Data Mining to Help Aphasic Quadriplegic and Coma Patients

Sanjana Sahayaraj[1], Shomona Gracia Jacob[2]

[1]B.E CSE, SSN College of Engineering, Rajiv Gandhi Salai, OMR Road, SSN Nagar, Kalavakkam, Tamil Nadu 603110, India

[2]Assistant Professor, SSN College of Engineering, Rajiv Gandhi Salai, OMR Road, SSN Nagar, Kalavakkam, Tamil Nadu 603110, India

Abstract: ***One of the major challenges in medical history has been differentiating between a brain dead person and a person in coma. Tests like apnoea can prove dangerous sometimes. The motive here is to apply data mining techniques to brain data that are obtained from EEG to differentiate between brain dead and coma patients. Here we also focus on another important application in the medical field, which is the ability to recognise the slightest response of a quadriplegic patient with aphasia to their surroundings. This is possible from EEG data. Data mining is done using new data set and present repository to obtain greater degree of accuracy.***

Keywords: EEG, Aphasia, Quadriplegia, Coma, Brain waves, WEKA data mining tools.

1. Introduction

All the activities of a human body are controlled by the brain which is considered the most important organ. Most fields of medicine have some form of measurements to evaluate illness. But as all the brain activities are characterized in terms of brain waves, we need techniques for mining the EEG signal data thus calibrating the brain activity to aid evaluation of normal function versus illness. The field of neuroscience needs to incorporate accurate calibrations to enable better treatments in future. Quadriplegia with aphasia is one of the cases where the patient's activity can be monitored accurately only by brain waves. Although very slight muscle activities might be present, the patients may still not be able to express much through it. So, it is important to apply the right data mining techniques to get information about the person's state of mind from the EEG data sets. Another important issue under consideration is the main difference between a patient in coma and a brain dead person. Brain activity is measured in terms of the four waves of EEG namely alpha, beta, theta and delta waves .The frequency varies for these 4 types: beta (>13 Hz), alpha (8-13 Hz), theta (4-8 Hz), delta (0.5-4 Hz).An apt data mining technique is required to compare data sets and arrive at the right decision regarding the brain functionality of the patient under examination.

2. Data Mining Using Weka

The EEG data obtained are stored in the form of attribute relation file format (*.arff) and is given as input to WEKA data mining tool. WEKA (Waikato Environment for Knowledge Analysis) is a collection of machine learning algorithms. It contains tools for clustering, association rules and visualising the behaviour of data set fed to it. We feed the patient data to the WEKA data mining tool and check whether the patient's brain waves are being generated normally or are there a possibility of brain death. For an Aphasic Quadriplegia patient, we compare the intensity of the waves between specific distances, or in other words the spikes in the waves to see if they are generated with balance. And furthermore the clustering and association helps us to identify the response of the quadriplegic patient with aphasia to our actions/stimuli. Also the EEG data can be extracted in such a way that the electrodes can be concentrated in specific regions (single-unit recording) to detect even the weakest of brain waves. This is again used for data mining with WEKA tools.

