



## A Comparative Analysis of Machine Learning Methods for Joint Attention Classification in Autism Spectrum Disorder Using Electroencephalography Brain Computer Interface

Elly Matul Imah<sup>1\*</sup>      Erina Seviyanti Dewi<sup>1</sup>      I. G. P. Asto Buditjahjanto<sup>2</sup>

<sup>1</sup>*Mathematics Department, Universitas Negeri Surabaya, Indonesia*

<sup>2</sup>*Electrical Engineering Department, Universitas Negeri Surabaya, Indonesia*

\* Corresponding author's Email: [ellymatul@unesa.ac.id](mailto:ellymatul@unesa.ac.id)

---

**Abstract:** Electroencephalography (EEG) is a method for recording the electrical activity of the brain. In the EEG various frequency signals can analyse the brain and the brain's behaviour. EEG can detect the abnormalities in the brain, one of the abnormality is Autism Spectrum Disorder (ASD). ASD is a condition where a person has a combination of neurological disorders in social communication and behavior, limited interest in something and sensory behavior. Joint attention (JA) is the ability in sharing attention between interactive social partner with third external elements as objects or events. Joint attention can also be said as social interaction behaviors to follow the attention of others, and direct attention of another. Joint attention is the key point affecting social communication in individuals with autism spectrum disorder (ASD). In this study, the detection of the ability of ASD sufferers to respond to instructions to see a targeted object based on the EEG signal recording was conducted. The dataset used is BCIAUT P-300 that is non-linear separable and imbalanced class with a ratio of 1: 8. In handling the imbalanced data, under-sampling was applied. The feature extraction method that compared are wavelet and principal component analysis. Based on the experiment result, Pz has the best channel to classify the Joint Attention of ASD using EEG P-300 data. The best accuracy is GRLVQ and the best G-mean is SVM. Over all the results, based on accuracy, g-mean, training and testing time show that GRLVQ is better performance than others, and SVM is the runner up. The differences of accuracy 9%, training time GRLVQ faster around 45 seconds than SVM, and testing time faster around 3 seconds.

**Keywords:** Electroencephalography, Autism spectrum disorder, Machine learning, Imbalanced data, Feature extraction.

---

### 1. Introduction

Autism Spectrum Disorder (ASD) is a condition in which a person has a combination of neurological disorders in social communication and behavior, limited interest in something and sensory behavior that can be detected from birth [1]. According to research conducted by Guthrie, a person who has ASD can be detected precisely before 24 months of age. Several factors that cause ASD are genetic defects or abnormal conditions during pregnancy [2]. Apart from these genetic factors, an autoimmune person has the potential to experience neurological disorders such as ASD [3]. Based on a study by

Siniscalco, it is known that the important comorbidity of ASD is the presence of immunological dysregulation, so that autoimmunity plays a role in the pathogenesis of ASD [4]. Environmental factors have an effect on increasing the risk of ASD, a study by Chun stated that there is a relationship between mothers who are exposed to air pollution and children with ASD [5].

Several countries in the world have confirmed an increase in the number of ASD cases, especially in men [1]. Based on a study by The NHS Information Center for Health and Social Care London in 2012, the prevalence of ASD cases occurring in adults in the United Kingdom reached 1.1% [6]. Meanwhile, according to a study by Saab, there was 1 ASD case

per 66 toddlers in Lebanon in 2018 [7]. A study on the prevalence of ASD cases was also carried out by Salhia in countries in the Arabian Gulf with a prevalence of ASD cases of 1.4 or 29 ASD cases per 10,000 people [8]. A high number of ASD cases also occurs in the United States, with a prevalence rate of 1 ASD case per 59 persons or one third of the total ASD cases in the world [11].

Due to the increasing number of ASD cases in various countries, it is necessary to understand the impact of having children with ASD on family social life because some of the characteristics of ASD patients, according to the American Psychiatric Association (APA), are the characteristics of ASD, namely limitations or disturbances in communication and social interaction [9]. One form of social communication is joint attention, which is important in social cognition in humans, such as sharing attention and following the attention of others. A study conducted by Eissa stated that someone with ASD has a deficiency in the ability of Joint Attention to the environment [10]. Jeniffer in his research in 2019 published that the ability of Joint Attention in an ASD can be improved through training and learning such as giving instructions and responding to things [11]. Therefore it is necessary to detect the response given by an ASD to an instruction.

Several previous studies have been conducted to classify brain responses based on EEG signals. The study was conducted by Mangia in testing the feasibility of the procedure in detecting the brain's response to EEG-based image instructions in patients with impaired consciousness [12]. Classification is done with a Support Vector Machine (SVM) to distinguish two mental tasks and detect answers to simple questions given so that an accuracy of above 0.8 is obtained. A study in predicting the response to EEG-based decision making was also carried out by Yajing Si in 2020 [13]. The combination of Discriminative Spatial Network Pattern (DSNP) and Linear Discriminate Analysis (LDA) is used to predict the responses of two different groups of individuals. The results of the accuracy in the response detection process were  $0.88 \pm 0.09$  for the first group, and  $0.90 \pm 0.10$  for the second group.

Based on the research that has been done in detecting responses, the use of Electroencephalography (EEG) based on the Brain-Computer Interface (BCI) has not been used in classifying the instructional response to an ASD. In addition, there is no classification using Generalized Relevance Learning Vector Quantization (GRLVQ). In this study, GRLVQ was used to detect the response of an ASD to instructions using the EEG signal based on BCI. Variations in the parameter values of the

GRLVQ are carried out to determine the best accuracy in the response detection process to an instruction.

The rest of this paper is organized as follow, second section describes the joint attention in ASD problem and related to EEG Brain Computer Interface. The data pre-processing and feature extraction explained in third section. Section fourth describe the classification methods that used. The experiment set up, evaluation measurement, experiment results, and discussion in fifth section. The last section is conclusion of this studies.

## 2. Joint attention in ASD and EEG brain computer interface

Based on the studies investigations from University of San Francisco and Boston University, EEG could be used as early biomarker for diagnosis of ASD [14]. The study by Carlos Amaral et al about social attention in ASD using EEG brain computer interface report that EEG brain computer interface can be used to train social cognition skill in people with ASD [15]. This section will be explained about Joint Attention in ASD and EEG brain computer interfaced.

### 2.1 Joint attention in ASD

Joint attention (JA) is the ability in sharing attention between interactive social partner with third external elements as objects or events. Joint attention can also be said as social interaction behaviors to follow the attention of others, and direct attention of another. People with autism spectrum disorder demonstrate difficulty with joint attention, affecting social and occupational performance [11]. Joint attention is the key point affecting social communication in individuals with autism spectrum disorder (ASD). There are two types of JA, initiated joint attention (IJA) and responded joint attention (RJA). IJA is intentional attempt to direct the partner attention to the external elements. How to be expressed through showing, pointing, or otherwise directing a social partner gaze to the objects or events. Responded joint attention is how the individual respond on attentional bids toward an external element. RJA is taking action implied by IJA or can be said as acknowledging the IJA or expresses through gaze-following towards the objects or events. JA is important skill in helping people communicate with each other in social life. Individual with autism have difficulty with JA or hard problem in ability to share focus on an object or area with the other persons [16].

