EEG Classification Analysis for Diagnosing Autism Spectrum Disorder based on Emotions

Nur Fadzilah Harun, Nabilah Hamzah, Norliza Zaini, Maizura Mohd Sani, Haryanti Norhazman and Ihsan Mohd Yassin
Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia
nurfadzilahharan93@gmail.com

Abstract—This research sets out to propose another method for the medical teams to diagnose Autism Spectrum Disorder (ASD) in children based on the analysis on Electroencephalography (EEG). Its main intention is to provide an effective and more time-saving method in diagnosing autism in suspected autistic children and to choose the best technique for classifying the EEG data to distinguish the Autistic traits from the normal ones. This research paper comprises of EEG data analysis on the brainwave activities of normal individuals and autism subjects to learn and compare the difference between their brain activity’s patterns. In addition, classification and analysis were also done to distinguish the different emotion projection of autistic and normal subjects and how their characteristics differ from each other. The signal processing techniques were performed on EEG data obtained from chosen subjects and two of the most well-known machine learning techniques which are Artificial Neural Network (ANN) and Support Vector Machine (SVM) were utilized in classifying the different classes of brainwave activities and signals. The efficiency of the two classifiers was then compared. For ANN, an experiment to determine the optimum value of hidden layers was also executed. The results obtained from this research provided classification accuracy that can be obtained from normal and autistic data classification as well as the ability to diagnose a new data using the trained ANN. Positive findings were obtained from this EEG analysis especially in classifying normal and Autistic patterns and also in classifying the different emotions. This outcome can very much help in the process of diagnosing ASD, where the whole process can be done in a more time-efficient manner and more accurate diagnosis can be made.

Index Terms—Autism; Classification and Analysis; Diagnosing; Electroencephalography; Emotions; Machine Learning.

I. INTRODUCTION

Autism and Autism Spectrum Disorder (ASD) are the terms used to define a group of complex disorders of brain development. These disorders are characterized by a variety and multiple levels of difficulties in social interaction, verbal and nonverbal communication and repetitive behaviors [1]. Autism is apparently found to take place in early brain development, where most signs and symptoms tend to emerge between the ages of 2 to 3 years old. In this context, [2] defines autistic children to be ones that are perceived to be living in their own world.

ASD or autism can be associated with intellectual disability, difficulties in motor movement abilities besides short attention spans and physical health issues such as sleep and gastrointestinal disturbances. These are a few common characteristics of autistic children [2]. The autistic children are often found to be having communication problem, where they might refuse to engage in conversation, unable to use proper language, repeating things that they hear and many of them have a low lingual understanding. They also tend to avoid social interaction. Most of them avoid making eye contact with people and are lacking in the ability to receive affection from people close to them as they usually prefer to be alone.

They also have unique playing patterns. They are usually absorbed in repetitive actions and can be over-possessive over their toys and belongings. Their movements can be peculiar to that of normal children. Some also have sensory disruption. These disruptions can be auditory, visual, touch, taste, vestibular and proprioceptive. These autistic children can be hypersensitive or hyposensitive, where they can get easily emotional. They sometimes have an unpredictable and always-changing mood.

However, some people with autism are found to excel in visual skills, music, math and art. Some of them also have a good memory and even a high IQ. General studies done on this topic have found out a few possible factors of autism. These factors can be a genetic factor, medicinal intake during pre and post-natal and maybe even polluted and toxic environment. In other words, the variety of characteristics and symptoms causes the process of diagnosing Autism Spectrum Disorder (ASD) to be more challenging. This is because there are many different signs and symptoms that are needed to be identified when performing the behavioral tests on the subjects.

A. Related Works

In this study, we explore and analyze the EEG signal of the human brain obtained from ASD subjects to interpret the different components distinguishing them from normal brainwave. Electroencephalography (EEG) has been proven to be an effective medium to study human brainwave in order to perform data classification on the obtained signals. EEG signal provides a non-invasive way to recognize the emotion of disabled people through EEG headset electrodes placed on their scalp [8]. A new study by researchers at Albert Einstein College of Medicine of Yeshiva University suggests that measuring how fast the brain responds to sights and sounds could help in objectively classifying people on the autism spectrum and may help diagnose the condition earlier [3, 6].

Currently, autism diagnosing in Malaysia still needs to be performed by psychologist, psychiatrist, neurologist, developmental paediatrician, or similarly qualified medical professional. There are also no medical tests performed on the subjects, the diagnosis is made based fully on the subjects’ history and symptoms. This comprehensive diagnosing process, which consists of (i) observation of the subjects and
(ii) a thorough interview with the subjects’ parents or caregivers, generally takes approximately 2 hours to complete [7].

