SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V12I1P105

Volume 12 Issue 1, 32-46, January 2025


*Original Article*

# Implementation of Dysarthria Identification Using MFCC and Multilayer Perceptron Algorithm

Abdul Fadlil[1], Latief Perdana[2], Ardi Pujiyanta[3], Herman[4], Haris Imam Karim Fathurrahman[5], Maulana Muhammad Jogo Samodro[6]

[1,5]*Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta 55164, Indonesia.*
[2,3]*Master of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta 55164, Indonesia.*
[4,6]*Master of Informatics, Universitas Ahmad Dahlan, Yogyakarta 55164, Indonesia.*

[1]*Corresponding Author : fadlil@mti.uad.ac.id*

Received: 03 November 2024 Revised: 05 December 2024 Accepted: 06 January 2025 Published: 25 January 2025

***Abstract*** *- Dysarthria is an inability of the child's muscles to pronounce certain vocabulary. One of the words that is often difficult to pronounce is the R sound. Therefore, it is important to identify R sound dysarthria as a preventive measure and can be used as a therapeutic reference. The study uses the phrase "laler menclok pager" as the basis for picking up voice data in children. In that sentence, there is a letter R that will be processed later. The processing method used is MFCC. The output from the extraction of the MFCC characteristics is inserted as the input material of the Multilayer Perceptron (MLP) artificial intelligence algorithm. The results of this study provide a high degree of accuracy, and the test data can be well identified as a whole. The results also obtained the MLP configuration of 16 input neurons and 8 hidden neurons with the highest accuracy as well as the lightest computing. With this result, further hardware can be developed to integrate the system for identifying dysarthria.*

***Keywords*** *- Dysarthria, MFCC, MLP, Neuron, R sound.*

## 1. Introduction

Dysarthria is a motor speech disease caused by neurological damage that impairs the muscles used in speech production [1]. Among the various speech sounds impacted by dysarthria, the pronunciation of the "R" sound poses a particular challenge for many individuals [2]. The "R" sound, or rhotic sound, requires precise coordination of the tongue, lips, and other articulators, making it one of the most complex phonemes to produce. Difficulty in articulating the "R" sound can significantly impact intelligibility and overall communication effectiveness, particularly for individuals with dysarthria [3, 4]. Nowadays, dysarthria can be identified using several approaches, including speech recognition. Speech processing has advanced significantly in recent years, notably in the field of Automated Speech Recognition (ASR) [5, 6]. Detecting dysarthria speech remains one of the most difficult difficulties that ASR systems confront, particularly in applications such as medical diagnostics, assistive technology, and law enforcement [7-9]. Dysarthria can be caused by a variety of diseases, such as neurological illnesses [10, 11], alcoholism [12, 13], or exhaustion, making its correct diagnosis critical for both clinical and practical purposes. Mel-Frequency Cepstral Coefficients (MFCCs) have emerged as one of the most useful characteristics for speech signal analysis due to their ability to describe sound's short-term power spectrum [14-17]. MFCCs are frequently employed in speech and speaker recognition applications because of their resilience and efficiency. This work uses MFCCs to extract essential information from speech signals and help identify dysarthria.

To improve detection accuracy, this research suggests employing a Multilayer Perceptron (MLP), a form of artificial neural network [18-20], to categorize speech as dysarthria or non-dysarthria using MFCC characteristics. MLPs are effective tools for pattern identification [21, 22] and classification [23, 24] because they can learn complicated, non-linear correlations between inputs and outputs. Training an MLP on MFCC characteristics collected from speech samples allows the model to accurately identify between dysarthria and non-dysarthria. Despite advances in speech processing and dysarthria identification, such as previous research, significant gaps remain in implementing identification systems. Current studies are often limited to specific datasets [25], lack robustness in real-world implementations [26], and lack exploration of feature extraction combinations beyond MFCC to improve accuracy [27]. Furthermore, while MLP is effective, its potential for implementation in lightweight, real-time applications has not been fully exploited.

**Table 1. List of acronyms and symbols**

| Acronyms | Definition |
|---|---|
| MFCCs | Mel-Frequency Cepstral Coefficients |
| ASR | Automated Speech Recognition |
| MLP | Multilayer perceptron |
| IDCT | Inverse Discrete Cosine Transform |
| FFT | Fast Fourier Transform |
| wav | Waveform Audio File |
| kHz | Kilo Hertz |
| bit | Binary Digit |

Furthermore, existing solutions often fail to provide affordable and accessible tools suitable for resource-constrained settings. Addressing these gaps could lead to the development of general and robust real-time systems that integrate advanced feature extraction techniques, adaptive modelling for personalization, and lightweight architectures for deployment in diverse and low-resource environments.

This study aims to create a robust dysarthria detection system that incorporates MFCC feature extraction and an MLP classifier. The suggested method includes gathering speech samples, extracting MFCCs, and training the MLP to recognize dysarthria patterns. This system's performance will be tested using a dataset containing both dysarthria and clear speech examples, emphasising determining the model's accuracy and generalizability. The rest of the paper is arranged as follows: Section 2 describes the method, which includes data collecting, MFCC extraction, and the use of MLP. Section 3 summarizes the experimental data and examines the suggested system's performance. Section 4 closes the report by outlining potential future research directions. Table 1 is a list of acronyms and symbols used throughout this article.

## 2. Methods

The development of a classification model to identify dysarthria using a combination of MFCC and MLP was carried out according to the research framework, as shown in Figure 1. The development of the classification model begins with collecting the audio dataset. This dataset is primary data collected conventionally by recording the pronunciation of words by individuals who are subjects of this research. The research subjects included people with speech disorders (dysarthria) and ordinary people. The audio recordings of each research subject are then stored in digital files as elements of the dataset.

The classification model is built using audio features extracted from the audio dataset. The feature extraction process uses the MFCC algorithm operated in the Python environment. This algorithm will internally undergo several process stages, including pre-emphasis, framing, windowing, FFT, filtering, and discrete cosine transformation. This study designs the feature extraction process to produce three groups of datasets. Each group will contain the same number of audio datasets but different sets of MFCC coefficients.

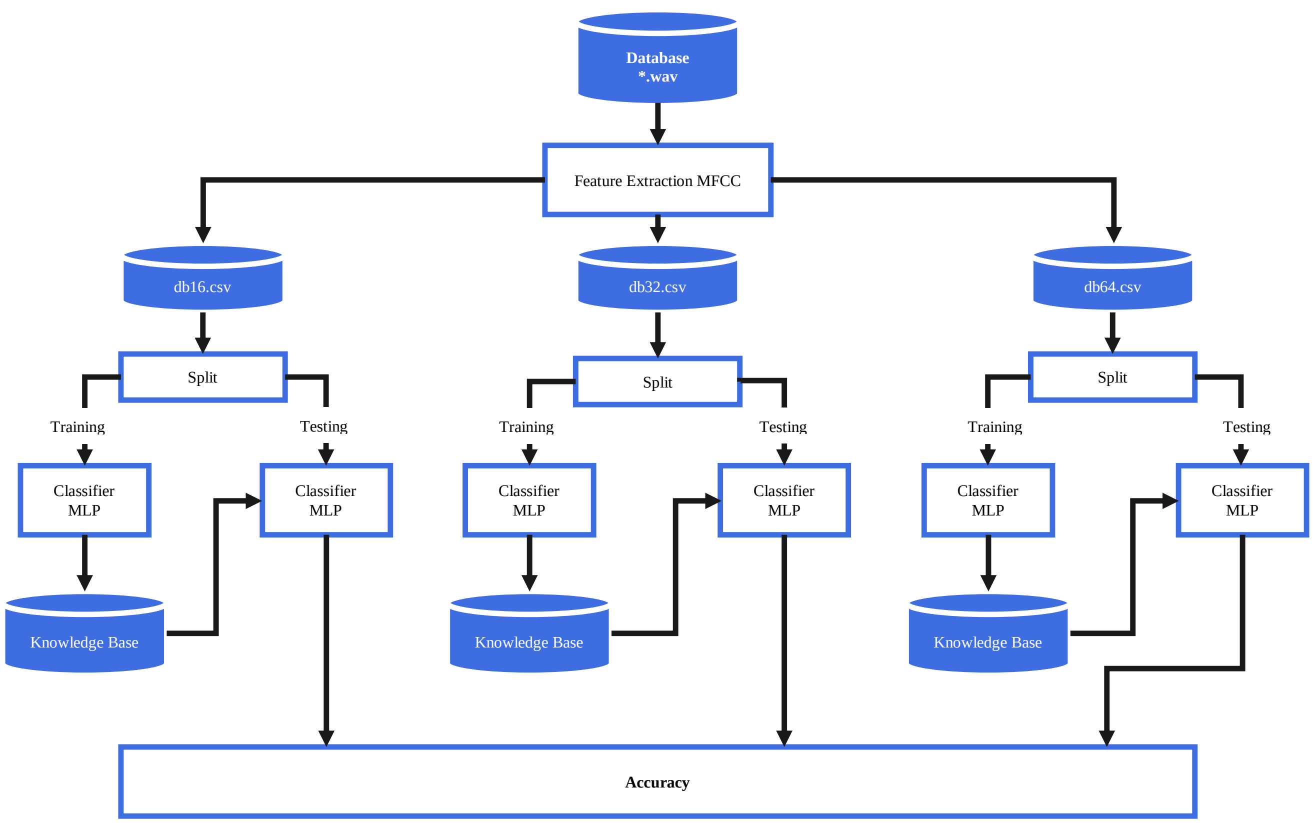



**Fig. 1 Research framework**

**Table 2. Sample voice data**

| No. | Class | Total |
|---|---|---|
| 1. | Dysarthria | 100 |
| 2. | Non-dysarthria | 100 |

