Applied Data Science and Analysis
Vol.2023, **pp**. 88–101
DOI: https://doi.org/10.58496/ADSA/2023/007 ; ISSN: xxx-xxxx
https://mesopotamian.press/journals/index.php/ADSA

Research Article

Multi-Tiered CNN Model for Motor Imagery Analysis: Enhancing UAV Control in Smart City Infrastructure for Industry 5.0

Z.T. Al-Qaysi[1,*], Mahmood M. Salih[1], Moceheb Lazam Shuwandy[1], M.A. Ahmed [1], Yazan S.M. Altarazi[2]

[1] *Department of Computer Science, Computer Science and Mathematics College, Tikrit University, Tikrit, Iraq.*

[2] *Aerodynamic, Heat Transfer & Propulsion Group, Department of Aerospace Engineering, University Putra Malaysia, Malaysia.*

ARTICLE INFO

Article History
Received 17 Jul 2023
Accepted 23 Aug 2023
Published 20 Sep 2023

Keywords
Brain-Computer Interface
Deep Learning
Motor Imagery
Transfer Learning
VGG-19
Neural Network Classifier
UAV
Short-time Fourier transform

ABSTRACT

The concept of brain-controlled UAVs, pioneered by researchers at the University of Minnesota, initiated a series of investigations. These early efforts laid the foundation for more advanced prototypes of brain-controlled UAVs. However, BCI signals are inherently complex due to their nonstationary and high-dimensionality nature. Therefore, it is crucial to carefully consider both feature extraction and the classification process. This study introduces a novel approach, combining a pretrained CNN with a classical neural network classifier and STFT spectrum, into a Multi-Tiered CNN model (MTCNN). The MTCNN model is applied to decode two-class Motor Imagery (MI) signals, enabling the control of UAV up/down movement. The experimental phase of this study involved four key experiments. The first assessed the MTCNN model's performance using a substantial dataset, resulting in an impressive classification accuracy of 99.1%. The second and third experiments evaluated the model on two different datasets for the same subjects, successfully addressing challenges associated with inter-subject and intra-subject variability. The MTCNN model achieved a remarkable classification accuracy of 99.7% on both datasets. In a fourth experiment, the model was validated on an additional dataset, achieving classification accuracies of 100% and 99.6%. Remarkably, the MTCNN model surpassed the accuracy of existing literature on two BCI competition datasets. In conclusion, the MTCNN model demonstrates its potential to decode MI signals associated with left- and right-hand movements, offering promising applications in the field of brain-controlled UAVs, particularly in controlling up/down movements. Furthermore, the MTCNN model holds the potential to contribute significantly to the BCI-MI community by facilitating the integration of this model into MI-based UAV control systems.

1. INTRODUCTION

In the last decade, Unmanned Aerial Vehicles (UAVs) have garnered increasing attention from the research community[1]. They represent a pivotal technological advancement that has seen a surge in interest. Moreover, the growing importance of drones in various applications has been remarkable due to their unparalleled role in executing aerial operations, especially in situations where piloted flights are unfeasible[2]. Notably, studies such as [29] have endeavored to develop brain-controlled Assistive Devices (AD) to empower individuals in exploring their environment using a computer and their thoughts. Leveraging the potential of a noninvasive Steady-State Visual Evoked Potential (SSVEP)-based Brain-Computer Interface (BCI) system, users can navigate a flying robot, commonly known as a UAV or drone, within 3D physical space. Within the domain of BCI-based UAV control, EEG signal modalities such as P300 evoked potentials, SSVEP, and Motor Imagery (MI) have emerged as prominent contenders [3]. However, it's essential to note that MI is distinct as it relies solely on spontaneous potential and doesn't necessitate external stimulation [4]. As a result, researchers have harnessed MI signals to aid individuals with disabilities in managing various equipment, such as robots and even self-driving cars [5]. Imagine being able to control a part of your body without physically moving it – this is the essence of MI [6]. EEG signals are generated by both real and imagined human movements. In motor imagery, EEG signals exhibit event-related

*Corresponding author. Email: ziadoontareq@tu.edu.iq

synchronization (ERS) and event-related desynchronization (ERD) features [7]. Each hemisphere of the human brain consists of four lobes, each with a unique function. Fissures divide these ear-like lobes (sulcus). In the context of the BCI system, the primary somatosensory cortex (parietal lobe) and the primary motor cortex (temporal lobe) are of paramount significance [8]. Mu and beta rhythms within the sensorimotor region of one hemisphere show a decrease or increase as individuals imagine or execute unilateral limb movements. These variations are known as event-related desynchronization (ERD) and event-related synchronization (ERS) [9].

