Feature Extracted Classifiers Based on EEG Signals: A Survey

Faris A. Abuhashish, Mohd S. Sunar, Hoshang Kolivand, Farhan Mohamed and Dzulkifli B. Mohamad

MaGIC-X (Media and Games Innovation Centre of Excellence), UTM-IRDA Digital Media Centre Universiti Teknologi Malaysia, 81310 Skudai, Johor MALAYSIA
Malaysia, shahrizal@utm.my

Abstract: This paper discusses and surveys feature extracted classification algorithms that used in the field of Brain-Computer Interface (BCI) systems, based on ElectroEncephaloGram (EEG) signals. It presents the most common algorithms which have been employed in the context of BCI, then describes and categorizes them depending on highlighted properties. On the basis of the literature of this study, we evaluated as well as summarized the algorithms in terms of accuracy and performance, in order to facilitate the choosing process of the best classifier, regarding an appropriate feature extracted in the castoff in EEG-based BCI investigation. A significant contribution of this study, as well as highlighting the current widely used methods which are posted in the tables in Appendix; we have discussed and contributed a significant way in order to facilitate choosing the best classifier to be used within the field of BCI. Widely used techniques and algorithms are categorized in a final table which could explore many ideas for the researchers who are looking for a most suited method.

Keywords: BCI; EEG; Feature Extraction; Classifier

1. Introduction

Brain computer interface BCI is the advance rebellion in the technology of the media of communication, in other word, it is known as a communication structure that does not need any activities of peripheral muscular (Wolpaw, 2002). Certainly, BCI arrangements permit a matter to direct the feature extraction to an electronic device with the help of brain activity (Vaughan, 2003). These states of features identified above denoted through the signals of ElectroEncephaloGraphy EEG. Where there is various mental activities have been considered through the investigators (emotions, movement, motor imagery and talking) (Petrantonas, 2010).

Meanwhile, the states of mental activity show a significant part in the daily existence of human creatures; it is essential and significance of the automatic mental activity recognition that has developed with growing part of the application of the human computer interface (Liu, 2010). The recognition of mental activity happens to many causes such as speech, text, gesture, facial expression, motor imagery and muscle movement or etc. Lately, extra researches were completed upon emotion recognition from EEG.

Auto mental activity recognition based on EEG signals; have to be more growth as much as more considerable of new brain based technology that reflects the interaction between human and digital media that is called human computer interaction HCI. This paper focuses on showing the recognition techniques of inside mental activity based on signals of EEG, with various mental activities might regulate and govern the humanoid languages. There are changed mental activity arrangements suggested by various researchers. Consequently, the foremost purpose of this paper is to evaluate the classifiers of extracting mental activity of altered feature algorithms in the castoff in EEG-based BCI investigation and assemble the outcome in tables.

3. Brain-Computer Interface Signals and Preprocessing

i. BCI

The area of Brain-Computer Interfaces (BCIs) has grown massive fame throughout the previous few ages. BCIs are currently offered for an extensive variety of presentations. By way of in several HCI application, BCIs can too profit starting adjusting their process to the mental activity state of the operator. It is more benefit of BCIs of taking contact to brain motion which can make available important vision into the operator's mental activity state. As an outcoming, it is conceivable to learn the mental activity based BCI through reviewing the extraction feature and arrangement, in order to put on the consequence in countless grounds i.e.: medical, games, engineering, etc.

ii. EEG Based BCI

Knowing that the electrical signals generated by brain is considered to be essential since it carry out the information of not only the brain function but also the information of the full body status and the entire state of mental activity, this kinds of assumptions are taken from an early stage (Liu, 2012). Furthermore, it offers the motivation to relate the procedures of advanced digital signal processing toward the signals.
of electroencephalogram (EEG) determined by the use of the human subject brain (Sanei, 2008).

By means of the recognition of the mental activity is almost fresh zone, a standard record of the signals of EEG for altered mental activity is desirable to be active, which might be castoff for additional research on the recognition of mental activity EEG-based. Recognition of the EEG-based mental activity algorithms permits perceiving of the inside human thoughts. It may possibly be castoff in the interfaces of the modern Human-Computer besides it can be applied in countless areas as entertainment, virtual collaborative spaces, education, etc. The main difficulty in BCI is the extrication of the signals control from the circumstantial EEG. Revelation and estimation of several structures in dissimilar domains will then afford the control signals. Therewith, in order to initiate the EEG signals it requires being preprocessed due to noisiness that affects the acquired signals through much external and internal interference. The conditioning of the information such as pre-whitening could too be needed before the execution of the algorithms of the source separation (Sanei, 2008). Refer to table 6.