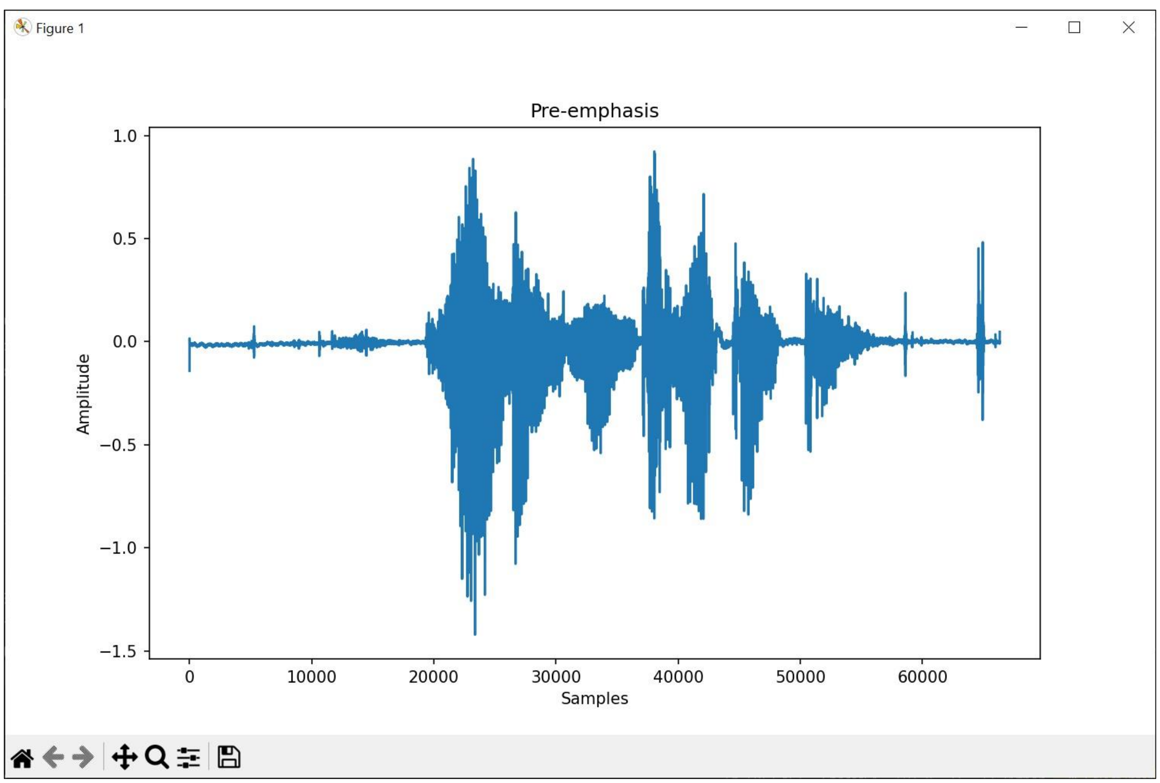
**Table 3. Subjects information**

| Characteristic | Dysarthria | Non-dysarthria |
|---|---|---|
| Age | 5-25 (years old) | 5-25 (years old) |
| Gender | 5 Female, 4 Male | 7 Female. 5 Male |
| Diagnosis | 9 | 12 |

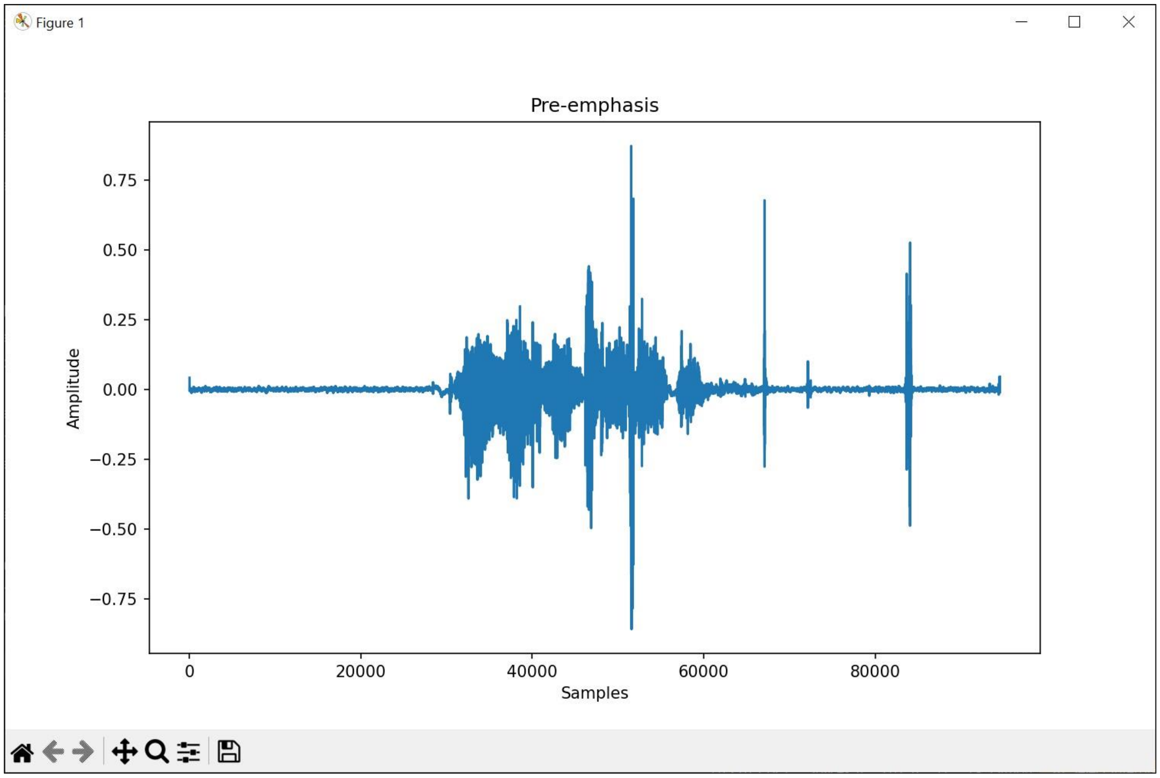
One group of datasets has 16 MFCC coefficients, another group has 32 MFCC coefficients, and the third group has 64 MFCC coefficients. Subsequently, this research conducted data training and testing using MLP modelling on each dataset group with different MFCC coefficients, namely, db16.csv, db32.csv, and db64.csv. The WEKA tools used in the modelling prepared each group's dataset into three training and testing data compositions. The compositions are, respectively, 80% training data and 20% testing data (80:20), 50% training data and 50% testing data (50:50), and lastly, 20% training data and 80% testing data (20:80). In addition to variations in the input layer according to the number of MFCC coefficients owned by the dataset, the MLP modelling in this study also applies three different hidden layers for each trained dataset. The first model has 8 hidden layers, the second model has 16 hidden layers, and the third model applies 32 hidden layers. Meanwhile, WEKA set two fixed parameters, 0.3 for learning rate and 0.2 for momentum.

### *2.1. Data Collections*

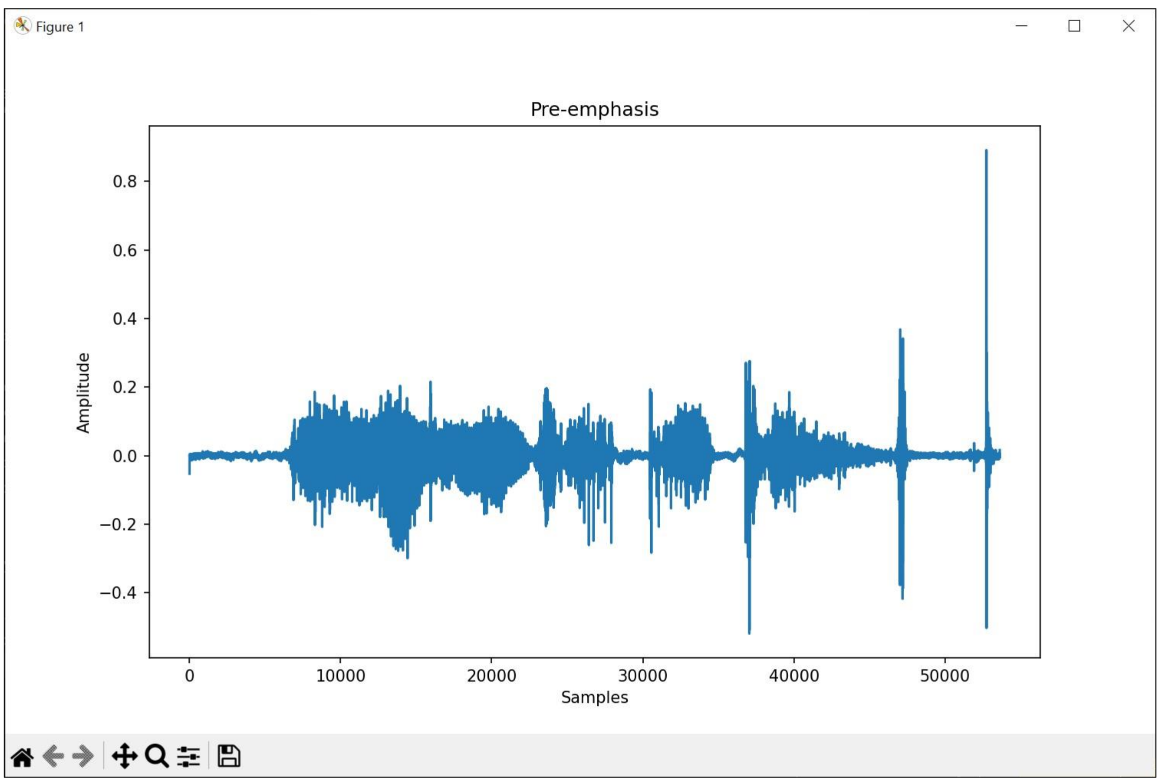
The process of collecting the audio dataset involved 200 research subjects, 100 data samples with dysarthria, and 100 data samples with non-dysarthria. The sum of the sample data can be seen in Table 2. Research subjects were selected randomly, either for dysarthria or non-dysarthria subjects. The number of subjects used in this study was 21 people divided into 9 non-dysarthria subjects and 12 with dysarthria. Details of the subject information can be seen in Table 3. It's important to note that, especially for dysarthria subjects, there was no prior knowledge about the detailed condition of the disorder or the level of dysarthria experienced by the subjects, ensuring the objectivity of the research. Research subjects were asked to say three examples of words combined into one phrase: "laler-menclok-pager". In the Indonesian language, these three words are commonly used phrases to help train people suffering from dysarthria. The pronunciation of these example words was recorded in a studio with minimal noise using a smartphone. The audio was recorded with a 16 kHz sample rate and 32-bit audio bit depth settings. Audio files are stored in Waveform Audio File (*.wav) format. The duration of each audio data is not determined fixedly but follows the duration of the pronunciation of the three sample words naturally by each research subject. Each audio data also has a backup to anticipate if an error occurs or the data is corrupted in the following process stage. Figure 2 shows audio data from voice recordings of three research subjects suffering from dysarthria, and Figure 3 shows three audio data from subjects with non-dysarthria speech.




