Computing and Monitoring various Biopotential signals using Machine Learning algorithms

Sri Vidya Devi P[1], *Sai* Krishna CH[1*], *Sai Srinivas* P[1] , *Ashraf* SK[1] ,*Sushith* K[2]

[1] EEE Dept, GRIET, Hyderabad, Telangana

Abstract. Nowadays health care units play a vital role of the human existence after the pandemic periods. It is very essential to monitor the potential signals of the human body for survival on regular basis. In this paper extracting the values of different biopotential signals produced in human body, monitoring and analysing them using various machine learning algorithms. Monitoring involves observing and checking the progress or quality of data over a period of time and keeping it under system review. The beauty of effective computing is to make machine more emphatic to the user. Machine with the capability of human electrical signal recognition can look inside the user's body. This paper generalises the view of training of the bio potentials signals data in the MATLAB software as well in python software. Analysis with different machine learning algorithms like K-Nearest Neighbours (KNN), Decision tree (DT), Logistic Regression (LR), Support Vector Machine(SVM) are used in the training ,testing and validation of the data. Better performance is achieved with these algorithms.

1 Introduction

Within biological tissue, there exist changes in potential between compartments known as bio potentials. The Na+/K+ pump is one active mechanism that keeps the concentration gradients of certain ions in the (bio)membrane that normally divides the compartments [1]. Hodgkin and Huxley were the first to model a biopotential (the action potential in the squid giant axon) with its electrical counterpart in 1952. Combining ordinary differential equations (ODEs) with a model of the nonlinear behaviour of ionic conductance in the axonal membrane led to an almost flawless interpretation of their findings. The physical laws that were used to generate the foundation ODE for the equivalent circuit are Nernst, Kirchhoff, and Ohm's laws [2]. An illustration of how to derive the differential equation in the membrane model for a single ion channel. Some of the key areas of focus include Electroencephalography (EEG), Electrocardiography (ECG), Electromyography (EMG), and Electrooculography (EOG).

EEG, ECG, and EMG biopotentials are produced by the volume conduction of currents created by assemblages of electrogenic cells. The electrical potential created by the combined actions of a significant number of neurons in the brain is known as EEG and is measured on

* Corresponding author: chellasaikrishna2000@gmail.com

the scalp. ECG and EMG are produced by the action potentials of cardiac muscle cells and skeletal muscle cell contractions, respectively. Other bio potentials, such as EOG and EGG, also come about as a result of ionic distribution or the combined effects of numerous electrogenic cell [3].The properties of almost all bio potentials, including EEG, ECG, and EMG, range over a very low frequency, usually less than 1 kHz. When measured by surface electrodes, their amplitude ranges from tens to hundreds of microvolts. Since EEG and ECG have a frequency range of less than 1 Hz, capturing these signals can be difficult due to EOVs, which can be as high as 100 mV and change slowly over time. There has been a significant amount of research on computing and monitoring biopotential signals using machine learning algorithms.

One popular application of machine learning in EEG analysis is for the detection of epileptic seizures. A study used a combination of wavelet transform and deep learning to accurately detect seizures in EEG signals [4]. Similarly, a study used a convolutional neural network (CNN) to classify EEG signals into normal, interictal, and ictal states with high accuracy [5]. In the field of ECG analysis, machine learning algorithms have been used for detecting cardiac arrhythmias. A study by Attia et al. (2019) used a deep learning approach to classify ECG signals into various arrhythmia types [6]. Another study used a support vector machine (SVM) algorithm to detect atrial fibrillation from ECG signals [7]. EMG signals are commonly used to monitor muscle activity, and machine learning algorithms have been used for various applications such as prosthetic control and rehabilitation. A study by Pan et al. (2018) used a combination of SVM and dynamic time warping (DTW) to accurately classify hand motions from EMG signals. Similarly, a study by Ameri et al. (2020) used a CNN to classify EMG signals for prosthetic hand control.

EOG signals are often used to monitor eye movements, and machine learning algorithms have been used to classify various types of eye movements. A study by Li et al. (2017) used a CNN to classify EOG signals for the detection of saccades, blinks, and smooth pursuits. Another study by Saboor et al. (2020) used a random forest algorithm to classify EOG signals for the detection of various eye movements during reading [8]. Overall, these studies demonstrate the potential of machine learning algorithms in the analysis and monitoring of various biopotential signals. Further research in this area could lead to improved diagnosis and treatment of various neurological and cardiovascular disorders [9].