The ability of JA in some individuals has developed in 6 months of age, when they look at what their mother head turn to a visible object. The IJA begin to show a desire for something at the age of 12 months. In children with autism, the ability of JA as a social communication disorder can be detected at the age of 1 year. Children with autism or mental retardation, however, exhibit less joint attention and other gestational behaviors than children who are developing normally. Joint attention has been found to be a good predictor of both concurrent and future language skills in children with autism. Joint attention skills were concurrently associated with language ability for both groups and predicted long-term gains in expressive language ability for the children with autism. In the present study, the concurrent associations between each of the three social attention skills—joint attention, social orienting, and attention to distress—and verbal language ability in young children with ASD. Joint attention would be the strongest concurrent predictor of language ability. If the child fails to pay attention to such cues, he or she will be missing important opportunities for acquiring and practicing joint attention skills.

Several studies have been conducted on whether social orientation disorder increases the ability to distinguish preschool-aged children with autism from those with developmental delay (DD) and developmental typical. The research results suggest that disruption is combined in JA and social orientation. Distinguishing children with autism from children of the same mental age without autism is better than does joint attention impairment alone. Furthermore, while JA was the best predictor of common language, both social orientation and attention to distress indirectly contributed to language skills through their relationship with JA. These results help to clarify the nature of attention disorders in autism, offer clues to possible development, and suggest specific targets for early intervention. This facilitates the development of mutual attention skills in autistic children by increasing their motivation to carry out social stimulation and triadic communicative exchanges. Such interventions are essential to prepare plans for future communicative skills development [17, 18].

## 2.2 EEG brain computer interface

A Brain-Computer Interface (BCI) is a system that interpret brain activity and translate it into commands that can be used to control devices or programs and provides a direct communication between the brain and a computer or external device.

The most common modality of neuroimaging techniques can be used to implement BCI is the electroencephalography (EEG), since it provides a portable, inexpensive, non-invasive solution to measure brain activity with high temporal resolution. Generating brain signal that can be interpreted and transform into commands by BCI. There are several approaches to generate brain signals, one of them is P300 approach. P300 first attempted by Farwell and Donchin in the 80s that used oddball paradigm of frequent and infrequent stimulus on interest [19]. The paradigm has been used for a positive deflection of the EEG measured in the central and posterior parts of the scalp is observed approximately around 300 ms after the infrequent stimulus of interest is presented. The BCI P300 potential is elicited whenever the letter the user is paying attention to flashes, and so the target letter can be identified by a P300 detection algorithm and then transmitted. This research using EEG joint attention data from ASD using BCI-P300 from the Amaral et all research [15].

Electroencephalography (EEG) is a method for recording the electrical activity of the brain. In the EEG various frequency signals can analyse the brain and the brain's behaviour [19]. Based on research by Abdolzaldegan, EEG can detect abnormalities in the brain, one of which is autism[20]. In recent years there have been several techniques of neuroimaging such as EEG being used to implement the Brain Computer Interface (BCI). BCI is a tool capable of measuring the activity of the central nervous system and changing it in the form of artificial output which can replace, increase, or increase its natural output [21]. BCI records signals in the brain then analyses them according to device commands [22]. In everyday life, BCI can be applied in various fields such as medical, neuro-ergonomics and smart environment, neuromarketing and advertising, education, games, and entertainment, as well as security and authentication fields [23]. In the medical field, EEG is often used in the diagnosis of several brain disorders such as epilepsy [24, 25].

## 3. BCIAUT-P300 data pre-processing and feature extraction

BCIAUT-P300 benchmark dataset on Autism Joint Attention using P300-Based Brain-Computer-Interfaces used in this study [19]. Before classified, the data has been pre-processed and extract the feature. Feature extraction is an important part to make the classification process work well. Good feature will lead the process to the better result as expected, but if the feature is not good enough, it will yield to negative result. There are many ways to do a

feature extraction processes. In this there are two feature extraction methods is compared. Two feature extraction methods are principal component analysis (PCA), and wavelet. Z-score data normalization has been applied before feature extraction. The detail about the methods as follow.

### 3.1 Data pre processing

The BCIAUT-P300 dataset on Autism Joint Attention is imbalanced data and has been used in Biomedical Mediterranean Conference on Medical and Biological Engineering and Computing (MEDICON) competition with some approaches among others deep learning in 2019 [19]. The dataset consist of EEG signals were obtained from 15 ASD (Autism Spectrum Disorder) subjects ranging in age from 16 to 38 years and male sex. The subjects used in this dataset have certain criteria, namely a positive ASD diagnosis result set by the Autism Diagnostic Interview-Revised, Autism Diagnostic Observation Schedule, and the Diagnostic and Statistical Manual of Mental Disorders, Five-Edition (DSM-5). There are 8 channels, namely C3, Cz, C4, CPz, P3, Pz, P4, and POz in the experimental process with 350 features. The dataset consists of train data and test data for each subject with 7 sessions. In this study, the EEG data used was in session 1 by taking signals from the Pz and Cz channel, based on the studied by M. Xu et all about EEG channel selection Pz and Cz has the best channel in EEG signal processing [26]. The dataset consists of two classes, namely False (0) and True (1). In this dataset, the subjects were instructed to pay attention to 8 objects available in the room. The object was irradiated alternately on 20

blocks with 10 repetitions, resulting in 1600 experiments. Furthermore, the subject is instructed to guess the object being illuminated, if the object identified is the same as the object illuminated, it is labelled as the targeted class (1). If the recognized object is different from the object being illuminated, it is labelled as non-targeted class (0). The data was normalize using z-score. In simple terms z-score can be written in Eq 1. After the data is normalized by z-score, the mean data value is 0. The sample result after z-score normalization can be seen on Fig. 1. The pattern of signal does not change but the signal range has been changes.

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \quad (1)$$

### 3.2 Principal component analysis (PCA) feature extraction

PCA is a method that can reduce the dimensions of the input data into smaller dimensions by considering the variance of the data. The principle of PCA is to find a linear transformation that can map the input data coordinate frame into a new coordinate frame orthonormal [27]. The linear transformation from original data to new data is called Principal Components (PCs). PCs are sequenced in descending order with the first few PCs values representing the variance of the data so that taking several PCs can represent important information on the data. The feature extraction algorithm using PCA as follow:

Step 1: Input data to be extracted features.

Step 2: Normalize the input data.