**(a)**




**(b)**




**(c)**

**Fig. 2 Audio data from dysarthria subject**

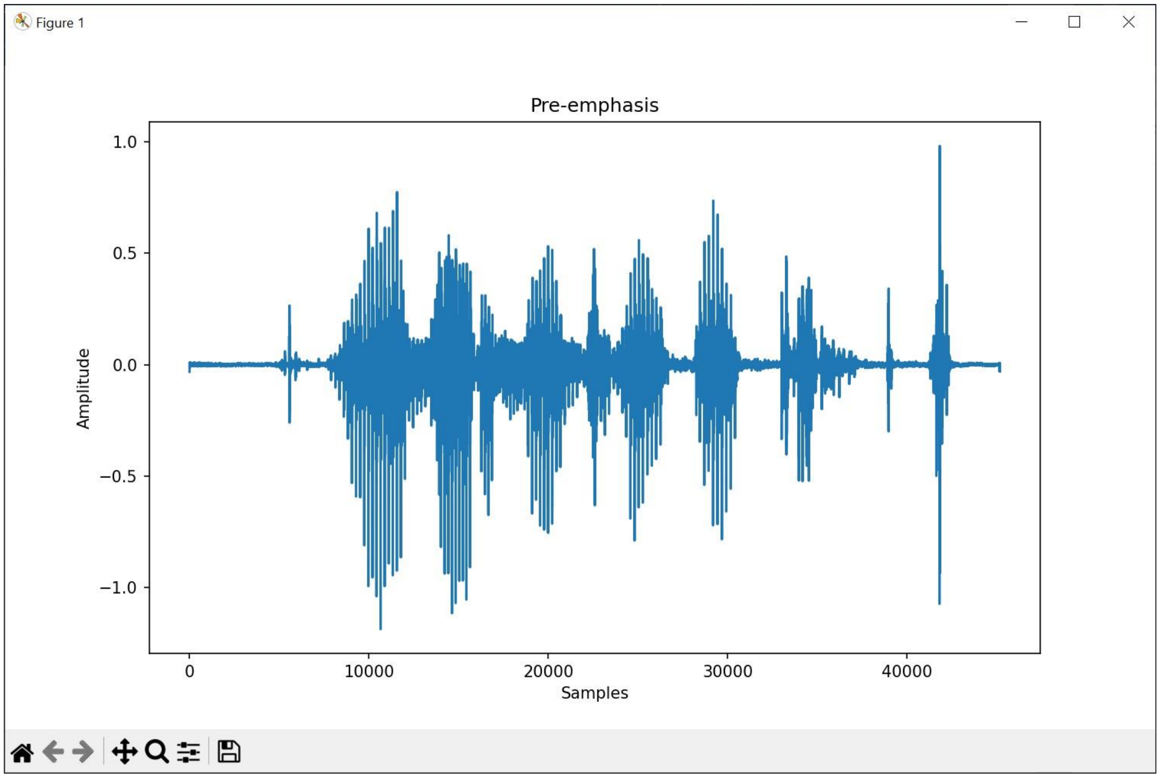


Figure 1
Pre-emphasis
Amplitude
Samples
1.0
0.5
0.0
−0.5
−1.0
0
10000
20000
30000
40000

**(a)**

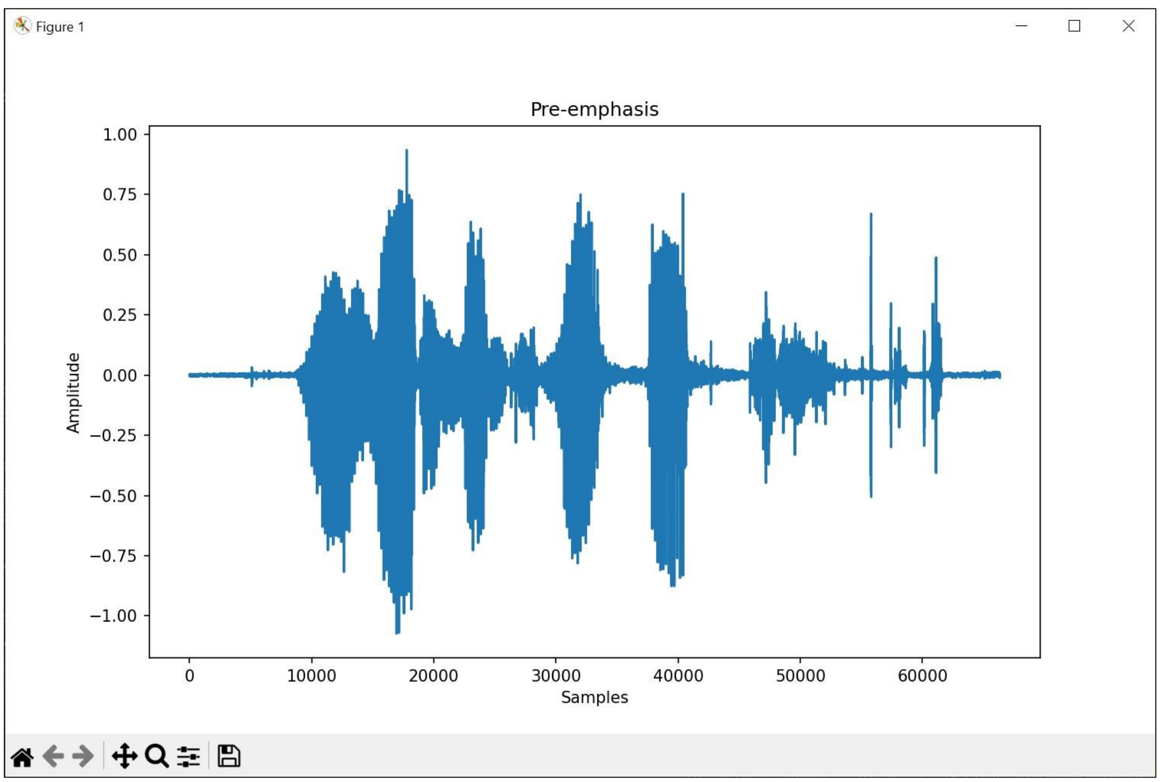
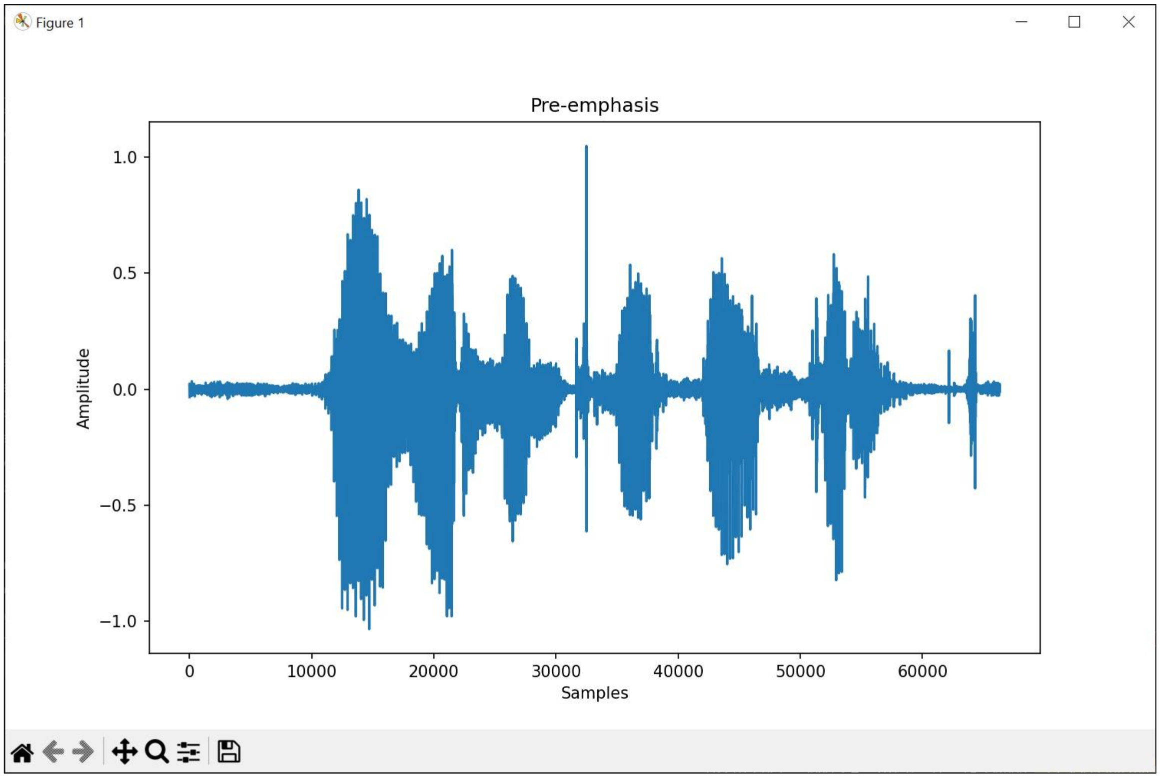


Figure 1
Pre-emphasis
Amplitude
Samples
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
10000
20000
30000
40000
50000
60000

**(b)**




**(c)**

**Fig. 3 Audio data from non-dysarthria subject**

### *2.2. MFCC Analyze*

Mel-Frequency Cepstral Coefficients (MFCC) is a prominent feature extraction approach in voice recognition [28]. The process of extracting MFCCs involves several stages:

#### *2.2.1. Pre-Emphasis*

This step uses a high-pass filter to boost higher frequencies in the audio signal [29]. It helps to balance the frequency spectrum and strengthens the subsequent processing processes. The equation used for the pre-emphasis is shown below:

$$S(n) = X(n) - a * X(n-1) \tag{1}$$

S(n) s(n) represents the sample output, x(n) represents the present sample, x(n-1) is the past sample, and a is value between 0.95 to 1.