2. Diversified Discussion of EEG Signal Preprocessing Techniques and Feature Extracted Classifiers

The assortment the furthermost suitable classifier aimed at a specific BCI arrangement, it is indispensable to obviously recognize what features are castoff, what their assets are and in what way they are handled. A countless change of features have been tried to enterprise BCI for example the value of the EEG signals largeness (Wolpaw, 2003), Band Powers (BP) (Peters, 2001), Power Spectral Density (PSD) ideals (Müller, G. R), Autoregressive (AR) and Adaptive Autoregressive (AAR) parameters (Müller-Putz, 2003), Time-frequency features (Pfurtscheller, 2005) and inverse model-based features (Vidalurre, 2006; Del, 2002.). Percentage of types measured in altered EEG directly frequency bands generally castoff in mental activity state arrangement aimed at the smallest elementary of mental activity stated above.

i. Techniques Based on Feature Extraction

This section evaluates the algorithms methods of EEG signal preprocessing castoff in types of extraction & arrangement. Particular researchers are governed by the various mental activities so as to achieve their desired. There are a cumulative amount of researches completed on the algorithms of EEG-based mental activity recognition. In work (Lin, 2009), Fourier Transform for short-time was applied for the extraction of the feature and SVM which means Support Vector Machine that tactic was active in the direction of categorizing the information into unlike mental activity approaches. The outcoming was 82.37% correctness to differentiate the sensation of mental activity. The rate of performance was 92.3% acquired in (Bos, 2006) by employing the analysis of Binary Linear Fisher’s Discriminate BLFD and the state of mental activity between positive to calm and negative to calm were contrasted.

The researcher in (Pati, 2010) suggested an algorithm of a subject-dependent utilizing the ratio of delta, theta, alpha, beta, gamma and mu (refer to table 6), in order to distinguish the control levels it high and low control of EEG. This line of attack would tolerate the mental activity recognition of real-time (Ismaahfezie, 2013) which is clear through the changed levels control. Discrete Fourier Transform has been used as a first method, and at that time the concentration of power spectral was calculated for the bands of frequency. Lastly the power ratio of the waves was planned as a hypothesis. To categorize the control levels of low and high dominance, we applied the classifier of Support Vector Machine (SVM). As an outcome for the algorithm anticipated that the correctness ranges vary from 73.64% to 75.17%.

Petronatokis et al. (2010) purposed at given that a new technique of feature extracted classification (Kharazi, 2013). The gifted presentation of the HAF-HOC faces the significance of the signals of EEG inside the endeavor of comprehending supplementary realistic, sentimental interface of human-machine. A novel filtering method which called Hybrid Adaptive Filtering (HAF) aimed at a proficient extraction of the EEG-characteristics emotion-related was advanced by enforcing Genetic Algorithms to the Experimental Approach of Decomposition-based demonstration of the EEG signals. In calculation, the Higher Order Crossings (HOCs) analysis was active for the realization of extraction of feature from the HAF-filtered signals. The outcoming of the HAF-HOC analysis and unlike classifiers castoff in is: (QDA 77.66), (SVM 85.17).

Yuen (2011) anticipated a numerical grounded preparation for human mental activity arrangement based on electroencephalogram (EEG). Leading technique aimed at the mental activity arrangement, common mental activity numerical features. The first one refers to the raw signals, second one refers to the standard deviation of the raw signals, third feature refers to the complete values of the leading changes of the raw signal, The complete values of the leading changes of the normalized signals is related to the fourth feature, The complete values of the second changes of the raw signals corresponds to fifth feature and final feature refers to the complete values of the second changes of the normalized signals. They are calculated from the EEG information. The next arrangement technique is by
means of the system of back-propagation neural. As an outcome the general amount accomplished as high as 95%.

