IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 14 November 2024, accepted 6 December 2024, date of publication 16 December 2024, date of current version 3 January 2025.

Digital Object Identifier 10.1109/ACCESS.2024.3517639

RESEARCH ARTICLE

A Spatiotemporal Feature Extraction Technique Using Superlet-CNN Fusion for Improved Motor Imagery Classification

NEHA SHARMA[1], MANOJ SHARMA[2], (Senior Member, IEEE), AMIT SINGHAL[3], (Senior Member, IEEE), NUZHAT FATEMA[4,5], VINAY KUMAR JADOUN[6], (Senior Member, IEEE), HASMAT MALIK[7,8], (Senior Member, IEEE), AND ASYRAF AFTHANORHAN[9,10]

[1]Department of Computer Science Engineering, Faculty of Engineering and Technology, SGT University, Gurugram 122505, India
[2]School of Artificial Intelligence, Bennett University, Greater Noida 201310, India
[3]Electronics and Communication Department, Netaji Subhas University of Technology, Delhi 110078, India
[4]Faculty of Management, Universiti Teknologi Malaysia (UTM), Johor Bahru 81310, Malaysia
[5]Intelligent Prognostic Pvt., Ltd., Delhi 110093, India
[6]Department of Electrical and Electronics Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
[7]Department of Electrical Power Engineering, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Skudai 81310, Malaysia
[8]Department of Electrical Engineering, Graphic Era (Deemed to be University), Dehradun 248002, India
[9]Artificial Intelligence for Islamic Civilization and Sustainability, Universiti Sultan Zainal Abidin (UniSZA), Kuala Nerus, Kuala Terengganu, Terengganu 21300, Malaysia
[10]Operation Research and Management Sciences, Universiti Sultan Zainal Abidin (UniSZA), Kuala Nerus, Kuala Terengganu, Terengganu 21300, Malaysia

Corresponding authors: Vinay Kumar Jadoun (vinay.jadoun@manipal.edu) and Hasmat Malik (hasmat.malik@gmail.com)

ABSTRACT In the realm of Brain-Computer Interface (BCI) research, the precise decoding of motor imagery electroencephalogram (MI-EEG) signals is pivotal for the realization of systems that can be seamlessly integrated into practical applications, enhancing the autonomy of individuals with mobility impairments. This study presents an enhanced method for the precise recognition of MI tasks using EEG data, to facilitate more intuitive interactions between individuals with mobility challenges and their environment. The core challenge addressed herein is the development of robust algorithms that enable the accurate identification of MI tasks, thereby empowering individuals with mobility impairments to control devices and interfaces through cognitive commands. Although there are many different methods for analyzing MI-EEG signals, research into deep learning and transfer learning approaches for MI-EEG analysis remains scarce. This research leverages the superlet transform (SLT) to transform EEG signals into a two-dimensional (2-D) high-resolution spectral representation. This 2-D representation of segmented MI-EEG signals is then processed through an adapted pretrained residual network, which classifies the MI-EEG signals. The effectiveness of the suggested technique is evident as the achieved classification accuracy is 99.9% for binary tasks and 96.4% for multi-class tasks, representing a significant advancement over existing methods. Through an intensive comparison with present algorithms assessed in variety of performance evaluating metrics the present study emphasize the exceptional ability of proposed approach to accurately classify the different MI categories from the EEG signals and which is a great contribution to the field of BCI research field.

INDEX TERMS Motor imagery (MI), deep neural network (DNN), superlet transform (SLT), brain–computer interface (BCI).

I. INTRODUCTION

Electroencephalography (EEG) signals are a gateway to the brain's electrical activity enabling the capture of the

The associate editor coordinating the review of this manuscript and approving it for publication was Joewono Widjaja.

dynamical oscillations at the center of neuronal interactions across various stages of consciousness, mental processes and stimuli response [1]. As a result, it has become a valuable tool in both medical diagnostics and cognitive neuroscience. MI-EEG, which forms the crux of neuroscientific researches, is devoted to the analysis of brain activity generated during the mental simulation of movements, from which the neural bases of motor planning can be deciphered. It creates the basis for future therapeutic practices and BCI (brain-computer interface) technologies [2]. BCIs represent a brand new technological frontier in bridging human mental states with external apparatus, thereby enabling immediate communication and regulation by neural signals decoding hence paving the way for new ways of interaction without body movement necessary. There are various techniques to record brain activity that range from invasive and semi-invasive to non-invasive approaches. Of all non-invasive methods like electroencephalography (EEG), Magnetoencephalography (MEG), Positron Emission Tomography (PET), Functional Magnetic Resonance Imaging (fMRI) and optical imaging, EEG is the outstanding choice. This preference is due to its non-invasiveness, portability potential of users, higher time resolution and lower cost [3]. BCIs can leverage MI-EEG signals in diverse ways to enhance control, holding significant promise in fields that necessitate decoding user thoughts for imagined actions, including gaming, neuro-prosthetics, and neurorehabilitation. For instance, utilizing EEG recordings of left- and right-hand motor imagery can enable the movement of a target, offering a novel communication pathway to compensate for lost motor function. This technology holds the potential to provide amyotrophic lateral sclerosis patients with a clear binary response to queries, thereby enhancing their overall quality of life [4]. Within the realm of scholarly literature, researchers have extensively explored signal processing techniques, encompassing methodologies grounded in time, frequency, and time-frequency analysis. Notably, only a limited subset of these approaches has exhibited the requisite robustness to merit in-depth consideration for further research. The commonly used feature extraction algorithms include wavelet transform (WT) [5], [6], [7] wavelet packet decomposition (WPD) [8], [9], common spatial patterns (CSP) [10], [11], empirical mode decomposition (EMD) [12], [13], empirical wavelet decomposition (EWT) [14], [15], Fourier decomposition method (FDM) [16], [17], and so on. Various differential evolution algorithms such as ant colony optimization (ACO) and artificial bee colony (ABC) were proposed by [18] for optimum feature selection. Traditionally extracted features often rely on manual design, necessitating extensive expertise. Therefore, the automatic identification of significant features from EEG signals is of considerable importance. Deep learning effectively handles non-linear and non-stationary data, automatically deriving useful features from the raw data. In recent years, some deep learning methods [19], [20], [21], [22] are employed for the classification of EEG signals where data is converted into its time frequency (TF) representation such as Short-time Fourier transforms (STFTs), continuous wavelet transforms (CWTs). The cutting edge technique transformer is also employed by numerous researcher in deep learning field [23], [24], [25], [26], [27], [28]. Authors has utilized spatiotemporal features for various classification tasks. However, the dataset available for deep learning techniques is limited, necessitating the need for more data to effectively train the models. This constraint guides us to employ transfer learning techniques to overcome it. Therefore, there are some methods combining traditional feature extraction methods with deep learning methods [29], [30]. In this approach, researchers leverage transfer learning by converting one-dimensional (1-D) data into two-dimensional (2-D) data, which is then trained and evaluated using pre-trained networks. In the proposed methodology, raw MI-EEG data is converted into 2-D data using superlet transform and then classified using pretrained residual CNN models for binary and multiclass classification. Contribution of the study:

- The proposed method distinguishes itself by utilizing superlet transform to generate distinct and informative features for motor imagery tasks. These features, when fed into a CNN architecture, demonstrably increase classification accuracy.
- To thoroughly assess the effectiveness of our algorithm, we implemented an ablation study using three pre-trained residual networks. This systematic analysis reveals the contributions of each component and conclusively demonstrates the algorithm's superior performance.
- Detailed evaluation on two different datasets reveals superior generalization capabilities compared to SoTA benchmarks, highlighting the robustness and adaptability of our proposed network.
- We introduce a pioneering method for motor imagery recognition in BCI by utilizing the superlet transform, offering a unique high-resolution approach.