#### *2.2.2. Framing and Overlapping*

The speech signal is divided into small overlapping frames (typically 20-40 milliseconds). This is done because the audio signal is considered to be quasi-stationary within short time windows [30]. When performing the Fourier transform, each frame is multiplied by a window function, often a Hanning or Hamming window, to minimize spectral leakage and edge effects.

$$S(n) = X(n) * W(n) \tag{2}$$

$$W(n) = 0.54 - 0.46\cos\left[\frac{2\pi n}{N-1}\right] 0 \le n \le N-1 \tag{3}$$

#### *2.2.3. Fast Fourier Transform*

Fast Fourier Transform (FFT) converts the windowed frames from the time domain to the frequency domain. This provides a frequency spectrum for each frame [31].

$$S(\omega) = fft\big(X(n)\big) \tag{4}$$

#### *2.2.4. Mel Filter Bank*

The Mel filter bank, which applies a number of filters spaced in accordance with the Mel scale, receives the frequency spectrum. A pitch perception scale called the Mel scale gauges how the human ear reacts to various frequencies. Below 1000 Hz, the frequency range of the Mel scale is linear, while above 1000 Hz, it is logarithmic.

$$F(mel) = 2595 * log_{10}\left(1 + \frac{f}{700}\right) \tag{5}$$

#### *2.2.5. Log and IDCT*

The output of the Mel filter bank is converted to a logarithmic scale to reflect the human auditory system's non-linear perception of amplitude. The log Mel spectrum is then transformed using the Inverse Discrete Cosine Transform

(IDCT).

*2.2.6. Energy*

The energy of all frames after IDCT is calculated. The result of the conversion carried out is called the Mel Frequency Cepstral Coefficient. The set of coefficients is called an acoustic vector. Therefore, every incoming speech data is converted into a series of acoustic vectors.

*2.2.7. Multilayer Perceptron*

MLP on Weka consists of a training set and a testing set. The training set is used to adjust classification parameters such as weights. Set testing is used to determine the overall performance of the classification. The backpropagation algorithm is used for network training [32]. Backpropagation is performed in several steps:

1. Input training data is to be computed, and the training output is to be computed.
2. For each output unit $k$
   $$\delta_k \leftarrow o_k(1 - o_k)(t_k - o_k) \quad (6)$$
3. For each hidden unit $h$
   $$\delta_h \leftarrow o_h(1 - o_h) \sum_{k \in outputs} \omega_{h,k} \delta_k \quad (7)$$
4. Update each network weight $\omega_{i,j}$
   $$\omega_{i,j} \leftarrow \omega_{i,j} + \Delta\omega_{i,j} \quad (8)$$

Where learning rate

$$\Delta\omega_{i,j} = \eta \delta_j \chi_{ij} \quad (9)$$

Momentum

$$\Delta\omega_{i,j}(n) = \eta \delta_j \chi_{ij} + \alpha \Delta\omega_{i,j}(n-1) \quad (10)$$

## 3. Results and Discussion

This section consists of research results divided into original voice data, extracting voice data characteristics, and identifying anomalies using artificial intelligence approaches.

### *3.1. Data Collections*

Feature extraction is the process of extracting unique variables that exist in data. In this study, the extraction of the sides was done using the MFCC kernel parameters. The parameters used are 16, 32, and 64. Figure 4 shows an example of a voice recording of 6 voiced data of dysarthria and non-dysarthria. It can be seen that in the process of extracting characteristics using MFCC, there is a difference between the classes of dysarthria and not-dysarthria.

