Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 132 (2018) 1523–1532

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2018)

Classification of EEG data for human mental state analysis using Random Forest Classifier

Damodar Reddy Edla[1], Kunal Mangalorekar[1], Gauri Dhavalikar[1], Shubham Dodia[1]

Department of Computer Science and Engineering, National Institute of Technology Goa, India

Abstract

Brain computer interface (BCI), has been one of the most popular domains in computing in the recent years. BCI is a pathway which allows communication between computers and the human brain. We acquire real time EEG data with the device, Neurosky Mindwave Mobile, which uses a single dry electrode. Experiment for acquisition of data is carried on 40 subjects (33 male and 7 female). Feature extraction of EEG signals are done by statistical measures such as mean, standard deviation, maximum and minimum amplitudes. In this paper we explore the approach of ensemble learning with classifiers such as random forest classifier to build a BCI model to predict mental states as concentration and meditation. Analysis and results of our proposed model shows an accuracy of 75% using the above methodologies. This model is further implemented in the field of Internet of Things (IoT), for the application of home automation.

Keywords: BCI; EEG; Random Forest Classifier; Mental States.

1. Introduction

Brain Computer Interface (BCI) [1] is a technology which facilitates the communication between the brain and external devices. A BCI system relies on real time recognition of different mental states from brain activity. The BCI reads different brain signals produced by the various parts of the human brain, translates these signals into actions and commands for various applications. There is no dependence on the peripheral nerves and muscles of a human body during any communication with BCI. The current development of BCI is focused for the people with severe disabilities that can render them to perform physical movements [2].

BCI is being used in many applications worldwide and one of them is in the medical field to help patients who are paralyzed or partially paralyzed. Through BCI based models, patients can control artificial prosthetic arms and

