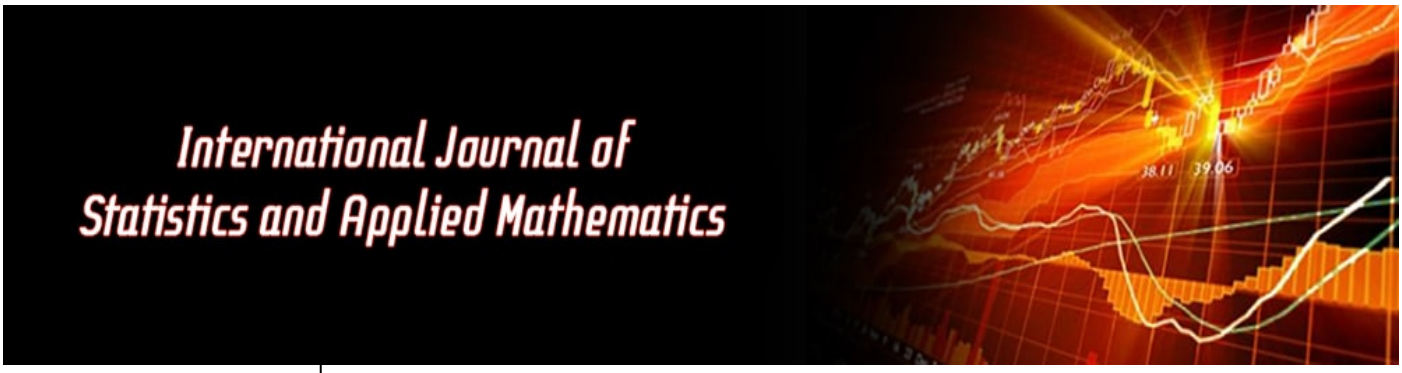

ISSN: 2456-1452
Maths 2023; 8(2): 13-18

https://www.mathsjournal.com
Received: 17-12-2022
Accepted: 20-01-2023

Fadya A Habeeb
Department of Mathematical, College of Education for Women, Tikrit University, Iraq

Kholood J Moulood
Department of Mathematical, College of Education for Women, Tikrit University, Iraq

Corresponding Author:
Fadya A Habeeb
Department of Mathematical, College of Education for Women, Tikrit University, Iraq

Speech recognition for Arabic numbers using empirical mode decomposition

Fadya A Habeeb and Kholood J Moulood

DOI: https://doi.org/10.22271/maths.2023.v8.i2a.940

Abstract
One of the main uses of artificial intelligence is automatic speech recognition. In this study, empirical mode decomposition is used to enhance the detection of speech dysarthria for speech recognition the problem in this paper is how to distinguish between the spoken words in Arabic we chose speech recognition by Empirical Mode Decomposition (EMD) to analyze the sound they were results as a confusion matrix results which indicates a classifier high accuracy (0.8966). Using 29 sounds to recognize by the EMD method the primary objective of this paper is to use EMD to Dysarthria speech recognize Arabic speech in real-time.

Keywords: Speech recognition, Automatic Speech Recognition (ASR), and Empirical Mode Decomposition (EMD), artificial intelligence

1. Introduction
Speech is the major form of communication between people, giving birth to the several subfields of speech processing that are now in use. Speech is an extremely non-linear, non-stationary signal, making it difficult to find this information. It is a signal generated by a highly developed psycho-acoustic system that humans have been using for thousands of years. Yet, it is more than just a tool for communication. It is a signal that contains a lot of information about the speaker, including their age, height, physical characteristics, general health, identity, mood, accent, and more [10]. Arabic is the official language of 22 countries and is spoken by more than 400 million people. It is regarded as being the fourth most widely spoken language online. Sharaf and Atwell MSA a variant of Arabic used in literary works and the Qur'an, which is written as well as used in official speeches, daily letters, and other informal situations. Machine Translation (MT) and Sentiment Analysis (SA) are the two tasks on which the majority of work has been done in SA over the past ten years [1]. A motor speech disease called dysarthria is caused by a number of abnormalities that affect the ability to control and carry out speaking movements. a loss of balance or a weakening of the speech-related muscles [6] .We used one of the applications of artificial intelligence automatic speech recognition to speech recognition of artificial intelligence empirical mode decomposition method. The problem in this paper is how to distinguish how to improve speech recognition for Arabic-speaking people who have dysarthria. We chose automatic speech recognition through the empirical mode decomposition method to solve that problem.

