DETECTION OF ALZHEIMER'S DISEASE USING MACHINE LEARNING ALGORITHMS

EI8711 CREATIVE AND INNOVATIVE PROJECT

Submitted by

VIGNESHWARAN SA 2014505561

VIMAL PRATAP R A 2014505562

SUMATHY B G 2014505554

In partial fulfilment for the award of the degree

Of

BACHELOR OF ENGINEERING

IN

ELECTRONICS AND INSTRUMENTATION ENGINEERING

DEPARTMENT OF INSTRUMENTATION ENGINEERING

MADRAS INSTITUTE OF TECHNOLOGY

ANNA UNIVERSITY:: CHENNAI 600044

OCTOBER 2017

MADRAS INSTITUTE OF TECHNOLOGY CAMPUS

ANNA UNIVERSITY : CHENNAI-600044

BONAFIDE CERTIFICATE

Certified that this project report **DETECTION OF ALZHEIMER'S DISEASE USING MACHINE LEARNING ALGORITHMS** is the bonafide work of **VIGNESHWARAN SA (2014505561), VIMAL PRATAP R A (2014505562), SUMATHY B G (2014505554)** who carried out the project work under my supervision.

SIGNATURE	**SIGNATURE**
DR.J.PRAKASH	**DR. D.MANAMALLI**
HEAD OF THE DEPARTMENT	**SUPERVISOR**
Dept. of instrumentation engineering	Associate Professor
Madras Institute of Technology	Dept. of Instrumentation Engg
Anna university	Madras Institute of Technology
.	Anna University

ACKNOWLEDGEMENT:

We wish to place on record, our deep gratitude and appreciation to our DEAN **Dr.A.Rajadurai** for providing the necessary infrastructure to carry out the project.

We are thankful to our HEAD OF THE DEPARTMENT, **Dr.J.Prakash,** who has constantly encouraged and motivated us in our endeavours.

We are extremely grateful to our project guide **Dr.D.Manamalli**, for his timely and thoughtful guidance without whom this project could not have been realized. I would like to thank committee member **Mrs.D.Kalpana** for guiding us.

We owe a debt of gratitude to all the staff members of the Department of Instrumentation Engineering and our friends who constantly offered constructive suggestions for improvement in various phases of the project.

VIGNESHWARAN SA

VIMAL PRATAP R A

SUMATHY B G

ABSTRACT

The detection of Alzheimer’s disease had always been a tedious task in the past, be it through the analysis of previous medical history or careful examination of the patient through various tests like brain imaging. Our proposed work utilizes the EEG signals of the patients to classify and the early detection of the disease. In our proposed work, we classify the surface EEG signals using advanced Machine Learning concepts like Evolutionary Extreme Learning Machine and Allocation Method using Multiple Classifiers and compare the results of the two methods to find the best one suited for the job. In Evolutionary Extreme Learning Machine(EELM)[1], there will be multiple hidden layers as against the Extreme Learning Machine(ELM)[2] which has a single hidden layer, being used has shown promising results on a variety of biological dataset including the Spiro-meter dataset with accuracies as high as 100%. The Allocation method, introduced by Sašo Karakatic,Vili Podgorelec[3], we developed an innovation as to improve the max accuracy provided by the conventional procedures. The proposed plan incorporates not more than 3 channels for data acquisition with the ultimate aim of reducing the computational cost and resources. The collected EEG data will be preprocessed using wavelet transform and the dimension of the resulting output will be reduced using PSO as in [4].

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS, ACRONYMS AND NOMENCLATURE

1. EEG – ElectroEncephalogram
2. ELM – Extreme Learrning Machine
3. E-ELM – Evolutionary Extreme Learning Machine
4. fMRI – functional Magnetic Resonance Imaging
5. CT – Computed Tomography
6. CART - Classification And Regression Tree
7. AD – Alzheimer's Disease
8. CNN - Convolutional Neural Network
9. Power Spectral Intensity(PSI)
10. Relative Intensity Ratio(RIR)
11. Petrosian Fractal Dimension(PFD)
12. Higuchi Fractal Dimension(HFD)
13. Approximate Entropy(ApEn)
14. Detrended Fluctuation Analysis(DFA)
15. SVM - Support Vector Machine
16. RBF- Radial Base Function.
17. OCSVM- OneClass Support Vector Machine
18. SMO-Sequential Minimal Optimization
19. Yma- majority dataset
20. Ymi- minority dataset

CHAPTER 1: INTRODUCTION

1.1 INTRODUCTION

Classification is the most commonly used supervised technique in machine learning. Since the dawn of ages of Machine Learning and Artificial Intelligence, several classification algorithms have been developed. Every classification algorithm worked on a specific principle, while a Naïve Bayes' Classifier belonged to a family of probabilistic classifiers, Decision Trees and CARTs worked on the process of decision making at every level of Classification; but as the need of man increases, so does the complexity of problem and data involved in it. This eventually led to a decline in the overall accuracy of micro classifiers, or to put it in other words, people and data scientists started expecting greater accuracy from the algorithms. Some classifiers worked perfectly on certain portion of the data while others work efficiently on other portions. Using an ensemble of these classifiers on the various portions of the same dataset, it is possible to obtain higher efficiencies. Several micro classifiers are combined to form a larger macro ensemble classifier. The main and most attractive advantage of these ensemble methods are that, they give a higher accuracy at shorter run time. These high rates of accuracies may seem like a likely choice if a non-vital dataset is considered. But when classifying medical and other biological datasets, it is absolutely necessary to have a perfect accuracy. Even a mere 2% error is not acceptable when it comes to medical data. That's when data scientists started developing more complicated ensembles and neural network approach for classifying the data. Neural Networks, conclusively gave the perfect accuracy of 100% but the drawback with the neural networks is that, they take a lot time consuming. Considering the factor of time and accuracy, various algorithms have been developed and employed for specific purposes. This project focuses on developing a novel ensemble method and a neural network for the classification of biological data (data used here is that of EEG signal from Alzheimer's patients) and comparing the results of both the methods.

Alzheimer's Disease has always been a matter of awe for the doctors, researches and the lame man. Alzheimer's is a dementia condition, prevailing in a large proportion of the entire population. All the conventional methods for the clinical detection of Alzheimer's may not be considered completely fool proof. These methods include a MRI/CT of the brain of the patient and the image processing of the collected data and a cross validation by a certified physician. There are two major drawbacks here,

1. MRI and CT scans are really expensive, making it non-affordable by the common. A typical MRI in India costs around ₹10,000.
2. Image processing of these MRIs and CTs are so computationally costly that it requires a week's time to generate the result even in the developed nations like Canada under non-emergency conditions.

