Machine Learning models for the diagnosis of Parkinson’s Disease based on audio signals

Author: Musteață Denisa Gabriela

Coordinator: Conf. dr. ing. Ferariu Lavinia-Eugenia

*Department of Automatics and Applied Informatics*

*Technical University Gheorghe Asachi, Iași*

Iasi, România

denisa-gabriela.musteata@student.tuiasi.ro

*Abstract*— **Parkinson's disease is one of the most common neural diseases among the population. An important discomfort that patients experience is related to changes in speech, with an impact on prosody, articulation, and phonation. This paper analyzes supervised machine learning methods using audio signals for the diagnosis of Parkinson’s disease. In this context, two main types of classifiers are designed: a Random Forest classifier using Fourier features, and a convolutional neural network that processes Mel-spectrograms. Mel-spectrograms provide a relevant description of the audio signals, with details on the low frequencies, in compatibility with the receptive range of the human ear. To ensure a proper configuration of the Mel-spectrograms, this paper proposes a genetic algorithm that determines the number of frequency bands and the frequency range. These parameters are the most influential according to preliminary experimental investigations. Thus, the genetic algorithm allows the generation of relevant data to the input of the neural model. The experimental results indicate better accuracy for convolutional neural models based on optimized Mel-spectrograms.**

Keywords — Artificial Inteligence, Aplications in medicine, Signal Processing9

# Introduction

Parkinson’s Disease was described by the author who gave it its name in 1817 known before as „agitated paralysis”, characterized by a hypertonic-hypokinetic syndrome with tremors [1]. The causes of this disease are not completely known. However, it has been found that several genetic and environmental factors contribute to the onset of Parkinson’s Disease. Research has shown that around 100-250 people per 100.000 suffer from Parkinson’s disease. [2]. In general, the disease is more frequent in people over the age of 50, although early symptoms have been observed in people between the ages 30 and 50. This condition is more prevalence in males The primary pathological disorder of the disease consists of the depletion of dopamine in the substantia niagra and neostriatum cells. [3]. The niagra and neostriatum are neuronal cells that produce dopamine in the central nervous system.

In terms of how this disease affects speech, a few modifications can be observed in the speech signals of patients. The speech difficulties of patients with Parkinson's disease are part of the category of non-motor symptoms, alongside other signs such as disturbances during sleep, difficulties with swallowing or chewing, etc. In the case of speech, prosody, articulation, and phonetics are affected, and based on these three factors, classification methods will be able to distinguish sick patients from healthy ones.

The analysis of the audio signals is a diagnostic method for Parkinson's disease of interest due to its non-invasive nature. Over time, numerous machine learning algorithms have been implemented for the detection of Parkinson's disease, with the aim of identifying the relationship between the features extracted from patients' audio recordings and the healthy/ill diagnosis. In the most recent studies, the detection of the disease has been attempted using datasets that includes handwriting or vocal recordings [4,5]. In this case, CNN models like AlexNet, GoogleNet, ResNet, etc., can be used to extract spatial features if the input data is formatted in a multi-dimensional array, or recurrent models, like LSTM, which ensure time analysis. [6].

In this paper, the focus is on methods based on convolutional neural networks (CNNs), along with Mel-spectrograms for a time-frequency domain representation. Our goal is to obtain better results through the utilization of different types of algorithms, combining machine learning techniques with various optimization methods like genetic algorithms and a visual representation of the audio signals in the form of Mel-spectrograms. This optimization method assists us in obtaining an optimal configuration for the parameters used to configure the Mel-spectrograms and improves the test and validation accuracy. The research phases followed in this paper are preprocessing, feature extraction to create a feature matrix, and classification using machine learning algorithms.

The structure of the chapters in this article is as follows: Section II highlights general information about the field of Artificial Intelligence and Mel-spectrograms. In Section III, we discuss details about the dataset and the application description based on the Random Forest Classifier. Section IV describes the method based on Mel-spectrograms and convolutional neural networks. Finally, the conclusions and future research steps will be highlighted.

# Theoretical foundations

## General Aspects of Artificial Intelligence Methods

Artificial Intelligence (AI) and a popular branch of it known as machine learning (ML) are used more often in medicine, clinical diagnosis, management or research.

Artificial intelligence field wants to understand human intelligence and to create, based on it, intelligent machines [9]. Machine learning utilizes an input dataset and outputs a dataset with a clear result. For example, a model can learn to predict whether a tumor exists or not, after being provided with a training dataset consisting of various images from both patients and healthy individuals. Therefore, we can conclude that machine learning is a part of artificial intelligence based on its ability to learn features from datasets, with or without human supervision, and make predictions based on them for multiple solutions [8].