MI-based Brain-Computer Interface (BCI) pattern recognition systems fundamentally involve three essential processes: preprocessing of the EEG signal, feature extraction, and classification [10]. Among these, feature extraction is the pivotal process in the MI EEG pattern recognition model. Extracted features serve the purpose of streamlining data processing by identifying the most relevant feature components within the signal. The feature extraction process can be executed in various signal processing domains, including the spatial domain, time domain, frequency domain, and time-frequency domain [10]. Notably, time-frequency representation (TFR) of MI features is widely employed for classification in BCI applications. TFR describes the signal's energy density and intensity at different time and frequency points by constructing a joint function of time and frequency [11]. Given that MI signals are 1-D in nature, methods like Continuous Wavelet Transform (CWT) and STFT are frequently utilized to convert the signal into a 2-D image. These methods are efficient and adept at capturing signal characteristics in both the time and frequency domains [12]. However, when the time interval is relatively short, EEGs may resemble a non-stationary signal. In such cases, the STFT method proves effective for extracting and computing the spectrum of the brain signal in the time-frequency domain [13]. Furthermore, STFT provides comprehensive information about the time-frequency domain while incurring a low computational cost [14].

In the same context, deep feature extraction using deep neural networks (DNNs) has recently demonstrated remarkable categorization capabilities across various applications, including computer vision, video processing, and speech recognition. The profound success of deep neural networks has inspired numerous academics to explore their potential in categorizing EEG signals [15]. In the field of Brain-Computer Interfaces (BCI), many researchers have begun to integrate deep learning into their applications, spanning from seizure detection and memory retrieval to MI categorization [16].

Convolutional Neural Networks (CNNs) have shown their ability to extract spatial and temporal features from MI data. They can effectively extract valuable features from both shallow and deep models, suggesting that significant features can be retrieved at various levels [5]. Nevertheless, one of the major challenges in categorizing MI EEG features using deep learning techniques is the limited availability of data, mainly due to patient fatigue during experiments [17]. Additionally, substantial individual variations among different subjects make it challenging to directly apply labeled data from other subjects to train a classifier for identifying specific individuals [18]. Furthermore, EEG data collection is costly, and obtaining a sufficient quantity of labeled samples is often a formidable task [19]. To address the issue of combining data from domains with different distributions, transfer learning has emerged as a valuable approach. Transfer learning encompasses techniques designed to transfer representations and knowledge from one domain to another [20]. In essence, it allows researchers to seamlessly incorporate new datasets into an already trained machine learning model. This capability is especially advantageous in BCI systems, where the available data is often insufficient for robust model training [21]. Notably, in BCI studies, the CNN-based subject-transfer technique has proven to outperform other methods. Subject-transfer strategies operate on the premise that the typical patterns of the target subject may be comparable to those of other subjects when performing the same activity [22].

The choice of a classifier significantly impacts the discrimination between two MI EEG mental commands, making classifier selection a critical aspect of the study. Traditional machine learning methods often rely on hand-crafted features for classification. In contrast, Deep Convolutional Neural Networks (DCNN) have the ability to perform classification by directly extracting features from raw data [23]. Previous studies, such as [24, 25], have used pretrained CNN models in conjunction with classical machine learning algorithms for classifying computer vision problems. Similarly, a study like [26] applied a similar technique for detecting epileptic seizures.

In this research, a hybrid approach was employed, combining a pretrained CNN with a classical neural network classifier and STFT spectrum analysis to decode two-class MI signals for controlling the up/down movement of a Unmanned Aerial

Vehicle (UAV). The methodology consisted of three tiers. In Tier 1, STFT was used to generate 2D images (spectrograms) from a 4-second trial, capturing both Event-Related Synchronization (ERS) and Event-Related Desynchronization (ERD) motor activity. This process produced six images related to the alpha and beta bands, extracted from a single EEG channel. In Tier 2, a VGG-19 model was applied to extract features from the motor imagery signals. The classification step was performed in Tier 3 by fusing the motor imagery signal features with the neural network (NN) classifier.

The paper is structured as follows: Section 3 details the methodology, Section 4 elaborates on the results and discussions, and Section 5 presents the findings and conclusions of the research.

2. RELATED WORK

In the literature, various Brain-Computer Interface (BCI) modalities have been explored for controlling UAVs such as Steady-State Visual Evoked Potential (SSVEP), P300, eye movement, and MI. For SSVEP-based UAV control, [29] focused on developing a brain-controlled Assistive Device (AD) using a noninvasive SSVEP-based BCI system. This system allowed users to control a UAV in 3D physical space. The proposed BCI system achieved an average Information Transfer Rate (ITR) of 10 bits per minute and a Positive Predictive Value (PPV) of 92.5%. The tests demonstrated the system's ability to control a drone in 3D space. Another study, [34], developed a practical SSVEP-based BCI for continuous control of a quadcopter from the first-person perspective. They used a classification algorithm based on task-related component analysis (TRCA) and linear discriminant analysis (LDA) to decode the commands. [36] explored a BCI system based on SSVEP for controlling quadcopters using electroencephalography (EEG) signals. They employed Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) models for classifying EEG data obtained from multi-flicker screens at different frequencies, each corresponding to a specific drone movement. The results showed high accuracy, particularly with the LSTM model for a 2-second segment, which was the unit of processing. Furthermore, [29] developed a brain-controlled Assistive Device (AD) using SSVEP-based BCI, enabling users to control a flying robot (UAV) in 3D physical space, assisting individuals in exploring their surroundings using their thoughts. These studies highlight the diverse applications and methods of using BCIs for controlling UAVs, with a particular focus on SSVEP-based systems for intuitive and precise control.