Following an introduction, the article progresses through a structured analysis: Section II details about the utilized Dataset, Section III provides a detailed description of the methodology used, while Section IV delivers the results and analysis., and Section V summarizes the key insights and also discuss the future scopes.

II. EXPERIMENTAL SETUP

The computational experiments were performed on a high-performance workstation featuring an Intel Core i7-10th generation 16-core processor running at 3.80 GHz, with a Linux 22.04 operating system, 128 GB of RAM, and an NVIDIA Quadro RTX 5000 GPU with 16 GB of memory. MATLAB 2022b was used for all experiments.

Research on MI-EEG signals leverages numerous datasets, each varying in subject count, electrode number, trial duration, total trials, sampling frequency, and MI task types. Among these, the most utilized datasets by researchers are BCI competition IVa and BCI IV 2a. This article utilized dataset IV-A from the BCI Competition-III, which is a

publicly available dataset of EEG signals for brain-computer interface research [31]. The dataset contains EEG data from five healthy subjects who performed right-hand (RH) and right-foot (RF) motor imagery (MI) tasks. EEG data were acquired from 118 out of 128 electrodes according to the international 10/20 system. Visual cues were displayed for 3.5 seconds to indicate the MI task. The unified dataset of 280 trials was split into training, validation, and test sets for individual subjects. The sampling frequency of the signal is 100 Hz. The timing diagram of dataset is shown in Figure 1.

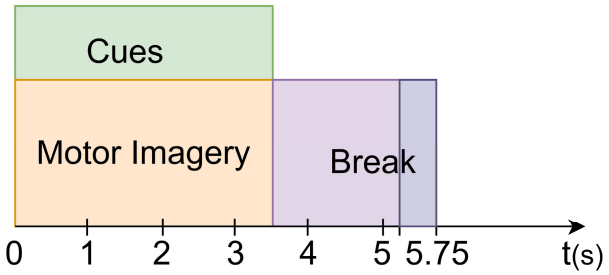

FIGURE 1. Timing diagram for binary class dataset.

The other adapted dataset for classifying multiclass motor imagery (MI) is BCI IV2a [32]. This study employed a dataset comprising 22 EEG channels and 3 EOG channels to observe the brain activity of 9 individuals as they imagined performing four distinct tasks: the main code: didgeridoo left hand (LH), right hand (RH), tongue (T), feet (F). Data gathering took place at 250Hz rate. The motor imagery task was conducted within a time frame of 2 to 6 seconds. In this study, segmentation was carried out at the 4-second mark, and a total of 22 channels were employed. The timing diagram of dataset is depicted in Figure 2.

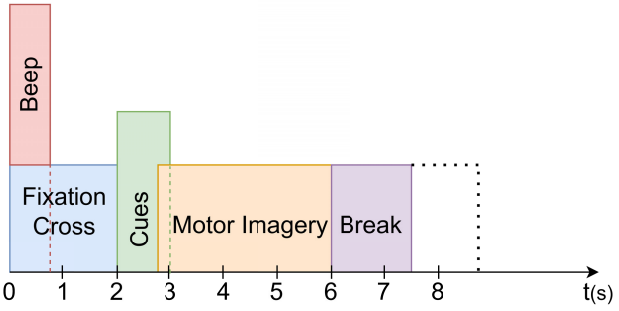

FIGURE 2. Timing diagram for multi class dataset.

In the study, the dataset was divided into three subsets to facilitate training, validation, and testing of the ResNet models. Sixty percent of the input data was allocated for training, ensuring that the model had sufficient data to learn meaningful patterns. Twenty percent of the data was used for validation, allowing the model's performance to be monitored during training, while the remaining 20% was reserved for testing to assess the model's generalization capability on unseen data. Segmentation of the EEG signals was tailored to the type of dataset being used. For the binary classification dataset, a segmentation window of 3.5 seconds was applied, while for the multiclass classification dataset, the window was set to 4 seconds. This segmentation ensures that the input data was divided into manageable portions that reflect meaningful temporal structures in the signals. The learning rate for fine-tuning the deep CNN model, particularly ResNet, was set to 1e-5, striking a balance between convergence speed and model performance. The network was trained for a maximum of 50 epochs, with each epoch consisting of 1,027 iterations. In total, 51,350 iterations were performed, which allowed the model ample opportunity to adjust its parameters and minimize the loss function effectively.

III. METHODOLOGY

The methodological framework of our study is outlined here. Subsection III-A provides a foundational understanding of the Time-Frequency (TF) representation technique, employed for feature extraction and signal characterization. Subsection III-B delves into the specifics of the deep learning architecture, unveiling its network topology, activation functions, and optimization algorithms. Here, a CNN based framework is proposed for classification of RH, LH, F, and T movements. The MI-EEG signal is segmented and transformed into 2D TF spectrogram using SLT. Then, these are applied to different residual pretrained network. Based on extracted features from pretrained network, the classifier identifies the label of the class.

The superlet transform offers a significant advantage in EEG signal analysis by enhancing time-frequency resolution through adaptive and variable bandwidths, enabling the capture of subtle oscillatory patterns and temporal variations critical for EEG classification. Similarly, the ResNet model, with its residual connections, addresses the vanishing gradient problem, allowing deeper architectures to efficiently learn complex features inherent in EEG signals. This combination improves the overall efficiency by leveraging precise feature extraction from the superlet transform and robust learning capabilities of ResNet, resulting in superior classification performance and faster convergence. Figure 3 shows the step-by-step approach used in the article.

A. SUPERLET TRANSFORM

STFT and CWT are established techniques for analysing the TF characteristics of signals. However, both approaches involve a trade-off between time and frequency resolution. The superlet transform (SLT) overcomes this limitation by employing a set of wavelets, offering improved TF resolution and reduced ''leakage'' compared to a single wavelet [33]. Table 1 summarizes the key advantages and disadvantages of each method, highlighting their suitability for different signal characteristics, resolution requirements, and desired representations. SLT provides a new spectral estimation which provides the information on high frequency components

