Identifying EEG Binary Limb Motor Imagery Movements using Thick Data Analytics

Vikas Trikha[1], Jinan Fiaidhi[2], Sabah Mohammed[3]

Abstract

Electroencephalography (EEG) is non-invasive technology that is widely used to record brain signals in brain computer interfacing (BCI) systems to control, motor imagery, in which movements signals occurring in limbs can control some services. Researchers have proposed numerous classification schemes of these motor imagery to incorporate it with various neurorehabilitation, neuroprosthetics and gaming applications. However, the existing classification schemes face the performance degradation caused by motor-imagery EEG signals with low signal to noise ratio. The paper's main objective is to use possible thick data analytics techniques to classify effectively the motor imagery EEG signals. Our attempt start with notable classifiers including Decision Trees, Extra Trees, Naive Bayes, Random Forest and SVM and move later to enhance classifications using variety of ensemble learning techniques including Bagging, Adaboost and Stacking. More techniques has been applied on the results of the ensemble learring to eliminate classification noise and supply more relevant features such as substituting outliers with mean value and exercising band-pass filter and Common Spatial Pattern (CSP). The thick data methods has been validated on a public dataset rendered by BCI competition II dataset III and was found to produce better classification performance metric which included performance metric parameters like accuracy, specificity, sensitivity, precision and recall when confronted with the existing work, thus projecting the usefulness of motor imagery BCI. The analytics is inclusive of Area Under the Curve (AUC) score and Mathews Correlation Coefficient (MCC) score to display an impactful analysis.

Keywords: EEG Brain Waves, EEG Motor Imagery, BCI Applications, Ensemble Learning, Thick Data Analytics.

1. Introduction

For the last five decades, computer scientist have tried to narrow the gap of communication between humans and computer technology with technologies like graphical and tangable user interfaces without much success as this process requires intensive interpretations and

1) (Professor) Department of Computer Science, Lakehead University, Thunder Bay, Ontario, Canada
  email: vikas@lakeheadu.ca
2) (Professor) Department of Computer Science, Lakehead University, Thunder Bay, Ontario, Canada
  email: jfiaidhi@lakeheadu.ca
3) (Professor, Corresponding Author) Department of Computer Science, Lakehead University, Thunder Bay, Ontario, Canada
  email: mohammed@lakeheadu.ca

interactions. However, the emerging brain-computer interfaces (BCI) is contributing to narrawing and simplifying this communication gap through enabling computers to be intentionally controlled via the monitoring of brain signal activity. Motor Imagery (MI) EEG signals can be captured via invasive or non-invasive BCI, which can be further employed as one of the most frequently used BCI applications that aim to help motor disable persons (who cannot move their leg, limbs or hands) and taking the intercommunication between human and machine to a further level. Motor Imagery is nothing but a usual mental engagement in which users are commanded to visualize themselves producing a specific motor activity such as hand or feet movement. For BCI development, subject produced EEG signals are analyzed, which would further differentiate and classify the Motor Imagery tasks, which would further process by BCI into device commands to communicate with prosthetic device to control movements[1]. Specific control instructions will be applied as assessment for motor disabled patients to interact along with outside environments to remotely control movements. If the MI tasks are presumably recognized by recognizing regular designs in EEG data, motor-impaired people could interact with a machine by producing these mental states' progressions. Therefore, an MI-based BCI presents encouraging administration and intelligent medians to characters experiencing motor disabilities. Thus, the discovery of MI tasks is fundamental for BCI construction to form control signals. [Fig. 1] casts a brief workflow of BCI application.
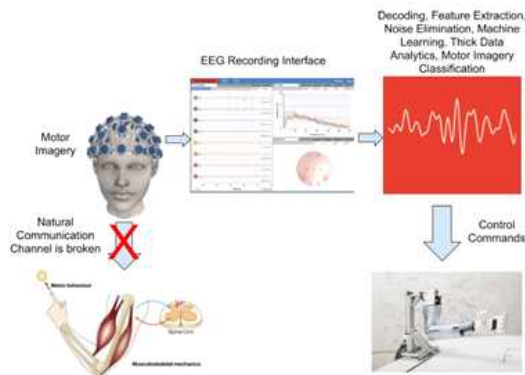

[Fig. 1] Workflow of a BCI Motor Imagery Application for Bionic Limb

The motivation for this research lies in a simple survey that asserts that nearly about 50,000 strokes are observed every year in Canada, and over 300 000 Canadians right-off live with the consequences of stroke[2]. This is where machine learning and artificial intelligence hops in to produce solutions for performing thick data analytics, but MI has its assemblage of challenges since it is challenging to quantify MI signals, which keeps MI a dynamic field of research. The aim of this paper is to review a wide analytics used in MI-based EEG systems, with a

particular focus on the state-of-the art regarding the use of thick data analytics role in feature extraction, feature selection and in providing more effective classification mechanisms.

