

Motor Imagery EEG Signal Classification Using Optimized Convolutional Neural Network

Abstract. Motor Imagery (MI) signals help the Brain-Computer Interface framework (BCI) to enable the binding of the human brain to external devices. Thus, both BCI and MI together are instrumental in enhancing the lives of patients affected by motor neuron disorders. A novel MI-Electroencephalography (EEG) signal identification and classification approach is proposed in this work. An error-free extraction algorithm is required to extract and classify the temporal and spatial features successfully. This paper proposes the Hilbert Transform (HT) for band energy analysis and Gabor Filter for the selection of optimal frequency band. In this work, the Wavelet Packet Decomposition (WPD) algorithm is used for feature extraction and it decomposes the signal into high and low-frequency components before extracting band coefficients. Moreover, the Convolution Neural Network (CNN) classifier is employed for the classification of MI-EEG tasks. The classification accuracy of the CNN classifier is enhanced using Sea Lion Optimization (SLno) algorithm. The approach is verified using MATLAB and the results are substantially better than those found in the current research, with an average classification accuracy rate of 96.44% by employing a smaller number of criteria, lessening resource consumption, and eliminating the influence of individual differences. The recommended method minimizes classification computation time while enhancing classification accuracy.

Streszczenie. Sygnały obrazu motorycznego (MI) pomagają strukturze interfejsu mózg-komputer (BCI) umożliwić wiązanie ludzkiego mózgu z urządzeniami zewnętrznymi. Zatem zarówno BCI, jak i MI razem odgrywają zasadniczą rolę w poprawie życia pacjentów dotkniętych zaburzeniami neuronu ruchowego. W tej pracy zaproponowano nowatorskie podejście do identyfikacji i klasyfikacji sygnałów MI-Elektroencefalografii (EEG). Do pomysłowego wyodrębnienia i sklasyfikowania cech czasowych i przestrzennych wymagany jest bezbłędny algorytm ekstrakcji. W artykule zaproponowano transformatę Hilberta (HT) do analizy energii pasma oraz filtr Gabora do wyboru optymalnego pasma częstotliwości. W tej pracy do ekstrakcji cech wykorzystano algorytm Wavelet Packet Decomposition (WPD), który rozkłada sygnał na składowe o wysokiej i niskiej częstotliwości przed wyodrębnieniem współczynników pasma. Ponadto do klasyfikacji zadań MI-EEG wykorzystuje się klasyfikator Convolution Neural Network (CNN). Dokładność klasyfikacji klasyfikatora CNN jest zwiększona dzięki zastosowaniu algorytmu Sea Lion Optimization (SLno). Podejście to jest weryfikowane przy użyciu MATLAB-a, a wyniki są znacznie lepsze niż w bieżących badaniach, ze średnim współczynnikiem dokładności klasyfikacji wynoszącym 96.44% przy zastosowaniu mniejszej liczby kryteriów, mniejszym zużyciu zasobów i wyeliminowaniu wpływu różnic indywidualnych. Zalecana metoda minimalizuje czas obliczeń klasyfikacyjnych, jednocześnie zwiększając dokładność klasyfikacji. (Klasyfikacja sygnału EEG z obrazowania ruchu przy użyciu zoptymalizowanej sieci neuronowej spłotowej)

Keywords: Hilbert transform, Gabor filter, Sea Lion optimization, Convolutional Neural Network.

Słowa kluczowe: Transformata Hilberta, filtr Gabora, optymalizacja lwa morskigo, konwolucyjna sieć neuronowa, dekompozycja pakietów falkowych.

Introduction

In the current era of artificial intelligence, computer-aided devices, and automated applications, the technology of Brain-Computer Interface (BCI) has intrigued many researchers on account of its ability to bind computers and the human mind. It is a blessing for patients affected by serious motor neuron disorder conditions, as it allows them to accomplish tasks that are previously unrealizable. It translates the brain activity (neural signals) into instructions, used for the control of wheelchairs, gaming applications, computers, speech synthesizers, robotic arms and home applications. Thereby, this alternative non-muscular communication approach is extremely useful for the rehabilitation of patients affected by brain injuries, epilepsy and paralysis [1][2][3][4]. The sensors are usually used to gather the neural signals from the brain in BCI and based on the positioning of these sensors, BCI is categorized into three major types (Invasive, Partially Invasive and Non-Invasive). The invasive and partially invasive BCI uses surgically implanted electrodes for obtaining signals with high resolution and high Signal-to-Noise-Ratio (SNR). But on the flip side, both these types are presented with certain safety and infection problems. Therefore, the non-invasive BCI is relatively safer, since it is surgery-free and uses a wearable EEG cap for gathering brain signals of high temporal resolution [5] [6].

In this study, the recognition and successful classification of Motor Imagery (MI), which is an Electroencephalography (EEG) based BCI paradigm is focussed. MI refers to the process in which, the imagination of a muscle movement leads to the generation of neural signals without performing the actual motor movement [7]. However, the small amplitude of MI-EEG signals is