In this context of research, this study utilized and compared the efficiency of two most common Machine Learning Techniques which are Artificial Neural Network (ANN) and Support Vector Machine (SVM). Artificial Neural Network (ANN) is a proven effective machine learning technique for data classification in many diseases diagnosing.

Based on the same purpose of disease diagnosis, ANN with radial basis functions is used to diagnose patients with multiple sclerosis [9, 11]. The diagnosis model of Smear Negative Pulmonary Tuberculosis based on ANN was constructed and the diagnosis efficiency was confirmed in [10] and the results indicate that it could be used as a new diagnosis method for the complex problem. Other related studies include [12] that used SVM for Breast Cancer Diagnosis using Screen Film Mammogram Data and [13] that used SVM and Image Processing Technique for Lung Cancer Diagnosis System.

B. Problem Statement

Based on the study done on related topics, there are a few problem statements that can be derived regarding the diagnosing processes of Autism Spectrum Disorder (ASD).

Firstly, behavioral tests on ASD suspects are found to be time-consuming. Presently, there are no medical tests that can diagnose autism [1]. When parents notice that their child is showing unusual behaviors such as the inability to make eye contact, not responding to his or her name or playing with toys in an unusual and repetitive ways, they will refer to their nearest medical assistance. Because there are no medical tests that can immediately identify autism traits in the subjects, the diagnosing process is done by specially trained physicians and psychologists who will administer autism-specific behavioral evaluations. These manual diagnosing processes require a lot of time in order to carefully observe the characteristics and behaviors of the suspected autistic children.

Another issue is that the Autism characteristics are widely variant and it is hard to determine the autistic level in suspected children. The Autism Spectral Disorder has a few different levels. There are a few distinct characteristics that can distinguish the autistic children from normal ones. However, it is more complex to determine the different level of the ASD itself. This is because not one autistic child is the same to another. ASD has different levels and classifications. For example, there are autistic children that can converse well, and there are some that could not even talk. Some autistic children are very affectionate, while some others are very elusive in social interactions. Some autistic children have high IQs, even higher than some normal people.

The point is, these autistic traits that these children possess are different from one another. It is safe to say that some of these autistic children are almost close to normal and some of them have a wider range of disabilities of limited abilities.

In the context of EEG analysis, no study is found to identify which analysis or classification method is the best option to detect Autism characteristics and distinguish them from the normal ones. In processing and classifying the EEG data obtained from autistic children’s brainwave activities, there are a few methods or techniques that can be used to distinguish the EEG data whether the brainwave belongs to an autistic kid or a normal one. Some of the most suitable techniques that can be used for data classification are Artificial Neural Network (ANN) and Support Vector Machine (SVM) [14]. However, there has been no comparison made to determine which one of the two techniques is the best option to use to classify EEG data of autistic children brainwave, in order to obtain more accurate results.

C. Objectives

Hence, this research is done in order to propose another convenient, accurate and time-efficient method for the medical teams to diagnose autism in children based on the analysis on Electroencephalography (EEG). Its main intention is to provide an effective and more time-saving method in diagnosing autism in suspected autistic children and to choose the best technique for classifying the EEG data to distinguish the Autistic traits from the normal ones. The specific aims of this study are defined as follows:

i. To design an EEG analysis specially to diagnose the occurrence of Autistic traits by differentiating the brainwave patterns of autistic children from normal brainwave.

ii. To apply the machine learning techniques to determine and classify the different levels of Autism Spectrum Disorder (ASD) in autistic children by studying their emotions as compared to emotions of normal subjects.

iii. To compare the different machine learning techniques, i.e. between ANN (Artificial Neural Network) and SVM (Support Vector Machine); and to decide which of these techniques is the best to differentiate brainwave signal of autistic children.

This study is scoped down to focus on the EEG analysis to compare the autistic and normal brainwave activities. The analysis was done on the emotion comparison between normal and autistic subjects. Comparisons were also made to compare the efficiency of the two machine learning techniques used which are ANN and SVM. This research focuses on experimenting with the optimum value of ANN hidden layers for classifying autistic and normal EEG data. The main mission of this study is to propose an EEG based diagnosing tools for the purpose of automatic diagnosing system development.

II. METHODOLOGY

This research project can be divided into three main parts; i) Signal Acquiring and Processing Phase, ii) Signal Classification Phase and iii) Data Analysis and Validation Phase. The process methodology flowchart is shown in Figure 1.