To overcome these drawbacks, we set foot into the signal processing approach of clinical detection of Alzheimer's Disease. Signal processing of EEG is comparatively less computationally costly and the EEG recordings are way cheaper as compared to the data acquisition through MRI and CT.

1.2 PROJECT OBJECTIVE:

The detection of Alzheimer's disease had always been a tedious task in the past, be it through the analysis of previous medical history or careful examination of the patient through various tests like brain imaging. Our proposed work utilizes the EEG signals of the patients to classify and the early detection of the disease. In our proposed work, we classify the surface EEG signals using advanced machine learning concepts like Evolutionary Extreme Learning Machine and Allocation Method using multiple classifiers and compare the results of the two methods to find the best one suited for the job.

CHAPTER 2: LITERATURE SURVEY

Alzheimer’s disease (AD) is an irreversible chronic neurodegenerative condition ranked as one of the top ten leading cause of death in the World. Moreover, AD is the most common cause of dementia, accounting for 60-70% of dementia cases round the world. Dementia is a broad name used for referring a wide range of cognitive problems (like memory loss, communication etc). In 2016, approximately 5.2 million people were diagnosed with AD, and this number is projected to grow to 13 million by 2050. Almost all the traditional techniques for Alzheimer’s detection was both ergonomically and computationally costly.

Even today, AD detection and diagnosis is done using medical history, laboratory tests and, brain imaging techniques like computed tomography (CT), and functional magnetic resonance imaging (fMRI). These clinical assessment methods, however, require experienced clinicians and exhaustive testing sessions, also the access to such techniques is limited even in developed countries (e.g., in Canada, the wait period for a non-emergency fMRI can be as high as six months). If this scenario exists in well developed nations, the condition of underdeveloped nations is pretty unquestionable.

To overcome these constraints, quantitative electroencephalography (qEEG) has been proposed as a promising tool to assist physicians in the diagnosis of AD [17]–[19]. Several effects of AD has been reported in the past literatures. The most distinctive of these effects include: i) shift in the EEG power spectrum (e.g., slowing of the EEG) [8]–[10]; ii) reduced inter-hemispherical spectral coherence [11]–[13]; iii) decreased EEG “complexity,” due to neuronal death or reduced communication between inter-cortical regions [19]; and more recently, iv) a change in neuro modulatory behaviour measured via EEG amplitude modulation analysis [17], [14]. Most of these effects are found to be related [15] and the diagnostic results are on par with more advanced neuroimaging techniques (e.g., fMRI) [16].

In the works of Zhe Chen et all [21], Non-Negative Matrix Factorisation Method is used for the early detection of Alzheimer's disease using clinical EEG recordings. But their empirical results are not persuasive enough to accept the procedure as a standard. Apart from this, the high dimensional EEG data is converted to low dimension using PSO for reducing the computation power required for classifying the various classes, as stated in the work of Adham Atyabi et al [23]. EEG signals are inherently prone to several artefacts and noise sources, like eye blinks, power grid interference, and hardware-inherent noise and sudden movement of the sensor, to name a few. Since artefacts may have detrimental effects on EEG-based AD diagnostics, a majority of the published works have utilized artefact-free EEG segments which have been selected by expert clinicians through exhaustive visual inspection manually which can never be as promising as that of an automated inspection. Thus, we plan to implement the Automated Artefact Removal (AAR) Algorithm as stated in [22].

Furthermore, in the proposed work, we are planning to use not more than 3 channels for EEG data acquisition which further reduces the computation time and the resources as proven in the work of Sabrina Ammar et al [24] for Seizure Detection. In the works of Ahmed Al-Ani et al [27], they have implemented Extreme Learning Machine and a max acuuracy of 89%. But the results of Pramila Vijayaraghavan et al proved an accuracy as high as 100% using Evolutionary Extreme Learning Machine on Spirometer dataset [25]. Based on the results of [25] we predict a higher accuracy using EELM. Also in the work of SašoKarakatic et al [26] they have proved that Allocation Method as a superior ensemble technique as compared to the conventional ensemble techniques of Bagging and Boosting. We have proposed an innovation in the proposed work of

Allocation Method using Feature Specific Outlier Detection which proved to be more efficient than the conventional Allocation Method proposed in [26].

The most recent literature on classification of Mild Cognitive Impaired and AD from Scalp EEG recordings is done using Deep CNN [28] and they have achieved an overall accuracy of 85% on AD-HC. Thus the proposed method with a predicted average accuracy of 95% can be safely stated to be a superior method.

CHAPTER 3 :MACHINE LEARNING IN PYTHON

Commercially there are a variety of tools for learning applications, but python is being the language of choice for most of the data scientists and engineers for a variety of reasons. First, python is very versatile and open source programming language. The developers need not spend a huge amount of money for python compared to the other learning approaches using MATLAB, Java, R etc. Second is that, python is a very high-level language, so the programmer can program with great ease. A program built in Python is too simple that, it's understandable to even a lame man. Next is that, Python is too mainstream and user friendly, an author can debug the program of another another author and vice versa. This might be possible in other programming languages like C, C++. The advantages of python is endless that, it is impossible to state a single reason for choosing python as the programming language for our application.

As per Wikipedia, Machine learning gives "computers the ability to learn without being explicitly programmed." [29]. Tom M. Mitchell provided a widely quoted, more formal definition of the algorithms studied in the machine learning field: "A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P* if its performance at tasks in *T*, as measured by *P*, improves with experience *E*." [30]. Machine learning tasks are typically classified into three broad categories, depending on the nature of the learning "signal" or "feedback" available to a learning system. These are [31]

- *Supervised learning:* The computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs.
- *Unsupervised learning:* No labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can be a goal

in itself (discovering hidden patterns in data) or a means towards an end (feature learning).

- *Reinforcement learning:* A computer program interacts with a dynamic environment in which it must perform a certain goal (such as driving a vehicle or playing a game against an opponent). The program is provided feedback in terms of rewards and punishments as it navigates its problem space.

Between supervised and unsupervised learning is *semi-supervised learning*, where the teacher gives an incomplete training signal: a training set with some (often many) of the target outputs missing.