regarded as its major limitation as it is highly susceptible to interferences in the form of heart rhythm, teeth grinding, muscular movements, eye movement and an eye blink, resulting in lower SNR [8]. The choice of an effective feature extraction and classifier technique is thereby instrumental in the accurate classification of the MI-EEG signal. The prominently used feature extraction techniques include Support Vector Machines [9], Linear Discriminant Analysis (LDA) [10], Independent Component Analysis (ICA) [11], Principal Component Analysis (PCA) [12], Filter-Bank Common Spatial Patterns (FBCSP) [13] and CSP [14]. In spite of the merits associated with these techniques, they are also bound with certain limitations. In the case of Fourier Transform (FT) [15] based feature extraction, the temporal information is completely lost and only the spatial resolution of the signal is preserved. The Autoregressive [16] based feature extraction is delimited due to its high noise content sensitivity, whereas the Power Spectral Density (PSD) [17] based feature extraction approach are affected by its liability to electrode locality. Hence, the WPD algorithm is used in this work for feature extraction and it is effective in decomposing the signal into high and low-frequency components before extracting specific band coefficients. In [18], the SVM classifier is employed with adaptive Riemannian approaches for classifying a BCI IV Ila dataset and a classification accuracy of 80% is achieved. Moreover, a classification accuracy of 85% is achieved using an SVM classifier compare to a logistic and multilayer perceptron classifier [19]. The SVM classifier in general is not suitable for handling non-linear problems and is highly sensitive to missing data. Furthermore, for certain complicated non-linear problems, the determination of kernel function is also very difficult [19]. In [20], the Decision

Tree (DT) classifier is used for the classification of MI tasks, and an optimal accuracy of 85.6% is accomplished using this technique. The combined application of K-nearest neighbor classifier and WPD yielded a classification accuracy of 92.8% in [21]. The implementation of Long Short-Term Memory (LSTM) [22] and Artificial Neural Network (ANN) classifiers [23] is limited to the high probability of their loss function descending into local optima. In this study, a CNN classifier is employed for the successful classification of MI-EEG signals. Moreover, the challenge of enhanced classification accuracy is solved by tuning the weights of the CNN classifier using Sea Lion Optimization (SLnO) technique.

A novel MI-EEG task recognition and classification using optimized CNN is presented in this study. The Hilbert Transform is employed for estimating the energy in each sub-band of the input signal before the selection of the optimal frequency band using the Gabor filter. The effective classification of an MI-EEG task requires suitable feature extraction and classification approaches. Thereby, the WPD approach is used for feature selection, while CNN is used for classification in this work. Moreover, the weights of CNN are tuned using SLnO algorithm and the efficacy of the entire approach is observed using MATLAB software.

Proposed System Description

The dynamic experience of MI refers to the mental simulation of a physical motor movement with the exception of any evident muscular mobility. The combination of both MI and BCI is effective in the transformation of brain neural signals into physical actions. Thereby, the MI-BCI is capable of providing mobility to patients who are previously deprived of it. The EEG signals which form the basis of MI-BCI have low SNR as it is easily affected by many disturbances due to their small magnitude. Hence an appropriate feature extraction and classification technique is crucial for the effective classification of the MI-EEG signals. Consequently, an effective MI-EEG classification technique based on an Optimized CNN classifier as seen in Fig. 1 is proposed in this work. Initially, the training and testing data are given as input to the proposed setup and the energy band of the given input signals are analyzed using Hilbert Transform.

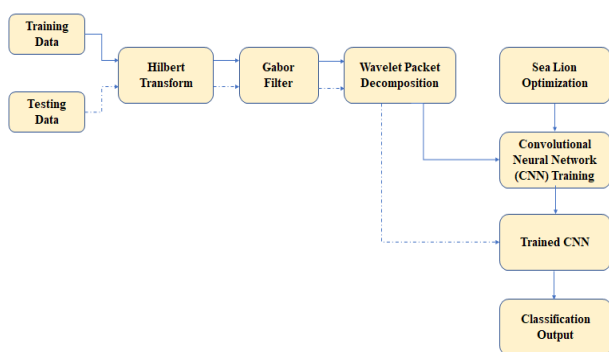


Fig. 1. Architecture of MI-EEG classification approach using Optimized CNN

The Hilbert Transform analyses the input signal and determines the energy level in each sub-band. The process of band selection follows Hilbert Transform-based band energy analysis. The required frequency band is selected using Gabor Filter and then the crucial process of feature extraction is carried out with the aid of WPD. Using WPD, the essential band coefficients are extracted after the signal is decomposed into high and low-frequency components. The SLnO CNN classifier finally classifies the MI-EEG

signals with excellent accuracy. The entire classification approach is executed in MATLAB and it is noted that both SLnO and WPD are effective in enhancing the classification accuracy of CNN.

Proposed System Modelling

Hilbert Transform

The energy band of the given input signals is analyzed using Hilbert Transform. For a real-time function with $x(t)$, the Hilbert Transform is expressed by,

$$(1) \quad \hat{x}(t) = H[x(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} x(\tau) \frac{1}{t-\tau} d\tau$$

The output $\hat{x}(t)$ is similarly a time-dependent function, as observed from (1), the transformation has no effect on the independent variable. Furthermore, the function $\hat{x}(t)$ is linear function of $x(t)$. It is obtained by applying convolution with $(\pi t)^{-1}$ to $x(t)$, as demonstrated in the relationship below:

$$(2) \quad \hat{x}(t) = \frac{1}{\pi t} * x(t)$$

On applying Fourier Transform the expression (2) becomes

$$(3) \quad F\{\hat{x}(t)\} = \frac{1}{\pi} F\left\{\frac{1}{t}\right\} F\{x(t)\}$$

Hence,

$$(4) \quad F\left\{\frac{1}{t}\right\} = \int_{-\infty}^{\infty} \frac{1}{x} e^{-j2\pi f x} dx = -j\pi \operatorname{sgn} f$$

Here,

$\operatorname{sgn} f$ is $+1$ for $f > 0$, 0 for $f = 0$ and -1 for $f < 0$

Thus (3) gives $x(t)$ Fourier's transform of its Hilbert transform as

$$(5) \quad F\{\hat{x}\} = -j \operatorname{sgn} f F\{x(t)\}$$

The result is then produced in a frequency domain by multiplying $x(t)$ spectrum by $j(+90^\circ)$ for negative and $-j(-90^\circ)$ for positive frequencies. By using an inverse Fourier transform, the time domain result is achieved. As a result, the harmonic conjugate of original function $x(t)$ is represented by its Hilbert transform.