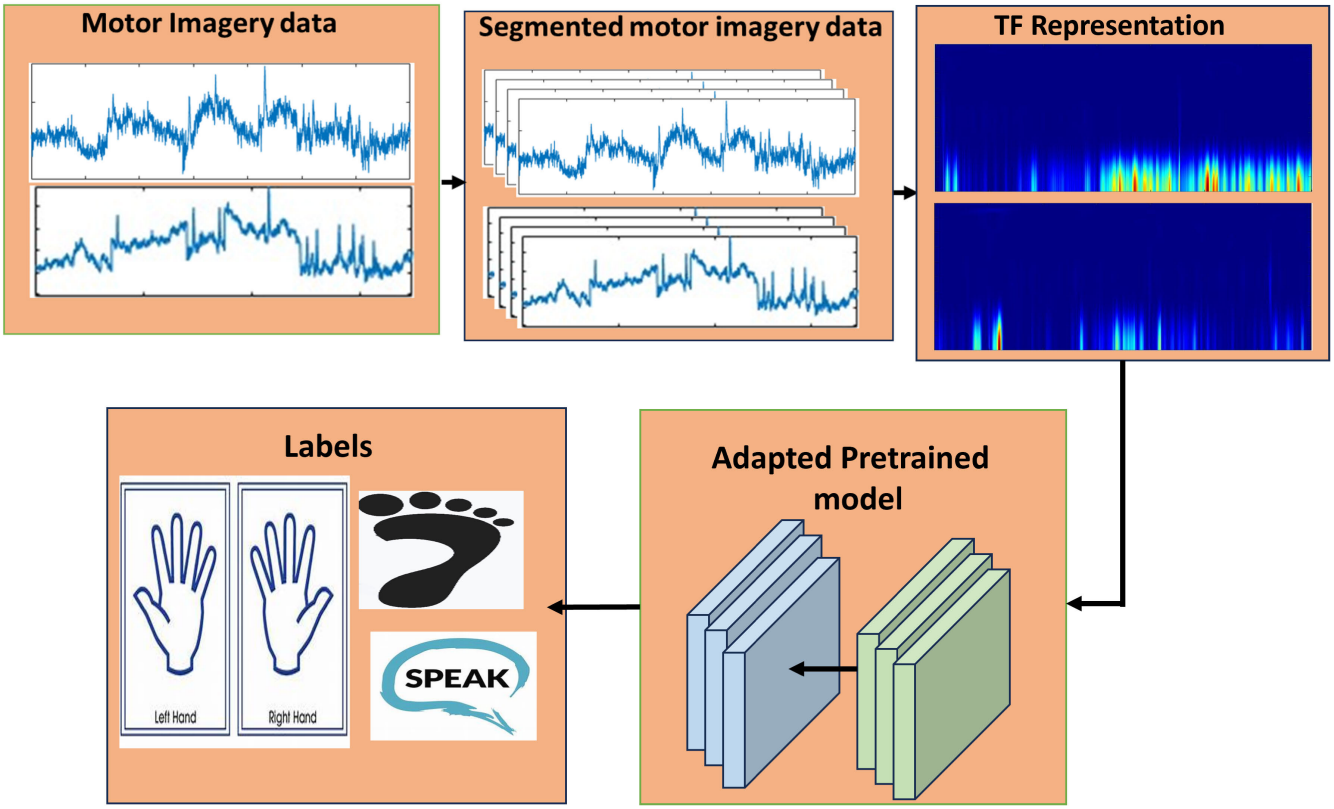

FIGURE 3. Pipeline of the suggested approach.

of signal [34]. SLT intrinsically enhances signal clarity by suppressing noise through the integration of responses from multiple wavelets. This noise suppression occurs naturally as a result of the combined wavelet responses. SLT offers superior adaptability and enhanced time-frequency resolution and has a relatively high computational complexity compared to traditional methods such as STFT and CWT. SLT's multi-scale nature, which allows for better signal analysis at various frequencies, requires multiple convolutions and scales to be computed simultaneously. This increased processing demands more computational resources and time, particularly when working with large datasets or real-time applications. The added computational overhead is a key consideration in practical deployment, and efforts to optimize SLT implementations for more efficient computation will be important for its wider adoption in resource-constrained and real-time environments. In our research, we employ the Morlet wavelet as the foundational mother wavelet. The Superlet Transform of a signal $g(t)$ can be mathematically represented as the geometric mean of the responses from the individual wavelets, which is defined as follows:

$$A\left[SL_{f,k}\right] = \sqrt[k]{\prod_{i=1}^{k}\left|\frac{\sqrt{2}}{s}\int_{-\infty}^{\infty} g(\tau)\,\psi^{*}_{f,c_i}\left(\frac{\tau - t}{s}\right)d\tau\right|} \quad (1)$$

The integration of multiple wavelets enhances TF representation in SL. The parameters t and s are scaling and shifting parameters in the signal. SL constitutes a finite group of wavelets sharing the same central frequency f and extending across multiple bandwidths with $c_1, c_2, \ldots, c_k$ indicating the number of cycles in the individual wavelets.

$$SL_{(f,k)} = \left[\psi_{(f,k)} | c = c_1, c_2, c_3, \ldots\ldots c_k\right] \quad (2)$$

where k represents the order of SL, and $\psi_{c_i}(t)$ is the i^{th} order wavelet that is given by:

$$\psi_{c_i}(t) = \frac{\sigma \cdot f}{c \cdot i \cdot \sqrt{2\pi}}\left(e^{-\frac{1}{2}\left[\frac{\sigma f t}{c \cdot i}\right]^2}\right)e^{j2\Pi f t} \quad (3)$$

where σ is the standard deviation in k^{th} order SL.

B. CNN MODELS

CNN models consist of feature extraction layers, such as convolutional layers, activation layer, batch normalization layers, and pooling layers, that learn to extract features and patterns from the input data and classification layer such as fully connected layers, softmax layers, and classification layer that gives the final decision of the input data [35]. In this study, we leverage the concept of transfer learning by utilizing the pre-trained ResNet architecture for the task of motion recognition. The ingenious design of ResNet lies in its utilization of residual connections, which significantly enhances the learning capability of deep neural networks by focusing on residual functions rather than direct mappings.

TABLE 1. Difference between various TFR methods.

Feature	STFT	CWT	SLT
Basic Concept	It divides the signal into short segments and applies the Fourier Transform to each.	It uses wavelets to analyze signals at various scales and positions.	It combines multiple wavelets to enhance TF representation.
Resolution	Fixed time and frequency resolution.	Variable time and frequency resolution.	Offers improved TF resolution by layering multiple wavelets.
TF Representation	Good frequency resolution but poor time resolution for given window size	Good time resolution but poor frequency resolution for given window size	Provides a detailed and refined TF representation as the output is the geometric mean of all the wavelets.
Complexity	Relatively simple and computationally efficient.	More complex and computationally intensive due to the need for multiple scales.	More complex, involving the combination of various wavelets at different scales.

This strategic approach effectively mitigates the issue of gradient vanishing or exploding, thus preserving the integrity of the learning process. The ResNet framework is distinguished by its incorporation of MaxPool and Average Pool layers. Specifically, ResNet 101, a deep convolutional neural network featuring 101 layers, exemplifies the advanced iteration of ResNet. It employs residual blocks to master residual functions, thereby overcoming learning degradation due to gradient issues. Structurally, ResNet 101 is organized into 33 convolutional layers divided into four distinct sets, characterized by varying filter counts and repetition frequencies. The sequence begins with a set of 64 filters repeated thrice, progresses to 128 filters repeated four times, advances to 256 filters repeated 23 times, and concludes with 512 filters repeated thrice. After every convolutional layer, there is a batch normalization layer and a ReLU activation function, guaranteeing effective and nonlinear processing. The architecture initiates with a 64-filter convolutional layer with a 7×7 kernel size, culminating in a fully connected layer equipped with 1000 nodes and a softmax function for classification. Capable of categorizing images into 1000 different objects, ResNet 101 demonstrates remarkable accuracy across a plethora of benchmark datasets and various computer vision challenges. Figure 4 represents the layered structure of ResNet 101.