Thick data analytics, however, is empowering big data and machine learning at scale with more qualitative informative materials, tools or techniques that help investigators gather granular, specific knowledge about their target users and learning goals. Thick data can be used to reveal people's underlying stories, emotions, motor imagery and behavioural models by using ethnographic, anthropological, qualitative, interpretive methods, conversational communities and practices as well as authomatic data labeling methods. [Fig. 2] illustrate some of the notable thick data analytics techniques used to empower variety of machine learning analytics.
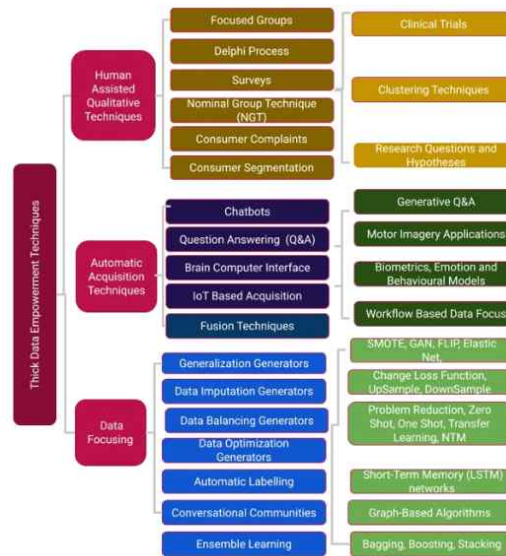
[Fig. 2] Thick Data Analytics Techniques for Machine Learning Empowerment

2. Related Research Work

The acquisition of EEG brain signals can be made either by invasive or non-invasive methods. However, invasive techniques are mostly used due to safety measures and ease of use. The first-ever human recorded EEG was conducted in 1929 by Hans Berger[3]. A lot of research in this domain has been ingested so far. However, EEG signals have low spatial interpretation and are variable and unstable. EEG beacons are receptive to artifacts produced by eye blinks, eye actions, pulsation, and muscular movements. After the signal acquisition, the fundamental

step of data pre-process that aims to remove all unwanted signals, and various approaches have been suggested by different authors so far. This section tends to define the related research made for EEG signal processing techniques. The peripheral of each paper focuses on fetching important information out of EEG signals and classifying motor imagery movements using different methods.

Feng Li et al. proposed a novel approach to discrimination algorithms based on deep learning by practicing CWT-SCNN to identify and to group left and right MI EEG signals. The proposed approach aims to refine the traditional pre-processing steps and supply pre-processed data to the upgraded CNN model to achieve better performance[4]. The proposed method comprised mainly of two stages, i.e. the first one is image generation of motor imagery EEG signals using Continuous Wavelet Transform (CWT). The following stage focuses on exercising SCNN to classify motor imagery movements to achieve better performance. The proposed approach was compared on the bases of Kappa value, and it produced the kappa value of 65.7 percent. The same aspects can be practiced using Fourier transform-based methods that can discriminate between temporal variations of EEG signals based on their MI tasks.

In their paper, Raoof Masoomi et al. have proposed LDA-based Classification of left and right-hand MI using BCI competition II dataset III and utilized wrapper sequential feature selection[5]. For signal pre-processing, the focus was retained onto remove artifacts and denoise the signal prominently notable at the 50-60 Hz power line. Among three channels, C3, C4 and Cz, authors prohibited Cz owing its presumable independency of obtaining the excellent separation. Sequential Forward Selection (SFS) and Sequential Backward Elimination (SBE) are two variants of sequential feature selection that add or remove the features from the full set until the criterion improves. LDA is the most elegant approach used in pattern recognition, machine learning and statistical evaluation. The proposed approach was evaluated on misclassification rate and it was reduced to 10% which outperformed the BCI competition II participants. Authors have implemented LDA to label every data window practicing its elected subset of features by LDA based SFS to compute the misclassification rate. The proposed methodology can be further aggregated with the quicker searching technique to procure the most optimal set of features and improve the proposed algorithm's efficiency.

Suraj et al., in their Hindawi publication entitled 'Classification of Two Class Motor Imagery Tasks Using Hybrid GA-PSO Based K-Means Clustering,' illustrates the transformation of laboratory conducted BCI experimentations to the real world to be applied synchronously with zero time constraints[6]. The authors recommended a hybrid GA-PSO based K-means Clustering procedure on classifying two MI tasks. The purpose stated to prefer a hybrid K-Means

technique over traditional evolutionary algorithms was less execution time and high computation power. The first operation step involved staying idle for 6 seconds and then thinking of moving left or right hand on the alert sound. The signal recording composed of six-channel recordings where three were bipolar cortex signals and the rest were EOG (Electrooculogram) and EEG level signals. The signals registered from all the channels were assorted into three divisions as left, right, and rest state, presenting eighteen assemblages of one-dimensional data. In the signal processing stage, the different features after computing, form a feature vector and then compute for the absolute feature vector using the feature types. The hybrid technology not only achieves a higher distribution of two-class 'left' and 'right' with fewer iterations, but the proposed clustering algorithm's performance can further be increased if it is executed for more number of repetitions and a more significant population. The asserted approach achieved the overall average performance of 57.40%, with a standard deviation of 1.736. The initial hypothesis observed for Friedman and Iman-Davenport feature type methods was rejected as the statistical values achieved were way beyond the minimum threshold.

Jinan Fiaidhi and Sabah Mohammed provided an apt introduction and need for thick data analytics that displays how big data alone is not self-sufficient to provide qualitative analytics[7]. Furthermore, thick data being a small data contains more attributes such as emotional quotient and quality parameters that can be utilized wisely to fetch meaningful analytics and the real-life example of this marvelous work is done by Netflix which boomed their profits by following thick data approach and tried to understand the customer needs rather than going for one parameter graph. Any analytics is just not fetching data from twitter and Facebook, but the importance is to be concentrated on the significance of recognizing the communication communities in social media platforms and bridge the gap between quantitative and qualitative analytics. Similarly, EEG signals not only contain the frequencies of different signals but also represent other parameters like emotions, movements, eye state, and many more, which need to be analyzed in a better way.