Researcher recommended (Khalili, 2009) a multimodal fusion among outlying signals and the brain and for detection common mental activity. It is all about outlying signals of the moving average filter. On the other hand filtered EEG signals and Genetic Algorithm is castoff toward resolving the trouble of feature space of high dimensional, meanwhile at every electrode determined the dimension of correlation is a resilient nonlinear feature. The outcome of the classification exactness for EEG, outlying and together brain and outlying indications of one participant is: for EEG 63.33%, for outlying 55%, for together brain and outlying indications 61.8%, rendering toward the examined consequences. Researcher in (Bos, 2006) designated a procedures of assignment directed to distinguish (feature classification & extraction) mental activity from brain signals determined through the Brainlquary EEG PET method. Hypothesis castoffs to happen in this paper are: the recognition of Modality (electrode placement), the recognition of mental activity (filtering of the bandpass, extraction of the feature, classification), positioning of the electrode, modality influence, EEG feature (by means of Fourier analysis: toward be situated on the way to look at precise regularity bands). The outcome of this paper (Bos, 2006) is, classification of the Binary modality rate is over 80% appears to be achievable. Aimed at binary arrangement rate displays most of over 90% for every one of the verified feature. So for a visual motivation it seems to be extra hard to classify than their audiovisual & audio. The maximum presentations classifications were gained by means of the band control.

Takahashi (2003) equated the success of neural system and SVM in categorizing the minimum of two mental activities which used Numerical features and neural system with SVM, correspondingly. By way of an outcome the rates of recognition attained for statistical features are 62.3% and 59.7% for neural network and SVM correspondingly.

Broek (2011) emerging the recognition system of the least mental activity by SVM to categorize least mental activity founded on numerical features calculated within the used raw signal, SVM based on Numerical Features. The rate of recognition is 41.7%. Work (Ishino, 2003) suggested a structure for sensing the valuation. Feature of the anticipated structure is acting an arrangement of the least mental activity through the relating of: Fourier Fast Transform (FFT), Principal Component Analysis (PCA), Wavelet Transform and Variance of Mean to acquire excerpt features of EEG information. Neural system is castoff for ordering of common mental activity. The outcome is maximum success rate is 67.7%. (Ryu, 1997) categorized the extraction of the mental activity through Fourier Fast Transform (FFT) based Support Vector Machine (SVM). The deduction consequence stands as the recognition rate 80% is accomplished.

Revising the classification algorithms castoff to scheme the BCI systems based on EEG. We momentarily show the frequently engaged algorithms and designate their serious assets. On behalf of the literature, we relate them in standings of presentation and deliver strategies to select the appropriate organization algorithms for an exact BCI. In the following, the most common classification algorithms used in BCI based preprocessing are systematic discussed.

ii. Classification Algorithm based on BCI

In order to select the furthermost suitable classifier for a specified set of structures, the possessions of the obtainable classifiers need to be identified. This section makes available of the classifier taxonomy. Similarly, it pacts through two difficulties of the classification exclusively related to the BCI research, to be exact, the dimensional curse and the tradeoff of Bias-Variance. Furthermore, this section reviews the arrangement algorithms castoff in BCI structures. They are separated into five dissimilar types; linear classifiers, nonlinear Bayesian classifiers, neural networks and nearest neighbor classifiers. The subsequent are momentarily designated and their most significant possessions for BCI presentations are underlined.

i. Linear Classifiers

These classifiers are perhaps the greatest general algorithms castoff in BCI presentations. Linear classifiers are discriminate procedures that practice linear occupations to differentiate between the classes. In BCI structures there are two categories; the Support Vector Machine (SVM) and Linear Discriminate Analysis(LDA).

a.1. Support Vector Machine; An SVM similarly that apply a discriminate hyperplane to detect periods (Burges, 1998; Bennett, 2005). The designated hyper-plane is the one that makes the most of the limitations, i.e., the space starting from the nearest preparation facts. Usually, a high achievement result will be accomplished, when applying the mentioned classifier on a large amount of synchronous BCI problems (Rakotomamonjy, 2005; Garrett, 2003; Blankertz, 2002). SVM have set very respectable outcomes for BCI presentations (Kaper, 2004). By way of LDA, SVM has been used to multiclass BCI difficulties by means of the OVR approach (Blankertz, 2002). SVM have numerous compensations.
Essentially, appreciations to the boundary enlargement and the regularization period, SVM are identified to have decent oversimplification possessions (Duda, 2012), to be oblivious to overtraining and also considering the curse-of-dimensionality. Lastly, SVM has insufficient hyper-parameters that are essential to be distinct through hand, specifically, the parameter C of regularization and the thickness of RBF if using kernel 2. These compensations are increased at the outflow of a little quickness of performance.