There are a few major stages that had to be undergone in order to classify the EEG signals of the subjects of this experiment. As shown in Figure 1, first, an experimental procedure was performed in order to acquire raw EEG signals from the chosen subjects. Then, the obtained signals went through a signal pre-processing stage which consists of noise and artefact removal, bandpass filter, frequency filtering. After the pre-processing stage, the signal has undergone feature extraction and normalization process before being classified using the chosen machine learning techniques. The final stage was the data validation stage, which was the analysis of the classified data obtained from the classification process performed.
EEG Classification Analysis for Diagnosing Autism Spectrum Disorder based on Emotions

Fig 1: Process methodology flowchart

A. Experimental Procedure to Acquire Raw EEG Data

The EEG data were obtained from the selected subjects by using the Emotiv device. Emotiv is an EEG reader that is wirelessly connected to the computer/laptop while data were being captured from the subjects. The EEG recording system consists of electrodes with sensor pads, amplifiers with built-in filters and 16-bit analog to digital converter, the sampling rate of 128 Hz and a computer. The EEG signal was recorded using 14 channels monopolar connection grounded behind both ears.

The EEG signals were recorded using the same device and computer for all subjects to ensure consistency in the analysis. Figure 2 shows the illustrations of EEG Emotiv reader and its connections to the subject’s scalp.

The subjects were chosen among children between the age range of 4 to 12 years old for both autistic and normal subjects. The autistic subjects were selected only from those with an approval from their parents. The subjects’ genders were also varied to obtain a more general data for both autistic and normal subjects.

Before EEG data is recorded from the subjects, they were introduced to the following procedures for data collection. First, the Emotiv device is attached to the subject’s scalp. The subject is then asked to close their eyes for 30 seconds and then open them again. Then, they are shown a video on neutral feeling simulation. The neutral feeling simulation consisted of calm and emotionless slideshows of images and videos such as animals and flowers with soothing background music. This stimulation aimed to calm the subjects and help them to relax and focus.

The subject is then asked to rest and then the process is repeated for happy video stimulation and sad video stimulation. The happy video stimulation shows images of happy, smiling, and laughing children with upbeat background music. The sad video stimulation is given extra attention and care in the process of image selection as some images can potentially be disturbing for the subjects. The process flows for the experimental procedure to acquire EEG data from the subjects are shown in Figure 3.

Figure 3: Experimental Procedure to Acquire EEG Data

Figure 4 shows the setting for raw EEG data collection. The EEG signals of each subject during all stimulations were then recorded as shown in Figure 5. The overall time for data collection was approximately 20 minutes per subject.

B. Signal Pre-Processing

The signal pre-processing stage was basically the removal of noises and non-required signals from the obtained EEG data of the subjects. This process is also known as noise and artefacts removal. The obtained raw EEG data from Emotiv software was in “.edf” file, which was then required to be converted into “.csv” (Excel) type of file. The appropriate tools provided by the Emotiv test bench software was used to convert the “.edf” file into “.csv” file to enable data to be imported into MATLAB software to be processed. In order to obtain artefact-free EEG signal, the data had to be filtered.

Figure 4: EEG Data Collection from a Normal Subject

Figure 5: EEG Data Analysis & Validation
using band filters. The bandpass filter was selected using four types of frequency bands which were alpha, beta, theta and delta band, according to the specified frequency range of the required data. The PSD (Power Spectral Density) mean values were chosen based on the four frequency bands and were used as the inputs to the classifiers. In other words, the EEG features used for classification is the output from the bandpass filtering process executed.

The greatest advantage of ANNs is their ability to be used as an arbitrary function approximation mechanism that learns from observed data. It is known as an effective technique for data classification, including pattern and sequence recognition, novelty detection and sequential decision-making. Hence, ANN is used in this study as one of the machine learning techniques in classifying the EEG brainwave data and classifying between the normal and the autistic brainwave, which will help to develop an automated diagnosing system for Autism.

Support Vector Machines (SVM) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier [9].

The distinctive characteristic of an SVM is the use of a separating hyperplane. The operation of SVM algorithm is based on finding the hyperplane that gives the largest minimum distance to the training examples. This distance receives the important name of margin within SVM’s theory. Hence, the optimal separating hyperplane maximizes the margin of the training data [7]. Figure 7 shows an example of how the hyperplane is assigned for a set of data on a 2D Cartesian plane.

The four different values (according to the ranges) of Power Spectral Density (PSD) features obtained from the signal pre-processing stage is used as the input for the ANN and SVM classifiers. An answer scheme was set as to train the classifiers to produce classification accuracy based on the provided values of either 0 or 1 for the two different types of input data.