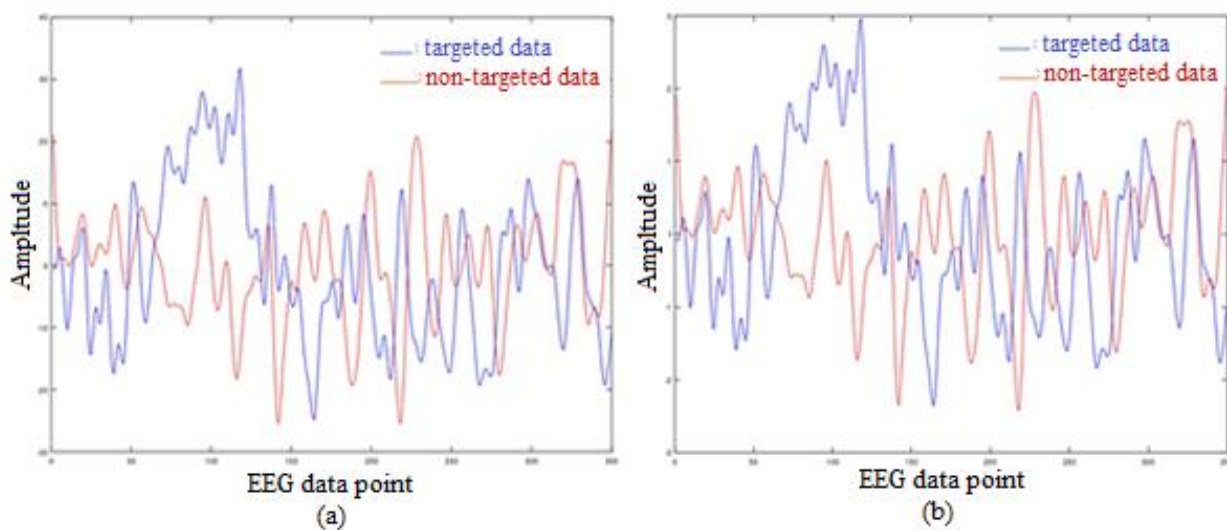


Figure 1. EEG Joint Attention: (a) data before Z-score Normalization and (b) after Z-score Normalization

Step 3: Calculate the covariance matrix

$$Cov(xy) = \frac{\sum xy}{n} - (\bar{x})(\bar{y}) \quad (2)$$

Stage 4: Calculating eigenvalues and eigenvectors

$$(A - \lambda I) = 0 \quad (3)$$

Step 5: Sorting eigenvalues descending and determining Principal Component

$$[A - \lambda I][X] = [0] \quad (4)$$

### 3.3 Wavelet feature extraction

In this study we use discrete wavelet transformation to extract features to compared with PCA. Wavelet is a method of removing noise in digital signals, compressing data, and detecting sudden discontinuities [28]. Wavelets describe the signal into various resolutions based on a function called the wavelet function. The discrete wavelet function that used is Daubechies 8 (db8). Daubechies wavelets are wavelets that can provide estimates of wavelet expansion. The Daubechies has the effectiveness of various wavelet functions for remove noise signal noise in normal patients compared to db4, db2, and Mayer [29].

The Wavelet Transform (WT) of a signal  $f(x)$  is defined as:

$$W_s f(x) = f(x)\psi_s(x) = \frac{1}{s} \int_{-\infty}^{+\infty} f(t)\psi\left(\frac{x-t}{s}\right) dt \quad (5)$$

Where  $s$  is scale factor,  $\Psi_s(x) = \frac{1}{s}\Psi\left(\frac{x}{s}\right)$  is the dilation of a basic wavelet  $\Psi(x)$  by the scale factor  $s$ . Let  $s = 2^j$  ( $j \in \mathbb{Z}$ ,  $\mathbb{Z}$  is the integral set), then the WT is called dyadic WT [30]. The dyadic WT of a digital signal  $f(n)$  can be calculated with Mallat algorithm as follows:

$$S_{2^j} f(n) = \sum_{k \in \mathbb{Z}} h_k S_{2^{j-1}} f(n - 2^{j-1}k) \quad (6)$$

$$W_{2^j} f(n) = \sum_{k \in \mathbb{Z}} g_k S_{2^{j-1}} f(n - 2^{j-1}k) \quad (7)$$

where  $S_{2^j}$  is smoothing operator,  $S_{2^j} f(n) = a_j$ ,  $a_j$  is the low frequency coefficients that approximate the original signal, while  $W_{2^j} f(n) = d_j$ ,  $d_j$  is high frequency coefficients represent detail of the original signals.

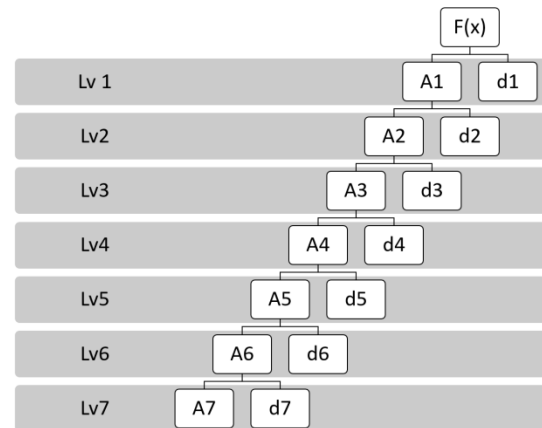


Figure. 2 Illustration of feature extraction using wavelet decomposition

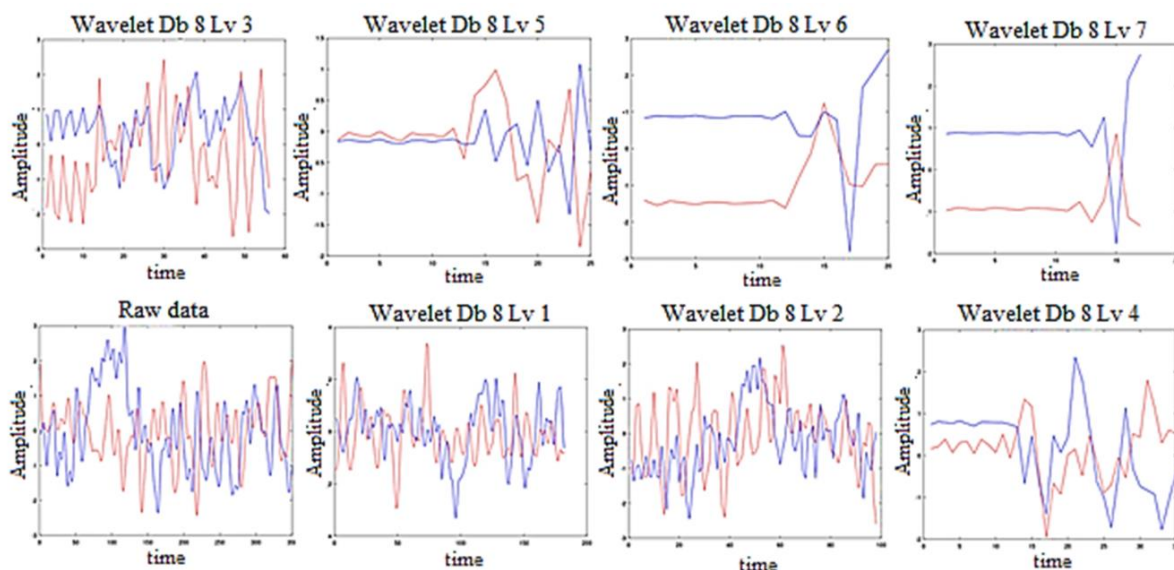


Figure. 3 EEG Signal in differences level of Wavelet Daubechies 8 decomposition

Selecting the appropriate mother wavelet and the number of decomposition level is an important part. The proper selection aims to retain the important part of information and still remain in the wavelet coefficients. The Mother wavelet that we used in this research is one member of the Daubechies families: Daubechies order 8. Throughout this research, EEG signal is decomposed from level 1 to 7. Using daubechies 8 (*db8*), every level generated 183, 98, 57, 35, 22, 20, and 17 coefficients respectively for both approximation (a) and detail (d). Fig. 2 shows the illustration of features extraction using wavelet decomposition level 1 to 7, which will be used in classification process. The EEG signal decomposition result can be seen on Fig. 3.