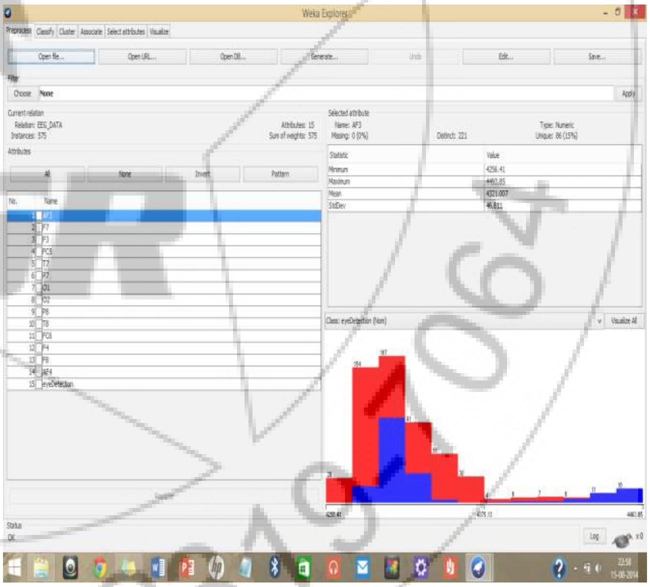

Figure 1: Pre-processing in WEKA

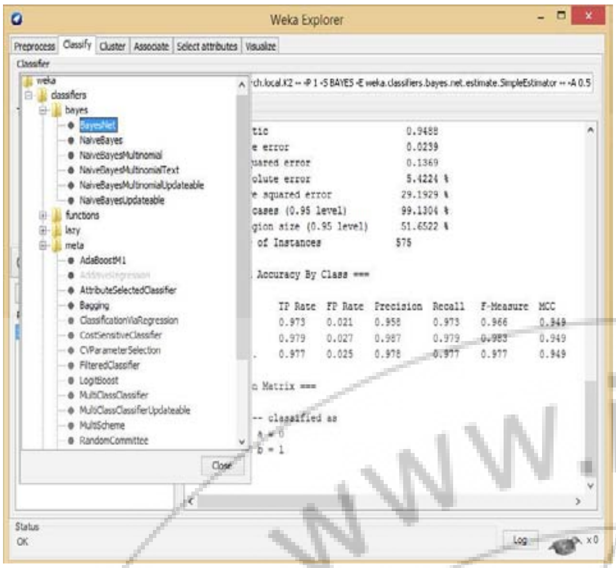

Figure 2: Classification in WEKA

3. Detect brain activity in Quadriplegic patients with Aphasia

Aphasic quadriplegic patients can listen, feel and reflect thoughts on things happening around them. But they cannot express themselves clearly because of the inability to move their limbs and the inability to express their thoughts in the form of articulation. But their inability to speak and express themselves can be overcome if the people around, try to understand their thoughts and feelings from their brain data. Such a patient will generate balanced brain waves. Even the higher order brain functions are performed with ease. In case, any imbalance is noticed in the form of a different brain wave pattern, the medical team can be immediately informed about the patient's condition. And nothing goes untold.

Humans react to some instances more than others. This can be identified by peaks in the waves. The spikes in the EEG wave produced can show some elevated response or action levels. This is used to give clues about an aphasic quadriplegic patient having active thoughts about a particular happening around him/ her.

From the wave:
z= mean of $((x_2-x_1) + (x_3-x_2) +....... (x_n-x_{n-1}))$, where each x_i corresponds to points on x axis bounding the flat regions. That is, x signifies a burst of electrical activity in the brain.
var = $\sum i = 1\ to\ n((x_i-x_{i-1}) / (x_i-z))$

This will give relative value of distance between spikes in the particular EEG wave.

The standard deviation is then taken:
S.D=$\sqrt{\text{var}}$ is the standard deviation , which will help bring the values of our patients close to the values obtained from normal people , aiding better characterization.

Now comparison is made between the brain characteristic data of our quadriplegic patient with aphasia and a normal person's data. That is, how the normal healthy people will react when he/she wants to express a thought inside their brain. This thought is one that is invoked by what happens around him/her. This aids betters characterization of the patient's emotion. If S.D of the quadriplegic patient approximately equals the S.D value of normal person, it means all their brain waves are being generated with a balance.

Brain state can be characterized in terms of psychophysical and EEG signal parameters. The EEG signal characteristics can be observed in brain lobes such as frontal lobe, parietal lobe, occipital lobe and temporal lobe. The activity in these regions is characterized by higher peaks in the brain waves generated.

For example, let us consider a scenario where the family members want to know which type of music would make an aphasic quadriplegic patient feel better. The EEG patterns are obtained while the patient listens to the different genres of music being played. The logistic regression algorithm in WEKA can be used to study how the patient feels about a particular genre of music played and predict which genre can make the patient happy. Few important attributes can be selected for this logistic regression. This will give some knowledge on which type of music is most preferred by the patient, and from now on that genre can be played.