In a paper titled, "Envisioning Insight-Driven Learning Based on Thick Data Analytics with Focus on Healthcare" conveys a more extensive understanding of how to make patients data more understandable to identify their pain points[8]. These insights can be envisioned with the help of thick data analytics, which requires actionable knowledge in this domain as the metrics obtained by big data alone do not fulfill the concept of understandability. Dr. Jinan Fiaidhi illustrates the concepts of Insight-Driven-Learning (IDL) methods that can be explored along with machine learning techniques to discover insights about the patient's routine. Parallelly,

feature extraction must be accompanied with more attributes in EEG signals to develop more understandable BCI applications for ones in need. IDL methods can fetch timestamps, conversation methods and more thick data observations that can further add more value to the data resulting in more qualitative analytics.

Sumantha Bhattacharyya and Manoj Kumar Mukhul, in their paper, proposed a linear convolutive mixing model to classify MI signals based on the cepstrum based method[9]. For data pre-processing, the band-pass filter was employed for filtering alpha and beta signals, further accompanied by cepstrum. The energy estimation is conducted on EEG signals of channel C3 and C4, where the maximum energy of uncommon cepstrum of the EEG signals has been exercised as a feature cast and is further supplied to the Naive Bayes classifier for classification. The proposed method achieves the misclassification error of 9.29% in a classification time of 6.70 seconds. This algorithm's computational time complexity seems very fast and has been compared with other algorithms like Temporal Relative Spectral Power (TRSP).

Indu Dokare and Naveeta Kant performed experimentation on BCI competition dataset II competition III[10]. They incorporated DWT for feature extraction. This experimentation's objective primarily focused on the distribution of left and right MI signals using the Support Vector Machine (SVM) classifier. SVM organizes data by securing the most accountable hyperplane that categorizes all data points of one class from those of the other class. The signals were decomposed in 4 levels using DWT based upon their frequency that is Cd4, Cd3, Cd2 and Cd1. The most reliable hyperplane for an SVM insinuates the one with the most significant boundary separating the two categories. The wavelet coefficients were handled as features, and the distribution accuracy was used as the fundamental peripheral, which ranged from 60-70 percent.

2.1 The Dataset

Data set rendered by the Department of Medical Informatics, Institute for Biomedical Engineering, University of Technology Graz. (Gert Pfurtscheller) is practiced for experimentation, as mentioned above. The dataset was recorded on an influential citizen through a feedback session where the request for left and right signals was arbitrary. The examination consists of 7 runs with 40 trials each. All trials were carried on a corresponding day with many minutes break in midway. The only C3, Cz and C4 were the EEG channels covered in this dataset[11].
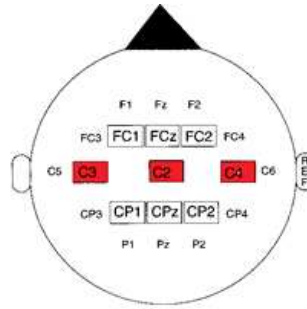
[Fig. 3] Pictorial Representation of Electrode Position

The EEG was considered including 128Hz; it was refined within 0.5 and 30Hz. The preliminaries for preparing and testing were arbitrarily chosen to stay away from the impacts of feedback. The data is accumulated in a MATLAB file format where the cue was conferred from time 3 seconds to 9 seconds. This dataset has been further used for BCI competition II dataset III, where the goal of the competition was to validate signal processing and provide classification.

2.1.1 Data Preparation

For data preparation, the wyrm toolbox is employed in this experimentation has its primary data structure named as Data. It is an uncomplicated data structure with an n-dimensional array to collect the data at its gist and a small set of meta information to describe the data sufficiently. Those additional attributes are names, axes, and units. Names attribute applied to collect the names of specific dimensions of the data, and axes represent the row plus columns of the data, furthermore units hold the physical units of the data in axes[12].

The dataset has 140 trials, 3 channels and 1152 data points per channel. This data can be stored in the form [class, time, channel] that is of shape [140,1152, 3]. Class labels replace the trials with the following notation.

names = ['Class', 'Time', 'Channel']

units = ['#', 'ms', '#']

The train and test data are prepared similarly. Except in test data, class labels do not replace the trials. Also, the sampling frequency value is added for the data as 128 Hz.

2.1.2 Data Preprocessing

Following data preparation, the signal must be processed, which may be known as a signal

enhancement. The first step was to find Nyquist frequency, which is nothing but half of the sampling frequency on which the EEG signal has been sampled. It describes the bandwidth of the sampled signal. Later, it is accompanied by Band Pass Filter.

2.1.3 Butterworth Band Pass Filter

It is recognized that each EEG beacons comprise the most critical knowledge is in the petty frequency bands. Those frequencies bands are primarily regained from the raw EEG signals. The band-pass filter's primary purpose is to limit the bandwidth of the EEG signal to the minimum necessary to convey data in the desired form. High pass and low pass filters are used to achieving this purpose. Precisely, Butterworth filter the data using 0.9 Hz high pass and 15 Hz low pass filter to flatten the response as much as possible. Further, the signals were subsampled at 16 Hz.