2. Automatic Speech Recognition Process
SR's principal objective is to make it possible for machines to hear, understand, and react to spoken information. Typically, the aim of ASR systems is to analyze, extract, classify, or recognize the information spoken by people. The four steps for the speech recognition system are: [9].

1. Speech analysis
2. 2-Feature extraction.
3. 3-Modeling.
4. 4-Testing

ASR is one of the most important uses of AI. The development of an ASR system for speech recognition necessitates the capture of speech signals, endpoint detection, feature extraction from speech signals, and a template matching algorithm. This post will discuss how to comprehend spoken Arabic numerals. Figure 1 displays the block diagram of the speech acquisition subsystem.

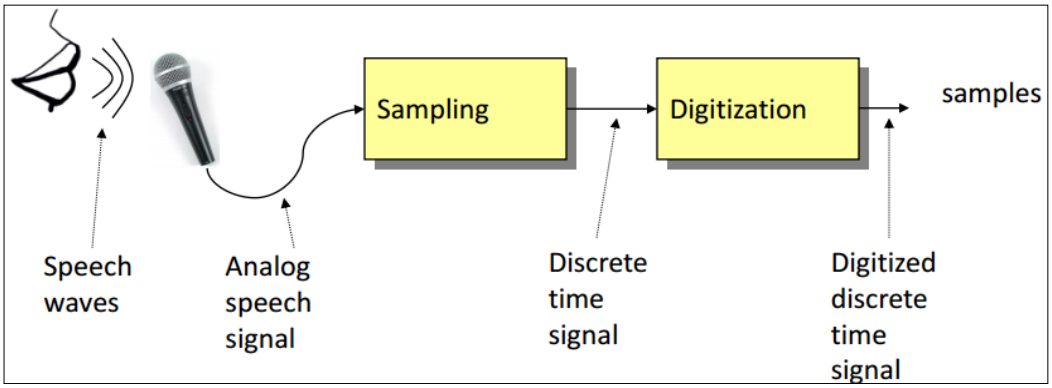

Fig 1: Speech Signal Acquisition [2]

To assure accurate recognition, Figure 2 depicts using signal energy in the endpoint detection process.

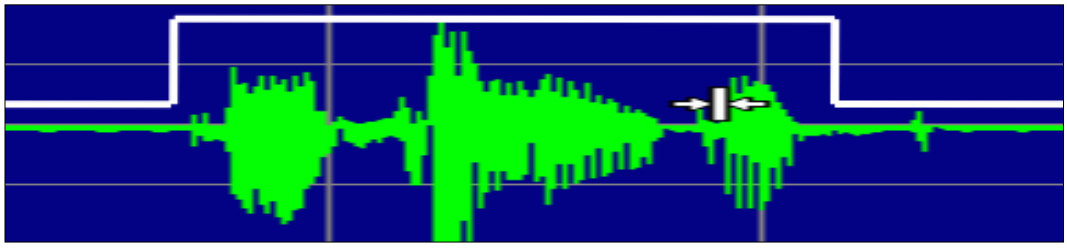

Fig 2: Endpoint detection [2]

3. Extraction of Features Using Empirical Mode Decomposition (EMD)

Described how to adaptively decompose non-stationary signals that use the EMD approach, which produces a residue and a set of intrinsic mode functions that represent the signal's numerous oscillations. The EMD technique is a very effective tool for signal analysis, de-noising, and filtering because EMD is a great visualization tool for the hidden sub-signals in a non-stationary signal [3].

