E-ISSN 2581 – 7957
P-ISSN 2277 – 3916
CVR Journal of Science and Technology, Volume 21, December 2021
DOI:10.32377/cvrjst 2107

Ensemble Classifier Model with Crow Search Feature Selection for Recognition of EEG Motor Imagery

V. Gokula Krishnan[1] and V. Divya[2]
[1]Assoc. Professor, CVR College of Engineering/CSIT Department, Hyderabad, India
Email: gokul_kris143@yahoo.com
[2]Research Scholar, Sathyabama Institute of Science and Technology/School of Electrical Engineering, Chennai, India
Email: vdivya6891@gmail.com

Abstract: **The patient with limb dyskinesia can do his daily activities and rehabilitation training with the help of Brain Computer Interaction (BCI) based on Electroencephalography (EEG). But the EEG feature extraction and classification faces the issues of low efficiency and accuracy due to large individual differences and low signal-to-noise ratio. To solve this problem, this paper proposes a recognition method of motor imagery EEG signal based on ensemble classifier. To increase the quality of EEG signal characteristic data, this method initially uses the short-time Fourier transform (STFT) and continuous Morlet wavelet transform (CMWT) by pre-processing the collected experimental datasets, which is based on time series characteristics. In order to achieve high quality features, ensemble classifier effectively recognize the EEG signals. The quality of EEG signal feature acquisition is improved by ensuring the high accuracy and precision of EEG signal recognition. Haralick features are extracted to improve the motor imagery, where Crow Search Algorithm (CSA) is used to select the optimal features. The experiments are carried out on laboratory measured data and BCI competition dataset. The results showed accuracy of this method for EEG signal recognition is better than existing techniques.**

Index Terms: **Brain computer interaction (BCI), short-time Fourier transform (STFT), continuous Morlet wavelet transform (CMWT), Crow search algorithm (CSA), Horlick features and Ensemble classifier.**

I. Introduction

BCI is a communication control system directly established between the brain and external devices (computers or other electronic devices), using signals generated during brain activity [1]. Instead of relying on the muscles and organs, the system directly builds communication between the brain and the machine. Electroencephalogram (EEG) is one of the most common signals used for building a BCI system because of its cost-effectiveness, non-invasive implementation, and portability. BCIs have shown potential in applying various fields such as communication, control, and rehabilitation [2]. Recent years have witnessed intense research into different types of BCI systems. According to the signal acquisition method, BCI technology can be divided into three types: non-implantable system, semi-implantable system, And an implantable system. Non-implantable BCI systems mainly use EEG to recognize human's intention. According to the signal generation mechanism, BCI systems can be divided into induced BCI systems and spontaneous BCI systems. The induced BCI systems are: steady-state visual evoked potentials (SSVEP) [3-4], slow cortical potentials [5], and the P300 [6-8] and the spontaneous BCI systems are: motor imagery (MI) [9-12].

Derived from this type of data, the curse of dimensionality problem [13] is usually present, as the features vastly outnumber the observations. In the particular case of this work, machine learning algorithms may lose the ability to generalize knowledge. A possible solution is Feature Selection (FS), which brings several benefits: noise and redundancy removal, reduced computational costs, and improved classification accuracy. FS is often highlighted in the existing literature on BCI applications [14], citing its importance in real-time performance or the understanding of the brain, among other benefits. However, FS is an NP-hard problem [15], which renders brute-force approaches unfeasible due to the size of the search space. The three main types of alternative methods are filter, wrapper and embedded [16]. Filter methods measure the relationship between features and the dependent class variable. Wrapper methods evaluate the performance of a classifier using different feature subsets. Embedded methods integrate FS into the classifier. The advantage of filters lies in their lower computational complexity, whereas wrapper and embedded approaches frequently achieve better results. In this paper, a wrapper method based on a Genetic Algorithm (GA) is employed. GAs are popular for BCI tasks, be it for FS [17–19] or other purposes [20].

Neural networks are a promising alternative to address the complexity of BCI data, since they are universal approximators and thus they can represent a wide variety of continuous functions. Besides standard Feed-Forward Neural Networks (FFNNs), which are the simplest kind of neural network in terms of structural design, there is growing interest in architectures that are able to leverage context. Convolutional Neural Networks (CNNs) extract local patterns through the convolution operator and have been successfully applied to EEG signals [21- 22]. Recurrent Neural Networks (RNNs) [23-24], which are not as widespread as CNNs yet, can dynamically store context to improve processing of individual bits of data. In EEG-based Motor Imagery (MI), many machine learning algorithms and feature extraction methods have been studied to try to overcome the limitations of small dataset and poor signal-to-noise ratios. The Support Vector Machine (SVM), despite its age, can still produce promising results when

Received on 16-06-2021, Revised on 22-07-2021, Accepted on 24-08-2021.

paired with the right features: in [25], mutual information is calculated from Common Spatial Pattern (CSP) features to select optimal frequency bands, and dimensionality is further reduced by means of Linear Discriminant Analysis (LDA) before finally classifying the patterns with SVM; in a later study [26], the same authors use LDA for spatial filtering and a Long Short-Term Memory (LSTM) network for temporal filtering before classifying again with SVM. Alternative approaches to classic machine learning also exist, such as classification by Riemannian geometry [27] or by a residual norm-based strategy [28].

II. Related Works

Saa and Cetin [29] have proposed the Filter Bank Common Spatial Pattern (FBCSP). After filtering the original EEG signal with a set of filters, the CSP method is used to extract features on each filtered frequency band. Finally, the feature selection algorithm is applied on the basis of the extracted features. The above work research has improved the accuracy of some motor imaging tasks to a certain extent. However, since it takes a lot of time to select the characteristics of the experimental data set for each experimental object, it is not universal. Suk and Lee [30] proposed a technique for selecting class and subject specific frequency bands on the basis of the analysis of channel frequency map, which is a channel-frequency matrix. They individually performed spatial filtering, feature extraction and classification for each frequency band. Thus, they considered the cumulated outputs at the end to feed as an input EEG. They found their proposed technique to surpass the performance of common spatial pattern (CSP) in terms of session-to-session transfer rate and cross-validation. For certain subjects, they found increased classification accuracy.