In the realm of the P300 and eye movement modalities, the study by [30] introduced an innovative method for drone control utilizing a P300-based brain-computer interface tailored for military applications as assistive technology. This research evaluated the user's calibration proficiency with the software and the program's efficacy in receiving and executing commands transmitted via EEG signals to control the drone. Another groundbreaking study, as documented in [35], proposed a wearable hybrid interface. This interface ingeniously integrated eye movements and mental concentration, directly influencing the control of a quadcopter in three-dimensional space. This noninvasive and cost-effective interface overcame limitations of earlier approaches by enabling users to accomplish complex tasks in a confined environment, where only visual feedback was available. Furthermore, in the study by [28], ocular pulses dominated the data. Employing Principal Component Analysis (PCA), the researchers extracted ocular components. They utilized classification methods, including Multiclass Support Vector Machine (SVM), Quadratic Discriminant Analysis (QDA), and Artificial Neural Networks (ANN), to assess these features independently as well as in combination. The findings underscored the superiority of spectral peaks and bandwidth in terms of classification accuracy among the three features. These studies exemplify diverse methodologies and technologies within BCI systems for drone control, showcasing advancements in P300-based BCIs and interfaces integrating eye movements and mental concentration to enhance UAV control in various environments. In the domain of MI-based BCI modalities, the study by [27] aimed to design and implement a MI based brain-computer interface (BCI) system enabling both disabled and able-bodied individuals to control a drone in a 3D physical environment using only their thoughts. They developed an enhanced version of the filter bank common spatial pattern (FBCSP) algorithm, which outperformed the winning FBCSP algorithm when tested on dataset 2a (4 class MI) from the BCI competition IV, achieving an accuracy of 68.5%. As documented in [1], a MI-based Brain Computer Interface (BCI) system was proposed to facilitate the user-friendly and stable control of a low-speed UAV for indoor target searching. The study leveraged an improved cross-correlation method (CC) for MI feature extraction and employed logistic regression (LR) for

MI feature classification and decision. The BCI system achieved an average classification accuracy rate of 94.36%. In [32], a brain-swarm interface system prototype was presented, capable of various applications using a visual imagery paradigm. The study reported a grand average classification accuracy exceeding the chance level accuracy. [33] implemented continuous control of a UAV in an indoor 2D space through MI tasks. This BCI system employed discriminative time- and frequency-dependent spatial filters for EEG feature extraction in MI tasks. The adaptive LDA method was used for feature classification. The calibration and actual indoor 2D space control experiments showcased the effectiveness and feasibility of employing this BCI system for continuous control of UAV in indoor 2D space. In a study reported by [37], even a mere 0.1% increase in classification accuracy is considered significant in the realm of BCI research due to the complexity of the signals involved. While several studies have sought to enhance classification accuracy, the results from the reviewed studies indicate that there is still room for improvement in this area.

3. METHODOLOGY

The methodological framework of the MTCNN model for decoding the MI signal is illustrated in Figure 1. This framework delineates the entire process of MI pattern detection, with further details provided in the subsequent subsections:

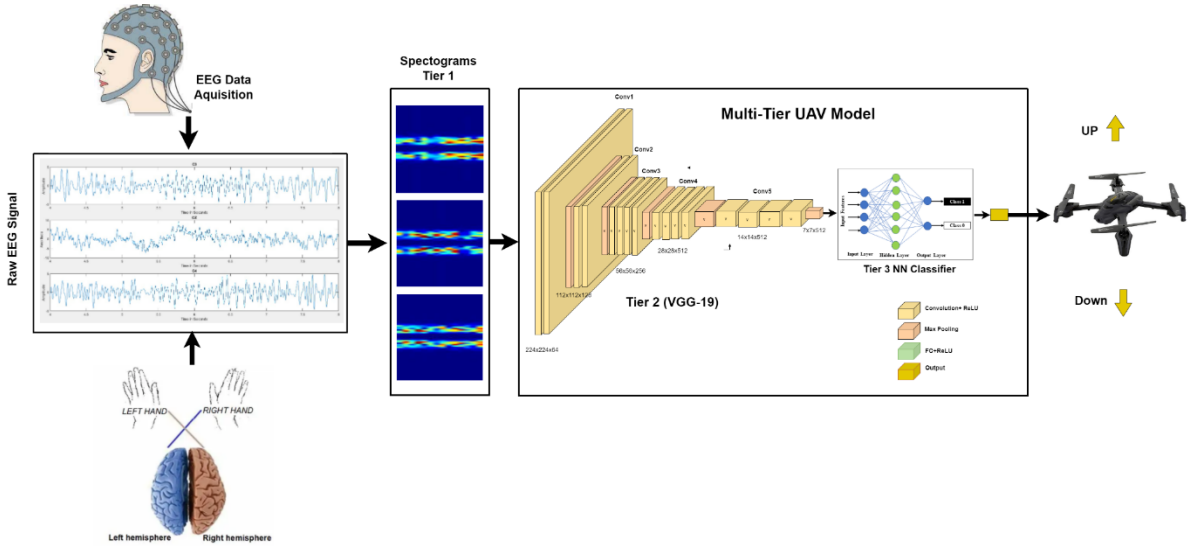

Fig. 1. Methodological Framework for MTCNN model

3.1 MI EEG Datasets

Developers often prefer a minimal number of channels when designing BCI-based systems, as it allows for easy and cost-effective integration into real-time applications [50]. Consequently, in this study, two MI EEG datasets recorded using just three channels were selected. These two datasets are sourced from the BCI competition datasets recorded at Graz University. Additional information about these datasets is provided in the following subsections.

The datasets are divided into two parts: the training part and the evaluation part. Both parts were used to assess the MTCNN model's performance, taking into account inter- and intra-subject differences. Due to the limitation of a large dataset, the data from all nine subjects were combined (union) to create an extensive dataset, encompassing all trials. This approach was adopted to develop a robust model capable of addressing the challenges posed by the complexity of brain signals.

3.1.1 Dataset-I (BCIC IV 2b dataset)

In this dataset, three EEG channels, namely C3, Cz, and C4, were utilized to capture signals related to two motor imagery tasks involving the left hand and right hand. The dataset was collected from nine subjects at a sampling frequency of 250 Hz. EEG data from a total of 160 trials were collected while the subjects watched a flat screen and sat in an armchair.
Two types of recording sessions were conducted: training without feedback and evaluation with smiley feedback. In the first two sessions, subjects received a short warning tone and were instructed to perform a required motor imagery task lasting four seconds. This task was based on a pointing arrow presented on a blank screen.
In the other three sessions, subjects were directed to move a gray smiley feedback symbol centered on the monitor either to the right or left direction after hearing a short warning beep. The smiley feedback was displayed for four seconds, with its color changing to red when moved in the wrong direction and to green when moved in the right direction. Figure 2 illustrates the timing scheme of these two types of sessions [51].

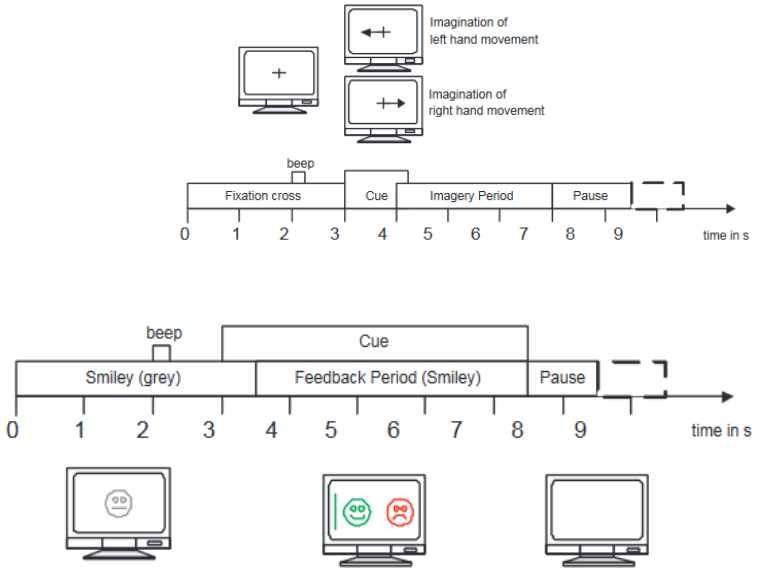

Fig. 2. Trials recording time scheme of BCIC IV 2b dataset (a) without feedback, (b) with smiley feedback.

3.1.2 Dataset-II (BCIC II dataset)

This dataset was collected from a single subject, a 25-year-old female. The EEG device utilized three EEG channels: C3, Cz, and C4, with a sampling frequency of 128 Hz. Each trial had a total duration of 9 seconds.
The recording protocol for this dataset followed the Graz protocol. During the first 2 seconds, the participant remained still. At t=2 seconds, an acoustic stimulus appeared on the screen, initiating the trial, and a cross '+' was displayed for 1 second. At the third second (t=3), a cue in the form of an arrow pointing right or left was presented. There was a total of 280 trials for two types of motor imagery hand movements: right and left. The entire signal of the dataset was filtered using a notch filter with a range from 0.5 to 30 Hz [49]. Figure 3 illustrates the timing scheme of the recording technique.