Detrending as well as numerous other uses. EMD adaptive analysis is well suited for defect identification in industrial flat surface products because it can discover short-term changes like minute flaws in a picture. The image's various spatial frequencies are separated by the following IMFs. The highest spatial frequency component in a picture is represented by the first IMF. The following high spatial frequency component is represented by the second IMF, and so on. The trend in the image is represented by the residence as an illumination trend. Sifting method the following sifting algorithm [4] is used to produce IMFs and residue

1. Identify the extrema of the input signal image $I(f, m)$
2. Interpolate the local maxima $e_{max}(f, m)$ of the upper envelope U(I) using the cubic spline
3. Interpolate the local minima $e_{min}(f, m)$ of the lower envelope L(I) using the cubic splie
4. Calculate the mean envelop of both U(I) and L(I):

$$em(f, m) = \frac{e_{max}(f,m)+e_{min}(f,m)}{2}$$

5. Subtract the mean envelop from the image $I(f, m)$:
$h(f, m) = I(f, m) - em(f, m)$
6. If h satisfies the IMF condition stop sifting else $I(f, m) = h(f, m)$.

To reconstruct the original image: $I(f, m) = r_L(f, m) + \sum_{j=1}^{L} h_j(f, m)$, where L is the number of Intrinsic mode functions [12].

4. Related work

Mohamed Sidi Yakoub in addition to others in 2020. Used Convolutional Neural Network (CNN), EMD, and Hurst-based mode selection (EMDH), in combination with deep learning architecture to enhance the recognition of dysarthric speech This technique was employed as a preprocessing step to enhance the quality of dysarthric speech. The EMDH-CNN method increases accuracy by 20.72% and 9.95%, respectively, in terms of total accuracy [7].

Surbhi Bhatia and other individuals in 2022 Giving the ReLU activation function to the Convolutional Neural Network (CNN) model was recommended as a testing technique for the model with adjusted parameters. It thus proved difficult. Those who are blind or have low vision use a Braille display with a solenoid drive to control the Braille pattern. Braille text is 84% more accurately recognized when spoken Arabic numerals are present [5].

Hemant A. Patil and Prasad A. Tapkir in 2018. There were issues with the risk of spoofing attacks in that paper. A brand-new EMD Cepstral Coefficient (EMDCC) feature set was suggested. By combining the proposed feature set with the linear frequency modified group delay cepstral coefficient (LFMGDCC) at the score level, we are able to reduce the EER to 18.36% [11].

The technique for detecting speech activity was proposed in the paper and is based on Ensemble Empirical Mode Decomposition (EEMD) in conjunction with the dual-threshold approach, which incorporates EEMD's decomposition. Hence, the VAD algorithm's anti-noise performance can be enhanced by the generation of state aliasing [13].

For the purpose of locating wind turbine bearing faults, a feature extraction approach was published in the study. The whole ensemble empirical mode decomposition with adaptive noise is used to extract time-domain features (CEEMDAN). [14].

In the study, a technique for an HIM recognition system was suggested. The maximum Lyapunov exponent is used to extract the features of multi-modal signals, empirical mode decomposition (EMD) is used to preprocess multi-modal signals, and the threshold segmentation approach is used to segment motion (MLE). The results are then subjected to a comprehensive non-linear data analysis [15].