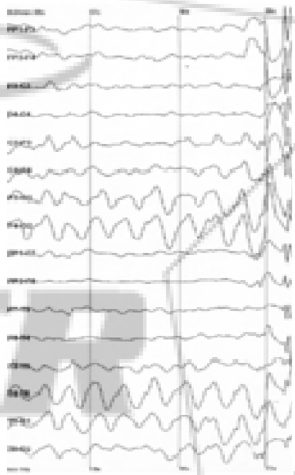

Figure 3: Wave discharges on an EEG

To detect the height or intensity of the emotion/response:
Data is pre-processed in WEKA
1) The data set that is obtained from the EEG when they are asleep and processed.
2) The data set is also collected when a person talks to the patient and expects the patient to feel something. This is processed too.

The height shall be a measurement of the intensity of the brain wave at some points, which in turn indicates responses at specific time instances. So, continuous recording of the brain waves of the patient is done.
val_1= (highest point of y axis of wave1)- (highest point at sleep) .
val_2= (highest point of y axis of wave 2) -(highest point at sleep). This is repeated for n waves obtained from the data set.
val =$(val_1)+(val_2)+(val_3)+...(val_n)$

Similarly we obtain a mean value from all the EEG waves of normal persons. The normal persons are tested under several levels of emotions such as happy to extremely happy. The same procedure is repeated for each level of happiness to get a value. This mean value is now used as a reference parameter to give hints on the patient's response. Cluster analysis plays an important role here:

[5] The greater the likeness (or homogeneity) within a group, and the greater the disparity between groups, the better or more distinct the clustering. Cluster analysis seeks as a result, a crisp classification of the data into non-overlapping groups. Now this value (val) of patient under examination is used to cluster. After required pre-processing, k means clustering is used. And once clustering is done , it can be found out to which group this person belongs. That is, the intensity of the emotion is known which shows whether the patient is extremely happy or just happy at that moment. We can also visualise the characteristics of each cluster.We can do this by using the 'visualise cluster assignments' option.

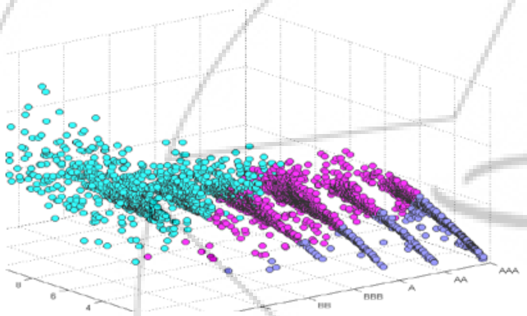

Figure 4: Visualise cluster assignments

Similar methodology is applied using clustering in WEKA for emotions such as sadness, boredom, etcetera .This will enable us to study and understand the intensity of various emotions of the patient. To facilitate this, data mining is done to identify which value closely matches. That is, to which cluster the patient's emotion belongs and from this, we can gain knowledge regarding the intensity of thoughts and feelings of the patient.

And furthermore, WEKA classifier algorithms such as Naïve Bayes classifier can be used to assess the patient's mood. This way, the patient's family members can understand the current state of mind of the patient much better and provide the needed moral support or company.

4. Differentiating the Brain Dead and a Coma Patient

There are some cases where the patient with coma eventually becomes brain dead. This means the person becomes entirely incapable of performing any more functions. To withdraw the life support from a coma patient who has been declared as brain dead, several tests have to be performed. And some tests might not be so effective and some other tests can even jeopardize the coma patient's life. But out of all of these, only the brain waves can give accurate information.

For this we need to perform data mining from repository and also from the patient under examination. We devise a formula here which decides whether the person is brain dead or not, based on relative value with respect to a threshold value set. Quantitative EEG (QEEG) applies multi channel measurements that can better determine spatial structures and localize areas with brain activity or abnormality. The results are often used for topographic brain mapping represented with colour maps in 2D and 3D to enhance visualization.