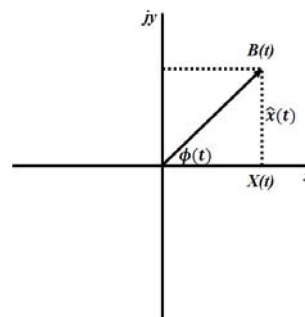


Fig. 2. Representation of Complex Envelope

When referring to the idea of an analytical signal or pre-envelope of real signal $x(t)$, the following expression is used to describe it:

$$(6) \quad y(t) = x(t) + j\hat{x}(t)$$

For envelope $B(t)$ of $y(t)$ the expression is defined as

$$(7) \quad B(t) = \sqrt{x^2(t) + \hat{x}^2(t)}$$

In complex plane the instantaneous phase angle is given by

$$(8) \quad \phi(t) = \arctan\left(\frac{\hat{x}(t)}{x(t)}\right)$$

The envelope generated using (7) has same slope and magnitude as original signal $x(t)$ at or around its local maxima, as illustrated in Fig. 2, when $\hat{x}(t) = 0$. This is because $B(t)$ and $x(t)$ share tangents and have same values at those sites where $\hat{x}(t) = 0$. Additionally, it is evident from (7) that $B(t)$ is always a positive function. As a result, the Hilbert transform provides the highest contribution to $B(t)$ at places where $x(t) = 0$. The process of band selection follows Hilbert Transform based band energy analysis.

Gabor Filter

The band selection process, which is used for selecting the optimal frequency band, comes after Hilbert Transform based band energy analysis. Here, the Gabor filters are used for band selection and these filters have important advantage of being insensitive to rotation, scale, translation, photometric disturbances, luminance, and image noise. A complex sinusoidal wave modulates a Gaussian kernel function, which is a Gabor filter and written as;

$$(9) \quad G(x, y) = \frac{f^2}{\pi\gamma} \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp(j2\pi f x' + \phi)$$

Here,

$$(10) \quad x' = x \cos\theta + y \sin\theta$$

$$(11) \quad y' = -x \sin\theta + y \cos\theta$$

From above equations the frequency of sinusoid signal is indicated as f , θ denotes the orientation of Gabor function's parallel stripes and normal stripes in relation to one another, Phase offset as ϕ , Gaussian envelope standard deviation as σ and γ specifies the spatial aspect ratio that describes Gabor function's support's ellipticity. Similarly, to identify significant structures, the obtained spectrograms were separated into five sub-images corresponding to frequency ranges of the beats. The resulting 8-bit grey scale images is then created from the spectrograms. This is based on bands suggested by segmenting spectrogram, but they only utilise four bands because gamma band was left out because the lost information was already accounted for. Delta (0-4Hz), Theta (4-8Hz), Alpha (8-12Hz), Beta (12-30Hz), and Gamma (30-50Hz) were the ranges used.

Feature Extraction by WPD

For EEG processing, feature extraction is crucial and it is accomplished using WPD, which allows multi-level time-frequency decomposition. It uses short-time and long-time intervals for high-frequency and low-frequency information respectively. It contains a variety of bases, and each base will produce a different classification performance. The input signal is divided into two orthonormal subspaces, V and W , by wavelet decomposition. The spaces V and W are complementary, with low frequency information contained in V and high frequency information contained in W . The low frequency subspace V is subjected to multiple decomposition, as seen in Fig. 3. The WPD is regarded as a more generalised version that enables high frequency band decomposition of signals retained in wavelet decomposition thereby separating only the frequency axis sharply towards low frequency. WPD results in a full wavelet packet tree, as depicted in Fig. 4, in which n^{th}

subspace of wavelet packet at j^{th} scale is specified as $U_{j,n}$ and its orthogonal basis corresponding is determined as $U_{j,k}^n(t)$. From which $U_{j,k}^n(t) = 2^{-\frac{j}{2}} u^n(2^{-j}t - k)$ where k indicates shift factor and satisfies (12) and (13).

$$(12) \quad U_{j,0}^n(t) = \sum_k h_0(k) u_{j-1,k}^i, \text{ for } n \text{ is even}$$

$$(13) \quad U_{j,0}^n(t) = \sum_k h_1(k) u_{j-1,k}^i, \text{ for } n \text{ is odd}$$

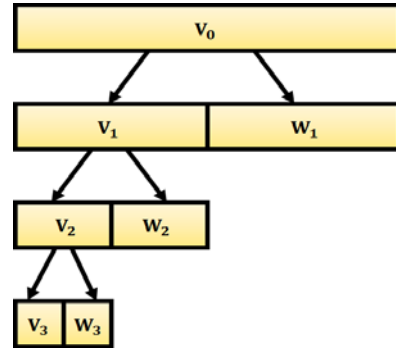


Fig. 3. Wavelet Decomposition Structure

$U_0^0(v_0)$							
$U_1^0(v_1)$				$U_1^1(w_1)$			
$U_2^0(v_2)$	$U_2^1(w_2)$	$U_2^2(w_2)$	$U_2^3(w_2)$	$U_2^4(w_2)$	$U_2^5(w_2)$	$U_2^6(w_2)$	$U_2^7(w_2)$
$U_3^0(v_3)$	$U_3^1(w_3)$	$U_3^2(w_3)$	$U_3^3(w_3)$	$U_3^4(w_3)$	$U_3^5(w_3)$	$U_3^6(w_3)$	$U_3^7(w_3)$
... ..							

