Investigation of Hybrid Feature Selection Techniques for Autism Classification using EEG Signals

S.Thirumal[1]
Research Scholor, Department of CSE
Hindustan Institute of Technology and Science
Chennai, India

J.Thangakumar[2]
Associate Professor, Department of CSE
Hindustan Institute of Technology and Science
Chennai, India

***Abstract*—Autism Spectrum Disorder (ASD), the non-uniform neurodevelopment condition that is characterized by the impairment of behaviour in communication and social interaction with some restricted their repetitive behaviour. Today, to measure the voltage created during brain activity is measured using electroencephalography (EEG). The wavelet transform is used for decomposing the time-frequency of the EEG signal. Feature Selection is the process that significantly reduces feature space dimensionality, while maintaining the right representation of their original data. In this work, metaheuristic algorithm is utilized for feature selection. The proposed feature selection is based on River Formation Dynamics (RFD) and a hybrid Greedy RFD is presented. Support Vector Machine (SVM) can be a concept consisting of a set of methods of supervised learning to analyze pattern recognition that is a successful tool in the analysis of regression and classification. Experimental results show the proposed Greedy RFD feature selection improves the performance of the classifiers and enhance the accuracy of classifying ASD.**

Keywords*—Autism spectrum disorder (ASD); electroencephalo graphy (EEG); feature selection; River Formation Dynamics (RFD); Support Vector Machine (SVM); hybrid greedy RFD

I. INTRODUCTION

All the existing medical conditions in human are recognised with the support of a medical professional [1]. But, owing to any variation in their physiological signals, the assessments can result in human error and are not similar while being met by a medical professional. Autism has been identified as a new spectrum disorder that can affect every individual in a different manner in different degrees. This is characterized by means of social deficits, poor communication, and repetitive behavioural patterns. Nevertheless, all problems with children that have a condition of Autism Spectrum Disorder (ASD) can contribute to this. Even though the ministry has been focusing on its early identification resulting in screening and intervention, there are some challenges faced. Different tools are available to diagnose the condition of ASD using this. Human brain have specific paths to send, receive and to interpret all information to bring about a response through its synaptic activation.. These pathways are known as sensory systems [2, 3].

The electroencephalogram (EEG) is a signal acquired form individuals that are scientifically and clinically sound. As a consequence, quantification has a vital role in the study of the human brain [4]. The frequency content and examination of EEG signals, in particular, were identified as a preponderant approach to the knowledge extraction problem. These were used in the research of brain processes related to Motor Imagery (MI). It shows there were tremendous development in the Brain –computer Interface (BCI) technology, which aims to replace compromised human neuromuscular system capability. The majority of BCIs depend on physiologically well-defined EEG properties, such as oscillations in neural networks or their potentials generated by specific stimuli. Various methods that assess the energy of the signal that is dispersed in this frequency, which is t-f or t-s domains, dominated the extraction of EEG features.The Discrete Wavelet Transform (DWT) further decomposes this EEG segment into various sub-bands [5]. Different statistical features and entropy functions were used for the extraction of features from sub-bands.

Multi-variate relations in data can be captured using Machine learning techniques [6] and are very suitable for the detection distribution and subtle differences of certain found in data. Therefore, when compared to the other univariate approaches, the machine learning approach will be able to perform better in terms of classifying the EEG, especially in terms of conditions such as ASD [7, 8]. Therefore, they hold plenty of promise in bringing about knowledge improvement in diagnosing ASD. The Support Vector Machines (SVM) tends to have some merits like higher accuracy and not needing large numbers of training samples that avoid overfitting. Therefore, the SVM has now aroused plenty of concern among the neuroimaging research community [9]. GBM iteratively builds by combining prediction from several weak learners and achieves good performance [10]. In this work, greedy RFD has been investigated to classify the EEG and ASD.

The main contribution of this work is:

The features of EEG and the behavioural data is fused to achieve better accuracy in classifying ASD.

The River Formation Dynamics (RFD) metaherustic is used for selecting features from the wavelet transforms.

The proposed Greedy RFD improves the performance of RFD.

II. RELATED WORK

Sudirman et al [11] aimed to build a new sensory profile using the EEG bio-signal and its potential for distinguishing between various sensory responses. All EEG signals needed here were useful in identifying various emotional states like super learning, light relaxation, and positive thinking. These were inside a frequency range falling between 8 and 12 Hertz. A total of 64 children were part of this research, and from among them, about 34 children were given vestibular sensory, taste, sound, and visual simulations. All raw EEG data had been filtered using an Independent Component Analysis (ICA), and wavelet transforms with the EEGLAB software. To build its sensory profile, standard deviations means and entropy approximation had been extracted from filtered EEG signals.

Herman et al [12] had performed a study which compares spectral signal representation like Power Spectral Density (PSD) techniques, continuous and discrete wavelets, atomic decompositions, and Time-Frequency (t-f) energy distributions. The main emphasis was on the identification of certain differentiated properties in the feature sets that represent EEG trials that are recorded at the time of imagination, that is, for the left-hand or the right-hand movement. Separating features can be quantified in an offline study by making use of the accuracy of classification based on a rate that is obtained using linear as well as non-linear classifiers. There are PSD approaches that demonstrate a consistent level of robustness along with effectiveness that is useful in certain distinctive spectral patterns. They are used to differentiate whether MI induced EEGs are belongs to left and right. The observation has been based on data analysis from a total of eleven subjects in two different sessions. Additionally, the capabilities of generalization of these classifiers found in their intersession performance have been discussed.

Cheong et al [13] studied the Discrete Wavelet Transform (DWT) for extraction of features from their EEG signals that were obtained based on a sensory response from children suffering from autism. For the purpose of this study, the DWT was used in order to decompose the filtered EEG signal into their components with a statistical DWT coefficient feature computed within the time domain. Such features were employed for training Multilayer Perceptron (MLP) based neural network for the classification of signals into three different classes based on the severity of autism (whether mild, moderate, or severe). The results of training in terms of accuracy had achieved about 92.3% with an MSE of 0.0362.