Wei et al., [31] have used EEG emotion data set SEED for emotion recognition research. The abstract features of EEG samples are automatically extracted based on the convolutional neural network in deep learning, eliminating the need for manual feature selection and dimensionality reduction. And with the most advanced methods at present, a considerable accuracy rate has been achieved. Liang et al., [32] have developed an online sequential algorithm called OS- ELM (online sequential extreme learning machine) for SLFNs with additive or radial basis function (RBF) hidden nodes in a consolidated framework. It has been designed to learn data either chunk- by- chunk or one-by-one. A detailed performance comparison of the proposed method was done with other renowned sequential learning algorithms on standard issues related to time series prediction, classification and regression. The experimental results proved OS-ELM to be faster and generated better generalization performance than its counterparts. Zhang et al., [33] an example of this, but the method was not robust to changes in data pre-processing in preliminary experiments conducted by the authors of this manuscript. In particular, their model appears to over fit to data which was overlapping between training and test sets, resulting in astoundingly high accuracies (98%). With clearer separation between the training and testing sets, we found their model to achieve approximately 36% accuracy on the four-class problem.

III. Proposed System

In this section, the explanation of proposed method is given briefly. Initially, the publicly available datasets are considered for this research work. In pre-processing, a filter is used to remove unwanted signals and two transformation models are included. Then, these outputs are given as input to Haralick feature extraction technique for relevant feature extraction. The optimal features are selected by using crow-search algorithm and finally, these optimal features are given as input to ensemble classifier for final classification. Figure 1 describes the working flow of proposed methodology.

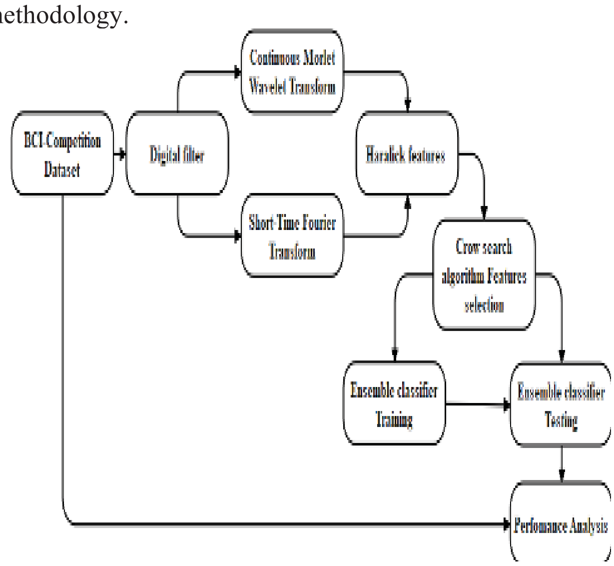

Figure 1. Proposed System model.

A. Dataset Description

Dataset 1: The first dataset was from BCI-Competition-III-IVa and was collected in a cue-based setting. Only cues for the classes "right" and "foot" are provided. This dataset was recorded from five healthy subjects (aa, al, av, aw and ay) at 100 Hz. The subjects sat in a comfortable chair with arms resting on armrests. The raw data were continuous signals of 118 EEG channels and markers that indicate the time points of 280 cues. Each sample was segmented from [0, 2.5] s by marks, then passed a band-pass filter (5-order Butterworth digital filter with cut-off frequencies at [8, 30] Hz) to remove muscle artifacts, line-noise contamination, and DC drifts. Under the condition that the positive sample and the negative sample were balanced, 100 samples were randomly selected as the training pool. The remaining samples were used as the test samples.

Dataset 2: The second dataset was from BCI-Competition-IV-1. The dataset was recorded from seven healthy subjects (a, b, c, d, e, f and g), including four healthy individuals (named "a," "b," "f," "g") and three artificially generated "participants" (named "c," "d," "e"). 59-channel EEG signals were recorded at 100 Hz. Two motor imagery classes were selected for each subject from the three classes: left hand, right hand, and foot. There were two subjects (a, f) whose motor imaging tasks were different from the others, so they were eliminated. Here we only used the calibration data because of the complete marker information. Each sample was segmented from [0, 2.5] s by marks, then passed a band-pass filter (5-order Butterworth digital filter with cut-off frequencies at [8, 30] Hz) to remove muscle artifacts, line-noise contamination, and DC drifts. After pre-processing, we obtained 200 samples for each subject. In this proposed system subject "g" is not considered. We randomly selected 100 samples as a training pool and the rest as test samples, like dataset 1.

B. Feature Extraction

These two phenomena are an important basis for distinguishing different types of EEG signals. Therefore, the time-frequency domain analysis method combines the two methods as one for the effective analysis methods [34], where the two methods includes Short-Time Fourier Transform and Continuous Morlet Wavelet Transform that are briefly discussed in the below section.

C. Haralick feature extraction

The extraction of distinguishing features from the given EEG data is carried out by Haralick features for better classification. According to the spatial dependencies of gray tone images, the quantifiable textural properties are obtained from the CMWT and STFT, which is utilized as statistical measures of textures. Haralick characteristics produce accurate extraction that is resilient in the various arm and leg movements. The transformation technique bidimensionally represents the spatial gray level distribution. After creating the co-occurrence matrices, the Haralick features may be extracted from the gray level histogram. The final feature vector is produced by the estimated Haralick features.