1.1 System Design

The EXG sensors are used to measure the electrical activity between the pair of electrodes. The sensor can be used in different ways like EMG, ECG, EOG, EEG.EMG: EMG means Electromyogram. It is a way of measuring electric signals produced by skeletal muscles while contraction of the muscle: ECG means Electrocardiogram. It is a way of measuring electrical potential produced by heart. It is measured with respect to time. It is also called as EKG.EOG: EOG means electrooculogram. It is a way of measuring electric potential between up and below the eyeball to determine eye orientation: EEG means electroencephalogram. It is a way of measuring the electrical activity in human brain [10]. The input voltage range of EXG sensor is 4.5-40 volts, it has one channel and two to three electrodes, it can sense biopotential signals like ECG, EMG, EOG, EEG.

Fig.1. EXG Bio Amp Pill

2 Machine Learning Algorithms

Machine learning is a branch of AI that use algorithms to review massive amounts of data. In order to understand the data, these algorithms process every potential data combination without taking into account human prejudice or time constraints [11]. Machine learning analytics also recognises the boundaries of important data. Machine learning algorithms like SVM (support vector machine), KNN (k closest neighbour), Fine Tree, etc. are available in the MATLAB software.

2.1 KNN (k-nearest neighbours)

A popular non-parametric machine learning technique for classification and regression applications is KNN (k-nearest neighbours). Here are some specifics on how KNN functions**:** The KNN machine learning algorithm memorises the training dataset and uses it to generate predictions. To forecast the output variable, KNN locates the k training instances that are closest to a particular test instance in the feature space. The selection of k influences the model's sensitivity to noise and capacity to identify fine patterns in the data. KNN makes the supposition that the distance metric used to compare instances is accurate and can be applied to infer the results of a class or regression [12]. where D is the distance between nearest neighbours

$$D(x, y) = \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2}$$

2.2 SVM (Support Vector Machine)

For tasks including classification, regression, and outlier detection, a technique known as SVM is used. To increase the margin—that is, the space between the hyperplane and the SVM identifies an ideal hyperplane that splits the data into distinct classes based on the nearest data points from each class. In order to separate the data, a linear hyperplane can be utilised to transform the input features into a higher-dimensional space using kernel functions in SVM. High-dimensional feature spaces and datasets with a lot of instances can be handled using SVM. The equations shows the co-efficient value of the SVM model.

$$* \ min_{w,b,\{\beta\}^1} \frac{1}{2} \left|\left|\omega\right|\right| 2 + c \sum \beta_n$$

2.3 LR (logistic regression)

A form of machine learning method called logistic regression (LR) is used for classification issues where the objective is to forecast a binary or categorical outcome based on a set of input features. It is a kind of linear regression that transforms the output into a probability value that may be utilised to get a binary choice using a sigmoid or logistic function.[10] The LR algorithm generates a probability score by first calculating the weighted sum of the input features and then passing this sum through the logistic function. The algorithm then compares this score to a threshold number and, if the score is higher than the threshold, predicts a positive result; otherwise, it predicts a negative result. The equation shows the Probability index of LR

$$\text{Prob(y=1)} = \frac{1}{1+e^{-x(\sum_{i=1}^{n} m_i x_i + b)}}$$

2.4 DT (decision tree)

Machine learning algorithms called Decision Trees (DT) are used to solve classification and regression issues [13]. It works by creating a structure that resembles a tree, with each leaf node standing in for a class label or numerical value, each interior node for a feature or attribute, and each branch for a rule based on the value of the feature [14]. The DT algorithm initially selects the most informative feature from the input data based on a specific criterion, such as information gain or Gini index. The data is then split into two or more subsets according to the value of the selected feature. Then, for each subset, this process is repeated recursively until a stopping criterion, such as a maximum depth or a minimum number of samples, is met. Both category and numerical data can be handled by the DT algorithm, which can also detect non-linear correlations between the input features and the output variable. It is quite simple to comprehend and visualise, and it can also manage missing numbers and outliers [15].

$$\text{E(s)} = \sum_{i=1}^{c} -p_i \log_2 p_i$$

3 Simulation analysis and Data sets collection

In the figure 1 the hardware of the proposed circuit is shown. As explained in the working principle, The proposed circuit monitors the instantaneous bio potential signals. The EXG sensor senses the signals from the movement of the human body. The data are collected in the Excel with the help of Arduino.