On the other hand, AI techniques generally aim to create machines or systems that mimic human cognitive processes, such as thinking, learning, perception, problem-solving, and identifying solutions. AI technology is used in a variety of fields, from medicine and the automotive industry to virtual assistance and video games. It has several branches, including machine learning, deep learning, convolutional neural networks, and computer vision, among others.

## Convolutional neural networks (CNN)

The Deep Learning method is a type of deep and automated learning (Machine Learning) inspired by the biological nervous system. It uses multiple layers of simple preprocessing units, called nodes, which are interconnected. Similar to the functions of neurons in the nervous system, deep learning algorithms automatically extract and progressively combine complex representations of an input to accomplish a given task [7]. There is a growing interest in deep learning methods due to their ability to uncover previously unknown relationships between data and perform complex tasks

Deep Learning is a specific machine learning technique that uses deep neural networks to analyze and understand complex data. Neural networks are also inspired by the biological nervous system, which consists of neurons and neural connections.

Convolutional Neural Networks (CNNs) are a type of neural network that take a multidimensional array as input and perform convolution operations on it. Most CNN architectures have been developed for various tasks such as classification, detection, and segmentation [7]. From a technical perspective, a CNN model consists of aggregation layers, convolution layers, fully connected layers, and classification or regression layers [11].

Convolution is the operation used in the initial layers to extract features from an input array or image. It can be applied using a set of filters to generate feature maps [10].

The convolution process is described by equation (1), where f and h represent the input image and the kernel, respectively. The row and column indices are denoted by m and n, respectively [10].

(1)

The pooling layer, used after the convolution layer in CNNs, reduces the image dimensions to accelerate computations. Fully connected layers, essential in deep networks, efficiently classify the extracted features. The fully connected input layer flattens the output from previous layers into a single vector, while the output layer, usually a softmax layer, calculates the probabilities for each label.

The mathematical description of a softmax layer is presented in equation (2) [10].

(2)

AlexNet is a CNN model known for winning the ImageNet competition in 2012. It is characterized by the inclusion of five convolutional layers and three fully connected layers, using the ReLU activation function and regularization techniques such as dropout to prevent overfitting. AlexNet was a pioneer in advanced visual processing, significantly influencing the further development of image recognition algorithms.

O imagine care conține diagramă, linie, Plan, Desen tehnic

Descriere generată automat

Figure 1 – AlexNet Model Architecture [14]

## Mel-Spectrograms and their utility

A spectrogram is a visual representation of the frequency spectrum of an audio signal as it varies over time. While a classic spectrogram is similar to a Mel spectrogram in definition, the key difference and main advantage of the latter is that the Mel spectrogram uses the Mel scale. This scale is perceptual and logarithmic, designed to align with the human auditory system, which has higher sensitivity to lower frequencies.

The Mel-Spectrogram is a visual representation, with multiple columns, each representing the frequency spectrum of a segment or time interval. The color intensity indicates the amplitude of the frequencies in that segment. Thus, we can say that a spectrogram of this type divides the audio signal into several frames, in order to achieve higher accuracy and greater similarity to what the human ear can perceive.

O imagine care conține diagramă, Plan, captură de ecran, Desen tehnic

Descriere generată automat

Figura 2 – Arhitectura funcției melSpectrogram [15]

# Diagnosis of Parkinson’s Disease Ursing Fourier Features and Random Forest

The application aims to diagnose Parkinson's disease based on audio signals, using machine learning classification methods. The core of the application is an algorithm that processes the data, extracts features, and classifies subjects into two classes: PC (1) for Parkinson's disease and HC (0) for healthy controls, based on audio recordings.

In the training dataset, recordings taken in various situations will be considered. In the recordings where the subject is asked to read a text, the sentence each one must pronounce is "The North Wind and the Sun." Depending on the subject's constitution, they may also be asked to read "Tech. Engin. Computer applications in geography snippet."

The dataset used for this study was created by King’s College London (KCL) Hospital, Denmark Hill, Brixton, London SE5 9RS, between September 26-29, 2017. The authors of the dataset are Hagen Jeger, Dhaval Trivedi, and Michael Stadtschhnitzer. This dataset consists of 35 recordings from individuals who are both Parkinson's disease patients and healthy controls. Each person has two types of recordings: one from reading a predefined text, previously mentioned, and the other captured during a spontaneous conversation. The recordings were made using a Motorola Moto G4 Smartphone. Although the recordings were directly made with the phone's microphone rather than a GSM (Global System for Mobile Communications), they are of high quality, with a sampling frequency of 44.1 kHz and 16-bit depth. The recordings are in wave (.wav) format.