N. Sivanandan [2] used interpolation technique in which k nearest neighbour electrodes were involved. The accuracy and the smoothness of the picture generated improves as k increases. The suggested methodology was to use one main electrode and k nearest electrodes .And a ground electrode is always needed to show the differential voltage.

But a technique can be used where k is set to 6 constantly and each electrode serves as the nearest neighbour to each of the remaining 5 electrodes at one point of time and single–unit recording is used. This gives a more accurate statistics regarding brain activity. All these 6 electrodes are concentrated over an area. An EEG signal between electrodes placed on the scalp consists of many waves with different characteristics.

By using groups of electrodes, we can split the entire portion into distinct small regions. And these regions are circular in nature. This can provide clues regarding a particular region where even slightly detectable brains waves are being generated. Minimal configuration for mono electrode measurement consists of one active, one reference and one ground electrode.

Here, the cerebral region is taken into consideration. The cerebrum is a major part of the brain, controlling emotions, hearing, vision, personality and much more. It controls all voluntary actions. Groups of 6 electrodes are chosen. We check for any active brain waves over the entire cerebral region, since "brain death" is synonymous with "cerebral death". And also, the highest influence to EEG comes from electric activity of cerebral cortex. Brain activity changes in a consistent and recognizable way when the general status of the subject changes, as from relaxation to alertness [7].

Let us consider 5 electrodes around one main electrode, under consideration at this time. Let d1, d2, d3, d4, d5 be the distance of the five other electrodes from the main electrode. The value of these parameters keeps changing for each trial. All are placed around a particular region. So the relative potential of the first main electrode is given by:

$r_1 = ((d1)pt_1+(d2)pt_2+(d3)pt_3+(d4)pt_4+(d5)pt_5)/ \text{potential}^2$

Here 'pt_i' refers to the potential of a particular neighbouring electode 'i' and 'potential' refers to the potential of the central electrode under consideration right now.

Now the main electrode is made as an electrode for comparison that is the neighbouring electrode. A new main electrode is chosen. And this way each time, r_2 ,r_3 ,r_4,r_5 and

r_6 are calculated which are the relative potentials of the other main electrodes.

Finally to characterize this circular region:

Activity $=\sum i = 1 \; to \; 6$ (r_i)/ 6 .

If this value is above a threshold , it means this EEG data obtained shows some evidence regarding brain waves being generated .The ARFF file has this value as one of the attributes along with brain data. Based on this the class of the .arff file can be @ATTRIBUTE brain_activity {0,1} where 0 indicates that the value is below threshold and 1 implies that value is above the threshold. This is provided for pre-processing in WEKA. The red portions shows which circular regions of the brain are active than the others. If a few regions come under red , then the person has at least some brain activity and is not brain dead. From this we can identify whether the person is really brain dead , or only in coma and has a chance of recovering. When all the circular regions are characterized by blue colour, it means all the brain functionalities are lost and there is no hope of recovery and can be declared a brain dead person. Rather than seeing activity around a point, this methodology concentrates on finding the activity in a particular region. And so even the faintest of brain waves around any circular region can be detected. If it is possible to successfully identify even the minute brain activity present, then there won't be any error in declaring a person as brain dead.

5. Other Applications and Future Work

The work can be extended in future to cases like Dementia, Delirium and Encephalopathy.

Dementia is a medical condition which causes the reduction in ability to learn reason or recall past experiences. All kind of dementia is caused by brain cell death or head injury. The signs and symptoms of Dementia are abnormal behaviour, memory loss, hallucination, agitation, confusion in decision making and learning disability. All these can be monitored over time and the patient's data studied through data mining. This can help monitor the evolution of the disease.

Delirium represents an organic decline from a previously attained baseline level of cognitive function. It is typified by fluctuating course, attentional deficits and generalized severe disorganization of behaviour. It typically involves other cognitive deficits, perceptual deficits, altered sleep-wake cycle, and psychotic features such as hallucinations and delusions. These cognitive functions are analysed through data mining from a large collection . Delirium not being a disease by itself but rather caused due to chemical substances used as part of the treatment for an underlying disease. By analysing the data sets, we can identify the main chemical substances giving rise to such a clinical syndrome. EEG data of patients who take these chemicals as part of treatment are used to extract data, they are classified using WEKA tool. Now the EEG data set of patients suffering from delirium is also analysed in various stages and compared to see how, at which stage and due to which organic chemicals, these disorders are caused.