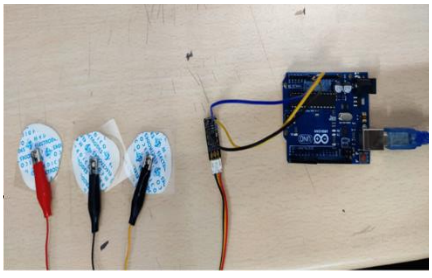

Fig.2. Hardware unit of proposed project for signal monitoring

The figure 3 shows that the data is extracted by using the EXG sensor and saved in the excel. Then the excel data is imported to the MATLAB software.

Table1. Range of Signals

Biopotential/Physiological Signal	**Measurement Range**
Electroencephalography (EEG)	0.005 ~ 0.2 mV
Electromyography (EMG)	0.1 ~ 5 mV
Electrooculography (EOG)	0.05 ~ 3.5 mV

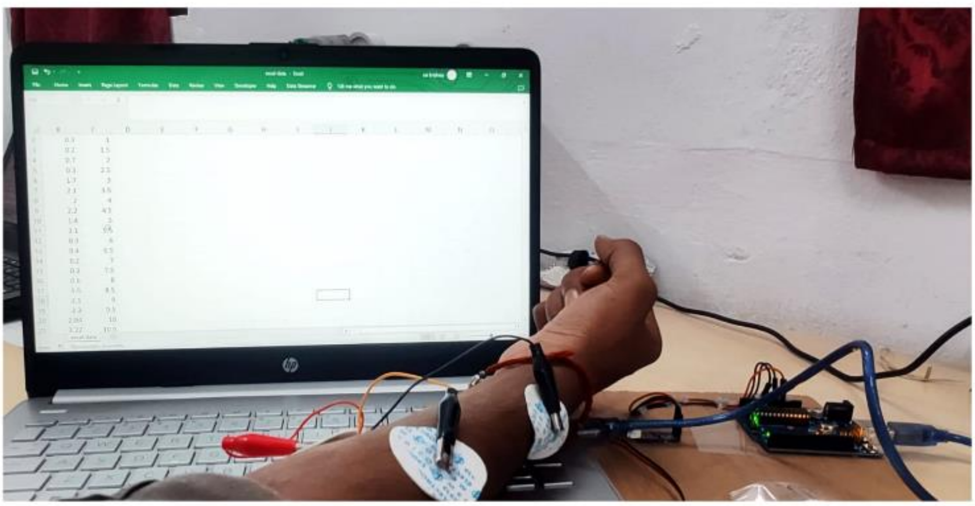

Fig.3.Data collected in the EXCEL sheets

The figure 4 is about the data training, by using classification learner, machine learning algorithm. In this algorithm neural networks like KNN are used to train the data. After the training of the data the accuracy of the data is calculated by using different neural networks. By using the collected data, the graph is plotted between pulse and time [16].

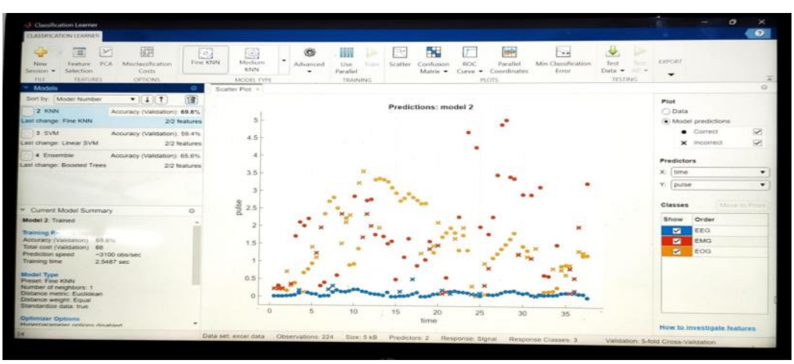

Fig.4.Scattered Plot in MATLAB software of various signal values for training

 https://doi.org/10.1051/e3sconf/202339101106

4 Training Testing and Validation

The datasets acquired from the hardware are used for the training to create an edge computing device. First data sets are imported in to the analysis where data pre-processing, imputation, training, testing & validation is carried for various bio potential signals Here the platform called Anaconda Cloud with the help of Jupiter lab is used to train the data and collected reports to draw the bar graph. From Fig 5 shown that plotting graph by time Vs pulse and obtained the results.