3.1 PYTHON LIBRARIES USED

1. Pandas
2. Numpy
3. Sklearn
4. Keras
5. Matplotlib

While Pandas and Numpy are used for data cleaning and retrieval, sklearn is used for machine learning and development of the ensembles, keras is used for neural network working on Tensorflow backend and Matplotlib is used for plotting of values.

CHAPTER 4: DATA ACQUISITION AND CLEANING

Data acquisition from the normal subjects is done using *Neurosky Mindwave EEG Headset.* Normal subjects had a sample size of N=15. The data acquired from the normal subjects are subject to standardization and confidentiality protocol, thus the details and identification of the patients were not recorded. The data from the affected patients was acquired from a secondary source of data. Since the data isn't completely reliable, a request for a primary data source is submitted to the public hospitals in the city. The request and the ethical clearance for the project is under process.

The data collected, a raw EEG signal was cleaned, amplified and converted to digital form and then transferred to the PC using the standard Bluetooth 2.4 protocol. The acquired digital signal again cleaned for any transmission noise and then signal processed. Digital Wavelet Transform is applied to the signal to convert the Raw EEG into its components, high alpha, low alpha, beta, theta, delta waves. A total of 16 parameters is used for the process of classification

1. High alpha
2. Low alpha
3. Beta
4. Theta
5. Delta
6. Power Spectral Intensity(PSI)
7. Relative Intensity Ratio(RIR)
8. Petrosian Fractal Dimension(PFD)
9. Higuchi Fractal Dimension(HFD)
10. Hjorth mobility and complexity
11. Spectral entropy
12. SVD entropy

13. Fisher information
14. Approximate Entropy(ApEn)
15. Detrended Fluctuation Analysis(DFA)
16. Hurst Exponent.

The signal parameters obtained so far is converted into a dataset for further processing of the signal and classification. The data obtained is classified into three classes depending on the severity of Alzheimmer's Disease:

1. Normal (class 1)
2. Early stage (class 2)
3. Severe onset (class 3)

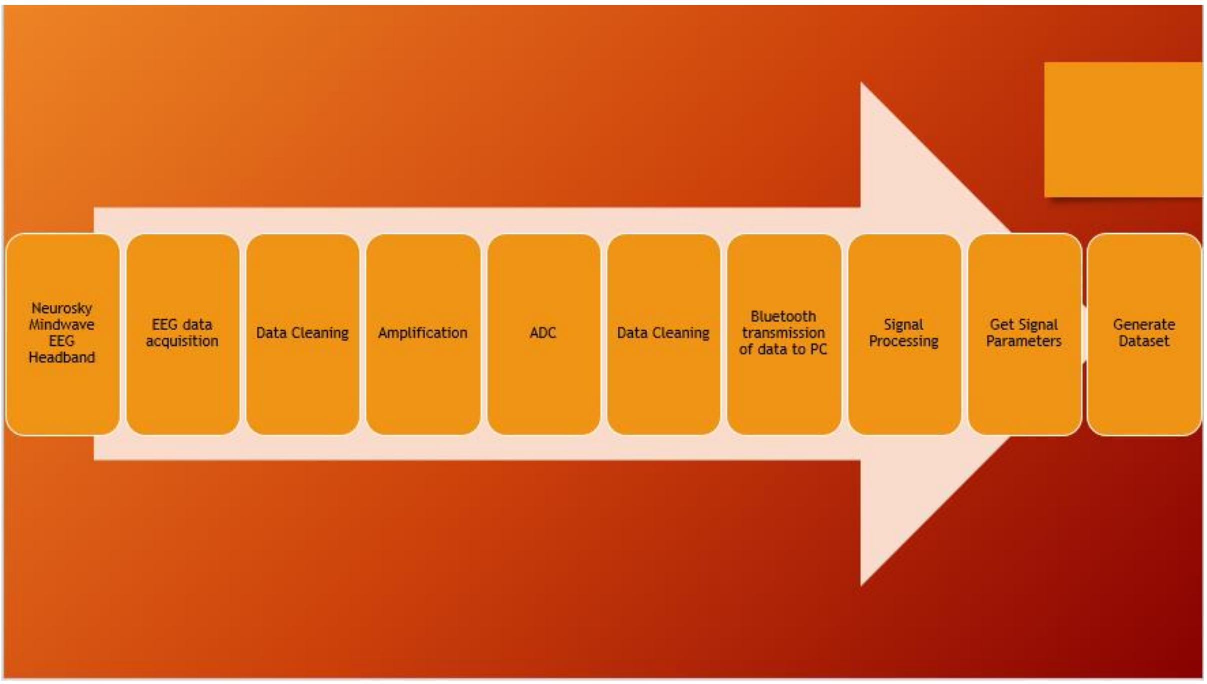

Fig 9: Data Acquisition and Cleaning

CHAPTER 5: PROPOSED WORK

The proposed ensemble technique is a combination of supervised, semi supervised and unsupervised algorithms. The contributions of this work are as follows:

- We propose a new Feature Segregator method for building a simple ensemble of three basic classifiers and one feature segregator. The feature segregator is based on distribution of data and uses gaussian distance between the subsequent points for outlier detection; it splits an original dataset into two disjoint subsets, of which each one these datasets into two disjoint subsets using a one-classSVM for anomaly detection, of which the larger subset is again split into two subsets using the same one-classSVM for fine anomaly detection, of which each one of the resulting three subsets are used to train one micro classifier using any of the existing classification algorithms.
- The feature segregator method proposes an easy-to-use approach, which does not require any further configuration or parameter setting from the user and works on any classification dataset. The feature segregator method thus ensures a simple and universal, yet competitive and scalable classification approach which provides very good classification performance.
- We perform extensive computational tests on a diverse set of bench mark datasets that demonstrate the strength of the proposed feature segregator method and show that it outperforms basic classifiers as well as standard classification ensemble methods.
- We discuss the possibility of extending our feature segregator method using different allocation approaches and algorithms, combinations of micro classifiers, and more than two micro classifiers in the ensemble.

Every dataset is composed of inliers and outliers. It is these outliers that cause a trouble during classification. While in most applications, these outliers can be eliminated from the original dataset, and then the cleaned dataset is worked upon, outlier is of major concern in a medical dataset. Outlier may be caused by some unknown or unnatural pathological or physiological event, or it might even be due to some movement artefact of the sensor. Thus, the outliers can never be neglected when it comes to medical data. In most of the previous applications, scientists have neglected outliers and worked on the inliers alone. In the proposed work, we identify the outliers and form a new subset of data for these outliers. Then an ensemble of classifiers is applied to both of the major subsets: Inliers and Outliers.