Figure 1. Shows SVM. CHRISTOPHER J.C. BURGES,"A Tutorial on Support Vector Machines for Pattern Recognition".

**a.2. Linear Discriminate Analysis;** the objective of LDA is to practice hyper-planes to isolate the information on behalf of the altered classes (Fukunaga, 1990; Pfurtscheller, 1999). Aimed at a two-class difficult, the course of an article vector rest on which crosswise of the hyper-plane the direction is there. The unraveling hyper-plane is attained through looking for the prediction that make the most of the coldness amid the two classes’ resources and diminish the interclass's modification. This method has an actual little computational condition which creates it to be appropriated for connected BCI structure. Furthermore this classifier is modest to apply and mostly delivers decent consequences. Therefore, LDA has been castoff through accomplishment in a countless amount of BCI arrangements for example motor descriptions founded BCI (Bostanov, 2004), speller of P300 (Scherer, 2004), multiclass (Garrett, 2003) or asynchronous BCI (Garcia, 2007).

**ii. Neural Networks**

The Neural Networks (NN), is composed through linear classifiers, the grouping of classifiers typically castoff in BCI investigation. Let us use the recollection that a NN is a gathering of numerous reproduction neurons which allows constructing nonlinear pronouncement limitations (Chiappa, 2004). Multilayer Perception, is the greatest extensively castoff NN for BCI. An MLP is collected of numerous coats of neurons; an involvement layer, conceivably one or quite a lot of unseen layers, and an amount produced layer. Neural Networks and consequently MLP are widespread estimates. Further to the circumstance, they can categorize any amount of sessions; this creates NN actual supple classifiers that can get used to a countless diversity of difficulties. Subsequently, MLPs which are the greatest widely held NN castoff in organization, have been functional to very nearly all BCI difficulties for instance binary (Rahman, 2012) or multiclass (Anderson, 1996), synchronous (Palaniappan, 2005) or asynchronous BCI (Rabiner, 1989).

**iii. Nonlinear Bayesian classifiers**

Here this paper will deliberate unique one of the classifier of nonlinear Bayesian that castoff in BCI. Hidden Markov Models (HMMs); are standard self-motivated classifiers in the ground of recognition speech (Friedman, 1997). An HMM is a generous of probabilistic machine that can deliver the chance of detecting an agreed arrangement of story vectors. Each national of the automaton could modelize the possibility of detecting an assumed story vector. HMMs are flawlessly appropriate algorithms for the organization of period sequence. HMMs are not much extensive inside the BCI public but these educations exposed that they stood talented classifiers for BCI structures.

Figure 2. Neural Networks algorithm (Lotte, 2006).

Figure 3. Hidden Markov model example (Huang, 2001)
iv. Nearest Neighbor Classifiers

The classifier obtainable in this section is comparatively simple. It contains in allocating a feature vector to a course agreeing to its adjacent neighbors. This neighbor could be a feature vector commencing the exercise regular as in the circumstance of k Nearest Neighbors (kNN). Where k is Nearest Neighbors; KNN algorithms are not actual widespread in the BCI public, perhaps since they are recognized to be actual subtle to the curse-of-dimensionality (Müller, 2004), which completed them miss the mark in numerous BCI experimentations (Borisoff, 2004). Though, when used in BCI arrangements through low-dimensional article vectors, kNN might demonstrate to be well-organized (Jain, 2000). On the other hand, the objective of this performance is to allocate a hidden fact the leading class between her k adjacent neighbors inside the exercise usual. For BCI, these adjacent neighbors are typically attained by means of a metric space, e.g. (Rakotomamonjy, 2005). Through an adequately extraordinary worth of k and sufficient exercise examples, kNN can estimated any purpose which allows it to harvest nonlinear pronunciation limitations.