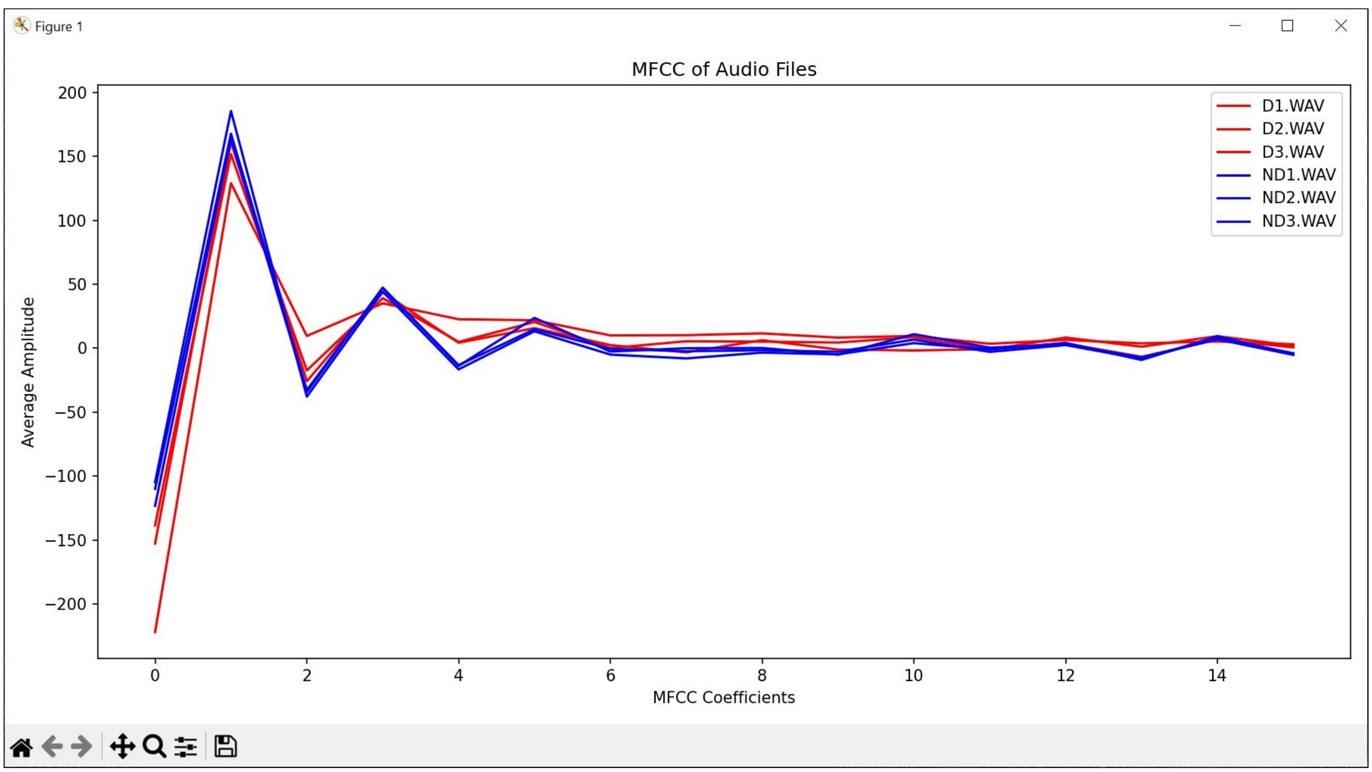



**Fig. 4(a) Feature extraction using 16 coefficient**

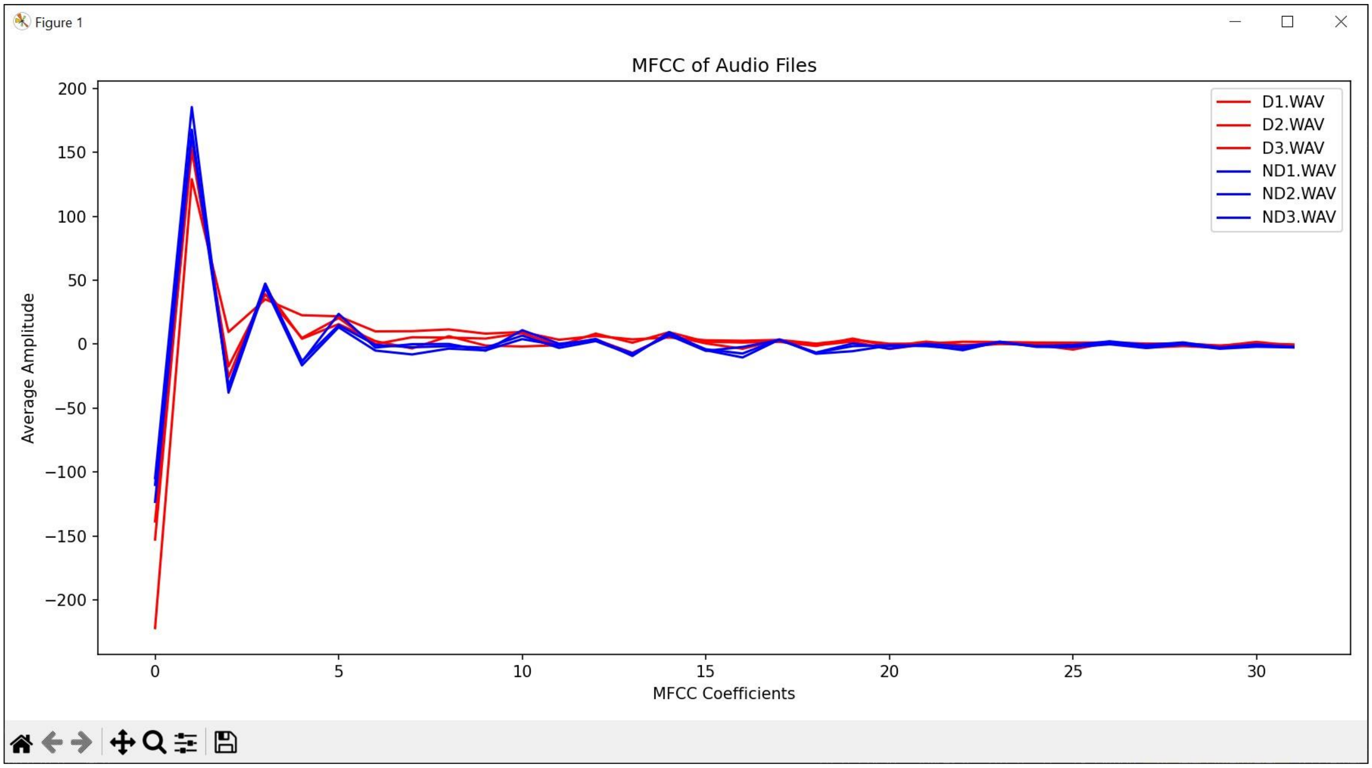



**Fig. 4(b) Feature extraction using 32 coefficient**

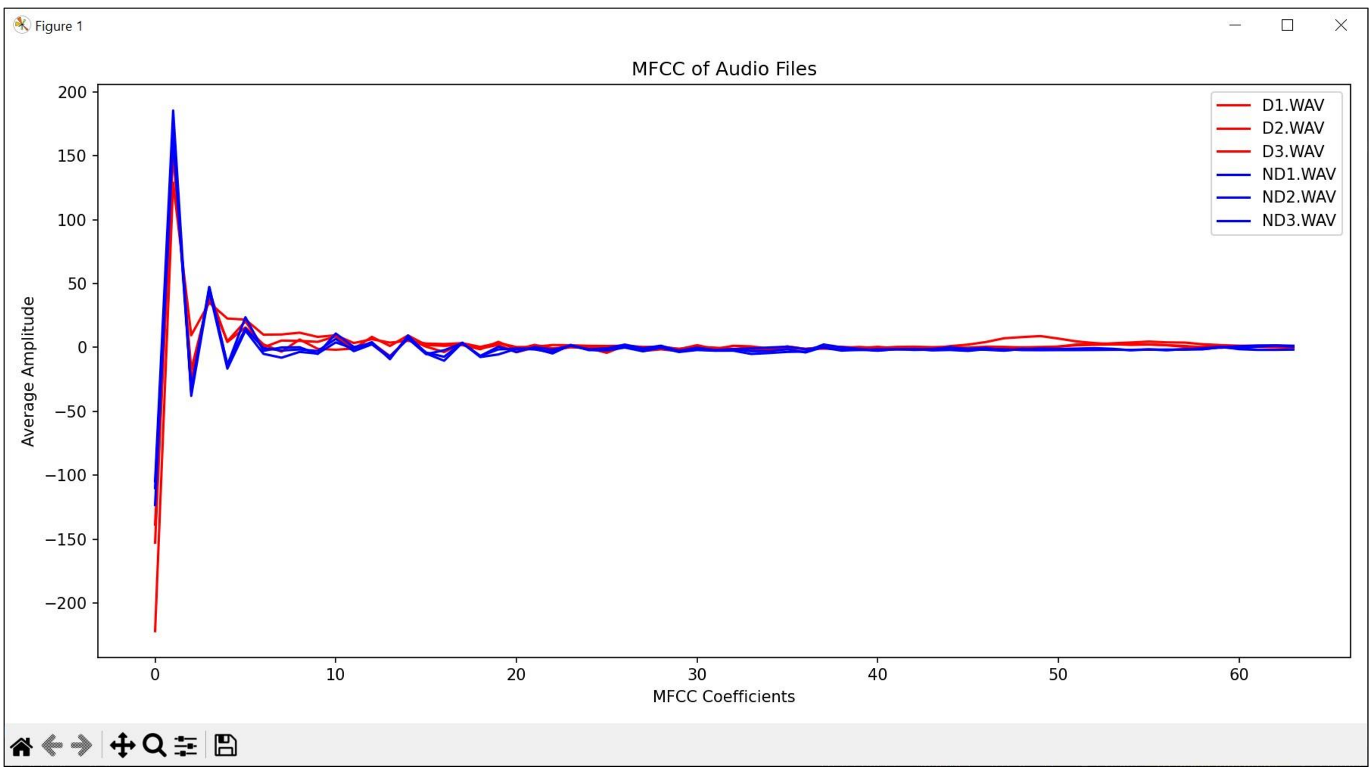



**Fig. 4(c) Feature extraction using 64 