A similar approach is applied to Encephalopathy. Encephalopathy does not refer to a single disease, but to a syndrome of global brain dysfunction. This syndrome can have many different organic and inorganic causes.

6. Conclusion

Thus by characterizing the data through mining and thorough analysis , we try to differentiate a person who is brain dead and who is still in coma and has hope to return to normalcy. By studying thoroughly the attributes through visual classification (identify attributes) and clustering (group those with similar values) the data mining enables us to understand the thoughts of people who have been affected by severe case of quadriplegia which is accompanied by aphasia.

References

[1] A collaborative filtering approach for quasi-brain-death EEG analysis.- Yili Xia ,Ling Li ; Jianting Cao ; Golz, M. ; Mandic, D.P
[2] Quantitative EEG Analysis Technique for Computerized Digital Brain Signals – N.Sivanandan
[3] Classification of EEG based Diseases using Data Mining – DeepikaKundra, Babita Pandey.
[4] Analyzing EEG of Quasi-Brain-Death Based on Dynamic Sample Entropy Measures -Li Ni,Jianting Cao, Rubin Wang.
[5] Comparison the various clustering algorithms of weka tools - Narendra Sharma, AmanBajpai , Mr. RatneshLitoriya .
[6] Intelligent computing methods for the interpretation of neuropsychiatric diseases - MohitGangwar, R. B. Mishra, R. S. Yadav.
[7] Fundamentals of EEG Measurement – M.Teplan.
[8] Ramani, R. G., & Jacob, S. G. (2013). Benchmarking Classification Models for Cancer Prediction from Gene Expression Data: A Novel Approach and New Findings. Studies in Informatics and Control, 22(2), 134.
[9] Ramani, R. G., & Jacob, S. G. (2013). Improved classification of lung cancer tumors Based on structural and physicochemical properties of proteins using data mining models. PloS one, 8(3), e58772.
[10] Ramani, R. G., & Jacob, S. G. (2013). Prediction of P53 mutants (multiple sites) transcriptional activity based on structural (2D&3D) properties. PloS one, 8(2), e55401.
[11] GeethaRamani, R., &Gracia Jacob, S. (2013). Prediction of cancer rescue p53 mutants in silico using Naïve Bayes learning methodology. Protein and peptide letters, 20(11), 1280-1291.
[12] Gracia, J. S., &RamaniGeetha, R. (2013). Design and Implementation of a clinical data classifier: A Supervised learning approach. Research Journal of Biotechnology Vol, 8, 2.
[13] Jacob, S. G., Ramani, R. G., & Nancy, P. (2013). Discovery of Knowledge Patterns in Lymphographic Clinical Data through Data Mining Methods and Techniques. In Advances in Computing and Information Technology (pp. 129-140). Springer Berlin Heidelberg.

Author Profile

Ms. Sanjana Sahayaraj is currently pursuing her B.E in the Department of Computer Science and Engineering(CSE) at SSN College of Engineering, Chennai , India . She is a recepient of gold medal for excelling in academics and a merit scholarship holder. Her areas of interest include Biological Data Mining , Brain Informatics and Algorithm Design.

Mrs. Shomona Gracia Jacob is Assistant Professor, Department of CSE, SSN College of Engineering, Chennai, India. She has submitted her Ph.D thesis at Anna University in the area of Biological and Clinical Data Mining. She has more than 25 publications in International Conferences and Journals to her credit. Her areas of interest include Data Mining, Bioinformatics, Machine Learning, and Artificial Intelligence. She has reviewed many research articles on invitation from highly reputed refereed journals. She is currently guiding under-graduate and post-graduate projects in the field of data mining and intelligent systems.