* Corresponding author.
E-mail address: dr.reddy@nitgoa.ac.in

1877-0509
10.1016/j.procs.2018.05.116

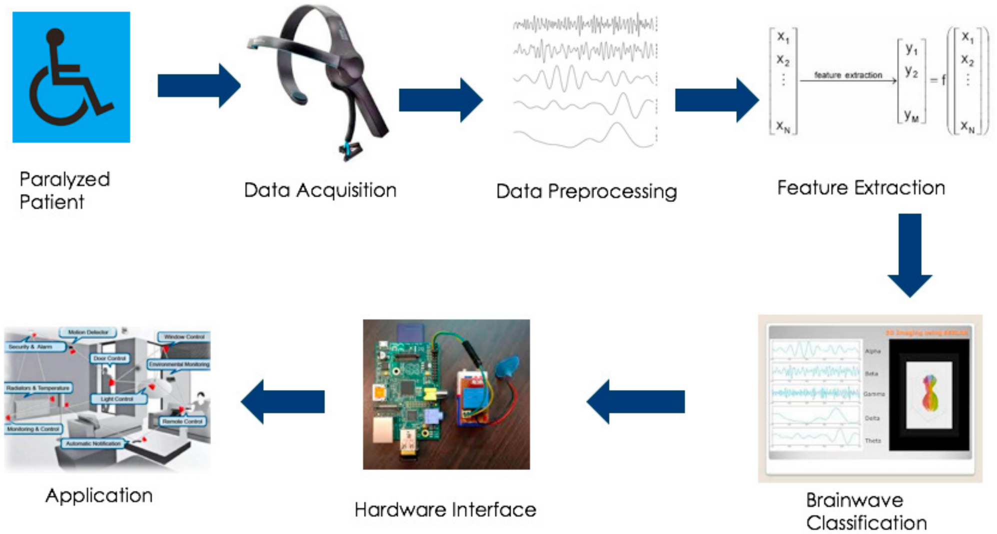

Fig. 1. General BCI model in the field of medical applications

limbs solely through their thoughts. Many such applications are possible with this new technology coming into light. Home automation and Internet of Things is another field which has tremendous application for medical purposes [11]. BCI model can be classified under the following three categories as dependability, invasiveness and synchronization. Dependability is further categorized as a dependent BCI wherein certain control is required by the subject and an independent BCI which does not require any control [6]. Signal acquisition can be performed mainly in two ways, by surgically implanting a BCI device into the human brain which is known as Invasive BCI acquisition technique or Non-Invasive BCI acquisition technique where electrodes are placed onto the scalp to collect the data [4]. A synchronous BCI system interacts at a certain period of time whereas in an asynchronous BCI system, also called as “self-placed”, the interactions take place at any period of time [7].

Any general BCI model contains the following steps, signal acquisition, data pre processing, feature extraction and classification [3]. The first step of a general BCI model is signal acquisition wherein signals are collected from the brain in the form of analog waves and then converted into digital numerical values. The acquired signals are then pre processed to remove noise and artifacts. Feature extraction is the next step where certain features are selected and extracted from the data for classification. A classifier takes as input these extracted features and predicts the corresponding class to which they belong to. Figure 1 shows a general BCI model for medical applications.

Electroencephalography (EEG) is a non-invasive method used to record the electrical activities generated by the brain from electrodes placed on the scalp. Frequency and Amplitude are the characteristics of these electrical signals acquired as EEG waves. Low frequency waves are generated when a person is asleep or in a relaxed state of mind whereas the frequency increases as the person starts responding to the outer environment. Attentive state of mind also gives rise to waves with higher frequencies [5].

2. Terminologies

2.1. NeuroSky Mindwave Mobile

The Neurosky Mindwave mobile is an EEG device which is used to capture electrical activity from the forehead of humans. The device connects wirelessly to any computer or mobile via bluetooth. The device has two electrodes, one which sits on the forehead and another reference electrode which clips to the ear. It incorporates a chip that preprocesses the raw EEG data obtained from the electrode and generates 8 values which are low gamma(31-40 Hz), high gamma(41-50 Hz), high alpha(10-12 Hz), low alpha(8-9 Hz), high beta(18-30 Hz), low beta(13-17 Hz), delta(1-3 Hz) and theta(4-7 Hz) every one second [12]. Figure 2 shows the Neurosky Mindwave Mobile device.

Fig. 2. Neurosky Mindwave Mobile

2.2. Ensemble Learning

There exist different algorithms used to predict mental states of humans by analyzing EEG signals. Different applications require different models and different modifications. The classification algorithms can be supervised or unsupervised based on the desired application. Unlike traditional machine learning algorithms, ensemble learning is an approach which uses multiple learning algorithms for learning in the aim of improving the predictive performance [14]. Ensemble learning falls under supervised learning and makes use of an ensemble of classifiers for learning. In general these ensemble classifiers such as run prediction over many basic classifiers such as svm, naive bayes and decision tree, and then take vote for final consideration of class [14].

2.3. Random Forest Classifier

Random forest classifier [20] uses an ensemble learning method for classification which uses multiple decision trees during training phase and outputs average prediction of individual trees. This classifier generates forests with random amount of trees. Normal decision tree algorithms are rule based and are solely based on some set of rules for prediction on data set. In contrast to this, random forest classifiers instead of using gini index [13] or information gain for calculation of root node, it finds the root node and splits the features randomly. The following Figure 3 [17] illustrates a general working of the classifier. x represents the input to the classifier. Random decision trees are generated namely $tree_1$, $tree_2$, ..., $tree_B$, whose corresponding outputs are k_1, k_2, ..., $tree_B$. Majority Voting is taken and class k is selected from k_1, k_2, ..., k_B. Output of the classifier is class k with majority of the votes.