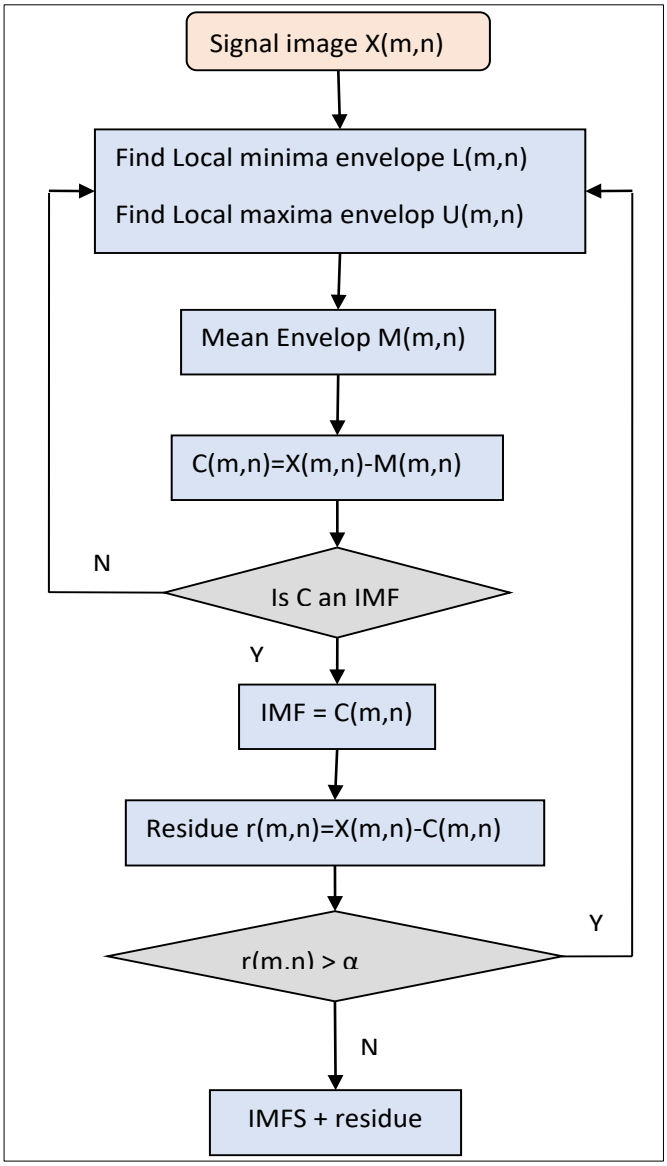

Fig 3: Flowchart of the Sifting Process [3]

5. Results and Discussion

We inserted the voice directly called Automatic speech recognition by Matlab language. The problem Patients with dysarthria speech produce sounds similar to those made by the first sound of a sound, however, it may be intermittent, intermittent, inaccurate, gasping, irregular, inaccurate, or monotonous, depending on the location of the damage Reading. In the work. The following figure 4, figure 5, and figure 6 shows the signal of each number. The entered data is saved in an array, and then the feature is extracted by EMD. The numbers that were entered are "اثنان" ,"واحد", and "ثلاثة", and after extracting the traits, we trained the data, and finally we tested the data. The number "ثلاثة" is the number with the most error rate because its letters are the most similar to the number "اثنان". We using 29 sounds to dividing into groups the first group represent the sound word (wahed) that numbers 10, the second group represents the sound (Ethneen) that numbers 10, and the third group represent the sound (Thalathah) that number 9. Access to the speech signal by direct audio recording at different times of the day, some in the morning when we wake up and some at other times when we are sick, or when we are in a state of joy or sadness. A comparison was made between the method used and two previous works, and we found that the method used has a better result than the two previous methods, which are Mohammed Sidi Yakoub, *et al*. (2020) [7] and Surbhi Bhatia *et al*. (2022) [5].