The application is developed in MATLAB. Below are the main components of the application.

## Preprocessing and future extraction

Preprocessing is a fundamental step in preparing a dataset for machine learning methods. This step involves a set of operations applied to raw data, making it suitable for further processing. It is essential to achieve the best possible accuracy results. Well-executed preprocessing ensures improved performance of the model during training and also enhances generalization for new, previously unseen datasets.

In the case of our dataset, preprocessing was divided into several stages: extracting the recordings and loading them into MATLAB, creating cell arrays to store the important data that will later be used in the design, splitting each audio signal into frames of a certain length, and applying filters to extract Fourier features.

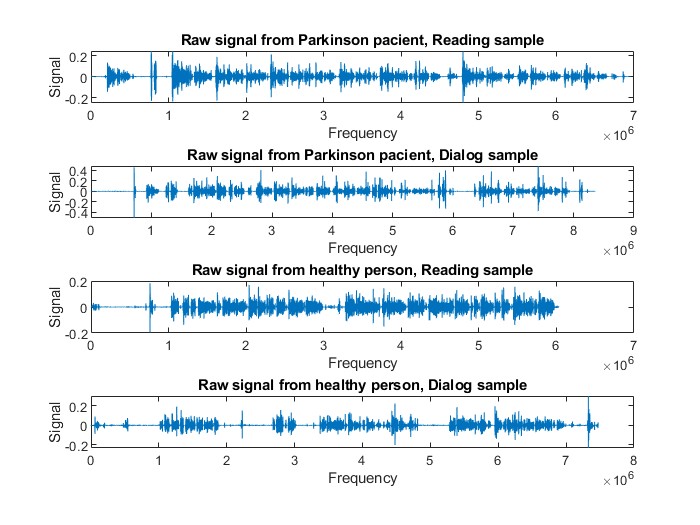
In Figure 3, the signals from a Parkinson's disease patient and a healthy individual are represented before being processed.

Figure 3 – Unprocessed signals from a ill and a healthy pacient

We can observe, even in the raw signal stage, that there is a difference between the audio signal representation of a Parkinson's disease patient and that of a healthy individual. To facilitate processing, each audio signal was segmented into frames of 10 seconds, with a 1-second overlap between adjacent frames. The choice of overlap was made to avoid the loss of significant information at the intersection between two consecutive frames.

The resulting frames were assigned the corresponding classes and were divided into datasets for training and testing.

In Figure 4, the values of the signal in the time domain are illustrated as histograms, specifically the means and variances, considering the frames of a Parkinson's disease patient and a healthy individual.

It is observed that the mean values of the samples are predominantly concentrated in the first interval, both for Parkinson's disease patients and healthy subjects, indicating a general uniformity in their responses. The minor differences observed may reflect subtle variations attributed to the impact of Parkinson's disease. As for the standard deviation, more pronounced differences between the two groups are highlighted. Healthy subjects show a distribution with moderate, constant variability, whereas patients display significantly greater dispersion, indicating increased variability in the values, most likely due to the neurodegenerative effects of Parkinson's disease on cognitive functions.

The analysis of the histograms suggests that, although there are differences between the studied groups, they are subtle enough that the classifier used cannot make significant distinctions based on the raw data. To achieve efficient classification, it is essential to extract the most relevant features that can accurately detect the sensitivities and variations between the two types of subjects.