C. EVALUATION METRICS

System performance can be assessed using a variety of performance metrics, such as accuracy, sensitivity, specificity, etc. A true positive (TP) correctly determines the presence of a condition or characteristic. A false positive (FP) claims erroneously that a condition or characteristic exists.TN is one that accurately identifies the absence of a condition. FN claims erroneously that a condition or characteristic does not exist [36], [37]. In Fig.5, the confusion matrix for the binary class is depicted.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Senstivity = \frac{TP}{TP + FN} \tag{5}$$

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$F - measure = \frac{2TP}{FN + FP + 2TP} \tag{8}$$

IV. RESULTS AND DISCUSSION

In this section, the performance of three residual pretrained networks is evaluated. Notably, ResNet 101 achieved the highest classification accuracy, reaching 99.9% for the binary dataset and 96.4% for the multiclass dataset. Throughout our experiments, the ADAM optimizer with a learning rate of 0.00001 was consistently employed. It's noteworthy that, for the purpose of this study, all samples related to foot and right-hand motion were amalgamated across all five subjects. The data were segmented into 3.5-second intervals for binary classification, while for the multiclass dataset, samples were combined and then segmented into 4-second intervals. The time-frequency representation (TFR) superlet transform was subsequently applied to the segmented data. Subsequently, the dataset was divided into training, validation, and test sets in an 80:20 ratio. The 2-D representation resulting from these processes served as input for various ResNet architectures.

The ResNet18 architecture demonstrated an accuracy of 82.8% in recognizing hand and foot motion, as illustrated in Figure 6 (a). The corresponding confusion matrix for the test dataset revealed that out of 1,595 hand and foot samples, 289 foot and 259 hand samples were misclassified, resulting in the overall test accuracy of 82.8%. In the case of the ResNet50 architecture Figure 6 (b), an impressive accuracy of 99.4% was achieved for hand and foot motion recognition. The confusion matrix indicated only 10 foot and 9 hand samples were misclassified out of the 1,595 samples in the test dataset. Furthermore, the performance of the ResNet101 architecture Figure 6 (c) reached an outstanding accuracy of 99.9%. In this case, all 1,595 foot samples were correctly identified, and out of the 1,595 hand samples, 1,592 were accurately classified as hand samples, with only 3 samples misclassified as foot samples.

The ResNet18 model achieved a notable accuracy rate of 95.3% in identifying MI-EEG signals, as depicted in Figure 7 (a). The analysis of the test dataset's confusion matrix shows accurate classification of 6017 LH, 6429 RH, 6317 F, and 5736 T samples, culminating in an overall test

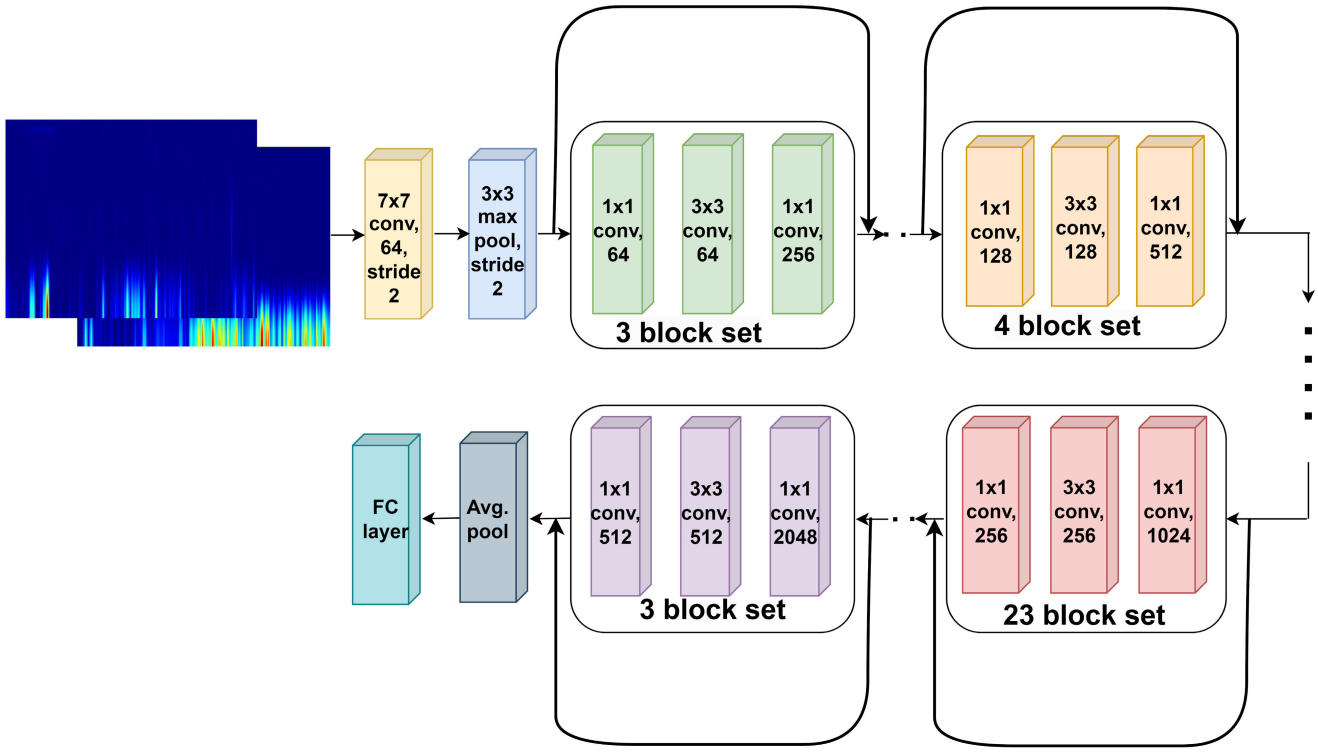

FIGURE 4. Architecture of Resnet 101.

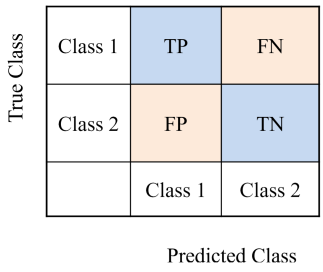

FIGURE 5. Confusion matrix.

accuracy of 95.3%. For the ResNet50 model, presented in Figure 7 (b), the accuracy slightly improved to 95.9% for MI-EEG signal classification. According to its confusion matrix, there were 324 LH, 230 F, and 342 T samples incorrectly classified, while RH samples were identified with perfect accuracy. The ResNet101 model, illustrated in Figure 7 (c), further enhanced the performance, reaching an accuracy of 96.4%. Misclassifications included 333 LH samples as T and RH, alongside 10 RH, 254 F, and 327 T samples incorrectly classified, showcasing the nuanced improvements in MI-EEG signal recognition across different ResNet architectures.

Simulation outcomes for the test datasets, employing the Superlet Transform (SLT) across the three ResNet models (ResNet 101, ResNet 50, and ResNet 18), are concisely summarized in Tables 2 and 3, covering both datasets. Out Of these, the ResNet 101 model exhibits superior performance,

TABLE 2. Detection summary of models using SLT and deep learning model for binary class dataset.

