



Alzheimer's disease diagnosis using genetic programming based on higher order spectra features

Mahda Nasrolahzadeh ^{a,b}, Shahryar Rahnamayan ^b, Javad Haddadnia ^{a,*}

^a Department of Biomedical Engineering, Hakim Sabzevari University, Sabzevar, Iran

^b Department of Electrical, Computer, and Software Engineering, Ontario Tech University, Oshawa, Canada



ARTICLE INFO

Keywords:

Alzheimer's disease
Spontaneous speech signal
Feature selection
Classification
Machine learning
Genetic programming

ABSTRACT

In Alzheimer's diagnosis field, Computer-Aided Diagnosis (CADx) technology can improve the work performance of medical researchers and practitioners since it gives early chances to patient's eligibility for clinical trials. The aim of this study is to develop a novel CADx system for the diagnosis of Alzheimer's disease (AD) by utilizing genetic programming (GP) as data-driven evolutionary computation based modeling. The proposed method invokes a majority voting based scheme to select a set of most discriminant features which leads to the highest diagnosis accuracy of the final classification. The effectiveness of GP in categorizing patients with Alzheimer's versus healthy group was revealed by developing models according to their performance in terms of higher-order spectra (HOS) features. The results show that the GP method achieved better performance compared to other the-state-of-the-art approaches. It is also found that the highest accuracy index was yielded by using the proposed data-driven modeling technique. The results of this study emphasize the practicality of GP-based method for developing CADx systems, on the basis of spontaneous speech analysis; can efficiently assist in the diagnosis of Alzheimer's disease.

1. Introduction

A recent survey of Alzheimer's disease (AD) showed that besides memory loss, one of the main issues is the loss of language skills (Gorji et al., 2020; McKhann et al., 2011). Since spontaneous speech is associated with the architecture of the human mind, it plays an important role in human life in terms of understanding, cognition, memory, attention, judgment, and decision-making (Jack & Holtzman, 2013). In addition, a great amount of information is embedded in the speech signal; the information is found directly linked to brain activity which affects the language skills in Alzheimer's or other dementia-related diseases. The speech impairment depends on the stages of the disease. This impairment defines the three stages of AD as pre-clinical stage, intermediate stage, and advanced stage (Hu et al., 2010). The pre-clinical stage is defined by difficulties in finding correct words in spontaneous speech that is often labeled as unfound. People with pre-clinical stage often cannot find the correct word to convey their meanings. In the intermediate stage, the vocabulary and language become poor in everyday use. The advanced stage is associated with limited answers, sometimes restricted to some words. In this stage, patient's ability to manage their daily activities is affected because of the severity of the disorder.

In this context, the analysis of spontaneous speech is a novel powerful noninvasive, and inexpensive approach for the early diagnosis of AD (López-de-Ipiña et al., 2013; Nasrolahzadeh et al., 2016a). However, this methodology may not be perfect, and it is difficult to interpret. The sensitivity of human speech screening is affected by the signal quality and the physician's level of proficiency. Fortunately, computer-aided diagnosis (CADx) technology can improve the efficiency of physicians (Pinheiro et al., 2018). In fact, since the classification implementation exactly influences the precision of the system, it is a remarkably significant part of a CADx system. Many classification methods have been used in the literature (López-de-Ipiña et al., 2013; Nasrolahzadeh et al., 2014, 2015a, 2015b). Each technique has some advantages and drawbacks to deal with classification problems. In this scenario, one of the most challenging topics is feature selection, as some of the features might be irrelevant or/and redundant. On the other hand, using all features might reduce the effectiveness of the classification accuracy. According to literature, by training a model that synthesizes data, Chaudhari et al. (2019) proposed the modified generator generative adversarial networks (MG-GAN) to handle imbalanced data, which is bound to happen in that most real-life signals, such as speech, gene expression data. Moreover, there are several research studies

The code (and data) in this article has been certified as Reproducible by Code Ocean: (<https://codeocean.com/>). More information on the Reproducibility Badge Initiative is available at <https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals>.

* Corresponding author.

E-mail addresses: ms.nasrolahzadeh@yahoo.com (M. Nasrolahzadeh), shahryar.rahnamayan@ontariotechu.ca (S. Rahnamayan), Haddadnia@hsu.ac.ir (J. Haddadnia).

presenting the performance of describing multimodal machine learning and multimodal deep neural nets as vibrant multidisciplinary fields of expanding significance and with phenomenal potential, particularly for the language task (Baltrušaitis et al., 2018; Joshi et al., 2021). Hence, if one selects them by a convenient method, the obtained CADx system may show better performance than other approaches.

Recently, higher order spectra (HOS) have been favored for exploring the behavior of non-linear complex systems in biological signals analysis such as speech signals, electrocardiogram (ECG), and electroencephalogram (EEG) (Acharya et al., 2017; Chua et al., 2010; Nasrolahzadeh et al., 2016b, 2018); moreover, they have been applied as a robust tool to evaluate the non-linear characteristics of biomedical signals, and they performed better even for noisy and poor signals (Chua et al., 2010; Mookiah et al., 2012). Accordingly, our proposed features, higher order spectra (HOS), have been applied for extracting discriminative information from spontaneous speech signals.

According to literature, genetic programming (GP) has proven its potential in several applications. For instance, the GP algorithm is employed for feature construction (FC) and learning (Bi et al., 2021c; Peng et al., 2021), ensembles for image classification (Bi et al., 2020, 2021a). Moreover, GP with a small number of training instances is used for few-shot image classification (Bi et al., 2021b). However, GP has been favored due to its significant application field, namely its application toward the symbolic regression (SR). It aims to find an optimal mathematical model that can describe and predict the observed response based on input data from the given system.

In order to minimize the various error criteria, SR explores the space of mathematical expressions. It can also concurrently search both the parameters associated with a function that is optimized for given data and the appropriate linear or non-linear structures. SR makes no hypothesis about the shape of the possible functions of curve fitting. In addition, SR has the ability to identify key predictions and combine them in the data. On the other hand, the best expressions optimized and searched by SR are less complex and, consequently, it is simple to deploy. Moreover, the models acquired are understandable to the user and able to explain discovered phenomena underlying the data. Also, GP can make its mathematical model based on a small size training set. Because of such attractive characteristics, GP has been used widely in a wide range of applications. For instance, it has been utilized in predictions biochemical systems based on developing soft-sensor (Sharma & Tambe, 2014), prediction of higher heating values of solid biomass (Ghugare et al., 2014), modeling of fermentation (Barm-palexis et al., 2011), as an optimizer in pharmaceutical formulation, as a classifier in diagnostic systems, the analysis of medical signals, detecting and predicting diseases in multiple areas, such as neurology, cardiology, and plethysmography (Bhardwaj et al., 2016; Lones et al., 2013; Sannino et al., 2015). Despite such attractive characteristics and significant potentials, GP has not been used as much as artificial neural networks (ANNs) and other machine-learning algorithms in the various fields of science, technology, and engineering.

Current research is focused on developing a novel CADx system for the early diagnosis of AD. The aim is to obtain the best prediction models for the early diagnosis of AD through HOS's extraction of features of speech data. More specifically, it aims at extracting the most relevant features from spontaneous speech signals of the healthy group and those with AD. However, a major drawback is the small number of available samples compared to the very high dimensionality of the feature space. To overcome this challenging problem, this paper investigates the performance of the GP as a very competitive and powerful technique.

The paper is structured as follows: In Section 2, the databases utilized in this study are described. In Section 3, the methods and the quantification analyses are presented. In Section 4, the results of the experiments and the discussions are reported. Finally, Section 5 presents the conclusions.

2. Database

In this study, to demonstrate the performance of the proposed approach, two databases are employed as follows:

2.1. Persian language database

2.1.1. Data

Persian language data in this study were collected from an old nursing home in Sabzevar, Iran, according to the accessibility of the special communication problems the patients faced depending on the relative stage of their disease for age-matched healthy normal subjects and patients with AD. The research was performed in accordance with the Institutional Review Board of all participants; it followed the ethical standards laid down in the Declaration of Helsinki (McKeith et al., 1996). After informing about the purposes of the study, all persons provided their agreement before inclusion.

2.1.2. Subjects

Spontaneous speech signals of 60 subjects were recorded while subjects were explicitly requested to tell elegant personal stories, express their feelings, and have a friendly conversation. The participants were a group of 30 healthy control (HC) subjects (15 women and 15 men) with the age range of 52–98 years old, and a group of 30 AD patients (14 women and 16 men) with the same age range, in three levels of Alzheimer's which included First Stage (FS), Second Stage (SS), and Third Stage (TS), where FS = 6, SS = 15, TS = 9.

Compared to the healthy individuals, the recording time was shorter for the subjects with Alzheimer's since they had difficulty in finding the words. They spoke more slowly with less clarity and paused longer. Their message was intermittent or unfinished, with more time needed for finding the words. Thus, they were tired early and wanted to end the conversation. An audio recorder was applied for recording the speech signals. The extraction of the recorded audio was done in WAV format with a frequency sampling rate of 16 kHz. Recording of the speech signal for the control and AD groups lasted 15 to 17 h, in the respective order. The recording was done in a friendly atmosphere. More details of the database can be found in our other published papers (Nasrolahzadeh et al., 2016a, 2018).

2.2. English language database

2.2.1. Data

In this paper, the Pitt Corpus as English data was utilized from the DementiaBank database.¹ The data was launched between 1983 and 1988 by the University of Pittsburgh, as a portion of the Alzheimer Research Program (MacWhinney et al., 2011). The details about the study cohort of this database are dealt with (Becker et al., 1994).

2.2.2. Subjects

The participants include a group of 167 who were diagnosed as suffering from "possible" or "probable" Alzheimer's disease and a group of 97 healthy control subjects. It is noteworthy that no distinction was made between individuals diagnosed with probable and possible AD in this study. In addition, they were particularly asked to describe the standard Cookie Theft picture from the Boston Diagnostic Aphasia Examination (Becker et al., 1994). After audio processing, 240 and 233 of the narrative samples from the AD and HC groups were yielded, respectively.

3. Methodology

Fig. 1 represents the overall structure of the proposed approach to AD diagnosis. The input data is spontaneous speech signals explained in Section 2. In the rest of this paper, the implementation of the proposed method is described in detail.

¹ (<https://dementia.talkbank.org>).

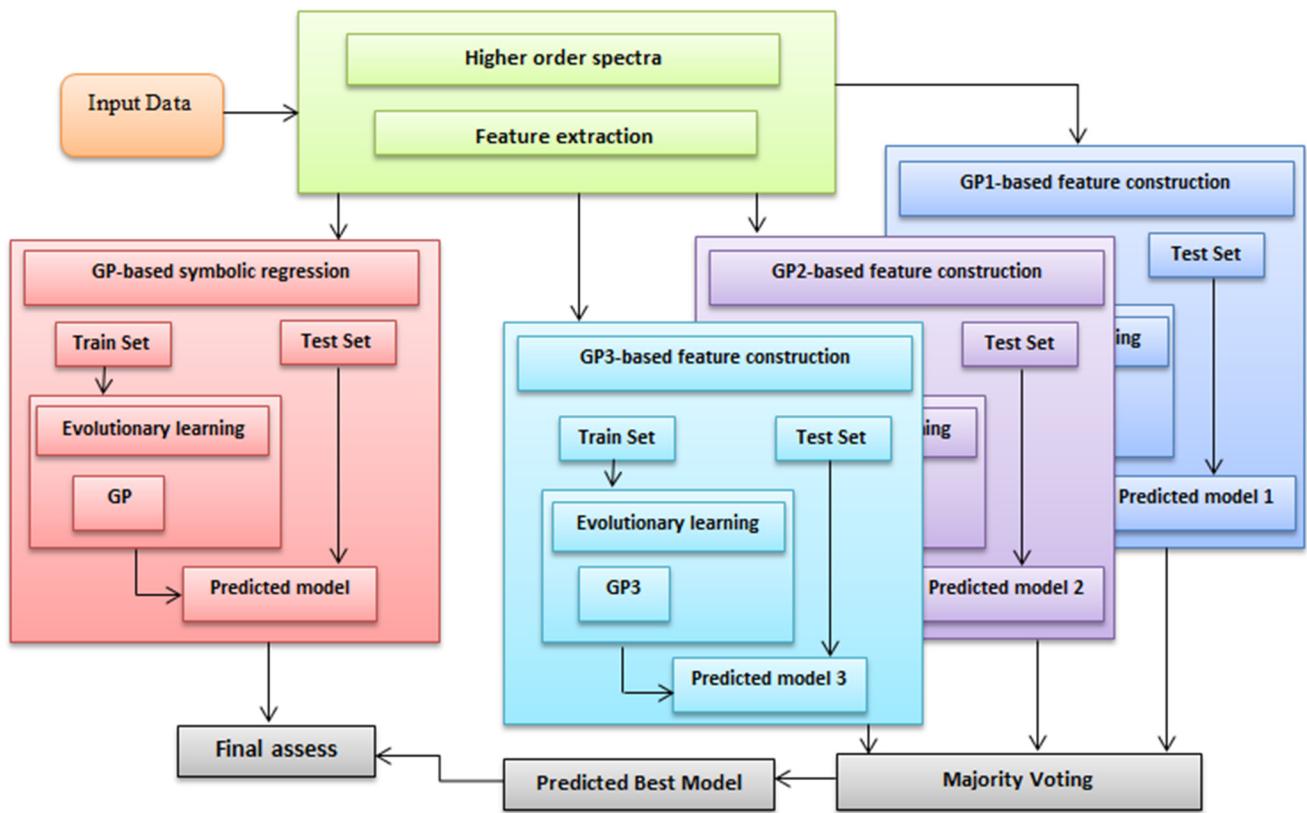


Fig. 1. Overall structure of the proposed approach to Alzheimer's disease diagnosis.

3.1. Feature extraction and selection method

The design of an accurate diagnostic system is highly dependent on the selection of the feature extraction method, since it can yield the most data-related knowledge to build a reliability model. In this spirit, if one elicits them correctly, the results of the proposed CADx system may perform superior to other conventional approaches. Therefore, a set of features based on HOSs were applied to extract discriminative information from spontaneous speech signals of patients with Alzheimer's and healthy subjects, as inputs to categorize these groups. Brief descriptions of the features, as well as how to extract them, are addressed in the following parts of this paper.