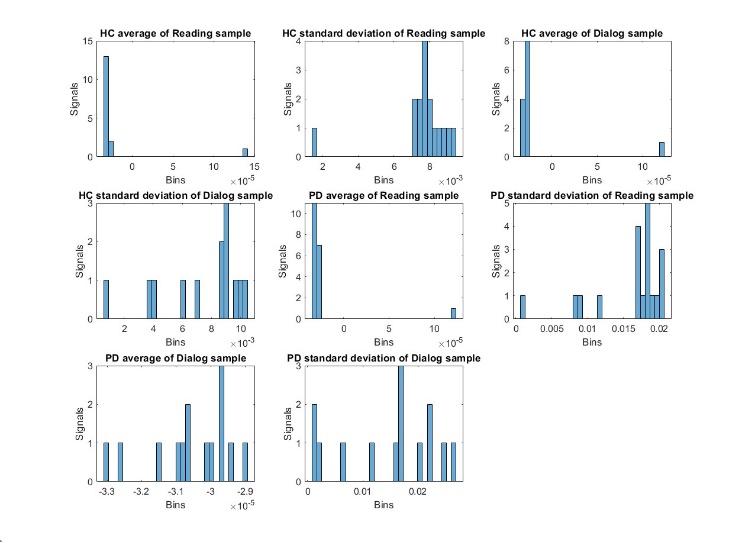


Figure 4 – Histograms – examples from a healthy and also from an ill pacient

The next step is the construction of a feature matrix. The feature matrix is a two-dimensional structure where each row corresponds to a frame, and each column corresponds to a feature or characteristic extracted for each example. This step is also essential in data analysis, as it simplifies the task for the classifier. Proper feature extraction can enhance the accuracy performance and the generalization ability of the model.

In the analysis of our dataset, the feature matrix was created using the Fourier transform. The Fourier transform (FT) is a well-known method in signal processing. It is a signal processing technique that decomposes a signal into a sum of sinusoids. The Discrete Fourier Transform (DFT) is a variant of the Fourier transform that operates on a discrete, equally spaced sequence of data corresponding to a continuous function. It is an extremely efficient method for signal analysis, providing relevant spectral information [9]. The Discrete Fourier Transform (DFT) is applied to audio signals to transform the vocal signal, represented in the time domain, into the frequency domain. The formula for transforming the vocal signal x(n)x(n)x(n) into the frequency domain using the Discrete Fourier Transform is given by the equation [10]:

(3)

In our application, we used the MATLAB function "fft" to obtain the Discrete Fourier Transform and analyze the frequency spectrum for constructing the feature matrix. In Figure 8, we have highlighted two graphical representations of signals from two different subjects (HC and PD) after applying the Fourier Transform. It can be clearly observed that there are many high-frequency sinusoids with amplitudes close to 0, which can hinder the training process without providing relevant information.

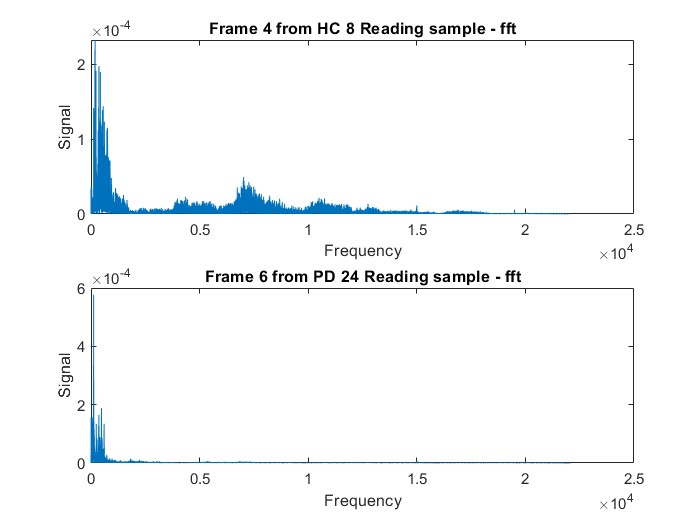


Figure 5 – Ilustrarea trăsăturilor Fourier pentru un cadru de la un pacient sănătos (imaginea de sus) și un pacient bolnav (imaginea de jos)

The process of creating the feature matrix involves several essential steps. Initially, harmonics with amplitudes close to zero are removed, as they are often associated with high frequencies and noise. Then, the signals are examined after applying the Fourier Transform to ensure they are nearly identical to the original signal. The spectral analysis was restricted to the lower half of the frequency spectrum because the signal becomes symmetric after applying the Transform. Finally, the Discrete Fourier Transform is applied to the entire dataset, maintaining a constant size for the frequency range retained to ensure uniformity and comparability of the extracted features. For applying the Discrete Fourier Transform, we used the “fft” function from the MATLAB library.

Setting the frequency range is an essential step in the signal analysis process. To validate the correctness of this range selection and to ensure that the data is suitable for subsequent training, we performed experimental tests using three different frequency values: 1,500 Hz, 11,500 Hz, and 21,500 Hz. After analyzing approximately 20 reconstructed signals, we concluded that these three values are the most appropriate for determining the ideal frequency range length. We opted for the median value, 11,500 Hz, along with two adjacent values, at a ±10,000 Hz variation, to ensure adequate coverage and a nearly identical representation of the reconstructed signal compared to the original one.