Fig. 4. WPD Structure

Here, $j, k \in Z, n = 0, 1, 2, \dots, 2^j - 1, h_0(k), h_1(k)$ represents couple of quadruple mirror filters, irrelevant to scales and also satisfies with (14)

$$(14) \quad h_1(k) = (-1)^{1-k} h_0(1-k)$$

From above equation the sample sequence of $f(t)(f(k\Delta t))$ is employed directly as coefficient of HH in approximation whenever the scale is just right. By using quadruple wavelet packet transformation, the coefficient of WPD at j^{th} level and k^{th} sample is represented by (15) and (16) respectively.

$$(15) \quad d_j^n(k) = \sum_m h_0(m-2k) d_{j-1}^{\frac{n}{2}}(m) \text{ for } n \text{ is even}$$

$$(16) \quad d_j^n(k) = \sum_m h_1(m-2k) d_{j-1}^{\frac{(i-1)}{2}}(m) \text{ for } n \text{ is odd}$$

The $(j-1)^{th}$ level is used to determine decomposition coefficient of j^{th} level, and sequential analogy is used to acquire the coefficients of all levels. The frequency ranges of all subspaces at level (U_j^n) , once it has been divided into j levels, are $\left\{ \left[0, \frac{f_s}{2^{j+1}} \right]; \left[\frac{f_s}{2^{j+1}}, \frac{2f_s}{2^{j+1}} \right]; \left[\frac{2f_s}{2^{j+1}}, \frac{3f_s}{2^{j+1}} \right]; \dots; \left[\frac{(2^j-1)f_s}{2^{j+1}}, \frac{f_s}{2} \right] \right\}$, where f_s specifies the sampling frequency. Currently, wavelet transform-based feature extraction for spontaneous EEG derives coefficients at the targeted frequency bands based on past knowledge. However, the EEG manufacturing mechanism is exceedingly intricate, making it difficult to simply gain accurate prior knowledge.

Average Coefficient

The total length of decomposition coefficients for any wavelet packet base is equal to the length of original discrete sequences, but since the new sequence concentrate coefficients, it makes it simple to extract most

important features. Initially there are l EEG channels with sampling and the value of $l = 1, 2, \dots, i, \dots, C$. The frequency of EEG signal is $0 \sim \frac{f_s}{2}$. The first feature is chosen from the sub band mean ($Mean_{j,n}$) at j^{th} level where frequency range is 0-50 Hz because, frequency of usable EEG is lower than 50 Hz.

$$(17) \quad (Mean_{j,n}) = \frac{2^N}{2^j} \sum_k d_j^n(k)$$

Each channel's EEG signal is calculated in accordance with (17); the feature vector created by all channels is presented

as $M = \{MEA_{j,0}^1, MEA_{j,1}^1, \dots, MEA_{j,0}^i, \dots, MEA_{j,0}^c, MEA_{j,1}^c, \dots\}$. The decomposition level (j) is chosen in accordance with sampling frequency and reality because, in theory, the greater frequency resolution is accomplished with higher decomposition level, but in practise, the feature space dimension is increased owing to computational complexity.

Sub-Band Energy

From an energy perspective, WPD distributes signal energy into multiple time-frequency plains, and the square amplitude integration of WPD is inversely proportional to signal power. Similar to sub-band mean selection rule, it selects the sub-band energy ($E_{j,n}$) at j^{th} level, with a frequency range of 0–50 Hz, as its initial features.

$$(18) \quad (E_{j,n}) = \sum_k (d_j^n(k))^2$$

Each channel's EEG signal is estimated in accordance with (18); the feature vector created by all channels is displayed as $N = \{E_{j,0}^1, E_{j,1}^1, \dots, E_{j,0}^i, \dots, E_{j,0}^c, E_{j,1}^c, \dots\}$, it can also be written as $N = \{n1, n2, n3, \dots\}$.

Sea Lion Optimized CNN

The process of SLnO CNN based classification follows the process of feature extraction using WPD. The CNNs are one of the prominently used techniques for the classification of EEG signals and its general architecture is given in Fig. 5. Initially, numerous WPD maps are created from the input signal after undergoing sliding window application and WPD operation estimation. The structure of the CNN is built using Keras toolbox with Sea lion Optimization, 8-fold cross validation and epochs set to 150. Moreover, it comprises of four convolution layers and in the first layer, trainable filter are employed to convolve the input data. This layer delivers 25 feature maps using 25 filters and the l^{th} feature map is given as,

$$(19) \quad h_{ij}^l = f(x) = f((W_k * X)_{ij} + b_l)$$

Where, Bayes bias is specified as b_l , weight matrix is specified as W_k and the input data is specified as X . The activation function used is rectified linear unit ($ReLU$), which is given as,

$$(20) \quad F(a) = ReLU(a) = \ln(1 + e^x)$$

The pooling layer that follows convolutional layer, entails two max pooling layers with filter sizes set at (2,1) and (3,1). Here, the problem of overfitting is avoided with the inclusion of dropout layer. For classifying MI-EEG tasks, categorical cross-entropy loss function is selected, whereas the appropriate CNN parameters are learned using back propagation.

Sea Lion Optimization (SLnO)

SLnO is proposed to address large-scale optimization. It imitates sea lions' hunting techniques, such as circling and seizing prey or employing their tail and whiskers. Sea lions

are intelligent mammals with the distinguishing trait to reciprocate instantly to the movement of fish. Additionally, they have remarkable senses that enable them to locate fish prey within the deepest, darkest waters. They significantly widen their pupils once they are focused on the prey to give their eyes clear underwater vision in the presence of light. In a murky environment, the sea lion's extremely sensitive whiskers are its most crucial attribute for locating the prey. If the prey escapes, it generates a wake or waves behind it to track the prey.