3.2. HOS-based features

Since the dynamics of human speech are chaotic, in spontaneous speech analysis, the HOS can be considered as a non-linear criterion capturing subtle changes in signals (Nasrolahzadeh et al., 2016b; Nikias & Raghubeer, 1987). Hence, this method can be employed on spontaneous speech signals for extracting non-linear features along with understanding the behavior of their dynamics for automatic classification.

HOS are extensively used as spectral representations in spectral systems analysis (Martis et al., 2013). They can provide more statistical information such as cumulants or moments of a random process to the power spectrum (Chua et al., 2008). Consider $x(n)$ as an nth-order stationary random process. The ith-order moment of this process can be computed as follows:

$$m_i^x(\tau_1, \tau_2, \dots, \tau_{i-1}) = E\{x(n)x(n+\tau_1)\dots x(n+\tau_{i-1})\} \quad (1)$$

In which ith-order moment spectrum denotes as $(i-1)$ dimensional Fourier transform (FR) of the nth-order moment of $x(n)$. The third-order spectrum of higher order spectra is based on a concept, called

the bispectrum, $B(f_1, f_2)$. The prefix bi denotes a signal with two frequencies. The bispectrum is an FR of the third moment and is given as follows:

$$B(f_1, f_2) = E[X(f_1)X(f_2)X^*(f_1 + f_2)] \quad (2)$$

where X is the FT of the signal x , X^* is the complex conjugate of X , and $X(f)$ is the discrete-time Fourier transform, respectively. In addition, $E[\cdot]$ denotes the expectation. The bispectrum consists of two frequency-variable functions with complex values, as shown in Eq. (2). Needless to say, the Fourier transform of a real value of the signal enjoys conjugate symmetry, as well as the power spectrum in the area of negative frequency, is redundant. Thus, bispectrum is computed in the non-redundant area is the resultant multiplying of the three Fourier coefficients, which also exhibits symmetry. Note that the Nyquist frequency is employed to normalize frequency "f" between zero and one (Nikias & Mendel, 1993; Ning & Bronzino, 1990). The values of the bispectrum are defined in the bispectral area by the following conditions:

$$0 \leq f_2, \quad f_2 \leq f_1, \quad f_1 + f_2 \leq 1 \quad (3)$$

The bispectrum region is also known as the triangular region, Ω , which can be calculated from the non-redundant region (Nikias & Mendel, 1993), as shown in Fig. 2.

Bispectrum can be computed through several methods (Li et al., 2005; Mendel, 1991; Xianda, 2002). In this paper, both direct (Fast Fourier Transform (FFT)-based) and parametric methods were estimated using the MATLAB toolbox. These methods are described in the following sections.

3.2.1. Direct (FFT-based) method

In this approach, the recorded signal is sampled first and divided into multiple overlapping frames (Rao & Gabr, 2012). The average is removed from each record, and then the Fast Fourier Transform is

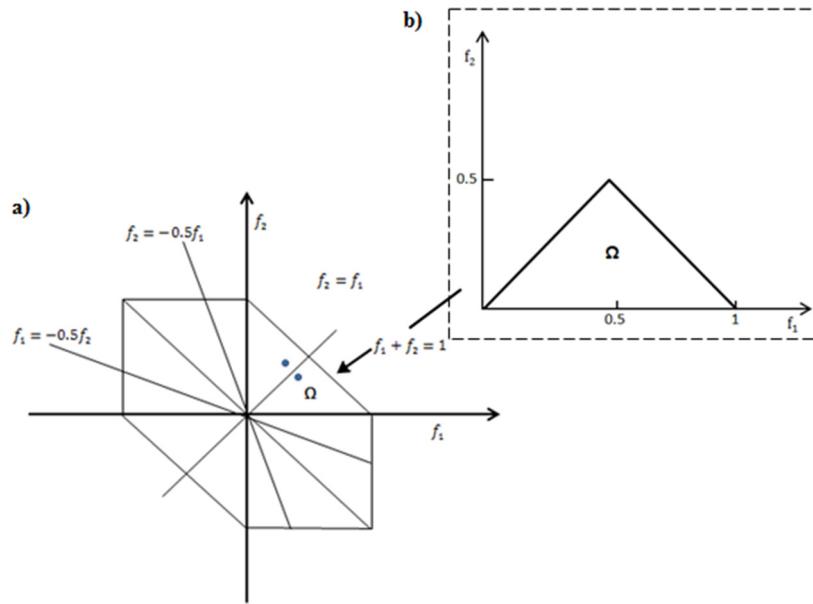


Fig. 2. The computation of the bispectrum for real signals. (a) The bispectrum region. (b) Non-redundant or triangular region Ω .

computed. Finally, the bispectrum is computed according to both the moment and cumulant spectrum. The estimation of bispectrum using the direct (FFT-based) method is calculated as follows (Childers, 1978):

$$B_N(k, l) = X_N(k) \cdot X_N(l) \cdot X_N^*(k+l) \quad (4)$$

where X_N is the Fast Fourier Transform of the N -th record. In order to reduce the calculation variance, the data is windowed and smoothed (Rao (1983)). In this study, the Rao–Gabr window is used (Childers, 1978). Moreover, in the direct (FFT-based) method, the percentage overlap between segments is set to zero, and the FFT length is set to 128.

3.2.2. Parametric method

The autoregressive (AR) parameters are required to calculate bispectrum using the parametric method. In the AR model, the coefficients are easily calculated in terms of a group of linear equations. If there is new data, the capability of updating the coefficients by using the Kalman filter equation (Ning & Bronzino, 1990; Sigl & Chamoun, 1994). In addition, the order of the AR model accord with the index where the singular values suggest the greatest decline (Swami et al., 1998). The AR model is defined as

$$y(n) + \sum_{i=1}^p a_i y(n-i) = x(n-i) \quad (5)$$

where $y(n)$ is a non-Gaussian function and p is the order of the model. In this study, the order of the AR model is set to 10.

In order to find a salient signature in the speech signal which permits to distinguish between HC subjects and patients with AD as well as analyze the effect of HOSs on the performance of the proposed method, a group of linear bispectral based features and also a group of non-linear bispectral based features have elicited from spontaneous speech signals in the non-redundant region Ω , separately. In what follows, the implementation of these two groups will be discussed.

- Linear bispectral based features: This group includes three linear features based on the energy of the bispectrums, namely bispectral mean, bispectral maximum, and bispectral minimum, which are calculated directly from the non-redundant region as follows:

1. Bispectral Mean

$$B_{avg} = \frac{1}{k} \sum_{\Omega} |B(f_1, f_2)| \quad (6)$$

where k denotes the total number of data points within the non-redundant region Ω .

2. Bispectral Maximum

$$B_{max} = \left| B(f_1, f_2) \right|_{\Omega}^{\max} \quad (7)$$

3. Bispectral Minimum

$$B_{min} = \left| B(f_1, f_2) \right|_{\Omega}^{\min} \quad (8)$$

- Non-linear bispectral based features: This group generally falls into two categories. The first category is assigned to frequency-relation based features. Note that features related to this category are calculated inside the bispectrum area. In this study, only one feature of the aforementioned category is used as follows:

1. The sum of logarithmic amplitudes of diagonal data points of the bispectrum (Acharya et al., 2008):

$$A = \sum_{\Omega} \log \left(\left| B(f_1, f_2) \right| \right) \quad (9)$$

The second category of non-linear features includes bispectrum features on the basis of the magnitude of data disturbance or entropy and also calculated in the bispectrum region Ω . The following equations used to determine spectral entropy-based HOS are defined as.

2. Phase entropy (Ph_e) of the bispectrum (Acharya et al., 2008):

$$Ph_e = \sum_n p(\psi_n) \log p(\psi_n) \quad (10)$$

where

$$p(\psi_n) = \frac{1}{M} \sum_{\Omega} I(\phi(B(f_1, f_2)) \epsilon \psi_n) \quad (11)$$

$$\psi_n = \begin{cases} \phi | -\pi + \frac{2\pi n}{N} \leq \phi < -\pi + \frac{2\pi(n+1)}{N} \end{cases}, \quad \forall n = 0, 1, \dots, N-1 \quad (12)$$

In which ϕ defines the phase angle of the bispectrum. $I(\cdot)$ is an indicator function (Mookiah et al., 2012) that has a value of 1 when the phase angle is within the range bin ψ_n , as expressed by Eq. (10).

3. Bispectral entropies (Mookiah et al., 2012): In this study, three features include P_{e1} , P_{e2} , and P_{e3} are derived from the bispectral Phase entropy Ph_e ; their definitions are given as follows:

$$P_{e1} = - \sum_k \frac{|\mathbf{B}(f_1, f_2)|}{\sum_{\Omega} |\mathbf{B}(f_1, f_2)|} \log \frac{|\mathbf{B}(f_1, f_2)|}{\sum_{\Omega} |\mathbf{B}(f_1, f_2)|} \quad (13)$$

$$P_{e2} = - \sum_i \frac{|\mathbf{B}(f_1, f_2)|^2}{\sum_{\Omega} |\mathbf{B}(f_1, f_2)|^2} \log \frac{|\mathbf{B}(f_1, f_2)|^2}{\sum_{\Omega} |\mathbf{B}(f_1, f_2)|^2} \quad (14)$$

$$P_{e3} = - \sum_n \frac{|\mathbf{B}(f_1, f_2)|^3}{\sum_{\Omega} |\mathbf{B}(f_1, f_2)|^3} \log \frac{|\mathbf{B}(f_1, f_2)|^3}{\sum_{\Omega} |\mathbf{B}(f_1, f_2)|^3} \quad (15)$$

As mentioned above, all eight features that we extracted from the spontaneous speech signals were calculated for the two aforementioned databases by using both direct (FFT-based) and parametric methods. Thus, we entirely utilized sixteen features for each segment of the signal.

3.3. GP

Genetic Programming (GP) is a well-known evolutionary algorithm and widely utilized in evolutionary based computations (Al-Sahaf et al., 2019; Poli et al., 2008). GP can be used to find global optima in a wrapped search space. It also can develop optimization algorithms inspired by Darwin's theory of evolution (Koza & Koza, 1992). GP searches for syntactic expressions by an evolutionary process using selection, crossover, mutation, and cloning operations, to find an expression, which yields better relevance between a group of independent (inputs) and dependent (output) variables (Emigdio et al., 2017). GP can optimize the model structure alone, and its solutions have symbolic nature. In addition, its representation is flexible. These important characteristics make GP a great approach for symbolic regression. On the other hand, the solutions evolved by the GP method provide powerful interpretability in how the features are learned or extracted from the signals, and what they are influential for classification (Bi et al., 2021c).

To the best of our knowledge, the present study is the first in which the GP approach has been used with the aim of predicting and early diagnosing AD, based on spontaneous speech data. In what follows, the details of two proposed schemes based on GP are discussed.

3.3.1. Scheme-I: GP-based SR

In the first approach followed in this study, GP-based SR is used to identify the best AD prediction model. The task of SR is to seek an optimal mathematical expression based on observing the input–output data in order to predict and describe a desirable system (Emigdio et al., 2017). In fact, SR is a suitable approach for making a linear/non-linear model and parameters related to a function, f , so that the best-fit represents the input and response data. In this process, each mathematical model includes N inputs (x_1, x_2, \dots, x_N) and one output (y) from the data set and its equation is defined as follow:

$$y = (x_1, x_2, \dots, x_N; \alpha_1, \alpha_2, \dots, \alpha_i) \quad (16)$$

where α is a function parameter (Ghugare et al., 2014). SR does not require any predetermined form and size of the model; it does not need to assume any distribution of the data either. In addition, its process is based on an iterative approach. At the beginning of the process, it starts with the formation of a population that consists of a certain number of predetermined random equations, including potential candidate solutions. During the optimization task, to find the most suitable relation between the input (predictor/independent/descriptor) and output (dependent) variables, in the current generation, they compete in the most efficient way for the best-fit model in the dataset. Usually, the SR method consists of four main operations: (1) fitness evaluation, (2) selection, (3) crossover, and (4) mutation (Kalkreuth et al., 2015). Here, mutation and crossover as primary biological search operators are responsible for the creation of the new candidate equations. In

this regard, not only does re-combining the previous equations lead to the generation of new candidates (called offspring), but also their sub-expressions change by mutation operator. In other words, the mutation occurs by flipping operators to other ones from operations set or arguments to other ones in inputs set the chosen chromosomes, with a probabilistically varying their sub-expressions. The mutation operator creates all the next generation individuals, and the SR algorithm also employs crossover to create the residual population, namely, the previous equations, called cloning. In fact, a couple of chosen chromosomes can exchange their information due to the crossover operator. It is believed that the suitability of a selective solution is defined as a quantitative estimation of how it adeptly fits the data set (Siedlecki & Sklansky, 1993). This means that chromosomes are more likely to survive, which have values with high fitness. In such procedures, after evaluating all the candidates in the current population, those candidate solutions that fit the data set comparatively better are preserved by the GP-based SR algorithm, while the unfavorable ones are rejected. Then, the selected solutions are applied to the crossover and mutation operations. Seeking in the high-dimensional input space runs iteratively, until a predetermined criterion is met. The commonly utilized criteria are as follows:

1. Achieve a desirable accuracy in a population with the best candidate solution.
2. Based on the maximum number of predetermined generations.
3. The value of fitness stays stable for the best candidate of the population, or its variations are only negligible in a vast number of generations.

Genetic operators are implemented on nodes in tree-formed chromosomes. Fig. 3a illustrates a hierarchically structured tree of a genetic operator to encode a mathematical equation. The operand and operator nodes are emanated from a root node. In this process, the depth of the trees is variable, so it possesses the ability to construct expressions with different lengths and complexities. Trees are easily capable of evaluating in a recursive way, pointing to the fact that the performance of the GP algorithm is based on a stochastic procedure; hence, this is likely done by running GP repeatedly. For example, the use of different initializations at random would result in different solutions. Consequently, to produce an acceptable model, it is necessary to analyze a set of input data of several runs (Gray et al., 1998). In this regard, Fig. 4 represents the flowchart of the GP-based SR, and the detailed steps can be described as follows:

• Step 1: Initial population

Generate the first population of candidate solutions at random with size N_p . Each candidate solution, based on a tree structure, is randomly generated using the function set and terminal set as inputs, which in our case are properties. A function set uses a set of mathematical operations, such as $F = \{+, -, \div, \wedge, \times, \sin, \cos, \exp\}$, while a terminal set is expressed as follows.

$$T = (\text{inputvariable descriptor/predictor})(x_N), \\ \text{ephemeral random constant } (\alpha_i)) \quad (17)$$

In this study, a terminal set called an “operand” node, involving eight features of bispectrum computed from the speech data in the area of Ω .