Figure 4. Usage of Nearest Neighbors Classification Sample (Huang, 2001).

v. Other Classifiers

In greatest identifications connected to BCI, the organization is attained by means of a particular classifier. A contemporary leaning, though, is to practice more than a few classifiers, combined in dissimilar habits. The classifier mixture approaches castoff in BCI submissions are the succeeding:

Stack Classifier; Loading contains in spending quite a few classifiers, every one of them sorting the participation feature direction. These classifiers are named level-0 classifiers. The production of every of these classifiers is then assumed as contribution to a so-named meta-classifiers (or level-1 classifiers) which creates the last result (Schapire, 1999). Loading has been castoff in BCI investigation by means of HMM by way of level-0 classifiers, in addition an SVM such as meta-classifiers (Nijboer, 2008).

Boost Classifier; Boosting involves in expending numerous classifiers in flow, every one classifier concentrating on the mistakes dedicated through the preceding ones. It can shape up an influential classifier out of numerous feeble ones, and it is improbable to over train. Inappropriately, it is workable to mislabel which might explicate the reason that it was not effective in one BCI education (Boostani, 2004). To date, in the ground of BCI, improving has been investigated through MLP besides Ordinary Least Square (OLS).

Vote Classifier; Despite the fact spending Voting, numerous classifiers are existence castoff, every one of them allocating the involvement feature vector toward a class. The last class will be that of the plurality (Duda, 2012). Elective is the furthest widespread method of unifying classifiers in BCI investigation, almost certainly as it is modest and capable. For the case in point, voting by means of LVQ NN (Qin, 2005), MLP (Ting, 2011) or SVM (Blankertz, 2002) have been tried.

A countless diversity of classifiers has been strained in BCI investigation. Their possessions are summarized in Table 2. It ought to be tense that particular well-known types of classifiers have not been tried in BCI research. The two greatest applicable classifiers are decision tree classifiers (Duda, 2012) and fuzzy classifiers (Lee, 2003). Additionally, dissimilar mixture arrangements of classifiers have been castoff, but numerous other resourceful and well-known ones can be originate in the collected works for example Arcing or Bagging (Bezdek, 1992). Such algorithms might demonstrate valuable as they all prospered in numerous other decoration acknowledgment difficulties. As an example, introductory consequences using a fuzzy classifier for BCI determinations are talented (Lotte, 2006).

4.Discussion

We can accomplish that on ordinary, EEG's signals appear to execute improved result than other signals of physiological. However the consequences of fusion among EEG and outlying are extra robust in equate to brain and outlying signals distinctly. Here after, we presented this discussion to the researchers in order to facilitate choosing the classifier regarding BCI filed. The use of feature extraction and classification algorithms obtained in this paper. It focuses on providing the researchers with guiding principle to help them regarding choosing a compatible classifier to their determined context. The BCI performances based on the classifiers that has been aforementioned and described in this paper are collected in tables within this paper. In the BCI
context, a numerous ways for measuring the efficiency have been suggested, such as Mutual Information (Barreto, 1996), Kappa coefficient (Palaniappan, 2005), classification accuracy, specificity (Kolivand, 2013b) and sensitivity. In the BCI context the classification accuracy considered one of the most common measuring ways, for instance, the ratio of correct feature vectors classification. Accordingly, this mentioned measure has to be mainly considered in this paper. Therefore, couple of dissimilar perspectives has been proposed. In context, the finest classifier determines a BCI category considered the first, while the finest classifier determines a feature category considered the second.

Finest classifiers that applies with BCI, in such context there are various proficient classifiers to be suitable for BCI category which already been used. Precisely, various outcomes have been noticed regarding BCI amongst asynchronous as well as synchronous.

The farthest commonly extend status is synchronous BCI. Here in this perspective we have a proved three classification methods that mainly proficient that are, combinations of classifiers, SVM and dynamic classifiers. But for the time being there is no preferable one amongst them. Hereafter, a justified discussion in term of efficient.

Asynchronous BCI domain; for the time being in context of asynchronous BCI a few experiments has been conducted, thus, surely there is no identified optimal classifier. Within this context, it is denoted that the dynamic classifier does not functioning better than static classifier (Vidaurre, 2006; Qin, 2005). In fact, to identifying each starting state of intellectual task in the experiments of asynchronous considered to be most difficult. As a result the dynamic classifiers may impossible use their related temporal skills in efficient way. The surprising thing is that neither classifiers combinations nor SVM have been used in BCI asynchronous yet.