To validate the correctness of the dataset and the selected frequency range, we developed a Random Forest classifier. As input data, the classifier requires a feature matrix and a class vector, each split into training and testing data. We created three feature matrices, each corresponding to one of the chosen frequency range values, and ran the algorithm for different tree count values: 50, 100, 150, and 200. The value to be retained will be the one for which the best average accuracy was obtained on the test set. This approach ensures an accurate evaluation of the classifier's performance under different conditions and helps us determine the most efficient configuration for future use.

## Random Forest experimental results

The Random Forest method is extremely popular in the field of machine learning. It belongs to the category of supervised learning algorithms, used for both classification and regression problems. This makes it suitable for an initial analysis of the dataset.

Initially, we split the feature matrix and the class vector into training and testing data. We defined 80% of the data for the training set, with the remaining 20% allocated for testing. We used the "cvpartition" function from the MATLAB library to ensure that the data partitioning was done randomly. For training, we used the "fitensemble" function, and predictions for the training and testing examples were subsequently calculated using the "predict" function. To verify that the dataset is correctly processed, we will calculate the model's accuracy by comparing the predictions with the actual labels, also calculating the percentage of correct predictions.

The analysis of the correctness and errors in signal interpretation will be carried out using a confusion matrix, generated during the testing process. This will be created using the "confusionmat" function from the MATLAB library.

Table 1 is organized as follows: the first column indicates the number of trees used for each of the four training runs. Adjacent to it, the two classes are displayed: "0" for healthy individuals and "1" for patients with Parkinson's disease. In the next two columns, the confusion matrix is presented, where the values for true negatives (00) – healthy patients (class 0) correctly identified as healthy, and false positives (01) – healthy patients misclassified as sick, are shown. Similarly, the values for false negatives (10) – sick patients (class 1) misclassified as healthy, and true positives (11) – sick patients correctly identified as sick, are displayed. The last column shows the accuracy achieved on the test dataset for each number of trees used in the Random Forest algorithm. This structure enables a clear analysis of the classification model's performance, highlighting its accuracy in detecting both states.

From the following table, it can be observed that there are no false positives or false negatives because the training accuracy was 100%. As a result, only the true negative and true positive values were interpreted.

Tabel 1 – Random Forest results – 21,500 Hz

| Nr. Of Trees | Class | Confusion Matrix | | | | Test Accuracy |
| --- | --- | --- | --- | --- | --- | --- |
| Test | | Train | |
| 0 | 1 | 0 | 1 |
| 50 | 0 | 125 | 16 | 508 | 0 | 80.00% |
| 1 | 28 | 51 | 0 | 327 |
| 100 | 0 | 126 | 15 | 508 | 0 | 80.45% |
| 1 | 28 | 51 | 0 | 327 |
| 150 | 0 | 125 | 16 | 508 | 0 | 79.55% |
| 1 | 29 | 50 | 0 | 327 |
| 200 | 0 | 126 | 15 | 508 | 0 | 80.91% |
| 1 | 27 | 52 | 0 | 327 |

The average accuracy obtained for the value of 21,500 Hz is 80.22%. This is the highest average achieved for the test set, which leads us to the decision to retain the value of 21,500 Hz for the frequency range length. The average accuracy for 11,500 Hz was 77.55%, and for 10,500 Hz, it was 79.88%.

The conclusion drawn with the help of the Random Forest classifier is that our data is relatively well processed, and better results can be obtained on testing after further processing and subsequent training. Therefore, in the next phase, the design of a convolutional neural network (CNN) model was considered, incorporating the AlexNet model, which will process Mel-spectrograms and be further optimized through the design of a genetic algorithm for more advantageous configuration of the input data.

# Clasification using CNN model AlexNet și Mel-spectrograms

The use of Convolutional Neural Networks (CNNs) is one of the most widespread methods in image-based diagnosis. Input data can range from radiographs or CT images to signals in the form of spectrograms. In this study, since we are working with audio signals, the data provided to the model will be in the form of Mel-spectrograms. The decision to use Mel-spectrograms instead of traditional spectrograms was made due to the logarithmic nature of the Mel scale, which leads to a more accurate approximation of the similarity and sensitivity perceived by the human ear. In previous research that focused on a similar study [10][16], the use of Mel-spectrograms in combination with VMD (Variational Mode Decomposition) and deep learning was observed, establishing standard values for parameters such as frame length, overlap length, and the number of Mel bands. One of the main objectives of this work was to automate the process of identifying the most optimal parameters, and for this purpose, we aimed to develop a model that allows them to be learned in an adaptable way, so they can be used effectively in the final training phase, while also avoiding the risk of overfitting that can occur during testing.