D. Short-Time Fourier Transform

The short-time Fourier transform first divides the entire time series into several time segments of equal length. Then calculate the frequency spectrum information in each time segment by Fourier transform. Obtain the change of each frequency component with respect to time from the surface. The calculation formula is as follows:

$$s(f,k) = \sum_{n-0}^{N-1} s(n)[W(n-k)e^{\frac{f2\pi fn}{-N}}] \tag{1}$$

Where, $S(n)$ represents the time series of EEG signals. $W(n)$ represents window function. N represents the number of time points recorded. k represents the index of different time windows. f represents the frequency component in the signal. n represents the time point. The length of the time window required to be divided in the formula is the same, which determines that the algorithm performs well when measuring high-frequency components. When measuring low-frequency components, it is often accompanied by distortion.

In order to effectively measure the change trend of the μ rhythm and β rhythm in the signal, this paper selects the time-frequency matrix obtained by the time window of $0.5s$ and the hamming window function. Combine the time-frequency matrices on the two channels C3 and C4. A three-dimensional tensor with a size of 33∗35∗2 is obtained as the input of the subsequent convolutional neural network. From the output matrix the 32 features are extracted such as autocorrelation, contrast, correlation, correlation, cluster output, cluster hue, difference, energy entropy, uniformity, uniformity, extreme likelihood, number of cubes, variance, normal sum, sum change, entropy of a sum, difference fluctuation, entropy conversion are highlighted. the evidence of correlation1, the informational proportion of correlation2, the inverse difference, the standardized inverse contrast and the normalized inverse of the change

E. Continuous Morlet Wavelet Transform

The Morlet wavelet transform uses a wavelet of finite length and attenuation as the base to measure the intensity of each rate component in the signal over time. The formula is as follows:

$$W(a,b) = \int_{-\infty}^{\infty} x(x)\frac{1}{\sqrt{a}}\psi(\frac{t-b}{a})dt \tag{2}$$

Where, $x(t)$ represents the signal sequence. $\psi(t)$ represents the wavelet basis.t represents the time point. The parameter a controls the scaling of the wavelet function. When a takes a value from small to large, the wavelet function gradually widens, so the low-frequency components can be better measured. And by adjusting the parameter b, the shift of the wavelet function is controlled to obtain the intensity information of each frequency band at different time domain positions. The calculation formulas of Morlet wavelet center time and time domain span are as follows:

$$\psi(x) = e^{-x^2}\cos(\pi\sqrt{\frac{2}{In\,2}}\,x) \tag{3}$$

$$t_0 = \frac{\int_{-\infty}^{\infty} t|\psi(t)|^2 dt}{\int_{-\infty}^{\infty} |\psi(t)|^2 dt} \tag{4}$$

$$\Delta t_\psi = \sqrt{\frac{\int_{-\infty}^{\infty}(t-t_0)^2|\psi(t)|^2 dt}{\int_{-\infty}^{\infty}|\psi(t)|^2 dt}} \tag{5}$$

The calculation formula for center frequency and bandwidth is as follows:

$$\omega_0 = \sqrt{\frac{\int_{-\infty}^{\infty}\omega|\psi(\omega)|^2 d\omega}{\int_{-\infty}^{\infty}|\psi(\omega)|^2 d\omega}} \tag{6}$$

$$\Delta\omega_\psi = \sqrt{\frac{\int_{-\infty}^{\infty}(\omega-\omega_0)^2|\psi(\omega)|^2 d\omega}{\int_{-\infty}^{\infty}|\psi(\omega)|^2 d\omega}} \tag{7}$$

Where, $\Psi(\omega)$ is the frequency component information obtained after $\psi(t)$ undergoes Fourier transform. It can be known from the above formula that when the wavelet transform measures high frequency components, because the wavelet used is narrow, a smaller time domain span can be obtained, but the frequency domain span will be enlarged accordingly. Therefore, in the output time-frequency matrix, the resolution of the frequency dimension of the high frequency part is relatively low, and the low frequency part is just the opposite. Similarly, the C3 and C4 channel position information are integrated, and a sample matrix of size 35×1152×2 is obtained as the input of the neural network. From the output matrix the 32 features are extracted. Which is mentioned in the above SIFT section. The extracted SIFT Horlick features and CMWT features are fused and are given to the feature selection.

F. Feature Selection using Crow search algorithm (CSA)

Naturally, every creature has its own behaviours and character. Each creature would have certain patterns of activities to fulfil its needs like food collection, reproduction and so on. Similarly, observing the characteristics and behaviours of crow reveals notable information [35-38]. Usually, crows live as a group also called flock. The study depicts that the crows [39] are very clever in nature and have strong memory power. A crow would remember the faces of the other crows as well as its hide spots (food sources) for a long period of time. It has its own way of communication techniques to share information with each mate. It also has the habit of learning from mate's by monitoring them. By this way, it identifies the mates hidden food sources and takes the Food when the owner is not present in its spot. The crows involved in this theft activity are always conscious of the other mates, and they often change their locations to avoid being a future sufferer.

The basic terminologies used in CSA are presented in this part. The current iteration is denoted as $k\,(k = 1,2,3\,\ldots.max\_iteration)$. The maximum number of iterations is denoted as $max\_iteration$. N denotes the flock size. There are two crows i and j, and their positions are shown in Fig. 2.

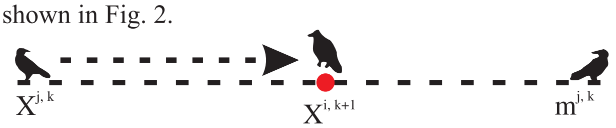

Figure 2. Crow position movement in CSA

The position of the crow i during the iteration k is denoted as $x^{i,k}(k = 1,2,3,\ldots,N)$. Each crow maintains the present best location in memory $m^{i,k}$. So, the crow would change its location to find better food spots than it has right now. Let a scenario while crow i decides to change its position, and at the same time a crow j flies be assumed. Crow i decides to follow the crow j to predict the hidden food source of crow j. The new position identification process has two possible cases.

Case 1: If crow j is not aware of the follower crow i, then the new position of the crow i will be obtained using Eq. (8).

$$x^{i,k+1} = x^{i,k} + r_i \times fl^{i,k} \times (m^{j,k} - x^{i,k}) \tag{8}$$

where r_i is a random value generated between 0 and 1. Based on the flight length (fl), local or global search is performed. If lower value is chosen for $fl\,(< 1)$, then the search will be performed in the local domain. When the higher value is chosen for fl it would lead the search in the global domain.