2.1.4 Common Spatial Pattern (CSP)

CSP is a broadly applied spatial filtering method for EEG-BCI. It boosts the difference of band-passed EEG signals from one class while limiting the sign fluctuation from the different classes[13]. It is sensitive to artifacts and electrode positions. In this MI thick dataset of right and left classes, the CSP calculation computes spatial channels that boost the proportion of the fluctuation of information originating from the two classes. This helps us to differentiate between the classes quickly. Also, it reduces the dimensionality of the data.

Further, the filtered data is supplied to variance function to check how spread out the dataset is, accompanied by a logarithm function and rectified channel that would undo the exponential effects and calculates the absolute value of data.

3. Thick Data Methods for Interpreting MI Signals

After extraction of desired data and features, it is available to be stationed in the predefined machine learning classifiers and advanced ensemble classifiers available in python such as Linear Discriminant Analysis (LDA), Random Forest, XGBoost, SVM, KNN, Logistic Regression, Naive Bayes and Decision Tree. Considering the principal part of machine learning is comprehensive of classification and explains what class an observation pertains to.

For better discrimination among categories, model selection has been made using grid search, which focuses on deducing the best results by tuning the optimum values for parameters. It avoids the exhaustive search of finding the best values for every classifier succeeding in

redeeming time, effort, and resources. The ensemble classifiers are a methodology of ensemble learning that produces diverse base classifiers from which a novel classifier is determined, aiming to perform better than other fundamental classifiers[14]. A few of the ensembled classifiers that we implemented in this experiment are Bagging, Boosting and Stacking.
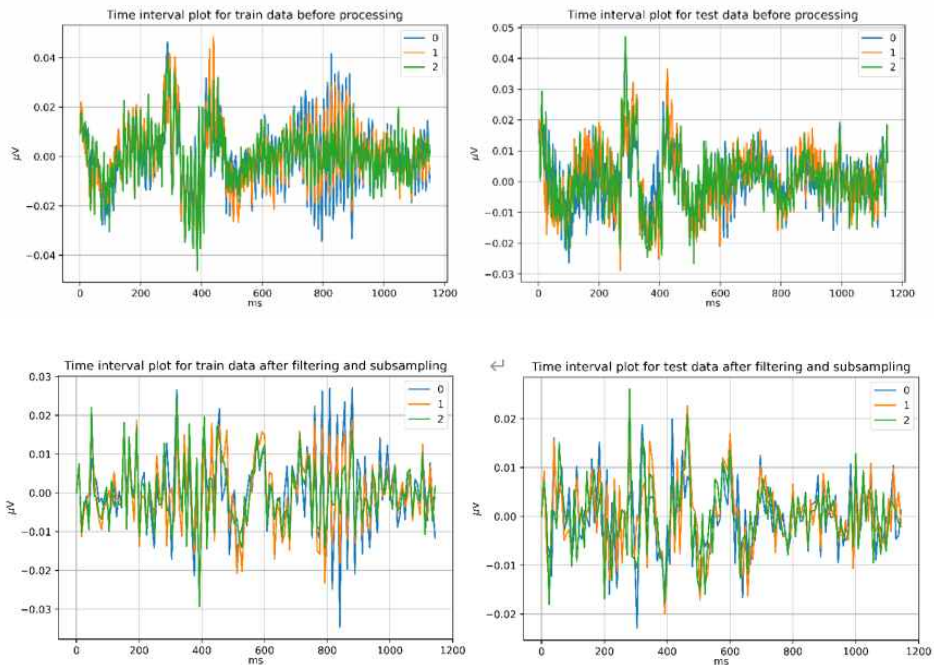
[Fig. 3] Visualization of Train and Test Data Before and After Signal Processing

3.1 Bagging

It is also referred to as bootstrap aggregation, aiming to reduce the variance of a decision tree. Various subsets are produced from the primary dataset among similar tuples, engaging observations with replacement. From each subset, a base model is deduced. Every model is studied in correspondence from every training set and is individualistic, and the final forecasts are produced by arranging the projections from all the models.

3.2 Boosting

It is an autodidactic procedure that determines by designating a weight for numerous items in the data. The boosting process commences by identifying identical weights. However, after each

model, each model has dispensed a weight depending on its presentation. It expects to reduce the inclination; that causes overfitting. Therefore, it is in the best interest to tune parameters to avoid overfitting. We have used Adaboost and XGBoost classifier for this experiment. In boosting, the model operates sequentially and is dependent on the previous models.

3.3 Stacking

In stacking, various layers of machine learning models stack upon each other, where each model passes the prediction to the model present above it. The training set follows the k-fold validation to build the final model. The indicators from all the models are utilized to anticipate the last forecast. The fundamental principle for selecting models should be based on a correlation on one another and the predictions. It utilizes the splitting strategy of stratified k-fold cross validation to evaluate model's prediction.

3.4 LDA

The simple and effective supervised method for binary classification where LDA offers predictions by assessing the probability that a different collection of inputs pertains to a specific class. The category that receives the most distinguished probability is the output, and a forecast is made. Further, it can also help in dimensionality reduction and visualization. It uses maximum likelihood rule and Bayesian rule to determine in which region x falls in.