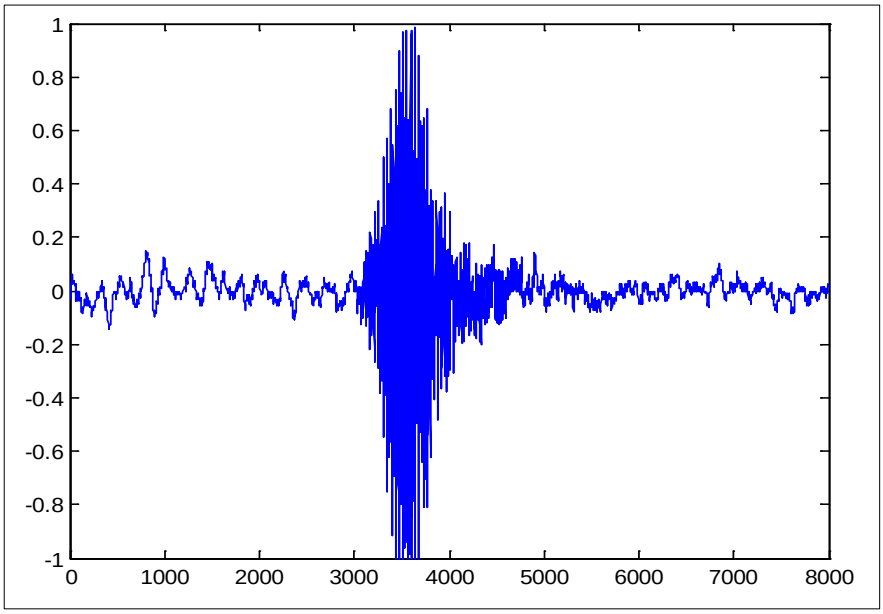

Fig 4: Wahid

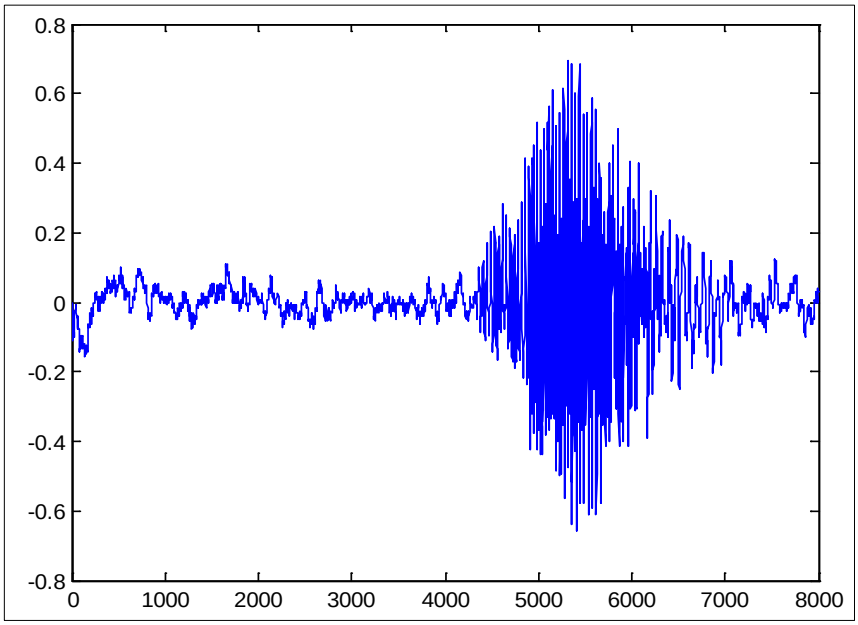

Fig 5: Ethneen

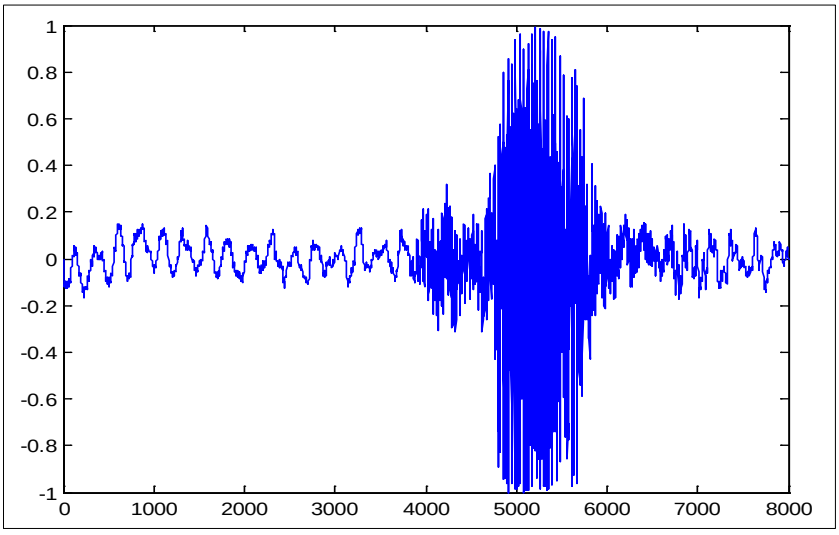

Fig 6: Thalatha

The proposed approach has been tested on a limited data set consisting of three words (10 Wahed, 10 Ethneen, and 9 Thalathah). EMD Decomposition of a spoken word into 11 IMFs As shown in the following figure(7).