Synchronous BCI domain; In the context of synchronous BCI, SVM has the top rank, this result has been proved in numerous experiments, that supposed to be in binary (Bishop, 2006; Garcia, 2003; Jayanna, 2009), or nonlinear form (Müller, 2005; Bostanov, 2004), or linear (Bishop, 2006; Palaniappan, 2005), or multiclass BCI, refer to Tables 1. SVM participates numerous properties with RFLDA, for instance, same properties of regularized and linear classifier are assigned to SVM as well as RFLDA. We may consider regularization as one of the various reasons of such a success. In fact, features of BCI are regularly may considered to be possibly containing outliers as well as noisy (Bishop, 2006, Kolivand 2013d). Regularization could vanquish aforementioned obstacle as well as increase the generalization classifier capabilities. Another reason of the aforementioned success could the SVM simplicity. Actually, the SVM resolution rule considered to be a modest linear function that controls the stability of SVM in the kernel space and thence, low variance have been fulfillment. Given that features of BCI are quite unsteady with the passage of time, getting in low variance state possibly will be also an answer key of errors within BCI low classification. The last explanation possibly is the durability of SVM in the sense of dimensionality. The SVM has been enabled by the aforementioned in order to obtain great outcome even among a little training as well as quite high dimensional feature vectors. Though, within BCI context the SVM is not free of disadvantages since they are in general slow comparing with other classifiers. Fortunately, in context BCI real-time they considered to be high speeding enough (Mangai, 2010; Kolivand, 2013a).

Aforementioned paragraph discussed the SVM in the context of the synchronous BCI, hereafter we will focus in our discussion on other type of classifiers. Dynamic classifiers, nearly constantly static outperformed during the experiments (see Table 1). Qin 2005 stated an exemption, however the authors admitted as well as stated that possibly will not be suitable of choosing HMM architecture (Jayanna, 2009).

Within BCI context dynamic classifiers possibly successful for the reason that they be able to catch the related temporal differences appears in the extracted features. Moreover, feature vectors sequence of low dimensional classifying, as a substitute of a quite high dimensional one, expedient, to solve a curse-of-dimensionality. Eventually, in synchronous BCI utilizing dynamic classifiers, as well solves the difficulty of finding the best possible instant in term of classification as the complete sequence of time is classified as well as not just a specific window time (Ishino, 2003).

This possibly will be explained by the Boost sensitivity to mislabels (Müller, 2001) in the meanwhile the truth about these all mislabels are be expected to appear as uncertain and noisy EEG signals data. Thence, such Stack or Vote combinations possibly preferred for the application of BCI.

As we discussed, the classifiers combination helps decreasing the classification error Variance component which mostly makes classifiers combinations extra efficient than their single one.

Moreover, this Variance will mostly reflect the sensitivity to the used training combination. Within experiments of BCI, variability of time could lead to Variance, either variability of object-object or variability of session-session. Thence, probably Variance may consider as an error main source. The
nonstationarity/variability obstacle may consider out of combining classifiers scope to be solved, which may well clarify its success.

We partially conclude that, without considering class numbers, SVM sound to be quite efficient. It is perfect properties explaining such success, these common properties are immunity, regularization and simplicity in context of dimensionality. In addition to SVM, dynamic classifiers sound to be quite promising as much as effective in context of synchronous BCI in the course of classifiers combination. Regarding the experiments of asynchronous, there is no such a classifier may considered as the best than other classifiers, of course with respect to the lack of available results.

No more than this notation that can be concluded because the dynamic classifiers sound losing their excellence within these experiments.

Choosing Classifier in accordance to specific features

In context of BCI features, based on the ability of coping with precise obstacles, we compare classifiers.

Time based; regarding experiments synchronous, the quite effective method regarding exploiting the feature temporal information considered to be dynamic classifier. The time based information could be effectively used by integration classifiers over time (Broek, 2011, Kolivand 2013c). In course of asynchronous experiments, there is no observation for any notability.

Dimensionally high of feature vectors; a quite proper classifier to deal with such cases could be SVM. If using a large amount of segment of time causes dimensionally high of feature vectors, this obstacle can be solved by using dynamic classifiers when considering the feature vectors sequence in place of high dimensionality for single vector. Namely, dynamic classifiers such as HMM (Garcia, 2003), SVM (Müller, 2005; Khalili, 2009) as well as TDNN (Huang, 2001) all of the mentioned considered as a perfect raw EEG data classifiers example. In context of directionality kNN not preferable to be used since it is sensitive to the mentioned context. Subsequently, using features selection and/or reduction of dimensionality methods considered highly recommend;

Small datasets training; LDA as an easy technique through not many parameters within little set of training might be used (Kaper, 2004).