Fan et al [14] had proposed a data-driven method to design the RFD. Speaking technically, the RFD was constructed using threshold responses using receptive fields from many candidates based on their distinctiveness and their correlations. By means of employing two types of such receptive fields (rectangular pooling and Gaussian pooling), two different binary descriptors, the RFDR and RFDG, were chosen. The experiments of image matching experiments on the Patch Dataset and the Oxford Dataset proved that the RFD was able to outperform them, and its work was comparable to the float-valued descriptors at a short time. Experiments on object recognition proved that the RFDR and RFDG were able to show better performance than their competitors.

The human genome can be used to extract vast amounts of data. Autism is a type of neurobehavioral disorder in which a person's capacity to interact and communicate is impaired. It has a solid genetic foundation. There are various gene variants linked to autism, and these changes can disrupt the functioning of the brain that begins before birth. Mutated genes are passed down from their immediate ancestors to their offspring, and this is a risk factor for autism. Reeta et al [15] had proposed another novel approach for ranking such diseased genes that are found in an autistic individual. This system will predict the autistic behaviour of individuals by means of comparing certain similarities among the genes of individuals and the diseased for training that is implemented by making use of the method known as the Naive Bayesian classification. For example, in case the DNA of an individual will be tested, and in case it is found that the DNA consists of certain diseased genes in the training set, autism may be predicted. The approach also makes the process of diagnosis of the condition simpler and earlier as well.

In Vaishali and Sasikala [16], the diagnosis dataset of the ASD will have 21 features that are obtained from the repository of the UCI machine learning, and this has been experimented with by using a swarm intelligence-based binary firefly feature selection wrapper by using the same, there was an identification made that about 10 features among a total of 21 ASD datasets were enough to distinguish between the patients that had ASD or did not. The obtained results, along with the proposed approach, were able to justify the new hypothesis by means of producing a certain amount of average accuracy within the range of about 92.12%-97.95% along with optimum feature subsets that were equal to the accuracy of the ASD.

In Alzubi et al [17], a hybrid mechanism to select feature which is more accurate was proposed to detect informative SNPs to be chosen for an optimal SNP subset. This method was based on the fusion of that of a filter and also a wrapper method along with the Conditional Mutual Information Maximization (CMIM) method and Support Vector Machine Recursive Feature Elimination (SVM-RFE). The proposed method's performance had been evaluated based on three different state-of-the-art methods, which are the Minimum Redundancy Maximum Relevance (mRMR), ReliefF, and the CMIM. It also used four classifiers which were the Naive Bayes (NB), SVM, Linear Discriminant Analysis (LDA), and the k Nearest Neighbors (kNN) on an ASD-SNP dataset that was obtained from the Gene Expression Omnibus and National Center for Biotechnology Information genomics data repository. The results of the experiment had demonstrated that the efficiency of this approach of feature selection outperformed other methods and achieved a classification accuracy of 89%.

III. Methodology

Feature selection refers to a technique that was employed for pre-processing data, and this is preferable at the time of performing machine learning. Selection by mean is to choose attributes and variables within a dataset that is fit into a particular model and is tested for performance. The section further details the extraction of features by making use of the Wavelet Transform, Support Vector Machine, RFD, Greedy RFD, Naïve Bayesian Classification, and K- Nearest Neighbours.

A. Datasets

The techniques are evaluated using autism dataset obtained from King Abdulaziz University (KAU) Brain Computer Interface (BCI) Group. The EEG data is recorded using all the electrodes with 16 channels. This dataset had been filtered using a band-pass filter along with pass band frequency (0.1–60Hz) and a notch filter that had stop band frequency (60Hz) and at 256Hz frequency sampling is digitized. The recording time of EEG is varied from 12 to 40 minutes among autistic subjects up to 173 minutes, and from 5 to 27 minutes up to 148 minutes for normal subjects.

The behavioural dataset related to autism screening of toddlers containing features that is used for further analysis to determine autistic traits. Ten behavioural features is recorded with other individuals characteristics.

B. Feature Extraction using Wavelet Transform (WT)

The These wavelets were in use in the recent decade for different tasks of image processing. For the purpose of image compression, fractals compression, fractals, resolution enhancement, denoising, and image enhancement frequency domain and time analysis is used. The basic idea behind all of this is the analysis of the signal in accordance with the scale. The main advantages of such Wavelet transforms are that compared to the Fourier transforms that represent functions; there are some discontinuities with sharp peaks that help in the accurate deconstruction or reconstruction of non-periodic, non-stationary finite signals. The images based on wavelet, enhancement, de-noising, and so on had better performance owing to the properties of sparsity or multi-resolution structure. Wavelet transform also has the trait of multi-resolution analysis aside from the ability to express a local feature of this signal in the domain of time and frequency. Therefore, it is found to be fit to detect the flash state or the irregularity of the signal and in setting out the composition [18]. The wavelet ψ having a compact support and vanishing moment n is as in (1):

$$\int_{-\infty}^{+\infty} t^k \psi(t) dt = 0, \text{ for } 0 \leq \mathrm{k} \leq \mathrm{n} \quad (1)$$

This has another function θ that has fast decay and is shown as (2).

$$\psi(t) = (-1)^n \frac{d^n \theta(t)}{dt^n} \quad (2)$$

Once this is done, the wavelet transform for signal f is given as in (3):

$$Wf(u,s) = s^n \frac{d^n}{du^n} (f * \overline{\theta}_s)(u) \quad (3)$$

Wherein, $\overline{\theta}_s(t) = s^{-1/2}\theta(-t/s)$, are the time and space coordinate and s the scale. The wavelet transforms Wf (u, s) is the nth order derivative of f that has $\overline{\theta}_s$ on a domain proportional to s.