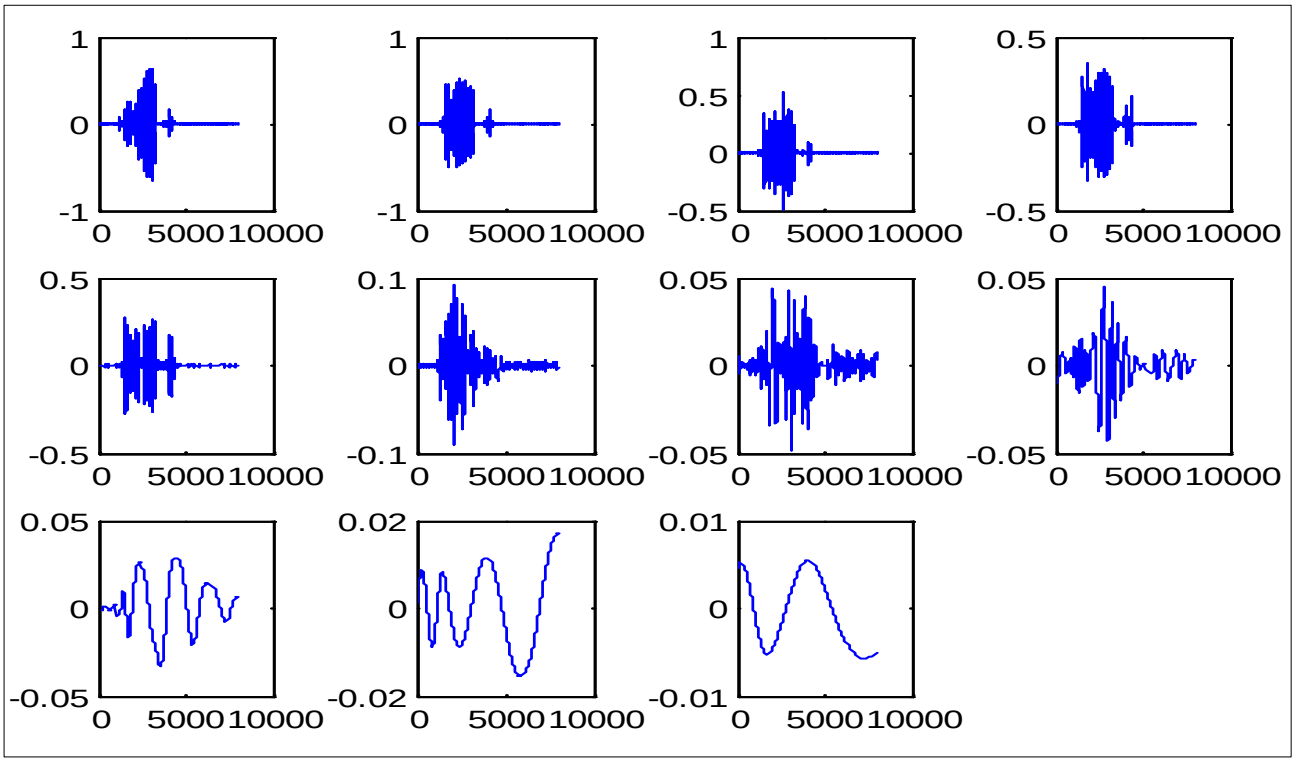

Fig 7: EMD Decomposition of a spoken word into 11 IMFs

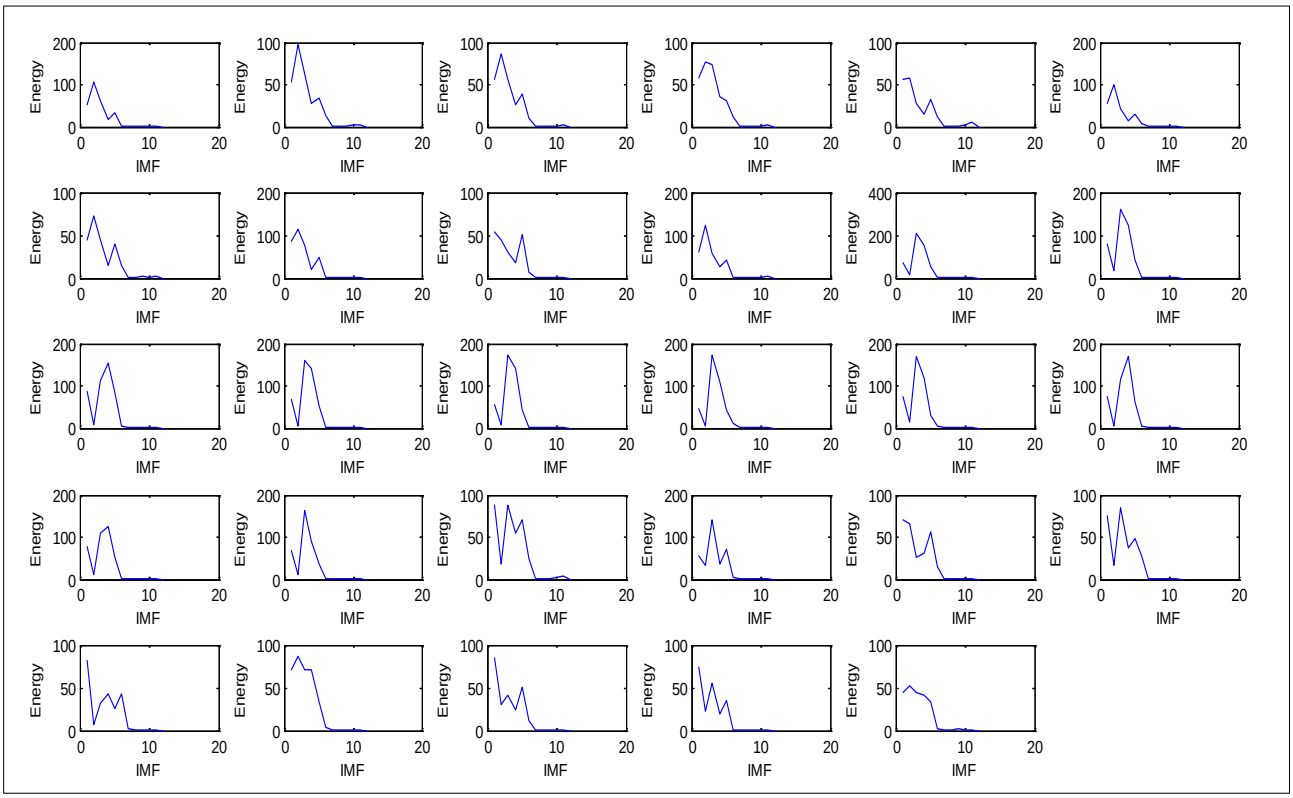

Fig 8: IMFs Energy Features extracted from the training set of 29 spoken words from three classes

To classify unseen spoken words a minimum distance classifier has been used according to the following formula: Assuming that z be the feature vector for the unknown input word, and let f_1, f_2, f_3 be templates (for Wahed, Ethnan and Thalatha) for the three classes. Then the error in matching z against f_k is given by $\| z - f_k \|$ .The vector u's norm is represented to here as $\| u \|$. A minimum-error classifier determines the class for which the error is minimum by computing $\| z - fk \|$ for k = 1 to 3. We refer to this as a minimum-istance classifier because $\| z - fk \|$ also represents the distance between z and fk. Obviously, a minimum-distance classifier is a template matching method. Testing Results for 29 sounds which belong to 3 spoken words. Figure (8) is summary that IMFs Energy Features extracted from the

training set of Summarizing the test results as a confusion matrix results which indicates a classifier high accuracy (0.8966).