3.5 XGBOOST

It is a scalable decision tree-based ensemble machine learning algorithm that effectively utilizes an optimized gradient boosting framework that helps to avoid overfitting or bias. XGBoost was engineered to contrive to compute time and memory resources. Furthermore, it can automatically supervise missing data values and promote the parallelization of tree construction. It can run on a cluster where it acknowledges dense and sparse matrix as the input and provide exceptional results on giving the best hyperparameters after tuning.

3.6 Logistic Regression

LR is the relevant regression analysis to administer when the dependent variable is binary and

is eventually an extension to the linear regression problem. The only thing that differentiates it from linear regression is the interpretation of weights; LR can handle probability scores, whereas linear regression values are above or below the threshold value. LR alters its yield using the strategic sigmoid capacity to yield a likelihood esteem, which would then be able to be mapped to at least two discrete classes.

3.7 SVM

SVM analyzes the coordinates by constructing a hypothetical hyperplane and tries to maximize the boundary of that hyperplane to build a clear fence for binary classification. It also helps up in introducing non-linearity to the existing data and helps in avoiding the issue of overfitting and supports a fast and dependable classification algorithm that performs exceptionally well with a small amount of data. It first transforms the data and sorts it into one of the two categories. It uses the subset of training points, thus consuming very little memory.

3.8 KNN

Based on the feature comparison procedure, KNN is a straightforward discrimination methodology. To predict the sample points classification, the figure of neighbours is the gist of the determinant. K is frequently an odd estimate in case of the presence of 2 classes. For getting most nearby neighbours, it receives the measure amongst points utilizing working distance measures such as Euclidean distance, Hamming distance, Manhattan distance, and Minkowski distance. Following determining the length, the last step is to vote for labels, which in our research model are binary.

3.9 Random Forest

Random forest, related to its alias, consists of an immense quantity of unique decision trees that present as an ensemble. All trees in the random forest assume a class prediction, and the class, including the total votes, enhances our model's prediction. A substantial mass of somewhat uncorrelated trees functioning as a committee will strike any individual constituent models. The coarse correlation among models is the core key. The preparation algorithm for this classifier exercises the standard bootstrap aggregating, to tree learners. It continually pulls a stochastic example with the replacement of the training set plus provides trees to individual

patterns. Random forest is an extension over bagging.

3.10 NAÏVE Bayes

Being a probabilistic classifier, it exercises Bayes theorem to pull out independent assumptions of the features, regardless of their correlations among the features. It is a secure and reliable algorithm to predict classes, especially in large datasets with independent features. In this experiment, as the values are continuous, the apt application of gaussian distribution has been implied. But in real-life cases, it is sporadic to find independent predictors.

3.10 Decision Tree

As a component of the supervised learning algorithm family, the Decision tree strives to produce a training pattern that can prognosticate the nature of the destination value by practicing simple decision rules induced from training data. The classification originates from the roots and travels forward by making a comparison with the corresponding records attribute. This comparison supports to understand the branches and likewise allows hopping to the next node. The hyperparameters should be controlled manually to avoid complexity. Decision trees are based upon divide and conquer algorithm that divides the datasets into repeatedly smaller datasets.

4. Thick Data Analytics

This research primarily focuses on predictive analytics by performing binary classification of motor imagery EEG signals. For effective analytics, we deduced the confusion matrix to provide meaningful statistics, which would help bridge the quantitative and qualitative analytics on the results available after exercising various operations on this thick dataset.

4.1 Confusion Matrix

A confusion matrix can be represented as 2-dimensional contingency table that compiles classification performance, as accuracy alone is not satisfactory to provide a quality approach[15]. It holds the amount of accurate and inaccurate predictions with count values slashed down in a class, further providing more insights about classification than accuracy alone. Four components of the confusion matrix are listed below:

[Fig. 4] Confusion Matrix

True Positive (TP): The resultant output is positive, and Forecast is positive.

False Negative (FN): The resultant output is negative, and Forecast is positive.

True Negative (TN): The resultant output is negative, and Forecast is negative.

False Positive (FP): The resultant output is positive, and Forecast is negative.

Based on the values obtained in the confusion matrix, we derive various statistical parameters such as accuracy, misclassification, sensitivity, specificity, precision, and f1 score[16].

4.1.1 Accuracy

Accuracy is an extent of precisely anticipated forecast of their absolute observations. The equation is outlined below:

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

4.1.2 Misclassification

Misclassification error refers to the number of records that belong to a category but are classified into different categories by the algorithm. Hence, it represents the negation of accuracy.

$$\text{Misclassification} = 1 - Accuracy$$

4.1.3 Sensitivity

Sensitivity, otherwise called True Positive Rate (TPR), is the quantity of right forecasts over the absolute number of positives.

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

4.1.4 Specificity

Specificity, otherwise called True Negative Rate (TNR), is the quantity of precise negative predictions over the all-out number of negatives.

$$\text{Specificity} = \frac{TN}{TN+FP}$$

4.1.5 Precision

Precision speaks to the all-out number of precise positive forecasts circulated by the complete number of positive expectations, thus known as Positive Predictive Value (PPV)

$$\text{Precision} = \frac{TP}{TP+FP}$$

4.1.6 F1 Score

F1 score utilizes both precision and sensitivity, thereby it speaks to the symphonious mean of precision and sensitivity, which further helps to demonstrate the balance between both.

$$\text{F1 Score} = 2\times\frac{\Pr ecision \times Recall}{\Pr ecision + Recall}$$

[Table 1] consists of the result observations of all the classifiers that have been deduced from this experiment. The best accuracy achieved was 79% by LDA and logistic regression, however logistic regression comes up with high f1 score of 79% whereas LDA comes up with f1 score of 67% which implies Logistic regression was competent of extracting a greater number of true positives.