The most important frequencies of EEG exist between 0.1 to 30 Hz. The standard EEG clinical bands are the delta (0.1 to 3.5 Hz), theta (4 to 7.5 Hz), alpha (8 to 13 Hz), and beta (14 to 30 Hz) bands. A sample EEG image and corresponding alpha, beta, delta and theta waveforms are shown in Fig. 1. 30Hz waves will termed as gamma waves.

In this work, the number of decomposition levels is taken as 5. Thus, the EEG signal is decomposed into D1-D5 details. As a result output coefficients of mean, standard deviation, variance, skewness and kurtosis feature vectors are used for classifying the signal.

C. River Formation Dynamics (RFD) Feature Selection

The algorithm of River Formation Dynamics was earlier used in solving problems that are NP-hard to find paths within a graph [19]. Furthermore, the algorithm is also capable of optimizing the distance of the path and also considers certain other dependencies like restrictions to acceleration, velocity, pathfinding tasks, and so on, which are considered to be NP-hard problems [20]. The RFD algorithm can be depicted as given below. The actual amount of soil that is assigned to each of the nodes will drop as they keep moving and erode the paths or depositing the carried sediment (thereby increasing node altitude). The descending slope makes more dependence on the probability of deciding the next node, and which can be relative to the actual the node's altitude dissimilarity, the altitude of its adjacent node, and the drop position. The environment created in the early stages will be flat. This means that throughout the process, the nodes will have the same height, except in the case of a target node with a height of zero. There is a blotch placed at the starting node to navigate the entire site, which determines the optimal path. At every step there is another set of droplets are traversed in space sequentially and erosion is marked at the nodes traversed.

The nodes represent features, and the routes between them reflect the decision of the next feature when utilising RFD to optimise feature selection. The drop travelling through with the least number of nodes visited that meets the stopping requirement is used to find the best feature subset.

The first step will be the initialization of the nodes of the algorithm and define the set that is formed using the cell decomposition of the site. Every node here will possess all information, and in case it consists of an obstacle aside from additional data, the determination of the time that is required for traveling across the entire distance to its goal is considered. When there is a drop initialization there may be a suitable number of drops that are kept to the first node. After this, the algorithm will be executed until such time the last condition is met.

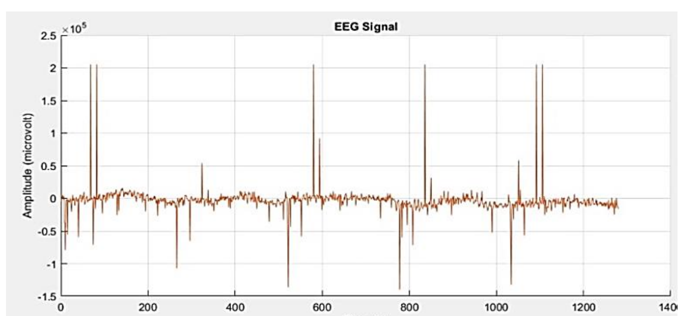

Fig. 1. Sample EEG Images.

The statements says that the drops were in the same path, and in addition to that, for reducing the time taken for computation, a upper limit on the actual number of iterations was introduced along with the condition that verifies whether the last n loops have improved the solution or not. Path analysis will involve identifying the right solution. For the purpose of discovering the best drop that can conduct any additional erosion, it has to be performed on travelled paths by means of reducing the altitude of the nodes. The final step will be to add some sediment to the nodes in order to overcome any circumstances in which all the altitudes are near 0, thus making the gradients can be ignored, which may destroy the paths formed. This will diminish slowly with each loop in the algorithm.

D. Proposed Greedy River Formation Dynamics Feature Selection

The Greedy algorithm is the one that can determine this problem by means of making a choice that appears to be the best at that moment. There are several problems of optimization that may be solved by means of using this algorithm, and some may not have an efficient solution. However, the Greedy algorithm can implement such a solution which is nearly optimal. The Greedy algorithm also reflects a problem-solving heuristic that makes a locally optimal choice in every stage, hoping to be able to find a global optimum. For most of these issues, the greedy strategy may not be able to produce an optimal solution, but it can bring about a locally optimal one that can approximate another globally optimal solution within a reasonable time frame.

```
    Initialize Nodes(), Drops(), Greedy (D, n)
    Solution <- 0
    while (not endConditionMet())
    For i<-1 to n do {
        S<-Select (D)
        If (Feasible (solution, s)) then
                Solution <-Union (Solution, s)}
        Return solution
    moveDrops()
    analyzePaths()
    erodePaths()
    depositSediments()
end while
```

E. K- Nearest Neighbors (KNN)

Computing the closest distance between the neighbours is depicted as the K value, and for making use of this algorithm, a few other elements in the initial subject set, which is the K number (of nearest neighbours), is observed, and the K parameter and distance were considered.

These training tuples have been described in the form of n attributes. In this, every tuple will indicate the point within the n-dimensional space. So, all training tuples will be stored within an n-dimensional pattern space. For a certain unknown tuple, the k-NN classifier will lookout for a pattern space in the k training tuples that are the closest to the unknown one. The k training tuples represent the k-nearest neighbours of this unknown tuple. “Closeness” has been defined as the distance metric like the Euclidean distance. Euclidean distance between that of two points or tuples will be$X = (x_1, x_2, \ldots., x_n)$ and $Y = (y_1, y_2, \ldots., y_n)$ is (4):

$$d(X, Y) = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2} \quad (4)$$

The step by step process of K-NN algorithm is given below:

- Computing the distance between that of the new sample and all earlier samples that have been grouped in the form of clusters;
- Sorting the distance in increasing order to choose k samples having smaller distance values;
- Applying a voting principle and adding a new sample to the largest cluster of the k samples.

