International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 11, pp 22-26, November-2023

https://doi.org/10.47001/IRJIET/2023.711004

Emotion Detection Using EEG and Voice Signals

Rustem Popa

Department of Electronics and Telecommunications, "Dunarea de Jos" University of Galati, Romania

***Abstract* - This paper explores the feasibility of creating a polygraph capable of discerning emotional states in individuals by analyzing skin impedance, EEG waves, and voice changes in response to "YES" and "NO" questions. Utilizing a laboratory MP36 BIOPAC Acquisition Unit with high-resolution A/D sampling, this polygraph employs three channels: Galvanic Skin Response (GSR), EEG signals, and audio input via a microphone. Emotion recognition is based on a majority voting circuit that compares the responses from all three channels. The variation in the amplitude of the skin conductivity is detected by comparing the derivative of the signal with a threshold value, and for the classification of EEG and vocal signals we use feedforward neural networks. To reduce the neurons in the input layers of the networks, the signals are processed using the Discrete Cosine Transform (DCT). Our findings reveal promising results in emotion detection via these multiple channels, offering potential applications in fields like human-computer interaction and emotional state monitoring.**

Keywords: polygraph, EEG signals, skin conductivity, Discrete Cosine Transform (DCT), neural networks, emotion detection, voice signals.

I. INTRODUCTION

A polygraph is a device that detects and records some electrophysiological signals or other psycho physiological variables of the human body, like the skin impedance, heart rate, respiratory rate or blood pressure. Although it is generally believed that polygraph is synonymous with lie detector, the literal meaning is "many measures" (poly - many, graph - write) [1]. In this paper, we will analyze the possibility of building a device of this kind, which can detect the emotions of the tested subject by measuring the impedance of the skin, the EEG waves and the changes in the voice when answering questions with "YES" or "NO".

Galvanic Skin Response and EEG signals were recorded using a laboratory MP36 BIOPAC Acquisition Unit using two channels with A/D sampling resolution of 24 bits, sample rate of 1000 samples/s, and the signal to noise ratio greater than 89 dB ([1], [2]). The third channel of the polygraph will process the signal recorded from a microphone. The final decision is made by comparing the answers of the 3 channels, using a majority voting circuit.

The EEG signals used in [2] were separated into known frequency bands: alpha (8-13 Hz), beta (13-30 Hz), theta (1-5 Hz) and delta (4-8 Hz). EEG signals were recorded from the parietal area (electrodes F4 and P4), during 60 seconds. These EEG signals have also been used in [2] and [5]. In this paper we will use the same signals collected in different emotional states, but without separating them into frequency bands. We will also use the Discrete Cosine Transform (DCT) to reduce the parameters that describe EEG signals, as was recommended in [3].

According to [4], EEG processing plays a vital role in emotion recognition. Emotions reflect human brain activity, and different emotional states have various EEG features. The commonly used methods for emotion recognition include Support Vector Machine (SVM), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Linear Discriminant Analysis (LDA), and Long Short-term Memory Neural Network (LSTM).

A method of change detection in EEG signals may be separation of the frequency bands of the EEG signals, computing the Root Mean Square (RMS) value inside that band, over a moving window, and thresholding. The threshold is set at the crossing of the two approximated Density of Probability Functions (DPFs), such as to minimize the discrimination error rate ([5]). The first time the method was tested on mice, so we are dealing with unsupervised discrimination. Various experiments on the modification of EEG signals in mice subjected to strong stress have been done also in [6]. Finally, reference [7] presents deep learning techniques in details, including deep belief networks, convolutional neural networks, and recurrent neural networks. The state-of-the-art applications of deep learning techniques for EEG emotion recognition are discussed in detail. The authors analyze the challenges and opportunities in this field and point out its future directions.

The most commonly used classification algorithms in recognition of speech are: Random Forest, Support Vector Machine (SVM) and K-nearest neighbour (KNN) ([8]). A new method to voice recognition is now being advised in [9], by coupling structured audio information with Long Short-term Memory Neural Network (LSTM). The features are extracted using Mel Frequency Cepstral Coefficients (MFCC) and Discrete Cosine Transform (DCT).