coefficient**

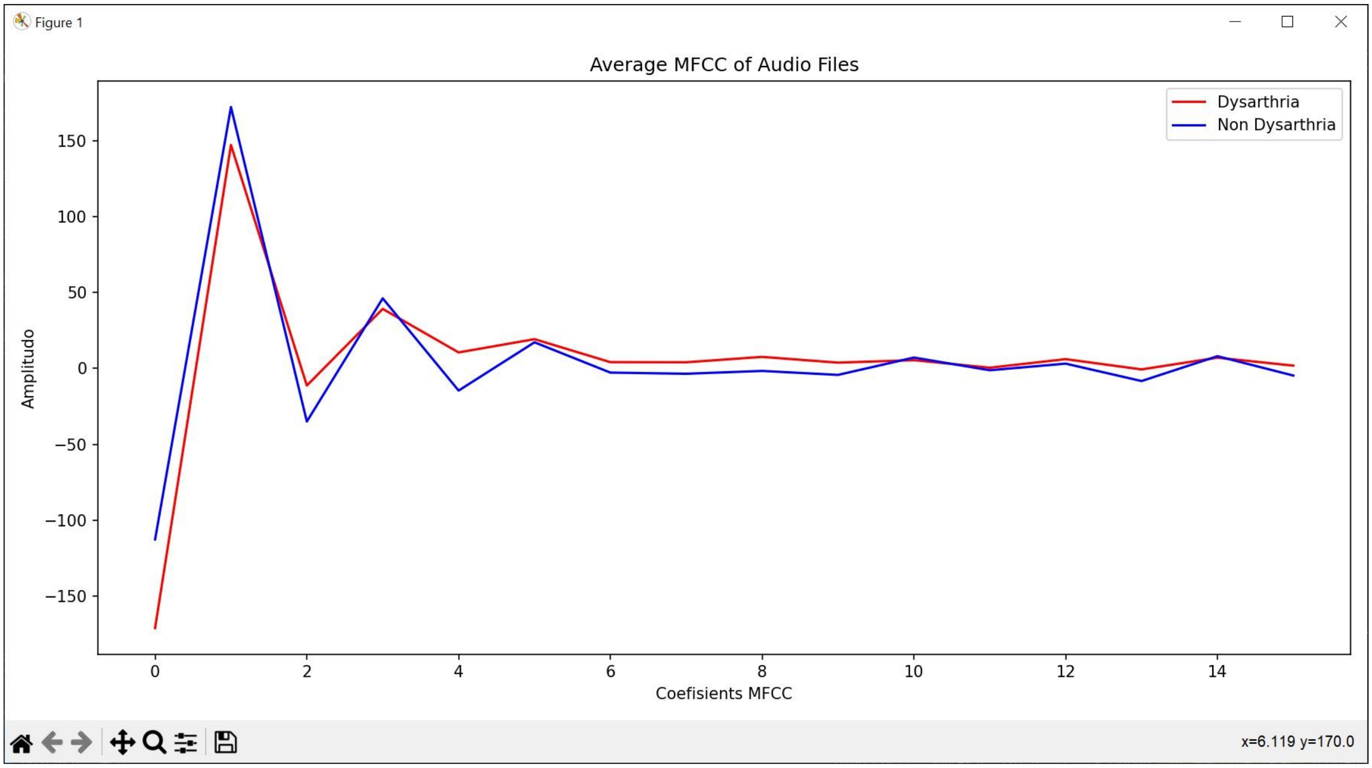



**Fig. 4(d) The average value of 16 coefficients**

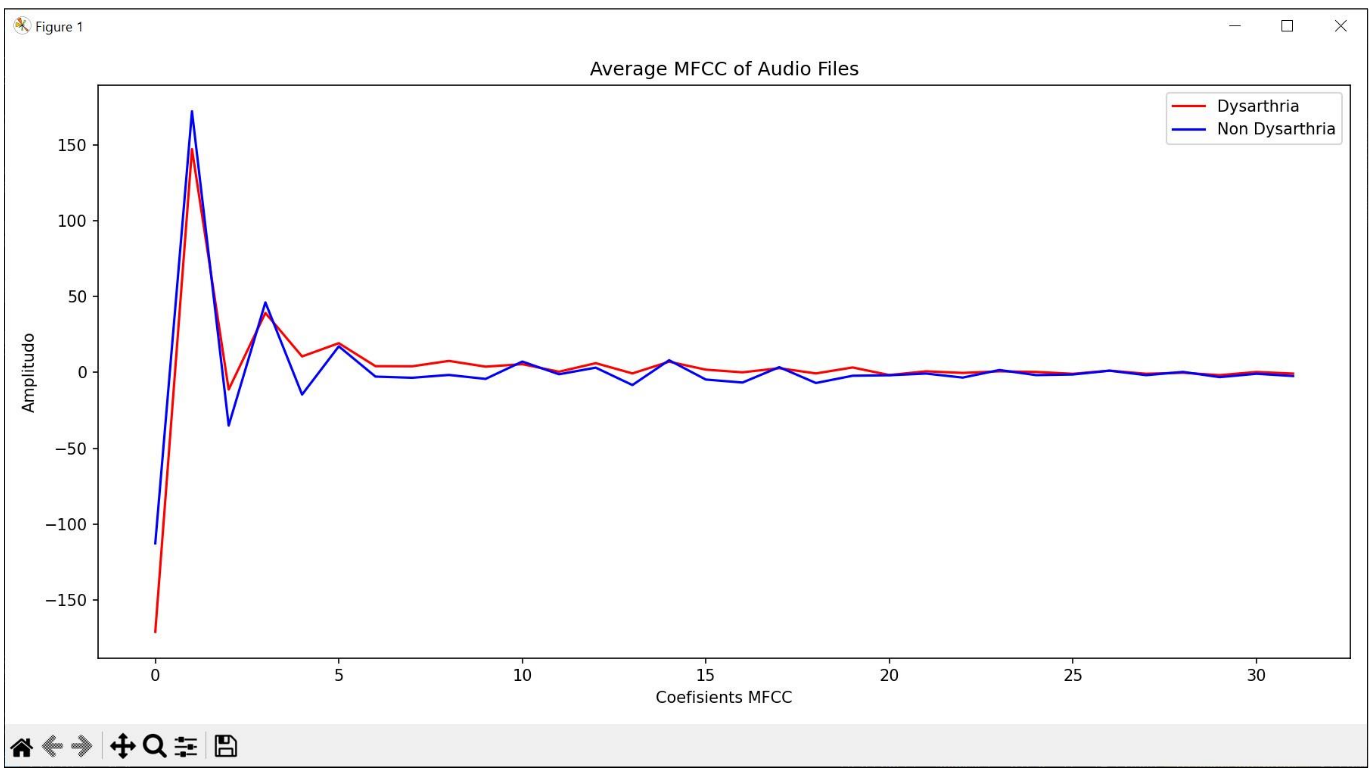



**Fig. 4(e) The average value of 32 coefficients**

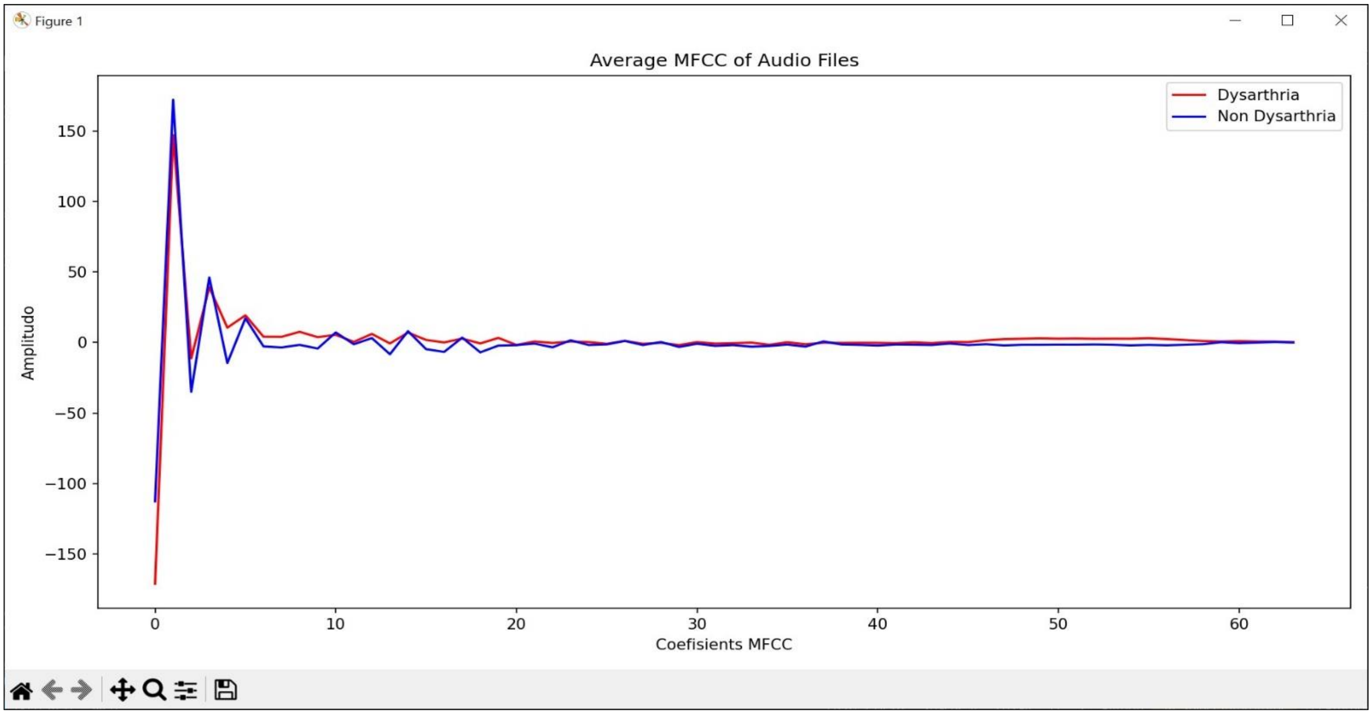
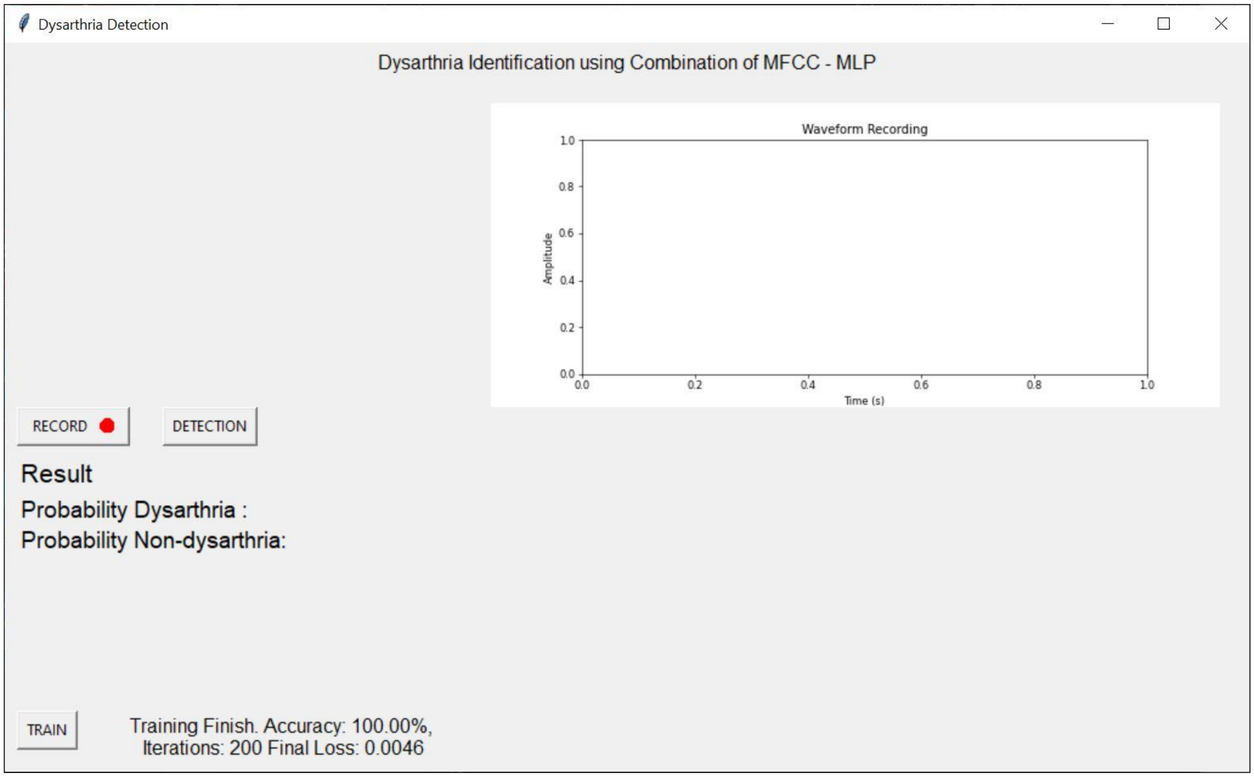