The following are the critical stages of sea lions' hunting behavior:

1. Locating and hunting the prey using their whiskers.
2. Calling for other subgroup members,
3. Swirling and chasing their prey
4. Attack towards the prey.

In this study, the sea lions' hunting strategy is mathematically analysed for SLnO algorithm development and optimization execution.

Detecting and tracking phase

Using uniform random distribution, SLnO generates N (the population's size) D -dimensional solutions (21) in the search space initially. Then they locate the prey within the sea lion swarm, gather the individuals who make up the subgroup, and set up the net. Prey is regarded as the best current option or closest to the ideal solution. These actions are exhibited in (2).

$$(21) \quad X_{i,j}^{ini} = X_{i,j}^{min} + rand_{i,j}(X_{i,j}^{max} - X_{i,j}^{min})$$

Where $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, D$. $X_{i,j}^{ini}$ is the initial position vector of i^{th} solution; the terms $X_{i,j}^{max}$ and $X_{i,j}^{min}$ refers to the minimum and maximum value for the j^{th} dimension of i^{th} solution; $rand$ specifies uniform random value in interval $[0, 1]$. The objective function is used to evaluate the fitness of the solution which is represented in (22)

$$(22) \quad X_{g+1} = X_{best} - C |2rX_{best} - X^g|$$

$$(23) \quad C = 2 \left(1 - \frac{g}{g_{max}}\right)$$

Where, the terms C, r, g_{max}, g, X^g and: X_{best} specifies a variable, a random value ranging between $[0, 1]$, maximum count of generation, present iteration of generations, sea lion in iteration g and the best solution position vector respectively. After updating, the search agent's new position is specified as X_{g+1} . The flowchart of SLnO is given in Fig. 6.

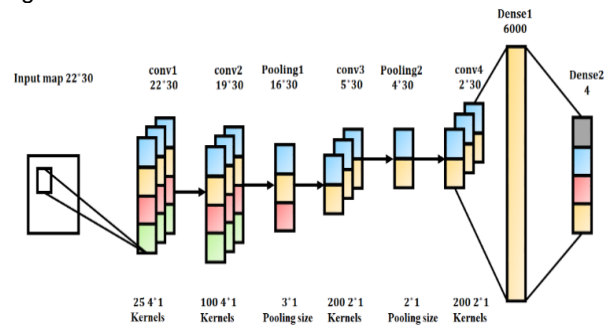


Fig. 5. CNN Architecture

Vocalization phase

When a cluster of prey is spotted by a sea lion, it signals other sea lions to assemble and form a capture net surrounding its prey. This particular sea lion is regarded as the group's leader, directing their movements and dictating

their conduct. These characteristics are represented mathematically in (24), (25), and (26). (30)

$$(24) \quad SP_{leader} = |V_1(1 + V_2)/V_2|$$

$$(25) \quad V_1 = \sin \theta$$

$$(26) \quad V_2 = \sin \phi$$

Where: SP_{leader} refers to the leaders command required to be abided by remaining sea lions of the group; the terms ϕ and θ refers to the angle of voice refraction and reflection in the water respectively. In this study, $\phi = 2\pi(r - 1)$ and $\theta = 2\pi r$, where r is a random number ranging between $[0, 1]$.

Attacking phase

The sea lions' hunting activities are divided into two stages,

- **Dwindling encircling technique:** This behaviour varies in accordance to the C value that progressively decreases during iteration from 2 to 0. Consequently, allowing the search space surrounding best position to contract, forcing remaining search agents in search space to update throughout the iteration.
- **Circling updating position:** In (7), where m is a random number within $[1, 1]$, sea lions begin their hunt at edges after chasing the fish's bait ball.

$$(27) \quad X_{g+1} = X_{best} + \cos(2\pi m) |X_{best} - X^g|$$

The second criterion is met by a piece of data from the best available global data, and the first need is met by data on a individuals historical behaviour combined with a random coefficient for the recently updated solution. Three vectors can be combined to create new solutions that can explore the search space as well as the best individual experiments and overall solution.

$$(28) \quad |dif_1 = (2 r_1 X_{best}^g - X^g)|$$

$$(29) \quad |dif_2 = (2 r_2 X_{local}^g - X^g)|$$

$$(30) \quad X^{g+1} = X^g + C \cdot dif_1 + C \cdot dif_2$$

Where: X^g_{local} is the personal best position up to the iteration g ; r_1, r_2 are random numbers in the range $[0, 1]$. Moreover, the difference between the previously best solution and current position is specified as dif_2 , while the difference between the best solution and present position is specified as dif_1 .

Searching for prey (Exploration phase)

Algorithm 1 presents the SLO pseudo-code and summarises the SLO's most crucial operations below.