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 11, pp 22-26, November-2023

https://doi.org/10.47001/IRJIET/2023.711004

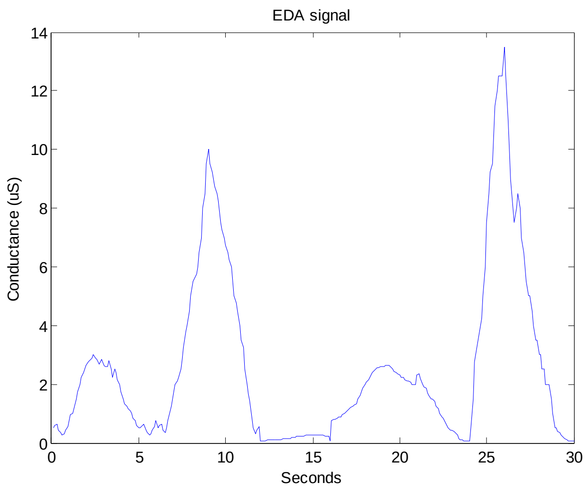

Figure 1: Variation of skin conductance in 30 seconds

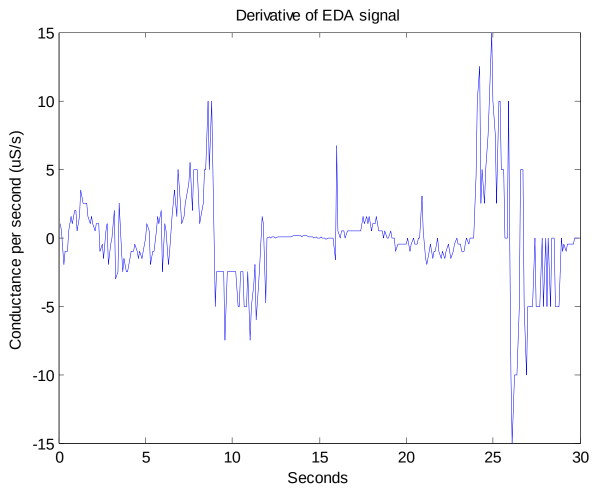

Figure 2: Derivative of the signal from Figure 1

The paper is organized as follows: section II describes the classification methods of the signals on these three channels, section III deals with results and discussions, and section IV concludes the paper.

II. CLASSIFICATION METHODS

In this section, we will present separately for each of the three channels the proposed solution for implementing the polygraph.

2.1 Classification based on skin conductivity

The human skin displays several forms of bioelectric phenomena, especially in the extremities, like the fingers or palms of the hands. When a week electric current is applied between two electrodes, the electrical resistance between them is called Galvanic Skin Resistance (GSR) and the value of this resistance depends on the emotional state of the subject. If these electrodes are connected to a voltage amplifier, but without any externally applied current, the voltage measured between them is called Galvanic Skin Potential (GSP), which also depends on the subject's emotional state. The combined changes in the GSR and GSP represent the Galvanic Skin Response, and this term was replaced with Electrodermal Activity (EDA), used to describe changes in the skin's ability to conduct electricity [1].

The signal represented in Figure 1 illustrates the variation of skin conductance between two electrodes placed on the fingers. Emotion causes a slight increase in perspiration on the skin, which decreases the electrical resistance between the two electrodes (that is, the conductance expressed in microsiemens (μS) increases).

A sudden variation of the conductance can be detected by estimating the slope of the increase or decrease of the signal, and the value of this slope is given by its derivative.

Figure 2 shows the derivative of the signal from Figure 1. We note here that for the derivative of the signal we can set the value of 10 μS/s, taken in module, i.e. + or – 10 μS/s, or even another value, depending on the experimental results for the type of signal. This threshold value can be a fixed value from the beginning, or it can be adapted to the signal using an optimization algorithm.

2.2 Classification of EEG Signals

For the discrimination of EEG signals in the two classes that interest us, we will use a feedforward neural network with a single hidden layer. The network entry data will be the DCT coefficients of the EEG signals, because they contain in a small number of samples all the essential information from the signal [3]. We will choose a number of 10 input neurons, each input value being an average of a number of samples, and 2 output neurons decide the membership classes for the input signals [2].