• Step 2: Fitness function

Evaluate the fitness value of every candidate solution to obtain the target SR. In the present generation, the evaluation of each candidate solution is expressed as follows:

- (1) In order to estimate a prediction error metric like the mean absolute error, compare the resultant outputs of input data in the candidate solution with the output values of their target.

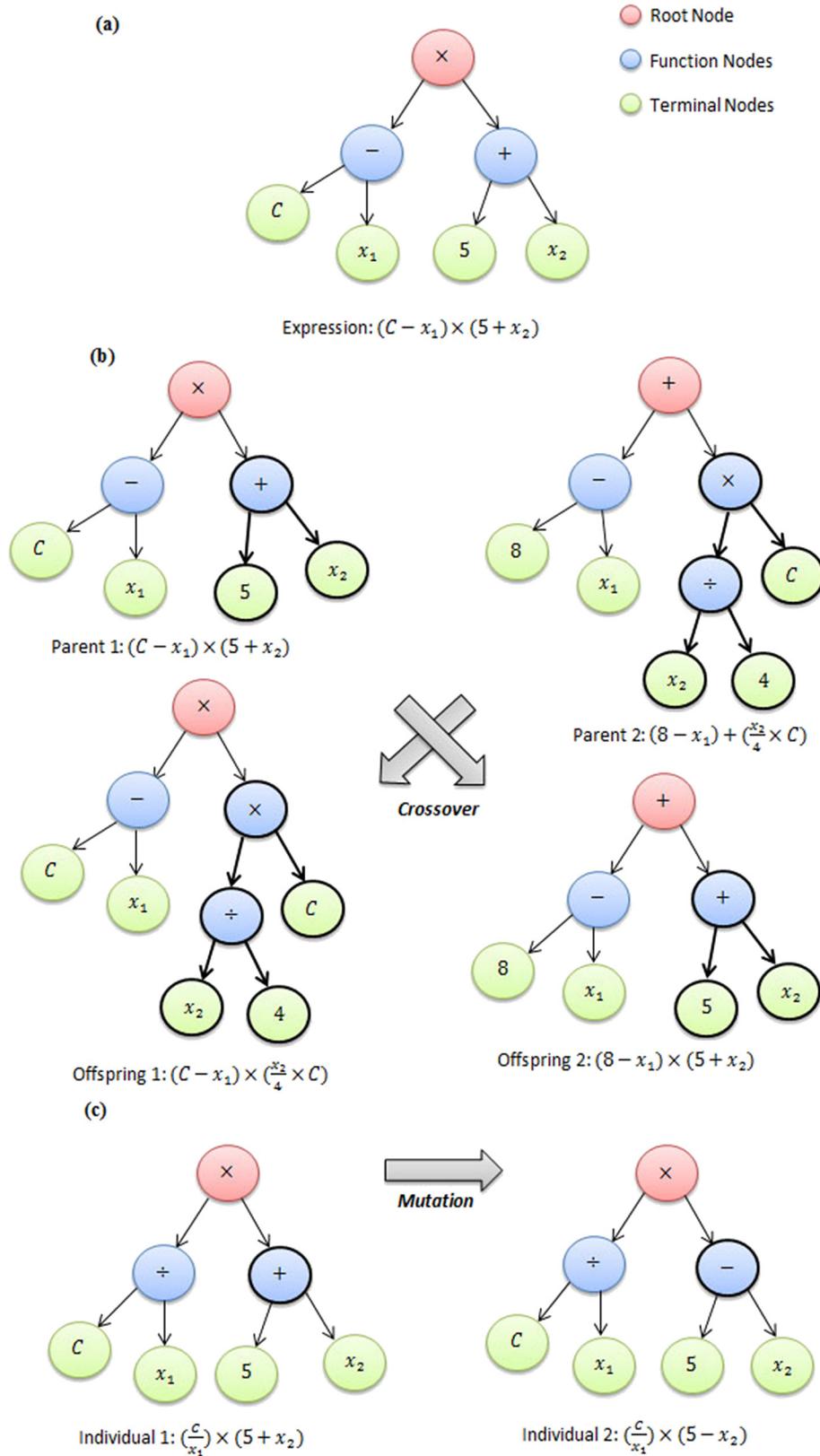


Fig. 3. Schematic diagram of the GP algorithm, (a) an original tree structure, (b) crossing operation, and (c) mutation operation within the evolutionary process of GP. The symbols used in the figure: Operators include $(-)$ subtraction, $(+)$ addition, (\div) division, and (\times) multiplication. Operands include x_1 , x_2 , and C .

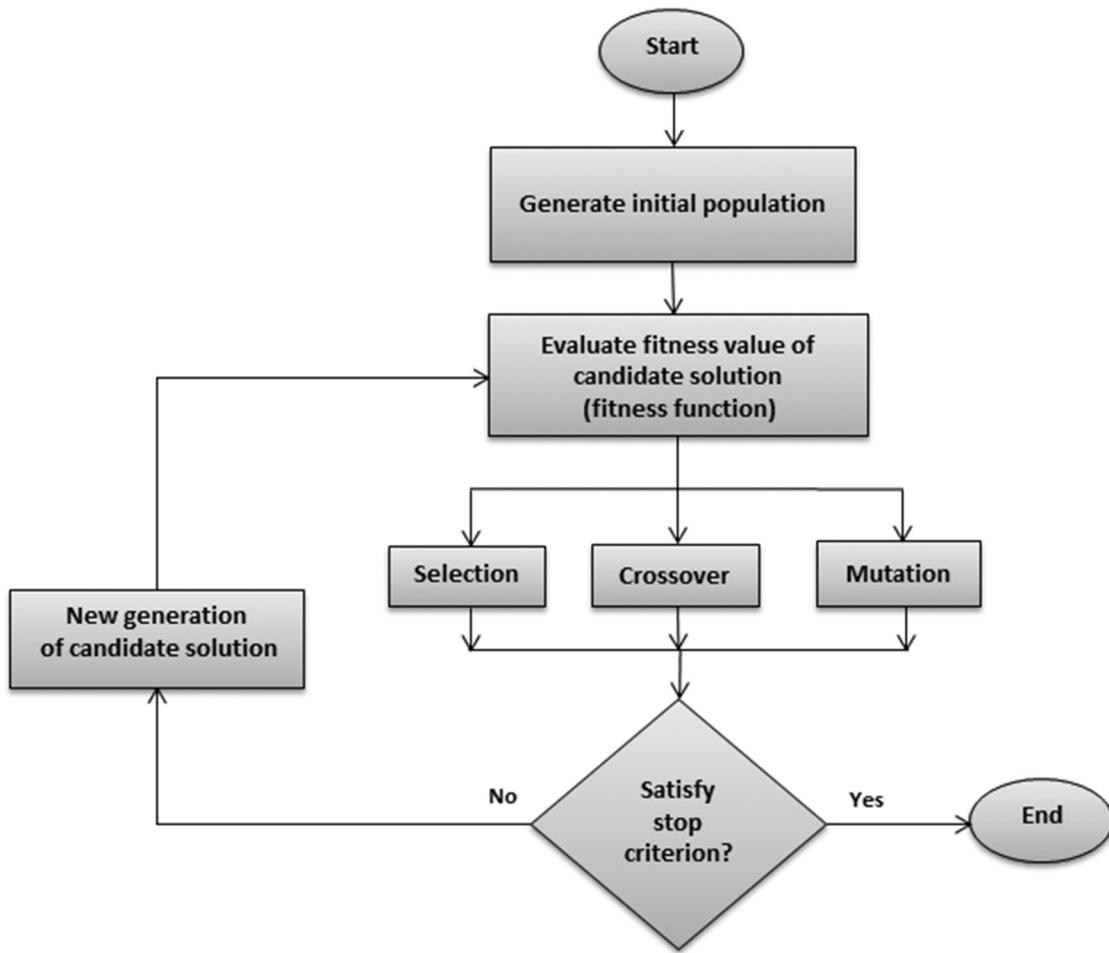


Fig. 4. The flowchart of the GP-based symbolic regression.

- (2) In order to calculate the fitness value of the candidate solution, the error metric value is used by a fitness function that is properly selected. This process runs cyclically for all candidate solutions in the present population. Additionally, the fitness function allocates a lower value of fitness to candidate solutions that perform incorrect output predictions and, vice versa, allocates a higher value of fitness to candidate solutions that perform close to correct output predictions.

• Step 3: Selection

Select the parent candidate solutions in terms of their fitness values. During the selection process, the candidates with higher fitness values, in the current generation, have more chances to enter the mating pool. For this stage, the available techniques are used, including Roulette-wheel, elitist mating, and tournaments (Yadavalli et al., 1999); the last one is the commonly used scheme.

• Step 4: Crossover

Conduct the generation over the union of the parents and offspring. A pair of parents is randomly chosen from the mating pool, and then two candidate solutions as offspring are produced. This means that a subcategory is chosen on a random basis from each parent tree, and these are exchanged reciprocally between the parent trees giving birth to two offspring (see Fig. 3b). The generation of a certain number of offspring is considered as a criterion for stopping this stage; otherwise, this process is repeated until it reaches the stop criterion.

• Step 5: Mutation

Replace the nodes of the trees as candidate solutions of offspring chosen randomly with low possibility, with the same kind of nodes at random (see Fig. 3c). In other words, in this process, variety is

preserved in the population of offspring. In addition, it avoids trapping the candidate's solutions at a local minimum. The population of offspring after the mutation reflects the candidate solutions of the next generation.

• Step 6: Termination

Stop if the criterion is satisfied; otherwise, $generation = generation + 1$ and go to Step 2.

Furthermore, for more information on how to calculate the GP formalism, readers are referred to Koza and Koza (1992), Langdon and Poli (2013), Poli et al. (2008) and Vyas et al. (2015).

In this study, at intensifying the execution of a novel CADx system for the early diagnosis of Alzheimer's employing the GP-based models, a software package was used termed Eureqa Formulize (Schmidt & Lipson, 2009). Since the Eureqa Formulize package is expected to enjoy a good generalization, it has been optimized to find a set of effective models. Besides, it suggests several options for pre-processing data and the generation of candidate expressions. Because of these important characteristics, the present study was designed to further investigate these effects at ensuring models with high predictive accuracy and generalization capability for the early diagnosis of AD using HOS-based features.

3.3.2. Scheme-II: GP for FC

In general, the quality of the feature enjoys of considerable importance for the early diagnosis of AD. Traditional methods to enhance the quality of extracted features employ common techniques such as feature selection, feature learning, and signal processing (López-de-Ipiña et al., 2013; Nasrolahzadeh et al., 2020, 2016a). In this regard, FC is an efficient technique in which the resultant set of high-level

and enlightening features is newly created based on the original low-level ones (Al-Sahaf et al., 2019). It has been shown that the features of spontaneous speech signals can be elicited from multi-records, and each record provides various properties (Nasrolahzadeh et al., 2016b). Nonetheless, to the best of our knowledge, FC, especially constructing features through multi-records, which can generate enlightening features to make better the AD diagnosis, has not been utilized in this field.

GP is an evolutionary computation approach that can be used as a suitable FC method (Al-Sahaf et al., 2019). Because its tree-based representation possesses a changeable length, GP can automatically make high-level features from low-level ones flexibly. In the feature construction process, GP evolves models based on original features and a set of mathematical operators like \times , $+$, and so on. These operators are applied to the original features and thus lead to the construction of new features (Peng et al., 2021).

As mentioned above, a small number of available samples compared to the very high dimensionality of the feature space cannot well demonstrate the distribution information of the class and cause weakens the performance of generalization. To address this issue, the GP algorithm is proposed in this paper to adaptively make the number of enlightening features from multi-records for comprehensive representing of data samples. Due to the non-stationary and nonlinear nature of spontaneous speech signals, three different records of each subject were considered for the extracted features. The extracted HOS features are represented by *Record1*, *Record2*, and *Record3*, respectively. Then, the converted sample data sets are divided into two training and test sets, and three GPs are employed separately to create high-level features from each single-record feature set. The function set involving $\{+, -, \div, \wedge, \times, \sin, \cos, \exp\}$ and the terminal set including eight features of bispectrum computed from the speech data in the area of Ω are used for FC from different records. It should be noted that to create a set of effective features that can comprehensively describe data samples and improve generalization performance, the features of these three records are constructed separately. Finally, the three constructed features sets are evaluated through the majority voting stage, described in the next section.

3.4. Majority voting

It is worth noting that the GP is based on a stochastic optimization process normally performed several times. Hence, it cannot be expected that even after simplification, each optimization will converge towards the same solution again. The resulting solutions seem different, and some of them possess a better fitness value than the other solutions. In general, each solution contains a set of terms; however, in most resultant solutions exist many of these terms. In addition, the major factors in the original expression are often made up of common terms. In this spirit, analysis of several separate optimizations runs indicates the possible components of the proposed model. On the other hand, majority voting causes common features to be found in selected individuals so that they are expected to be the most useful features. In fact, the performance of majority voting is similar to local search, which can increase the potential exploitation of the search algorithm (Mahdavi et al., 2019). Therefore, in order to raise the accuracy of the proposed CADx system, the use of the GP method to select the most discriminant features and subsequently to execute the final classification conforming to a majority vote decision-based strategy is proposed. Fig. 5 represents the schematic diagram of the majority voting procedure. When a test feature undergoes the GP-based model, the three outcomes predicted by the GP will assign a label to the test feature, respectively. In the labeling process of features yielded from GP, a threshold is assigned to each class. Then, the majority threshold is set for above K% of it and takes a vote among outcomes predicted by the GP on each feature, separately. The feature that has occurred more often on the same majority threshold is selected as the most promising feature for

the corresponding class. As a result, it gets a “+” label. “+” means that the feature predicted by the GP belongs to the corresponding class. While “-” means that the feature predicted by the GP does not belong to the corresponding class. The possible labels are $(+, +, +)$ or $(-, +, +)$ or $(+, -, +)$, and so on. As such, if we have two or more than labels of “+”, the feature predicted by the GP is recognized as the corresponding class. In other words, when at least two outcomes predicted by the GP agree on the same decision, the output of the voting decision is executed.