Case 2: If crow j is identified to be spying the activity of crow i, then, crow j will divert the path. In this way, crow j safeguards its food source. In this case, crow i would take a random position. The above two cases are denoted in Eq. (9)

$$x^{i,k+1} = \begin{cases} x^{i,k} + r_i \times fl^{i,k} \times (m^{j,k} - x^{i,k}) & r_j \geq AP^{i,k} \\ a\ random\ position & otherwise \end{cases} \tag{9}$$

Awareness probability [39] is denoted as AP which is used to minimize the imbalance between intensification and diversification. The flow chart of the CSA is defined in the below figure 3.

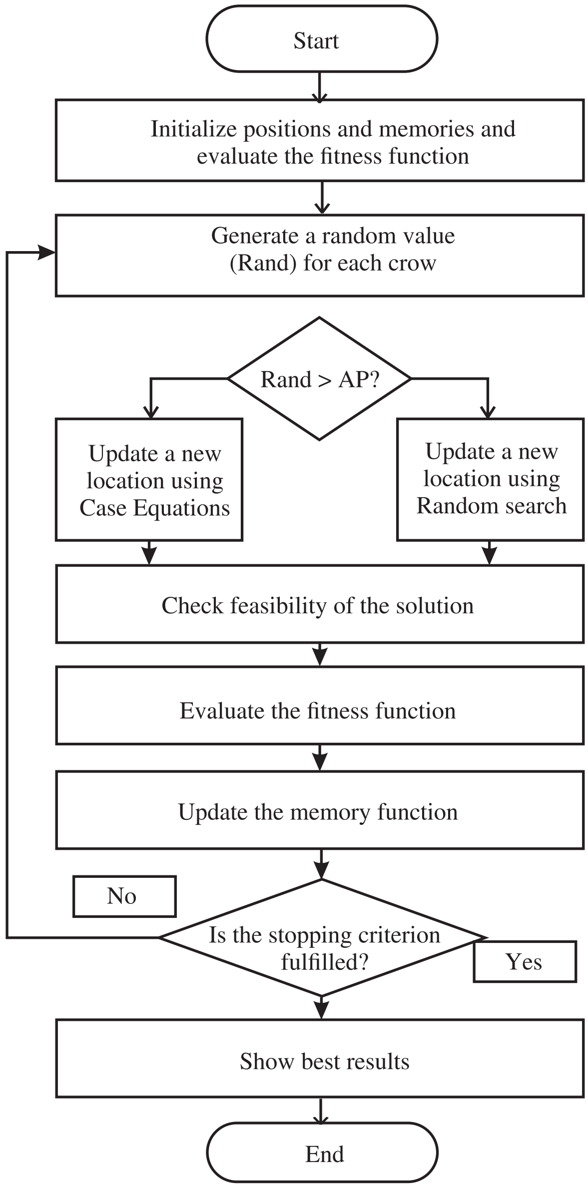

Figure 3. Flow chart of the Crow search algorithm.

IV. Classification

Supervised machine learning approaches are used for the classification of Motor Imagery. There are various steps involved in the supervised learning approaches. For classification purposes, many classifiers have been used in the method. Some commonly used classification methods are Artificial Neural Networks (ANN), Bayesian classification, K-Nearest Neighbor (KNN) classifiers, and Support Vector Machine (SVM).The classification might provide the solution whether the underwater images are consist of metal or not.

A. KNN Algorithm

KNN is a simple algorithm that stores all available cases and classifies new classes based on a similarity measure (e.g. distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970's as a non-parametric technique. The algorithm works as follows. A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbor measured by a distance function. If $K = 1$, then the case is simply assigned to the class of its nearest neighbour. There are three distances used for continuous variables: Euclidean (10), Manhattan (11), Minkowski (12).

$$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \quad (10)$$

$$\sum_{i=1}^{k}|x_i - y_i| \quad (11)$$

$$\left(\sum_{i=1}^{k}(|x_i - y_i|)^2\right)^{1/q} \quad (12)$$

$$D_H = \sum_{i=1}^{k}|x_i - y_i|^2 \quad (13)$$

In instance of categorical variables, Hamming distance (5) must be used. It also brings up the issue of standardization of the numerical variables between 0 and 1 when there is a mixture of numerical and categorical variables in the dataset. Choosing the optimal value for K is best done by first inspecting the data. In general, a large K value is more precise as it reduces the overall noise but there is no guarantee. Cross validation is another way to retrospectively determine a good K between 3-10.

B. SVM Algorithm

SVM is a popular classification technique discovered in 1995. A classification task usually involves separating data into training and testing sets. Each instance in the training set contains one "target value" and several "attributes". The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes.

Given a training set of instance-label pairs$(x_i, y_i), i = 1, \dots, l$ where $x_i \in R^n$and $y \in \{1-1\}^l$, the SVM require the solution of the following optimization problem:

$$\min_{w,b,\xi} \frac{1}{2}w^T w + C\sum_{l=1}^{l}\xi_i$$

$$subject\ to\ {}^{y_i(w^T\phi(x_i)+b)\geq 1-\xi_i}_{\xi_i \geq 0} \quad (14)$$

Here training vectors $x_i$are mapped into a higher (maybe infinite) dimensional space by the function ϕ. SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. $C > 0$ is the penalty parameter of the error term. Furthermore, $K(x_i, x_j) \equiv \emptyset(x_i)^T\phi\,(x_i)$is called the kernel function. The following are four basic kernels:

Linear:

$$K(x_i, x_j) = x_i^T x_j \quad (15)$$

The algorithm requires that each data instance is represented as a vector of real numbers. Hence, if there are categorical attributes, it has to be first converted into numeric data. Another thing to do before applying SVM is performing scaling. The main advantage of scaling is to avoid attributes in greater numeric ranges dominating those in smaller numeric ranges. Another advantage is to avoid numerical difficulties during the calculation.

