denoted $Vsx_{40k,mx}$. The acoustic parameters matrix size is equal to the order denoted q_{40mx} of the parametric reference vector of the correspondent consonant and it is expressed as follows:

$$Vsx_{40k,mx} = (\beta_{1k}, \ldots, \beta_{1i}, \ldots\ldots, \beta_{40k}) \tag{4}$$

m_x: number of acoustic segments concatenated to form a parametric vector of a possible back obstruent consonant.

$$Msx = \begin{pmatrix} \beta_{11} \cdots & \beta_{21} & \cdots \beta_{i1} & \cdots\cdots & \beta_{401} \\ \cdots\cdots\cdots & & & & \\ \beta_{1k} \cdots & \beta_{2k} & \cdots \beta_{ik} & \cdots\cdots & \beta_{40k} \\ \cdots\cdots\cdots & & & & \\ \beta_{1m_x} \cdots & \beta_{2m_x} & \cdots \beta_{im_x} & \cdots\cdots & \beta_{40m_x} \end{pmatrix} \tag{5}$$

Msx: parametric matrix. It is obtained by the concatenation of a set of short-term parametric vectors of 'm_x' segments of the vocal continuum.

The acoustic segmentation of the vocal continuum was performed automatically by scanning this continuum from the left to the right. It was done chronologically by concatenating "m_x" short-term parametric vectors with a step of 10 ms.

5. Distance Classifiers

The job of a classifier is to learn the relationship between the given input audio features, and their corresponding text or phonemes. They are first trained using the training data, which should be big enough for a classifier to recognize the specific patterns present in the speech signal and their correspondence to the output phonemes.

In the realm of artificial intelligence, particularly within the field of machine learning and pattern recognition, classification techniques serve as fundamental building blocks. These techniques are used to categorize data into predefined classes based on input features. One pivotal aspect of classification is the distance metric, which are mathematical functions used to measure the similarity or dissimilarity among the samples within data and quantifies the separation between two points in a feature space.

For the acoustic classification process, a metric criterion, called decision rule assigns acoustic feature matrices of phonemes to one of existent classes [105].

The distance function between two vectors *x* and *y* is a function *d*(*x*, *y*) that defines the distance between both vectors as a non-negative real number. There are numerous distance metrics, such as Manhattan also called city block, Minkowski, Mahalanobis, Canberra, squared Euclidean and so on [91, 106].

Minkowski distance measures is a family of distances that includes three distance metrics that are special cases corresponding to different values of *p* for this power distance. The Minkowski distance, which is also known as *Lp* norm, is a generalized metric. It is defined as:

$$D(x,y)=\left(\sum_{i=1}^{m}\left|x_i-y_j\right|^{p}\right)^{\frac{1}{p}} \tag{6}$$

where *p* is a positive value.

When *p*=2, the distance becomes the Euclidean distance. When *p*=1 it becomes Manhattan distance. Chebyshev distance is a variant of Minkowski distance where *p*=∞. As we increase *p*, it gives more emphasis to large differences in individual dimensions [107].

The Mahalanobis distance accounts for unequal variances as well as correlations between features, it will adequately evaluate the distance by assigning different weights or importance factors to the features of data points [91, 108].

Euclidean distance is the most popular among them [109]. Recent studies suggest improving the performances for the distance metrics to make machine-learning algorithms more accurate [108, 109].

For this work, among various distance metrics the Euclidean distance is chosen as distance measurement to represent the GA evaluation function because it is more adaptable and applicable to voice data. For calculating the distance between two sample x and y with m attributes, it calculates the Euclidian distance using the following Equation (7) [110]:

$$D(x,y)=\left(\sum_{i=1}^{m}\left|x_i-y_j\right|^{2}\right)^{\frac{1}{2}} \tag{7}$$

Once vectors representation are obtained, several traditional classifiers established in vector spaces can be used. For categorical or mixed features, city block distance is used, and for continuous features, the Euclidean distance is chosen. In the case of datasets with more categorical attributes Euclidean, Squared Euclidean and Manhattan metrics generate relatively more reliable classes than the others [110, 111]. It is formulated as follows:

$$d_E=d\left(Msx,\ Mrx\right)=\left(\sum_{i=1}^{40}\sum_{k=1}^{m_x}\left|\left(\beta_{ik}-\alpha_{ik}\right)/\sigma\right|^{2}\right)^{\frac{1}{2}} \tag{8}$$

σ: standard deviation; distance normalization coefficient.

When comparing distance metrics mentioned in this section, the best performances are achieved by using Euclidean and Mahalanobis distance metric. They performed the same recognition rate beacause when the features are uncorrelated, the distance under a Mahalanobis distance metric is identical to that under the Euclidean distance metric [109]. In addition, the classification accuracy achieved by using city block distance metric is very close to Euclidean and squared Euclidean ones "Table 4".

Table 4. Comparison of average recognition rates according to the distance Metric types under the best GA adjustment.

Distance metric types	Overall rate of phonemes classification
Euclidean distance	93.95 %
City block distance	93.85 %
Mahalanobis distance	93.94 %
Squared euclidean distance	92.25 %
Canberra distance	90.75 %

Regarding the impact on the execution time of our algorithm. The city block distance metric requires less execution time for similarity score calculation and phoneme classification as compared with the other proposed distance metrics.

6. Genetic Algorithm Modelling

GAs are general methods of problem solving by searching solutions [111].

6.1. Genetic Encoding of Speech

Chromosome representation encodes problem solutions as individuals. Although initial versions of GA uses binary numbers for encoding of chromosomes but their major drawbacks such as lack of precision and existence of "Hamming-cliff" problem impel to look for another type of encoding scheme. To overcome these difficulties related to binary encoding of continuous parameter optimization problems, Real-Encoding GA (REGA) of chromosomes is used. It is used to represent a gene in terms of values. It is a case of Direct Value Encoding (DVE). It are used to code genes in a chromosome and maximize the number of schemata available for genetic processing [58].

Real GA is most suitable in a continuous search space since it uses the direct representations of the design parameters and avoids any intermediate encoding and decoding steps When the representation of the solutions is very close to the natural formulation of many problems, e.g., there are no differences between the genotype (coding) and the phenotype (search space), increasing the GA's speed [58].

In addition, it preserves problem structure and reduces search space. It has many attractive properties such as high precision, effective large space search, simple computing, fast convergence and suitable coding scheme (one parameter represents one gene) [59].

Each potential solution is represented in the form of a string, which is dubbed a chromosome or an individual. A pool of strings forms a population. We used a fixed length chromosome for the GA, where each chromosome contains, as genes, the coefficients of several segments noted S_k of the vocalic continuum of the chosen corpus.

The GA individuals are chosen relatively to the entity to be recognized. The search space are set of all possible individuals, which refers to the candidate solutions collection to a problem. Each individual is represented by a chromosome that can be either a matrix or a vector [59]. Once encoded, these entities will constitute the population of our GA. Thus, for each back obstruent consonant, we obtain specific and distinct individuals.

6.2. Genetic Evaluation Function

Once the problem is encoded in a chromosomal manner a measure evaluation function or fitness function is chosen. Each chromosome provides a potential solution to the problem at hand. However, these solutions do not have the same degree of relevance at all. The evaluation function causes the population to evolve in a way that is beneficial to finding the best solution. In other words, it must be able to assign each individual a positive indicator, representing his or her relevance to the problem we are trying to solve. The evaluation function assigns scores to individuals, guiding the GA towards the optimum solution. Consequently, the influence of this function on GA performance is significant. Therefore, it is imperative to choose this function wisely [113].

In this work, to evaluate the population, the fitness of each individual or chromosome is calculated. It represents the differences between the sound to be recognized and each sound in the database. Whenever the distance value is near to zero then the phoneme is recognized. Its formula is derived from the Euclidean distance as shown in Equation (6), it is applied to the evaluation function using the Equation (7). If the fitness value is close to one, the best match occurs between the parametric matrix Msx of the test and the parametric matrix Mrx of the reference files and vice versa [114].

6.3. Genetic Operators of Reproduction

Every search algorithm needs to address the exploration and exploitation of a search space. Exploration is the process of visiting entirely new regions of a search space, whilst exploitation is the process of visiting those regions of a search space within the neighborhood of previously visited points. In order to be successful, a search algorithm needs to establish a good ratio between exploration and exploitation. Too much exploration can lead to a slow convergence, as the algorithm spends too much time on less promising areas of the search space. Conversely, too much exploitation can cause the algorithm to converge prematurely on suboptimal solutions A good balance between them maintains the diversity of the search space permitting the GA to jump out of local or suboptimal solutions to avoid premature convergence and therefore to avoid reducing the system's performance [115, 59].