Evaluation Parameters	ResNet-18	ResNet-50	ResNet-101
Accuracy (%)	82.8	99.4	99.9
Recall (%)	83.7	99.4	99.8
Specificity (%)	81.8	99.3	100
F-1	.82	.99	.99

TABLE 3. Detection summary of ResNet-101 using SLT and deep learning model for multi-class dataset.

Class	Performance measures			
	Accuracy (%)	Precision (%)	Recall (%)	F-1
Class1 (LH)	97.9	94.8	96.8	0.96
Class2 (RH)	99.9	99.8	99.9	1
Class3 (F)	98.5	96.1	97.9	0.97
Class4 (T)	96.4	94.7	90.9	0.93

achieving an accuracy rate of 99.9% for the binary class dataset and 96.4% for the multiclass dataset.

As shown in Figure 8, SLT proves to be a more reliable and effective method for accurately identifying the MI-EEG signals. As compared to CWT and STFT, SLT has attained better classification accuracy. SLT achieves an impressive error or misclassification rate of only 0.1%, which is

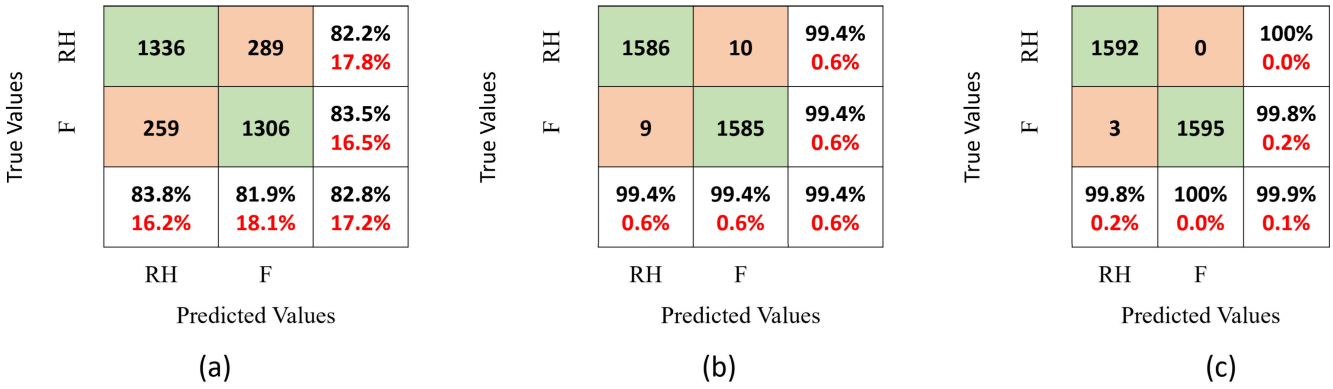

FIGURE 6. Confusion matrix of pretrained ResNet model for binary class dataset.

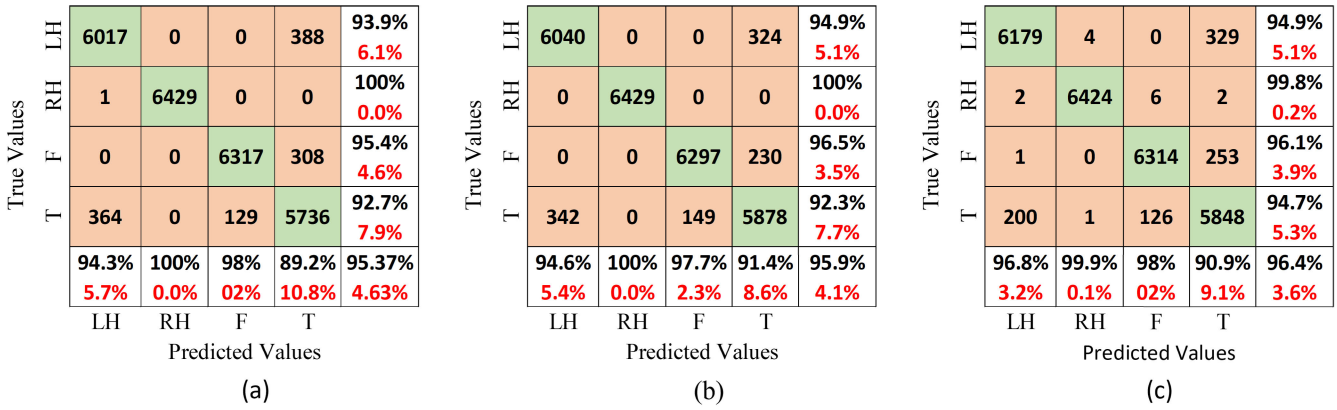

FIGURE 7. Confusion matrix of pretrained ResNet model for multi class dataset.

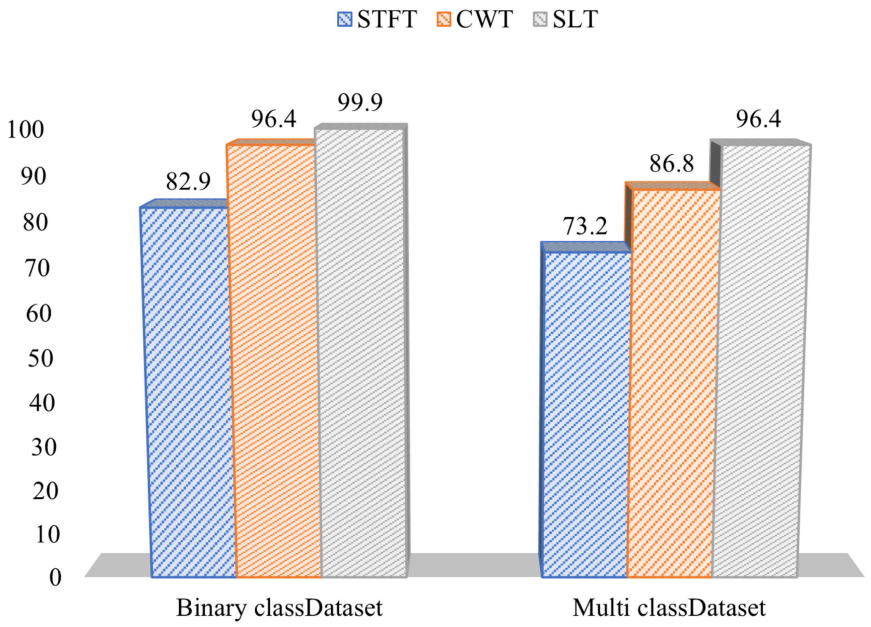

FIGURE 8. Classification accuracy (%) of different of TFR methods.