F. Naive Bayesian Classification

Naive Bayes’ is very popular method in the categorization of the texts and in identifying documents of a certain type as to whether they are legitimate or spam. The Naive Bayes’ Classifier can be very scalable and needs linear parameters in the variables (or predictors) for a learning problem. The probability models with strong assumptions of independence of Naïve Bayes classification using diseased gene classification and is a conditional model.

$$P(C_k|x) = P(C_k) \,.\, P(x|C_k)\, P(x) \quad (5)$$

If there is a genetic instance that has to be classified and is represented by a vector $x = (x_1, x_2 \ldots\ldots x_n)$ will represent n genes with assigned probabilities. $P(C_k, x_1 \ldots\ldots x_n)$ It is for every possible k outcome or class Ck.

G. Support Vector Machine (SVM)

The SVM refers to a new form of supervised learning algorithm for rightly categorizing a target result that uses independent variables which is present inside the dataset. Another new SVM can be the maximum margin classifier, and this will further maximize any separation between the n classes of data. The SVMs are useful, especially when there is a boundary between the groups that are non-linear and owing to this feature, they are normally used for problems of classification where there is a distinction made between groups that are non-linear. The SVM algorithms are used for classifying individuals based on diagnosis, neuroimaging, genes, standardized assessments, and some more measurements [21]. The SVM classification algorithm makes use of training instances to predict other new ones using two different class label −1, 1. As per Fig. 1, a hyperplane is $w^T x + b = 0$, wherein, $w \in R^n$ is orthogonal to a hyperplane and $b \in R^n$is constant. With training data D, equation (6) is derived:

$$D = \{(\vec{x}_i, \vec{y}_i) | \vec{x}_i \in R^m, \vec{y}_i \in \{-1, +1\}\}_{i=1}^{n} \quad (6)$$

In which x_i is the m-dimensional real vector, y_i the input vector class, x_i either -1 or $+1$. The SVM looks for a hyperplane to maximize the margin between two sample classes in D.

$$y_i(\vec{w}^T\vec{x}+b) \geq 1 \tag{7}$$

The aim of the SVM is to increase the distance between two different hyperplanes. One will compute the distance between both hyperplanes $\frac{1}{||\vec{w}||}$. The SVM training in a non-separable case will be solved with a problem of quadratic optimization as in Equation (8):

$$min\,i\,mize\text{: P}(\vec{w},\text{b},\xi) = \frac{1}{2}||\vec{w}|| + C\sum_{i=1}^{n}\xi_i$$

$$subject\ to\text{: y}(\vec{w}.\phi(\vec{x})\text{+b}) \geq \text{1-}\xi_i\ ,\ \xi_i \geq 0 \tag{8}$$

H. Gradient Boosting Machine (GBM)

An ensemble technique, boosting will help to reduce the bias which is dependent and also generalized error in an ensemble. Another technique of boosting which will repeatedly combine 30 base (weak) learners that have low variance and high bias like the stumps in the decision tree. These base learners are combined to ensure ensemble bias, and this will reduce the variance remaining the same, thus reducing its net ensemble error. For every boosting step or iteration, the GBM will construct yet another new base learner to a negative gradient of the loss function with the observed data in order to ensure the focus of the new base learner is on the model and its weakness. This means a functional approximation for the model has been made by bringing about a consecutive improvement with the negative direction to the loss function [22].

Normally, the GBM algorithm will have better results in the case, for every iterative step, there can be an added decision tree and its contribution, which is reduced by using a parameter for shrinkage α which is known as the rate of learning. The main idea behind this method of shrinkage in the GBM will be that it has more steps, and these small ones will result in better accuracy compared to a less number of larger steps. This parameter of learning α will fall between 0 and 1, and the smaller the value, the more its accuracy.

For each iterative step, as opposed to making use of a complete training dataset, a randomly chosen (that does not have a replacement) subsample that will fit a decision tree will be used. If there are many observations a default fraction of this data will ½=0.5. This means about half i.e.,50%) of the dataset can be used. Additionally, algorithms computation cost will be reduced by subsampling to the means of a factor that is equal to the subsampling factor [23] and improves the accuracy of Gradient boost machine model. The algorithm used for gradient boosting is as given below:

- Initialize the predictions with one simple decision tree.
- Calculate the residual – and this will be its (actual-prediction) value.
- Build a shallow decision tree to predict the residual based on independent values.
- Update its original prediction with another one multiplied by its rate of learning.
- Iterate the step 2 to 4 for certain number of times which should be equal to the tree count.

I. Proposed Greedy RFD GBM

The proposed Greedy RDF is used to optimize the GBM's hyperparameters. The GBM model hyperparameter of number of trees, tree depth, learning rate, Minimum number of observations in terminal nodes is optimized using the Greedy RFD optimization algorithm. In the proposed Greedy RFD GBM, the initial solutions are created randamly. The range specified for number of trees is 1000 to 5000, tree depth 1 to 10, learning rate 0.05–0.3, Minimum number of observations in terminal nodes 5-15. The Root Mean Square Error (RMSE) is used as the objective function. On iterations of the Greedy RBF, the optimal set of GBM hyperparameters to classify ASD is obtained.

IV. Result and Discussion

The techniques were evaluated for two scenarios, using features from EEG data only and using features from both EEG and behavioural data. In the latter, the features of EEG and the behavioural data is fused. In this section, the RFD feature methods with classifiers such as KNN, NB, SVM and GBM are evaluated. Section 4.1 presents the results for EEG data without feature fusion and section 4.2 the results for feature fusion.

A. Without Feature Fusion

Tables I to IV and Fig. 2 to 5 shows accuracy of the classification, precision, recall and F measure for both normal and ASD features.