The EEG signals are recorded for 60 seconds, and we will divide them into intervals of 500 ms to obtain 120 sets of data for each signal. With these data we will train and test the neural network. A fragment of 500 samples from the EEG signal in a state of relaxation is represented in Figure 3, and the DCT transform of this signal is represented in Figure 4. The 10 inputs to the neural network are obtained by averaging each group of 5 samples from the first 50 samples of the DCT transform. We notice from Figure 4 that they contain the essential information from the signal.

The database thus obtained has two matrices: one consisting of 10 rows and 240 columns (rows are the coefficients of discrete cosine transformation, and the columns are given by the 240 signals of 500 ms each) and a binary matrix of 2 rows and 240 columns representing the output values. Network training is done with a set of 204 signals, and testing with the remaining 36 signals [2].

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 11, pp 22-26, November-2023

https://doi.org/10.47001/IRJIET/2023.711004

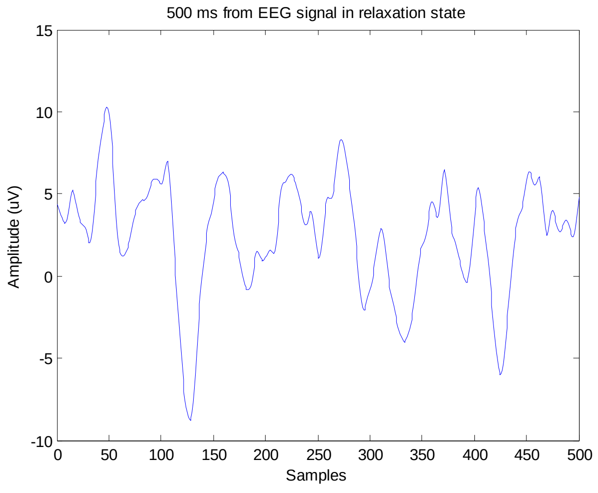

Figure 3: A 500 ms fragment of the EEG signal in a state of relaxation

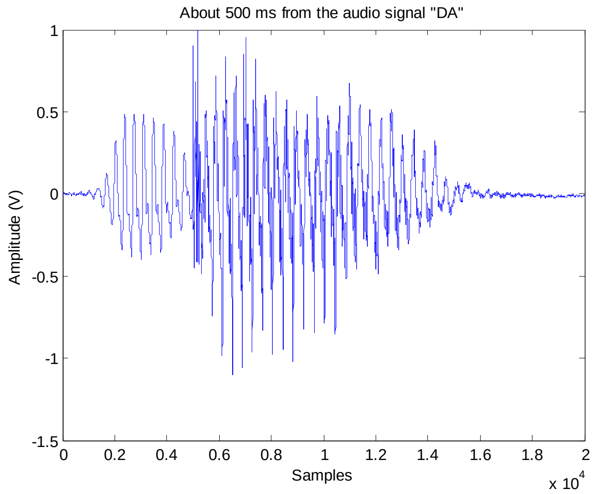

Figure 5: The "DA" signal pronounced without emotion

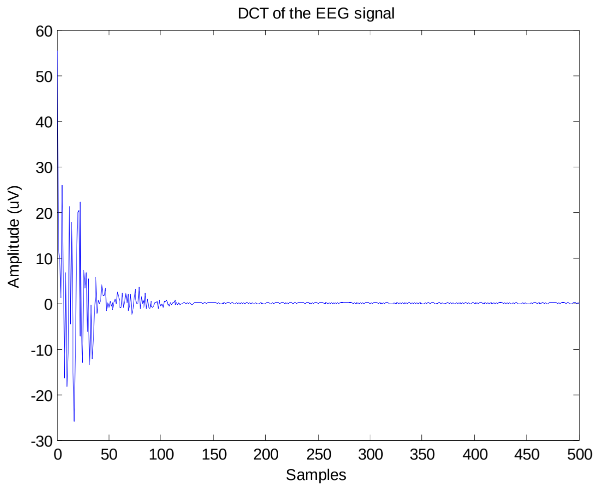

Figure 4: DCT transform of the signal from Figure 3

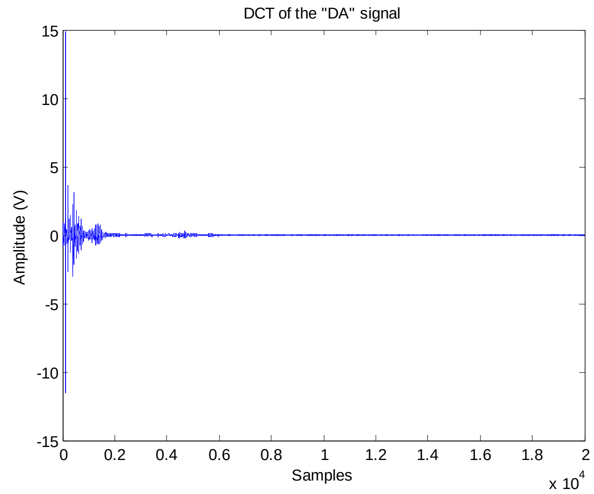

Figure 6: DCT transform of the "DA" signal

2.3 Classification of voice signals

We will also use a feedforward neural network with a hidden layer that will have a variable number of neurons. The input data in the network will be also the coefficients of the DCT transform of the "DA" and "NU" audio signals (that is, the words "YES" and "NO" in Romanian). Figure 5 shows the signal generated by the pronunciation of the word "DA" in about 500 ms, and Figure 6 contains the DCT transform of this signal.

For the best possible result, we should have as many example sets as possible for training the network. Unfortunately, we had a small database and limited ourselves to only 40 words pronounced with emotion and 40 words pronounced without emotion, both for "YES" and for "NO". The sampling frequency is 44100 Hz, standard for recording audio signals. Pronouncing a "YES" or "NO" word takes about 500 ms, that is, about 20,000 samples. To have a reduced number of features, we calculated the DCT transform for each word.

We obtain a total of 80 data sets with which we train the network. With some of them we will test the response of the trained network for this data set.