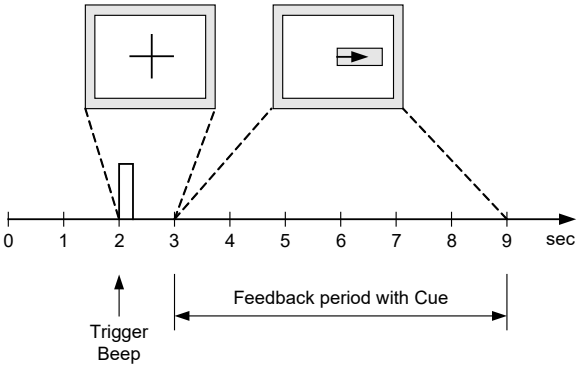

Fig. 3. Trials recording time scheme of BCIC II dataset

3.2 Pre-processing

Undoubtedly, the EEG-MI signal is susceptible to contamination from various sources, including body movements, eye blinks, facial muscle activity, and environmental artifacts like electromagnetic fields produced by electrical devices [3]. Given that the framework is based on deep learning, minimal preprocessing is employed. Frequency filtering is conducted to improve the signal-to-noise ratio of the raw brainwaves and enhance the pertinent signal information. Specifically, a fourth-order Butterworth filter is applied within the range of 8-30 Hz, as the MI EEG signals are reliant on the alpha (8-13 Hz) and beta (14-30 Hz) rhythms.

3.3 VGG-19

The classification of EEG signals poses a challenge that requires high-dimensional features to represent the latent characteristics of brain signals. CNN relies on the convolution process to extract dominant features by employing multiple kernels, also known as filters [23]. Transfer learning is a method that involves using a previously trained network to solve a new classification problem by retraining a few of its final layers. This approach saves a significant amount of time and requires fewer training samples compared to training the network from scratch. The pretrained VGG structure was initially developed by Andrew Zisserman and Karen Simonyan and trained on a dataset comprising 14 million images spanning one thousand classes from the ImageNet dataset. This deep learning framework has found applications in various research domains, delivering superior results in two key computer vision problems: localization and classification [52].

VGG-19 is an extension of the standard VGG architecture that maximizes the feature extraction process with an increased kernel size from 64 to 512. The network consists of units, each comprising a convolution layer followed by a pooling layer with a stride size of 2x2. The Rectified Linear Unit (ReLU) is employed as the activation function in this network, and max-pooling is used for downsampling. The input image size is 224x224, and there are three fully connected layers with 4096 neurons in each, following the order of 4096, 4096, and 1000 layers. The classification concept in this CNN network is based on probability, utilizing the softmax function for multi-class problems [53]. Figure 4 provides an overview of the internal architecture of the VGG-19 neural network. The key components of VGG-19 include three types of layers: convolution, pooling, and fully connected. Further details about this network are outlined below:

1) **The Convolution layer**: This layer is responsible for performing the convolution operation, applying filters to the input image to extract feature maps. This operation helps to identify specific features and aspects in the image's topographical map. The CNN network establishes spatial connectivity among neurons, known as local connectivity, which is useful for tasks like blurring, sharpening, and edge detection [54].

2) **The Pooling layer**: This layer is in charge of reducing the dimension of feature maps through downsampling. This process is essential for preventing overfitting and underfitting issues. The two common techniques for

pooling are max pooling and average pooling, chosen based on the presence of features in specific patches. In VGG-19, max pooling is predominantly used [55].

3) **The Fully Connected Layer**: This layer is the final unit in VGG-19 and consists of three layers. It receives input from the previous layers, namely the pooling and convolution layers. The input to this unit is flattened, and a matrix multiplication process is carried out along with the addition of a bias offset to produce the final output [56].