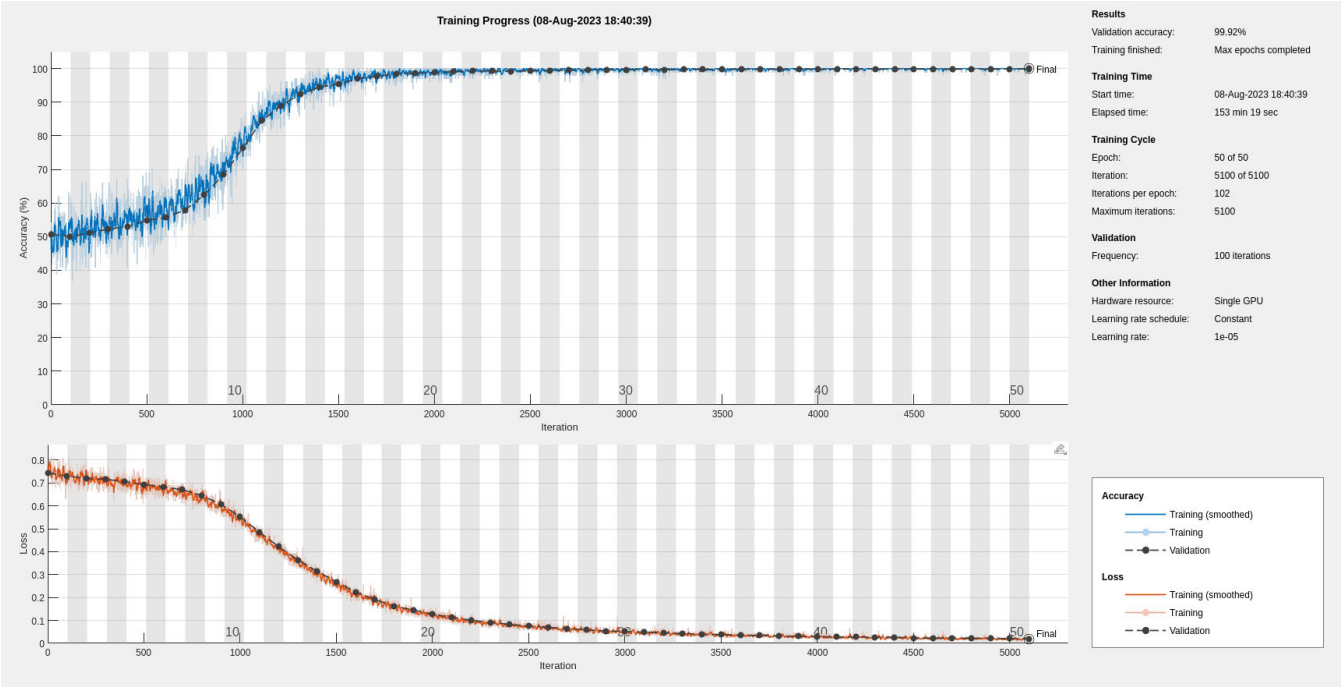

FIGURE 9. Traininng plot for binary class data for Resnet 101.

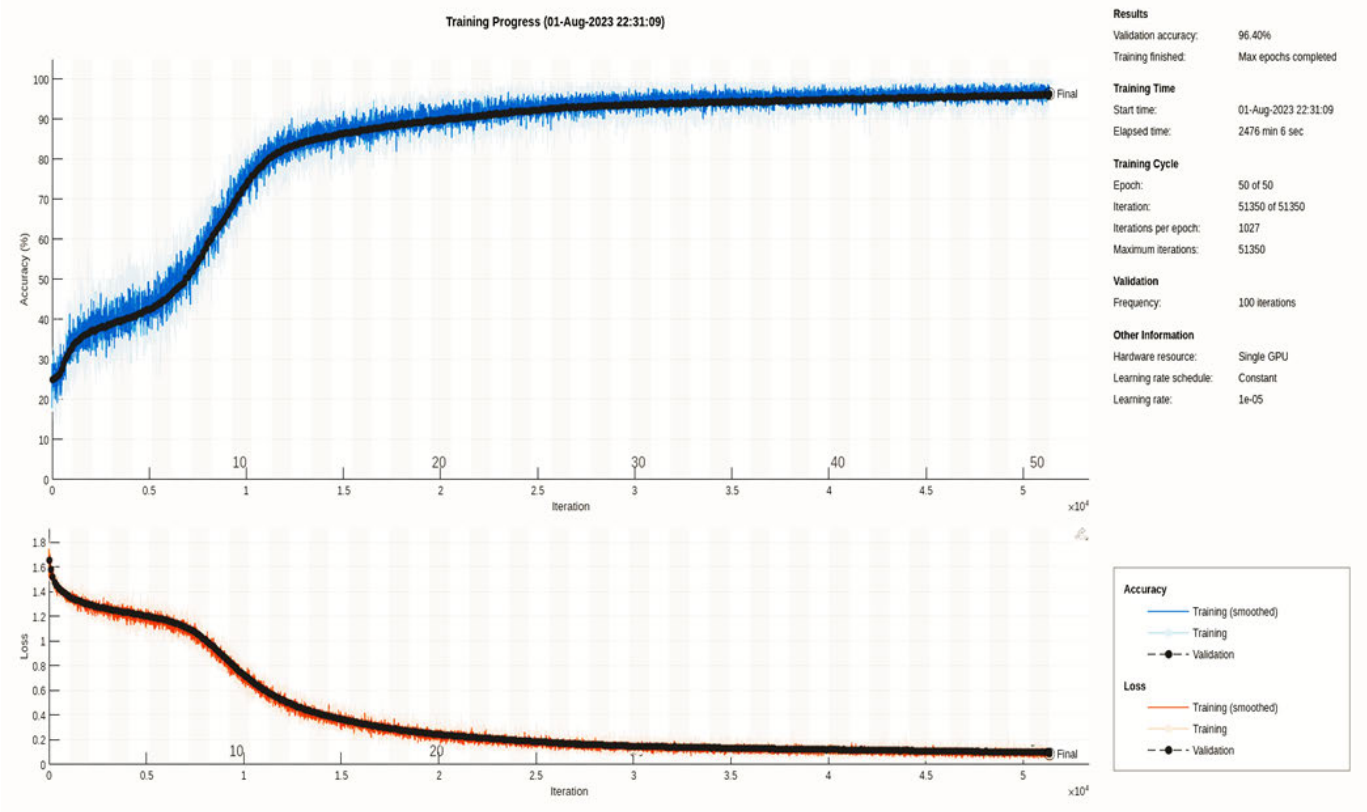

FIGURE 10. Traininng plot for multiclass data for Resnet 101.

almost negligible. Compared to STFT and CWT, which have

TABLE 4. Comparative analysis of the proposed method and current leading-edge techniques for binary classification datasets.

S. No	Author	Methodology	Accuracy					
			aa (%)	al (%)	av (%)	aw (%)	ay (%)	Avg. (%)
1.	Xiong et al. [13]	EA	82	94	70	87	95	85.6
2.	Kervic et al. [38]	WPD	96	92.3	88.9	95.4	91.4	92.8
3.	Sadiq et al. [39]	EWT	99.4	98.7	92.2	85.7	100	95.3
4.	Rashid et al. [40]	CSP	94.6	94.6	95	93.2	90.1	93.6
Proposed Methodology			**100**	**100**	**99.8**	**99.9**	**100**	**99.9**

TABLE 5. Comparative analysis of the proposed method and current leading-edge techniques for multiclass classification datasets.

S. No.	Author	Methodology	Input Formulation	Accuracy (%)
1.	Chaudhary et al. [21]	WT	2-D (image)	85.6
2.	Xu et al. [41]	NFBCSP	Time -series	76.8
3.	Sakhavi et al. [42]	CNN+FBCSP	2-D (image)	74
4.	Riyad et al. [43]	EEGNet	2-D (image)	74
Proposed methodology				**96.4**

classification accuracies of 82.9% and 96.4% respectively, SLT outperforms both methods by attaining an accuracy of 99.9% and 96.4%.