Extremes & noisiness; SVM sounds suitable in order to deal with such cases. Within context of BCI systems, in order to deal with extremes, even Muller et. al. suggested to steadily regularize the classifier. Although, in context of extremes or noisiness it discussed that distinctive classifiers perform quite better than generative classifiers;

Instability; in context of variance reducing, a classifiers combination may probable be used to solve this obstacle. Either SVM or LDA also could be used as stable classifiers.

Table 1 illustrates the time line (stated as columns for date, paper title, method used, aim of using the methods and the accuracy result column) for the classification methods and also shows the accuracy result that has been acheived with the related used classification methods, this helps the researchers to easily choose the best classifier with the best accuracy rate.

Table 1. Common extraction & classification algorithms in term of accuracy result

<table>
<thead>
<tr>
<th>DATE</th>
<th>Paper Title</th>
<th>Method Used</th>
<th>Accuracy Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>EEG-based Dominance Level Recognition for Emotion-enabled Interaction</td>
<td>DFT, Power Spectral Density</td>
<td>from 73.64% to 75.17%</td>
</tr>
<tr>
<td>2010</td>
<td>Emotion Recognition from Brain Signals Using Hybrid Adaptive Filtering and Higher Order Crossings Analysis</td>
<td>HAF based applying Genetic Algorithm (QDA 77.66), (SVM 85.17)</td>
<td>85.17%</td>
</tr>
<tr>
<td>2010</td>
<td>Classification of Human Emotions from EEG Signals using Statistical Features and Neural Network</td>
<td>Classifying mental activity based on emotion</td>
<td>95%</td>
</tr>
<tr>
<td>2009</td>
<td>Emotion Recognition System Using Brain and Peripheral Signals: Using Correlation Dimension to Improve the Results of EEG</td>
<td>MAF, Genetic Algorithm, Correlation Dimension, EEG 63.33%, for peripheral 55%, for both 61.8%</td>
<td>61.8%</td>
</tr>
<tr>
<td>2006</td>
<td>EEG-based Emotion Recognition</td>
<td>using Fourier analysis</td>
<td>&lt; 80%</td>
</tr>
<tr>
<td>2005</td>
<td>Remarks on Emotion Recognition from Multimodal Bio-Potential Signals</td>
<td>Statistical Analysis, NN and SVM</td>
<td>between 62.3% and 59.7%</td>
</tr>
<tr>
<td>2004</td>
<td>Remarks on SVM-Based Emotion Recognition from Multi-modal Bio-Potential Signals</td>
<td>SVM with Statistical Analysis</td>
<td>41.7%</td>
</tr>
<tr>
<td>2003</td>
<td>A Feeling Estimation System Using a Simple Electroencephalograph</td>
<td>FFT, Wavelet Transform, PCA, Statistical Analysis and NN</td>
<td>67.7%</td>
</tr>
<tr>
<td>1997</td>
<td>An Estimation of the Correlation Dimension for the EEG in the Emotional States</td>
<td>SVM</td>
<td>80%</td>
</tr>
</tbody>
</table>
Table 2 illustrates each classifier and its property in term of statistical classification that used to solve and classify extracted features, the rows stated the classifier name meanwhile the each corresponding column of the statistical characteristic that belong to. This helps the researcher to choose easily the corresponding statistical classifier with the related feature.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Linear</th>
<th>Non-Linear</th>
<th>Generative</th>
<th>Discriminant</th>
<th>Dynamic</th>
<th>Static</th>
<th>Regularized</th>
<th>Stable</th>
<th>Unstable</th>
<th>High Dimensional</th>
</tr>
</thead>
<tbody>
<tr>
<td>k-NN</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mahalanobis distance</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMM</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IODM</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayes quadric</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayes graphical network</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLDA</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFLDA</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>linear-SVM</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RBF-SVM</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLP</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILR NN</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALN NN</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TDNN</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIRNN</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDNN</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaussian NN</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LVQ NN</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceptron</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RBF-NN</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PoGNC</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuzzy ARTMAP NN</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This paper tried to collect all comparisons between the widely used classifiers to serve the field of BCI as well as the researchers. Preferably, before choosing a classifier it is better to be examined in the same course, same circumstances and the protocol. At present, it is considered as a critical obstacle in the context of BCI. So, a general purpose of BCI systems has been proposed by other researchers i.e. BCII2000 toolkit (Boostani, 2004) to be more flexible in use. It is based on modulation the framework that facilitates the preprocessing, feature extraction or classification in term of modules. So that, the classifier testing is become more flexible and easy with variant features.