TABLE I. Classification Accuracy for Greedy RFD-Greedy RFD GBM without Feature Fusion

Techniques Used	Classification Accuracy
RFD-KNN	92.31
RFD-NB	92.69
RFD-SVM	93.27
RFD-GBM	95.19
Greedy-RFD-GBM	95.58
Greedy RFD-Greedy RFD GBM	96.35

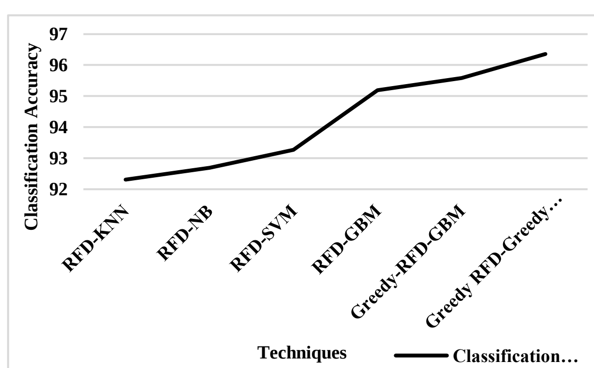

Fig. 2. Classification Accuracy for Greedy RFD-Greedy RFD GBM without Feature Fusion.

From Fig. 2, it can be observed that the Greedy RFD-Greedy RFD GBM has higher classification accuracy by 4.3%, by 3.9%, by 3.25%, by 1.21% and by 0.8% for RFD-KNN, RFD-NB, RFD-SVM, RFD- GBM and Greedy-RFD-GBM, respectively. The selection of optimal hyperparameters of GBM shows improved performance.

TABLE II. RECALL FOR GREEDY RFD-GREEDY RFD GBM WITHOUT FEATURE FUSION

Techniques Used	Recall for normal	Recall for ASD
RFD-KNN	0.9214	0.9237
RFD-NB	0.9286	0.9263
RFD-SVM	0.9357	0.9316
RFD-GBM	0.95	0.9526
Greedy-RFD-GBM	0.95	0.9579
Greedy RFD-Greedy RFD GBM	0.9571	0.9658

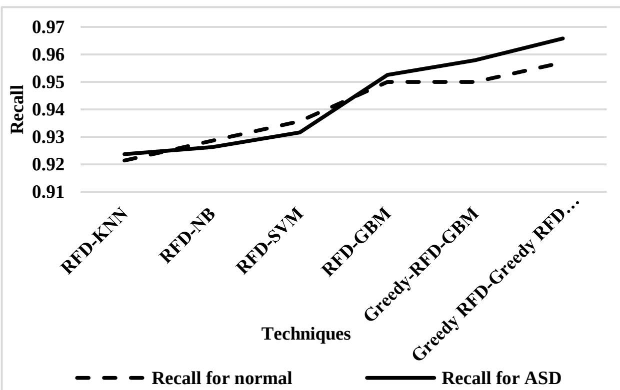

Fig. 3. Recall for Greedy RFD-Greedy RFD GBM without Feature Fusion.

From Fig. 3, it can be observed that the Greedy RFD-Greedy RFD GBM has higher recall by 3.8%, by 3.02%, by 2.26%, by 0.74% and by 0.74% for RFD-KNN, RFD-NB, RFD-SVM, RFD- GBM and Greedy-RFD-GBM, respectively, for normal. Similarly, the Greedy RFD-Greedy RFD GBM has higher recall by 4.46%, by 4.18%, by 3.6%, by 1.38% and by 0.82% for RFD-KNN RFD-NB, RFD-SVM, RFD- GBM and Greedy-RFD-GBM, respectively for ASD. It is observed that the feature selection significantly improves the classification of the ASD. The proposed Greedy RFD feature selections achieves the best performance.

TABLE III. PRECISION FOR GREEDY RFD-GREEDY RFD GBM WITHOUT FEATURE FUSION

Techniques Used	Precision for normal	Precision for ASD
RFD-KNN	0.8165	0.9696
RFD-NB	0.8228	0.9724
RFD-SVM	0.8344	0.9752
RFD-GBM	0.8808	0.981
Greedy-RFD-GBM	0.8926	0.9811
Greedy RFD-Greedy RFD GBM	0.9116	0.9839

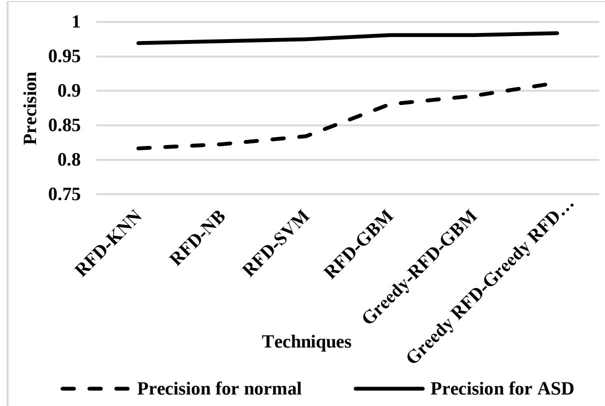

Fig. 4. Precision for Greedy RFD-Greedy RFD GBM without Feature Fusion.

From Fig. 4, it can be observed that the Greedy RFD-Greedy RFD GBM has higher Precision by 11%, by 10.24%, by 8.84%, by 3.44% and by 2.1% for RFD-KNN, RFD-NB, RFD-SVM, RFD- GBM and Greedy-RFD-GBM, respectively for normal. Similarly, the Greedy RFD-Greedy RFD GBM has higher Precision by 1.46%, by 1.2%, by 0.89%, by 0.3% and by 0.29% for RFD-KNN RFD-NB, RFD-SVM, RFD- GBM and Greedy-RFD-GBM, respectively for ASD. The optimization of the feature selection and classifier hyperparameters has significant improvement in the performance. The proposed Greedy RFD is effective when compared to RFD.