In this way, the voting starts after the termination of the GP evolution process to produce an optimum model. Next, a counter is assigned to determine the majority of every class. The voting ends when the counter reaches the threshold of the majority. Finally, the final decision is made. For that, we call this process decision-making based on majority voting.

3.5. Implementation, parameter settings, and measures

A software package called Eureqa Formulize is utilized to implement the GP-based AD prediction models (Schmidt & Lipson, 2009). It is optimized to detect parsimonious models owning good generalizability. This package provides some options for generating candidate expressions and data pre-processing. The effectiveness of these options is meticulously studied in both case studies with the aim of assuring models enjoying high AD prediction accuracy and generalization ability.

The parameter settings of Eureqa Formulize are presented as follows. This software possesses several options for the pre-processing of data including outlier removal, smooth the data, normalization of scale and offset, handle missing values, and data filtering. All of these options are selected for this study. Moreover, Eureqa Formulize can select the operators. The arithmetic operators involving $+, -, \div, \wedge, \times, \sin, \cos, \exp$ are used in this study. The Formulize software is capable of data splitting (i.e., choice of training and test sets). In the present study, training and test sets are set to 70% and 30%, respectively. Approximate and final GP models are made using the corresponding training sets. Several modelings runs are performed so that the effect of changes in the above-mentioned Formulize options on the converged solution is studied in detail. This is conducted to assess the generalization applicability on the convergent model.

The performance of the proposed system is assessed by the percentages of three statistical measurements, i.e., mean squared error (MSE), R^2 goodness of fit (R^2GF), and mean absolute error (MAE). The definitions of these metrics are described as:

- **MSE:** Measuring the difference between predicted values by the model and the desired objective for the three stages of AD and HC groups; its definition is expressed as below:

$$MSE = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)^2 \quad (18)$$

where X_i and \hat{X}_i denote observed values and predicted values, respectively. In addition, n is the number of data samples.

- **R^2GF :** Measuring the difference between predicted values by the model and the desired objective for the three stages of AD and HC groups. The degree of each linear correlation between observed output and predicted output is determined by R^2 . Besides, a particular linear correlation should only be considered to evaluate the goodness of fit. Therefore, it can be uttered using the following equation:

$$O_i = 1 \times \hat{O}_i + 0 \quad (19)$$

where O_i and \hat{O}_i denote observed output and predicted output, respectively.

- **MAE:** Measuring the difference between predicted values by the model and the desired objective for the three stages of AD and HC groups; its definition is expressed as below:

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (20)$$

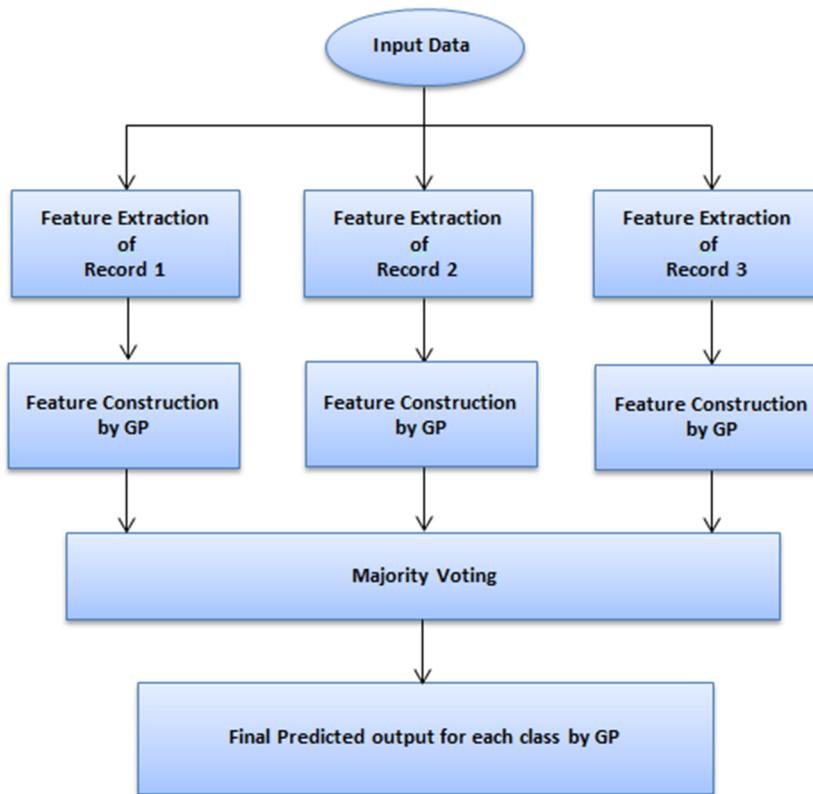


Fig. 5. The schematic diagram of the proposed majority voting scheme.

where y_i and x_i denote prediction and true value. Moreover, n is the total number of data samples.

The models were accepted, which satisfy the respective conditions: (1) The quantities of large and comparable of R^2 GF, and (2) the quantities of small and comparable of MSE and MAE for the datasets of training, validation, and test.

In addition, to further look into the robustness of the proposed method, we have evaluated the classification accuracy measure as the fitness function, which is one of the most popular benchmark functions of a CADx system and is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (21)$$

where TP, TN, FP, and FN are the number of true positive, true negative, false positive, and false negative answers, in the respective order. This recognizes that a true positive prediction happens when both the physician and the classifier suggested the presence of a positive detection. A true negative prediction happens when the absence of a positive detection of the classifier and the physician occur simultaneously. A false positive decision happens when the classifier incorrectly labels a negative case as a positive. In contrast, a false negative decision happens when the classifier incorrectly labels a positive case as a negative.

4. Experimental results and discussion

The purpose of this study was to evaluate the GP algorithm as a significant alternative to multilinear regression, artificial neural networks, and other models of data-driven formalism to the development of the CADx system. To do this, the performance of the proposed method was analyzed more deeply over two datasets. In order to investigate the results, these datasets were considered as two case studies: datasets in the first case study had two thousand data points consisting of healthy subjects and three stages of AD, and in the second case study, they had

more than two hundred thousand recorded sample data of two groups, namely, Alzheimer's and healthy subjects.

Figs. 6 and 7 illustrate one of the main outcomes of our proposed approach by using two databases. The same observations are yielded in these representations. As depicted in Fig. 6, the first row (i) shows representative plots of spontaneous speech signals recorded from an HC subject and three different stages of AD. It is easy to observe the poor quality of the speech signals for the patient with Alzheimer's because of the poor language skills revealed in speaking, understanding, thinking, and relationship with daily life. The segmentation of spontaneous speech signals was essential for HOS quantification assessment to be conducted on the speech signals. More information on how to calculate the processing and the segmentation of the speech signals were presented in our earlier works (Nasrolahzadeh et al., 2016a, 2014). Bispectrum was estimated for each speech segment via both parametric and direct (FFT-based) methods. In this paper, bispectral analysis was run using the MATLAB toolbox. In the second row (ii), the bispectrum estimated using the parametric method over the spontaneous speech is shown, and in the third row (iii), we show the bispectrum estimated by the FFT method. Regarding the bispectra of the four groups, it is obvious that the dynamical patterns of the spontaneous speech signals are completely different among various stages of AD and HC subjects. Moreover, the colors red and blue represent the highest rise and the highest decrease in amplitude of the bispectrum, respectively. Each data sample offers the quantity of bispectrum for the speech signal at the bifrequency (f_1, f_2), as illustrated in Fig. 6. The results show that the bispectral analysis can reveal additional information non-acquirable from the power spectrum and may impact the attribute components of human speech in various psychological stages of Alzheimer's differentially.

According to Figs. 6 and 7, the bispectrums are significantly different among the groups in the two mentioned databases. It can be seen that the bispectrums during different stages are capable of capturing subtle changes in the signal and the dynamical behavior of the speech

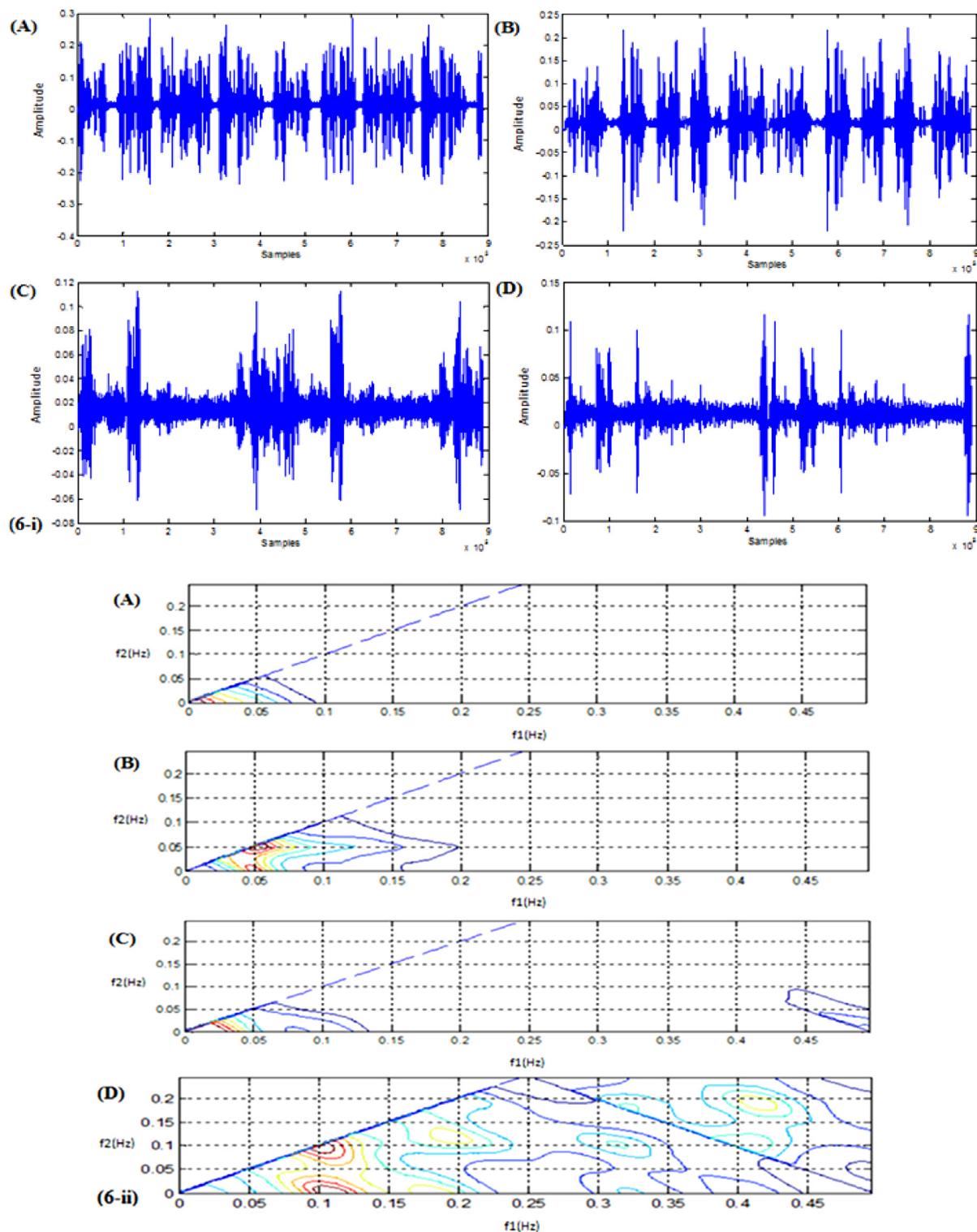


Fig. 6. Representative (6-i) spontaneous speech signals from Persian language database, (6-ii) bispectrum estimation through the parametric method, and (6-iii) bispectrum estimation through the FFT method belonging to (A) a healthy control subject (B), (C), and (D) three stages of Alzheimer's patients, respectively. As can be seen in (6-i), there is considerable poverty in the signals of patients with Alzheimer's compared to a healthy subject. Note that poverty in the signal is more obvious in the more advanced stages of the disease. This is because the language skills are reduced in Alzheimer's patients. The bi-spectral results calculated using both parametric and FFT methods are entirely different in structure, as illustrated in (6-ii) and (6-iii). Each range of bi-spectral values is associated with a specific color. . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

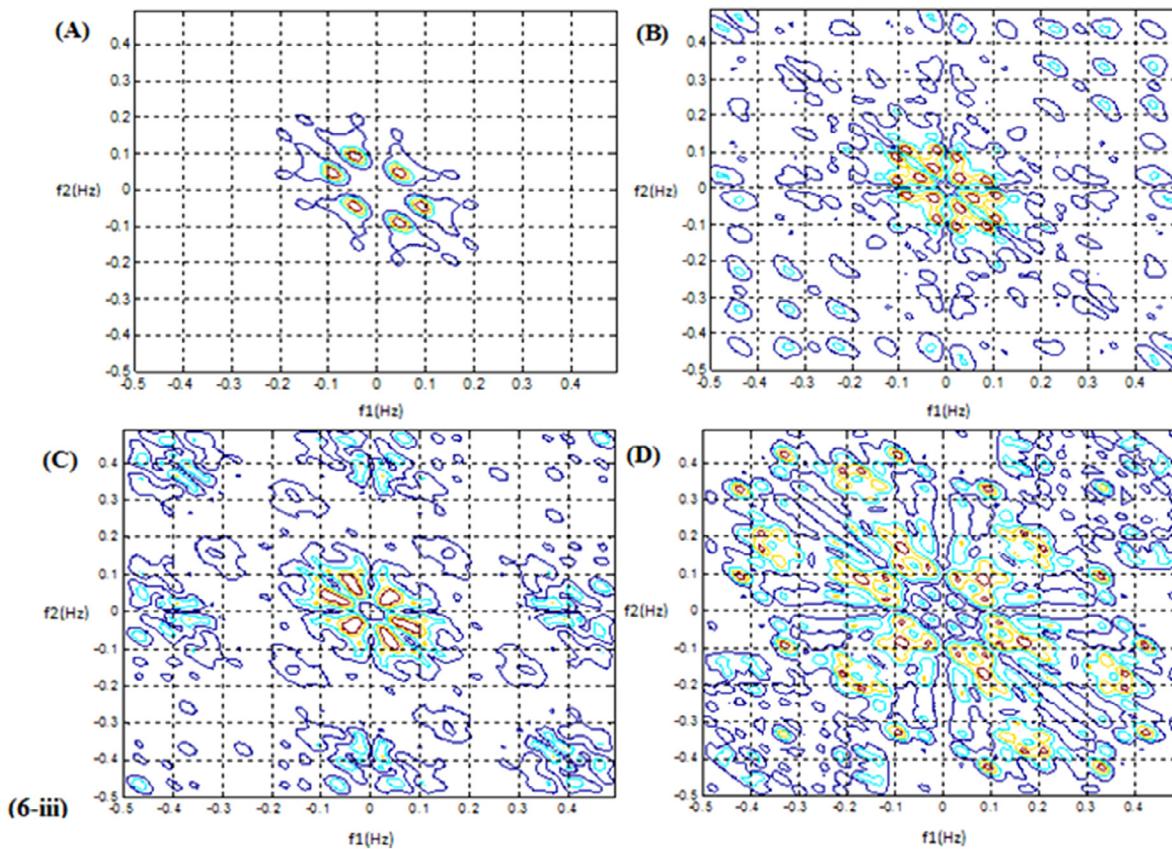


Fig. 6. (continued).

signal is completely different among the individuals in the various groups. This is the property to be extracted to categorize a subject as being a patient or a healthy. Furthermore, the results suggest that the proposed approach has a significant effect on the characteristics of speech signals in various levels of AD.