C. ANN Algorithm

Generally speaking, ANN can be considered as an information processing system which is composed of a network of interconnected simple processing elements, i.e. neurons. Determined by the connections between these neurons and the associated parameters, ANN can exhibit complex global behaviour to generate expected outputs via supervised or unsupervised learning. Inspired by the biological nervous system, the learning process is to adjust the connection strength or weights between the neurons. Each neuron forms a node in the whole network and after training each node is assigned with a determined bias or threshold. For each interconnection between two nodes, a weight is also assigned to represent the link-strength between the neurons.

Let $x = (x1, x2, \dots, xd)^T$ be an input vector and $w = (w1, w2, \dots, wd)^T$ the weight vector, the output of a single neuron z.

$$Z = g(w^T X - b) = g(\textstyle\sum_{i=1}^{d} w_i X_i - b) \quad (16)$$

where$g()$ is namely an activation function to decide whether the perceptron should fire or not. The sigmoid function $Sig(x) = (1 + ex)\,1$ is the most popular used activation function, others include $tanh$ and step functions, etc. Using the same process as to determine the output of a single neuron, the output of the whole network can be also calculated in a topological manner. This means that for each neuron its inputs from other neurons need to be computed before determining its output. As seen, the weight vector and the bias associated to each connection and each node will influence the outputted results, and they can be determined in training or learning process as follows. First of all, the topology of the ANN needs to be specified, and feedforward ANN is adopted as it has been widely applied for the classification of MCCs. A feed-forward ANN is a multilayer perceptron (MLP) which contains three or more layers of neurons, i.e. one input layer, one output layer and at least one hidden layer. With a given training set, a specified activation function and a learning ratio c where $c\ 2\ (0,1)$, the learning process for supervised training using the well-known back-propagation algorithm.

V. Result and Discussion

The proposed system is experimented using MATLAB (version 2018a) with 3.0 GHz Intel i5 processor with 1TB hard disc and 8 GB RAM.

A. Evaluation Metrics

The challenge evaluation metrics are used for evaluating the both segmentation and classification performance of our method. For the segmentation, the evaluation criteria include sensitivity (SE), specificity (SP), accuracy (AC), Recall (R) and Precision (P). The performance criteria are defined is as:

$$SE = \frac{tp}{tp + fn} \quad (17)$$

$$AC = \frac{tp + tn}{tp + fp + tn + fn} \quad (18)$$

$$SP = \frac{tn}{tn + fp} \quad (19)$$

Where tp, tn, fp and fn denote the number of a true positive, true negative, false positive and false negative. As for the classification, there are four evaluation criteria, including sensitivity (SE), specificity (SP) and accuracy (AC).

B. Performances analysis

The performance of the proposed system has validated in various ways such, which are discussed in the below section, where two datasets are divided into five and six subjects, which is given as follows: Dataset 1 (subjects such us, aa, al, av, aw, ay) and dataset 2 (subjects such us, a, b, c, d, e, f, g). Initially, table 1 describes two subjects of dataset 1.

TABLE I.
CLASSIFICATION PERFORMANCE FOR aa AND al.

	Dataset 1 Subject aa			Dataset 1 Subject al		
Method	SE (%)	SP (%)	ACC (%)	SE (%)	SP (%)	ACC (%)
Without Feature extraction ANN	60.33	81.33	60	58.44	79.67	57
With Feature extraction ANN	63.51	84.33	66.89	77.23	80.11	63.40
With CSA selection Proposed ANN	69.67	80.66	75.99	69.72	75.64	68.87
Without Feature extraction Features KNN	59.41	79.13	61.12	63.90	70.14	59.41
With Feature extraction KNN	62.60	76.46	67.90	69.62	84.33	64.19
With CSA selection KNN	65.58	77.76	76.31	71.25	75.17	70.21
Without Feature extraction Features SVM	67.26	78.04	63.09	59.13	58.15	60.45
With Feature extraction SVM	68.15	78.27	78.62	80.32	73.06	67.34
With CSA selection SVM	69.62	79.67	89.69	62.60	80.66	78.35

Here, ensemble classifiers such as ANN, KNN and SVM are tested with feature extraction and without feature extraction as well as for feature selection. The two subjects such as aa and al for dataset 1 provides better accuracy, while implementing with feature extraction as well as feature selection. For instance, KNN achieved only 61.12% without feature extraction, where the same technique achieved 76.31% of accuracy, when implementing with CSA selection algorithm. In addition, SVM achieved better accuracy than other two classifiers, for sample, SVM with feature extraction achieved 78.62% of accuracy, where ANN and KNN with feature extraction achieved nearly 66% to 67% of accuracy. Table 2 and 3 explains the proposed method's performance for remaining subjects of dataset 1. Figure 4 depicts the average classification accuracy for all subjects of dataset 1.

TABLE II.
CLASSIFICATION PERFORMANCE FOR av AND aw.

	Dataset 1 Subject av			Dataset 1 Subject aw		
Method	SE (%)	SP (%)	ACC (%)	SE (%)	SP (%)	ACC (%)
Without Feature extraction ANN	62.54	84.33	63	54.33	88.33	68
With Feature extraction ANN	72.19	70.23	66.34	69.43	70.51	70.94
With CSA selection Proposed ANN	62.60	79.67	79.86	63.51	79.67	73.67
Without Feature extraction Features KNN	63.24	80.19	69.26	69.58	71.25	70.61
With Feature extraction KNN	65.86	80.43	73.87	69.62	71.33	72.97
With CSA selection KNN	69.30	81.24	77.45	79.15	81.60	76.09
Without Feature extraction Features SVM	59.98	84.33	66.51	59.44	75.29	68.67
With Feature extraction SVM	69.62	79.67	75.66	76.35	72.18	74.01
With CSA selection SVM	70.26	76.53	86.46	81.27	80.75	83.12

TABLE III.
CLASSIFICATION PERFORMANCE FOR *ay*.