#### 4. Classification of EEG joint attention in ASD

In this study, the machine learning classification algorithms that is used for study comparison are Generalized Learning Vector Quantization (GRLVQ), Support Vector machine (SVM), Random Forest, and Backpropagation. We are not used deep learning because in EEG data deep learning does not appropriate and high computational cost[31], [32]. The details of each algorithm are as follows:

##### 4.1 Generalized relevance learning vector quantization (GRLVQ)

The GRLVQ algorithm is a learning algorithm modified by the Generalized Learning Vector Quantization (GLVQ) algorithm and Relevance Learning Vector Quantization (RLVQ). This algorithm reduces the dimensions of the input data but still retains the important information contained in the data [33]. The advantage of the GRLVQ algorithm is that it is efficient, this is due to the formation of adaptive metrics through the process of scaling the input dimensions. In addition, this algorithm can classify data well. The GRLVQ algorithm is as follows:

- Step 1: Stage 1: Input training data along with Weight ( $w$ ), Iteration ( $n$ ), Learning Rate ( $\alpha$ ), and Factor Relevance ( $\lambda$ ) with  $\sum \lambda = 1$
- Step 2: Calculating the distance between training data and weights using Euclidean distance
- Step 3: Determine and where is the shortest distance between the training data and the weight of the wrong class and the correct class.
- Step 4: Calculating the value of the misclassification error ( $\mu$ ) and the value of the activation function ( $f$ )

$$d = \sum_{i=1}^N \lambda_i (x_i - w_i)^2 X(t+1) = X_p(t) - A.D \quad (8)$$

$$\mu = \frac{d_j - d_k}{d_j + d_k} \quad (9)$$

$$f = \frac{1}{(1 + e^{(-\mu t)})} \quad (10)$$

Step 5: Updating the Relevance Factor with the formula

$$\lambda_m := \lambda_m - \alpha \frac{df}{d\mu} \cdot b \quad (11)$$

Where :

$$b = \frac{d_k}{(d_j + d_k)^2} (x_m^I - w_m^j)^2 - \frac{d_j}{(d_j + d_k)^2} (x_m^i - w_m^k)^2$$

Step 6: Updating the Initial Weight with the formula

$$\begin{cases} w_j + \alpha \frac{df}{d\mu} \frac{d_k}{(d_j + d_k)^2} \lambda (x - w_j); & \text{1st class weighted} \\ w_k - \alpha \frac{df}{d\mu} \frac{d_j}{(d_j + d_k)^2} \lambda (x - w_k); & \text{2nd class weighted} \end{cases} \quad (12)$$

Step 7: Reduce the learning rate to  $n$  iterations, and repeat Stages 2 to 7 as many as  $n$  iterations

##### 4.2 Support vector machine (SVM)

This study used non-linear SVM. Non-linear SVM classifies data between classes by finding the best hyperplane that is built from set of data in separator between classes. The set data are points of each class with the closest distance from the hyperplane is called support vector [34]. The best hyperplane determined by calculating the maximum margin. Margin is a value that represents the distance of a class to another class. The application of the SVM algorithm in this study uses 2 parameters, namely C, and kernel. The parameter C in SVM states the penalty limit for Lagrange multiplier or also called the threshold. The kernels used to compare with the other machine learning methods is poly-kernel, because this kernel show better accuracy and g-mean than others. The kernel function that used in the experiment are linear kernel, poly-kernel and RBF kernel which are formulated as follows:



Linier Kernel:

$$K(x, y) = (x^T y) \quad (13)$$

Poly-kernel:

$$K(x, y) = (x^T y + c)^d \quad (14)$$

RBF kernel:

$$K(x, y) = \exp\left\{\frac{\|x - y\|^2}{2\sigma^2}\right\} \quad (15)$$

where  $x, y$  is the input data,  $c$  is constant,  $d$  is the degree of polynomial, and  $\sigma$  is standard deviation. If there is a dataset with two classes e.g.  $y_i \in \{-1, +1\}$ ,  $x_i \in R^n, n > 1, i = 1, 2, 3, \dots, I$  and the *hyperplane*  $g(x) = \langle w, x \rangle + c$  so the *decision rule* defined by :

$$f(x) = \begin{cases} +1, & g(x) \geq 1 \\ -1, & g(x) < 1 \end{cases} \quad (16)$$

Furthermore, to find the best hyperplane it is necessary to minimize the following values:

$$\frac{1}{2} \|w\|^2 = \frac{1}{2} (w_1^2 + w_2^2) \quad (17)$$

with the following conditions:

$$(w_1 \cdot x_i + w_2 \cdot x_i + c) \geq 1 \quad (18)$$

After obtaining the hyperplane optimal, the output of the sign decision function  $f(x)$  is calculated for determine the data class.

$$f(x_d) = \sum_{i=1}^m \alpha_i y_i K(x_i, x) + y \quad (19)$$

Where  $m$  is a lot of support vectors,  $\alpha_i$  is the weight of each data, and  $K(x_i, x)$  is kernel function.

### 4.3 Random forest

Random forest is classification algorithm that is basically based on random tree. Every input feature vector is compared to the one stored in the train dataset to find the best matching pattern. Growing an ensemble of random trees for recognition using a probabilistic scheme is called random forest of trees. Recognition accuracy is as high as the trees vote for the most popular class. Trees drawn at random from a set of possible trees is called random tree. Random

tree is a decision tree that considers  $k$  randomly chosen attributes at each node. The class probabilities on each node are based on back fitting with no pruning [35]. The steps involved in growing a random tree are as follow:

1. The training set for growing the tree is obtained by selecting  $N$  cases at random but with replacement from original dataset.
2. A random number of attributes  $m$  are chosen for each tree. The attributes from the nodes and leaves using standard tree building algorithms. The best split on  $m$  is used to split the nodes and  $m$  is held constant.
3. Each tree is growing to the fullest extent possible without pruning.

A new object is classified using its input vector down each of the trees in the forest. The forest chooses the class with the most vote, the new object input vector is classified.