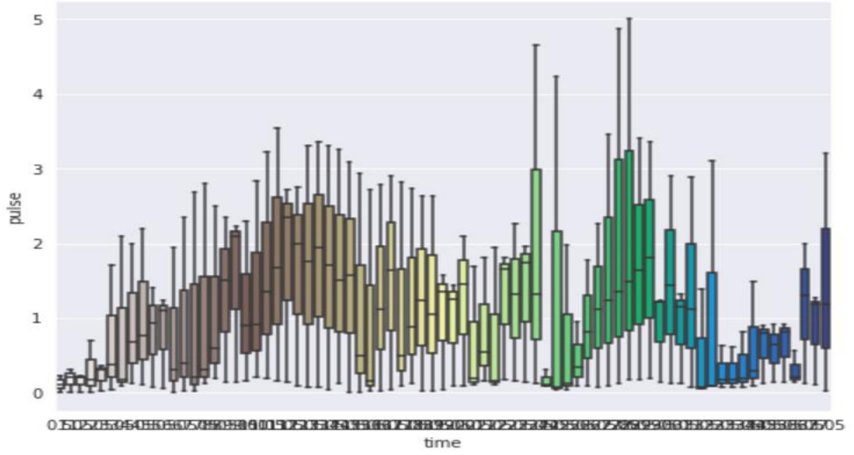

Fig.5.Box Plot between Pulse and Time

4.1 Training and Testing of EEG

EEG is a way of measuring the electrical activity in human brain.The receiver operating characteristic curve (ROC curve) is a graph that displays how well a classification model performs across all categorization levels.

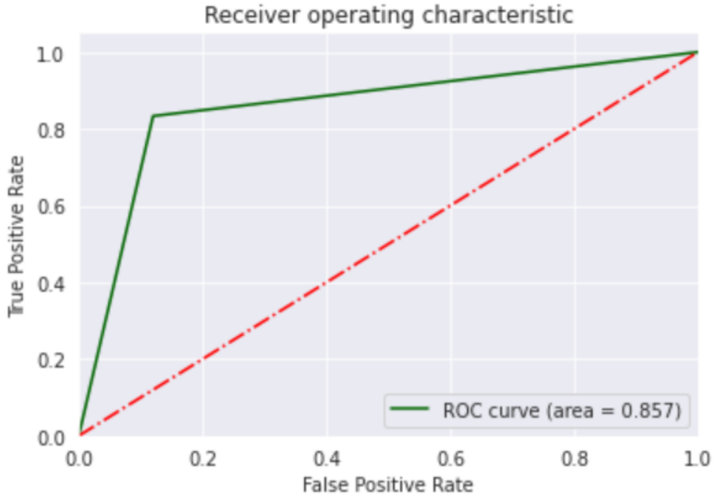

Fig.6. ROC curve of EEG signals

4.2 Training and Testing of EMG

EMG is a way of measuring electric signals produced by skeletal muscles while contraction of the muscle**.**

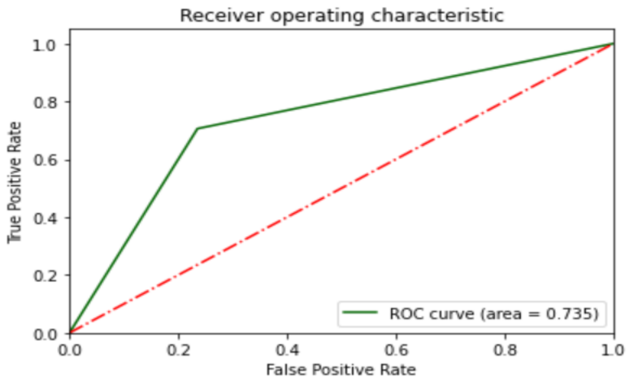

Fig.7. ROC curve of EMG signals

4.3 Training and Testing of EOG

EOG is a way of measuring electric potential between up and below the eyeball to determine eye orientation.