The signal energy is contained approximately in the first 2000 samples of the DCT transform. We will consider only 100 coefficients obtained by averaging groups of 20 neighboring samples from the 2000 representative samples. The values of these averaged samples will be inputs to the neural network, so we set a number of 100 input neurons. In the hidden layer we will consider a variable number of neurons, from 5 to 20, and at the output we will have 2 neurons, one for each of the 2 classes.

For each of the two words "YES" and "NO" we generate a database which contain 2 matrices, the first consisting of 100 lines and 80 columns (the number of input neurons and the number of training examples, respectively), and the next matrix having 2 lines and 80 columns, the value on each line establishing the degree of belonging to one of the two classes.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 11, pp 22-26, November-2023

https://doi.org/10.47001/IRJIET/2023.711004

III. RESULTS AND DISCUSSIONS

The results of the classification of EEG signals are represented in Figures 7 and 8. Figure 7 shows the performance evolution with the number of epochs in the 3 situations: network training, testing and validation, the minimum error of $4.936 \cdot 10^{-8}$ being obtained at iteration 56. Figure 8 shows the confusion matrix indicating the percentage result of the classification in the two classes. The two classes are given by the state of relaxation, when the subject tells the truth, and the state of frustration, when the subject lies. We notice that the sum of the percentages in the squares marked with green (also found in the blue square), of 94.4%, represents the number of correct classifications for our set of signals with 5 neurons in hidden layer.

A similar analysis is done for the neural network intended for the classification of audio signals. Figure 9 shows the performance evolution with the number of epochs in the 3 situations: network training, testing and validation, the minimum error of 0.081983 being obtained at iteration 24. The classification performance for the "DA" signal varies from 66.67% for 5 neurons in the hidden layer, to 83.33% for 10 neurons and to 87.50% for 15 neurons. The volume of data is unfortunately much too small for proper network training. On the other hand, better results are obtained if we use Mel Frequency Cepstral Coefficients (MFCC) as input data instead of time samples. In future experiments we will increase the database with greater attention to the pronunciation of words and use MFCC as input data. Almost identical results are obtained in the case of pronouncing the word "NU".