TABLE IV. F MEASURE FOR GREEDY RFD-GREEDY RFD GBM WITHOUT FEATURE FUSION

Techniques Used	F measure Normal	F Measure ASD
RFD-KNN	0.8658	0.9461
RFD-NB	0.8725	0.9488
RFD-SVM	0.8822	0.9529
RFD-GBM	0.9141	0.9666
Greedy-RFD-GBM	0.9204	0.9694
Greedy RFD-Greedy RFD GBM	0.9338	0.9748

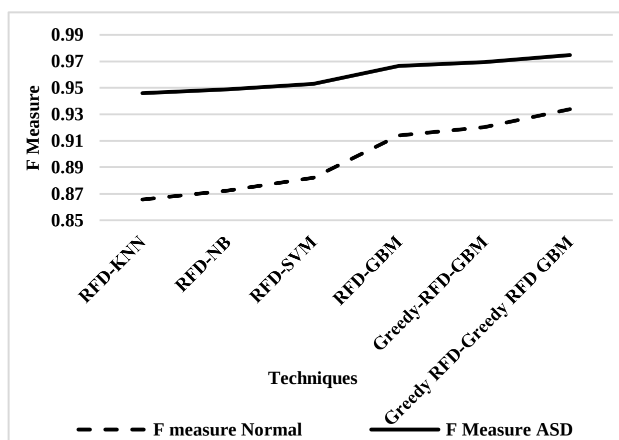

Fig. 5. F measure for Greedy RFD-Greedy RFD GBM without Feature Fusion.

From Fig. 5, it can be observed that the Greedy RFD-Greedy RFD GBM has a higher F measure by 7.6%, by 6.79%, by 5.68%, by 2.13% and by 1.45% for RFD-KNN, RFD-NB, RFD-SVM, RFD- GBM and Greedy-RFD-GBM respectively for normal. Similarly, the Greedy RFD-Greedy RFD GBM has higher F measure by 2.99%, by 2.7%, by 2.3%, by 0.84% and by 0.56% for RFD-KNN, RFD-NB, RFD-SVM, RFD- GBM and Greedy-RFD-GBM, respectively for ASD.

B. *After Feature Fusion*

Headings, The classification accuracy, recall, precision and F Measure for both normal and ASD features as shown in Table V to VIII and Fig. 6 to 9.

TABLE V. CLASSIFICATION ACCURACY FOR GREEDY RFD FEATURE SELECTION-GREEDY RFD GBM

Techniques	Classification accuracy
RFD-Feature Selection - KNN	94.08
RFD-Feature Selection -Naïve Bayes	94.64
RFD-Feature Selection - SVM	95.21
RFD Feature Selection - GBM	95.78
Greedy RFD feature selection GBM	96.56
Greedy RFD feature selection - Greedy RFD GBM	97.15

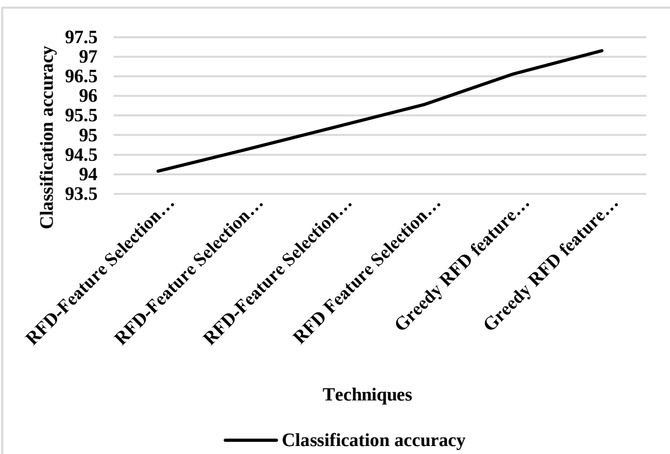

Fig. 6. Classification Accuracy for Greedy RFD Feature Selection-Greedy RFD GBM.

From Fig. 6, it can be observed that the Greedy RFD feature selection-Greedy RFD GBM has higher classification accuracy by 3.21% for RFD-feature selection-KNN, by 2.62% for RFD-feature selection-Naive Bayes, by 2.01% for RFD-feature selection-SVM, by 1.42% for RFD-feature selection-GBM and by 0.61% for Greedy RFD feature selection GBM, respectively.

From Fig. 7, it can be observed that the Greedy RFD feature selection-Greedy RFD GBM has higher recall for normal by 2.94% for RFD-feature selection-KNN, by 2.22% for RFD-feature selection-Naive Bayes, by 1.48% for RFD-feature selection-SVM, by 1.48% for RFD-feature selection-GBM and by 0.74% for Greedy RFD feature selection GBM respectively. The Greedy RFD feature selection-Greedy RFD GBM has higher recall for ASD by 3.3% for RFD-feature selection-KNN, by 2.77% for RFD-feature selection-Naive Bayes, by 2.22% for RFD-feature selection-SVM, by 1.39% for RFD-feature selection-GBM and by 0.55% for Greedy RFD feature selection GBM, respectively.

From Fig. 8, it can be observed that the Greedy RFD feature selection-Greedy RFD GBM has higher precision for normal by 8.33% for RFD-feature selection-KNN, by 6.99% for RFD-feature selection-Naive Bayes, by 5.63% for RFD-feature selection-SVM, by 3.56% for RFD-feature selection-GBM and by 1.44% for Greedy RFD feature selection GBM, respectively. The Greedy RFD feature selection-Greedy RFD GBM has higher precision for ASD by 1.12% for RFD-feature selection-KNN, by 0.84% for RFD-feature selection-Naive Bayes, by 0.56% for RFD-feature selection-SVM, by 0.56% for RFD-feature selection-GBM and by 0.27% for Greedy RFD feature selection GBM, respectively.