[Table 1] Consists of the Result Observations of All the Classifiers

Statistical Parameters	LDA	Decision Tree	Random Forest	XGBoost	SVM	Logistic Regression	KNN	Naïve Bayes	Mean
Accuracy (%)	79	71	76	71	74	79	75	71	74.5
Misclassification (%)	21	29	24	29	26	21	25	29	25.5
Sensitivity (%)	77	81	73	83	80	77	69	59	74.88
Specificity (%)	81	60	80	60	67	81	81	83	74.13
Precision (%)	81	67	78	67	71	81	79	77	75.13

F1 Score (%)	67	74	76	74	75	79	73	67	73.13
MCC	0.586	0.424	0.515	0.44	0.475	0.558	0.504	0.427	0.491

[Table 2] consists of the performance measurements of all advanced ensembled classifiers used in this experiment. Out of which, stacking classifiers turn out to be the most effective one with the highest accuracy of 80% with an f1-score of 79%. The results for logistic regression and ensembled stacking classifiers were quite identical, whereas it outperforms LDA in aspects of the F1 score. Hence, logistic regression and stacking are the best performers when it comes to deriving correlations among predictors and calculating the best accuracy. However, it turns out that the accuracy mean of all ensembled classifiers is 77%, whereas the accuracy means of fundamental or traditional classifiers turn out to be 74.5%. The mean of f1 score in ensembled and fundamental classifiers turns out to be 73.125% and 75.33%. Similarly, all the ensembled classifiers can showcase the encouraging results and overcoming the gaps that can be observed in the fundamental classifiers.

[Table 2] Performance Metric of Ensembled Classifiers

| Statistical Parameters	Bagging	AdaBoost	Stacking	Mean
Accuracy (%)	75	76	80	77
Misclassification (%)	25	24	21	23.33
Sensitivity (%)	69	67	77	71
Specificity (%)	81	86	81	82.67
Precision (%)	79	82	81	80.67
F1 Score (%)	73	74	79	75.33
MCC	0.533	0.533	0.601	0.556

Hence, several parameters can be used to check for the robustness of the binary classification model, and accuracy is the simplest one among all. But the F1 score does play the role of data that is having the essence of class imbalance. [Fig. 6] displays the model comparison of all accuracies attained from all classifiers. Additionally, classifiers' performance majorly depends upon the tuning of their hyperparameters, for which we exercised grid-search to evaluate and display their best performing hyperparameters.

For better analytics, Area Under the Curve- Receiver Operating Characteristics (AUC- ROC) curve has been visualized, displaying how much each model can distinguish between classes. It

is one of the best performance measurements, which straightly implies that more the AUC, more the model can predict and discriminate between categories. The ROC bend is plotted with True Positive Rate (TPR) against False Positive Rate (FPR), where TPR is on the y-axis, and FPR is on the x-axis. Out of all classifiers, the best AUC score is obtained by ensembled stacking classifier, i.e. 0.813, which suggests a probability of 81.3% to discriminate among classes whereas LDA, the best performer among traditional classifiers got AUC of 0.812 with a probability of 81.2%. If we decrease the threshold, we will get more positive values, which increase TPR and FPR and vice-versa. Conclusively, the more the score is closer to 1, the more the model's probability of discriminating among classes for outstanding performance.
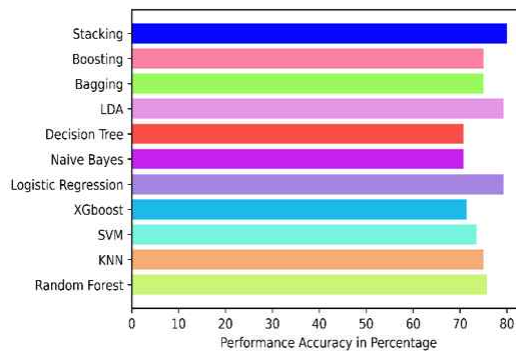
[Fig. 5] Visualization of Model Comparison
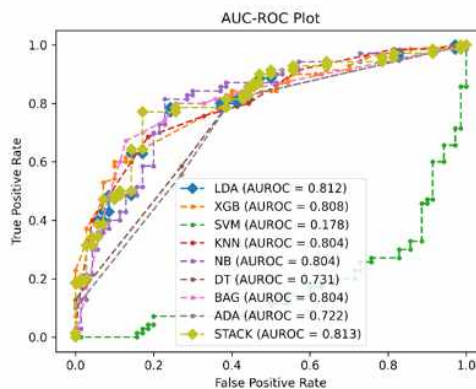

[Fig. 6] AUC-ROC Curve Plot

Another analytical measure that we have introduced for this experiment is MCC, a more reliable statistical rate that produces a high score in case of good predictions and is regularly viewed as a proportion of quality in case of occurrence of 2 categories. MCC score lies

between -1 to +1 in which +1 represents the perfect model, whereas -1 represents a poor model. MCC can be computed as:

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

It employs all 4 values of a confusion matrix, and a high value designates that both classes are predicted well. [Fig. 8] shows that the best MCC score has been deduced by ensembled stacking classifier, i.e. 0.601 among all ensembled techniques. On the other hand, LDA produced an MCC score of 0.586 among all traditional classifiers. Thereby, the ensembled stacking classifier outperforms all different classifiers to achieve the best results.
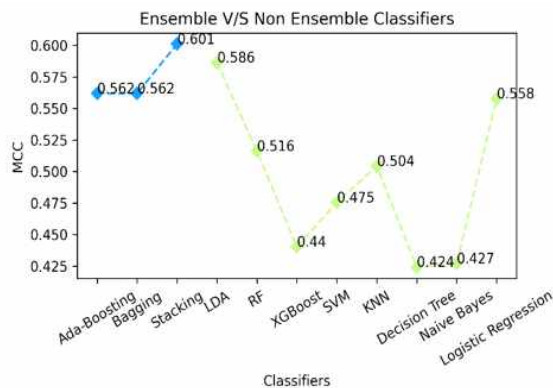
[Fig. 7] Mathews Correlation Coefficient Plot

5. Comparison to Other Related Research Work

Jason Adair et al. in their experiments used feature extractors like entropy, mutual information and maximum relevance minimal redundancy to extract the essential features and supplied it to classifiers like KNN and SVM to achieve the accuracy of 60-73% and other notable parameters that they deduced were the number of features and CVE[17]. Indu Dokare et al., in their research, exercised the wavelet transform methodology to extract features after pre-processing the data and decompose the signal into 6 wavelet coefficients. The decomposition was done based on frequencies, and SVM was invoked along with 5-cross validation on it to attain the accuracy metric for extracted wavelet coefficients and was able to accomplish the accuracy in between 67-71.42% only. Noting down the above researches, in our proposed system, we implemented the Butterworth band-pass filter in our intermediate pre-processing

stage. In this stage, the brain signals between the frequency 0.9 Hz high pass and 15 Hz low pass filter was processed to flatten the response. Further, in the same stage, CSP (Common Spatial Pattern) was utilized for dimensionality reduction of data, resulting in quick class differentiation. The accuracy obtained with KNN and SVM is 74% and 75%, respectively, with a low misclassification rate ranging between 24 – 26%.

J.Machado et al. performed their research on the same dataset and exercised signal processing for better distinction. Energy difference was considered as the primary feature, and based on that Priori probability was derived, which was further given to Naïve Bayes classifier. The experiment was able evaluated based on the average hit rate and was able to achieve an average of 80%[18].

A. Konar et al. used KNN in their experiment on the BCI dataset to distinguish between left and right movements. The results were evaluated on accuracy with and without using PCA. For data pre-processing, band-pass filter was used, after which they applied wavelet feature extraction techniques. This experiment achieved an accuracy of 75.71% without PCA and stretched it to 77.14% with the application of PCA and KNN[19]. All the previous research has mentioned the sole accuracy percentage obtained with several classifiers and has emphasized more on one parameter for judging classification. In this paper, we have described all the elements for a detailed performance metric computed to form a comparative study of all classifiers considered, such as LDA, Decision Tree, Random Forest, XGBoost, SVM, KNN, Logistic Regression, and Naïve Bayes. The statistical parameters evaluated for an exceptional case study are Accuracy, Misclassification, Precision, Sensitivity, Specificity and F1-Score. The mean accuracy obtained by combining all the accuracy is 74% and can be boosted by integrating feature extraction techniques and wavelet transformations with the above-mentioned pre-processing techniques.

Therefore, to conclude this section we would like to mention the techniques used to improve the system in comparison with the previous researches. In the data pre-processing stage, we have applied the band-pass filter and CSP to the data to reduce noise and to discover meaningful information out of thick dataset aiding the impactful analysis. On the other hand, the previous researches preferred applying CWT and PCA on the similar dataset due to its advantages such as dimensionality reduction and removal of corelated features. The previous researchers focused on evaluating the pre-defined classifiers on the dataset for achieving minimum misclassification error and decent accuracy score. In this proposed research we did the comparative analysis of existing classifiers with our proposed ensemble learning model and obtained a higher accuracy than all the previously tested models. The performance of ensemble

models seemed to be consistent throughout the analytics in comparison to the previously studied models. Discussing the post-processing of the model to validate the robustness, we have exercised non-typical parameters like MCC and model stability like AUC to achieve correlation coefficient and display ensemble models ROC respectively.

6. Conclusions

Ensemble classifiers embark on the performance by imitating human nature and adopting fusion methods, which mainly focuses on improving predictive accuracy before making any critical distinction. Classifiers like bagging and boosting utilizes weighting mechanism, whereas stacking employs meta-learning mechanism to stack up multiple classifiers and uses the outputs of previous classifiers as input to the other. In this experiment, the stacking not only performs the best among all ensemble classifiers in every aspect and performance metrics but also outperforms every different traditional classifier, especially in thick data. The peripheral of ensemble classifier revolves around to weigh individual models and combine them to improve predictive performance. Boosting provides sequential learning of the predictors, and classifiers like Adaboost have been used in this experiment in which the first one learns from the dataset, and the next one learns from the performance of the previous ones. On the contrary, stacking works utterly different from boosting and bagging. It performs two-level learning in which the first level is known as a base learner which different models learn from the dataset and the second level creates a new dataset that holds the output of each model and combines them up to build a stack model learner, which is why stacking is able to achieve accuracy of 80%, and MCC of 0.601 whereas bagging and boosting were not significant performers with an accuracy of 75% and 76%. However, this approach has a potential refinement and holds the capability of enhanced performance if employed with other feature extraction and selection techniques. PCA could be used to select the principal components and reduce the dimensionality of the dataset to be secure and robust.