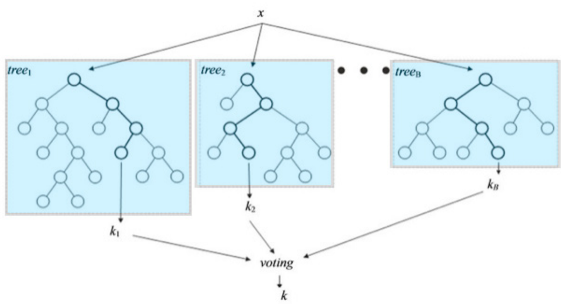

Fig. 3. General working of random forest classifier

3. Proposed Approach

3.1. Overview

Our proposed module comprises of three main steps - data preprocessing, feature extraction and classification. Once the data has been collected, the first step is data preprocessing which includes noise reduction and filters which is performed by the EEG device, NeuroSky Mindwave Mobile, used for data acquisition. The next step is feature extraction wherein we extract statistical features like mean, standard deviation and difference of maximum and minimum of the preprocessed data. The final step consists of classification of the extracted features using a suitable classifier for the collected set of real time data. In the following model we propose to classify the given data using random forest classifier which uses an ensemble learning approach.

3.2. Data Acquisition

There are various systems for the placement of electrodes on the scalp. The most widely used system for research purpose is the 10/20 system which comprises of 21 electrodes [9]. The positioning of the EEG electrodes is done with the help of two landmarks which are the nasion and inion. The area between the eye and above the nose is the nasion area whereas the area at the lowest point of the skull from the back of the head is the inion area. Figure 4 shows the position of the electrodes in the 10/20 system [10].

The acquisition of EEG signals was done by a non-invasive method, using the Neurosky MindWave Mobile device. This device uses a single electrode which sits on the forehead of a person, and records the electrical activity from the frontal lobe of the human brain. The electrode used by Neurosky Mindwave Mobile is the Fp1 electrode according to the 10/20 system which is also marked in the above Figure 3 and a reference electrode is used which is the A1 electrode. Two states of mind namely attention and meditation can be recorded using this device, also the blink strength of the eye is recorded. Corresponding to each attention and meditation signal, the device outputs other nine types of wave signals [12].