TABLE VI. RECALL FOR GREEDY RFD FEATURE SELECTION-GREEDY RFD GBM

Techniques	Recall for Normal	Recall for ASD
RFD-Feature Selection – KNN	0.9348	0.9431
RFD-Feature Selection -Naïve Bayes	0.9416	0.9482
RFD-Feature Selection - SVM	0.9485	0.9534
RFD Feature Selection - GBM	0.9485	0.9613
Greedy RFD feature selection GBM	0.9556	0.9694
Greedy RFD feature selection - Greedy RFD GBM	0.9627	0.9748

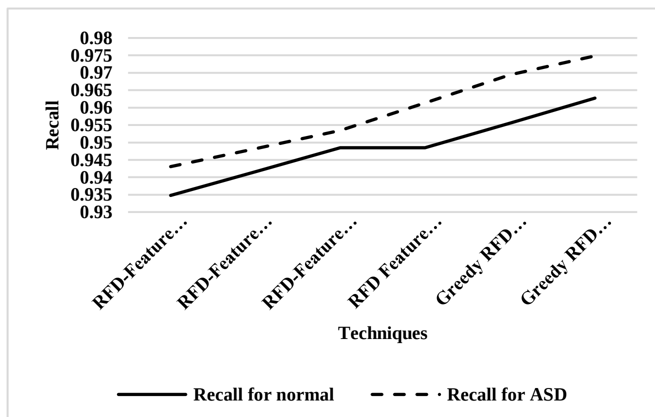

Fig. 7. Recall for Greedy RFD Feature Selection-Greedy RFD GBM.

TABLE VII. PRECISION FOR GREEDY RFD FEATURE SELECTION-GREEDY RFD GBM

Techniques	Precision for Normal	Precision for ASD
RFD-Feature Selection - KNN	0.86	0.9748
RFD-Feature Selection -Naïve Bayes	0.8716	0.9775
RFD-Feature Selection - SVM	0.8836	0.9803
RFD Feature Selection - GBM	0.9021	0.9803
Greedy RFD feature selection GBM	0.9214	0.9831
Greedy RFD feature selection - Greedy RFD GBM	0.9348	0.9858

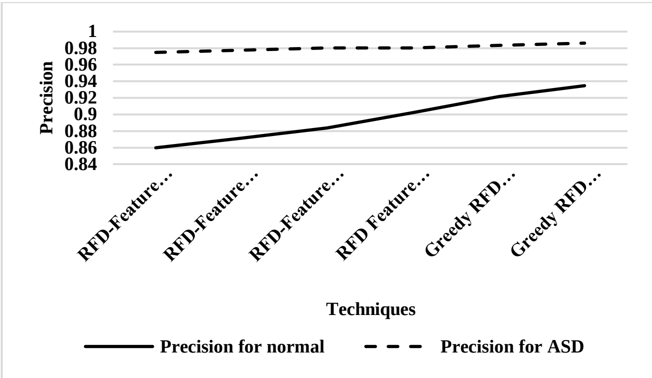

Fig. 8. Precision for Greedy RFD Feature Selection-Greedy RFD GBM.

TABLE VIII. F MEASURE FOR GREEDY RFD FEATURE SELECTION-GREEDY RFD GBM

Techniques	**F Measure for Normal**	**F Measure for ASD**
RFD-Feature Selection - KNN	0.8958	0.9587
RFD-Feature Selection -Naïve Bayes	0.9052	0.9626
RFD-Feature Selection - SVM	0.9149	0.9666
RFD Feature Selection - GBM	0.9247	0.9707
Greedy RFD feature selection GBM	0.9381	0.9762
Greedy RFD feature selection - Greedy RFD GBM	0.9485	0.9802

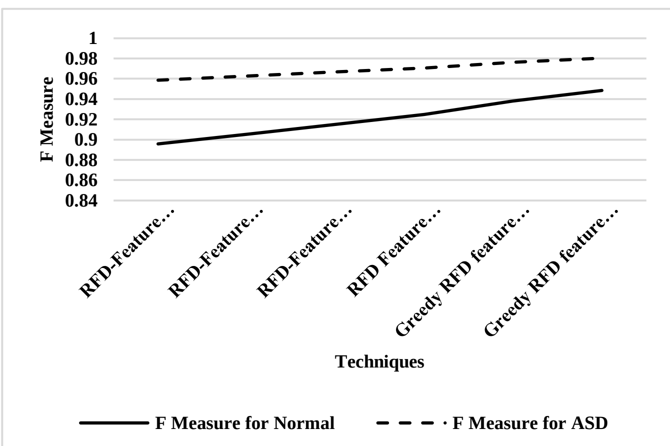

Fig. 9. F Measure for Greedy RFD Feature Selection-Greedy RFD GBM.

From Fig. 9, it can be observed that the Greedy RFD feature selection-Greedy RFD GBM has higher f measure for normal by 5.71% for RFD-feature selection-KNN, by 4.67% for RFD-feature selection-Naive Bayes, by 3.61% for RFD-feature selection-SVM, by 2.54% for RFD-feature selection-GBM and by 1.09% for Greedy RFD feature selection GBM respectively. The Greedy RFD feature selection-Greedy RFD GBM has higher f measure for ASD by 2.22% for RFD-feature selection-KNN, by 1.82% for RFD-feature selection-Naive Bayes, by 1.39% for RFD-feature selection-SVM, by 0.98% for RFD-feature selection-GBM and by 0.41% for Greedy RFD feature selection GBM, respectively.

It is observed that the feature fusion using EEG and behavioural data has higher classification of ASD of 97.15% when compared to 96.35% accuracy with only EEG data. Similarly, precision, recall and f measure achieved is better when feature fusion is used. Though the classification is enhanced, the sampling used is small. The proposed methods needs to be evaluated with larger dataset.