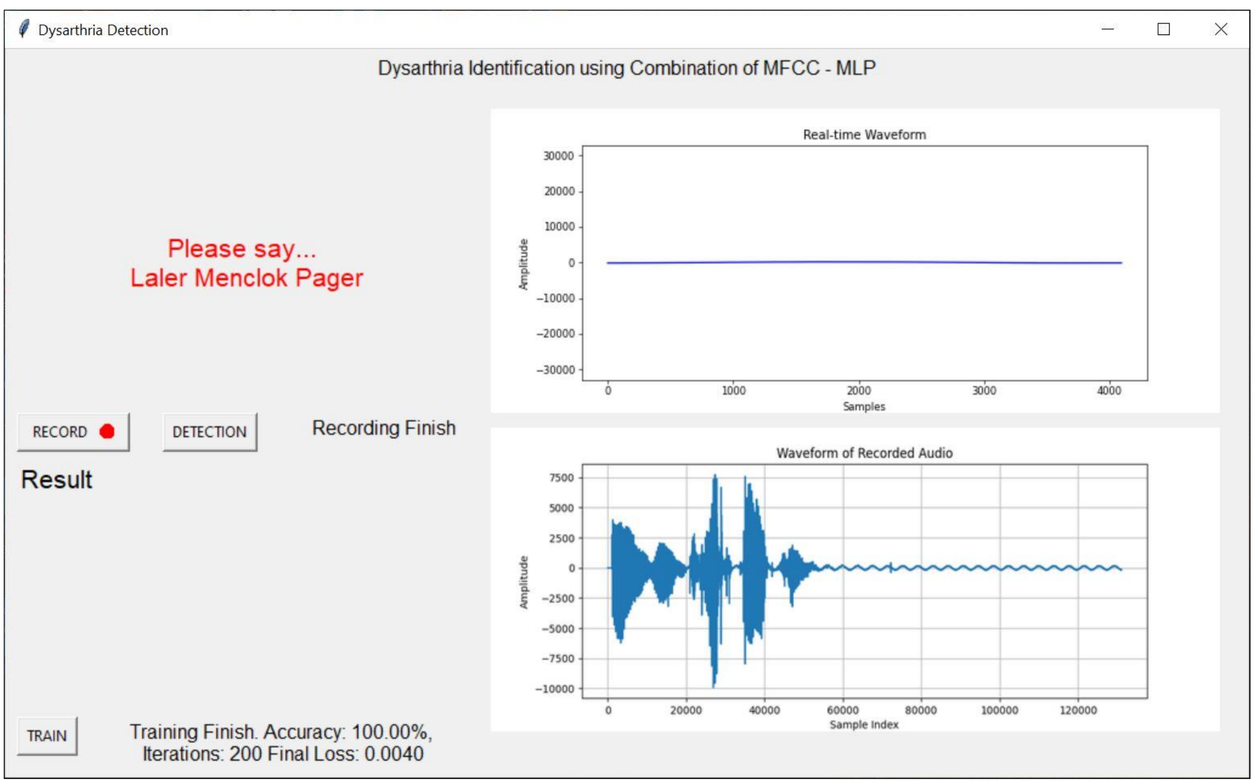

**Fig. 4(f) The average value of 64 coefficients**

**Table 4. MLP results**

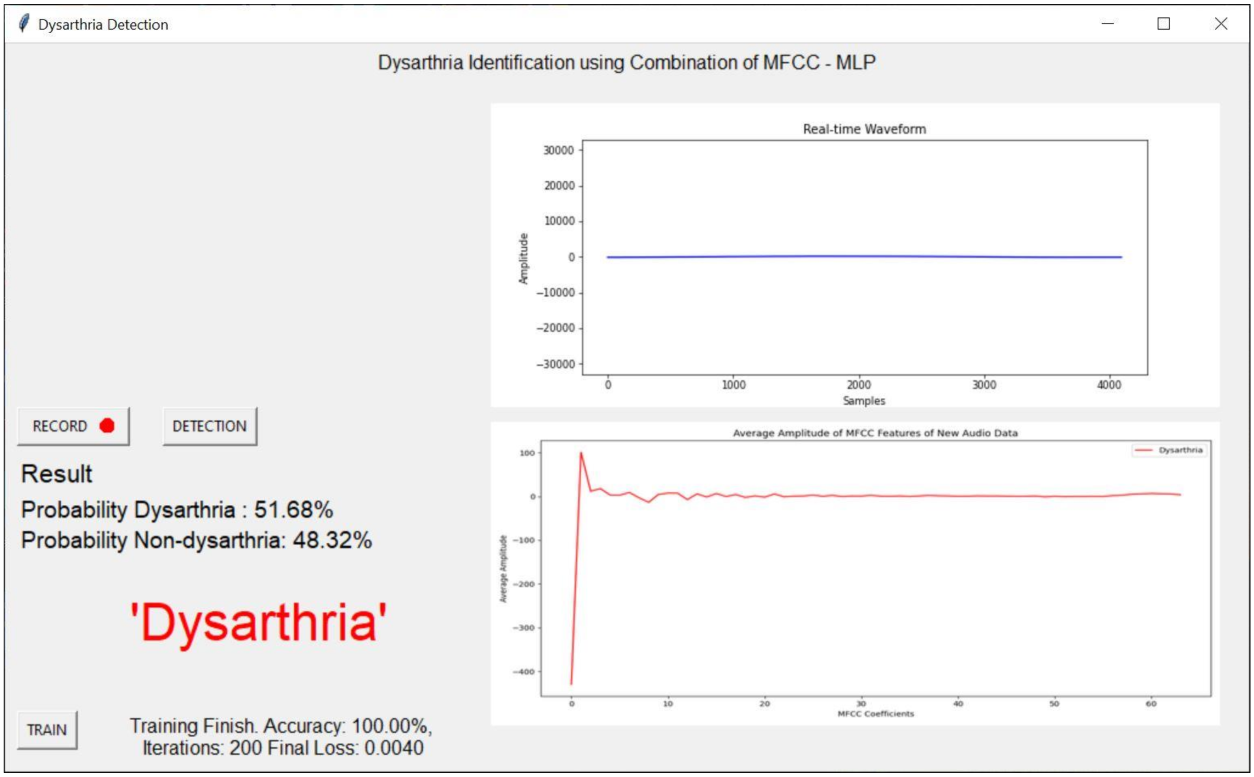
| Data (Training: Testing) | Neuron Input | Hidden Layers | Accuracy (%) |
|---|---|---|---|
| 80:20 | 16 | 8 | 100* |
| 80:20 | 16 | 16 | 100 |
| 80:20 | 16 | 32 | 100 |
| 50:50 | 16 | 8 | 97 |
| 50:50 | 16 | 16 | 97 |
| 50:50 | 16 | 32 | 97 |
| 20:80 | 16 | 8 | 96.875 |
| 20:80 | 16 | 16 | 95.625 |
| 20:80 | 16 | 32 | 94.375 |
| 80:20 | 32 | 16 | 100 |
| 80:20 | 32 | 32 | 100 |
| 80:20 | 32 | 64 | 100 |
| 50:50 | 32 | 16 | 97 |
| 50:50 | 32 | 32 | 97 |
| 50:50 | 32 | 64 | 97 |
| 20:80 | 32 | 16 | 96.875 |
| 20:80 | 32 | 32 | 96.25 |
| 20:80 | 32 | 64 | 96.25 |
| 80:20 | 64 | 32 | 100 |
| 80:20 | 64 | 64 | 100 |
| 80:20 | 64 | 128 | 100 |
| 50:50 | 64 | 32 | 100 |
| 50:50 | 64 | 64 | 100 |
| 50:50 | 64 | 128 | 100 |
| 20:80 | 64 | 32 | 98.75 |
| 20:80 | 64 | 64 | 100 |
| 20:80 | 64 | 128 | 100 |
| *best model | | | |


Dysarthria Detection
Dysarthria Identification using Combination of MFCC - MLP
Waveform Recording
Amplitude
Time (s)
RECORD
DETECTION
Result
Probability Dysarthria :
Probability Non-dysarthria:
TRAIN
Training Finish. Accuracy: 100.00%,
Iterations: 200 Final Loss: 0.0046

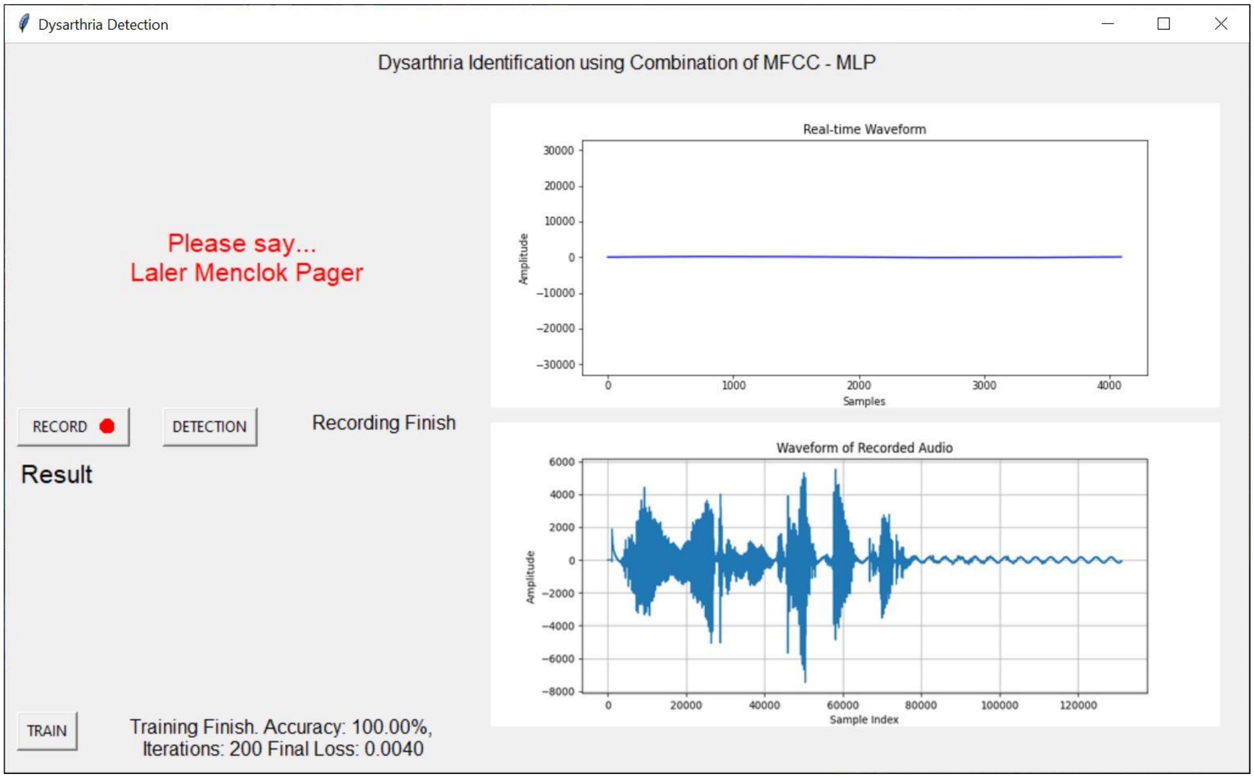
**(a)**


Dysarthria Detection
Dysarthria Identification using Combination of MFCC - MLP
Please say...
Laler Menclok Pager
Real-time Waveform
Amplitude
Samples
RECORD
DETECTION
Recording Finish
Result
Waveform of Recorded Audio
Amplitude
Sample Index
TRAIN
Training Finish. Accuracy: 100.00%,
Iterations: 200 Final Loss: 0.0040

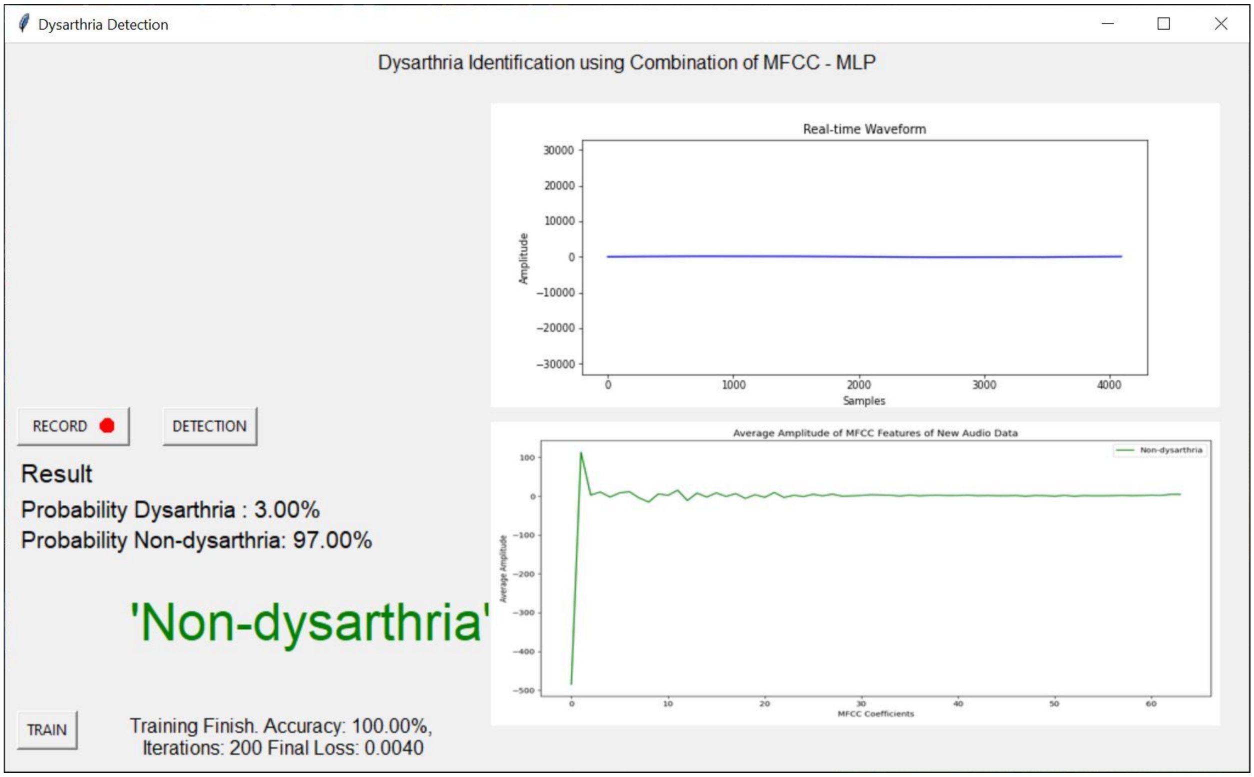
**(b)**


Dysarthria Detection
Dysarthria Identification using Combination of MFCC - MLP
Real-time Waveform
Amplitude
Samples
RECORD
DETECTION
Average Amplitude of MFCC Features of New Audio Data
Dysarthria
Average Amplitude
MFCC Coefficients
Result
Probability Dysarthria : 51.68%
Probability Non-dysarthria: 48.32%
'Dysarthria'
TRAIN
Training Finish. Accuracy: 100.00%,
Iterations: 200 Final Loss: 0.0040

**(c)**


Dysarthria Detection
Dysarthria Identification using Combination of MFCC - MLP
Please say...
Laler Menclok Pager
Real-time Waveform
Amplitude
Samples
RECORD
DETECTION
Recording Finish
Result
Waveform of Recorded Audio
Amplitude
Sample Index
TRAIN
Training Finish. Accuracy: 100.00%,
Iterations: 200 Final Loss: 0.0040

**(d)**




**(e)**

**Fig. 5 (a)-(e) Graphic user interface for implementation of dysarthria identification.**

**Fig. 6 Hardware implementation**

### *3.2. Multilayer Perceptrons*

The artificial intelligence approach in this study uses Multilayer Perceptron (MLP). Data extraction features that become outputs are used as learning materials in MLP. Within MLP, input neuron configurations and hidden layers imply the speed of data training. This study uses three types of input neurons, 16, 32 and 64, while hidden layers use configurations 8, 16 and 32. In connection with this, the separation of training and testing data is also used to compare values. Based on Table 4, the data division uses the divisions 80:20, 50:50, and 20:80. The accuracy level demonstrates the development of a system that performs effectively in detection using the given dataset. The findings of this study are consistent with earlier research using a more sophisticated artificial intelligence models [33].