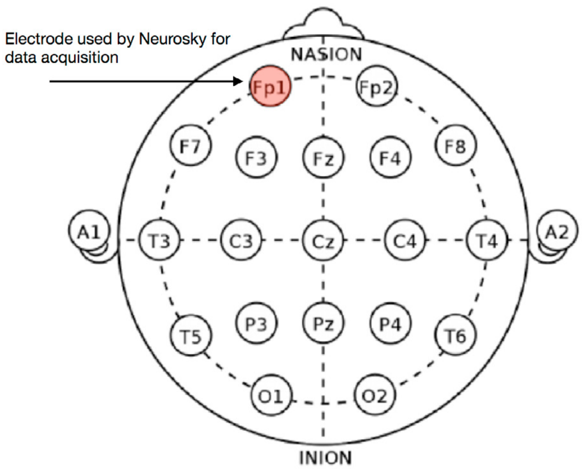

Fig. 4. Neurosky Electrode position

Data was collected from a total of 40 subjects out of which 33 were male and 7 were female. All the subjects fell into the age group of 18 to 22 years. All the volunteers had voluntarily participated in this experiment and a consent form was signed from each of them. In the experiments conducted, subjects were asked to solve a mathematics problem mentally, for one minute of time, thus recording their attentive state of mind. Later to test the relaxed state of mind, the subjects were allowed to meditate by closing their eyes and relaxing their mind for one minute. Each subject repeated the above set of tasks for one more time. The experiment was conducted in a quiet environment so as to maintain the focus of the subject and reduce the signal noises which could have interfered otherwise. Figure 5 and Figure 6 show sample output of the signals generated by the device for a subject while concentrating. The first column shows the time, the next three columns are the eSense meters [12] which is a proprietary technology of Neurosky. The next 8 columns are the values of low gamma, high gamma, high alpha, low alpha, high beta, low beta, delta and theta, which are used to processing and extracting features.

```
0.223282, 0, 0, 23, 81, 8937, 6940, 76118, 79103, 18533, 12158, 696162, 93085
0.235613, 0, 0, 17, 83, 7955, 6565, 80958, 30461, 29424, 11996, 878224, 79426
0.248556, 0, 0, 29, 90, 8533, 5166, 165632, 4992, 9496, 20372, 684713, 51322
0.260686, 0, 0, 26, 97, 9882, 5774, 128346, 27683, 7490, 38910, 536523, 106986
0.273253, 0, 0, 24, 100, 5655, 5099, 178006, 25035, 18537, 37871, 1009111, 98297
0.284367, 0, 0, 21, 100, 1430, 734, 41823, 4050, 4171, 1448, 122099, 100400
0.295358, 0, 0, 1, 94, 4071, 3447, 109181, 11670, 25793, 27243, 1050484, 356201
0.307311, 0, 0, 1, 74, 1550, 1047, 32942, 24847, 3029, 8991, 403443, 10384
0.32024, 0, 0, 8, 83, 11647, 3263, 107862, 22124, 25350, 62875, 2074645, 141240
0.332271, 0, 0, 13, 75, 1362, 482, 33146, 16693, 3075, 9469, 331789, 29385
0.344342, 0, 0, 16, 84, 1361, 781, 28748, 24978, 4657, 8392, 43287, 25620
0.356528, 0, 0, 17, 100, 985, 1368, 44559, 16307, 3898, 10610, 259040, 71712
0.368625, 0, 0, 14, 87, 11332, 8747, 59482, 31231, 13207, 19942, 1633629, 141190
0.381007, 0, 0, 7, 87, 521, 599, 40898, 3518, 2520, 5257, 94168, 53872
0.392666, 0, 0, 1, 81, 4241, 4531, 37143, 54408, 14702, 24618, 438092, 73932
0.403498, 0, 0, 7, 75, 6445, 1606, 27406, 20026, 18484, 43958, 900346, 237191
0.415479, 0, 0, 16, 60, 18899, 6445, 35028, 68932, 18583, 47638, 696614, 119983
0.42686, 0, 0, 24, 60, 634, 445, 4539, 1298, 949, 2291, 40740, 7271
0.439263, 0, 0, 38, 56, 1339, 594, 16357, 6961, 2295, 5184, 276759, 24516
```

Fig. 5. Output of the device for meditation state of mind

```
0.167812, 0, 0, 90, 84, 1508, 949, 3203, 15514, 8016, 10146, 133736, 38691
0.180342, 0, 0, 84, 70, 3783, 2777, 3485, 10768, 5870, 18148, 573495, 18388
0.193154, 0, 0, 96, 56, 9032, 7513, 15279, 14077, 11496, 13129, 1179111, 17200
0.205996, 0, 0, 100, 64, 29317, 7823, 25351, 61559, 25780, 92004, 436123, 377586
0.216809, 0, 0, 100, 75, 4104, 3243, 3700, 9609, 10694, 10152, 1084586, 36726
0.229007, 0, 0, 100, 66, 10283, 3179, 8039, 5281, 12855, 12045, 70447, 32239
0.240852, 0, 0, 100, 61, 33718, 37640, 10555, 24906, 25508, 23732, 1293020, 166294
0.252404, 0, 0, 90, 37, 1238, 1470, 1246, 2192, 3295, 937, 28083, 14104
0.265383, 0, 0, 100, 40, 28517, 88436, 80344, 30035, 19225, 23309, 1436290, 177130
0.276735, 0, 0, 91, 54, 4307, 29775, 4808, 10617, 3281, 11809, 1148265, 29867
0.288082, 0, 0, 69, 61, 11526, 90991, 20050, 30755, 17000, 29933, 727766, 84383
0.29801, 0, 0, 69, 75, 16287, 77329, 32048, 8935, 18732, 20974, 1024286, 95003
0.31089, 0, 0, 64, 83, 15751, 71115, 63879, 1378, 11293, 12388, 1365915, 59958
0.320636, 0, 0, 69, 90, 6730, 4119, 21590, 14780, 23021, 29032, 807542, 179603
0.330916, 0, 0, 77, 74, 4956, 4589, 25318, 7055, 25464, 8370, 64830, 40866
0.342872, 0, 0, 90, 77, 938, 928, 10497, 18319, 4187, 2472, 120615, 57520
0.355476, 0, 0, 70, 84, 1192, 781, 21972, 11615, 4556, 6700, 186409, 36881
0.367898, 0, 0, 63, 84, 7839, 3518, 12183, 6902, 6836, 30233, 1025026, 53015
0.380892, 0, 0, 57, 91, 6580, 3994, 46386, 97096, 17512, 45730, 1874658, 172174
```