	Dataset 1 Subject ay		
Method	SE (%)	SP (%)	ACC (%)
Without Feature extraction ANN	50.33	90.33	65
With Feature extraction ANN	59.41	69.26	72.45
With CSA selection Proposed ANN	69.62	84.33	76.71
Without Feature extraction Features KNN	58.54	63.47	67.40
With Feature extraction KNN	68.25	70.45	71.78
With CSA selection KNN	63.51	79.67	74.89
Without Feature extraction Features SVM	61.27	67.34	70.51
With Feature extraction SVM	69.04	76.38	77.45
With CSA selection SVM	69.62	79.67	81.56

CVR Journal of Science and Technology, Volume 21, December 2021
DOI:10.32377/cvrjst 2107

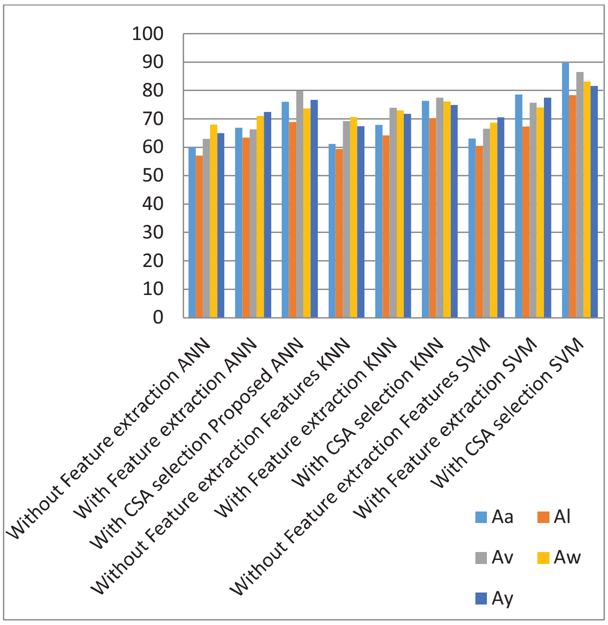

Figure 4. Average classification Accuracy for dataset-1.

While testing the ensemble classifiers, SVM with CSA selection algorithm achieved better accuracy for other three subjects. For instance, SVM with CSA achieved 86.46% of accuracy for av subjects, 83.12% of accuracy for aw subjects and 81.56% of accuracy for ay subjects. But, the KNN with CSA achieved 77.45% of accuracy for av subjects, 76.09% of accuracy for aw subjects and 74.89% of accuracy for ay subjects. The other set of experiments i.e. Table 4 shows the performances of the proposed methods for two subjects of dataset 2 are depicted as follows:

TABLE IV.
CLASSIFICATION PERFORMANCE FOR a AND b.

	Dataset 2 Subject a			Dataset 2 Subject b		
Method	**SE (%)**	**SP (%)**	**ACC (%)**	**SE (%)**	**SP (%)**	**ACC (%)**
Without Feature extraction ANN	51.33	57.33	60	43.68	77.39	59
With Feature extraction ANN	59.41	60.12	62.54	63.51	65.20	60.39
With CSA selection Proposed ANN	63.76	71.43	73.54	63.51	72.26	74.54
Without Feature extraction Features KNN	62.60	79.67	67.32	68.16	69.30	71.02
With Feature extraction KNN	62.60	84.33	70.51	62.60	72.70	76.32
With CSA selection KNN	78.24	77.45	81.05	63.51	67.69	77.45
Without Feature extraction Features SVM	63.51	79.67	75.08	68.01	68.34	70.25
With Feature extraction SVM	62.40	69.04	72.18	69.62	80.66	82.14
With CSA selection SVM	66.67	79.62	82.45	78.03	82.57	85.05

Various techniques are tested with ANN and results are tabulated, which shows that ANN with feature extraction achieved 60.12% of SP and 62.54% of ACC, where ANN without feature extraction achieved 57.33% of SP and 60% of ACC. But, while implementing with CSA selection technique, ANN achieved 71.43% of SP and 73.54% of ACC for a subjects. This proves that CSA provides better performance for ensemble classifiers. In addition, KNN with CSA achieved 81.05% of ACC for a subjects and 77.45% of ACC for b subject. However, SVM achieved better performance for two subjects than other ensemble classifiers. For sample, SVM with CSA achieved 82.45% of ACC for a subject and 85.05% of ACC for b subject. Table 5 provides the results for subject c and d in terms of ACC, SE and SP.

TABLE V.
CLASSIFICATION PERFORMANCE FOR c AND d.

	Dataset 2 Subject c			Dataset 2 Subject d		
Method	**SE (%)**	**SP (%)**	**ACC (%)**	**SE (%)**	**SP (%)**	**ACC (%)**
Without Feature extraction ANN	62.01	79.67	69.04	63.51	69.64	70.13
With Feature extraction ANN	69.33	86.33	80.3	72.25	83.24	81.00
With CSA selection Proposed ANN	62.60	70.24	77.45	69.62	74.85	81.58
Without Feature extraction Features KNN	63.98	64.33	78.05	70.14	75.47	77.59
With Feature extraction KNN	65.70	75.16	79.60	69.16	72.35	79.04
With CSA selection KNN	72.06	81.25	84.31	73.64	79.18	82.05
Without Feature extraction Features SVM	69.62	79.67	80.37	59.98	84.33	66.51
With Feature extraction SVM	68.10	75.25	80.86	69.62	79.67	75.66
With CSA selection SVM	70.62	79.36	82.56	71.03	75.24	83.46

KNN with feature extraction achieved 75.16% of SP and 79.60% of ACC, where KNN without feature extraction achieved 64.33% of SP and 78.05% of ACC. But, while implementing with CSA selection technique, KNN achieved 81.25% of SP and 84.31% of ACC for a subjects. This proves that CSA provides better performance for ensemble classifiers. In addition, ANN with CSA achieved 77.45% of ACC for a subject and 81.58% of ACC for b subject. However, SVM achieved better performance for two subjects than other ensemble classifiers. For sample, SVM with CSA achieved 82.56% of ACC for a subject and

83.46% of ACC for b subject. Table 6 describes the classification performance of ensemble classifiers for remaining subjects of dataset 2, where Fig. 5 shows the average classification accuracy for all subjects of dataset 2.