In this regard, the performance of the GP has been assessed by developing a CADx system for the prediction and early diagnosis of AD using two sets of speech data, namely, the Persian database (case study-I) and the English database (case study-II). To this end, GP was accurately run several times to investigate the effectiveness of different options provided by Eureqa Formulize software and the compound of mathematical operators on the generalization performance and prediction accuracy of the obtained models in each case study. In addition, GP-based models were developed using HOS-based features as inputs.

As mentioned above, all speech signals were pre-processed for the two mentioned databases. Then, bispectrum was computed as features for each speech segment via both direct (FFT-based) and parametric models using the MATLAB toolbox. For simplicity, the bispectrum calculated via the parametric approach was displayed by Bi-AR and the one calculated via the direct (FFT-based) method by Bi-FFT.

Eight features were extracted using Bi-AR and Bi-FFT methods for developing the GP-based models. To establish the *P*-value for each feature, a one-way analysis of variance (ANOVA) was performed to assess the applicability of the suggested features in differentiating the groups. According to the results, the most outstanding features were observed in different groups for both databases with the best potential discrimination at a *P* significant level of 0.001. *P*-values of the sixteen extracted features for both Bi-AR and Bi-FFT methods from the two databases are tabulated in Tables 1 and 2, respectively. As can be seen, the features extracted by both methods enjoy statistically remarkable discernment among different groups. Then, these features were fed into the GP-based models for evaluating the efficiency of bispectrum features and GP algorithms to identify HC groups and those with AD.

Table 1

Features and their corresponding *p* values for Bi-FFT and Bi-AR methods from the Persian database.

Features name	<i>P</i> value (Bi-AR)	<i>P</i> value (Bi-FFT)
B_{avg}	5.1423e-007	2.5464e-005
B_{max}	4.2157e-005	0.0056
B_{min}	0.0025	0.0027
A	6.5478e-007	5.2142e-006
P_{e1}	0.0065	3.8951e-004
P_{e2}	0.0011	0.0018
P_{e3}	5.0458e-005	0.0034
Ph_e	0.0087	8.0654e-004

Table 2

Features and their corresponding *p* values for Bi-FFT and Bi-AR methods from the English database.

Features name	<i>P</i> value (Bi-AR)	<i>P</i> value (Bi-FFT)
B_{avg}	2.5423e-005	3.1486e-004
B_{max}	5.3457e-006	0.0064
B_{min}	0.0038	4.2537e-006
A	5.4278e-006	6.3142e-007
P_{e1}	3.2751e-004	0.0058
P_{e2}	5.2317e-006	0.0023
P_{e3}	5.1158e-006	4.3275e-005
Ph_e	0.0077	7.0065e-005

In addition, in this research, it has been tried to enhance the performance of the proposed CADx system using the GP method to choose the most relevant attributes and afterward to conduct the final classification in accordance with a majority voting-based strategy. The optimal number of votes needed to put a subject into different stages of Alzheimer's and HC groups was selected by the prediction of the individual class by GP (described in Section 3.4).

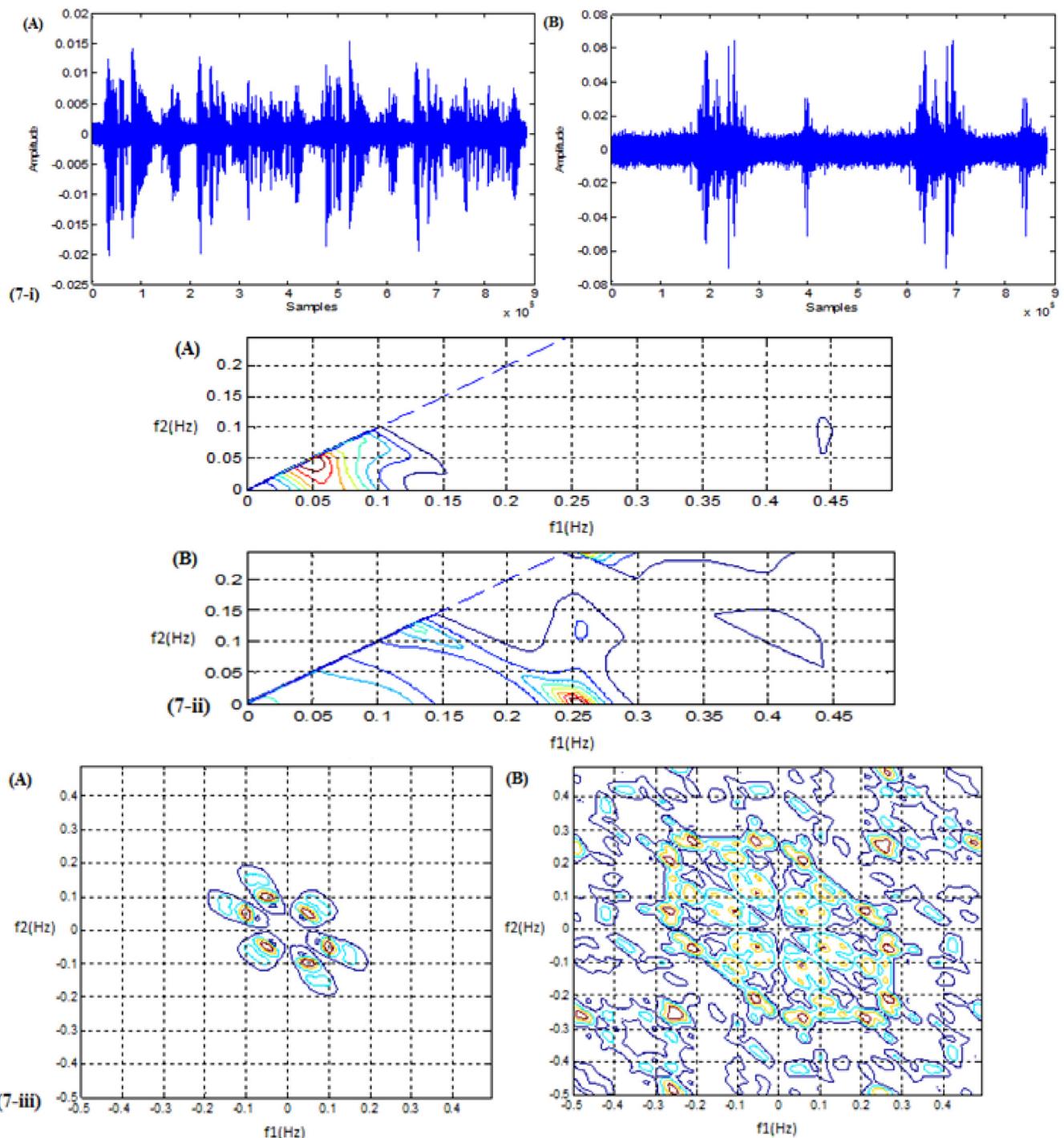


Fig. 7. Representative (7-i) spontaneous speech signals from English language database, (7-ii) bispectrum estimation using the parametric method, and (7-iii) bispectrum estimation using the FFT method belonging to (A) a healthy control subject and (B) an Alzheimer's patient, respectively. As shown in (7-i), it is apparent that the speech signals of a patient with Alzheimer's are poorer than those of a healthy subject. This is mainly because of a loss of language skills Alzheimer's patients face in difficulty speaking, comprehension, and communicating with the environment. Both (7-ii) and (7-iii) illustrate bi-spectral results obtained using both parametric and FFT methods, respectively, which provide a detailed understanding of the spontaneous speech distinctions between a healthy subject and an Alzheimer's patient. Note that, the colors depict the relative changes in the amplitude of bi-spectral values. . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.1. Case study-I: Evaluating of GP-based models for predicting AD using spontaneous speech signals from the persian database

We considered spontaneous speech signals of three stages of AD and HC groups. To evaluate the performance of our proposed method, a study is carried out on 500 segments of each group from the Persian database explained in Section 2.

To demonstrate the performance of GP-based models, 40% of the input patterns are randomly utilized as the training set, 30% of them as the validation set, while the residual 30% is utilized as the testing set. Training set is introduced to the GP-based models during training, and the GP is regulated relevant to its error. Validation set is utilized to estimate the generalization of GP-based models; when improving, the generalization stops, to halt training. Since training enjoys no gain from the test set, an independent estimate of the performance of

$$f = \exp\left(\frac{1.2}{(B_{min} + P_h^{B_{min}} - 1.2B_{max}^2)^{C_3} + \sin(0.56 + 1.2B_{min} + 1.2B_{max}B_{min} + \sin(1.5 + B_{max} + B_{min})^5)}\right) \quad (23)$$

Box I.

the GP-based models provides during and after the training (Hai-Jew, 2014). In this research, the validation set is utilized to avoid the overfitting problem. To evade over-fitting and find the best formula, it is essential to utilize several techniques such as cross-validation, early stopping, and regularization. The early stopping method is employed in this study. For avoiding over-fitting by using this technique, existing data are categorized into three subsections, namely training, test, and validation. During the training task, the error is supervised on the validation set. The validation error will usually reduce as much as the error in the training set during the primary stage of training. Nevertheless, when the GP-based models start to over-fit the data, the error will commonly start to increase on the validation set. The training is discontinued when, for a certain number of repetitions, the validation error has risen, and the mutation and crossover operators for the creation of the next generation members at the minimum of the validation error turned back (Ghugare et al., 2014; Hai-Jew, 2014).

In this case study, to evaluate the applicability of the obtained models for AD prediction, several GP experiments were conducted by careful investigation of the effects of different options. In this process, the eight features described above make up the input space of the GP-based models. Note that, the best models yielded by the algorithm do not require that it necessarily includes all the eight features. In other words, the resultant model is a subset of all eight features which offer high predictive accuracy of Alzheimer's and generalizability. In fact, the GP algorithm is conducting feature selection tasks as part of its optimization process. Therefore, in the current research, the two GP-based SR models with high-performance employed a subset of eight features as inputs, and the corresponding equations are given as:

1. Two features (A, Ph_e) for Bi-FFT based model

$$f = 1.16 + 7.98e^{-13H^2} + \frac{0.0047}{\sin(1.14 + 7.98e^{-13H^2})} - 7.28e^{-7H} - 2.41e^{-26H^4} - 1.93e^{-28H^2}P_h^2 \quad (22)$$

2. Three features (B_{max}, B_{min}, Ph_e) for Bi-AR based model (see Eq. (23) which is given in Box I)

It can be inferred that Eqs. (22) and (23) as two indicator models produced through the GP, utilizing a subset of the eight features in the statistically significant differences among the four groups as inputs (see Table 1). In fact, it demonstrates one of the most significant GP-based SR capabilities, which lead to the use of the top powerful features in the input space. Table 3 presents the comparison where the statistical quantities and a subset of the features utilized for each model are listed. As can be seen, the values of the R^2GF related to the prediction of AD, which is generated by the two indicator models, are rise at around (≈ 0.95) for all three datasets, including training, test, and validation. In addition, the values of MSE and MAE are down. Especially, the MSE values associated with predicting two GP-based models are constantly below 1%. The small and comparable MSE and MAE values, and large and comparable R^2GF values in connection with the generated predictions clearly demonstrate that the two models separately contain an excellent predictive accuracy of Alzheimer's and generalizability.

The values of each of the three measurements (R^2GF, MSE , and MAE) are comparable for both GP-based models. As can be seen, the Bi-FFT based features present a kind of better performance than the Bi-AR based features.

Fig. 8 shows the two-dimensional t-SNE embedding of the features selected by the GP algorithm, with visually severable clusters. Each data point indicates an individual. The assignment of the clusters is

Table 3
The prediction and generalization performance of the GP-based models for AD.

GP-model	No. of selected features	R^2GF	MSE	MAE	Accuracy (%)
Bi-FFT based feature	2	0.99	0.013	0.057	98.7
Bi-AR based feature	3	0.96	0.068	0.137	94

reflected by the colors. The similarity of individuals to each other is shown through the distance between the points, which is measured according to the proportion of their subject. In the figure, it is feasible to visualize how the data sample could be separated into four clear clusters, belonging to healthy individuals and three different stages of Alzheimer's. This verifies the importance of the bispectrum based features in distinguishing healthy subjects and various stages of AD. In addition, observing these features visually may assist us to grasp some new aspects of the signals regarding the normal and abnormal cases, which could be described more explicitly. Furthermore, these visual representations allow us to employ the GP algorithm as a valuable tool for the early diagnosis of AD.