Algorithm 1: Sea Lion Optimization Algorithm (SLO)

Input: Population size N , the maximum number of generations g_{max}

Output: The best solution X_{best}

Initialize the Sea lion population $X_i (i = 1, 2, \dots, n)$ randomly

Sort the population value by its fitness value and find the global best solution X_{best}

$g = 1$

While $g < g_{max}$

 Calculate the value of C by (23)

 Calculate SP_{leader} using (22)

For $i < N$

If $SP_{leader} < 1.0$ then

If $C > 1$ then

 Calculate dif_1 and dif_2 using (28) and (29)

 Update the location of the current search agent using

```

Else
    Create a new solution using (22)
    Create its opposite solution  $X_{oppo}^{g+1}$  using (23)
    Calculate the fitness of both solution
    Compare and keep the location of the better one as the
    new position for the current
    individual.
End If
End If
Else
If  $rand() < 0.5$  then
    update the location of the current search agent by (27)
Else
    update the location of current search agent by levy -
    flight(20)
End If
End If
End If
Check the bound and calculate the fitness of the new solution
Replace the old solution by the new one if it has a better
fitness value
Sort the population by its fitness values and update the global
best solution  $X_{best}$ 
 $g = g + 1$ 
End While
Return:  $X_{best}$ 

```

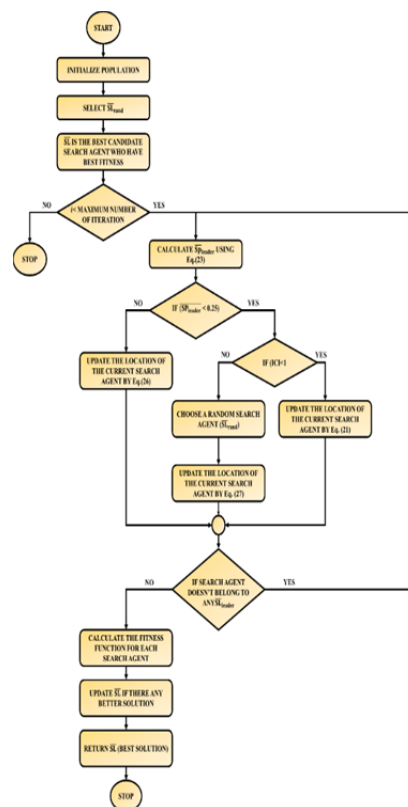


Fig. 6. Flowchart of SLO algorithm

Thus, the SLO algorithm is used for the effective optimization of the weights in the convolutional layer of the CNN classifier. Consequently, both WPD feature extraction and SLO contributes the enhanced classification accuracy of CNN in classifying MI-EEG tasks.

Result and Discussion

The proposed work is validated using BCI completion IV dataset 2a with 9 subjects. It involves four motor imaginary movement, like left hand, right hand, foot and tongue. Each classes consist of 72 trials and considering both the training and evaluation session all together 576 trials were available. For this work 144 trials from the training and 144 trials from the evaluation for two classes were consider.

To improve convergence, the model has been trained using SLnO optimizer. Additionally, classification cross-entropy loss reduction is achieved in order to train the model. Fig. 7 illustrates iterative variations in loss and accuracy during training phase through the iteration curves. The training set's accuracy converges to near to 1 and loss is near to 0, demonstrating the successful convergence of the model.

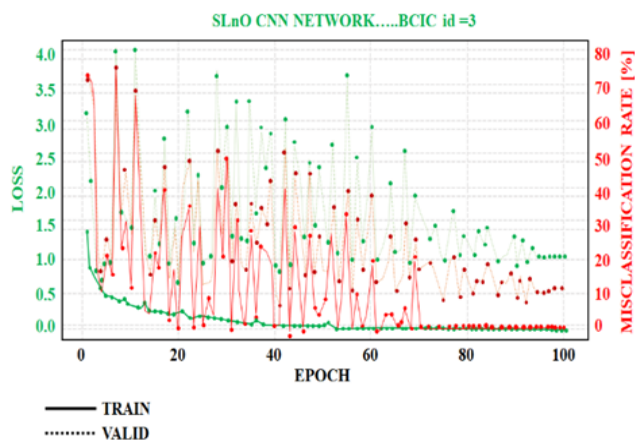


Fig. 7. Iteration Result

From Fig. 7, the green solid line and dotted line show changes in loss of validation and training sets throughout the iterative process, while dashed line and red solid line show changes in accuracy of validation and training sets, respectively. It is observed that, from Table 1 and Fig. 7 our technique performs better in terms of decoding accuracy for both datasets. Using the BCI competition IV dataset 2a, our model outperforms an average accuracy of classification of 96.44%.

Table 1. Classification Accuracy Vs Subject

Subject	ConvNet	DeepConvNet	LSTM	CARNet	Proposed SLnO-CNN
A01	75.12	68.74	84.77	80.61	98.84
A02	50.44	50.89	60.56	65.33	98.66
A03	80.53	86.45	91.35	87.53	99.23
A04	65.31	62.84	63.51	69.05	95.52
A05	57.97	51.03	60.88	50.73	95.27
A06	51.74	52.38	63.84	87.53	96.81
A07	84.68	86.45	84.32	80.24	91.85
A08	76.34	82.45	74.95	78.50	94.39
A09	74.62	82.98	77.38	70.17	97.41
Mean	68.52	69.35	73.50	74.41	96.44

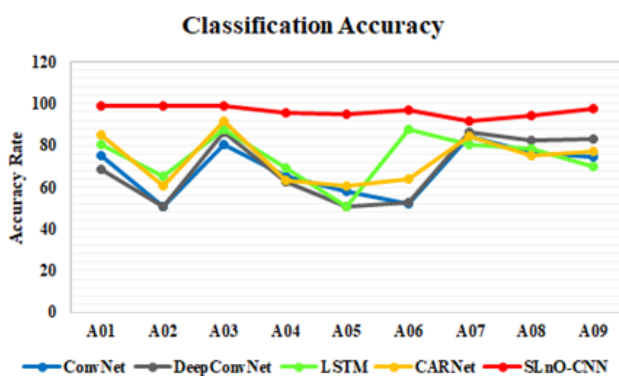


Fig. 8. Classification Accuracy on BCI Competition IV dataset 2a

From Fig. 7, the green solid line and dotted line show changes in loss of validation and training sets throughout the iterative process, while dashed line and red solid line show changes in accuracy of validation and training sets, respectively. It is observed that, from Table 1 and Fig. 7 our technique performs better in terms of decoding accuracy for

both datasets. Using the BCI competition IV dataset 2a, our model outperforms an average accuracy of classification of 96.44%.