The goal of our research was to test whether our technique of segregation is a valid method for solving classification problems and how it performs on different datasets and using different classification algorithms.

CHAPTER 6: METHODOLOGY

6.1 OUTLIER CLASSIFICATION

Spatial distribution of the dataset gives a general idea of inliers and outliers. Box plotting the dataset gives a general picture of this spatial distribution. The following plots indicates the data distribution for every class for all of the 16 features.

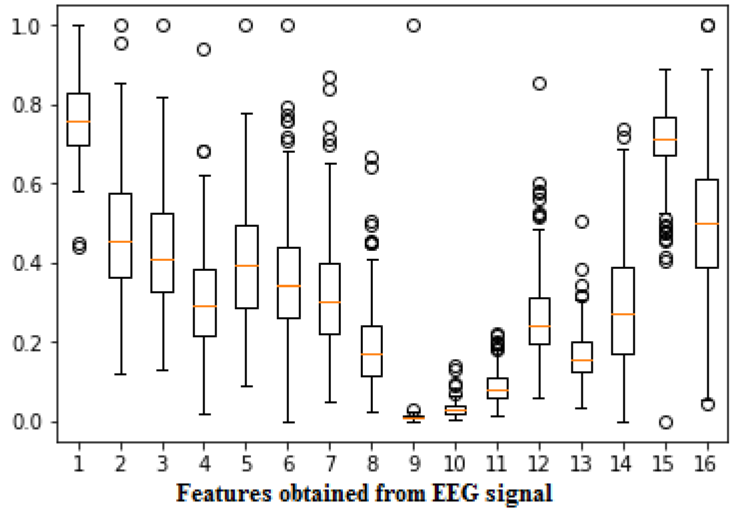

fig1: Box plot of class 1 data samples indicating spatial distribution

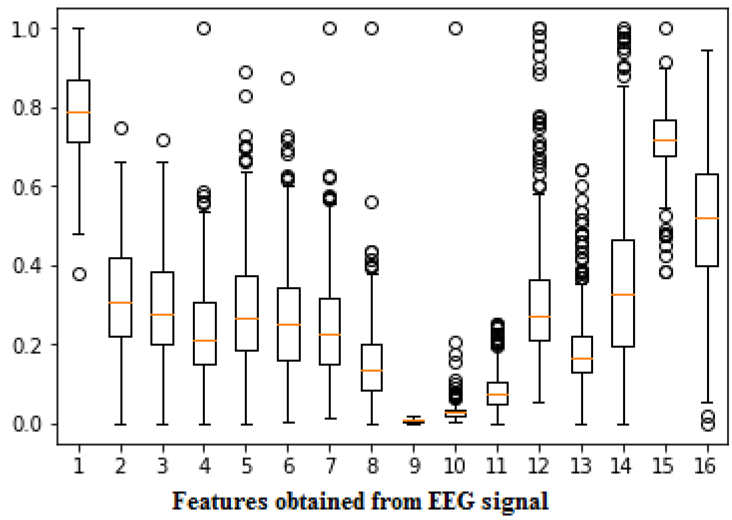

fig2: Box plot of class 2 data samples indicating spatial distribution

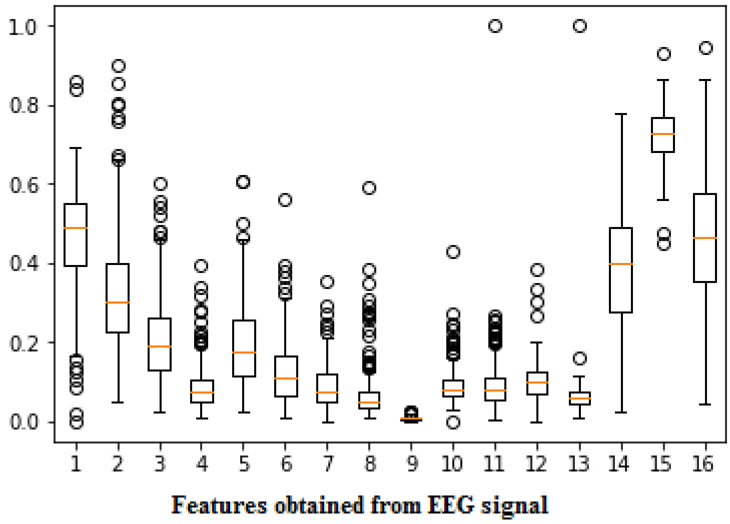

fig3: Box plot of class 3 data samples indicating spatial distribution

From these plots, it is clear which instances contribute to the inlier subclass and which instance to the outlier subclass. Any point lying outside the whisker is considered a outlier and grouped into separate subclass

ALGORITHM

1. Plot the box plot for every class present in the problem and identify the minimum and maximum values.
2. min ⟵ minimum; max ⟵maximum
3. For i in range (0, no. of class -1)

```
    For j in range (0, no. of features-1)
        For k in range (0, no. of samples)
            if (data[i][j][k] < min or data[i][j][k] > max)
                outlier.append(data[i][j[k]),
                dataset.del(data[i][j][k])
            end
        end
    end
end
```

fig4: Algorithm for identifying and classifying outliers into seperate subclass

6.2 FEATURE SEGREGATION

The subsets thus obtained are classified with an ensemble of micro classifiers. For the anomaly detection algorithm, we use one class support vector machine (SVM) algorithm and expanded the original SVM. The idea with one-classSVM is to teach the algorithm what the normal data looks like and to determine how far from the normal set of data is still considered normal. All instances outside of this normal buffer area are labelled as anomalies. Let us consider the training instances $x1$, $x2$,..., $xn \in X(1)$ from set X that is contained in the feature space F where $n \in N$ is the size of the X. The goal of the one-classSVM is to find the hyperplane, from feature space F that contains the normal data and represents the buffer zone for normality. As the hyperplane sometimes cannot be found with the use of feature space F, we use the feature map Φ that maps the set X to richer feature space F'. From numerous kernel functions available, we used the Gaussian radial basis function (RBF) for OCSVM.