Fig. 6. Output of the device for concentration state of mind

3.3. Data Preprocessing

NeuroSky Mindwave Mobile, the EEG device we have used for data acquisition, for our proposed model, preprocesses the raw data in the following way. Firstly, the raw EEG data obtained from the front electrode is given as 8 different signals which are low gamma, high gamma, high alpha, low alpha, high beta, low beta, delta and theta. The data is received every single second and preprocessed using a single chip present in the NeuroSky Mindwave Mobile device. These 8 signals represent the various states of a human brain mainly focusing on the concentration and meditation state of mind [12].

3.4. Feature Extraction

The features are taken from the 8 values generated by the EEG device we have used. For extraction of features we used statistical tools such as mean, standard deviation and difference of maximum and minimum value of the data.

For implementation of these in the code, we use python libraries such as numpy to calculate the above quantities. The following equations are used to compute the statistical features of signals [8].

1. Mean: Mean, $\bar{x}$ is the expected value of a sample in the data set. The mean of all features x_i is calculated per trial, for each of the 8 features of the subject.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N}(x_i) \tag{1}$$

2. Standard Deviation: The standard deviation, σ of a data set tells us how the data varies with respect to the mean. The standard deviation of each feature x_i in the data set of 8 features is calculated per trial of every subject.

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{2}$$

3. Maximum - Minimum Amplitude: The difference between the highest and lowest value. Each of the features represent amplitude values of which the maximum and minimum is taken for each feature x_i , in every trial of the subject.

$$Max(x_i) - Min(x_i) \tag{3}$$

The data set is given by x_i , i=1,2,3...N.

All of these statistical features would be calculated on our data set of 8 values and appropriate class labels would be given in the classification stage.

3.5. Classification

Real time data is very complex and varies with a lot of factors unlike traditional data sets. Traditional classifiers in classification of brainwaves have been able to give good results for all kinds of data in the field of BCI in the recent years. Some of them are namely svm, naive bayes, k-NN, lda, and so on. Thus experimenting with ensemble learning gives us an alternate approach for learning from real time data obtained. Classification in our model is binary which means that our model predicts whether the state of mind of the person is attentive or relaxed state. In our model, for classification, we have incorporated random forest classifier which uses an ensemble learning approach towards prediction [20]. Random forest classifier works in a similar way as the decision tree classifier only with an ensemble learning approach added to it. The first step is creation of many random decision trees each predicting a particular class according to the features given to it. Once each tree predicts a class, a voting is carried out to take into consideration the final class according to majority. The output is then the class which has the majority voting as explained in the Figure 7 below.

4. Results and Discussion

For testing the model, we have taken 10% of the total data as testing data and 90% of the data for training the model. Therefore for 40 subjects a total of 600 data set was collected and 60 was used for testing. The results and performance of our model is given in the three subsections as follows.

4.1. Confusion Matrix

For a set of true data and predicted data, the performance of a classification model, is represented using confusion matrix [15] in the form of a table. The confusion matrix obtained for our classification model is given in Table 1 below.