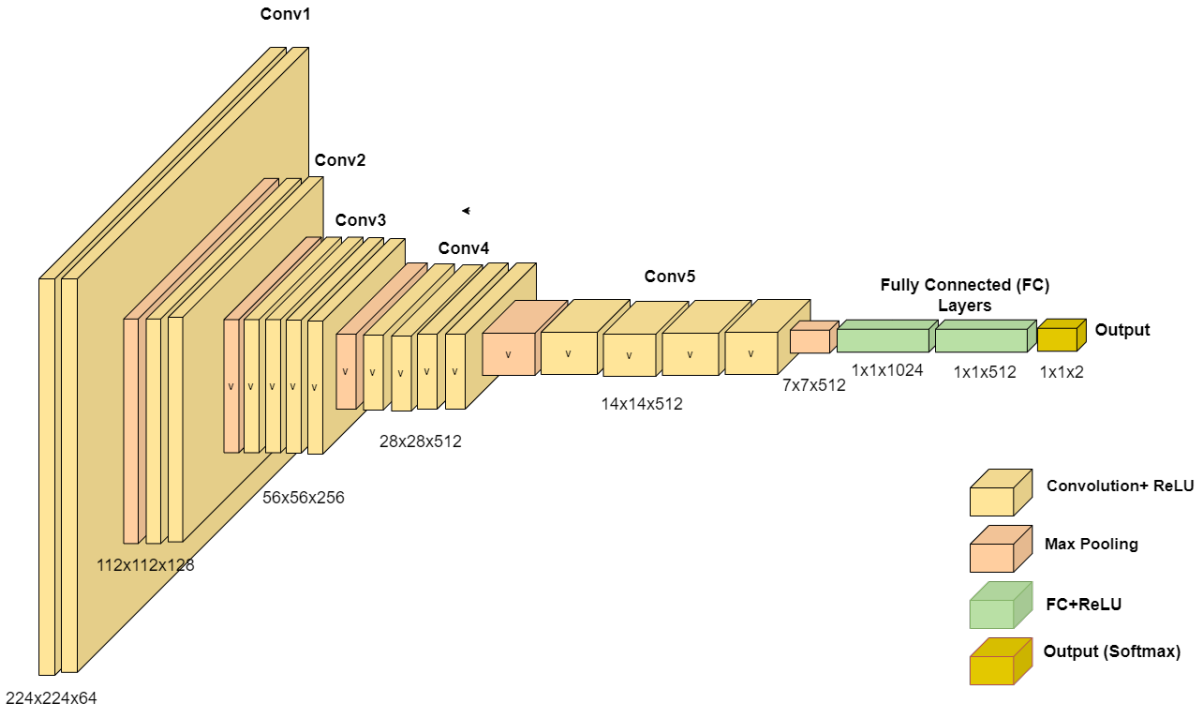

Fig. 4. VGG-19 Architecture

3.4 Short Time Fourier Transform for EEG Image Formulation

The STFT, introduced by Gabor in 1946, stands as one of the most widely employed signal processing techniques for analyzing non-linear and non-stationary signals. Its strength lies in its capacity to capture both the phase and magnitude of a signal that varies over time and frequency [57]. The STFT dissects a long signal into segments, each using the same window size, and subsequently applies the Fourier transform to each of these segments [58]. This process represents an advanced form of Fourier analysis that presents a signal in a way that allows complete estimation in both the time and frequency domains. STFT leverages a window function to extract a portion of the time-domain signal, subsequently applying the Fourier transform to this section to unveil various characteristics of the signal [12]. In the context of STFT, the processed EEG signal x(t) is convolved with a short-time window that slides along the time axis. The output comprises a series of windowed signal segments. Each of these segments then undergoes Fourier transformation, resulting in two-dimensional time-frequency representations of the original signal. The mathematical formulation for STFT is defined as follows [38]:

$$STFT\,(\tau, \omega) = \int_{-\infty}^{+\infty} x(t)w(t - \tau)e^{-j\omega t}dt \qquad (1)$$

In equation (1), *w(t)*and τ *represents* a fixed window size with a limited number of non-zeros on the time axis respectively. STFT method helps in understanding the embedded EEG signal features by consider the signal in two domains namely, the time domain and the frequency domain concurrently. The raw MI signals are defined as E = {(Xi, yi)|i = 1, 2,..., N}, where Xi ∈ $^{RC\times K}$ is a two-dimension matrix that represents the i-th MI trial in the dataset for a given C channels and K samples. The total number of samples in the dataset denoted as N, and the Xi corresponds to the total number of trials, and Yi corresponds to the label for each Xi trial. They get their values from L set that compromises M classes MI tasks. The total number of classes in this study is two class, and their label set denoted as: L = {l1 = "left t", l2 = "right"}. Studies such as [17] reported the efficiency of the STFT for creating 2D images (spectrograms) for 4 s length to be fed then to the CNN as an input image. Therefore, were selected 4 s length which means a total number of 1000 samples for each of the MI signals in *Xi* trial. Then we select a window size of 64 samples with 50 samples of an overlapping. The output of this process is in image capturing the power spectral density (PSD) of any given MI signal and their values measured in Hertz. Therefore, three images are produced for data collected using three electrodes. But, in this study, we aim to capture the alpha and beta frequency bands corresponding to the ERS and ERD motor activity, therefore, the output of this process is six images for each MI trail.

3.5 STTL_RF Evaluation Metrics

The performance of the proposed MTCNN model has been assessed using seven key metrics: accuracy, precision, sensitivity, specificity, F1 score, LogLoss, and AUC. Below, you can find Table I, which presents the mathematical equations for each of these metrics, along with a brief description of each [59]. In the table, the following abbreviations are used: TPL (true positive), TN (true negative), FP (false positive), and FN (false negative).

TABLE I. THE EMPLOYED PERFORMANCE METRICS

Evaluation metric	Mathematical equation	Explanation
Classification accuracy	$CA = \frac{TP + TN}{TP + FP + FN + TN}$	the ratio of the number of correctly classified samples to the total number of the same class input samples
Precision	$Precision = \frac{TP}{TP + FP}$	The number of correctly classified samples among all the classified samples. It tests the classifier's ability to reject irrelevant subjects.
Recall (Sensitivity)	$Recall = \frac{TP}{TP + FN}$	The number of correctly classified samples from all the positive representations.
F1-score	$F1\text{-score} = \frac{2 * TP}{2 * TP + FP + FN}$	The F1 score can be described as a weighted average of precision and recall, where an F1 score achieves its best value at 1 and the worst value at 0.
Specificity	$Specificity = \frac{TN}{TN + FP}$	Assesses a model's ability to detect true negatives of each category.
Log Loss	$\text{LogLoss} = -\frac{1}{n}\sum_{i=1}^{n} [y_i \log_e(\hat{y}_i) + (1 - y_i)\log_e(1 - \hat{y}_i)]$	Log loss is the crucial classification metric based on probabilities. It defines the probability outputs of a classifier instead of its discrete predictions.