### 4.4 Backpropagation

Backpropagation is one of Artificial Neural Network (ANN) algorithms that develop from Multi Layers Perceptron (MLP) by adding delta rule as backward phases. The basic idea of Backpropagation is to efficiently compute partial derivatives of an approximating function realized by the network and toward the entire processing element (neuron) [36]. It is an adjustable weight vector for a given value of input vector. Using nonlinear activation function can be easy to classify non-linearly separable data. Backpropagation and many ANN algorithms have special advantaged for solving multiclass classification task. Architecture Network of Backpropagation in this study has 3 layers, input, hidden, and output layers. Output layers consist of 5 neurons.

## 5. Results and discussion

This study classifies BCIAUT-P300 EEG Joint Attention of ASD using PCA and wavelet feature extraction, and some classification algorithm that is Generalized Relevance Learning Vector Quantization (GRLVQ) algorithm, Random Forest (RF), Multilayer Perceptron (MLP), and Support Vector Machine (SVM). The scatter plot of data can be seen on Fig. 4. The data is non-linear separable. It is hard classification problems.

The classification results are used to analysis the machine learning algorithm performances. Every algorithm run many times to get the best result. The GRLVQ algorithm run with the variation of learning rate and the number of iterations. Meanwhile, RF

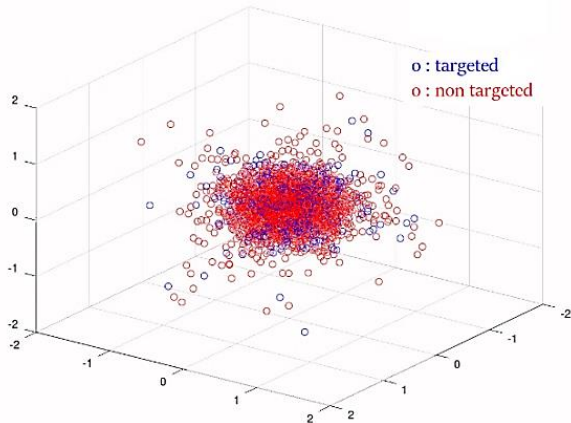


Figure. 4 Scatter plot of BCIAUT P-300 EEG data

uses a variety of feature parameters (NumFeatures). MLP uses variations in the number of hidden layers, and SVM uses variations in the value of C and kernel function. The experimental results are then evaluated using accuracy, geometrical mean accuracy, and F-Measure.

The test results yield the number of TP (True Positive), FP (False Positive), FN (False Negative), and TN (True Negative) which are used to evaluate system performance. Accuracy is the amount of data classified according to the correct class compared to the total data tested. The accuracy value is calculated using Eq. (20). Geometrical Mean Accuracy (G-mean) is a method for measuring the success of classification on data where the amount of data between classes is not balanced. GMA is measured based on the number of TP and TN classification data. The G-mean formula is in Eq. (21). F-Measure is an evaluation technique that calculates the average of

recall and precision values. The F-Measure formula is in Eq. (24).

Table 1. Full feature data classification result without under-sampling

Methods	Accuracy	G-mean	Confusion matrix	
<b>Random Forest</b>	0.8746	0.0333	6296	4
<b>Back-propagation</b>	0.875	0	899	1
<b>SVM</b>	0.875	0	6300	0
<b>GRLVQ</b>	0.6789	0.4293	900	0
			6300	0
			900	0
			4664	1636
			676	224

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{20}$$

$$G - mean = \sqrt{\frac{TP \times TN}{(TP + FN)(TN + FP)}} \tag{21}$$

$$Fmeasure = 2 \times \frac{Recall \times Precision}{Recall + Precision} \tag{22}$$

The BCIAUT P-300 EEG Joint attention is imbalanced data. The classification result in imbalanced data without under sampling can be seen on Table 1. The accuracy in Table 1 is high more than 0.87 in Random Forest, Backpropagation, and SVM but extremely low G-mean. The lowest g-mean indicate that minor class does not recognized.