In addition, this study aims at intensifying the detection rate improvement of the proposed CADx system based on the GP algorithm. It is worth noting that model validation is a significant section of each modeling procedure. As mentioned above, GP is based on a stochastic optimization process, and so it is that frequent implementation of GP will afford different solutions. In the process of any optimization, the convergence of the solutions is not expected to be the same even after simplification. The resultant solutions seem different, and also the fitness value some of them enjoy better than others. In general, each solution includes a set of terms, and in most solutions occurs many of these terms. Furthermore, prevalent terms often constitute the major factors of the original expression. In this spirit, examining several optimizations separately reveals the possible components of the proposed system. Thus, the results of model validation can be utilized to refine the GP algorithm, as an essential step in the process of developing the iterative model. Hence, it would be expected that using a technique to validate the proposed model would improve the results. Improving the detection rate is consistent with the fact that to create a suitable model, it is necessary to analyze a set of several runs. As a result, all the training data points to be properly categorized. In other words, any training sample that is misclassified possesses a very less penalty in this case. Therefore, the enhancement of detection rate can be considered as the outcome of the more robust model in the solution space. In order to achieve the best results, in this research, we provide a novel scheme for GP regarding a concept, called decision making based on majority voting, as described in Section 3.4.

Table 4 shows the results of applying a majority voting-based decision to the GP using the Bi-FFT and Bi-AR based features, respectively. It can be seen that the majority voting process has resulted in the enhancement of the diagnosis rates. These models are expressed as follows:

1. One feature (A) for Bi-FFT based model

$$f = 2.6 + 2.82e^{-7H} + \frac{2.37e^5}{H - 3.33e^6} + 3.63e^{-13H}^{2\sin(0.887 + \sin(7.89e^{-13A^2}))} - 1.98 \sin(0.887 + \sin(7.83e^{-13H^2})) \quad (24)$$

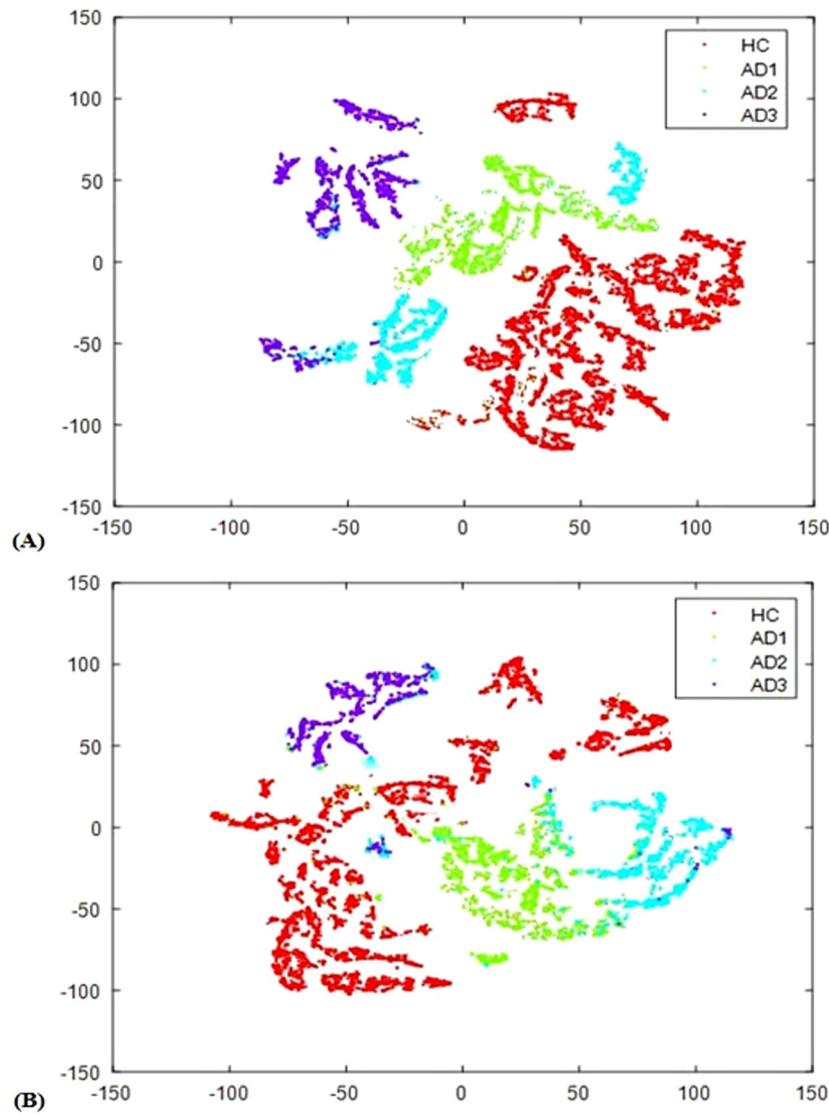


Fig. 8. Visualization results of t-SNE embedding for Persian language database: (A) two selected features using GP algorithm on the basis of Bi-FFT based feature. (B) three selected features using GP algorithm on the basis of Bi-AR based feature. As can be seen, there is a clear distinction between healthy clusters and the three stages of Alzheimer's. This confirms the importance of selected features using GP-based SR models, which can provide sufficient information for distinguishing among the four classes. . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. Three features (B_{max} , B_{min} , P_{e1}) for Bi-AR based model

$$f = 1.32 + B_{max}^5 - 7.98B_{min} - 1.42 \cos(7.28B_{max}^{(-1.05B_{min})})^{(P_1^2)} \\ \times \cos(0.23 + 1.93B_{max}^4 - 2.09B_{min}) \quad (25)$$

The results indicate that the proposed CADx system produces the highest accuracy according to a majority voting GP-based models than the GP-based models used alone. It is also obvious that, constructed model according to a majority voting-based decision leads to the utilization of a more relevant group of feature vectors, which would be capable of increasing the detection rates. The utilized subset of features and the statistical quantities for each model are also listed in the table. As shown, the performance on the Bi-AR based features is outperformed by the Bi-FFT based features in terms of the AD predictions. In addition, the Bi-FFT based features yield an MSE of 0.009, an R^2GF 0.99, an MAE 0.04, and an Accuracy of 99.09%. On the other hand, the Bi-AR based features result in an MSE of 0.047, an R^2GF 0.97, an MAE 0.15, and an Accuracy of 96%. In addition, the majority voting procedure provides top performance for AD prediction, as shown in

Table 4

The prediction and generalization performance of the majority vote GP-based models for AD.

GP-model	No. of selected features	R^2GF	MSE	MAE	Accuracy (%)
Bi-FFT based feature	1	0.99	0.009	0.04	99.09
Bi-AR based feature	3	0.97	0.047	0.15	96

Table 4. Furthermore, according to Tables 3 and 4, comparing with GP-based models, the majority voting GP-based model can reduce the prediction errors and improve the generalization performance of the CADx system. For example, MSE was improved from 0.013 to 0.009, and MAE was improved from 0.057 to 0.04 in Bi-FFT based features. However, R^2GF did not change.

In our previous work (Nasrolahzadeh et al., 2018), the performance of the proposed CADx systems was compared by using four classifiers, namely SVM, DT, KNN, and NB, in classifying healthy and AD subjects. In this work, we used Bi-FFT and Bi-AR based features to extract discriminative information from the spontaneous speech signals of people with and without Alzheimer's. For more information about how

Table 5

Classification results obtained using four classifiers vs. GP-based models based on Bi-FFT features.

Classifiers	SVM	DT	KNN	NB	GP-based models (Proposed method)
Accuracy (%)	92.08	95.42	93.75	95	99.09

Table 6

Classification results obtained using four classifiers vs. GP-based models based on Bi-AR features.

Classifiers	SVM	DT	KNN	NB	GP-based models (Proposed method)
Accuracy (%)	96.46	94.18	97.71	95.94	96

to assess these classifiers to consider the robustness of the HOS-based features, see our previous work (Nasrolahzadeh et al., 2018). Tables 5 and 6 present the comparison among the four mentioned classifiers in our previous work and GP-based models from the viewpoints of the classification results using Bi-FFT and Bi-AR based features. Since the database and utilized features extraction method is the same for both studies, our proposed CADx system is comparable with our previous reported system from the viewpoint of diagnostic performance. As can be seen, the proposed method yielded better results than the other four classifiers used in Nasrolahzadeh et al. (2018).

4.2. Case study-II: Evaluating of GP-based models for predicting AD using spontaneous speech signals from the English database

Some 473 narrative samples of English data pertaining to 97 HC subjects and 167 AD patients were used in the GP-based modeling. In order to implement the proposed scheme, all data were divided at random into the training, test, and validation sets. 40% of the narrative samples were allocated to train, 30% of them were dedicated to validation, and the other 30% were aside for the testing set.

To find the suitable models, the proposed GP was trained utilizing the training set. As described earlier, the early stopping method was used in this paper so that over-training was avoided. In such a procedure, the training method would be able to stop to avoid over-training when enhancement of the accuracy on the validation set was stopped. Moreover, the mutation and crossover operators for the creation of the next generation members are being altered until the best topology, which yields the best accuracy and the minimum amount of the validation error, is found. Not to mention the fact that to generate all the next generation, GP not only applies the mutation operator but also employs crossover to create the rest of the population. In this process, a couple of chosen chromosomes swap their information by the crossover operator. The chromosomes were more probably to remain alive, which were enjoyed higher ranks of fitness. In the number of generations of GP, different from actual biological communities, remaining alive for an individual is feasible. This means that the genetic characteristics of such elites were passed on to the newborn population unchanged. In this way, the search is cyclically run in the high-dimensional input space until a stop criterion is met (Ghugare et al., 2014; Hai-Jew, 2014). After finding the best model, the testing samples as unseen data were fed into the GP model, and its performance was evaluated in terms of three statistical measurements, i.e., *MSE*, *R²GF*, and *MAE*.

To implement the GP-based Alzheimer's predictive models, the eight features previously described are employed as input space. The two GP-based SR models with high-performance yielded employing the Eureqa Formulize package are expressed as follow:

1. Two features (B_{max} , Ph_e) for Bi-FFT based model

$$f = 1.63e^{-2B_{max}} + \frac{2.06e^{-5Ph}}{\sin(-3.02Ph)} + 0.56P_h^4 + 1.32B_{max} \sin(14.1P_h^4) - 2.1 \sin(1.5B_{max}^4) \quad (26)$$

Table 7

The prediction and generalization performance of the GP-based models for AD.

GP-model	No. of selected features	<i>R²GF</i>	<i>MSE</i>	<i>MAE</i>	Accuracy (%)
Bi-FFT based feature	2	0.95	0.072	0.16	93
Bi-AR based feature	4	0.89	0.112	0.189	89

2. Four features (B_{avg} , B_{min} , P_{e1} , Ph_e) for Bi-AR based model

$$f = 1.65B_{avg} + 14.95e^{-6B_{min}P_2} + 91.53e^{\frac{-5}{P_2}} \cos(98.7P_h^2) + 34.9e^{-5B_{min}^2} - 1.57P_h \sin(1.9P_h^{-8}) - 121.3B_{avg}^2 \quad (27)$$

From the above equations, it can be observed that based on characteristics derived from bispectrum, the obtained GP models are non-linear. It is also noticed that in this case study, the GP-based model has utilized the two features derived from the Bi-FFT based model, namely B_{max} and Ph_e . Besides, the three features derived from Bi-AR based model include B_{avg} , B_{min} , P_{e2} , and Ph_e made the other model (see Table 3). The presented results demonstrate the ability of GP-based SR to use the most influential features in model construction.

Table 7 shows the *R²GF*, *MSE*, and *MAE* values achieved by the two GP-based models in this case study. According to this Table, the two-descriptor model possesses a higher index of prediction accuracy. However, both models enjoy generalizable. The expressed model for Alzheimer's prediction has been obtained in high *R²GF* (≈ 0.90) based on three datasets, including training, validation, and test. The corresponding *MSE* and *MAE* values were also low 1%. In addition, the small values of *MSE* and *MAE*, and the large amounts of *R²GF* associated with the yielded predictions clearly indicate that the two models separately have excellent Alzheimer's predictive accuracy and generalizability. The results indicate that the GP-model predicted AD values of the spontaneous speech signals can reveal valuable information elicited from bispectral and may suggest new insights regarding the origin of the formation of such abnormalities of speech in people with Alzheimer's disease.

Moreover, the results reveal that the performance of the Bi-FFT based features outperforms the Bi-AR based features in terms of predictive accuracy of constructed models for the two mentioned groups in this case study.

We have also investigated the visualization of the two and four features utilized to make the GP model using t-SNE (Fig. 9). The visualization results of t-SNE generated distinct clustering for subjects with and without Alzheimer's. It can be inferred that bispectrum based features selected by GP enjoy an important role in discriminating HC subjects and patients with AD. This is why the GP-based SR model provides us a good choice for the early diagnosis of Alzheimer's.

In this case study, our proposed approach was evaluated using a majority voting. The results of the obtained models using the majority voting algorithm are shown in Table 8. The two models obtained enjoy the best performance, which is based on Bi-FFT and Bi-AR based features. These models are given as follows:

1. One feature (B_{max}) for Bi-FFT based model

$$f = 0.99 + 2.7e^{-13B_{max}^2} - (1.64e^{-5} + 1215.34B_{max}^2) - 2.2e^{-7B_{max}} \sin(1.96e^{-12B_{max}^2}) \quad (28)$$

2. Three features (B_{avg} , B_{min} , Ph_e) for Bi-AR based model

$$f = \frac{11.57P_h}{B_{avg}} + 61.53e^{-6B_{min}} + \cos(72.95B_{min}) + 0.55e^{-5} \left(\frac{40.33B_{min}}{(\cos(-96.01B_{avg}) + 1.4P_h^2)} \right) - 3.02e^{-6P_h B_{min}} \quad (29)$$

It can be inferred from Eqs. (28) and (29) that the proposed CADx system generates the highest accuracy according to a majority voting.