The AlexNet neural network cannot receive audio signals or numerical data as input, so we opted for using Mel-spectrograms. These are based on various parameters that can influence the results of training. The steps followed by the MATLAB function when constructing such a spectrogram are as follows: signal extraction, segmentation into smaller frames, applying the FFT function, applying Mel-type filters, and summing the results. It is important to note that the “melSpectrogram” function does not apply the traditional Fourier Transform through the FFT function; instead, it converts the frequency spectrum into a Mel spectrum. Therefore, it is essential that before calling this function, our signal should have already undergone a Fourier filtering, transforming the signal from the time domain to the frequency domain. Finally, I saved the spectrograms with the ".mat" extension to allow reading them without further format conversions.

## Preliminary analysis of the model

The first step in the automation process, aimed at identifying the optimal parameters, involved observing and determining the most influential factors. During this phase, several experimental investigations were carried out by manipulating the values for window length (Window Length) and the number of Mel bands (NumBands). Additionally, the overlap length was considered to be dependent on the window length, being set at a quarter of this value. The experimental investigations were conducted by allocating values for these two parameters, chosen as powers of 2. Specifically, the window length ranged from 256 to 4096, and the number of bands ranged from 16 to 64. These configurations were evaluated in the context of using the AlexNet model, where a learning rate of 0.0001 was applied

I conducted training for a number of 50 epochs and a batch size of 100, thus avoiding overfitting the network. These tests were designed as preliminary tests, not final trainings, providing an initial evaluation of the model's behavior.

Tabel 2 – AlexNet results – MBS – 100, NEP – 50

| Nr.  Crt. | Window  Length | Num  Bands | AlexNet train results | | |
| --- | --- | --- | --- | --- | --- |
| Train | Test | Validation |
| 1. | 256 | 16 | 88.64% | 81.82% | 76.36% |
| 2. | 256 | 32 | 96.14% | 85.45% | 81.82% |
| 3. | 256 | 64 | 95.34% | 80.91% | 83.64% |
| 4. | 512 | 16 | 92.05% | 79.09% | 81.82% |
| 5. | 512 | 32 | 97.05% | 84.55% | 83.64% |
| 6. | 512 | 64 | 94.32% | 80.91% | 78.18% |
| 7. | 1024 | 16 | 92.39% | 89.09% | 76.36% |
| 8. | 1024 | 32 | 95.00% | 81.82% | 86.36% |
| 9. | 1024 | 64 | 94.43% | 78.18% | 78.18% |
| 10. | 2048 | 16 | 90.23% | 78.18% | 81.91% |
| 11. | 2048 | 32 | 94.32% | 81.82% | 79.09% |
| 12. | 2048 | 64 | 94.32% | 85.45% | 82.73% |
| 13. | 4096 | 16 | 86.48% | 76.36% | 74.55% |
| 14. | 4096 | 32 | 88.52% | 72.73% | 77.27% |
| 15. | 4096 | 64 | 89.43% | 77.27% | 78.18% |

The results of these preliminary trainings are described in Table 2. The conclusions drawn will be elaborated in the following sections.

Depending on the number of bands allocated for each window size, the observations vary. For a window size of 256, the results are generally positive, with a notable exception when the number of bands is 16. For a window size of 512, performance drops below 80% for both testing and validation when the number of bands is 16 and 64. With a window size of 1024, the weakest results are recorded when the number of bands is 64. For a window size of 2048, the results are generally good, with only minor decreases observed for 16 and 32 bands. Finally, with a window size of 4096, the results fall below the desired average, regardless of the number of bands used.

For NumBands = 16, we observe that the algorithm learns more efficiently with an average number of windows. Accuracy decreases both in training and in testing and validation when the number of windows is changed. For example, with a training accuracy of 92.05%, accuracy drops to 79.09% in testing and 81.82% in validation, indicating unsatisfactory results. Increasing the number of bands leads to a significant decrease in accuracy across all evaluation stages, although values remain above 80% when NumBands is set to 32 for a small number of windows. However, for NumBands = 64, satisfactory results are obtained only for window sizes of 256 and 2048. These findings highlight the importance of correctly selecting the number of bands and the window size in optimizing the algorithm's performance.