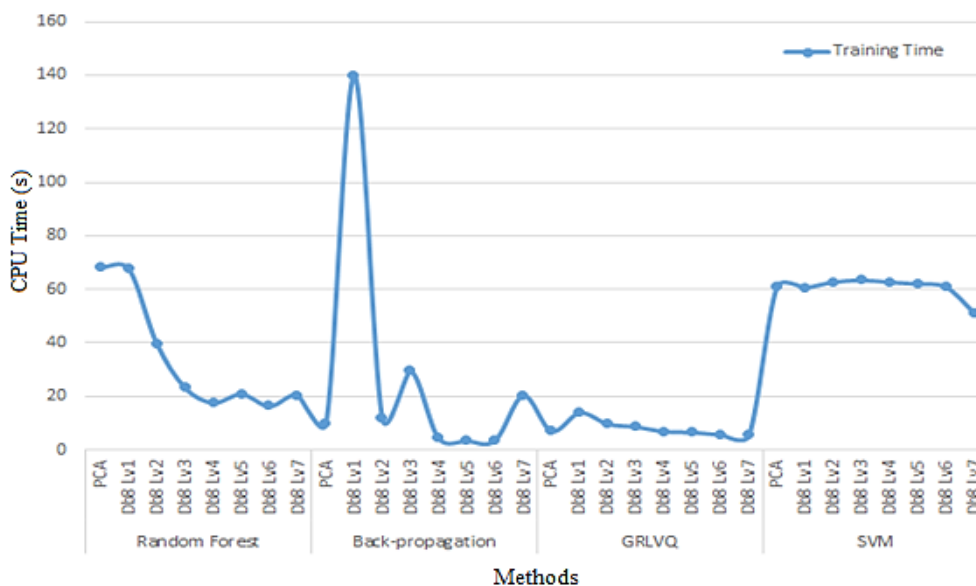


Figure 5. CPU time to build the model



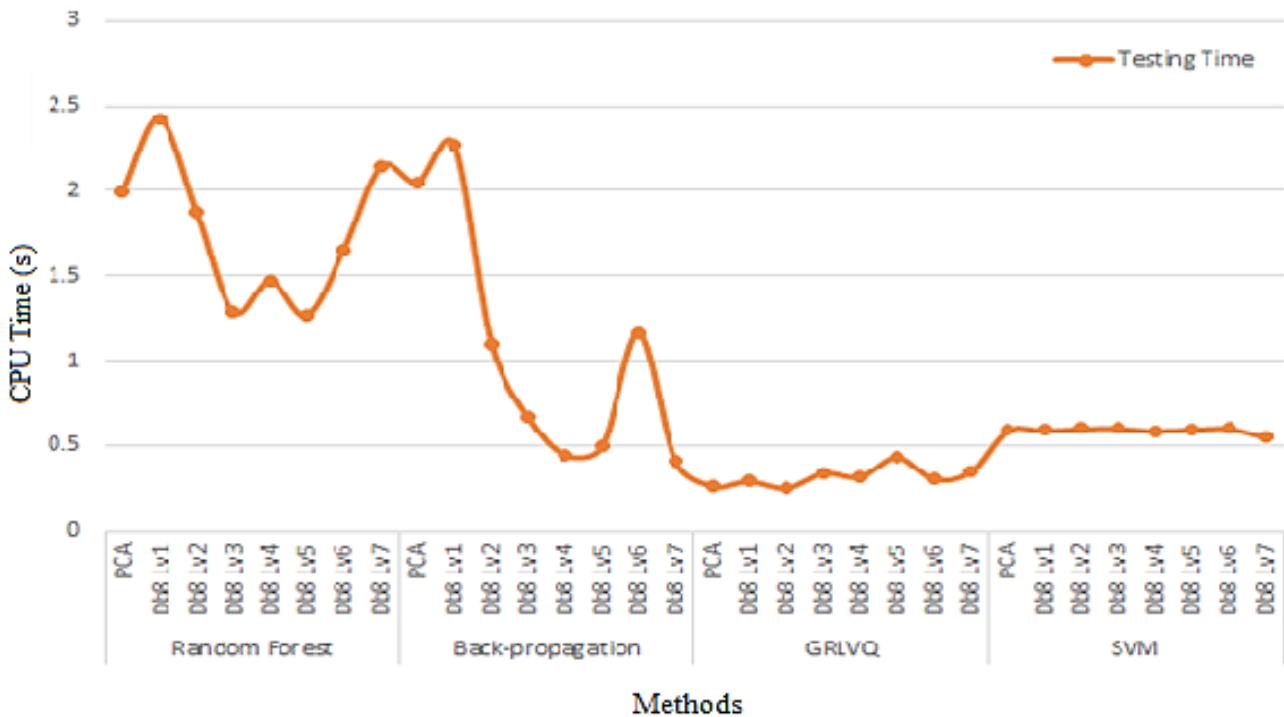


Figure 6. CPU time for testing process

GRLVQ show the highest G-mean results than Random Forest, Backpropagation, and SVM

In handling the imbalanced class the raw data was under-sampling. The major class (non-targeted class) has been reduced near to the minor class (targeted class). The data also have been extracting the feature using PCA and Wavelet Db 8. The classification result can be seen on Table 2. Based on the accuracy and G-mean the wavelet feature extraction shows better performance than PCA. The best performance is GRLVQ using wavelet Db 8 level 6 for feature extraction with the accuracy value is 0.7117 and G-mean is 0.529. This accuracy better than the some results of MEDICON competition [37] [19].

Fig. 5 and 6 show the CPU time for testing and build the model. The analysis of machine learning performance not only evaluate using accuracy, g-mean and so on, but also evaluate how long the model was build and how long the model able to classify the unseen data. We can see that the fastest algorithm to build the model and to classification the unseen data is GRLVQ. The CPU time is robust does not depend on the number of features. GRLVQ is simple algorithm that find the best centre or prototype models in every class. Difference with the others machine learning methods.

The overall results with deference channel of EEG Joint Attention data can be seen on Table 3. We

used Channel Pz and Cz based on [26]. The result show that Pz has the best channel to classify the Joint Attention of ASD using EEG P-300 data. The best accuracy is GRLVQ and the best G-mean is SVM. From all the result we can said that GRLVQ show the better performance, and SVM is the other ones.

GRLVQ is very simple but show the best accuracy because this algorithm able to selecting the best feature using relevance factor coefficient.

Table 3. The classification result in difference channel EEG

Channel	Methods	F-measure	G-mean	Accuracy
<b>Pz</b>	GRLVQ	0.5056	0.4311	<b>0.726</b>
	RF	0.4978	0.5907	0.605
	SVM	0.5177	<b>0.6042</b>	0.6354
	BP	<b>0.5509</b>	0.5795	0.7136
	<b>avg</b>	<b>0.518</b>	<b>0.5514</b>	<b>0.67</b>
<b>Cz</b>	RF	0.4355	<b>0.5193</b>	0.5238
	SVM	0.5052	0.4888	0.6871
	GRLVQ	0.5155	0.4496	<b>0.7329</b>
	BP	<b>0.5173</b>	0.4894	0.7101
	<b>avg</b>	<b>0.4934</b>	<b>0.4868</b>	<b>0.6635</b>

Table 2. Evaluation Results for the Classification of All Algorithms

Methods	Feature	Accura-cy	G-mean	Confusion matrix		Methods	Feature	Accura-cy	G-mean	Confusion matrix		
Random Forest	PCA	0.4293	0.5128	2493	3807	SVM	PCA	0.6104	0.5896	3888	2412	
				302	598						393	507
	Wavelet Db 8 Lv 1	0.5597	0.5683	3508	2792		Wavelet Db 8 Lv 1	0.6082	0.5914	3866	2434	
				378	522						387	513
	Wavelet Db 8 Lv 2	0.5683	0.5798	3556	2744		Wavelet Db 8 Lv 2	0.6282	0.6018	4011	2289	
				364	536						388	512
	Wavelet Db 8 Lv 3	0.5714	0.5674	3608	2692		Wavelet Db 8 Lv 3	0.6354	0.6042	4066	2234	
				394	506						391	509
Wavelet Db 8 Lv 4	0.605	0.5907	3841	2459	Wavelet Db 8 Lv 4	0.6279	0.5892	4033	2267			
			385	515					412	488		
Wavelet Db 8 Lv 5	0.5965	0.5739	3804	2496	Wavelet Db 8 Lv 5	0.6215	0.5967	3966	2334			
			409	491					391	509		
Wavelet Db 8 Lv 6	0.5846	0.5753	3702	2598	Wavelet Db 8 Lv 6	0.6096	0.6025	3855	2445			
			393	507					366	534		
Wavelet Db 8 Lv 7	0.5135	0.5274	3205	3095	Wavelet Db 8 Lv 7	0.5136	0.5484	3158	3142			
			408	492					360	540		
Back-propagation	PCA	0.6756	0.5547	4474	1826	GRLVQ	PCA	0.875	0	6300	0	
				510	390						900	0
	Wavelet Db 8 Lv 1	0.5894	0.5482	3795	2505		Wavelet Db 8 Lv 1	0.5454	0.5445	3438	2862	
				451	449						411	489
	Wavelet Db 8 Lv 2	0.6839	0.5819	4497	1803		Wavelet Db 8 Lv 2	0.5399	0.5375	3406	2894	
				473	427						419	481
	Wavelet Db 8 Lv 3	0.5668	0.5289	3646	2654		Wavelet Db 8 Lv 3	0.5731	0.551	3655	2645	
				465	435						429	471
Wavelet Db 8 Lv 4	0.6751	0.5729	4442	1858	Wavelet Db 8 Lv 4	0.5821	0.5725	3687	2613			
			481	419					396	504		
Wavelet Db 8 Lv 5	0.7056	0.5599	4702	1598	Wavelet Db 8 Lv 5	0.5465	0.5423	3452	2848			
			522	378					417	483		
Wavelet Db 8 Lv 6	0.7022	0.5942	4623	1677	Wavelet Db 8 Lv 6	0.7117	0.529	4793	1507			
			467	433					569	331		
Wavelet Db 8 Lv 7	0.5696	0.5511	3626	2674	Wavelet Db 8 Lv 7	0.5232	0.5121	3319	2981			
			425	475					452	448		