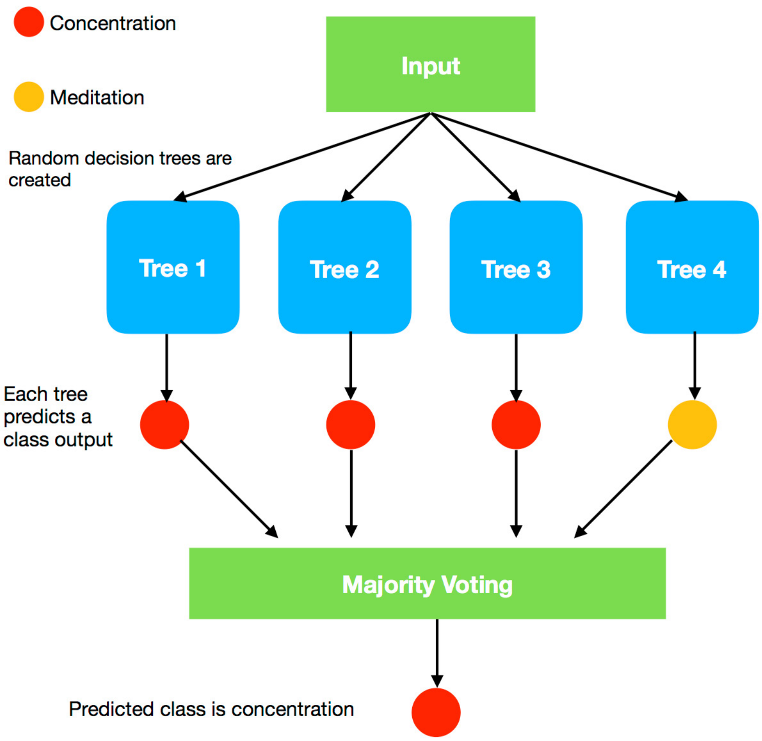

Fig. 7. Schematic diagram of random forest classifier

Table 1. Confusion Matrix for testing data.

Classes	Concentration	Meditation
Concentration	24	9
Meditation	6	21

The number of correctly classified samples as concentration state are 24 whereas number of correctly classified samples for meditation state are 21. The number of falsely classified samples as concentration state are 9 whereas the number of falsely classified samples as meditative state are 6. Thus, we observe that around 75% of the total data set is predicted into correct classes.

4.2. Performance Measures

Classification performance is measured by using certain performance measures [16] such as precision, recall, support etc. Some of these measures given are listed in the Table 2 below.

Table 2. Classification performance measures

Classes	Precision	Recall	Support
Class 1 (Concentration)	0.80	0.73	33
Class 2 (Meditation)	0.70	0.78	27

True positives (TP) are the samples which are correctly classified and correctly predicted to the class concentration. True negatives (TN) are the samples which are correctly classified and correctly predicted to class meditation. False negatives (FN) are the the samples which are wrongly classified as concentration and false positives (FP) are the samples which are wrongly classified as meditation.

1. Precision: It is the ratio of true positives to the total positives.

$$\frac{TP}{TP + FP} \tag{4}$$

2. Recall: It is defined as the ratio of true positives to the total positives classified.

$$\frac{TP}{TP + FN} \tag{5}$$

3. Support: It is defined as the number total samples classified in the respective class. The equations are given as follows.
   For positive class (concentration):

$$TP + FN \tag{6}$$

   For negative class (meditation):

$$FP + TN \tag{7}$$

Survey shows that many papers have used ensemble learning for classification of brainwaves for various applications. Ensemble classifiers have been use widely for classification tasks such as motor imagery. Prediction analysis from [18] shows an accuracy of upto 90% in predicting classes with ensemble approach. The model was built on real time datasets generated by collecting EEG data from various subjects.
Various traditional classifiers have been used for classifying EEG signals. Neural networks and svm have been used in [19] for EEG data classification, analysis of which show an accuracy of upto 88% for neural networks and upto 82% for svms. These predictions were done on real time data sets generated by collecting EEG data from medical patients. In our model we have applied classification on real-time EEG data from healthy students to do a binary classification of mental state of mind namely concentration and meditation. The ensemble approach of learning used in random forest classifier show an accuracy of 75% in predicting the classes accurately.