5. Conclusion

This paper has surveyed classification algorithms used to design Brain-Computer Interfaces (BCI). Based on features extracted this paper surveyed the classification algorithms founded on the current available methods in the context of brain-computer interface (BCI) systems. In context of features classifications we surveyed the five categories of the classifications algorithms, i.e. nearest neighbor classifiers, nonlinear Bayesian classifiers, linear classifiers, neural networks and combinations of classifiers. In a BCI context, the previously obtained results subjected to analysis as well as compare regarding providing the readers of this review paper, guidelines of choosing a classifier within a BCI context. In context of synchronous BCI the study showed the SVM proved efficiency. It is possibly caused by the property of regularization. Moreover, dynamic classifiers as well as classifiers combinations sound extremely efficient the experiments in course of synchronous. This paper, highlighted on reviewing and studying widely used classifiers in brain-computer interface (BCI). Nevertheless, this paper focused only on the most common classifiers, so that there are many additional existent classification methods, which are not presently used in the course of BCI. Moreover, BCI nowadays takes place in the virtual environment in specific games world as well as augmented reality, it seems to be promising area of research. So far, many experiments regarding EEG signals in order to choose best classifier that serve this prominent domain still in progress. Subsequently, new state properties should be put in consideration when exploring new classifiers, for instance the availability of EEG signals dataset within variability long term. Unfortunately, lack of published research papers in terms of comparing classifiers considered the main barrier that encountered in such survey study. Optimally, testing as well as comparing between classifiers should be done within the identical course of action.

The key note is an explanation of the common offered algorithms is posted in Table 1. The situation outcome that accomplished is presented in
Table 2 depending on our deep study. In the meanwhile, how each algorithm does work was discussed within the context. This study highlighted the properties, advantages and drawbacks of the algorithms. Finally, it was suggested that it would be beneficial to know which algorithm is best to use in different situations. In Table 2 we stated the common classification algorithms in term of properties. The contribution of this paper is explaining and summarizing the main common feature of classifiers algorithms, and evaluate them in term of accuracy as well as classify them in term of properties. In addition of providing and facilitating of choosing the correct classifier regarding the features extracted.

Finally, through this study noticed that most of researchers were using SVM as well LDA which considered a common classifiers, to have more accurate of analysis result in the context of EEG based on BCI. In the most experimental studies the accuracy is getting higher to be close to a 100% or less a bit. Meanwhile, the researchers ignore the Fast Fourier transforms (FFTs) classifier that also may reach the same percentage accuracy.

Our future analysis will be conduct on the Fast Fourier transforms (FFTs) as an old classifier as well on a new used but explored since decades classifier, it is Higuchi algorithm that used for real-time EEG classification and we will make a deep comparison with each other as well with the already used classifiers.

Acknowledgements:
This research is supported by ViCubelab Department of Computer Graphics, Universiti Teknologi Malaysia (UTM).

Corresponding Author:
Mohd Shahrizal Sunar
MaGIC-X (Media and Games Innovation Centre of Excellence), UTM-IRDA Digital Media Centre Universiti Teknologi Malaysia, 81310 Skudai, Johor MALAYSIA
shahrizal@utm.my

References
ICASSP 2009. IEEE International Conference on (pp. 489-492). IEEE.


Algorithms and Architectures for Parallel Processing (pp. 274-283). Springer Berlin Heidelberg.


http://scikit-learn.org/0.11/auto_examples/neighbors/plot_classification.html/example-neighbors-plot-classification-py, Date access 20 Feb 13


[72] Comparison of linear, nonlinear, and feature selection methods for eeg signal classification


[77] Ekkel, T. Searching Robust Acoustical Indicators for Hypoxia-Related Disorders in Neonates for Classification with Neural Networks.


7/8/2013