Genetic reproduction is implemented using three operators: selection, recombination, and mutation. These operators represent the main tools to achieve this good ratio. A balance between the two explorative and exploitative features of GA, which are characterized by the two operators, crossover and mutation, is imperative in order to speed up the search process to produce quality GA solutions [58, 116].

The key to a successful implementation of GAs primarily depends on the efficient crossover and mutation search operators to guide the system toward global optima by locating the global minimum or maximum of the GA evaluation functions [59].

A basic process of a GA begins by randomly generating solutions or chromosomes to the problem. Posteriorly, an iterative process is carried out in which, at each step, the good solutions are selected and the crossing between species is carried out. Occasionally we can have mutations on certain solutions. Through the selection of good solutions in the iterative process, the computer will develop more better solutions [120, 121].

The selection operation is a fundamental process in GAs that determines which individuals from the current population will contribute to the creation of the next generation. The selection operation must balance exploration (searching through new and diverse feature subsets) and exploitation (focusing on the best-performing subsets found so far) [121].

The primary purpose of the selection operation is to promote the survival and reproduction of individuals with higher fitness scores, while less fit individuals are less likely to be chosen [118, 121]. Several selection methods are commonly used in GAs, each with its own characteristics and suitability for different types of problems: roulette wheel selection: also known as fitness-proportionate selection, this method assigns a selection probability to each individual based on its fitness score. Individuals with higher fitness have a greater chance of being selected, similar to how segments of a roulette wheel are sized according to the fitness values. While this method is straightforward, it may suffer from issues when fitness values are very similar, leading to slow convergence. Tournament selection: in this method, a subset of individuals is randomly chosen from the population, and the one with the highest fitness is selected for reproduction. This process is repeated until the required number of individuals is selected. Tournament selection is easy to implement and can be adjusted by changing the tournament size, offering a balance between exploration and exploitation. Rank-based selection: instead of using raw fitness scores, individuals are ranked based on their fitness, and selection probabilities are assigned based on these ranks [59, 122].

Crossover operators combine the genomes of two or more solutions with the possibility that good chromosomes may generate better ones. Crossover's main objective is to ensure that the genes are shared and that the children inherit the parent's genes. The crossover operator preference has a greater impact on GA's results. Through choosing suitable 'breeding' operators, the premature convergence can be avoided. Crossover operators may be categorized into crossover for binary problem representation, crossover for real code problem representation and crossover that apply to both types [58, 123].

The real crossover method applies in continuous variables. The crossover acts to generate at least one offspring based on two or more parent chromosomes in the population. The real crossover operation is applied in a variety of forms, as the elements are a real-valued vector [58, 124].

This general crossover method can be used in binary and real value problem representation. Although different crossover operators apply to different types of representation conversion between binary and real-value vectors is relatively easy [22, 125].

In this work, we have used the general crossover method. Among the various types of general crossover, we studied single point, two-point and uniform crossover. Single point crossover, detects one crossover point at random before splitting parents at this crossover point thereby producing offspring by exchanging tails. The common crossover probability is within the range of 0.5-1.0 [127].

N-point crossover is a generalization of the single point crossover. The n crossover points are randomly identified, after which they split along those points and convene, and then alternating between parents [127].

Uniform crossover assigns 'heads' to one parent, 'tails' to the other. A coin is flipped for each gene of the first child and an inverse copy of the gene is created for the second child [119]

Mutation adds randomness to solutions and should be scalable, drift-less, and reach each location in solution space. The genotype or chromosome of a solution is mapped to a phenotype, the real solution; before it can be evaluated on the fitness function. The role of mutation in GAs is that of restoring lost or unexplored genetic material into the population to prevent the premature convergence of GA to suboptimal. When the genetic algorithm is close to the optimal solution neighborhood through the crossover operator, the local random search ability using the mutation operator can accelerate the convergence to the optimal solution. Obviously, the probability of mutation in this case should be a small value; otherwise the building blocks close to the optimal solution will be destroyed by the mutation. The second is to enable genetic algorithms to maintain group diversity to prevent immature convergence [129].

In this work and for real-coded GAs, we have used among various types of mutation: Random or uniform, Non-uniform, standard Gaussian and Multi-non-uniform mutation. The uniform mutation is a common mutation operator based on Gaussian distribution whereby the user specifies a range of uniform random value to replace the value of the chosen gene [133]. Directed mutation or standard Gaussian mutation is based on extrapolation. The directed mutation finds a new point in the population using the information applied in the previous generations [130]. Non-uniform mutation possesses a fine-tuning capability whereby its action depends on the number of the population in order to reach equilibrium between exploration and exploitation. The search was performed uniformly at the beginning and very locally towards the end of the search. Multi-non-uniform mutation escalates the genetic diversity of a candidate individual. It can perform a uniform search with local fine-tuning thus increasing the exploitative capability [131, 129].

6.4. Genetic Algorithm Tuning

The success of the GA is dependent on the parameter values used, and identifying suitable values to use is a difficult task. It is necessary to make choices regarding how to evolve individuals. In addition, we want to determine which mutation and which crossover to choose, and what rates (frequencies of action) to associate with them? First, we must decide on the types of reproduction operators, as well as their rates. Basically, mutation can increase the diversity of chromosomes in GAs to exploit the solution space, while crossover, in order to help GAs to converge to optima, needs to identify, inherit and protect good common genes shared by chromosomes, and at the same time to recombine non-common genes [132].

Due to the stochastic nature of mutation and crossover, it is not an easy task to design efficient evolutionary operators free of the feasibility problem. The influence of the mutation and crossover operators' types was taken into account to determine which one among them produces the best model performance. When the defining lengths of schemata is low, one-point and two-point crossover are better than the uniform one. When the defining length is high, uniform crossover is the best [133, 134].

Uniform crossover is probably the most wildly used crossover operator because of its efficiency in not only identifying, inheriting and protecting common genes, but also re-combining non-common genes [95, 135].

For our task, uniform crossover and multi-non-uniform mutation operators were opted. Table 5 shows that the mutation operator can produce a significantly better solution by approaching the global optimum. Mutation rate (probability) determines how many chromosomes should be mutated in one generation when crossover rate (probability) represents the number of times a crossover occurs for chromosomes in one generation [136, 94].

Table 5. Comparison of nominal recognition rate according to types of crossover and mutation operators.

Nominal recognition rate according to crossover and mutation operators types (%) (p_c=0.8 and p_m=0.2)					
		Mutation Type			
		Standard Gaussian	**Uniform**	**Non-uniform**	**Multi-non-uniform**
Crossover type	**Single point**	86.65	89.05	88.78	91.75
	Uniform	89.15	92.74	93.64	93.95
	Two-point	88.67	91.74	92.80	92.97

A series of experiments were carried out in order to find the best values of the different parameters of the recognition GA. These values allows us to avoid premature convergence and improve the performance of our AG. Consequently, we set the crossover rate at 0.8 and the mutation one at 0.2 "Table 6".

Table 6. Comparison of recognition rate according to crossover and mutation probabilities.

Recognition rate according to crossover and mutation probabilities (%)						
		Crossover probability (Pc)				
		0.20	**0.40**	**0.75**	**0.80**	**0.85**
Mutation probability (Pm)	0.05	60.3	63.7	72.56	74.74	72.40
	0.10	62,20	69.85	91.75	92.85	89.79
	0.20	63.14	70.23	92.90	93.95	92.97
	0.40	60.08	63.45	77.12	86.18	85.25
	0.60	54.17	57.89	71.18	81.23	77.84
	0.80	0.31	0.34	0.45	0.55	0.53

Table 6 shows that the mutation rate denoted P_m and defined as the probability that any given feature will be mutated is a critical parameter in GAs. A too low mutation rate may lead to insufficient exploration, causing the algorithm to converge prematurely. Conversely, a high mutation rate may disrupt the inheritance of good traits, reducing the overall effectiveness of the GA.

The diversity measure is calculated as the ratio of the number of distinct individuals to the total number of individuals in the population. Thus, a diversity equal to 1 corresponds to the initialization of the algorithm (all individuals were randomly generated), while a diversity equal to zero corresponds to the convergence of the algorithm towards an optimal solution [137].

Table 7 shows that the low level of diversity in the population is obtained with a probability of mutation equals to 0.2 and the probability of crossover denoted Pc equals to 0.8. Furthermore, it shows that the optimal level of selective pressure is achieved with the optimal values of Pc and Pm equal to 0.8 and 0.2 respectively.

Table 7. Evolution of diversity rate according to crossover and mutation probabilities.