## 6. Conclusion

Based on the experimental results in the Social Joint Attention of ASD using BCI P-300 data, the best channel that representation to classify the EEG data is Channel Pz. The fastest algorithm to build the model is GRLVQ because this algorithm very simple

and able to select the best feature using relevance factor coefficient. The dataset is non-linear separability and imbalanced. If we used all training data to build the models, the minor class is not able to classify correctly. The accuracy is good enough more than 0.87 but this accuracy only good for major class, the minor class accuracy is zero. In handling

imbalanced class problem, under sampling has been implemented in this study. The result show that g-mean is increasing, but the accuracy is reduced. This condition is much better because the minor class can be classified is more importance. The best feature extraction method is Wavelet than PCA. And overall the combination of wavelet and GRLVQ show better performance than the other, and the runner up is combination of wavelet feature extraction and SVM. This studies still challenging on handling imbalanced data for the best g-mean result, and low computational cost and memory.

### Conflicts of Interest

In accordance with IJIES policy and our ethical obligation as researchers, we are reporting that We do not have a financial and/or business interests related to this topic, and do not receive funding from a company that may be affected by the research reported in the enclosed paper. I have disclosed those interests fully to IJIES, and have in place an approved plan for managing any potential conflicts arising from this arrangement. We have no conflicts of interest to disclose.

### Author Contributions

The authors contribution as follow: conceptualization, methodology, EMI; software, EMI and ESD; validation, EMI and ESD; formal analysis, EMI; investigation, EMI; data curation, EMI and ESD; writing—original draft preparation, EMI; proofreading IGPAB.

### Acknowledgments

This work is supported by Grant of National Competitive on Applied Research with agreement No. B//11606/UN38.9/LK.04.00/2020 funding by the Ministry of Research, Technology, and Higher Education.

### References

- [1] J. Baio, L. Wiggins, D. L. Christensen, M. J Maenner, J. Daniels, Z. Warren, M. Kurzius-Spencer, W. Zahorodny, C. R. Rosenberg, T. White, M. S. Durkin, P Imm, L. Nikolaou, M. Yeargin-Allsopp; L. C. Lee, R. Harrington, M. Lopez R. T. Fitzgerald, A. Hewitt, S. Pettygrove, J. N. Constantino, A. Vehorn, J. Shenouda, J. Hall-Lande; K. V. N. Braun, and N. F. Dowling, “Prevalence of autism spectrum disorder among children aged 8 Years - Autism and developmental disabilities monitoring network, 11 Sites, United States, 2014”, *MMWR Surveill. Summ.*, Vol. 67, No. 6, 2018, doi: 10.15585/mmwr.ss6706a1.
- [2] H. V. Ratajczak, “Theoretical aspects of autism: Causes-A review”, *J. Immunotoxicol.*, Vol. 8, No. 1, pp. 68–79, 2011, doi: 10.3109/1547691X.2010.545086.
- [3] T. Zou, J. Liu, X. Zhang, H. Tang, Y. Song, and X. Kong, “Autoantibody and autism spectrum disorder: A systematic review”, *Res. Autism Spectr. Disord.*, Vol. 75, No. May, p. 101568, 2020, doi: 10.1016/j.rasd.2020.101568.
- [4] D. Siniscalco, S. Schultz, A. L. Brigida, and N. Antonucci, “Inflammation and neuro-immune dysregulations in autism spectrum disorders”, *Pharmaceuticals*, Vol. 11, No. 2, pp. 1–14, 2018, doi: 10.3390/ph11020056.
- [5] H. K. Chun, C. Leung, S. W. Wen, J. McDonald, and H. H. Shin, “Maternal exposure to air pollution and risk of autism in children: A systematic review and meta-analysis”, *Environ. Pollut.*, Vol. 256, p. 113307, 2020, doi: 10.1016/j.envpol.2019.113307.
- [6] T. Brugha, S. A. Cooper, S. McManus, S. Purdon, J. Smith, F. J. Scott, N. Spiers, and F. Tyrer, “Estimating the prevalence of Autism Spectrum Conditions in adults: extending the 2007 adult psychiatric morbidity survey”, *NHS Inf. Cent. Heal. Soc. Care*, pp. 1–31, 2012.
- [7] D. Saab, M. Chaaya, and R. M. Boustany, “National Prevalence and Correlates of Autism: A Lebanese Cross-Sectional Study”, *Autism. Open. Access*, Vol. 08, No. 01, pp. 1–7, 2018, doi: 10.4172/2165-7890.1000223.
- [8] H. O. Salhia, L. A. Al-Nasser, L. S. Taher, A. M. Al-Khathaami, and A. A. El-Metwally, “Systemic review of the epidemiology of autism in Arab gulf countries”, *Neurosciences*, Vol. 19, No. 4, pp. 291–296, 2014.
- [9] A. Pereira, A. Peterman, A. N. Neijhoft, R. Buluma, R. A. Daban, A Islam, E. T. V. Kainja, I. F. Kaloga, T. Kheam, A. K. Johnson, M. C. Maternowska, A. Potts, C. Rottanak, C. Samnang, M. Shawa, M. Yoshikawa, and T. Palermo, “Disclosure, reporting and help seeking among child survivors of violence: A cross-country analysis”, *BMC Public Health*, Vol. 20, No. 1, Jul. 2020, doi: 10.1186/s12889-020-09069-7.
- [10] M. A. Eissa, “The Effectiveness of A Joint Attention Training Program On Improving Communication Skills Of Children With Autism Spectrum Disorder”, Vol. 4, No. 3, pp. 3–12, 2015.
- [11] J. Kourassanis-Velasquez and E. A. Jones, “Increasing Joint Attention in Children with