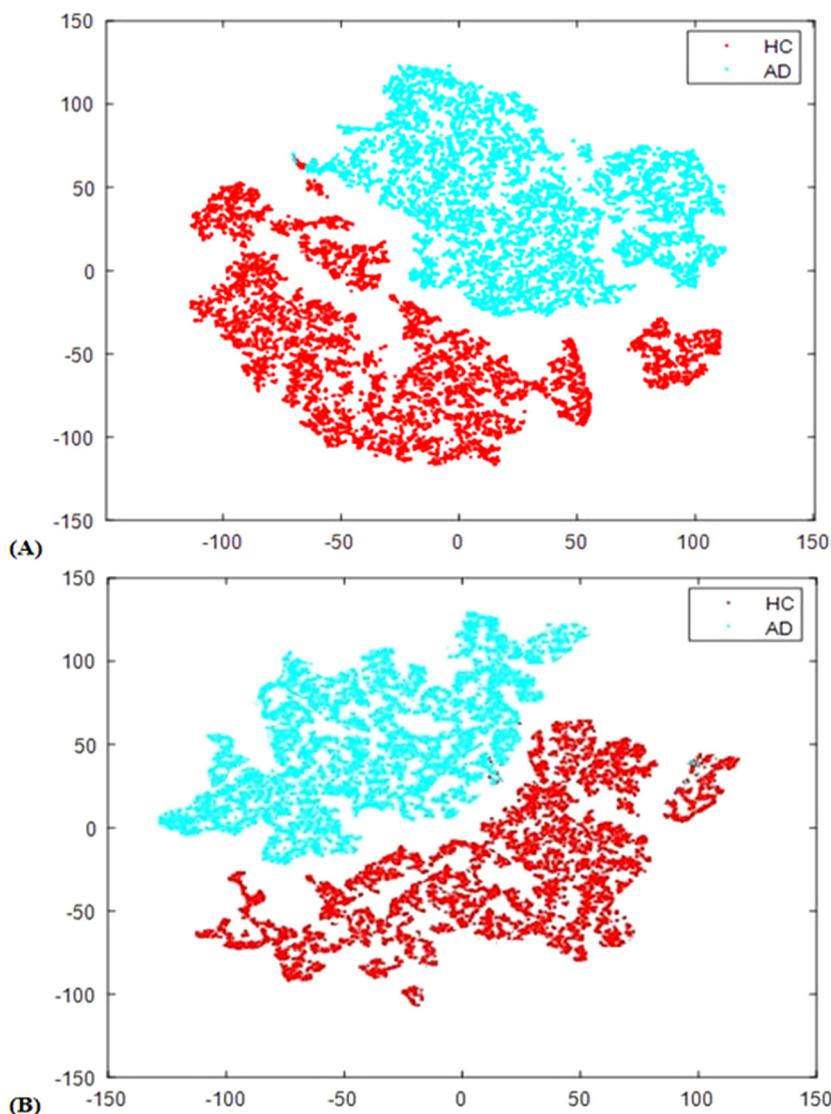


Fig. 9. Visualization results of t-SNE embedding for English language database: (A) two selected features using GP algorithm on the basis of Bi-FFT based feature. (B) four selected features using GP algorithm on the basis of Bi-AR based feature. It can be seen that there is a clear distinction between healthy clusters and those with Alzheimer's. The t-SNE results confirm the significance of selected features using GP-based SR models, providing valuable information for distinguishing between the two classes.

Table 8

The prediction and generalization performance of the majority vote GP-based models for AD.

GP-model	No. of selected features	R^2GF	MSE	MAE	Accuracy (%)
Bi-FFT based feature	1	0.96	0.038	0.148	96
Bi-AR based feature	3	0.91	0.078	0.173	92

The number of utilized features and the corresponding statistical values are tabulated in Table 8 for each model. From this table, it is observed that the GP-based models in terms of majority voting corresponding to the AD predictions yielded by the most relevant features. As can be seen, Bi-FFT based features perform better than Bi-AR based features in terms of Alzheimer's prediction. Table 8 provides a comparison of statistical quantities according to AD predictions performed by the two models. According to this table, the Bi-FFT based features yield an MSE of 0.038, an R^2GF 0.96, an MAE 0.148, and an Accuracy of 96%. Moreover, the Bi-AR based features result in an MSE of 0.078, an R^2GF 0.91, an MAE 0.173, and an Accuracy of 92%.

It is worth noting that the majority voting method also provides high performance for predicting AD in this case study. According to Tables 7 and 8, the models obtained by using majority vote according to the Alzheimer's predictions can reduce prediction errors and improve generalized performance. For instance, MSE was improved from 0.072 to 0.038, and MAE was improved from 0.16 to 0.148. Also, R^2GF was improved from 0.95 to 0.96 in Bi-FFT based features.

Additionally, Table 9 compares our proposed method and the other CADx algorithms from the viewpoint of diagnostic performance briefly. Although our results do not indicate improvement in terms of predictive accuracy in the second case study compared to the first case study, overall, in the current study, fewer features are employed compared to other studies, which is a superiority of this study. Moreover, we have tried to evaluate those systems that mainly focused on accurately measuring some fundamental parameters as the descriptors of quality and feature variation in Alzheimer's speech. Note that the fundamental parameters are usually used along with other non-linear features such as Fractal dimension and Lyapunov exponents because of their great potential for AD diagnosis; hence the CADx systems that only utilize such features are considerably rare. Furthermore, it is a difficult task to compare systems accurately because various studies have used different

Table 9

Comparison of the proposed CADx systems with other CADx systems.

Reference	Database	Description and Number of features	Number of classes	Accuracy (%)
Our proposed method (Case study-I)	Private	HOS-based features	4	99.09
Our proposed method (Case study-II)	General	HOS-based features	2	96
Nasrolahzadeh et al. (2020)	Private	Energy and entropy features	2	98.33
Nasrolahzadeh et al. (2018)	Private	HOS-based features	4	97.71
Nasrolahzadeh et al. (2014)	Private	Acoustic and the voice quality features/ more than 100	4	97.96
López-de-Ipiña et al. (2013)	Private	Acoustic and the voice quality and duration features + Fractal dimension/more than 100 feature	4	97.7
König et al. (2015)	Private	Vocal markers/6	3	81
Bucks et al. (2000)	Private	Linguistic features including part-of-speech (POS) tag frequencies and measures of lexical diversity/8	2	87.5
Thomas et al. (2005)	General	Character n-gram-based techniques	2	94.5
Guinn and Habash (2012)	Private	POS tags and measures of lexical diversity	2	79.5
Meilán et al. (2014)	Private	Temporal and acoustic features/5	2	84.8
Jarrold et al. (2014)	Private	Acoustic features, POS features, and psychologically-motivated word	2	88
Rentoumi et al. (2014)	General	Frequency unigrams and excluded binary unigrams, syntactic complexity features, measures of vocabulary richness, and information theoretic features	2	75
Fraser et al. (2016)	General	Linguistic variables	2	81
Hernández-Domínguez et al. (2018)	General	Combination of all linguistic and information coverage features	2	94

databases. According to Table 9, the proposed method yields better results compared to other state-of-the-art methods. As shown, the GP-based models employ fewer features as pattern descriptors represent the subtle changes in signals, which are comparable with other reported systems. In addition, its accuracy is significantly higher than those reported in the original paper (Nasrolahzadeh et al., 2018). Thus, we can assert that the present study obtained a more effective performance than the other study. Moreover, the proposed method may be helpful to the automatic diagnosis and analysis of other clinical pathologies that affect the language skills of individuals with disorders such as Parkinson's disease.

5. Conclusion

This paper presented a new CADx system based on spontaneous speech signals intending to achieve optimal prediction accuracy and good generalization capability for the early diagnosis of AD. For that, it was analyzed and compared on both Persian and English language databases in depth. The genetic programming (GP) was utilized as advanced artificial intelligence-based data-driven formalistic modeling to develop the AD prediction using HOSs as the descriptors of the properties of human speech. The GP-based data-driven modeling formalism was applied to the HOS feature vector pool to find the optimal model to predict and describe AD based on monitored input-response data. Then, a majority voting-based strategy was employed to determine the best model for the final prediction of AD. The results suggest that the best model was yielded when the GP algorithm based on the majority voting was used to predict the AD values. The proposed method using GP-based majority voting with Bi-FFT based features managed to discriminate the subjects in the four classes with an accuracy index of 99.09% for the Persian language database and high accuracy of 96% in discriminating between the subjects in the two classes for the English language database as well. Compared to other studies that also utilized the DementiaBank database, our proposed method outperformed better results. The results suggest that the performance of models built using only the GP algorithm was poorer than those built using the GP-based on majority voting. Moreover, our findings revealed that the suggested method could offer a superior AD prediction and generalization performance; thus, this method could be a good option to diagnose AD early. In future work, the authors plan to extend the

proposed CADx system to an Alzheimer's prediction system where it would be required to perform appropriate modifications to identify the person in pre-clinical (i.e., before earliest symptoms) and prognostic (i.e., some symptoms very early but without dementia) stages.

CRediT authorship contribution statement

Mahda Nasrolahzadeh: Conceived of the presented idea, Carried out the experiments, Wrote the manuscript, Discussed the results and contributed to the final manuscript. **Shahryar Rahnamayan:** Conceived of the presented idea, Verified the analytical methods, Wrote the manuscript, Discussed the results and contributed to the final manuscript. **Javad Haddadnia:** Conceived of the presented idea, Verified the analytical methods, Wrote the manuscript, Discussed the results and contributed to the final manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

At first, special gratitude to Prof. Haddadnia, professor at Hakim Sabzevari University, Sabzevar, Iran, for supervising the implementation of this project. Thanks to Ontario Tech University for making it possible for Dr. Nasrolahzadeh as a Visiting Research Scholar to study at that university. Also, special thanks to Prof. Rahnamayan, professor at Ontario Tech University, Oshawa, Canada, for support in using NICI Lab's high performance computing resources. Moreover, appreciate to Prof. Haddadnia and Prof. Rahnamayan for their invaluable feedback without which this work would not have been possible. Finally, thanks to all authors who contributed to this research. The contributions of this paper are as follows. Prof. Haddadnia, Prof. Rahnamayan, and Dr. Nasrolahzadeh designed the whole study together. Dr. Nasrolahzadeh carried out the experiments. Prof. Haddadnia and Prof. Rahnamayan verified the results of the analytical method. All authors discussed the results and contributed to the final manuscript. Dr. Nasrolahzadeh wrote the whole paper with support from Prof. Rahnamayan and Prof. Haddadnia.