4.3. IoT application

Development of home automation and IoT is moving forward towards the future in creating the ideal smart home environment. Smart homes and IoT has helped many paralyzed patients to control their home appliances effortlessly. Our model predicts two states of mind namely concentration and meditation which can be mapped to electrical signals corresponding to switch on and switch off. The integration is facilitated by a micro processor Raspberry Pi [21] which will translate the model output into appropriate on and off signals. The logic flow of the IoT application is given in Figure 8 below.

Raspberry Pi has the advantage of facilitating python code for programming its hardware pins and working, therefore making it easier for the user to integrate the predictive model and hardware efficiently.The model outputs either class concentration or meditation which is mapped by the raspberry pi to the GPIO pins to directly send the electrical signals. GPIO module in python is used to call appropriate functions for generating either “ON” or “OFF” [21]. The hardware architecture of the entire setup is shown in Figure 9.

5. Conclusion

In this paper, we have proposed a system for human mental state classification into classes namely concentration and meditation. The model uses statistical feature extraction technique and random forest classifier. An experiment

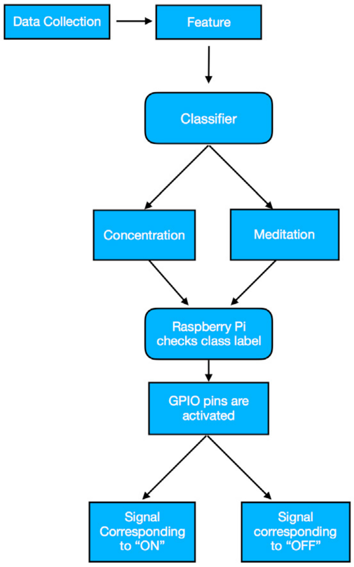

Fig. 8. Logic flowchart for the hardware setup

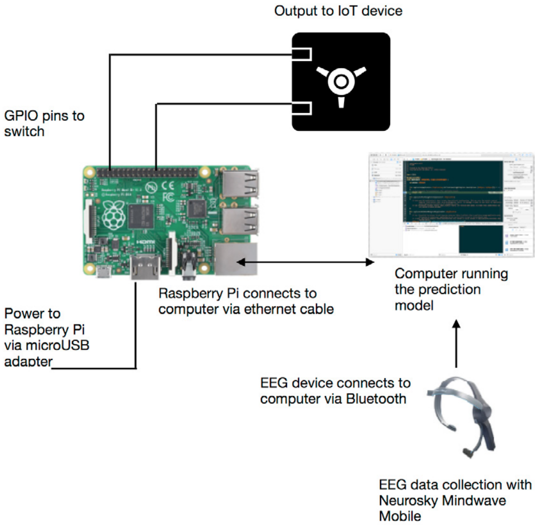

Fig. 9. Logic flowchart for the hardware setup

was conducted to collect EEG data from the human brain of 40 subjects and a machine learning model was built, which resulted in an accuracy of 75% on this real time data. This binary classification could be used for numerous purposes such as, in the field of IoT to switch on and switch off devices and to control other similar home appliances, only using your concentration and meditation state of mind. The predictive model we built learns the users' states of mind over time and thus can improve and help to control IoT devices with better precision.

References

[1] Kaur, M., Ahmed, P. and Rafiq, M.Q. (2012) "Technology development for unblessed people using bci: A survey." *International Journal of Computer Applications*, **40** (1).