Evolution of diversity rate according to crossover and mutation probabilities (%)						
		Crossover Probability (Pc)				
		0.20	**0.40**	**0.60**	**0.80**	**0.90**
Mutation probability (Pm)	0.05	0.75	0.66	0.58	0.55	0.63
	0.10	0.45	0.35	0.33	0.20	0.23
	0.20	0.25	0.24	0.18	0.13	0.15
	0.30	0.48	0.45	0.35	0.31	0.35
	0.40	0.67	0.67	0.60	0.58	0.57
	0.50	0.73	0.72	0.62	0.59	0.60
	0.80	0.89	0.85	0.84	0.80	0.82

In addition, Table 7 shows that these values of probabilities, permitted to obtain the highest score of phoneme recognition. It shows the great influence of diversity in the performance of the recognition GA. Thus, controlling diversity is a key aspect of optimizing the search process in GAs [137].

When diversity measure is low the selective pressure is high, the search process tends to converge rapidly towards high-fitness solutions. On the other hand, when the selective pressure is low, the search process maintains a higher level of diversity in the population while this may slow down the convergence speed, it allows for a more extensive exploration of the search space [140].

Selective pressure refers to the degree to which the selection process in a GA favors fitter individuals over less fit ones. It determines the intensity of the competition among individuals to be selected for reproduction and survival in the next generation. The higher the selective pressure, the more the selection process favors the fittest individuals, while lower selective pressure allows for a more diverse selection of individuals. Selective pressure is essential in driving the search towards optimal solutions by promoting the survival and reproduction of high-quality individuals [59, 137].

Elitism is a powerful concept in GAs that control selective pressure and ensures the survival of the best individuals from one generation to the next. The motivation behind elitism is to prevent the loss of high-quality solutions during the selection process. By preserving the fittest individuals, elitism helps maintain the best genetic material and guides the search towards optimal solutions. Elitism can be incorporated into selection strategies in various ways [138, 139].

One common approach is to combine elitism with other selection methods, such as tournament selection or fitness-proportionate selection. In this case, the elite individuals are first selected and added to the next generation, and then the remaining population undergoes the chosen selection method [140, 141].

In the literature, there are several selection methods like Roulette Wheel Selection (RWS), stochastic universal sampling, the Tournament Selection (TS) and elitism.

Tournament selection is an important selection mechanism for GAs, where a set of solutions is selected randomly and within this competition subset, the best solutions are finally selected as new parents. Tournament selection offers a positive probability for each solution to survive, even if it has worse fitness values than other solutions [141]. It is simple to code, easy to implement, robust in the presence of noise, and has adjustable selection pressure. It is adaptable for any tournament size or noise level [59, 142].

Roulette wheel also known as fitness proportional selection selects parental solutions randomly with uniform distribution. The probability for being selected depends on the fitness of a solution. For this sake, the relative fitness of solutions normalized with the sum of all fitness values in a population, usually by division. This fraction of fitness can be understood as probability for a solution of being selected. The advantage of fitness-proportional selection operators is that each solution has a positive probability of being selected [141].

For this work, fitness proportional selection and Tournament one methods were studied.

Population sizing is one of the important topics to consider in evolutionary computation. The size of the population affects the convergence of a GA. In the literature, there are various studies and results about the appropriate GA population size [141, 142]. Larger populations enhance diversity but may slow convergence and reduce efficiency due to longer execution times. Smaller populations risk getting stuck in local minima. Typically, population sizes range from 75 to 175 individuals, depending on the problem and computational resources [142, 143].

Table 8 shows that a population of 100 individuals by generation permits to reach a high performance of 93.95 % for the best individual. It shows that the optimal population size is equal to 100.

Table 8. Evolution of the overall phoneme classification rate depending on the size of the population.

Population size (individuals)	**Overall rate of phoneme classification**
50	60.250 (%)
100	93.950 (%)
150	93.127 (%)
200	83.150 (%)
300	77.400 (%)

Comparing the results obtained, we can notice that the population performance increases with population size up to 100 and then decreases over generations "Table 7". Indeed, a population of 50 individuals achieves a performance of 60.25% for the best individual, while a population of 100 individuals produces the best one.

Therefore, implementing these optimal values and optimal population size ensures fast convergence of GA to the solutions, making it more efficient.

7. Results and Interpretations

The objective of this study is to demonstrate the effectiveness of the proposed MSA recognition model and its ability to classify the eight MSAs behind obstruent consonants in continuous speech, where a recognition GA has been developed during this work. We divided the vocal signal of the corpus to be processed into a set of acoustic segments of the same size to which we applied cepstral analysis representation method. Real vector GA preserves problem structure and reduces search space, ensuring new individuals respect problem constraints.

The behaviour of our AG was studied according to the type of reproduction and selection operators and to its parametric adjustment. To aid the calibration of GAs, it is proposed that there exists an optimal number of GA generations for a given problem, where from the number of generations, the population size can be determined from the total function evaluations that are available "Figure 1".

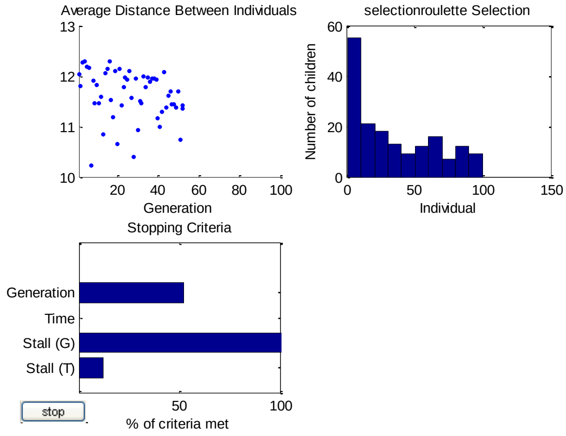

Figure 1. Evolution of the best individuals as a function of population size by generation.

Table 9 shows that the use of the genetic model allowed us to obtain, with an optimal population size equal to 100, the best values of GA evaluation function, after evolving only 100 generations, and then it stabilizes at this rate without any notable progression over the generations since the values of the GA fitness function are close to one (0.94). The result shows the rapid convergence of the GA model toward the solutions. For both large and small populations, the results show that the fitness values are low, which reduces the efficiency of GA in reaching global optima and negatively affects the ASR rate [141, 142]

Table 9. Evolution of GA evaluation value according to population size and generations.

Evolution of GA evaluation value according to population size and GA generations						
		GA Generations				
		50	**100**	**150**	**300**	**500**
Population size	25	0.32	0.54	0.55	0.54	0.52
	50	0.45	0.60	0.61	0.61	0.60
	100	0.63	0.93	0.94	0.94	0.93
	150	0.60	0.93	0.94	0.94	0.92
	300	0.55	0.90	0.91	0.91	0.91
	500	0.54	0.85	0.86	0.85	0.84
	1000	0.50	0.89	0.90	0.89	0.89

The GAs are stopped when the population converges toward the optimal solution. Unfortunately, evolution often looks like a never-ending process. GAs cannot be expected to stop spontaneously on the global optimum solution as if by magic. GAs are not even guaranteed to find the global optimum solution. So evolution has to be stopped at some point according to a pre-determined criterion. In fact there are many ways to decide to stop the algorithm. The simplest one is to stop evolution after a fixed number of iterations. A better solution consists of continuing the iteration for as long as any improvements are noticeable. In our case, The GA is stopped after 100 iterations because a notable improvement has occurred and almost all members of the population have become similar [141].

Regarding the mutation crossover rates or probabilities, Tables 5 and 6 show that when they are set in the range of 0.08 to 0.2 and 0.75 to 0.85 respectively, a high degree of diversity is maintained, which helps to avoid getting stuck in the same local optima and thus avoid reducing the system performance. However, the results obtained show that when the mutation rate is lower than 0.08 or higher than 0.2, our GA converges prematurely. The simulations are carried out on a population of 100 individuals subjected to a mutation whose probability is equal to 0.1. The results show that the best crossover rates that significantly influence the performance of individuals are between 75% and 85%.

It is deduced that the recognition rate is influenced by the respective values of the crossover and mutation probabilities. A crossover rate value of 0.8 allows the GA to converge toward a global optimum after a certain number of iterations.

Table 10 shows that the recognition rates of back obstruent obstruents of MSA differ from one sound to another depending on the articulatory mode and to a very small extent on the place of articulation of these speech sounds. This shows that the place of articulation does not influence recognition rates inversely to the articulatory mode.

Table 10. Recognition rates according to the type of MSA Back obstruent phonemes.

MSA back consonants (API)	**Classification rates of MSA phonemes (API) (%)**
[?]	94,88
[ε]	89,18
[γ]	90.45
[h]	89,95
[ħ]	89,10
[X]	90.85
[q]	97,25
[k]	97,75

The recognition rates of the three occlusive consonants the voiceless velar /k/, the voiceless uvular /q/ and the voiceless glottal /?/ are high. As can be seen, this reflects the good performance of GA used as a classifier due to the good adjustment of GA parameters during implementation.