References

- Acharya, R., Chua, C. K., Ng, E. Y. K., Yu, W., & Chee, C. (2008). Application of higher order spectra for the identification of diabetes retinopathy stages. *Journal of Medical Systems*, 32(6), 481–488. <http://dx.doi.org/10.1007/s10916-008-9154-8>.
- Acharya, U. R., Sudarshan, V. K., Koh, J. E., Martis, R. J., Tan, J. H., Oh ..., S. L., & San Tan, R. (2017). Application of higher-order spectra for the characterization of coronary artery disease using electrocardiogram signals. *Biomedical Signal Processing and Control*, 31, 31–43. <http://dx.doi.org/10.1016/j.bspc.2016.07.003>.
- Al-Sahaf, H., Bi, Y., Chen, Q., Lensen, A., Mei, Y., Sun, Y., & Zhang, M. (2019). A survey on evolutionary machine learning. *Journal of the Royal Society of New Zealand*, 49(2), 205–228. <http://dx.doi.org/10.1080/03036758.2019.1609052>.
- Baltrušaitis, T., Ahuja, C., & Morency, L. P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423–443. <http://dx.doi.org/10.1109/TPAMI.2018.2798607>.
- Barmpalexis, P., Kachrimanis, K., Tsakonas, A., & Georgarakis, E. (2011). Symbolic regression via genetic programming in the optimization of a controlled release pharmaceutical formulation. *Chemometrics and Intelligent Laboratory Systems*, 107(1), 75–82. <http://dx.doi.org/10.1016/j.chemolab.2011.01.012>.
- Becker, J. T., Boiler, F., Lopez, O. L., Saxton, J., & McGonigle, K. L. (1994). The natural history of alzheimer's disease: description of study cohort and accuracy of diagnosis. *Archives of Neurology*, 51(6), 585–594. <http://dx.doi.org/10.1001/archneur.1994.00540180063015>.
- Bhardwaj, A., Tiwari, A., Krishna, R., & Varma, V. (2016). A novel genetic programming approach for epileptic seizure detection. *Computer Methods and Programs in Biomedicine*, 124, 2–18. <http://dx.doi.org/10.1016/j.cmpb.2015.10.001>.
- Bi, Y., Xue, B., & Zhang, M. (2020). Genetic programming with a new representation to automatically learn features and evolve ensembles for image classification. *IEEE Transactions on Cybernetics*, 51(4), 1769–1783. <http://dx.doi.org/10.1109/TCYB.2020.2964566>.
- Bi, Y., Xue, B., & Zhang, M. (2021a). A divide-and-conquer genetic programming algorithm with ensembles for image classification. *IEEE Transactions on Evolutionary Computation*, <http://dx.doi.org/10.1109/TEVC.2021.3082112>.
- Bi, Y., Xue, B., & Zhang, M. (2021b). Dual-tree genetic programming for few-shot image classification. *IEEE Transactions on Evolutionary Computation*, <http://dx.doi.org/10.1109/TEVC.2021.3100576>.
- Bi, Y., Xue, B., & Zhang, M. (2021c). *Genetic programming for image classification: an automated approach to feature learning* (Vol. 24). Springer Nature, doi: 10.1007978-3-030-65927-1.
- Bucks, R. S., Singh, S., Cuerden, J. M., & Wilcock, G. K. (2000). Analysis of spontaneous, conversational speech in dementia of alzheimer type: Evaluation of an objective technique for analysing lexical performance. *Aphasiology*, 14(1), 71–91. <http://dx.doi.org/10.1080/026870300401603>.
- Chaudhari, P., Agrawal, H., & Kotecha, K. (2019). Data augmentation using MG-GAN for improved cancer classification on gene expression data. *Soft Computing*, 1–11. <http://dx.doi.org/10.1007/s00500-019-04602-2>.
- Childers, D. G. (Ed.), (1978). *Modern spectrum analysis*. IEEE Computer Society Press.
- Chua, K. C., Chandran, V., Acharya, U. R., & Lim, C. M. (2008). Cardiac state diagnosis using higher order spectra of heart rate variability. *Journal of Medical Engineering & Technology*, 32(2), 145–155. <http://dx.doi.org/10.1080/03091900601050862>.
- Chua, K. C., Chandran, V., Acharya, U. R., & Lim, C. M. (2010). Application of higher order statistics/spectra in biomedical signals—A review. *Medical Engineering & Physics*, 32(7), 679–689. <http://dx.doi.org/10.1016/j.medengphy.2010.04.009>.
- Emigdio, Z., Abatal, M., Bassam, A., Trujillo, L., Juárez-Smith, P., & El Hamzaoui, Y. (2017). Modeling the adsorption of phenols and nitrophenols by activated carbon using genetic programming. *Journal of Cleaner Production*, 161, 860–870. <http://dx.doi.org/10.1016/j.jclepro.2017.05.192>.
- Fraser, K. C., Meltzer, J. A., & Rudzicz, F. (2016). Linguistic features identify alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease*, 49(2), 407–422. <http://dx.doi.org/10.3233/jad-150520>.
- Ghugare, S. B., Tiwary, S., Elangovan, V., & Tambe, S. S. (2014). Prediction of higher heating value of solid biomass fuels using artificial intelligence formalisms. *BioEnergy Research*, 7(2), 681–692. <http://dx.doi.org/10.1007/s12155-013-9393-5>.
- Gorji, H. T., Khoei, T. T., & Kaabouch, N. (2020). Biomarkers selection toward early detection of alzheimer's disease. In *2020 ieee international conference on electro information technology (eit)* (pp. 487–494). IEEE, <http://dx.doi.org/10.1109/EIT4899.2020.9208258>.
- Gray, G. J., Murray-Smith, D. J., Li, Y., Sharman, K. C., & Weinbrenner, T. (1998). Non-linear model structure identification using genetic programming. *Control Engineering Practice*, 6(11), 1341–1352. [http://dx.doi.org/10.1016/S0967-0661\(98\)00087-2](http://dx.doi.org/10.1016/S0967-0661(98)00087-2).
- Guinn, C. I., & Habash, A. (2012). Language analysis of speakers with dementia of the alzheimer's type. In *2012 aaai fall symposium series*.
- Hai-Jew, S. (Ed.), (2014). *Enhancing qualitative and mixed methods research with technology*. IGI Global.
- Hernández-Domínguez, L., Ratté, S., Sierra-Martínez, G., & Roche-Bergua, A. (2018). Computer-based evaluation of alzheimer's disease and mild cognitive impairment patients during a picture description task. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10, 260–268. <http://dx.doi.org/10.1016/j.adm.2018.02.004>.
- Hu, W. T., McMillan, C., Libon, D., Leight, S., Forman, M., Lee, V. Y., & Grossman, M. (2010). Multimodal predictors for alzheimer disease in nonfluent primary progressive aphasia. *Neurology*, 75(7), 595–602. <http://dx.doi.org/10.1212/WNL.0b013e3181ed9c52>.
- López-de Ipiña, K., Egiraun, H., Sole-Casals, J., Ecay, M., Ezeiza, A., Barroso, N., & Martinez-de Lizárdiz, U. (2013). Feature extraction approach based on fractal dimension for spontaneous speech modelling oriented to alzheimer disease diagnosis. In *International conference on nonlinear speech processing* (pp. 144–151). Berlin, Heidelberg: Springer, http://dx.doi.org/10.1007/978-3-642-38847-7_19.
- Jack, C. R., Jr., & Holtzman, D. M. (2013). Biomarker modeling of alzheimer's disease. *Neuron*, 80(6), 1347–1358. <http://dx.doi.org/10.1016/j.neuron.2013.12.003>.
- Jarrold, W., Peintner, B., Wilkins, D., Vergry, D., Richey, C., Gorno-Tempini, M. L., & Ogar, J. (2014). Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. In *Proceedings of the workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality* (pp. 27–37). <http://dx.doi.org/10.3115/v1/W14-3204>.
- Joshi, G., Walambe, R., & Kotecha, K. (2021). A review on explainability in multimodal deep neural nets. *IEEE Access*, <http://dx.doi.org/10.1109/ACCESS.2021.3070212>.
- Kalkreuth, R., Rudolph, G., & Krone, J. (2015). Improving convergence in cartesian genetic programming using adaptive crossover, mutation and selection. In *2015 ieee symposium series on computational intelligence* (pp. 1415–1422). IEEE, <http://dx.doi.org/10.1109/SSCI.2015.201>.
- König, A., Satt, A., Sorin, A., Hoory, R., Toledo-Ronen, O., Derreumaux, A., & David, R. (2015). Automatic speech analysis for the assessment of patients with predementia and alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 1(1), 112–124. <http://dx.doi.org/10.1016/j.adm.2014.11.012>.
- Koza, J. R., & Koza, J. R. (1992). *Genetic programming: on the programming of computers by means of natural selection* (Vol. 1). MIT Press.
- Langdon, W. B., & Poli, R. (2013). *Foundations of genetic programming*. Springer Science & Business Media.
- Li, Z., Wu, Z., He, Y., & Fulei, C. (2005). Hidden Markov model-based fault diagnostics method in speed-up and speed-down process for rotating machinery. *Mechanical Systems and Signal Processing*, 19(2), 329–339. <http://dx.doi.org/10.1016/j.jmssp.2004.01.001>.
- Lones, M. A., Smith, S. L., Alty, J. E., Lacy, S. E., Possin, K. L., Jamieson, D. S., & Tyrrell, A. M. (2013). Evolving classifiers to recognize the movement characteristics of parkinson's disease patients. *IEEE Transactions on Evolutionary Computation*, 18(4), 559–576. <http://dx.doi.org/10.1109/TEVC.2013.2281532>.
- MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). Aphasiabank: Methods for studying discourse. *Aphasiology*, 25(11), 1286–1307. <http://dx.doi.org/10.1080/02687038.2011.589893>.
- Mahdavi, S., Rahnamayan, S., & Mahdavi, A. (2019). Majority voting for discrete population-based optimization algorithms. *Soft Computing*, 23(1), 1–18. <http://dx.doi.org/10.1007/s00500-018-3530-1>.
- Martis, R. J., Acharya, U. R., Mandana, K. M., Ray, A. K., & Chakraborty, C. (2013). Cardiac decision making using higher order spectra. *Biomedical Signal Processing and Control*, 8(2), 193–203. <http://dx.doi.org/10.1016/j.bspc.2012.08.004>.
- McKeith, I. G., Galasko, D., Kosaka, K., Perry, E. K., Dickson, D. W., Hansen, L. A., & Perry, R. H. (1996). Consensus guidelines for the clinical and pathologic diagnosis of dementia with lewy bodies (DLB): report of the consortium on DLB international workshop. *Neurology*, 47(5), 1113–1124. <http://dx.doi.org/10.1212/wnl.47.5.1113>.
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack Jr, C. R., Kawas, C. H., & Phelps, C. H. (2011). The diagnosis of dementia due to alzheimer's disease: recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & Dementia*, 7(3), 263–269. <http://dx.doi.org/10.1016/j.jalz.2011.03.005>.
- Meilán, J. J. G., Martínez-Sánchez, F., Carro, J., López, D. E., Millán-Morell, L., & Arana, J. M. (2014). Speech in alzheimer's disease: Can temporal and acoustic parameters discriminate dementia? *Dementia and Geriatric Cognitive Disorders*, 37(5–6), 327–334. <http://dx.doi.org/10.1159/000356726>.
- Mendel, J. M. (1991). Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications. *Proceedings of the IEEE*, 79(3), 278–305. <http://dx.doi.org/10.1109/5.75086>.
- Mookiah, M. R. K., Acharya, U. R., Lim, C. M., Petznick, A., & Suri, J. S. (2012). Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features. *Knowledge-Based Systems*, 33, 73–82. <http://dx.doi.org/10.1016/j.knosys.2012.02.010>.
- Nasrolahzadeh, M., Haddadnia, J., & Rahnamayan, S. (2020). Multi-objective optimization of wavelet-packet-based features in pathological diagnosis of alzheimer using spontaneous speech signals. *IEEE Access*, 8, Article 112393-112406. <http://dx.doi.org/10.1109/ACCESS.2020.3001426>.
- Nasrolahzadeh, M., Mohammadpoori, Z., & Haddadnia, J. (2016a). Analysis of mean square error surface and its corresponding contour plots of spontaneous speech signals in alzheimer's disease with adaptive wiener filter. *Computers in Human Behavior*, 61, 364–371. <http://dx.doi.org/10.1016/j.chb.2016.03.031>.
- Nasrolahzadeh, M., Mohammadpoori, Z., & Haddadnia, J. (2016b). A novel method for early diagnosis of alzheimer's disease based on higher-order spectral estimation of spontaneous speech signals. *Cognitive Neurodynamics*, 10(6), 495–503. <http://dx.doi.org/10.1007/s11571-016-9406-0>.

- Nasrolahzadeh, M., Mohammadpoory, Z., & Haddadnia, J. (2018). Higher-order spectral analysis of spontaneous speech signals in alzheimer's disease. *Cognitive Neurodynamics*, 12(6), 583–596. <http://dx.doi.org/10.1007/s11571-018-9499-8>.
- Nasrolahzadeh, M., Mohhamadpoori, Z., & Haddadnia, J. (2014). Optimal way to find the frame length of the speech signal for diagnosis of alzheimer's disease with PSO. *Asian Journal of the Mathematics and Computational Research*, 2(1), 33–41.
- Nasrolahzadeh, M., Mohhamadpoori, Z., & Haddadnia, J. (2015a). Adaptive neuro-fuzzy inference system for classification of speech signals in alzheimer's disease using acoustic features and non-linear characteristics. *Asian Journal of the Mathematics and Computational Research*, 3(2), 122–131.
- Nasrolahzadeh, M., Mohhamadpoori, Z., & Haddadnia, J. (2015b). Alzheimer's disease diagnosis using spontaneous speech signals and hybrid features. *Asian Journal of the Mathematics and Computational Research*, 7(4), 322–331.
- Nikias, C. L., & Mendel, J. M. (1993). Signal processing with higher-order spectra. *IEEE Signal Processing Magazine*, 10(3), 10–37.
- Nikias, C. L., & Raghubeer, M. R. (1987). Bispectrum estimation: A digital signal processing framework. *Proceedings of the IEEE*, 75(7), 869–891. <http://dx.doi.org/10.1109/PROC.1987.13824>.
- Ning, T., & Bronzino, J. D. (1990). Autoregressive and bispectral analysis techniques: EEG applications. *IEEE Engineering in Medicine and Biology Magazine*, 9(1), 47–50. <http://dx.doi.org/10.1109/51.62905>.
- Peng, B., Bi, Y., Xue, B., Zhang, M., & Wan, S. (2021). Multi-view feature construction using genetic programming for rolling bearing fault diagnosis [application notes]. *IEEE Computational Intelligence Magazine*, 16(3), 79–94. <http://dx.doi.org/10.1109/MCI.2021.3084495>.
- Pinheiro, P. R., Tamanini, I., Pinheiro, M. C. D., & De Albuquerque, V. H. C. (2018). Evaluation of the alzheimer's disease clinical stages under the optics of hybrid approaches in verbal decision analysis. *Telematics and Informatics*, 35(4), 776–789. <http://dx.doi.org/10.1016/j.tele.2017.04.008>.
- Poli, R., Langdon, W. B., McPhee, N. F., & Koza, J. R. (2008). A field guide to genetic programming. Lulu.com. <https://www.gp-field-guide.org.uk> ISBN 978-1-4092-0073-4.
- Rao, T. S. (1983). The bispectral analysis of nonlinear stationary time series with reference to bilinear time-series models. *Handbook of Statistics*, 3, 293–319.
- Rao, T. S., & Gabr, M. M. (2012). *An introduction to bispectral analysis and bilinear time series models* (Vol. 24). Springer Science & Business Media.
- Rentoumi, V., Raoufian, L., Ahmed, S., de Jager, C. A., & Garrard, P. (2014). Features and machine learning classification of connected speech samples from patients with autopsy proven alzheimer's disease with and without additional vascular pathology. *Journal of Alzheimer's Disease*, 42(s3), S3–S17. <http://dx.doi.org/10.3233/jad-140555>.
- Sannino, G., De, F., & De, P. (2015). Non-invasive estimation of blood pressure through genetic programming. In *Presented at the biodevices 2015–8th international conference on biomedical electronics and devices, proceedings; part of 8th international joint conference on biomedical engineering systems and technologies, biostec 2015* (pp. 241–249). <http://dx.doi.org/10.5220/0005318002410249>.
- Schmidt, M., & Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science*, 324(5923), 81–85. <http://dx.doi.org/10.1126/science.1165893>.
- Sharma, S., & Tambe, S. S. (2014). Soft-sensor development for biochemical systems using genetic programming. *Biochemical Engineering Journal*, 85, 89–100. <http://dx.doi.org/10.1016/j.bej.2014.02.007>.
- Siedlecki, W., & Sklansky, J. (1993). A note on genetic algorithms for large-scale feature selection. In *Handbook of pattern recognition and computer vision* (pp. 88–107). [http://dx.doi.org/10.1016/0167-8655\(89\)90037-8](http://dx.doi.org/10.1016/0167-8655(89)90037-8).
- Sigl, J. C., & Chamoun, N. G. (1994). An introduction to bispectral analysis for the electroencephalogram. *Journal of Clinical Monitoring*, 10(6), 392–404. <http://dx.doi.org/10.1007/BF01618421>.
- Swami, A., Mendel, J. M., & Nikias, C. L. (1998). *Higher-order spectral analysis toolbox, Vol. 3* (pp. 22–26). The Mathworks Inc.
- Thomas, C., Keselj, V., Cercone, N., Rockwood, K., & Asp, E. (2005). Automatic detection and rating of dementia of alzheimer type through lexical analysis of spontaneous speech. In *Ieee international conference mechatronics and automation, 2005, Vol. 3* (pp. 1569–1574). IEEE, <http://dx.doi.org/10.1109/ICMA.2005.1626789>.
- Vyas, R., Goel, P., & Tambe, S. S. (2015). Genetic programming applications in chemical sciences and engineering. In *Handbook of genetic programming applications* (pp. 99–140). Cham: Springer, http://dx.doi.org/10.1007/978-3-319-20883-1_5.
- Xianda, Z. (2002). *Modern signal processing*. Bei Jing: Tsinghua University Press.
- Yadavalli, V. K., Dahule, R. K., Tambe, S. S., & Kulkarni, B. D. (1999). Obtaining functional form for chaotic time series evolution using genetic algorithm. *Chaos. An Interdisciplinary Journal of Nonlinear Science*, 9(3), 789–794. <http://dx.doi.org/10.1063/1.166452>.