The anomaly detection algorithm splits data to "normal" in-stances and "anomalies" and then forwards them each to the corresponding micro classifier. fig5 gives the general idea of how these micro classifiers are made. As we start with all of how these micro classifiers are made. The training data for this anomaly detection model is the majority class. The algorithm shown in Fig.6 calculates the distribution of the classes in the dataset and returns the labels of the majority classes. As is shown in the pseudocode(seeFig.6), the majority consists of the half (orhalf+1, in case of odd number of class labels) of the class labels that represent the largest portion of the whole dataset. Following always applies:

$$
\begin{aligned}
&X_1 \cup X_2 \cup \ldots \cup X_m = X \\
&X_i \cap X_j = \emptyset, \quad \forall i = 1, 2, \ldots, m, \quad \forall j = 1, 2, \ldots, m, \quad i \neq j \\
&X_i \subseteq X, \quad \forall i = 1, 2, \ldots, m
\end{aligned}
$$

Where *Xi* is set of instances of the same class, *X* is the whole data set and *m* is the number of all classes in the whole dataset. As we split the original dataset in the majority and minority datasets, this applies:

$$\Upsilon_{Ma} \cup \Upsilon_{Mi} = X, \quad \Upsilon_{Ma} \cap \Upsilon_{Mi} = \emptyset$$
$$X_1 \cup X_2 \cup \ldots \cup X_{mN} = \Upsilon_{Ma}, \quad X_{mN+1} \cup X_{mN+2} \cup \ldots \cup X_m = \Upsilon_{Mi}$$
$$mN = \left\lceil \frac{m}{2} \right\rceil, \quad mN \geq m - mN, \quad |\Upsilon_{Ma}| \geq |\Upsilon_{Mi}|$$

Yma is the majority dataset, *Ymi* is the minority dataset, *mN* is the number of majority classes and *m-mN* is the no. of minority classes.

Algorithm 1: GetMajorityClasses

```
Data: classification dataset X
Result: array of indexes of majority classes as majority_classes
begin
    num_of_classes ← NumberOfClasses(X)
    class_ratios ← DistributionOfClasses(X)
    half ← num_of_classes / 2

    if isOdd(num_of_classes) then
        half ← half + 1
    end

    majority_classes ← [ ]

    foreach i ∈ [1, half] do
        largest_class_i ← GetLargestClass(class_ratios)
        majority_classes[i] ← largest_class_i
        RemoveLargestClass(largest_class_i)
    end
end
```

fig.6: pseudocode for returning the majority class from the dataset

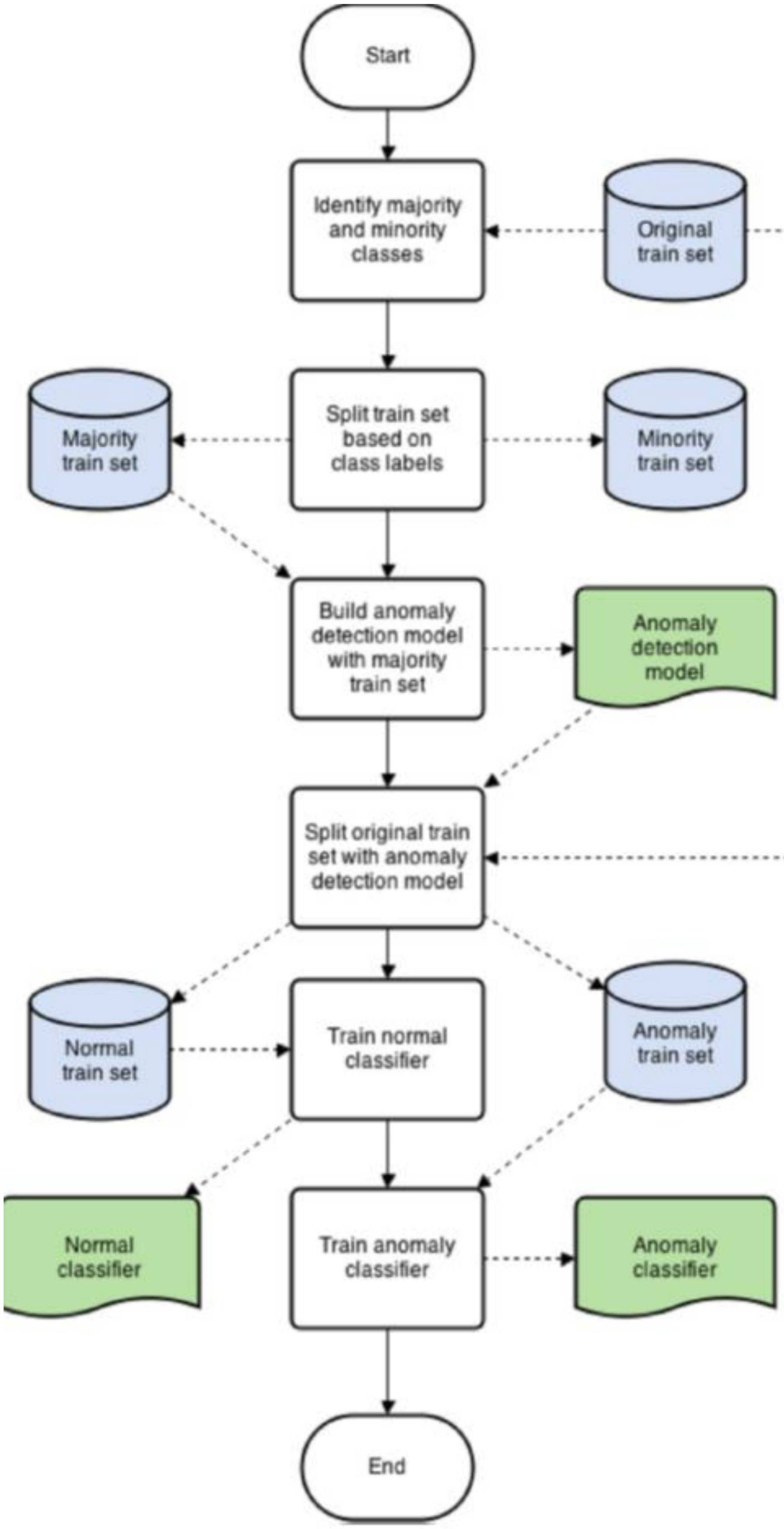

fig.5: The flow diagram for training of the anomaly detection model and the classifiers.