In conclusion, the worst results were obtained for a window length of 4096, while the network performed the best in training, testing, and validation with window lengths of 256, 512, and 2048. These results suggest that smaller window sizes are more efficient in capturing audio signal features without overloading the model with unnecessary information, whereas a window size that is too large can lead to performance degradation.

## Integration of the Genetic Algorithm for Optimizing the Configuration of Mel-Spectrograms

The concept behind genetic algorithms is derived from the theory of evolution of species, first proposed by Charles Darwin in 1859. This theory states that species undergo a continuous process of evolution, in which variations are passed down to offspring. According to this theory, individuals that are better adapted to their environment have the highest chances of survival and development. These principles are applied in genetic algorithms to optimize solutions to various computational problems through evolutionary computing, and they are generally used for solving complex optimization and search problems, where classical methods may prove inefficient. Additionally, this algorithm is particularly accessible and preferred because it does not rely on gradient-based calculations.

Regarding evolutionary computing, it transposes the theory of natural selection and the theory of evolution into numerical algorithms. Thus, the natural model is adopted in a simplified version. Evolutionary algorithms operate on a population of structures that evolve over multiple generations, with the best-adapted structures surviving to the next generations and contributing to the production of new offspring that are increasingly well-adapted..

The steps that the genetic algorithm goes through are structured as follows: initialization of solutions and chromosome representations, selection of the best solutions, crossover and mutation to produce new solutions, and insertion to create the population for the next generation. In more detail, the algorithm models the solutions to the problem as a population of individuals, with each individual representing a potential solution. These individuals are represented by chromosomes, which are strings of characters (e.g., 0/1 bits) encoding the solution's parameters. In our algorithm, a population consists of Nind individuals, and was generated using reduced Nind values (4-8).

Selection is performed based on the adaptation level (fitness) of each individual, with individuals having poor performance being eliminated — in our case, approximately 50% of the population. The next important step is crossover, a process that combines the chromosomes of two parents to generate offspring, by swapping sections of their chromosomes at randomly generated crossover points. The probability of this process occurring between two parent chromosomes in our algorithm is 0.7. Mutation, another key mechanism, helps maintain genetic diversity in the population and prevents the algorithm from stagnating around the same solutions. This process is random and involves altering one or more genes (bits) in the chromosome, such as changing bits from 0 to 1 and vice versa. The mutation probability in our algorithm is 0.1. Finally, the iteration and convergence process involves repeating these procedures for multiple generations until the algorithm finds a satisfactory solution or until the pre-established maximum number of generations is reached.

In this case, a genetic algorithm is implemented to identify the optimal parameters that will contribute to the construction of the Mel-spectrograms. Once the optimal solution is identified, these parameters will be passed to the function responsible for creating the spectrograms, which, in turn, will be used in the neural network training process. The decision variables are two exponents, e1 and e2, which will determine the number of bands and the window size, 2^e1 and 2^e2, respectively.

The objective function is considered to be the training accuracy of a neural network model trained for a limited number of epochs with the resulting spectrograms. In the implementation, "evalChrom" ensures the calculation of the objective values. Within this function, each solution is decoded to create the Mel-spectrograms, then the AlexNet neural network is trained with a reduced number of epochs (NEP) and using mini-batches of size MBS. Tests were conducted for different configurations of these parameters, with MBS values of 25, 50, and 75, and for epoch numbers of 50, 100, and 150. The final training focused on two specific combinations of these parameters: MBS = 50 and NEP = 100, and MBS = 100 and NEP = 50. These experimental investigations used a learning rate of 0.0001 for the AlexNet model, and although they are still in progress, Table 3 presents some results from which a few satisfactory conclusions can be drawn.

Tabel 3 – Rezults of AlexNet after optimization with Genetic Algorithm

| MBS/  NEP  (GA) | Nr ind | FctObj | Phen [*e*1,*e*2] | MBS | NEP | Rezultate antrenări AlexNet | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Train | Test | Validation |
| 25/50 | 4 | 0.1364 | [9,7] | 100 | 50 | 97.05% | 80.91% | 88.18% |
| 25/50 | 4 | 0.1364 | [9,7] | 50 | 100 | 99.77% | 87.27% | 81.82% |
| 50/100 | 4 | 0.0375 | [10,4] | 100 | 50 | 93.52% | 80.91% | 90.00% |
| 50/100 | 4 | 0.0375 | [10,4] | 50 | 100 | 99.20% | 84.55% | 85.45% |
| 75/150 | 4 | 0.0364 | [9,5] | 100 | 50 | 96.93% | 83.64% | 84.55% |
| 75/150 | 4 | 0.0364 | [9,5] | 50 | 100 | 99.43% | 80.00% | 85.55% |
| 25/50 | 8 | 0.0432 | [7,5] | 100 | 50 | 96.36% | 83.64% | 80.00% |
| 25/50 | 8 | 0.0432 | [7,5] | 50 | 100 | 99.66% | 90.00% | 89.09% |
| 50/25 | 8 | 0.1284 | [8,7] | 100 | 50 | 94.77% | 88.18% | 81.82% |
| 50/25 | 8 | 0.1284 | [8,7] | 50 | 100 | 98.64% | 87.27% | 83.64% |