For binary class data, the training process for ResNet 101 is represented by Figure 9 which was conducted over 50 epochs with a total of 5,100 iterations, averaging 102 iterations per epoch. The training duration was 153 minutes and 19 seconds with a constant learning rate of 1e-5. By the end of the training, the model achieved an impressive validation accuracy of 99.92%, indicating strong performance on the validation set. The accuracy curve shows a steady increase from around 50% to almost 100%, with the model converging around iteration 3,500. Similarly, the loss decreased consistently from approximately 0.7 to just above 0.1. Validation occurred every 100 iterations, and both accuracy and loss curves demonstrate stable and effective model training across the entire process.

For multi class data, the training process for ResNet 101 is represented by Figure 10 which was conducted over 50 epochs, with a total of 51,350 iterations, averaging 1,027 iterations per epoch. The training duration was 2476 minutes with a constant learning rate of 1e-5. The validation accuracy achieved was 96.40%, indicating that the model performed well on unseen data. The accuracy steadily improved throughout the training from around 20% to 96%, while the loss decreased from approximately 1.6 to below 0.2. Validation was performed every 100 iterations, and both the accuracy and loss curves show a consistent convergence over time, confirming the model's stability and effectiveness.

Tables 4 and 5 compare the effectiveness of the suggested model with leading-edge methods across binary and multiclass datasets. Kervic and Subasi employed WPD on a constrained sample set, securing a 92.8% accuracy [38]. Sadiq et al. explored 2D modelling with EWT, achieving a 95.3% classification accuracy [39]. Utilizing CSP for feature extraction, Rashid et al. reported a 93.6% accuracy [40]. A strategy involving Euclidean alignment (EA) with either an LDA or LR classifier by Xiong and Wei yielded an 85.6% accuracy [13]. Analysing signals with CSP, Xu et al. and Sakhavi et al. reported accuracies of 76.8% and 74%, respectively [41], [42]. Chaudhary and Agrawal's application of wavelet transform resulted in an 85.6% accuracy [21], while Riyad et al., utilizing EEGNet, achieved 74% [43]. The proposed methodology achieved notably higher accuracy when compared to Leading-edge techniques as clearly demonstrated in the Table 4 and 5.

V. CONCLUSION

In our research, we introduced the application of the Superlet Transform (SLT) for analysing motor imagery EEG data in time-frequency space and assessed the efficacy of three distinct residual CNN models for both binary and multiclass classification tasks. Utilizing transfer learning techniques on a pretrained network, we evaluated model performance through accuracy, sensitivity, F-1 score, and precision metrics, derived from confusion matrices on test datasets. Our comparative analysis spanned model depth, layer count, parameter volume, approaches for TF signal representation, along with training and evaluation durations. Findings revealed that SLT-enhanced feature extraction notably boosts classification outcomes over current leading methods, with the residual CNN architectures showing superior accuracy rates. Specifically, ResNet 101 stood out, delivering an exceptional 99.9% accuracy for binary

classifications and 96.4% for multiclass dataset. A key consideration for achieving such high accuracy involved the optimal selection of order and time spread parameters for the SLT process.

By incorporating time-frequency (TF) representation, researchers can achieve a deeper and more precise understanding of the intricate, dynamic changes in brain activity across both temporal and spectral domains. This approach facilitates the identification of subtle patterns linked to distinct motor imagery tasks, enhancing analytical accuracy. Moreover, the integration of pretrained networks through transfer learning accelerates the analytical process by leveraging extensive knowledge from large-scale datasets, overcoming the constraints of traditional methods. This powerful combination has the potential to drive significant advancements in brain-computer interface (BCI) technologies, enabling more intuitive and adaptive control systems for assistive devices, thus greatly enhancing the quality of life for individuals with mobility challenges. Additionally, it paves the way for innovations in neurorehabilitation, cognitive neuroscience, and neurofeedback applications, fostering broader progress in the understanding and application of neural mechanisms.

ACKNOWLEDGMENT

The authors would like to express sincere gratitude to Intelligent Prognostic Private Limited Delhi, India for funding this research work. The authors extend their appreciation to the SGT University, India, Bennett University, India, Netaji Subhas University of Technology, India, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India, Universiti Sultan Zainal Abidin (UniSZA) Malaysia for providing research facility.

REFERENCES

[1] E. A. Mohamed, M. Z. Yusoff, A. S. Malik, M. R. Bahloul, D. M. Adam, and I. K. Adam, "Comparison of EEG signal decomposition methods in classification of motor-imagery BCI," *Multimedia Tools Appl.*, vol. 77, no. 16, pp. 21305–21327, Aug. 2018.

[2] R. Janapati, V. Dalal, and R. Sengupta, "Advances in modern EEG-BCI signal processing: A review," *Mater. Today, Proc.*, vol. 80, pp. 2563–2566, Jan. 2023.

[3] N. Sharma, M. Sharma, A. Singhal, R. Vyas, H. Malik, A. Afthanorhan, and M. A. Hossaini, "Recent trends in EEG-based motor imagery signal analysis and recognition: A comprehensive review," *IEEE Access*, vol. 11, pp. 80518–80542, 2023.

[4] A. Kawala-Sterniuk, M. Pelc, R. Martinek, and G. M. Wójcik, "Editorial: Currents in biomedical signals processing—Methods and applications," *Frontiers Neurosci.*, vol. 16, Jul. 2022, Art. no. 989400.

[5] N. Bajaj, J. R. Carrión, F. Bellotti, R. Berta, and A. De Gloria, "Automatic and tunable algorithm for EEG artifact removal using wavelet decomposition with applications in predictive modeling during auditory tasks," *Biomed. Signal Process. Control*, vol. 55, Jan. 2020, Art. no. 101624.

[6] H. Göksu, "BCI oriented EEG analysis using log energy entropy of wavelet packets," *Biomed. Signal Process. Control*, vol. 44, pp. 101–109, Jul. 2018.

[7] V. Gupta, T. Priya, A. K. Yadav, R. B. Pachori, and U. R. Acharya, "Automated detection of focal EEG signals using features extracted from flexible analytic wavelet transform," *Pattern Recognit. Lett.*, vol. 94, pp. 180–188, Jul. 2017.

[8] C. Uyulan, "Development of LSTM&CNN based hybrid deep learning model to classify motor imagery tasks," *Commun. Math. Biol. Neurosci.*, vol. 2021, pp. 1–26, Jan. 2021, doi: 10.28919/cmbn/5265.

[9] A. Echtioui, A. Mlaouah, W. Zouch, M. Ghorbel, C. Mhiri, and H. Hamam, "A novel convolutional neural network classification approach of motor-imagery EEG recording based on deep learning," *Appl. Sci.*, vol. 11, no. 21, p. 9948, Oct. 2021.

[10] M. Dai, D. Zheng, S. Liu, and P. Zhang, "Transfer kernel common spatial patterns for motor imagery brain–computer interface classification," *Comput. Math. Methods Med.*, vol. 2018, no. 1, pp. 1–9, 2018.

[11] K. Darvish Ghanbar, T. Yousefi Rezaii, A. Farzamnia, and I. Saad, "Correlation-based common spatial pattern (CCSP): A novel extension of CSP for classification of motor imagery signal," *PLoS ONE*, vol. 16, no. 3, Mar. 2021, Art. no. e0248511.