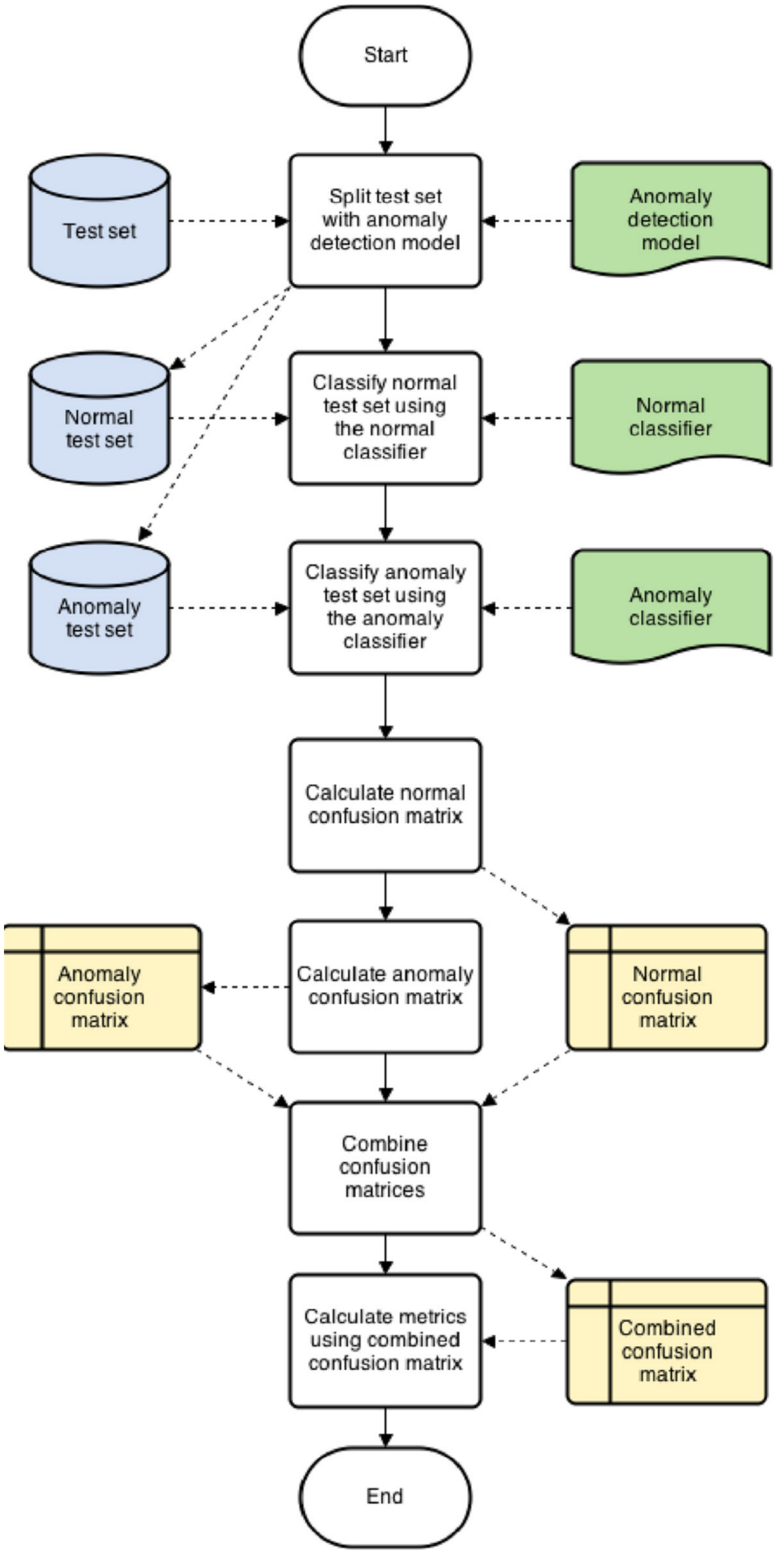

fig.7: The flow diagram for using and testing resulting solution with three models.

The data acquired is first converted into two subclasses: Inlier and Outlier subclass. Each of the subclass is given to the oneclassSVM and split into normal, anomaly and fine anomaly subsets. Each of the subsets is then given to a micro classifier and accuracy specific to each of the subset is obtained. The same procedure is applied for the outlier subset too. The accuracy of the two major subsets is then combined to get the accuracy of the ensemble method.

6.3 EVOLUTIONARY EXTREME LEARNING MACHINE

The signal parameters obtained from the feature segregator is used as the input neurons for the eelm neural network. One of the main advantages of using a eelm neural network is, it is comparatively fast as compared with the other neural networks with very high accuracy. A *sequential* model is used with the eelm for classification. Table 2 gives the eelm parameters

Table 2 : E-ELM Parameters

Parameters	**Values**
Model	Sequential
Input layer neurons	16
Number of hidden layers	5
Hidden layer neurons	48
Output layer neurons	2
Input layer activation function	Relu
Hidden layer activation function	Relu
Output layer activation function	Sigmoid
Loss function	Sparse categorial cross entropy
Optimiser	Rmsprop
Number of epochs	2
Batch size	7

CHAPTER 7: RESULTS AND CONCLUSIONS

Five well known classification algorithms have been used throughout the experiment: NaïveBayes, CART, SMO, OneR, and NBTree. These five classifiers were used as a basis for the following ensemble methods: MultiBoost, bagging, AdaBoost, and our proposed feature segregator. For the sake of comparison, the original classification results, obtained by the six classification algorithms, are presented also. In this manner, five different methods have been compared in the experiment: (original, allocation, MultiBoost, bagging, AdaBoost).

Table 1 summarises the results of the proposed method and proves that the feature segregator works efficiently than the other conventional methods.

Table1
Average accuracy results obtained by various ensembles of different classifiers over test data.

Method →	Original (none)		Allocation		MultiBoost		Bagging		AdaBoost	
Classifier ↓	Avg	Rank	Avg	Rank	Avg	Rank	Avg	Rank	Avg	Rank
Naïve Bayes	74.29	16.99	**97.9**	**14.30**	75.82	17.31	74.70	17.13	76.89	17.37
CART	80.08	18.90	**98.2**	**12.89**	80.09	18.85	80.35	19.10	79.91	19.11
SMO	79.99	15.94	**98.4**	**12.91**	80.16	16.64	80.14	15.91	80.66	16.03
OneR	63.39	22.59	**97.8**	**19.09**	64.50	23.65	65.01	22.40	64.49	23.26
NBTree	81.26	16.40	**97.9**	**12.03**	83.12	17.01	83.56	17.24	82.50	18.09
All classifiers	76.66	18.69	**98.0**	**14.51**	77.63	19.35	77.65	19.00	77.87	19.27

Average *f*-score results obtained by various ensembles of different classifiers over test data.