The unrecognized phonemes are essentially due to the segmentation difficulties and to the coarticulation phenomenon problems. Back consonant that occur in the medium position are strongly influenced by the preceding and the following sounds [23, 144].

Table 11 shows that occlusive obstruent sounds are better recognized than fricative sounds. It can be explained by the fact that the fricative sounds are more difficult to be categorized and characterized acoustically and their phonetic features are less distinctive to guaranty their systematic identification [145, 128].

Table 11. Comparison of average recognition rates according to the type of MSA back obstruents.

Back obstruent types	**Overall rate of phonemes classification**
Occlusive	96,18 %
Constrictive	89.60 %

In comparison of these results with those of works using classical models such as statistical or neural ones, it appears that our genetic recognition model is also efficient, but if we compare them to the results of HMM-ANN, HMM-DNN, ANN/AG, DNN/AG hybrid classifiers proposed in literature, we note that our model is almost as efficient “Table1”.

8. Conclusions

In this work, a genetic model based on a dynamic and adaptable genetic evaluation function was developed to rigorously recognize the eight MSAs back obstruent consonants in continuous speech whatever their positions in syllables. The application of Genetic Algorithms for ASR, as a genetic model for classifying acoustic vectors of speech sequences, is very recent. In addition, we aimed at overcome the premature convergence problem that traditional GA suffers, due to the loss of genetic diversity in the candidate solution population.

For our task, an efficient parametric representation of the back obstruents based on the choice of a set of distinctive relevant and discriminative parameters, was conceived. It is worth noting the difficulties encountered in choosing the appropriate parameters for the different operators (mutation, crossover, selection, and replacement) and that the rejection rate is controlled by the choice of these types of parameters.

It is concluded that the results are too satisfactory since the overall recognition rate of MSA back obstruent consonants in the vocal corpus can reach the threshold of 93.95 % under the best conditions of adjustment of the used GA parameters. Comparing the obtained level of accuracy with the results of the other authors is not indicative, since systems use different dictionaries and different natural languages. Future work will be oriented towards speech recognition using hybrid models to increase recognition rates.

References

[1] Abdelhamid A., Alsayadi H., Hegazy I., and Fayed Z., "End-to-End Arabic Speech Recognition: A Review," *in Proceedings of the 19th Conference of Language Engineering*, Alexandria, pp. 26-30, 2020. https://www.researchgate.net/publication/344799361_End-to-End_Arabic_Speech_Recognition_A_Review

[2] Abdelmaksoud E., Hassen A., Hassan N., and Hesham M., "Convolutional Neural Network for Arabic Speech Recognition," *Egyptian Journal of Language Engineering*, vol. 8, no. 1, pp. 27-38, 2021. https://doi.org:10.21608/ejle.2020.47685.1015

[3] Ahmed A., Hifny Y., Shaalan K., and Toral S., *Computational Linguistics, Speech and Image Processing for Arabic Language*, World Scientific, 2019. https://doi.org:10.1142/9789813229396_0011

[4] Aissiou M. and Guerti M., "Genetic Supervised Classification of Standard Arabic Fricative Sounds," *International Journal of Speech Technology*, vol. 12, pp. 139-147, 2009. https://doi.org/10.1007/s10772-009-9061-5

[5] Aissiou M., "A Genetic Model for Acoustic and Phonetic Decoding of Standard Arabic Vowels in Continuous Speech," *International Journal of Speech Technology*, vol. 23, pp. 425-434, 2020. https://doi.org:10.1007/s10772-020-09694-y

[6] Ait-Mait H. and Aboutabit N., "HMM-GMM Acoustic Modeling for Arabic Speech Recognition System," *in Proceedings of the 3rd ICMDS'24: Machine Learning, Inverse Problems and Related Fields*, Khouribga, pp. 1-13, 2025. https://doi.org/10.1007/978-3-031-94802-2_1

[7] Al-Abdullah A., Al-Ajmi A., Al-Mutairi A., Al-Mousa N., and et al., "Artifcial Neural Network for Arabic Speech Recognition in Humanoid Robotic Systems," *in Proceedings of the 3rd International Conference on Bio-Engineering for Smart Technologies*, Paris, pp. 1-4, 2019. https://doi.org/10.1109/BIOSMART.2019.8734261

[8] Al-Anzi F. and AbuZeina D., "Literature Survey of Arabic Speech Recognition," *in Proceedings of the International Conference on Computing Sciences and Engineering*, Kuwait, pp. 1-6, 2018. DOI:10.1109/ICCSE1.2018.8374215

[9] Albadr M., Tiun S., Ayob M., and Al-Dhief F., "Genetic Algorithm Based on Natural Selection Theory for Optimization Problems," *Symmetry*, vol. 12, no. 11, pp. 1-31, 2020. https://doi.org/10.3390/sym12111758

[10] Alfeilat H., Hassanat A., Lasassmeh O., Tarawneh A., and et al., "Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review," *Big Data*, vol. 7, no. 4, pp. 1-28, 2019. https://doi.org/10.1089/big.2018.0175

[11] Al-Fraihat D., Sharrab Y., Alzyoud F., Qahmash A., and Maaita A., "Speech Recognition Utilizing Deep Learning: A Systematic Review of the Latest Developments," *Human-centric Computing and Information Sciences*, vol. 14, pp. 1-33, 2024. https://doi.org: 10.22967/HCIS.2024.14.015

[12] Alghamdi M., *Arabic Phonetics and Phonology*, Altawbah Library, 2015. https://www.researchgate.net/publication/280798546_Arabic_Phonetics_and_Phonology_in_Arabic

[13] Algihab W., Alawwad N., Aldawish A., and AlHumoud S., "Arabic Speech Recognition with Deep Learning: A Review," *in Proceedings of the Social Computing and Social Media. Design, Human Behavior and Analytics*, Orlando, pp. 15-31, 2019. DOI:10.1007/978-3-030-21902-4_2

[14] Alhijawi B. and Awajan A., "Genetic Algorithms: Theory, Genetic Operators, Solutions, and Applications," *Evolutionary Intelligence*, vol. 17, no. 3, pp. 1245-1256, 2024. https://doi.org/10.1007/s12065-023-00822-6

[15] Ali A., Der Spiegel J., and Mueller P., "Acoustic-Phonetic Features for the Automatic Classification of Stop Consonants," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 833-841, 2001. DOI:10.1109/89.966086

[16] Alim S. and Rashid N., *From Natural to Artificial Intelligence-Algorithms and Applications*, IntechOpen, 2018. http://doi:10.5772/intechopen.80419

[17] Alotaibi Y., Selouani S., Yakoub M., Seddiq Y., and Meftah A., "A Canonicalization of Distinctive Phonetic Features to Improve Arabic Speech Recognition," *Acta Acustica United with Acustica*,

vol. 105, no. 6, pp. 1269-1277, 2019. https://doi.org/10.3813/AAA.919404

[18] AlShourbaji I., Helian N., Sun Y., and Alhameed M., "A Novel HEOMGA Approach for Class Imbalance Problem in the Application of Customer Churn Prediction," *SN Computer Science*, vol. 2, no. 6, pp. 1-13, 2021. https://doi.org/10.1007/S42979-021-00850-Y

[19] Anyasi B., Babarinde O., and Iloene G., "Acoustic Analysis of Obstruents in Some Igbo Dialects," *Theory and Practice in Language Studies*, vol. 10, no. 12, pp. 1510-1527, 2020. http://dx.doi.org/10.17507/tpls.1012.02

[20] Bäck T., "Selective Pressure in Evolutionary Algorithms: A Characterization of Selection Mechanisms," *in Proceedings of the 1st IEEE Conference on Evolutionary Computation*, Orlando, pp. 57-62, 1994. https://doi/10.1109/ICEC.1994.350042

[21] Bala M., Bansal S., and Sharma C., "Improved Speech Recognition Using Vector Quantization, Genetic Algorithm and Neural Networks," *in Proceedings of the International Conference on Computer, Electronics and Electrical Engineering and their Applications*, Srinagar Garhwal, pp. 1-5, 2023. DOI:10.1109/IC2E357697.2023.10262606

[22] Ball M., Howard S., and Miller K., "Revisions to the extIPA Chart," *Journal of the International Phonetic Association*, vol. 48, no. 2, pp. 155-164, 2018. https://doi.org/10.1017/S0025100317000147

[23] Bani-Salameh M., "Phonemic Consonant Sounds in Modern Standard Arabic," *Linguistics and Culture Review*, vol. 5, no. S2, pp. 1643-1658, 2021. https://doi.org/10.21744/lingcure.v5nS2.2257

[24] Barman T. and Deb N., "State of the Art Review of Speech Recognition Using Genetic Algorithm," *in Proceedings of the IEEE International Conference on Power, Control, Signals and Instrumentation Engineering*, Chennai, pp. 2944-2946, 2017. https://doi:10.1109/ICPCSI.2017.8392264

[25] Bendahmane A., Acoustic Study of Standard Arabic fricatives (Algerian Speakers), Master Thesis, *École Doctorale Des Humanités*, University of Strasbourg, 2013.