- Autism and Their Peers”, *Behav. Anal. Pract.*, Vol. 12, no. 1, pp. 78–94, 2019, doi: 10.1007/s40617-018-0228-x.
- [12] A. L. Mangia, M. Pirini, L. Simoncini, and A. Cappello, “A feasibility study of an improved procedure for using EEG to detect brain responses to imagery instruction in patients with disorders of consciousness”, *PLoS One*, Vol. 9, no. 6, 2014, doi: 10.1371/journal.pone.0099289.
- [13] Y. Si, F. Li, K. Duan, Q. Tao, C. Li, Z. Cao, Y. Zhang, B. Biswal, P. Li, D. Yao, and P. Xu, “Predicting individual decision-making responses based on single-trial EEG”, *Neuroimage*, Vol. 206, No. November 2019, p. 116333, 2020, doi: 10.1016/j.neuroimage.2019.116333.
- [14] M. O’Neill and T. Shear, “EEG for Diagnosis of Autism Spectrum Disorder”, *Pediatr. Neurol. Briefs*, Vol. 32, no. 0, p. 13, 2018, doi: 10.15844/pedneurbriefs-32-13.
- [15] C. Amaral, S. Mouga, M. Simões, H. C. Pereira, I. Bernardino, H. Quental, R. Playle, R. McNamara, G. Oliveira, and M. Castelo-Branco, “A feasibility clinical trial to improve social attention in Autistic Spectrum Disorder (ASD) using a brain computer interface”, *Front. Neurosci.*, Vol. 12, no. JUL, pp. 1–13, 2018, doi: 10.3389/fnins.2018.00477.
- [16] D. Ambrose, D. E. MacKenzie, P. Ghanouni, and H. F. Neyedli, “Investigating joint attention in a guided interaction between a child with ASD and therapists: A pilot eye-tracking study”, *Br. J. Occup. Ther.*, 2020, doi: 10.1177/0308022620963727.
- [17] B. Siposova and M. Carpenter, “A new look at joint attention and common knowledge”, *Cognition*, Vol. 189, No. March, pp. 260–274, 2019, doi: 10.1016/j.cognition.2019.03.019.
- [18] G. Dawson, K. Toth, R. Abbott, J. Osterling, J. Munson, A. Estes, and J. Liaw, “Early Social Attention Impairments in Autism: Social Orienting, Joint Attention, and Attention to Distress”, *Dev. Psychol.*, Vol. 40, no. 2, pp. 271–283, 2004, doi: 10.1037/0012-1649.40.2.271.
- [19] M. Simões, D. Borra, E. Santamaría-Vázquez, GBT-UPM, M. Bittencourt-Villalpando, D. Krzemiński, A. Miladinović, Neural\_Engineering\_Group, T. Schmid, H. Zhao, C. Amaral, B. Direito, J. Henriques, P. Carvalho, and M. Castelo-Branco, “BCIAUT-P300: A Multi-Session and Multi-Subject Benchmark Dataset on Autism for P300-Based Brain-Computer-Interfaces”, *Front. Neurosci.*, Vol. 14, no. September, 2020, doi: 10.3389/fnins.2020.568104.
- [20] N. Friedman, T. Fekete, K. Gal, and O. Shriki, “EEG-based prediction of cognitive load in intelligence tests”, *Front. Hum. Neurosci.*, Vol. 13, no. June, 2019, doi: 10.3389/fnhum.2019.00191.
- [21] M. Elgendi and C. Menon, “brain sciences Review Assessing Anxiety Disorders Using Wearable Devices: Challenges and Future Directions”, 2019, doi: 10.3390/brainsci9030050.
- [22] N. J. Hill and J. R. Wolpaw, *Brain-Computer Interface* ☆, No. July 2015. Elsevier Inc., 2016.
- [23] H. Ayaz, *Advances in Neuroergonomics and Cognitive Engineering*, AHFE 2019. Gewerbestrasse: Springer Nature Switzerland AG, 2020.
- [24] F. E. Abd El-Samie, T. N. Alotaiby, M. I. Khalid, S. A. Alshebeili, and S. A. Aldosari, “A Review of EEG and MEG Epileptic Spike Detection Algorithms”, *IEEE Access*, Vol. 6, pp. 60673–60688, 2018, doi: 10.1109/ACCESS.2018.2875487.
- [25] M. A. S. Ali and M. Abd-Elfattah, “A hybrid model (SVM-LOA) for epileptic seizure detection in long-term EEG records using machine learning techniques”, *International Journal of Intelligent Engineering and Systems*, Vol. 11, no. 5, pp. 162–172, 2018, doi: 10.22266/IJIES2018.1031.15.
- [26] M. Xu, H. Qi, L. Ma, C. Sun, L. Zhang, B. Wan, T. Yin, and D. Ming, “Channel Selection Based on Phase Measurement in P300-Based Brain-Computer Interface”, *PLoS One*, Vol. 8, no. 4, 2013, doi: 10.1371/journal.pone.0060608.
- [27] E. J. Cushion, J. Warmenhoven, J. S. North, and D. J. Cleather, “Principal component analysis reveals the proximal to distal pattern in vertical jumping is governed by two functional degrees of freedom”, *Front. Bioeng. Biotechnol.*, Vol. 7, no. AUG, pp. 1–11, 2019, doi: 10.3389/fbioe.2019.00193.
- [28] R. Ramos, B. Valdez-Salas, R. Zlatev, M. S. Wiener, and J. M. B. Rull, “The discrete wavelet transform and its application for noise removal in localized corrosion measurements”, *Int. J. Corros.*, Vol. 2017, 2017, doi: 10.1155/2017/7925404.
- [29] M. Mamun, M. Al-Kadi, and M. Marufuzzaman, “Effectiveness of wavelet denoising on electroencephalogram signals”, *J. Appl. Res. Technol.*, Vol. 11, no. 1, pp. 156–160, 2013, doi: 10.1016/S1665-6423(13)71524-4.

- [30] D. Popov, A. Gapochkin, and A. Nekrasov, “An algorithm of daubechies wavelet transform in the final field when processing speech signals”, *Electron.*, Vol. 7, no. 7, pp. 1–10, 2018, doi: 10.3390/electronics7070120.
- [31] S. Jefferies, “Deep Learning: The Bad, the Ugly, and the Good”, *VAHPERD J.*, Vol. 28, no. 1, p. 8, 2006.
- [32] G. Li, C. H. Lee, J. J. Jung, Y. C. Youn, and D. Camacho, “Deep learning for EEG data analytics: A survey”, *Concurr. Comput.*, No. December 2018, pp. 1–13, 2019, doi: 10.1002/cpe.5199.
- [33] M. Kästner, B. Hammer, M. Biehl, and T. Villmann, “Generalized functional relevance learning vector quantization”, In: *Proc. of ESANN 2011 - 19th Eur. Symp. Artif. Neural Networks*, Vol. 15, pp. 93–98, 2011.
- [34] E. M. Imah, F. Al Afif, M. Ivan Fanany, W. Jatmiko, and T. Basaruddin, “A comparative study on Daubechies Wavelet Transformation, Kernel PCA and PCA as feature extractors for arrhythmia detection using SVM”, In: *Proc. of IEEE Region 10 Annual International Conference, TENCON*, 2011, pp. 5–9.
- [35] N. Belgacem, “ECG Based Human Authentication using Wavelets and Random Forests”, *Int. J. Cryptogr. Inf. Secur.*, Vol. 2, No. 2, pp. 1–11, Jun. 2012, doi: 10.5121/ijcis.2012.2201.
- [36] Pablo Lanillos, Daniel Oliva, Anja Philippsen, Yuichi Yamashita, Yukie Nagai, Gordon Cheng, “A review on neural network models of schizophrenia and autism spectrum disorder”, *Neural Networks*, Vol. 122, pp 338-363, 2020, <https://doi.org/10.1016/j.neunet.2019.10.014>.
- [37] V. S. Adama, B. Schindler, and T. Schmid, “Using Time Domain and Pearson’s Correlation to Predict Attention Focus in Autistic Spectrum Disorder from EEG P300 Components”, In: *Proc. of MEDICON 2019*, pp. 1890–1893, 2020.