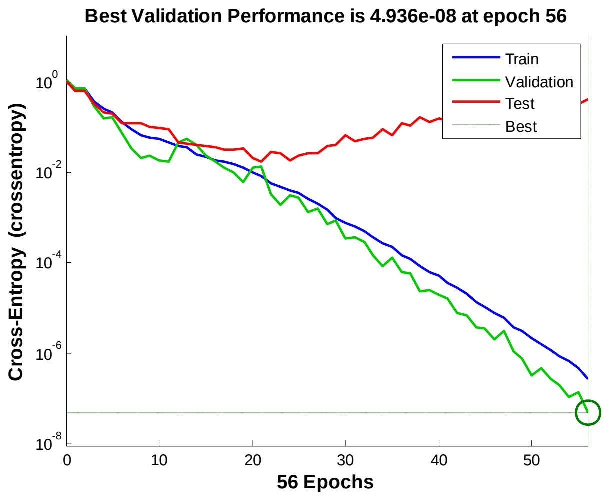

Figure 7: Best validation performance of the network

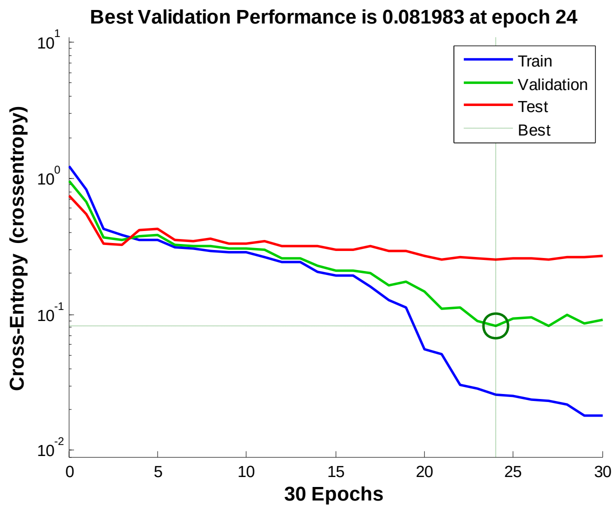

Figure 9: Best validation performance of the network

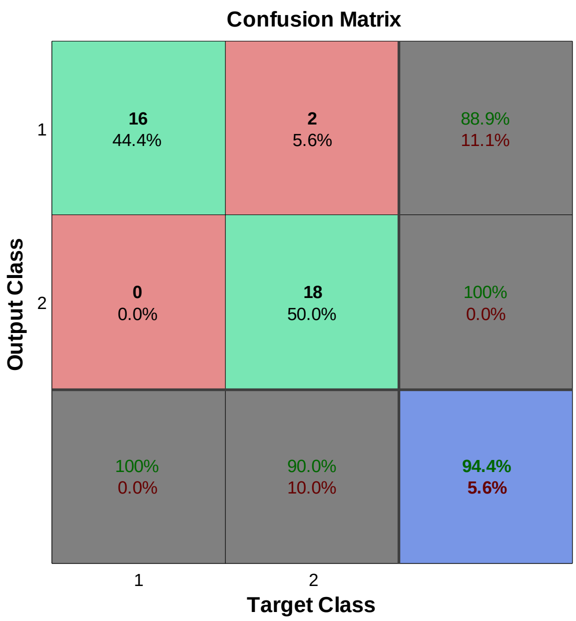

Figure 8: Confusion matrix for 5 neurons in hidden layer

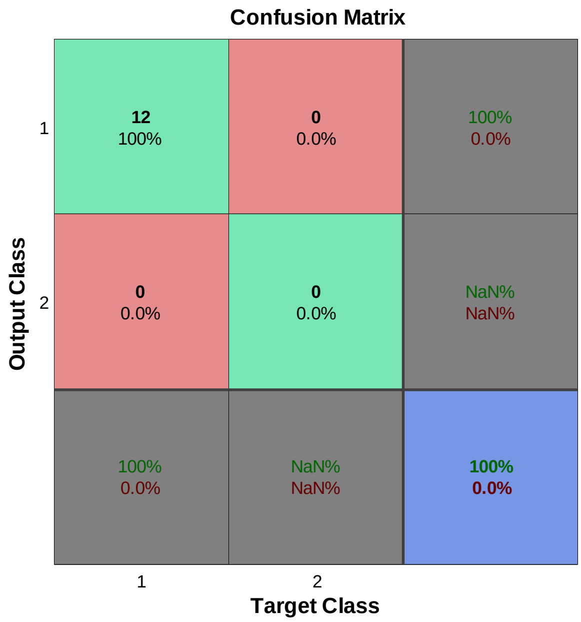

Figure 10: Confusion matrix for 5 neurons in hidden layer

ISSN (online): 2581-3048
Volume 7, Issue 11, pp 22-26, November-2023
https://doi.org/10.47001/IRJIET/2023.711004