[26] Benmachiche A. and Makhlouf A., "Optimization of Hidden Markov Model with Gaussian Mixture Densities for Arabic Speech Recognition," *WSEAS Transactions on Signal Processing*, vol. 15, pp. 85-94, 2019. https://wseas.com/journals/sp/2019/a225114-677.pdf

[27] Besdouri F., Zribi I., and Belguith L., "Arabic Automatic Speech Recognition: Challenges and Progress," *Speech Communication*, vol. 163, no. 1, pp. 103110, 2024. https://doi.org/10.1016/j.specom.2024.103110

[28] Bhatia S., Devi A., Alsuwailem R., and Mashat A., "Convolutional Neural Network Based Real Time Arabic Speech Recognition to Arabic Braille for Hearing and Visually Impaired," *Frontiers in Public Health*, vol. 10, pp. 1-10, 2022. https://doi.org/10.3389/fpubh.2022.898355

[29] Bhatt S., Dev A., and Jain A., "Hindi Speech Vowel Recognition Using Hidden Markov Model," *in Proceedings of the 6th Workshop on Spoken Language Technologies for Under Resourced Languages*, Gurugram, pp. 201-204, 2018. https://doi.org/10.21437/SLTU.2018-42

[30] Boriah S., Chandola V., and Kumar V., "Similarity Measures for Categorical Data: A Comparative Evaluation," *in Proceedings of the SIAM International Conference on Data Mining*, Atlanta, pp. 243-254, 2008. https://doi.org/10.1137/1.9781611972788.22

[31] Bouchakour L. and Debyeche M., "Improving Continuous Arabic Speech Recognition over Mobile Networks DSR and NSR Using MFCCs Features Transformed," *International Journal of Circuits, Systems and Signal Processing*, vol. 12, pp. 379-386, 2018. https://www.naun.org/main/NAUN/circuitssystemssignal/2018/b082005-afe.pdf

[32] Brest J., Maucec M., and Boskovic B., "Single Objective Real-Parameter Optimization: Algorithm jSO," *IEEE Congress on Evolutionary Computation Conference*, Donostia, pp. 1316-1318, 2017. https://doi.org/10.1109/CEC.2017.7969456

[33] Brodbeck C. and Simon J., "Continuous Speech Processing," *Current Opinion in Physiology*, vol. 18, pp. 25-31, 2020. https://doi.org/10.1016/j.cophys.2020.07.014

[34] Cantineau J., *Cours de Phontique Arabe*, Librairie C. Klincksieck, 1960. https://archive.org/details/Cantineau-CoursDePhonetiqueArabe1960

[35] Chang L., Ouzrout Y., Nongaillard A., and Bouras A., "Optimized Hidden Markov Model Based on Constrained Particle Swarm Optimization," *arXiv Preprint*, vol. arXiv:1811.03450v1, pp. 1-5, 2018. https://doi.org/10.48550/arXiv.1811.03450

[36] Chauhan N., Isshiki T., and Li D., "Speaker Recognition Using LPC, MFCC, ZCR, Features with ANN and SVM Classifier for Large Input Database," *in Proceedings of the 4th International Conference on Computer and Communication Systems*, Singapore, pp. 130-133, 2019. https://doi.org/10.1109/CCOMS.2019.8821751

[37] Chen S., Cao J., Chen F., and Liu B., "Entropy-Based Fuzzy Least Squares Twin Support Vector Machine for Pattern Classification," *Neural*

Processing Letters, vol. 51, pp. 41-66, 2020. http://dx.doi.org/10.1007/s11063-019-10078-w

[38] Chouhan k., Singh A., Shrivastava A., Agrawal S., and et al., "Structural Support Vector Machine for Speech Recognition Classification with CNN Approach," *in Proceedings of the 9th International Conference on Cyber and IT Service Management*, Bengkulu, pp. 1-7, 2021. https://doi.org/10.1109/CITSM52892.2021.9588918

[39] Crepinsek M., Liu S., and Mernik M., "Exploration and Exploitation in Evolutionary Algorithms: A Survey," *ACM Computing Surveys*, vol. 45, no. 3, pp. 1-33, 2013. https://doi.org/10.1145/2480741.2480752

[40] Dabbabi K. and Mars A., "Spoken Utterance Classification Task of Arabic Numerals and Selected Isolated Words," *Arabian Journal for Science and Engineering*, vol. 47, pp. 10731-10750, 2022. https://doi.org/10.1007/s13369-022-06649-0

[41] Danaa A., Daabo M., and Abdul-Barik A., "An Improved Hybrid Algorithm for Optimizing the Parameters of Hidden Markov Models," *Asian Journal of Research in Computer Science*, vol. 10 no. 1, pp. 63-73, 2021. https://doi.org/10.9734/ajrcos/2021/v10i130235

[42] Danaa A., Nawusu Y., Mashud A., and Diyawu M., "Hidden Markov Model and Deep Neural Network Hybrid Model for Enhanced Speech Recognition," *Journal of Mathematical Sciences and Computational Mathematics*, vol. 5, no. 4, pp. 405-416, 2024. DOI:10.15864/jmscm.5403

[43] Deshmukh A., "Comparison of Hidden Markov Model and Recurrent Neural Network in Automatic Speech Recognition," *European Journal of Engineering and Technology Research*, vol. 5, no. 8, pp. 958-965, 2020. https://doi.org:10.24018/ejeng.2020.5.8.2077

[44] Donkers T., Loepp B., and Ziegler J., "Sequential User-based Recurrent Neural Network Recommendations," *in Proceedings of the Eleventh ACM Conference on Recommender Systems*, New York, pp. 152-160, 2017. https://doi.org/10.1145/3109859.3109877

[45] Droua-Hamdani G., Selouani S., and Boudraa M., "Algerian Arabic Speech Database (ALGASD): Corpus Design and Automatic Speech Recognition Application," *Arabian Journal for Science and Engineering*, vol. 35, no. 2, pp. 157-166, 2010.

[46] Dua M., Akanksha., and Dua S., "Noise Robust Automatic Speech Recognition: Review and Analysis," *International Journal of Speech Technology*, vol. 26, pp. 475-519, 2023. https://doi.org/10.1007/s10772-023-10033-0

[47] El Hindi K., "Specific-Class Distance Measures for Nominal Attributes," *AI Communications*, vol. 26, no. 3, pp. 261-279, 2013. http://doi:10.3233/AIC-130565

[48] Elfahm Y., Abajaddi N., Mounir B., Elmaazouzi L., and et al., "Classification of Arabic Fricative Consonants According to Their Places of Articulation," *International Journal of Electrical and Computer Engineering*, vol. 12, no .1, pp. 936-945, 2022. http://doi.org/10.11591/ijece.v12i1.pp936-945

[49] Elharati H., Alshaari E., and Këpuska V., "Arabic Speech Recognition System Based on MFCC and HMMs," *Journal of Computer and Communications*, vol. 8, no. 3, pp. 28-34, 2020. https://doi.org/10.4236/jcc.2020.83003

[50] Eray O., Sezai T., and Serdar I., "An Application of Speech Recognition with Support Vector Machines," *in Proceedings of the 6th International Symposium on Digital Forensic and Security*, Antalya, pp. 1-6, 2018. https://doi.org/10.1109/ISDFS.2018.8355321

[51] Ettaouil M., Lazaar M., and En-Naimani Z., "A Hybrid ANN/HMM Models for Arabic Speech Recognition Using Optimal Codebook," *in Proceedings of the 8th International Conference on Intelligent Systems: Theories and Applications*, Rabat, pp. 1-5, 2013. https://doi.org/10.1109/SITA.2013.6560806

[52] Faizal E. and Hamdani H., "Weighted Minkowski Similarity Method with CBR for Diagnosing Cardiovascular Disease," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 12, pp. 305-310, 2018. https://dx.doi.org/10.14569/IJACSA.2018.091244

[53] Falkenauer E., "The Worth of the Uniform [Uniform Crossover]," *in Proceedings of the Congress on Evolutionary Computation-CEC99*, Washington (DC), pp. 776-782, 1999. DOI:10.1109/CEC.1999.782011

[54] Gaffar A., Malani R., Supriadi., Wajiansyah A., and Putra A., "A Multi-Frame Blocking for Signal Segmentation in Voice Command Recognition," *in Proceedings of the International Seminar on Intelligent Technology and its Application,* Surabaya, pp. 299-304, 2020. DOI:10.1109/ISITIA49792.2020.9163761

[55] Gales M. and Young S., "The Application of Hidden Markov Models in Speech Recognition," *Foundations and Trends® in Signal Processing*, vol. 1, no. 3, pp. 195-304, 2008. http://dx.doi.org/10.1561/2000000004

[56] Ghazeli S., Back Consonants and Backing Coarticulation in Arabic, Ph.D. Dissertation, University of Texas at Austin, 1977. https://www.scirp.org/reference/referencespapers?referenceid=991987

[57] Goldberg D., "Real-coded GAs, Virtual Alphabets, and Blocking," *Complex Systems*, vol. 5, no. 2 pp.