The Receiver Operating Characteristics (ROC) curve is employed to assess the performance of the models. This curve allows us to evaluate the classification model's performance at various threshold settings, effectively measuring the degree of separability between classes.

To evaluate how well the model will perform on unseen MI EEG inputs, k-fold cross-validation was employed in this study. In k-fold cross-validation, the data is divided into k subsets. Out of these subsets, k-1 are used for training the model, and the remaining one subset is reserved for testing the model. This process is repeated k times (folds), ensuring that each subset is used as the validation data at least once. The results from the k-folds can be averaged to determine the accuracy of estimation [60].

4. RESULTS AND DISCUSSION

This section essentially presents and discusses the results of developing the MTCNN model for brain-controlled UAV using the Multi-Tiered technique. In the experimental part of this study, four experiments were conducted. The first experiment aimed to evaluate the hybridization of VGG-19 with the neural network (NN) classifier in order to assess the MTCNN model over a large dataset. The results indicated that the hybrid model achieved an impressive classification accuracy of 0.991.

TABLE II. MTCNN MODEL TRAINING PART OVER DATASET-I (TRAINING PART)

Subjects	Performance Metrics								
	Training Time	Testing Time	AUC	CA	F1	Precision	Recall	Logloss	Specificity
S1	295.350	14.389	0.997	0.997	0.997	0.997	0.997	0.022	0.997
S2	334.561	17.496	1.00	0.997	0.997	0.997	0.997	0.010	0.997
S3	309.507	14.945	0.994	0.997	0.997	0.997	0.997	0.095	0.997
S4	238.119	13.967	0.994	0.997	0.997	0.997	0.997	0.042	0.997
S5	324.393	14.134	1.00	1.00	1.00	1.00	1.00	0.005	1.00
S6	322.377	14.507	1.00	0.997	0.997	0.997	0.997	0.009	0.997
S7	289.967	15.152	1.00	0.997	0.997	0.997	0.997	0.012	0.997
S8	270.951	15.037	1.00	0.997	0.997	0.997	0.997	0.007	0.997
S9	297.584	15.451	1.00	0.997	0.997	0.997	0.997	0.009	0.997
Mean	298.089	15.008	0.998	0.997	0.997	0.997	0.997	0.023	0.997

The second experiment, conducted on "Dataset I training part," involved nine subjects and aimed to evaluate the MTCNN model's ability to overcome the problem of inter-subject variations. Since different subjects exhibit varying complexities in their brain signals, this experiment assessed the model's performance. The results of this experiment are presented in Table II, and the mean accuracy across the nine subjects was an impressive 0.997.

Furthermore, to evaluate the MTCNN model's performance in addressing intra-subject issues, such as variations between recording sessions due to physiological factors and recording protocols, experiment 3 utilized "Dataset I Evaluation Part." This dataset incorporated feedback recording protocol, in contrast to the feedback-free protocol in experiment 2. The results of this experiment are presented in Table III, and the mean accuracy across the nine subjects remained at 0.997.

TABLE III. MTCNN MODEL EVALUATION OVER DATASET-I (EVALUATION PART)

Subjects	**Performance Metrics**								
	Training Time	**Testing Time**	**AUC**	**CA**	**F1**	**Precision**	**Recall**	**Logloss**	**Specificity**
S1	343.023	16.314	1.00	0.994	0.994	0.994	0.994	0.014	0.994
S2	299.476	15.069	1.00	1.00	1.00	1.00	1.00	0.004	1.00
S3	302.900	14.949	0.998	0.997	0.997	0.997	0.997	0.021	0.997
S4	312.821	15.201	1.00	1.00	1.00	1.00	1.00	0.004	1.00
S5	368.516	17.950	1.00	0.997	0.997	0.997	0.997	0.015	0.997
S6	309.955	16.397	0.999	0.997	0.997	0.997	0.997	0.020	0.997
S7	318.149	16.799	1.00	0.997	0.997	0.997	0.997	0.008	0.997
S8	279.849	15.132	0.998	0.997	0.997	0.997	0.997	0.021	0.997
S9	316.591	17.454	1.00	1.00	1.00	1.00	1.00	0.003	1.00
Mean	316.808	16.140	0.999	0.997	0.997	0.997	0.997	0.012	0.997

TABLE IV. MTCNN MODEL EVALUATION OVER DATASET-II (TRAINING PART AND TESTING PART)

Dataset-I	Performance Metrics								
	Training Time	Testing Time	AUC	CA	F1	Precision	Recall	Logloss	Specificity
Training Part	28.915	15.488	1.00	1.00	1.00	1.00	1.00	0.004	1.00
Testing Part	371.380	14.732	1.00	0.996	0.996	0.996	0.996	0.014	0.996

Comparing the results of Table II and Table III, it is evident that the challenges associated with brain signal complexity can be observed in the differences in mean values for training and testing times of the MTCNN model over two datasets: "Dataset II training part" and "Dataset II Evaluation Part" from the same subjects. This highlights the difficulties posed by session-to-session data collection. The longer time required for the MTCNN model to understand the distinctive features of the brain signal and distinguish between left and right commands underscores these challenges. However, despite the differences in training and testing times, the performance metrics of the MTCNN model remain consistent, with only slight variations in their values across the two recording sessions.