7. Acknowledgments

References

[1] Siuly, Wang H., Zhang Y., Detection of motor imagery EEG signals employing Naïve Bayes based learning process, Measurement, **(2016)**, Vol.86, pp.148-158. DOI: https://doi.org/10.1016/j.measurement.2016.02.059

[2] Health Canada, Tracking Heart Disease and Stroke in Canada, https://www.phac-aspc.gc.ca/publicat/2009/cvd-avc/pdf/cvd-avs-2009-eng.pdf, Feb 1 **(2020)**.

[3] Jonathan R. Wolpaw, Niels Birbaumer, William J. Heetderks, Dennis J. McFarland, P. Hunter Peckham, Gerwin Schalk, Emanuel Donchin, Louis A. Quatrano, Charles J. Robinson, Theresa M. Vaughan, Brain‐Computer Interface Technology: A Review of the First International Meeting, IEEE Trans. on Rehabilitation Engg, **(2000)**, Vol.8, No.2, pp.164-173.

[4] Li Feng, He Fan, Wang Fei, Zhang Dengyong, Xia Yi, Li Xiaoyu, A Novel Simplified Convolutional Neural Network Classification Algorithm of Motor Imagery EEG Signals Based on Deep Learning, Applied Sciences, **(2020)**. Vol.10, No.5, https://doi.org/10.3390/app10051605

[5] Masoomi R, Khadem A, Enhancing LDA-based discrimination of left and right hand motor imagery: Outperforming the winner of BCI Competition II, 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI), IEEE, **(2015)**, November 5; Tehran, Iran.

[6] Suraj, P. Tiwari, S. Ghosh, R. K. Sinha, Classification of Two Class Motor Imagery Tasks Using Hybrid GA-PSO Based K - Means Clustering, Computational Intelligence and Neuroscience, **(2015)**, Vol.2015, https://doi.org/10.1155/2015/945729

[7] J. Fiaidhi, S. Mohammed, Thick Data: A New Qualitative Analytics for Identifying Customer Insights, IT Professional, **(2019)**, Vol.21, No.3, pp.4-13.

[8] J. Fiaidhi, Envisioning Insight-Driven Learning Based on Thick Data Analytics With Focus on Healthcare, in IEEE Access, **(2020)**, Vol.8, pp.114998-115004, doi: 10.1109/ACCESS.2020.2995763.

[9] S. Bhattacharyya, M. K. Mukul, Cepstrum Based Algorithm for Motor Imagery Classification, International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE), **(2016)**, September 22-23; Ghaziabad, India, doi: 10.1109/ICMETE.2016.140.

[10] Dokare I, Kant N, Classification of eeg signal for imagined left and right hand movement for brain computer interface applications, International Journal of Application or Innovation in Engineering & Management, **(2014)**, pp.291-294.

[11] BCI Competition II, http://www.bbci.de/competition/ii/, Jul 20 **(2020)**

[12] Venthur, B., Dähne, S., Höhne, J., Heller, H., Blankertz, B., Wyrm: A Brain-Computer Interface Toolbox in Python, Neuroinformatics, **(2015)**, Vol.13, No.4, pp.471-486, https://doi.org/10.1007/s12021-015-9271-8

[13] Xygonakis I, Athanasiou A, Pandria N, Kugiumtzis D, Bamidis PD, Decoding motor imagery through common spatial pattern filters at the EEG source space, Computational Intelligence and Neuroscience, **(2018)**, Vol.2018, https://doi.org/10.1155/2018/7957408

[14] Tarun Acharya, Advanced Ensemble Classifiers, https://towardsdatascience.com/advanced-ensemble-classifiers-8d7372e74e40, Aug 10 **(2020)**

[15] Jason Brownlee, What is a Confusion Matrix in Machine Learning, https://machinelearningmastery. com/confusion-matrix-machine-learning/, Aug 7 **(2020)**

[16] Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures, https://blog.exsilio.com/ all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/, Mar 21 **(2020)**

[17] Adair J, Brownlee AE, Ochoa G, Mutual information iterated local search: A wrapper-filter hybrid for feature selection in brain computer interfaces, International Conference on the Applications of Evolutionary Computation, **(2018)**, April 3-7; Parma, Italy

[18] J. Machado, A. Balbinot, A. Schuck, A study of the Naive Bayes classifier for analyzing imaginary movement EEG signals using the Periodogram as spectral estimator, 2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC), **(2013)**, February 18-20; Rio de Janerio, Brazil, pp.1-4, doi: 10.1109/BRC.2013.6487514.

[19] Khasnobish Anwesha, Bhattacharyya Saugat, Konar Amit, Tibarewala D. N., K-Nearest Neighbor Classification of Left-Right Limb Movement Using EEG Data, Academia, https://www.academia.edu/20886132/K_Nearest_neighbor_classification_of_left_right_limb_movement_using_EEG _data, April 30 **(2020)**