139-167, 1991. https://content.wolfram.com/sites/13/2018/02/05-2-2.pdf

[58] Goldberg D., Deb K., and Clark J., "Genetic Algorithms, Noise, and the Sizing of Populations," *Complex Systems*, vol. 6, no. 4, pp. 333-362, 1992. https://content.wolfram.com/sites/13/2018/02/06-4-3.pdf

[59] Goldberg D., *Genetic Algorithms for Search, Optimization, and Machine Learning*, Addison-Wesley, 1989. https://www.amazon.com/Genetic-Algorithms-Optimization-Machine-Learning/dp/0201157675

[60] Guerid A. and Houacine A., "Recognition of Isolated Digits Using DNN-HMM and Harmonic Noise Model," *IET Signal Processing*, vol. 13, no. 2, pp. 207-214, 2019. https://doi.org/10.1049/iet-spr.2018.5131

[61] Haboussi S., Oukas N., Zerrouki T., and Djettou H., "Arabic Speech Recognition Using Neural Networks: Concepts, Literature Review and Challenges," *Journal of Umm Al-Qura University for Applied Sciences*, vol. 11, pp. 1-23, 2025. https://doi.org:10.1007/s43994-025-00213-w

[62] Hamid O., "Frame Blocking and Windowing Speech Signal," Journal of Information, Communication, and Intelligence Systems, vol. 4, no. 5, pp. 87-94, 2018. https://www.researchgate.net/publication/331635757_Frame_Blocking_and_Windowing_Speech_Signal

[63] Hassanat A., Almohammadi K., Alkafaween E., Abunawas E., and et al., "Choosing Mutation and Crossover Ratios for Genetic Algorithms-A Review with a New Dynamic Approach," *Information*, vol. 10, no. 12, pp. 1-36, 2019. http://doi.org :10.3390/info10120390

[64] Hayes B., Shier J., Fazekas G., McPherson A., and Saitis C., "A Review of Differentiable Digital Signal Processing for Music and Speech Synthesis," *Audio and Acoustic Signal Processing*, vol. 3, pp. 1-194, 2023. https://doi.org/10.3389/frsip.2023.1284100

[65] Herrera F., Lozano M., and Sánchez A., "A Taxonomy for the Crossover Operator for Real-Coded Genetic Algorithms: An Experimental Study," *International Journal of Intelligent Systems*, vol. 18, no. 3, pp. 309-338, 2003. https://doi.org/10.1002/int.10091

[66] Holland J., *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, MIT Press, 1992. https://doi.org/10.7551/mitpress/1090.001.0001

[67] Hussain A. and Muhammad Y., "Trade-off Between Exploration and Exploitation with Genetic Algorithm Using a Novel Selection Operator," *Complex and Intelligent Systems*, vol. 6, no. 1, pp. 1-14, 2020. https://doi.org/10.1007/s40747-019-0102-7

[68] Ibrahim A., Seddiq Y., Meftah A., Alghamdi M., and et al. "Optimizing Arabic Speech Distinctive Phonetic Features and Phoneme Recognition Using Genetic Algorithm," *IEEE Access*, vol. 8, pp. 200395-200411, 2020. https://doi.org/10.1109/ACCESS.2020.3034762

[69] Ibrahim Y., Odiketa J., and Ibiyemi T., "Preprocessing Technique in Automatic Speech Recognition for Human Computer Interaction: An Overview," *International Conference on Decision Aid Sciences and Applications*, vol. 15, no. 1, pp. 186-191, 2017. DOI:10.1109/DASA54658.2022.9765043

[70] Jalili F. and Barani M., "Speech Recognition Using Combined Fuzzy and Ant Colony Algorithm," *International Journal of Electrical and Computer Engineering*, vol. 6, no. 5, pp. 2205-2210, 2016. http://doi.org/10.11591/ijece.v6i5.pp2205-2210

[71] Jebari K. and Madiafi M., "Selection Methods for Genetic Algorithms," *International Journal of Emerging Sciences*, vol. 3, no. 4, pp. 333-344, 2013. https://www.researchgate.net/publication/259461147_Selection_Methods_for_Genetic_Algorithms

[72] Kadyan V., Mantri A., Aggarwal R., and Singh A., "A Comparative Study of Deep Neural Network Based Punjabi-ASR System," *International Journal of Speech Technology*, vol. 22, no. 1, pp. 111-119, 2019. https://doi.org:10.1007/s10772-018-09577-3

[73] Kaminski K. and Dobrowolski A., "Automatic Speaker Recognition System Based on Gaussian Mixture Models, Cepstral Analysis, and Genetic Selection of Distinctive Features," *Sensors*, vol. 22, no. 23, pp. 1-25, 2022. https://www.mdpi.com/1424-8220/22/23/9370

[74] Kanke R., Gaikwad R., and Baheti M., "Enhanced Marathi Speech Recognition Using Double Delta MFCC and DTW," *International Journal of Digital Technologies*, vol. 2, no. 1, pp. 49-58, 2023. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=pLqwjO8AAAAJ&citation_for_view=pLqwjO8AAAAJ:9yKSN-GCB0IC

[75] Kate R., "Using Dynamic Time Warping Distances as Features for Improved Time Series Classification," *Data Mining and Knowledge Discovery*, vol. 30, pp. 283-312, 2016. https://doi.org/10.1007/s10618-015-0418-x

[76] Katoch S., Chauhan S., and Kumar V., "A Review on Genetic Algorithm: Past, Present, and Future," *Multimedia Tools and Applications*, vol. 80, pp. 8091-8126, 2021. https://doi.org/10.1007/s11042-020-10139-6

[77] Kaur A., Singh A., Sachdeva R., and Kukreja D., "Automatic Speech Recognition Systems: A Survey of Discriminative Techniques," *Multimedia Tools and Applications*, vol. 82, pp. 13307-13339, 2023. https://doi.org/10.1007/s11042-022-13645-x

[78] Kaur G., Srivastava M., and Kumar A., "Genetic Algorithm for Combined Speaker and Speech Recognition Using Deep Neural Networks," *Journal of Telecommunications and Information Technology*, vol. 2, no. 2, pp. 23-31, 2018. https://doi.org/10.26636/jtit.2018.119617

[79] Këpuska V. and Elharati H., "Robust Speech Recognition System Using Conventional and Hybrid Features of MFFCC, LPCC, PLP, RASTA-PLP and Hidden Markov Model Classifier in Noisy Conditions," *Journal of Computer and Communications*, vol. 3, no. 6, pp. 1-9, 2015. https://doi.org/10.4236/jcc.2015.36001

[80] Kheddar H., Hemis M., and Himeur Y., "Automatic Speech Recognition Using Advanced Deep Learning Approaches: A Survey," *Information Fusion*, vol. 109, pp. 1-22, 2024. https://doi.org/10.1016/j.inffus.2024.102422

[81] Khelifa M., Elhadj Y., Abdellah Y., and Belkasmi M., "Constructing Accurate and Robust HMM/GMM Models for an Arabic Speech Recognition System," *International Journal of Speech Technology*, vol. 20, no. 4, pp. 937-949, 2017. https://doi.org/10.1007/s10772-017-9456-7

[82] Khwaileh F., Flipsen P., Hammouri H., and Alzoubi F., "Acoustic Characteristics of Arabic Pharyngealized Obstruents in Children with Cochlear Implants," *Journal of the Acoustical Society of America*, vol. 146, no. 2, pp. 893-908, 2019. https://doi.org/10.1121/1.5119355

[83] Kim J., Han M., and Shin S., "Development of a Mutation Operator in a Real-Coded Genetic Algorithm for Bridge Model Optimization," *KSCE Journal of Civil Engineering*, vol. 28, no. 5, pp. 1822-1835, 2024. https://doi.org/10.1007/s12205-024-2480-7

[84] Kramer O., *Genetic Algorithm Essentials*, Springer, 2017. https://link.springer.com/book/10.1007/978-3-319-52156-5

[85] Kubicki M. and Figurowski D., "An Introduction to a Novel Crossover Operator for Real-Value Encoded Genetic Algorithm: Gaussian Crossover Operator," *in Proceedings of the International Interdisciplinary PhD Workshop*, Swinoujscie, pp. 85-90, 2018. doi:10.1109/IIPHDW.2018.8388331

[86] Kumar V., Chhabra J., and Kumar D., "Parameter Adaptive Harmony Search Algorithm for Unimodal and Multimodal Optimization Problems," *Journal of Computational Science*, vol. 5, no. 2, pp. 144-155, 2014. https://doi:10.1016/j.jocs.2013.12.001

[87] Latif S., Zaidi A., Cuayahuitl H., Shamshad F., and et al., "Transformers in Speech Processing: A Survey," *arXiv Preprint*, vol. arXiv:2303.11607v2, pp. 1-27, 2023. https://doi.org/10.48550/arXiv.2303.11607

[88] Lekshmi k. and Sherly E., "Automatic Speech Recognition Using Different Neural Network Architectures-A Survey," *International Journal of Computer Science and Information Technologies*, vol. 7, no. 6, pp. 2422-2427, 2016.