### *3.3. Implementation*

The implementation of this dysarthria identification system is designed in the form of a Graphical User Interface (GUI) that provides an easily understandable and interactive display. In the initial GUI screen, as shown in Figure 5(a), the system is still in an empty state before any data input. After the data is entered, as illustrated in Figures 5(b) and 4(c), the data collection process begins with the phrase "*Laler Menclok Pager*," which is used to analyze speech patterns and detect the presence of dysarthria. The results of this classification process can be seen in Figures 5(d) and 4(e), which display the output as a classification category that has been processed, providing an overview of the severity level of dysarthria in the user based on the voice input provided. The research's findings also include hardware implementation. As shown in Figure 6, this hardware is combined with a microphone input that serves as input data for the MFCC and MLP-based system to identify.

## 4. Conclusion

The results show that R sound pronunciation anomalies or dysarthria can be identified using a feature extraction method integrated with multilayer perceptron. The degree of accuracy in the identification of dysarthria reaches 100% in the MLP network of 16 inputs and 8 hidden layers with 80%-20% data training-testing. Research in the future can implement models to hardware devices so that it can facilitate preventive action of R pronunciation deviations.

## Funding Statement

This work was funded by Grand Number: DRTPM Dikti 107/E5/PG.02.00.PL/2024; Universitas Ahmad Dahlan 001/PFR/LPPM UAD/VI/2024.

## Acknowledgements

The authors would like to express their gratitude to the Directorate General of Higher Education, Ministry of Education and Culture of the Republic of Indonesia, for the financial support and facilities provided under the basic research program scheme.

## References

[1] M. Bourqui et al., "The Encoding of Speech Modes in Motor Speech Disorders: Whispered Versus Normal Speech in Apraxia of Speech and Hypokinetic Dysarthria," *Clinical Linguistics and Phonetics*, pp. 1-22, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Sherine R. Tambyraja, Kelly Farquharson, and Laura M. Justice, "Phonological Processing Skills in Children with Speech Sound Disorder: A Multiple Case Study Approach," *International Journal of Language and Communication Disorders*, vol. 58, no. 1, pp. 15-27, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Iida Aakko et al., "Auditory-Perceptual Evaluation with Visual Analogue Scale: Feasibility and Preliminary Evidence of Ultrasound Visual Feedback Treatment of Finnish [r]," *Clinical Linguistics and Phonetics*, vol. 37, no. 4-6, pp. 345-362, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Elizabeth Roepke, "Assessing Phonological Processing in Children with Speech Sound Disorders," *Perspectives of the ASHA Special Interest Groups*, vol. 9, no. 1, pp. 1-21, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Hamza Kheddar, Mustapha Hemis, and Yassine Himeur, "Automatic Speech Recognition Using Advanced Deep Learning Approaches: A Survey," *Information Fusion*, vol. 109, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Wenyi Yu et al., "Connecting Speech Encoder and Large Language Model for ASR," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 12637-12641, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Pranav Kumar, Md. Talib Ahmad, and Ranjana Kumari, "HPO Based Enhanced Elman Spike Neural Network for Detecting Speech of People with Dysarthria," *Optical Memory and Neural Networks*, vol. 33, no. 2, pp. 205-220, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Kodali Radha, Mohan Bansal, and Venkata Rao Dhulipalla, "Variable STFT Layered CNN Model for Automated Dysarthria Detection and Severity Assessment Using Raw Speech," *Circuits, Systems, and Signal Processing*, vol. 43, no. 5, pp. 3261-3278, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] C. Zander et al., "Freiburg Neuropathology Case Conference: 68-Year-Old Patient with Slurred Speech, Double Vision, and Increasing Gait Disturbance," *Clinical Neuroradiology*, vol. 34, no. 1, pp. 279-286, 2024. [CrossRef] [Publisher Link]

[10] Huma Nasir, and Muhammad Arslan Zahid, "Chlorpromazine-Induced Neurological Symptoms Mimicking Stroke in an Elderly Patient with Intractable Hiccups: A Case Report," *Journal of Health and Rehabilitation Research*, vol. 4, no. 1, pp. 995-999, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] Malik Saad, Quratulain Maha, and Muhammad Talal, "Disseminated Salmonella Typhi Infection Presenting with Slurred Speech and Encephalopathy: An Unusual Presentation," *National Journal of Health Sciences*., vol. 9, no. 2, pp. 131-136, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Alan Wayne Jones, "Dubowski's Stages of Alcohol Influence and Clinical Signs and Symptoms of Drunkenness in Relation to A Person's Blood-Alcohol Concentration-Historical Background," *Journal of Analytical Toxicology*, vol. 48, no. 3, pp. 131-140, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Panduranga Vital Terlapu, and R. Prasad Reddy Sadi, "Real-time Speech-based Intoxication Detection System: Vowel Biomarker Analysis with Artificial Neural Networks," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 1637-1666, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Ran Zhou et al., "MFCC Based Real-Time Speech Reproduction and Recognition Using Distributed Acoustic Sensing Technology," *Optoelectronics Letters*, vol. 20, no. 4, pp. 222-227, Apr. 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Manjit Singh Sidhu, Nur Atiqah Abdul Latib, and Kirandeep Kaur Sidhu, "MFCC In Audio Signal Processing for Voice Disorder: A Review," *Multimedia Tools and Applications*, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Siba Prasad Mishra, Pankaj Warule, and Suman Deb, "Speech Emotion Recognition Using MFCC-Based Entropy Feature," *Signal, Image Video Processing*, vol. 18, no. 1, pp. 153-161, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Nur Aishah Zainal et al., "Integration of MFCCs and CNN for Multi-Class Stress Speech Classification on Unscripted Dataset," *IIUM Engineering Journal*, vol. 25, no. 2, pp. 381-395, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Wittaya Jitchaijaroen et al., "Machine Learning Approaches for Stability Prediction of Rectangular Tunnels in Natural Clays Based on MLP and RBF Neural Networks," *Intelligent Systems with Applications*, vol. 21, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Nikhil B. Gaikwad et al., "Hardware Design and Implementation of Multiagent MLP Regression for the Estimation of Gunshot Direction on IoBT Edge Gateway," *IEEE Sensors Journal*, vol. 23, no. 13, pp. 14549-14557, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] Amjad Alsirhani et al., "A Novel Approach to Predicting the Stability of The Smart Grid Utilizing MLP-ELM Technique," *Alexandria Engineering Journal*, vol. 74, pp. 495-508, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Qiang Gao et al., "Electroencephalogram Signal Classification Based on Fourier Transform and Pattern Recognition Network for Epilepsy Diagnosis," *Engineering Applications of Artificial Intelligence*, vol. 123, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[22] J. Naskath, G. Sivakamasundari, and A. Alif Siddiqua Begum, "A Study on Different Deep Learning Algorithms Used in Deep Neural Nets: MLP SOM and DBN," *Wireless Personal Communications*, vol. 128, no. 4, pp. 2913-2936, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[23] Ahmad Abbaskhah, Hamed Sedighi, and Hossein Marvi, "Infant Cry Classification by MFCC Feature Extraction with MLP and CNN Structures," *Biomedical Signal Processing and Control*, vol. 86, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[24] Yuanyuan Wei et al., "AE-MLP: A Hybrid Deep Learning Approach for DDoS Detection and Classification," *IEEE Access*, vol. 9, pp. 146810-146821, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[25] Joomee Song et al., "Detection and Differentiation of Ataxic and Hypokinetic Dysarthria in Cerebellar Ataxia and Parkinsonian Disorders Via Wave Splitting and Integrating Neural Networks," *PLoS One*, vol. 17, no. 6, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[26] Gennaro Tartarisco et al., "Artificial Intelligence for Dysarthria Assessment in Children with Ataxia: A Hierarchical Approach," *IEEE Access*, vol. 9, pp. 166720-166735, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[27] Adolfo M. García et al., "Cognitive Determinants of Dysarthria in Parkinson's Disease: An Automated Machine Learning Approach," *Movement disorders*, vol. 36, no. 12, pp. 2862-2873, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[28] Mayur R Gamit, and Kinnal Dhameliya, "Isolated Words Recognition Using MFCC, LPC and Neural Network," *International Journal of Research in Engineering and Technology*, vol. 4, no. 6, pp. 146-149, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[29] Amit Moondra, and Poonam Chahal, "Improved Speaker Recognition for Degraded Human Voice using Modified-MFCC and LPC with CNN," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 4, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[30] V. Anil Kumar, Ch. V. Rama Rao, and N. Leema, "Audio Source Separation by Estimating the Mixing Matrix in Underdetermined Condition Using Successive Projection and Volume Minimization," *International Journal of Information Technology*, vol. 15, no. 4, pp. 1831-1844, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[31] Lei Fan et al., "Accurate Frequency Estimator of Sinusoid Based on Interpolation of FFT and DTFT," *IEEE Access*, vol. 8, pp. 44373-44380, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[32] Sumam Sebastian, "Performance Evaluation by Artificial Neural Network Using WEKA," *International Research Journal of Engineering and Technology (IRJET)*, vol. 3, no. 3, pp. 1459-1464, 2016. [Google Scholar] [Publisher Link]

[33] Dong-Her Shih et al., "Dysarthria Speech Detection Using Convolutional Neural Networks with Gated Recurrent Unit," Healthcare, vol. 10, no. 10, pp. 1-14, 2022. [CrossRef] [Google Scholar] [Publisher Link]