Method →	Original (none)		Allocation		MultiBoost		Bagging		AdaBoost	
Classifier ↓	Avg	Rank	Avg	Rank	Avg	Rank	Avg	Rank	Avg	Rank
Naïve Bayes	65.62	20.62	**97.75**	**20.26**	67.42	20.39	66.12	20.79	68.78	20.99
CART	72.48	22.34	**97.62**	**19.67**	72.45	22.36	72.57	22.43	72.46	22.40
SMO	68.98	22.92	**98.31**	**20.55**	69.58	23.11	69.01	23.01	72.14	20.98
OneR	50.64	27.68	**98.12**	**25.88**	51.17	28.45	51.73	28.08	51.77	28.38
NBTree	72.92	21.83	**98.56**	**19.40**	76.07	21.07	75.85	22.33	75.51	21.69
All classifiers	67.21	23.72	**97.84**	**21.30**	68.61	24.04	68.32	24.11	69.42	23.75

Fig.7 illustrates the aggregated accuracy results from all six classifiers. It can be seen that all three existing ensemble methods (MultiBoost, bagging and AdaBoost) scored very similar accuracy results in the case of all six classifiers

and improved the results of original classifiers (when no ensemble method is used) by approximately one percent on average. On the other side, it can be also seen that our feature segregator outperforms these three ensemble methods in the case of all six classifiers. In this manner, it further improved the results by approximately four percent on average.

Fig.8 illustrates the aggregated *f*-score results from all six classifiers. It can be seen that all three existing ensemble methods still outperform the original classifiers, but their *f*-score results are less similar than accuracy. Bagging improved the original classification results by approximately one percent on average, MultiBoost by approx. one and half percent on average and AdaBoost by approx. two percent on average. On the otherside, our feature segregator again outperforms all three ensemble methods in the case of all six classifiers. Its improvement of original classification results is thus more than seven percent on average.

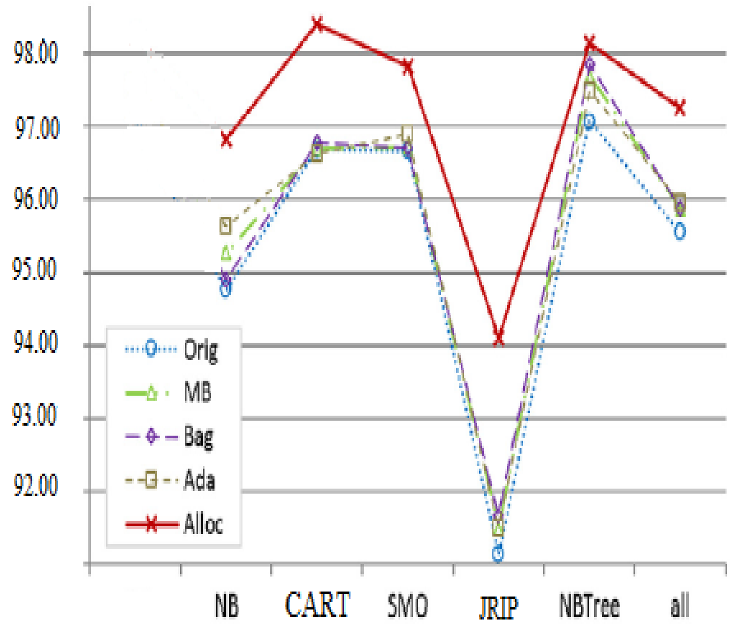

fig.7 : accuracy

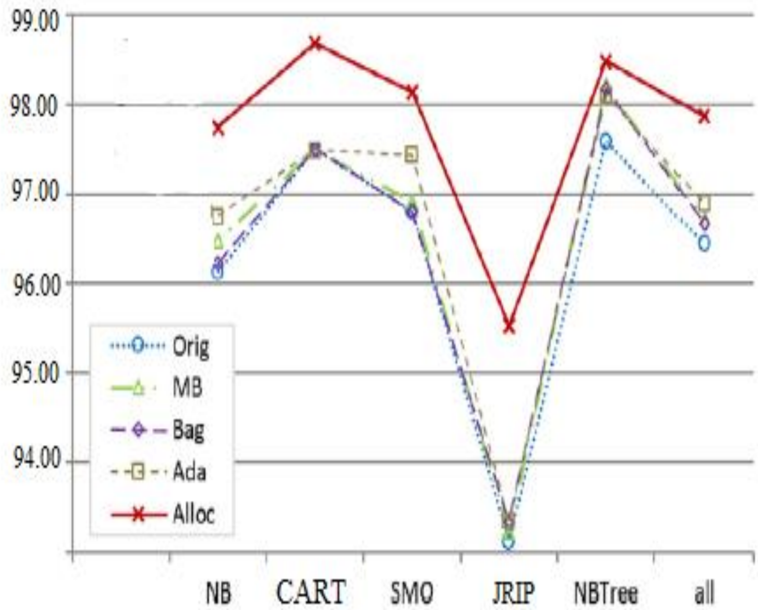

fig.8 : f-score

CHAPTER 8: REFERENCES

[1] Qin-Yu Zhu, A.K. Qin, P.N. Suganthan, Guang-Bin Huang, Rapid and brief communication Evolutionary extreme learning machine, Pattern Recognition 38 (2005) 1759 – 1763, Elseiver

[2] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: a new learning scheme of feed forward neural networks, in: Proceedings of the International Joint Conference on Neural Networks (IJCNN2004), Budapest, Hungary, 25–29 July 2004.

[3] SašoKarakatic, ViliPodgorelec, Improved classification with feature segregator and multiple classifiers, Information Fusion 31 (2016) 26–42, Elseiver.

[4] Adham Atyabi, Martin Luerssen, Sean Fitzgibbon, David M. W. Powers, Dimension Reduction in EEG Data using Particle Swarm Optimization, WCCI 2012 IEEE World Congress on Computational Intelligence June, 10-15, 2012 - Brisbane, Australia.