[89] Lerato L. and Niesler D., "Feature Trajectory Dynamic Time Warping for Clustering of Speech Segments," *Journal on Audio, Speech, and Music Processing*, vol. 2019, no. 6, pp. 1-9, 2019. https://doi.org/10.1186/s13636-019-0149-9

[90] Li C. and Li H., "A Survey of Distance Metrics for Nominal Attributes," *Journal of Software*, vol. 5, no. 11, pp. 1262-1269, 2010. DOI:10.4304/jsw.5.11.1262-1269

[91] Liao Y., Hsu H., Juang Y., and Yu L., "On the Investigation of Population Sizing of Genetic Algorithms Using Optimal Mixing," *in Proceedings of the Genetic and Evolutionary Computation Conference*, Prague, pp. 820-828, 2019. https://doi.org/10.1145/3321707.3321798

[92] Lim S., Sultan A., Sulaiman M., Mustapha A., and Leong K., "Crossover and Mutation Operators of Genetic Algorithms," *International Journal of Machine Learning and Computing*, vol. 7, no. 1, pp. 9-12, 2017. https://doi:10.18178/ijmlc.2017.7.1.611

[93] Lim S., Sultan A., Sulaiman M., Mustapha A., and Leong K., "Crossover and Mutation Operators of Genetic Algorithms," *International Journal of Machine Learning and Computing*, vol. 7, no. 1, pp. 9-12, 2017. https://doi :10.18178/ijmlc.2017.7.1.611

[94] Lin W., Lee W., and Hong T., "Adapting Crossover and Mutation Rates in Genetic Algorithms," *Journal of Information Science and Engineering*, vol. 19, no. 5, pp. 889-903, 2003. https://api.semanticscholar.org/CorpusID:15262029

[95] Little C., Choudhury S., Hu T., and Salomaa K., "Comparison of Genetic Operators for the Multiobjective Pickup and Delivery Problem," *Mathematics*, vol. 10, no. 22, pp. 1-21, 2022. https://doi.org/10.3390/math10224308

[96] Liu D., "Mathematical Modeling Analysis of Genetic Algorithms under Schema Theorem," *Journal of Computational Methods in Sciences and Engineering*, vol. 19, no. 2, pp. 1-7, 2019. https://doi:10.3233/JCM-191019

[97] Lyu S., Tian X., Li Y., Jiang B., and Chen H., "Multiclass Probabilistic Classification Vector Machine," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 10,

pp. 3906-3919, 2020. doi:10.1109/TNNLS.2019.2947309

[98] Mahin M., Islam M., Khatun A., and Debnath B., “A Comparative Study of Distance Metric Learning to Find Sub-Categories of Minority Class from Imbalance Data,” *in Proceeding of the International Conference on Innovation in Engineering and Technology*, Dhaka, pp. 1-6, 2018. DOI: 10.1109/CIET.2018.8660777

[99] Malik A., “A Study of Genetic Algorithm and Crossover Techniques,” *International Journal of Computer Science and Mobile Computing*, vol. 8, no. 3, pp. 335-344, 2019. https://www.ijcsmc.com/docs/papers/March2019/V8I3201947.pdf

[100] Manamela P., Manamela M., Modipa T., Sefara T., and Mokgonyane T., “The Automatic Recognition of Sepedi Speech Emotions Based on Machine Learning Algorithms,” *International Conference on Advances in Big Data, Computing and Data Communication Systems*, Durban, pp. 1-7, 2018. https://doi.org/10.1109/ICABCD.2018.8465403

[101] Manzoni L., Mariot L., and Tuba E., “Balanced Crossover Operators in Genetic Algorithms,” *Swarm and Evolutionary Computation*, vol. 54, pp. 1-26, 2020. https://doi.org/10.1016/j.swevo.2020.100646

[102] Mao S., Tao D., Zhang G., Ching P., and Lee T., “Revisiting Hidden Markov Models for Speech Emotion Recognition,” *in Proceedings of the IEEE international Conference on Acoustics, Speech and Signal Processing*, Brighton, pp. 6715-6719, 2019. https://doi.org:10.1109/ICASSP.2019.8683172

[103] Michalewicz Z., *Genetic Algorithms+Data Structures=Evolution Programs*, Springer, 1992. https://doi.org/10.1007/978-3-662-02830-8

[104] Mirjalili S., *Evolutionary Algorithms and Neural Networks: Theory and Applications*, Springer, 2019. https://link.springer.com/book/10.1007/978-3-319-93025-1

[105] Mooi S., Lim S., Sultan M, Bakar A., and et al., “Crossover and Mutation Operators of Genetic Algorithms,” *International Journal of Machine Learning and Computing*, vol. 7, no. 1, pp. 9-12, 2017. https://doi:10.18178/ijmlc.2017.7.1.611

[106] Mustafa M., Allen T., and Appiah K., “A Comparative Review of Dynamic Neural Networks and Hidden Markov Model Methods for Mobile On-Device Speech Recognition,” *Neural Computing and applications,* vol. 31, pp. 891-899, 2019. https://doi.org/10.1007/s00521-017-3028-2

[107] Muxamediyeva D., Nilufar N., and Oxanovna I., “Application of Genetic Algorithm in Training Automatic Speech Recognition,” *Journal of Advanced Research Design*, vol. 127, no. 1, pp. 1-15, 2025. https://doi.org/10.1016/j.procs.2021.01.060

[108] Nahar K., Shquier M., Wasfi G., Al-Khatib M., and et al., “Arabic Phonemes Recognition Using Hybrid LVQ/HMM Model for Continuous Speech Recognition,” *International Journal of Speech Technology*, vol. 19, no. 3, pp. 495-508, 2016. https://doi.org/10.1007/s10772-016-9337-5

[109] Naqvi F. and Shad M., “Seeking a Balance between Population Diversity and Premature Convergence for Real-Coded Genetic Algorithms with Crossover Operator,” *Evolutionary Intelligence*, vol. 15, no. 4, pp. 2651-2666, 2021. https://doi /10.1007/s12065-021-00636-4

[110] Nassif A., Shahin I., Attili I., Azzeh M., and Shaalan K., “Speech Recognition Using Deep Neural Networks: A Systematic Review,” *IEEE Access*, vol. 7, pp. 19143-19165, 2019. https://doi.org/10.1109/ACCESS.2019.2896880

[111] Novoa J., Wuth J., Escudero J., Fredes J., and et al., “DNN-HMM Based Automatic Speech Recognition for HRI Scenarios,” *in Proceedings of the 13th ACM/IEEE International Conference on Human-Robot Interaction*, Chicago, pp. 150-159, 2018. https://doi.org/10.1145/3171221.3171280

[112] Ouisaadane A. and Safi S., “A Comparative Study for Arabic Speech Recognition System in Noisy Environments,” *International Journal of Speech Technology*, vol. 24, no. 3, pp. 761-770, 2021. https://doi.org/10.1007/s10772-021-09847-7

[113] Papazoglou G. and Biskas P., “Review and Comparison of Genetic Algorithm and Particle Swarm Optimization in the Optimal Power Flow Problem,” *Energies*, vol. 16, no. 3, pp. 1-25, 2023. https://doi.org/10.3390/en16031152

[114] Peng S., Wang W., Chen Y., Zhong X., and Hu Q., “Regression-based Hyper Parameter Learning for Support Vector Machines,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 12, pp. 18799-18813, 2024. DOI:10.1109/TNNLS.2023.3321685

[115] Permanasari Y., Harahap E., and Ali E., “Speech Recognition Using Dynamic Time Warping (DTW),” *in Proceedings of the 2nd International Conference on Applied and Industrial Mathematics and Statistics*, Kuantan, pp. 1-6, 2019. https://doi.org/10.1088/1742-6596/1366/1/012091

[116] Prabakaran D. and Sriuppili S., “Speech Processing: MFCC Based Feature Extraction Techniques- An Investigation,” *in Proceedings of the Recent Advances in Smart System Automation, Computing and Communication*, Villupuram, pp. 1-9, 2021. https://doi:10.1088/1742-6596/1717/1/012009

[117] Rahman A., Kabir M., Mridha M., Alatiyyah M., and et al., “Arabic Speech Recognition: Advancement and Challenges,” *IEEE Access*, vol.