6. Conclusion

The EMD speech technique is used to improve the quality of dysarthria speech one of the most difficult languages to distinguish between its words is Arabic, so we chose this problem and try to find the best solution. We look forward to expanding our work in the future by taking similar words and having different movements"كسرة" "ضمة" "فتحة" .Future work In fact opens up a new idea with Populistic Neural Networks (PNN) and we think that in the future can use PNN.

7. References

1. Guellil I, Saâdane H, Azouaoua F, Gueni B, Nouveld D. Arabic natural language processing: An overview, 2021 Journal of King Saud University: Computer and Information Sciences. 2021;33(5):497-507.
2. http://www.cs.cmu.edu/~bhiksha/courses/11756.asr/spring2014/lectures/Intro.class1.pdf.
3. Flandrin P, Rilling G, Goncalv`esv P. Empirical Mode Decomposition as a filter bank, IEEE Signal Processing Letters. 2004;11(2):112-114.
4. Yip L. Realtime Empirical Mode Decomposition for Intravascular Bubble Detection, BSc. Thesis EG4011/EG4012, James Cook University; c2010.
5. Bhatia S, Devi A, Alsuwailem RI, Mashat A. Convolutional Neural Network Based Real Time Arabic Speech Recognition to Arabic Braille for Hearing and Visually Impaired, Frontiers in Public Health. 2022;10:1-10. doi:0.3389/fpubh.2022.898355.
6. Enderby P. In Handbook of Clinical Neurology (110 ed.) Disorders of communication: Dysarthria (Elsevier B. V). 2013;110:273-281. https://www.sciencedirect.com/science/article/pii/B9780444529015000228. https://doi.org/10.1016/B978-0-444-52901-5.00022-8
7. Yakoub MS, Selouani S, Zaidi BF, Bouchair A. Improving dysarthric speech recognition using empirical mode decomposition and convolutional neural Network, EURASIP Journal on Audio, Speech, and Music Processing. 2020;2020(1):1-7.
8. Gulzar T, Singh A, Kumar D, Rajoriya, Farooq N. A Systematic Analysis of Automatic Speech Recognition: An Overview, International Journal of Current Engineering and Technology. 2014;4(3):1664-1675.
9. Sharmaa R, Vignolob L, Schlotthauerc G, Colominasc MA, Rufinerb LH, Prasanna SRM. Empirical Mode Decomposition for adaptive AM-FM analysis of Speech: A Review Speech Communication. Journal – Elsevier. 2017;88:39-64.
10. Tapkir P, Patil H. Novel Empirical Mode Decomposition Cepstral Features for Replay Spoof Detection, Hyderaba; c2018.
11. Demir B, Erturk S. Empirical Mode Decomposition Pre-Process for Higher Accuracy Hyperspectral Image Classification, IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium. 2008;2:939-941.
12. Shan M, Ablimit M, Hamdulla A. EEMD and Double Thresholds Integrated Voice Activity Detection. 2022 3rd International Conference on Pattern Recognition and Machine Learning (PRML). IEEE; c2022. p. 273-279.
13. Li H, Jiang Z, Shi Z, Han Y, Yu C, Mi X. Wind-speed prediction model based on variational mode decomposition, temporal convolutional network, and sequential triplet loss. Sustainable Energy Technologies and Assessments. 2022;52:101980.
14. Xue Y, Yu Y, Yin K, Li P, Xie S, Ju Z. Human In-hand Motion Recognition Based on Multi-modal Perception Information Fusion. IEEE Sensors Journal. 2022;22(7):6793-6805.