IV. CONCLUSION

In this study, we have explored the development of a multi-channel polygraph capable of detecting and classifying emotional states by analyzing Galvanic Skin Response (GSR), EEG signals, and voice responses to "YES" and "NO" questions. Our investigation has yielded valuable insights and promising results in the domain of emotion recognition and its potential applications. The combination of these channels allows for a more robust and comprehensive assessment of an individual's emotional state, capturing both physiological and psychological aspects.

The study acknowledges certain limitations, including the relatively small database size and the challenges in distinguishing emotions with subtle differences in voice responses. These limitations provide opportunities for further research and improvements in the future. The findings in this paper have significant implications for various fields, including human-computer interaction interfaces, lie detection, and emotional state monitoring. The multi-channel polygraph has the potential to enhance user experience in interactive systems and improve the accuracy of emotion detection.

In conclusion, our research demonstrates the feasibility of a multi-channel polygraph for emotion detection, by integrating multiple channels, coupled with advanced classification techniques, offers a promising avenue for accurate emotion recognition. As technology continues to advance, the potential applications of such a device are vast, and this study lays the foundation for further research and development in this field.

REFERENCES

[1] R. Pflanzer, and W. McMullen, "BIOPAC Student Lab, Lesson 9, Electrodermal Activity and Polygraph", *BIOPAC Systems Inc., CA, USA*, 2019.

[2] R. Popa, "Emotion Detection using EEG Signals", *Int. Journal of Progressive Sciences and Technologies*, vol. 24, no. 1, pp. 175-182, 2020.

[3] D. Birvinskas, V. Jusas, I. Martisius, and R. Damasevicius, "EEG Dataset Reduction and Feature Extraction using Discrete Cosine Transform", 2012 *UKSim-AMSS 6th European Modelling Symposium*, pp. 199-204, DOI 10.1109/EMS.2012.88.

[4] C. Yu, M. Wang, "Survey of emotion recognition methods using EEG information", *Cognitive Robotics*, vol. 2, pp. 132-146, 2022.

[5] L. Frangu and R. Popa, "Change Detection in EEG Signals," in *6th International Symposium on Electrical and Electronics Engineering* (ISEEE), Galati, Romania, 2019, pp. 1-6, doi: 10.1109/ISEEE48094.2019.9136145.

[6] R. Douglas Fields, "Electric Brain. How the New Science of Brainwaves Reads Minds, Tells Us How We Learn, and Helps Us Change for the Better", BenBella Books Inc., Dallas, TX, USA, 2020.

[7] X. Wang, Y. Ren, Z. Luo, W. He, J. Hong, and Y. Huang, "Deep learning-based EEG emotion recognition: Current trends and future perspectives", *Frontiers in Psychology*, 14:1126994, 2023, doi: 10.3389/fpsyg.2023.1126994.

[8] X. Wang, "Emotion Recognition in Human Voice Speech Based on Machine Learning", in Y. Chen et al. (Eds.): ICMETSS 2022, ASSEHR 693, pp. 149–157, 2023.

[9] R. Leelavathi, S. Aruna Deepthi, V. Aruna, "Speech Emotion Recognition using LSTM", *International Research Journal of Engineering and Technology*, vol. 9, issue 1, pp. 586-594, 2022.

AUTHOR'S BIOGRAPHY

Rustem Popa received the B.S. degree in electronics engineering from Bucharest Politehnica University in 1984, and Ph.D. degree at "Dunarea de Jos" University in Galati, Romania, in 1999. His research interests include digital electronics, medical electronics and soft computing. He is the author and co-author of 7 books and over 60 journal and conferences papers.

Citation of this Article:

Rustem Popa, "Emotion Detection Using EEG and Voice Signals" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 7, Issue 11, pp 22-26, November 2023. Article DOI https://doi.org/10.47001/IRJIET/2023.711004

****