Fig. 8 provide a visual comparison of our technique's classification accuracy with those of other standard techniques. On proposed BCI competition IV datasets 2a, the decoding performance has improved, proving applicability of the proposed methodology. Additionally, 4-fold cross validation of the data shows that the proposed model lowers the EEG signal variation impact across different participants to some extent and enhances the model's general resilience and stability, which is significant for application of motor imagery.

Classification Analysis

In Table 2, the classification outcomes are displayed. With a maximum loss of less than 0.40 and an average accuracy of 96.44%. Average testing duration is less than 2s, and training time is estimated finding the product of every epoch time and number of epochs. The classification pace is not slowed down when we achieve high classification accuracy. The classifying level of every subject is largely balanced because determination of appropriate frequency band is based on band's energy.

Table 2. Classified Results

Subject No.	Loss	Accuracy (%)	Training Time (s)	Testing Time (s)
A01	0.08	98.84	725	2
A02	0.07	98.66	751	1
A03	0.17	99.23	763	1
A04	0.16	95.52	744	1
A05	0.25	95.27	750	1
A06	0.34	96.81	721	2
A07	0.15	91.85	752	2
A08	0.20	94.39	763	1
A09	0.9	97.41	759	2
Average	0.25	96.44	748	1.4

Analysing the Impact of Automatic band Selection in WPD performance

Automatic band selection uses various band energy percentages to extract bands. Table 3 includes results as well as a comparison of classification performance of WPD and analyses the impact of automatic band selection over its performance. Automatic band selection has a considerable impact on classification performance, as seen in the table. First, the Hilbert Transform is used to determine the energy %. For this, the raw signal is split into a descriptive and approximation component, and then energy of each subband is determined. The WPD maps are then created using subband with highest proportion of energy. The network's input shape without automatic band selection is $22 * 50$, and input data shape is $22 * 30$. The classification parameters, kappa value, training time, and testing time are all compared in table 3.

Table 3. Automatic Band Selection Comparison of with and without WPD

Parameters	WPD	WPD Without Automatic Band Selection
Training Parameters	154626	171426
Average Training Time	748	1045
Average Test Time	1.4	3.8
Average Loss	0.25	0.28
Average Kappa Value	0.92	0.84

The classification results' reliability is assessed by kappa value. When compared to same network without intelligent

band selection, the WPD performs significantly better. It is necessary for adaptive portion to decrease low-energy band, eliminate low-frequency interference, and eliminate high-frequency noise in order to increase the loss and accuracy. Both networks' average kappa values are more than 0.85, which is not significantly different from one another. Automatic band selection helped to shorten the runtimes during training and testing. Networks typically take 748 and 1045 seconds to train, respectively, which results in a time reduction of 25%. Similarly, testing times are 1.4 and 3.8 seconds, resulting time reduction of 40%. The CNN parameters are accordingly reduced as a result of reduction in WPD band pass, which is a novelty of proposed approach that significantly enhances the testing and training classification effectiveness of CNN.

Conclusion

A robust classification approach using optimized CNN classifier is proposed for the accurate classification of MI-BCI tasks. In this work, Hilbert transform is used for estimating the band energy distribution, while Gabor filter is applied for band selection. Moreover, the process of feature selection is carried out using WPD and SLnO optimized CNN is used as the classifier. From the experimental results, it is noted that both WPD and SLnO are effective in improving the classification accuracy to an excellent value of 96.44%. The WPD is effective in significantly reducing the number of parameters and selecting the optimal features required for classification. With the reduction of large number of parameters, the proposed approach is effective in delivering an excellent real-time performance. The proposed classification approach is also effective in classifying huge data in quick runtime. The testing and training classification of CNN is also enhanced significantly using the proposed approach.

Conflict of Interest. The authors declare that there is no conflict of interest.

Authors. Dr. Thiyam Deepa Beeta, Vel Tech Rangarajan Dr. Sagunthala R & D Institute of Science and Technology, Chennai, Tamilnadu, Email: thiyamdeepa@gmail.com; Shelishiyah Raymond; Vel Tech Rangarajan Dr. Sagunthala R & D Institute of Science and Technology, Chennai, Tamilnadu, Email: shelishiyah.raymond@gmail.com. Padmanabha Sarma Avasarala, Vel Tech Rangarajan Dr. Sagunthala R & D Institute of Science and Technology, Chennai, Tamilnadu, Email: a.padmanabhasarma@gmail.com;