[5] Alzheimer's Association, "2013 Alzheimer's disease facts and figures," Journal of the Alzheimer's Association, vol. 9, no. 2, pp. 208–245, 2013.

[6] M. Hurd, P. Martorell, A. Delavande, K. J. Mullen, and K. M. Langa, "Monetary costs of dementia in the United States," New England Journal of Medicine, vol. 368, no. 14, pp. 1326–1334, 2013.

[7] WHO and Alzheimer's Disease International, "Dementia: a public health priority," World Health Organization, Tech. Rep., 2012.

[8] R. Brenner, R. Ulrich, D. Spiker, R. Sclabassi, C. Reynolds, R. Marin, and F. Boller, “Computerized EEG spectral analysis in elderly normal, demented and depressed subjects,” Electroencephalography and clinical europhysiology, vol. 64, no. 6, pp. 483–492, 1986.

[9] L. Coben, W. Danziger, and L. Berg, “Frequency analysis of the resting awake EEG in mild senile dementia of Alzheimer type,” Electroencephalography and clinical neurophysiology, vol. 55, no. 4, pp. 372–380, 1983.

[10] S. Giaquinto and G. Nolfe, “The EEG in the normal elderly: a contribution to the interpretation of aging and dementia,” Electroencephalography and clinical neurophysiology, vol. 63, no. 6, pp. 540–546, 1986.

[11] C. Besthorn, H. F¨orstl, C. Geiger-Kabisch, H. Sattel, T. Gasser, and U. Schreiter-Gasser, “EEG coherence in Alzheimer’s disease,” Electroencephalography and clinical neurophysiology, vol. 90, no. 3, pp. 242–245, 1994.

[12] T. Locatelli, M. Cursi, D. Liberati, M. Franceschi, and G. Comi, “EEG coherence in Alzheimer’s disease,” Electroencephalography and clinical neurophysiology, vol. 106, no. 3, pp. 229–237, 1998.

[13] E. P. Sloan, G. W. Fenton, N. S. Kennedy, and J. M. MacLennan, “Neurophysiology and SPECT cerebral blood flow patterns in dementia,” Electroencephalography and clinical neurophysiology, vol. 91, no. 3, pp. 163–170, 1994.

[14] F. J. Fraga, T. H. Falk, L. R. Trambaiolli, E. F. Oliveira, W. H. Pinaya, P. A. Kanda, and R. Anghinah, "Towards an EEG-based biomarker for Alzheimer's disease: Improving amplitude modulation analysis features," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013, pp. 1207–1211.

[15] J. Dauwels, K. Srinivasan, M. Ramasubba Reddy, T. Musha, F.-B. Vialatte, C. Latchoumane, J. Jeong, and A. Cichocki, "Slowing and loss of complexity in Alzheimer's EEG: two sides of the same coin?", International journal of Alzheimer's disease, vol. 2011, 2011.

[16] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "Alzheimer's disease: models of computation and analysis of EEGs," Clinical EEG and Neuroscience, vol. 36, no. 3, pp. 131–140, 2005.

[17] T. H. Falk, F. J. Fraga, L. Trambaiolli, and R. Anghinah, "EEG amplitude modulation analysis for semi-automated diagnosis of Alzheimer's disease," EURASIP Journal on Advances in Signal Processing, vol. 2012, no. 1, pp. 1–9, 2012.

[18] C. Babiloni, R. Lizio, F. Vecchio, G. B. Frisoni, M. Pievani, C. Geroldi, F. Claudia, R. Ferri, B. Lanuzza, and P. M. Rossini, "Reactivity of cortical alpha rhythms to eye opening in mild cognitive impairment and Alzheimer's disease: an EEG study," Journal of Alzheimer's Disease, vol. 22, no. 4, pp. 1047–1064, 2010.

[19] J. Jeong, "EEG dynamics in patients with Alzheimer's disease," Clinical neurophysiology, vol. 115, no. 7, pp. 1490–1505, 2004.

[20] M. Sarazin, L. de Souza, S. Leh´ericy, and B. Dubois, “Clinical and research diagnostic criteria for Alzheimer’s disease,” Neuroimaging clinics of North America, vol. 22, no. 1, pp. 23–32, 2012.

[21] Zhe Chen, Andrzej Cichocki and Tomasz M. Rutkowski, Constrained Non-Negative Matrix Factorization Method for EEG Analysis In Early Detection Of Alzheimer Disease, Icassp 2006.

[22] Raymundo Cassani, Tiago H. Falk, Francisco J. Fraga, Paulo A. Kanda, Renato Anghinah, Towards Automated EEG-Based Alzheimer’s Disease Diagnosis Using Relevance Vector Machines, 2014

[23] Adham Atyabi, Martin Luerssen, Sean Fitzgibbon, David M. W. Powers, Dimension Reduction in EEG Data using Particle Swarm Optimization,
WCCI 2012 IEEE World Congress on Computational Intelligence.

[24] Sabrina Ammar and Mohamed Senouci, Seizure Detection with Single-Channel EEG using Extreme Learning Machine, 17th international conference on Sciences and Techniques of Automatic control & computer engineering - STA'2016, Sousse, Tunisia, December 19-21, 2016

[25] Pramila Vijayaraghavani, Mahesh Veezhinathan, Performance assessment of Optimized Extreme Learning Machine based on Evolutionary Computing for Spirometric Data Classification.

[26] SašoKarakatic, ViliPodgorelec, Improved classification with feature segregator and multiple classifiers, Information Fusion 31 (2016) 26–42.

[27] Ping TAN, Weiping SA, Lingli YU, Applying Extreme Learning Machine to Classification of EEG BCI, The 6th Annual IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems June 19-22, 2016, Chengdu, China.

[28]Francesco Carlo MORABITO, Maurizio CAMPOLO, Cosimo IERACITANO, Javad Mohammad EBADI, Deep Convolutional Neural Networks for Classification of Mild Cognitive Impaired and Alzheimer's Disease Patients from Scalp EEG Recordings, 2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI.

[29] https://en.wikipedia.org/wiki/Machine_learning#Types_of_problems_and_tasks

[30] Mitchell, T. (1997). Machine Learning. McGraw Hill. p. 2. ISBN 0-07-042807-7.

[31]Russell, Stuart; Norvig, Peter (2003) [1995]. Artificial Intelligence: A Modern Approach (2nd ed.). Prentice Hall. ISBN 978-0137903955.