[2] Jonathan R Wolpaw, Niels Birbaumer, William J Heetderks, Dennis J McFarland, P Hunter Peckham, Gerwin Schalk, Emanuel Donchin, Louis A Quatrano, Charles J Robinson, Theresa M Vaughan. (2000) "Brain-computer interface technology: a review of the first international meeting." *IEEE transactions on rehabilitation engineering* **8** (2):164–173.

[3] Steven G. Mason, and Gary E. Birch. (2003) "A General Framework for BrainComputer Interface Design." *IEEE Transactions on neural systems and rehabilitation engineering* **11** (1): 70–85.

[4] Anupama.H.S, N.K.Cauvery, Lingaraju.G.M. (2012) "Brain computer interface and its types - a study." *International Journal of Advances in Engineering and Technology* **3** (2): 739–745.

[5] Balkis Solehah Zainuddin, Zakaria Hussain, Iza Sazanita Isa. (2014) "Alpha and Beta EEG Brainwave Signal Classification Technique: A Conceptual Study." *IEEE 10th International Colloquium on Signal Processing & its Applications (CSPA2014)*

[6] B. Allison, B. Graimann, A. Grser. (2007) "Why use a BCI if you are healthy, BRAINPLAY Brain-Computer Interfaces Games Work." *Advances in Computer Entertainment Technology Conferences*.

[7] A. Bashashati, R.K. Ward, G.E. Birch. (2007) "Towards development of a 3-state self-paced brain-computer interface." *Computational Intelligence and Neuroscience*

[8] V. Abootalebi, M. H. Moradi, M. A. Khalilzadeh. (2009) "A new approach for eeg feature extraction in p300-based lie detection." *Computer methods and programs in biomedicine* **94** (1): 48–57.

[9] V. Jurcak, D. Tsuzuki, and I. Dan. (2007) "10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems." *Neuroimage* **34** (4): 1600–1611.

[10] Mohammed Hassan Alnemari. (2017) "Integration of a Low Cost EEG Headset with The Internet of Thing Framework." *UC Irvine Electronic Theses and Dissertations*

[11] Wanli Min and Gang Luo. (2009) "Medical Applications of EEG Wave Classification." *CHANCE* **22** (4): 14–20.

[12] R Robbins and M Stonehill. (2014) "Investigating the NeuroSky MindWave¢ EEG Headset." *Transport Research Foundation* **1** (28): 14–20.

[13] Hegelich, S. (2016) "Decision Trees and Random Forests: Machine Learning Techniques to Classify Rare Events". *European Policy Analysis* **2** (1): 98–120.

[14] Lotte, Fabien, et al. (2007) "A review of classification algorithms for EEG-based braincomputer interfaces." *Journal of neural engineering*

[15] Subasi, Abdulhamit. (2007) "EEG signal classification using wavelet feature extraction and a mixture of expert model." *Expert Systems with Applications* **32** (4): 1084–1093.

[16] Guo, Lei, et al. (2011) "Classification of mental task from EEG signals using immune feature weighted support vector machines." *IEEE Transactions on Magnetics* **47** (5) : 866–869.

[17] Nguyen C. , Wang Y. , & Nguyen H. N. (2013) "Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic." *Journal of Biomedical Science and Engineering* **6** (05): 551–560.

[18] Kevric, Jasmin, and Abdulhamit Subasi. (2017) "Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system." *Biomedical Signal Processing and Control* **1** (31) : 398–406.

[19] Kousarrizi, Mohammad Reza Nazari, et al. (2009) "Feature extraction and classification of EEG signals using wavelet transform, SVM and artificial neural networks for brain computer interfaces." *International Joint Conferences on Bioinformatics, Systems Biology and Intelligent Computing* .

[20] Fraiwan, Luay, et al. (2012) "Automated sleep stage identification system based on timefrequency analysis of a single EEG channel and random forest classifier." *Computer methods and programs in biomedicine* **108** (1): 10–19

[21] Richardson, Matt, and Shawn Wallace. (2012) "Getting started with raspberry PI", *O'Reilly Media, Inc.*