REFERENCES

- [1] D. Wu, Y. Xu, and B. L. Lu, "Transfer Learning for EEG-Based Brain-Computer Interfaces: A Review of Progress Made since 2016," *IEEE Trans. Cogn. Dev. Syst.*, vol. 14, no. 1, pp. 4–19, 2022, doi: 10.1109/TCDS.2020.3007453.
- [2] M. T. Sadiq, X. Yu, Z. Yuan, and M. Z. Aziz, "Identification of motor and mental imagery EEG in two and multiclass subject-dependent tasks using successive decomposition index," *Sensors (Switzerland)*, vol. 20, no. 18, pp. 1–25, 2020, doi: 10.3390/s20185283.
- [3] R. Zhang *et al.*, "A BCI-based environmental control system for patients with severe spinal cord injuries," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1959–1971, 2017, doi: 10.1109/TBME.2016.2628861.
- [4] Y. Zhou, S. He, Q. Huang, and Y. Li, "A Hybrid Asynchronous Brain-Computer Interface Combining SSVEP and EOG Signals," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 10, pp. 2881–2892, 2020, doi: 10.1109/TBME.2020.2972747.
- [5] O. Ali, M. Saif-ur-Rehman, S. Dyck, T. Glasmachers, I. Iossifidis, and C. Klaes, "Enhancing the decoding accuracy of EEG signals by the introduction of anchored-STFT and adversarial data augmentation method," *Sci. Rep.*, vol. 12, no. 1, pp. 1–19, 2022, doi: 10.1038/s41598-022-07992-w.
- [6] M. Sreeja, S. R., Rabha, J., Samanta, D., Mitra, P., & Sarma, "Classification of motor imagery based EEG signals using sparsity approach," *Intell. Hum. Comput. Interact. 9th Int. Conf. IHCI 2017*, vol. 9, pp. 47–59, 2017.
- [7] W. Yi, S. Qiu, H. Qi, L. Zhang, B. Wan, and D. Ming, "EEG feature comparison and classification of simple and compound limb motor imagery," *J. Neuroeng. Rehabil.*, vol. 10, no. 1, pp. 1–12, 2013, doi: 10.1186/1743-0003-10-106.
- [8] Z. Zhang *et al.*, "A Novel Deep Learning Approach with Data Augmentation to Classify Motor Imagery Signals," *IEEE Access*, vol. 7, pp. 15945–15954, 2019, doi: 10.1109/ACCESS.2019.2895133.
- [9] S. Selim, M. M. Tantawi, H. A. Shedeed, and A. Badr, "A CSP\IAM-BA-SVM Approach for Motor Imagery BCI System," *IEEE Access*, vol. 6, pp. 49192–49208, 2018, doi: 10.1109/ACCESS.2018.2868178.
- [10] R. K. Bhatnagar, M., Gupta, G. S., & Sinha, "Linear discriminant analysis classifies the EEG spectral features obtained from three class motor imagination.," *2018 2nd Int. Conf. Power, Energy Environ. Towar. Smart Technol.*, pp. 1–6, 2018.
- [11] X. Wu, B. Zhou, Z. Lv, and C. Zhang, "To Explore the Potentials of Independent Component Analysis in Brain-Computer Interface of Motor Imagery," *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 3, pp. 775–787, 2020, doi: 10.1109/JBHI.2019.2922976.
- [12] L. Cheng, D. Li, G. Yu, Z. Zhang, X. Li, and S. Yu, "A motor imagery EEG feature extraction method based on energy principal component analysis and deep belief networks," *IEEE Access*, vol. 8, pp. 21453–21472, 2020, doi: 10.1109/ACCESS.2020.2969054.
- [13] H. Wang *et al.*, "Diverse Feature Blend Based on Filter-Bank Common Spatial Pattern and Brain Functional Connectivity for Multiple Motor Imagery Detection," *IEEE Access*, vol. 8, pp. 155590–155601, 2020, doi: 10.1109/ACCESS.2020.3018962.
- [14] S. Chaudhary, S. Taran, V. Bajaj, and A. Sengur, "Convolutional Neural Network Based Approach Towards Motor Imagery Tasks EEG Signals Classification," *IEEE Sens. J.*, vol. 19, no. 12, pp. 4494–4500, 2019, doi: 10.1109/JSEN.2019.2899645.
- [15] G. Rodríguez-Bermudez and P. J. García-Laencina, "Automatic and adaptive classification of electroencephalographic signals for brain computer interfaces," *J. Med. Syst.*, vol. 36, no. SUPPL.1, 2012, doi: 10.1007/s10916-012-9893-4.
- [16] K. Polat and S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform," *Appl. Math. Comput.*, vol. 187, no. 2, pp. 1017–1026, 2007, doi: 10.1016/j.amc.2006.09.022.
- [17] A. Gupta *et al.*, "On the Utility of Power Spectral Techniques with Feature Selection Techniques for Effective Mental Task Classification in Noninvasive BCI," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 51, no. 5, pp. 3080–3092, 2021, doi: 10.1109/TSMC.2019.2917599.
- [18] F. Yger, M. Berar, and F. Lotte, "Riemannian Approaches in Brain-Computer Interfaces: A Review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1753–1762, 2017, doi: 10.1109/TNSRE.2016.2627016.
- [19] R. Chatterjee, T. Bandyopadhyay, D. K. Sanyal, and D. Guha, "Comparative analysis of feature extraction techniques in motor imagery EEG signal classification," *Smart Innov. Syst. Technol.*, vol. 79, no. February, pp. 73–83, 2018, doi: 10.1007/978-981-10-5828-8_8.
- [20] R. Chaudhary, P., & Agrawal, "Non-dyadic wavelet decomposition for sensory-motor imagery EEG classificationTitle," *Brain-computer interfaces*, vol. 7, no. 1–2, pp. 11–21, 2020.
- [21] A. Kevric, J., & Subasi, "Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system," *Biomed. Signal Process. Control*, vol. 31, pp. 398–406, 2017.
- [22] P. Wang, A. Jiang, X. Liu, J. Shang, and L. Zhang, "LSTM-based EEG classification in motor imagery tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 11, pp. 2086–2095, 2018, doi: 10.1109/TNSRE.2018.2876129.
- [23] M. Hamed, S. H. Salleh, A. M. Noor, and I. Mohammad-Rezazadeh, "Neural network-based three-class motor imagery classification using time-domain features for BCI applications," *IEEE TENSYPMP 2014 - 2014 IEEE Reg. 10 Symp.*, no. Mi, pp. 204–207, 2014, doi: 10.1109/tenconspring.2014.6863026.