TABLE VI.
CLASSIFICATION PERFORMANCE FOR *e* AND *f*.

	Dataset 2 Subject e			Dataset 2 Subject f		
Method	**SE (%)**	**SP (%)**	**ACC (%)**	**SE (%)**	**SP (%)**	**ACC (%)**
Without Feature extraction ANN	62.99	76.24	70.05	59.56	80.39	61
With Feature extraction ANN	63.04	78.15	72.45	63.51	72.85	76.32
With CSA selection Proposed ANN	65.38	80.25	76.32	62.99	79.67	80.16
Without Feature extraction KNN	79.67	84.33	76.66	65.28	68.10	72.38
With Feature extraction KNN	63.51	69.15	72.16	67.20	73.04	79.12
With CSA selection KNN	70.14	75.85	82.09	69.62	73.07	82.47
Without Feature extraction Features SVM	69.01	72.47	70.43	69.04	72.45	71.32
With Feature extraction SVM	65.34	74.08	73.14	69.95	84.33	77.45
With CSA selection SVM	69.62	79.67	85.45	72.01	78.04	84.32

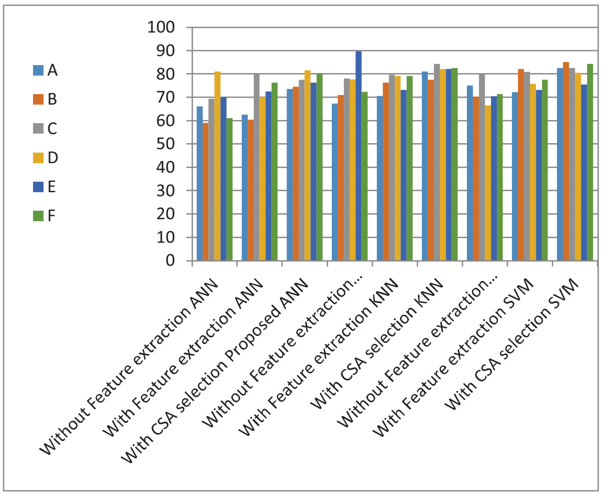

Figure 5. Average classification Accuracy for dataset-2.

Here, ensemble classifiers such as ANN, KNN and SVM are tested with feature extraction and without feature extraction as well as for feature selection. The two subjects such as e and f for dataset 2 provides better accuracy, while implementing with feature extraction as well as feature selection. For instance, KNN achieved only 76.66% of ACC without feature extraction, where the same technique achieved 82.09% of accuracy, while implementing with CSA selection algorithm for e subjects. In addition, SVM achieved better accuracy than other two classifiers, for example, SVM with feature extraction achieved 73.14% of accuracy, where ANN and KNN with feature extraction achieved nearly 72% of accuracy for e subjects. While testing the ensemble classifiers, SVM with CSA selection algorithm achieved better accuracy for other three subjects. For instance, SVM with CSA achieved 85.45% of accuracy for e subjects, 84.32% of accuracy for f subjects. But, the KNN with CSA achieved 82.09% of accuracy for e subjects and 82.47% of accuracy for f subjects.

VI. CONCLUSIONS

In summation, several interesting results regarding EEG motor imagery classification protocols is presented. Firstly, while the compared ANN, Bayesian classification, KNN classifiers networks may improve MI classification in some specific applications, the implementation of the classifier is still the most reliable of the three. Secondly, the efficacy of STFT and CMWT has been evaluated with the ensemble classifier model. Additionally, to improve the motor imagery Haralick features are extracted, and which are selected with the help of Crow search algorithm (CSA). Finally, the proposed method is validated based on the BCI competition dataset and laboratory measured data. The overall results shows feature selection algorithms CSA provides much better performance with EEG motor imagery classification. In future work could take a more in-depth look at the differences between networks of the same type but varying sizes, generalize this analysis to different datasets, or tackle the tuning of neural network hyper parameters by using other techniques, such as Bayesian optimization to guide the search in a more informed (and thus, perhaps more efficient) way.

REFERENCES

[1] Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckham PH, Schalk G, Donchin E, Quatrano LA, Robinson CJ, Vaughan TM (2000) Brain–computer interface technology: a review of the first international meeting. IEEE Trans Rehabilit Eng 8(2):164–173.
[2] Abdulkader, S. N., Atia, A., and Mostafa, M.-S. M. (2015). Brain computer interfacing: applications and challenges. Egypt. Inform. J. 16, 213–230.
[3] Friman, O., Volosyak, I., and Graser, A. (2007). Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces. IEEE Trans. Biomed. Eng. 54, 742–750.
[4] Ko, L.-W., Chikara, R. K., Lee, Y.-C., and Lin, W.-C. (2020). Exploration of user's mental state changes during performing brain-computer interface.
[5] Beuchat, N. J., Chavarriaga, R., Degallier, S., and Millán, J. D. R. (2013). "Offline decoding of upper limb muscle synergies from EEG slow cortical potentials," in 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (Lausanne: Swiss Federal Institute of Technology), 3594–3597.
[6] Yin, E., Zeyl, T., Saab, R., Hu, D., Zhou, Z., and Chau, T. (2016). An auditory-tactile visual saccade-independent p300 brain-computer interface. Int. J. Neural Syst. 26:1650001.
[7] Yu, Y., Zhou, Z., Liu, Y., Jiang, J., Yin, E., Zhang, N., et al. (2017). Self-paced operation of a wheelchair based on a hybrid brain-computer interface combining motor imagery and p300 potential. IEEE Trans. Neural Syst. Rehabil. Eng. 25, 2516–2526.
[8] Chikara, R. K., and Ko, L.-W. (2019). Neural activities classification of human inhibitory control using hierarchical model. Sensors 19:3791.
[9] Choi, K., and Cichocki, A. (2008). "Control of a wheelchair by motor imagery in real time," in International Conference on Intelligent Data Engineering and Automated Learning (Springer), 330–337.