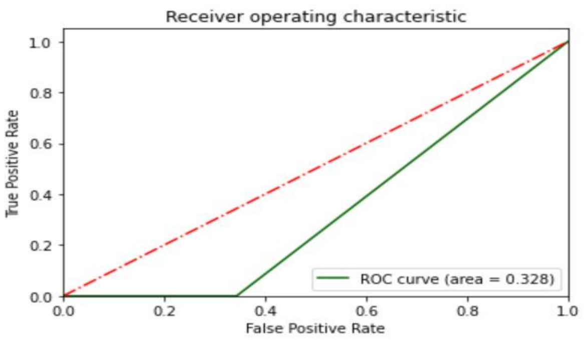

Fig.8. ROC curve of EOG signals

Table 2. Report of Accuracy and Score of signals

TYPES OF SIGNALS	ARTIFICIAL NEURAL NETWORK ML ALGORITHM METRICS							
	ACCURACY				SCORE			
	KNN	SVM	DT	LR	KNN	SVM	DT	LR
EEG	0.88	0.84	0.88	0.87	0.83	0.76	0.83	0.77
EMG	0.75	0.66	0.69	0.62	0.59	0.3	0.51	0.19
EOG	0.75	0.65	0.68	0.65	0.62	0.45	0.58	0.0

A measure of accuracy is the percentage of correctly classified items among all the items. If the classes in a dataset are balanced, accuracy is a meaningful metric. Depending on the context, the term "score" can refer to a wide range of metrics, but in general, it is used to summarise how well a model or system is working. Score is sometimes used to refer to a particular metric, such as the F1 score, which combines precision and recall into one measurement.

Table 3. Report of Precision and Recall

TYPES OF SIGNALS	**ARTIFICIAL NEURAL NETWORK ML ALGORITHM METRICS**							
	PRECISION				**RECALL**			
	KNN	**SVM**	**DT**	**LR**	**KNN**	**SVM**	**DT**	**LR**
EEG	0.74	0.71	0.76	0.83	0.95	0.81	0.9	0.71
EMG	0.7	0.55	0.57	0.37	0.5	0.2	0.45	0.12
EOG	0.63	0.47	0.57	0.0	0.6	0.43	0.62	0.0

A measurement of precision is the percentage of relevant results that are returned. Even if some relevant items are missed, a high precision score shows that the system or model is good at recognising just important elements Recall is a metric that gauges how many pertinent things a model or system was able to retrieve. Even if the system or model also detects some irrelevant objects, a high recall score shows that it is good at detecting all relevant items.

5 Conclusion

In this paper monitoring of different biopotential signals has been computed. Datasets collection and analysing the data of various biopotential signals produced during the muscle contraction /relaxation or during eye blink etc. The paper provides the clear view on the working of the EXG sensor pill. The data acquisition from the prototype, carried out by using microcontroller. Training of the data and prediction of the data is done by using the MATLAB/anaconda python platform. Different Machine Learning (ML) algorithms shows the better accuracy and precision in analysis. In further analysis an edge computing device can be build for the signal prediction in human.

References

1. Pivovarov AS, Calahorro F, Walker RJ,**0221,** IEEE access(2018)
2. Qinyi Lv, Yazhou Dong, Yongzhi Sun, Changzhi Li, Lixin Ran, IEEE access,**7303791**,(2015)
3. auzani.N Jamaluddin, Siti A. Ahmad, Samsul Bahari Mohd Noor, Wan Zuha Wan Hasan, IEEE access(2014)
4. S. Kiranyaz, IEEE access(2016)
5. Yang et al. Gate Research (2018)
6. Attia et al., article(2019)
7. MTRibeiro ,Tongshuang Wu ,C.Guestrin, SameeSingh, IEEE access(2020)
8. Ratna Astuti Nugrahaeni1 , Kusprasapta Mutijarsa2, IEEE access, **0493-2137** ,(2016)

9. F. Lotte et al, IEEE access (2018)
10. F. Moody and S. Moody, IEEE access,**13**, (2017)
11. H. Khandoker , IEEE access (2018)
12. A. Amiri, IEEE access (2018)
13. K. Ang, IEEE access (2015)
14. S. Garg, IEEE access (2019
15. V. Palshikar, IEEE access (2012)
16. Masoud MALEKI, Negin MANSHOURI, Temel KAYIKÇIOöLU, IEEE access 2017