[12] T. Mwata-Velu, J. Ruiz-Pinales, J. G. Avina-Cervantes, J. J. Gonzalez-Barbosa, and J. L. Contreras-Hernandez, "Empirical mode decomposition and a bidirectional LSTM architecture used to decode individual finger MI-EEG signals," *J. Adv. Appl. Comput. Math.*, vol. 9, pp. 32–48, May 2022.

[13] W. Xiong and Q. Wei, "Reducing calibration time in motor imagery-based BCIs by data alignment and empirical mode decomposition," *PLoS ONE*, vol. 17, no. 2, Feb. 2022, Art. no. e0263641.

[14] J. Gilles, "Empirical wavelet transform," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3999–4010, Aug. 2013.

[15] A. Bhattacharyya, L. Singh, and R. B. Pachori, "Fourier–Bessel series expansion based empirical wavelet transform for analysis of non-stationary signals," *Digit. Signal Process.*, vol. 78, pp. 185–196, Jul. 2018.

[16] P. Singh, S. D. Joshi, R. K. Patney, and K. Saha, "The Fourier decomposition method for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 473, no. 2199, Mar. 2017, Art. no. 20160871.

[17] N. Sharma, M. Sharma, A. Singhal, R. Vyas, H. Malik, M. A. Hossaini, and A. Afthanorhan, "An efficient approach for recognition of motor imagery EEG signals using the Fourier decomposition method," *IEEE Access*, vol. 11, pp. 122782–122791, 2023.

[18] M. Z. Baig, N. Aslam, H. P. H. Shum, and L. Zhang, "Differential evolution algorithm as a tool for optimal feature subset selection in motor imagery EEG," *Expert Syst. Appl.*, vol. 90, pp. 184–195, Dec. 2017.

[19] P. Kant, S. H. Laskar, J. Hazarika, and R. Mahamune, "CWT based transfer learning for motor imagery classification for brain computer interfaces," *J. Neurosci. Methods*, vol. 345, Nov. 2020, Art. no. 108886.

[20] R. Zhang, Q. Zong, L. Dou, X. Zhao, Y. Tang, and Z. Li, "Hybrid deep neural network using transfer learning for EEG motor imagery decoding," *Biomed. Signal Process. Control*, vol. 63, Jan. 2021, Art. no. 102144.

[21] P. Chaudhary and R. Agrawal, "Non-dyadic wavelet decomposition for sensory-motor imagery EEG classification," *Brain-Comput. Interfaces*, vol. 7, nos. 1–2, pp. 11–21, Apr. 2020.

[22] Y. Zhang, Y. Liu, W. Kang, and R. Tao, "VSS-Net: Visual semantic self-mining network for video summarization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 4, pp. 2775–2788, Apr. 2024.

[23] Y. Zhang, C. Wu, W. Guo, T. Zhang, and W. Li, "CFANet: Efficient detection of UAV image based on cross-layer feature aggregation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5608911.

[24] J. Xie, J. Zhang, J. Sun, Z. Ma, L. Qin, G. Li, H. Zhou, and Y. Zhan, "A transformer-based approach combining deep learning network and spatial–temporal information for raw EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2126–2136, 2022.

[25] Y. Zhang, Y. Liu, and C. Wu, "Attention-guided multi-granularity fusion model for video summarization," *Expert Syst. Appl.*, vol. 249, Sep. 2024, Art. no. 123568.

[26] N. Sharma, A. Upadhyay, M. Sharma, and A. Singhal, "Deep temporal networks for EEG-based motor imagery recognition," *Sci. Rep.*, vol. 13, no. 1, p. 18813, Nov. 2023.

[27] Y. Ma, Y. Song, and F. Gao, "A novel hybrid CNN-transformer model for EEG motor imagery classification," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2022, pp. 1–8.

[28] Y. Zhang, T. Zhang, C. Wu, and R. Tao, "Multi-scale spatiotemporal feature fusion network for video saliency prediction," *IEEE Trans. Multimedia*, vol. 26, pp. 4183–4193, 2024.

[29] N. Mammone, C. Ieracitano, and F. C. Morabito, "A deep CNN approach to decode motor preparation of upper limbs from time–frequency maps of EEG signals at source level," *Neural Netw.*, vol. 124, pp. 357–372, Apr. 2020.

[30] H. Li, M. Ding, R. Zhang, and C. Xiu, "Motor imagery EEG classification algorithm based on CNN-LSTM feature fusion network," *Biomed. Signal Process. Control*, vol. 72, Feb. 2022, Art. no. 103342.
[31] B. Blankertz, K.-R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, J. R. Millán, M. Schröder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, Jun. 2006.
[32] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, pp. 6–55, Jul. 2012.
[33] V. V. Moca, H. Bârzan, A. Nagy-Dăbâcan, and R. C. Mureşan, "Time-frequency super-resolution with superlets," *Nature Commun.*, vol. 12, no. 1, p. 337, Jan. 2021.
[34] P. M. Tripathi, A. Kumar, M. Kumar, and R. Komaragiri, "Multilevel classification and detection of cardiac arrhythmias with high-resolution superlet transform and deep convolution neural network," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–13, 2022.
[35] S. Chaudhary, S. Taran, V. Bajaj, and A. Sengur, "Convolutional neural network based approach towards motor imagery tasks EEG signals classification," *IEEE Sensors J.*, vol. 19, no. 12, pp. 4494–4500, Jun. 2019.
[36] T. N. Alotaiby, S. A. Alshebeili, T. Alshawi, I. Ahmad, and F. E. A. El-Samie, "EEG seizure detection and prediction algorithms: A survey," *EURASIP J. Adv. Signal Process.*, vol. 2014, no. 1, p. 183, Dec. 2014.
[37] P. M. Tripathi, A. Kumar, R. Komaragiri, and M. Kumar, "A review on computational methods for denoising and detecting ECG signals to detect cardiovascular diseases," *Arch. Comput. Methods Eng.*, vol. 29, no. 3, pp. 1875–1914, Oct. 2021.
[38] J. Kevric and A. 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, Jan. 2017.
[39] M. T. Sadiq, X. Yu, Z. Yuan, and M. Z. Aziz, "Motor imagery BCI classification based on novel two-dimensional modelling in empirical wavelet transform," *Electron. Lett.*, vol. 56, no. 25, pp. 1367–1369, Dec. 2020.
[40] M. Rashid, B. S. Bari, M. J. Hasan, M. A. M. Razman, R. M. Musa, A. F. A. Nasir, and A. P. P. A. Majeed, "The classification of motor imagery response: An accuracy enhancement through the ensemble of random subspace k-NN," *PeerJ Comput. Sci.*, vol. 7, Mar. 2021, Art. no. e374.
[41] S. Xu, L. Zhu, W. Kong, Y. Peng, H. Hu, and J. Cao, "A novel classification method for EEG-based motor imagery with narrow band spatial filters and deep convolutional neural network," *Cognit. Neurodyn.*, vol. 16, no. 2, pp. 379–389, Apr. 2022.
[42] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain–computer interface using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5619–5629, Nov. 2018.
[43] M. Riyad, M. Khalil, and A. Adib, "MI-EEGNET: A novel convolutional neural network for motor imagery classification," *J. Neurosci. Methods*, vol. 353, Apr. 2021, Art. no. 109037.