**Data Analysis and Validation:** After the signal-features values had been classified using the two machine learning techniques, all the data obtained were analyzed and then validated. Confusion matrices were used to validate the data. There are four terms used in a confusion matrix: True Positive (TP), False Negative (FN), False Positive (FP) and True Negative (TN).

The data analysis and validation phase were performed on a few categories of data classification. The main goal of this project is to analyze the difference between normal and ASD subjects EEG signal and brainwave activities. The second goal is to analyze the difference between the emotion classification of normal subjects and autistic subjects, i.e. Neutral vs Sad or Neutral vs Happy. This explains the different emotion stimulation during the experimental procedure when acquiring the raw EEG data.

**Feature Extraction**

After the signal pre-processing stage, the signals were now artefact-free. However, the signal was not guaranteed to be good enough for it to be efficiently processed by the MATLAB classifiers. Hence, feature extraction process needs to be performed on the data to help the MATLAB classifier to efficiently classify the different tendency of the brainwave characteristics based on the different patterns captured in the EEG signal of the subjects.

**Normalization Process**

The normalization process is required to ensure that the value of the data lies only in between desired range. For example, in Artificial Neural Network (ANN) machine learning technique, the data has to be normalized before the neural network is trained. The normalization technique will be used to minimize the high value of power band ratio data.

**Data Classification using Machine Learning Techniques**

Artificial Neural Network (ANN) is a machine learning technique that consists of an interconnected group of nodes, similar to the vast network of neurons in human brain [8]. Here, each circular node represents an artificial neuron and an arrow represents a collection from the output of one neuron to the input of another. Figure 6 shows the illustration of an artificial neural network structure.

**Figure 5:** A sample of raw EEG data obtained from subjects

**Figure 6:** Artificial Neural Network Architecture

**Figure 7:** Hyperplane on a 2D Cartesian plane
Each of the procedures used in this project played their own crucial roles respectively in achieving the goals of this research. The experimental procedures were necessary in order to obtain the EEG data that will be used as the input and the training of the classifiers while signal processing stage is crucial in making sure that the data was free from noises and in the right range suitable for classification. The classification process played a major role in classifying the input values and generating new data. The data analysis and validation phase were also crucial parts in translating all the results obtained and making valid conclusions from them.

III. RESULT AND ANALYSIS

In this study, we had performed classification analysis based on a few categories of data. This analysis includes the Normal vs Autistic classification, Autism Emotion Classification and Normal Emotion Classification. In addition, an analysis and comparison of the optimum hidden layer (patternet value of ANN classifier were also carried out. The results and findings of each analysis are demonstrated and discussed accordingly next.

A. Normal vs. Autism Classification

As explained in the earlier parts of this paper, the classification process for normal and autistic EEG data was executed using two different machine learning techniques, which are ANN and SVM. The accuracy of the classifiers using the same sets of input data are shown in Table 1.

Table 1

<table>
<thead>
<tr>
<th>Normal vs. Autism Classification Using ANN and SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine Learning Techniques</td>
</tr>
<tr>
<td>Artificial Neural Network (ANN)</td>
</tr>
<tr>
<td>Support Vector Machine (SVM)</td>
</tr>
</tbody>
</table>

These classifications were performed using the EEG data of the subjects’ neutral emotion for the use of autism diagnosing. For ANN classifier, the pattern recognition tools, i.e. ‘nptool’ in the NN toolbox were utilized in training the ANN classifier. There were two classes defined as the variable used for the ANN classifier output neuron, which was normal and autistic. The output for each sample target was set as ‘0’ for normal and ‘1’ for autistic. The number of hidden layer (patternet) used for all ANN classification was set as 55 as it was found and proven as the most optimum value of hidden layers. This will be explained in another section of this paper.

As shown in Table 1, the value of accuracy obtained from ANN classifier is significantly higher than the accuracy value of data classification obtained using SVM when using the exact same data sets. Hence, it can be concluded that ANN is the better classifier in classifying the EEG data sets. This can be due to the fact that ANN has its hidden layers processing mechanisms that can aid in making more accurate predictions. As for SVM, the accuracy value is still good enough for data classification but the error percentage is considerably high when compared to ANN classifier’s accuracy.

B. Emotions Classification for Normal and Autistic Subjects

This part of research compares and classifies the different emotions of both normal and autistic subjects. This is done for the purpose of studying the autistic subjects’ different emotions towards different feelings stimulation. It also aims to compare the difference between autistic emotions to the normal emotions and how they differ in terms of the abilities to react and process information from the emotion stimulators. The obtained results are summarized in Table 2.