12, pp. 39689-39716, 2024. https://doi:10.1109/ACCESS.2024.3376237

[118] Rajakumar B. and George A., "APOGA: An Adaptive Population Pool Size based Genetic Algorithm," *AASRI Procedia*, vol. 4, pp. 288-296, 2013. https://doi.org/10.1016/j.aasri.2013.10.043

[119] Ramya R., Preethi M., and Rajalakshmi R., "Genetic Algorithm-Based Optimization for Speech Processing Applications," *Optimization Techniques in Engineering, Advances and Applications*, pp. 215-230, 2023. https://doi.org/10.1002/9781119906391.ch13

[120] Roger V., Farinas J., and Pinquier J., "Deep Neural Networks for Automatic Speech Processing: A Survey from Large Corpora to Limited Data," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2022, no. 19, pp. 1-15, 2022. https://doi.org/10.1186/s13636-022-00251-w

[121] Rossi M., Benatti S., Farella E., and Benini L., "Hybrid EMG Classifier Based on HMM and SVM for Hand Gesture Recognition in Prosthetics," *in Proceedings of the IEEE International Conference on Industrial Technology*, Sevilla, pp.1700-1705, 2015. https://doi.org/10.1109/ICIT.2015.7125342

[122] Salhi M. and Alenezi A., "Voice Recognition Enhancement by Genetic Algorithm," *Heliyon*, vol. 10, no. 23, pp. 1-13, 2024. https://doi.org/10.1016/j.heliyon.2024.e38845

[123] Sastry K., Goldberg D., and Kendall G., *Search Methodologies, Introductory Tutorials in Optimization and Decision Support Techniques*, Springer New York, 2014. https://link.springer.com/book/10.1007/978-1-4614-6940-7

[124] Shancheng J., Kwai-Sang C., Long W., Gang Q., and Tsui K., "Modified Genetic Algorithm-Based Feature Selection Combined with Pre-Trained Deep Neural Network for Demand Forecasting in Outpatient Department," *Expert Systems with Applications*, vol. 82, pp. 216-230, 2017. https://doi.org/10.1016/j.eswa.2017.04.017

[125] Sheikhan M., "Synthesizing Supra Segmental Speech Information Using Hybrid of GA-ACO and Dynamic Neural Network," *in Proceedings of the 5th Conference on Information and Knowledge Technology*, Shiraz, pp. 175-180, 2016. https://doi: 10.1109/IKT.2013.6620060

[126] Shosted R., Fu M., and Hermes Z., "Arabic Pharyngeal and Emphatic Consonants," *The Routledge Handbook of Arabic Linguistics*, pp. 48-61, 2018. https://doi.org :10.4324/9781315147062-4

[127] Singh C., Venter M., Muthu R., and Brown D., *Intelligent Speech Signal Processing*, Academic Press, 2019. https://doi.org/10.1016/B978-0-12-818130-0.00003-9

[128] Someya H., "Striking a Mean- and Parent-Centric Balance in Real-Valued Crossover Operators," *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 6, pp. 737-754, 2013. DOI:10.1109/TEVC.2012.2200255

[129] Sonkamble B. and Doye D., "Hidden Markov Model for Speech Recognition Using Modified Forward-Backward Re-Estimation Algorithm," *International Journal of Computer Science Issues*, vol. 9, no. 2, pp. 242-247, 2012. https://www.ijcsi.org/papers/IJCSI-9-4-2-242-247.pdf

[130] Sood M. and Jain S., "Speech Recognition Employing MFCC and Dynamic Time Warping Algorithm," *in Proceedings of the Innovations in Information and Communication Technologies Conference*, Erbil, pp. 235-242, 2021. https://doi.org/10.1007/978-3-030-66218-9_27

[131] Sosiawan A., Nooraeni R., and Sari L., "Implementation of Using HMM-GA in Time Series Data," *Procedia Computer Science*, vol. 179, pp. 713-720, 2021. https://doi.org/10.1016/j.procs.2021.01.060

[132] Taha Z, Abdullah A., and Rashid T., "Optimizing Feature Selection with Genetic Algorithms: A Review of Methods and Applications," *arXiv Preprint*, vol. arXiv:2409.14563v1, pp. 1-40, 2024. https://doi.org/10.48550/arXiv.2409.14563

[133] Tang P. and Tseng M., "Adaptive Directed Mutation for Real-Coded Genetic Algorithms," *Applied Soft Computing*, vol. 13, no. 1, pp. 600-614, 2013. https://doi.org/10.1016/j.asoc.2012.08.035

[134] Tebbal I. and Hamida A., "Effects of Crossover Operators on Genetic Algorithms for the Extraction of Solar Cell Parameters from Noisy Data," *Engineering, Technology and Applied Science Research*, vol. 13, no. 3, pp. 10630-10637, 2023. https://doi.org/10.48084/etasr.5417

[135] Temby L., Vamplew P., and Berry A., "Accelerating Real-Valued Genetic Algorithms Using Mutation-with-Momentum," *in Proceedings of the 18th Australian Joint Conference on Artificial Intelligence*, Sydney, pp. 1108-1111, 2005. https://doi.org/10.1007/11589990_149

[136] Wahyuni E., "Arabic Speech Recognition Using MFCC Feature Extraction and ANN Classification," *in Proceedings of the 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering*, Yogyakarta, pp. 22-25, 2017. https://doi.org/10.1109/ICITISEE.2017.8285499

[137] Xu N., Wang C., and Bao J., "Voice Activity Detection Using Entropy-Based Method," *in Proceedings of the 9th International Conference on Signal Processing and Communication Systems*,

Cairns, pp. 1-4, 2015. DOI:10.1109/ICSPCS.2015.7391751

[138] Xue Y., "Speaker Recognition System Using Dynamic Time Warping Matching and Mel-Scale Frequency Cepstral Coefficients," *in Proceedings of the 9th International Conference on Communications, Signal Processing, and Systems*, Changbaishan, pp. 961-967, 2021. https://doi.org/10.1007/978-981-15-8411-4_127

[139] Yang M., Yang Y., and Chen Y., "Evaluation of Fitness Functions of GA Classification," *in Proceedings of the Companion Publication of the Annual Conference on Genetic and Evolutionary Computation*, Vancouver, pp. 1479-1480, 2014. https://doi.org/10.1145/2598394.2602267

[140] Yang M., Yang Y., Su T., and Huang K., "An Efficient Fitness Function in Genetic Algorithm Classifier for Landuse Recognition on Satellite Images," *The Scientific World Journal*, vol. 2014, pp. 1-12, 2014. https://doi:10.1155/2014/264512

[141] Yue L., Hu P., Chu S., and Pan J., "Genetic Algorithm for High-Dimensional Emotion Recognition from Speech Signals," *Electronics*, vol. 12, no. 23, pp. 1-14, 2023. https://doi.org/10.3390/electronics12234779

[142] Zaidi B., Selouani S., Boudraa M., and Yakoub M., "Deep Neural Network Architectures for Dysarthric Speech Analysis and Recognition," *Neural Computing and Applications*, vol. 33, no. 15, pp. 9089-9108, 2021. https://doi.org/10.1007/s00521-020-05672-2

[143] Zarrouk E. and Benayed Y., "Hybrid SVM/HMM Model for the Arab Phonemes Recognition," *The International Arab Journal of Information Technology*, vol. 13, no. 5, pp. 574-582, 2016.

[144] Zitouni A., Falek L., Amrouche A., Dahou B., and Abbas M., "Design and Construction of 14 Arabic Fricatives Dataset, Classification and Characterization Using CRNN, Tansformers, and H-CRNN," *Multimedia Tools and Applications,* vol. 83, pp. 77187-77217, 2024. https://doi.org/10.1007/s11042-024-18355-0

Aissiou Mohamed received PhD in the field of electronic science and technology from the national higher polytechnic school of Algiers, Algiers, Algeria. His research interests included electronics, automatic speech processing, human-machine interaction, pattern recognition and evolutionary algorithms. Currently, he is working as a Senior Lecturer of electronics at National Higher School of advanced Technologies, Algiers, Algeria. He has direct research experience through his research activities, in the Signal & speech research laboratory of National Polytechnics School of Algiers, and after in the communication research laboratory of Houari-Boumediene University of Sciences and Technology (USTHB) and actually in laboratory of Applied Sciences of National Higher School of advanced Technologies (ENSTA). He published in reputed international journals. His publishing name is M. Aissiou.