CVR Journal of Science and Technology, Volume 21, December 2021
DOI:10.32377/cvrjst 2107

[10] Belkacem, A. N., Nishio, S., Suzuki, T., Ishiguro, H., and Hirata, M. (2018). Neuromagnetic decoding of simultaneous bilateral hand movements for multidimensional brain-machine interfaces. IEEE Trans. Neural Syst. Rehabil. Eng. 26, 1301–1310.

[11] Chen, C., Zhang, J., Belkacem, A. N., Zhang, S., Xu, R., Hao, B., et al. (2019). G-causality brain connectivity differences of finger movements between motor execution and motor imagery. J. Healthcare Eng. 2019.

[12] Wang, K., Xu, M., Wang, Y., Zhang, S., Chen, L., and Ming, D. (2020). Enhance decoding of pre-movement eeg patterns for brain-computer interfaces. J. Neural Eng. 17:016033.

[13] Raudys SJ, Jain AK. Small sample size effects in statistical pattern recognition: Recommendations for practitioners. IEEE Trans Pattern Anal Mach Intell. 1991;13(3):252–264.

[14] Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A, et al. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. J Neural Eng. 2018;15(3).

[15] Amaldi E, Kann V. On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. Theor Comput Sci. 1998;209(1-2):237–260.

[16] Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res. 2003;3(Mar):1157–1182.

[17] Ortega J, Asensio-Cubero J, Gan JQ, Ortiz A. Classification of motor imagery tasks for BCI with multiresolution analysis and multiobjective feature selection. Biomed Eng Online. 2016;15(73). pmid:27454531.

[18] Corralejo R, Hornero R, Alvarez D. Feature selection using a genetic algorithm in a motor imagery-based Brain Computer Interface. In: International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2011. p. 7703–7706.

[19] Abootalebi V, Moradi MH, Khalilzadeh MA. A new approach for EEG feature extraction in P300-based lie detection. Comput Methods Programs Biomed. 2009;94(1):48–57.

[20] Kee CY, Ponnambalam SG, Loo CK. Multi-objective genetic algorithm as channel selection method for P300 and motor imagery data set. Neurocomputing. 2015;161:120–131.

[21] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. J Neural Eng. 2018;15(5).

[22] Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comput Biol Med. 2018;100:270–278.

[23] Davidson PR, Jones RD, Peiris MTR. EEG-based lapse detection with high temporal resolution. IEEE Trans Biomed Eng. 2007;54(5):832–839.

[24] Wang P, Jiang A, Liu X, Shang J, Zhang L. LSTM-based EEG classification in motor imagery tasks. IEEE Trans Neural Syst Rehabil Eng. 2018;26(11):2086–2095.

[25] Kumar S, Sharma A, Tsunoda T. An improved discriminative filter bank selection approach for motor imagery EEG signal classification using mutual information. BMC Bioinformatics. 2017;18.

[26] Kumar S, Sharma A, Tsunoda T. Brain wave classification using long short-term memory network based OPTICAL predictor. Sci Rep. 2019;9(1).

[27] Gaur P, Pachori RB, Wang H, Prasad G. A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry. Expert Syst Appl. 2018; 95:201–211.

[28] Jiao Y, Zhang Y, Chen X, Yin E, Jin J, Wang X, et al. Sparse Group Representation Model for Motor Imagery EEG Classification. IEEE J Biomed Health Inform. 2019;23(2):631–641. pmid:29994055

[29] Saa, J. F. D., and Cetin, M. (2012). A latent discriminative model-based approach for classification of imaginary motor tasks from EEG data. J. Neural Eng. 9:026020.

[30] Suk HI, Lee SW. Subject and class specific frequency bands selection for multiclass motor imagery classification. Int J Imag Syst Tech 2011; 21:123–30.

[31] Wei, C., Chen, L., and Zhang, A. (2019). EEG emotion recognition based on integrated convolutional neural network. J. East China Uni. Sci. Technol. (Natural Science Edition) 45, 614–622.

[32] Liang NY, Huang GB, Saratchandran P, Sundarajan N. A fast and accurate online sequential learning algorithm for feed forward networks. Neural Networks IEEE Trans. 2006;7(6):1411–23.

[33] Zhang R, Lan Y, Huang GB, Soh YC. Extreme learning machine with adaptive growth of hidden nodes and incremental updating of output weights. Auto Intell Syst LNCS 2011; 6752:253–62. Jan.

[34] Padfield N, Zabalza J, Zhao H, Masero V, Ren J. EG-based brain-computer "Interfaces using motor-imagery: techniques and challenges," sensors (MDPI) in press 2019.

[35] Askarzadeh A (2016) A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm. Comput Struct 169:1–12.

[36] Manogaran G, Varatharajan R, Lopez D, Priyan MK, Sundarasekar R, Thota C (2018) A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system. Future Gener Comput Syst 82:375–387.

[37] Emery NJ, Clayton NS (2004) The mentality of crows: convergent evolution of intelligence in corvids and apes. Am Assoc Adv Sci 306(5703):1903–1907.

[38] Manogaran G, Vijayakumar V, Varatharajan R, Kumar PM, Sundarasekar R, Hsu CH (2018) Machine learning based big data processing framework for cancer diagnosis using hidden Markov model and GM clustering. Wirel Pers Commun 102(3):2099–2116.

[39] Zolghadr-Asli B, Bozorg-Haddad O, Chu X (2017) Crow search algorithm (CSA), advanced optimization by nature-inspired algorithms, 2017, pp 143–149.