By referring to Table 2, it can be observed that the accuracy values for autistic emotions classifications are significantly lower than that of the normal emotions when using both ANN and SVM classifiers. The accuracy value of a classifier mainly indicates how easy it is for the provided data to be classified most probably due to the distinct differences in patterns and characteristics between the two classes.

Hence, the lower accuracy value of the classifier indicates that the data is more complex and therefore harder to classify into two clear parts as the values of the data are inconsistent. Therefore, it can be said that the autistic subjects have a more complex emotion as compared to normal subjects. This fact explains why autistic children are usually showing minimum emotion when interacting with people or unable to differentiate their own feelings towards something.

C. Optimum Hidden Layer Value for ANN

This study also included an experimentation of hidden layer (patternet) values to determine the most optimum value of the hidden layer to be used for ANN classification in producing the highest accuracy.

A number of values ranging from 5 to 100 with the interval of fives, e.g. 5, 10, 15,………..100 were set and experimented with different hidden layer values for ANN. For each value of the hidden layers, the ANN classifier is executed ten times using the same data sets, which is the data for Normal vs Autistic (Neutral Emotion). The ten values of accuracy obtained are averaged out and the hidden layer value with the highest average accuracy is chosen as the hidden layer value for ANN classifier.

From the performed hidden layer values testing, it was shown that the hidden layer value of 55 provided the highest average accuracy with 89.05%, while the hidden layer value of 35 is in second place with 87.62% average accuracy. From these findings, it can be concluded that the optimum value for ANN hidden layer in classifying normal and autistic data would be higher than the usual hidden layer values for classifying other types of data.

D. New Data Classification for Autism Diagnosing

After all of the study and comparison made between the two machine learning techniques, i.e. ANN and SVM, it was concluded that ANN provided a higher accuracy in classifying normal and autistic data. Hence, ANN will be recommended as the primary method to diagnose a new data in this research. This means that the trained ANN can be used to classify a new data and determine whether the data belongs
to the normal or autistic group. This is in order to achieve the primary goal of this project, which is to recommend a diagnosing method for Autism Spectrum Disorder (ASD).

Therefore, a sample program was constructed in order to develop the basic idea of the intended diagnosing system. The trained ANN is imported and then executed to ensure its optimum accuracy value (this is because ANN classifier can produce different values when executed numerous times) and from the trained ANN, the highest threshold value was chosen as the set point for the new data classification. This was done to make sure that the classification of the new data would be using the highest value of accuracy from the trained ANN.

Figure 8 shows a sample of the confusion matrix obtained as a reference before the new data classification was executed. The accuracy obtained, as shown in the confusion matrix was first made sure to be optimum before the trained classifier was used to classify new data.

New data were then chosen randomly from one of the normal subjects and one of the autistic subjects to be classified into their expected categories. The classifications of the new data provided manually from an excel file were successful as the program produced the right output for the respective normal and autistic data. Figure 9 shows the correct output of the program for both normal and autistic data classification.

IV. CONCLUSION

In completing this study, a number of appropriate Electroencephalography (EEG) classification analysis was successfully carried out to differentiate the brainwave of autistic children from normal children’s brainwave patterns. This finding can be a stepping stone in helping to simplify the time-consuming manual diagnosing processes using the behavioral tests and observations. This classification analysis can be employed as part of an ASD diagnosing system, where a more accurate diagnosis can be determined. This analysis can also help us to better understand the different traits of the different levels of ASD. A comparison between different machine learning techniques was also done to analyze their performance in classifying the different emotions in autistic and normal children. Another contribution of this research is an experiment on the different values of ANN-classifier’s hidden layers in determining the optimum value of hidden layers, i.e. in producing the highest classification accuracy. In addition, this research also concluded the comparison between the two different techniques under machines learning, which are Artificial Neural Network (ANN) and Support Vector Machine (SVM) in terms of accuracy in classifying the brainwave signals of the autistic children and used the trained ANN to diagnose new data. The best technique proposed for the ASD diagnosing process is ANN due to its higher value of accuracy as compared to SVM classifier.

ACKNOWLEDGMENT

We would like to extend our acknowledgment to Universiti Teknologi MARA (UiTM) and to those who have directly and indirectly contributed to our project. This research is funded by Ministry of Higher Education Malaysia via Research Acculturation Grant Scheme (RAGS/2013/UiTM/ICT06/1) and Fundamental Research Grants Scheme Research (FRGS/1/2014/TK03/UiTM/02/14).

REFERENCES

[9] Á. Gutiérrez, “Influence of Wavelets and Boundary Conditions on the Diagnosis of Multiple Sclerosis Using Artificial Neural Networks,” in