V. CONCLUSION

Autism Spectrum Disorder (ASD) can be a lifelong condition that is quite serious and characterized by restricted but repetitive behaviour along with deficits in terms of communication or its reciprocal social interaction. The EEG and behavioural data is used for identifying the ASD. In this work, the wavelet transform is used to extract features from EEG. Features of the EEG and behavioural data is fused. Feature selection is achieved using proposed RFD and Greedy RFD algorithms. SVM can take a collection of input data to make a prediction where the binary class input will make the SVM a two-class linear classifier that is non-probabilistic. GBM iteratively builds by combining prediction from several weak learners. The hyperparameters of the GBM is optimized using Greedy RFD to improve the classification of the ASD. The results have proved that Greedy RFD-Greedy RFD GBM with fused features from both EEG and behavioural data had a better accuracy of classification than the RFD-KNN, RFD-NB, RFD-SVM, RFD-GBM, and the Greedy-RFD-GBM. The findings show that fusing of EEG and behavioural data does improve the classification of ASD. Future research can focus on evaluating the proposed methods using larger dataset.

REFERENCES

[1] Daniel P. Howsmon , Uwe Kruger, Stepan Melnyk, S. Jill James, Juergen Hahn , "Classification and adaptive behavior prediction of children with autism spectrum disorder based upon multivariate data analysis of markers of oxidative stress and DNA methylation", PLOS Computational Biology, March 2017.

[2] Bone, D., Goodwin, M. S., Black, M. P., Lee, C. C., Audhkhasi, K., & Narayanan, S. (2015). Applying machine learning to facilitate autism diagnostics: pitfalls and promises. *Journal of autism and developmental disorders*, *45*(5), 1121-1136.

[3] Nogay, H. S., & Adeli, H. (2020). Machine learning (ML) for the diagnosis of autism spectrum disorder (ASD) using brain imaging. *Reviews in the Neurosciences*, *31*(8), 825-841.

[4] Brodu, N., Lotte, F., & Lécuyer, A. (2011, April). Comparative study of band-power extraction techniques for motor imagery classification. In *2011 IEEE symposium on computational intelligence, cognitive algorithms, mind, and brain (CCMB)* (pp. 1-6). IEEE.

[5] Ibrahim, S., Djemal, R., & Alsuwailem, A. (2018). Electroencephalography (EEG) signal processing for epilepsy and autism spectrum disorder diagnosis. *Biocybernetics and Biomedical Engineering*, *38*(1), 16-26.

[6] Raj, S., & Masood, S. (2020). Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques. *Procedia Computer Science*, *167*, 994-1004.

[7] Thabtah, F., & Peebles, D. (2020). A new machine learning model based on induction of rules for autism detection. *Health informatics journal*, *26*(1), 264-286.

[8] Hyde, K. K., Novack, M. N., LaHaye, N., Parlett-Pelleriti, C., Anden, R., Dixon, D. R., & Linstead, E. (2019). Applications of supervised machine learning in autism spectrum disorder research: a review. *Review Journal of Autism and Developmental Disorders*, *6*(2), 128-146.

[9] Kang, J., Han, X., Song, J., Niu, Z., & Li, X. (2020). The identification of children with autism spectrum disorder by SVM approach on EEG and eye-tracking data. *Computers in biology and medicine*, *120*, 103722.

[10] Katuwal, G. J., Cahill, N. D., Baum, S. A., & Michael, A. M. (2015, August). The predictive power of structural MRI in Autism diagnosis. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 4270-4273). IEEE.

[11] Sudirman, R., Hussin, S. S., Airij, A. G., & Hai, C. Z. (2018, July). Profile Indicator for Autistic Children using EEG Biosignal Potential of Sensory Tasks. In *2018 2nd International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)* (pp. 136-141). IEEE.

[12] Herman, P., Prasad, G., McGinnity, T. M., & Coyle, D. (2008). Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *16*(4), 317-326.

[13] Cheong, L. C., Sudirman, R., & Hussin, S. S. (2015). Feature extraction of EEG signal using wavelet transform for autism classification. *ARPN Journal of Engineering and Applied Sciences*, *10*(19), 19.

[14] Fan, B., Kong, Q., Trzcinski, T., Wang, Z., Pan, C., & Fua, P. (2014). Receptive fields selection for binary feature description. *IEEE Transactions on Image Processing*, *23*(6), 2583-2595.

[15] Reeta, R., Pavithra, G., Priyanka, V., & Raghul, J. S. (2018, April). Predicting Autism Using Naive Bayesian Classification Approach. In *2018 International Conference on Communication and Signal Processing (ICCSP)* (pp. 0109-0113). IEEE.

[16] Vaishali, R., & Sasikala, R. (2018). A machine learning based approach to classify Autism with optimum behaviour sets. *International Journal of Engineering & Technology*, *7*(4), 18.

[17] Alzubi, R., Ramzan, N., & Alzoubi, H. (2017, August). Hybrid feature selection method for autism spectrum disorder SNPs. In *2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)* (pp. 1-7). IEEE.

[18] Chaudhari, Y. P., & Mahajan, P. M. (2017). Image denoising of various images using wavelet transform and thresholding techniques. *Int. Res. J. Eng. Technol.(IRJET)*, *4*(2).

[19] Redlarski, G., Pałkowski, A., & Dąbkowski, M. (2013). Using river formation dynamics algorithm in mobile robot navigation. In *Solid State Phenomena* (Vol. 198, pp. 138-143). Trans Tech Publications Ltd.

[20] Pałkowski A., Use of swarm intelligence methods in mobile robot navigation [in Polish: Wykorzystanie metod inteligencji rojowej w nawigacji robotów mobilnych], MSc thesis, Gdansk University of Technology, Gdansk, 2012.

[21] Sharma., A, Nagpal., D & Kaur., S, "Comparative Study of Machine Learning Techniques Used in Autism Spectrum Disorder", International Journal for Research in Engineering Application & Management (IJREAM) ISSN : 2454-9150 Vol-05, Issue-10, Jan 2020.

[22] Das, K., & Behera, R. N. (2017). A survey on machine learning: concept, algorithms and applications. *International Journal of Innovative Research in Computer and Communication Engineering*, *5*(2), 1301-1309.

[23] Touzani, S., Granderson, J., & Fernandes, S. (2018). Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy and Buildings*, *158*, 1533-1543.