The conclusion that can be drawn after optimization using the genetic algorithm is that it offers advantages in the design of the neural network model. The training accuracy is close to 100%, while the results for testing and validation are above 80%. It can be observed that the best results were obtained for the following combinations of exponents: [9,7], [10,4], [7,5], and [8,7].

# Conclusions and future developments

This paper presents an application for diagnosing Parkinson's disease using audio signals. Two working approaches are considered. The first one is based on Fourier features and the Random Forest algorithm, while the second approach uses Mel-spectrograms and convolutional neural networks. For the second method, an optimization variant of the Mel-spectrogram configuration is proposed, utilizing a genetic algorithm with binary encoding.

Preliminary experiments indicate promising results in establishing significant differences between the vocal signals of patients with Parkinson's disease and healthy individuals, for solving the classification problem. All investigated variants provided good accuracy performance, however, the best results were achieved using the AlexNet model and Mel-spectrograms optimized with the help of the genetic algorithm.

Future investigations will also focus on optimizing other working parameters using the genetic procedure, considering the possibility of expanding the dataset, as well as potential additional preprocessing steps that may be required.

##### Bibliography

1. L. Pendefunda, Neurologie, Iasi: *Editura U.M.F*., 1996.
2. A. E. Lang and A. M. Lozano, "Parkinson's disease," *New England J. Med., vol. 339,* pp. 1130-1143, 1998.
3. A. H. Kaye, Essential Neurosurgery, Third Edition, Melbourne, Australia: *Blackwell Publishing*, 2005.
4. Este
5. C. D. Rios-Urrego, J. C. Vasquez-Correa, J. F. Vargas-Bonilla, E. Noth, F. Lopera și R. Orozco-Arroyave, „Analysis and evaluation of handwriting in patients with Parkinson’s disease using kinematic, geometrical, and non-linear features,” *Computer Methods and Programs in Biomedicine,* pp. 43-52, 2019.
6. C. Chen, Z. Hua , R. Zhang, G. Liu și W. Wen, „Automated arrhythmia classification based on a combination network of CNN and LSTM,” *Biomedical Signal Process Control,* 2020.
7. S. Mutasa, S. Sun și R. Ha, „Understanding artificial inteligence based radiology studies: CNN arhitecture,” *Clinical Imaging*, pp. 72-76, 2021.
8. T. Cristina, S. Asne și M. Patrick, „Machine Learning (ML) diffusion in the design process: A study of Norwegian design consultancies,” *Technological Forecasting & Social Change,* 2023.
9. K. Presannakumar și A. Mohamed, „Deep learning based source identification of environmental audio signals using optimized convolutional neural network,” *Applied Soft Computing*, 2023.
10. M. B. Er, E. Isik și I. Isik, „Parkinson's detection based on combined CNN and LSTM using enhanced speech signals with Variational mode decomposition,” *Biomedical Signal Processing and Control*, 2021.
11. C. Quan, K. Ren, Z. Luo, Z. Chen and Y. Ling, "End-to-end deep learning approach for Parkinson's disease detection from speech signals," *Biocybernetics and biomedical engineering,* pp. 556-574, 2022.
12. D. Cireundefinedan, U. Meier, J. Masci, L. Gambardella și J. Schmidhuber, „Flexible, High Performance Convolutional Neural Networks for Image Classification,” *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence,* pp. 1237-1242, 2011.
13. Z. Sun, G. Wang, P. Li, H. Wang, M. Zhang și X. Liang, „An improved random forest based on the classification accuracy and correlation measurement of decision trees,” *Expert Systems With Applications*, 2024.
14. Data Set (online free to use): [https://zenodo.org/records/2867216#.XeTbN59R2BZ](about:blank)
15. [https://medium.com/@tuncerergin/convolutional-neural-network-convnet-yada-cnn-nedir-nasil-calisir-97a0f5d34cad](about:blank)
16. MathWorks