Additionally, in experiment 4, the hybrid model was validated on another dataset, which in this study consisted of one subject but with two parts: the training part and the testing part. These parts represented data collected from two different recording sessions. The results for the MTCNN model in this experiment were highly promising, with a classification accuracy of 1.00 for the training part and 0.996 for the testing part, as presented in Table V, along with other performance metrics.

TABLE V. RESULTS COMPARISON WITH STATE-OF-THE-ART STUDIES RELATED TO DATASET-I

Year	Study	Method	Accuracy
2015	[61]	LDA-based wrapper SFS	90%
2016	[62]	STFT with KNN	83.57%
2016	[63]	WT+SE using SVM and KNN	86.4%
2016	[64]	MEMD + STFT with KNN	90.71%
2017	[65]	Fuzzified Adaptation with SVM	81.48%
2019	[66]	Genetic Algorithm with FKNN	84%
2019	[58]	STFT with CNN	89.73%
2019	[67]	CWT with 1D CNN	92.9%
2020	[49]	WPT+CWT with CNN	95.71%
2021	[68]	WTTD + CWT with CNN	96.43%
2022	This Study	Proposed Method	99.7%

TABLE VI. RESULTS COMPARISON WITH STATE-OF-THE-ART STUDIES RELATED TO DATASET-II

Year	Study	Method	Accuracy
2014	[43]	Hjorth parameter-LDA	79.1%
2015	[40]	CSP-EMD	72.30%
2018	[41]	CSP- autoregressive model	77%
2018	[47]	WDPSD	89.36%

2018	[48]	A normalization model with one contralateral EOG channel	96.86%
2019	[45]	a separated channel convolutional network	83%
2019	[38]	STFT-VGG16	71.2%
2020	[42]	multi-domain features	79%
2020	[46]	CNN with hybrid convolution scale	87.6%
2020	[44]	Hilbert transform (HT)-SVM	82.50%
2020	[49]	CWT-VGG19	97.06%
2021	[39]	CWT-CNN	71.25
2022	This Study	Proposed Method	100%, 0.99%

It is worth noting that when comparing the accuracy of the MTCNN model over dataset-I and dataset-II, the efficiency of the model becomes evident. In fact, the model has outperformed the accuracy reported in the literature, as presented in both Table V and Table VI. This comparative result underscores the model's capability in decoding MI brain signals related to both left- and right-hand movements, making it suitable for integration into the field of brain-controlled UAVs as a control command for managing the up and down movements of the aircraft. Furthermore, the MTCNN model holds the potential to make valuable contributions to the BCI-MI community by facilitating the deployment of this proposed model in MI-based control systems.

5. CONCLUSION

The primary objective of this study is to develop a hybrid feature learning model capable of addressing the complexities of EEG signals when decoding MI signals associated with left- and right-hand movements for controlling the up and down actions of a UAV. To achieve this, the study employs a novel approach, known as the Multi-Tiered CNN (MTCNN) model, which combines a pretrained CNN with a classical neural network classifier and utilizes STFT spectrograms. The experimental phase of the study involved four key experiments. In the first experiment, the MTCNN model was evaluated using a large dataset, achieving an impressive classification accuracy of 0.991. The second and third experiments sought to assess the model's performance on two different datasets involving the same subjects. These experiments aimed to test the model's ability to overcome challenges related to inter-subject and intra-subject variations stemming from differences in brain complexities. Remarkably, the MTCNN model achieved a classification accuracy of 0.997 on both datasets, underlining its robustness. In the fourth experiment, the model was validated using another dataset, achieving classification accuracies of 100% and 0.996 on the training and testing parts, respectively. Notably, the MTCNN model outperformed the accuracy reported in the existing literature based on two BCI competition datasets. In conclusion, the MTCNN model demonstrates its capacity to decode MI signals associated with both left- and right-hand movements, making it a promising tool for controlling the up and down movements of a brain-controlled UAV. This innovation holds potential significance for advancing the field of brain-computer interfaces and MI-based UAV control systems.

References

[1] T. Shi, et al., "Brain Computer Interface system based on indoor semi-autonomous navigation and motor imagery for Unmanned Aerial Vehicle control," Expert Systems with Applications, vol. 42, pp. 4196-4206, 2015.

[2] A. Nourmohammadi, et al., "A survey on unmanned aerial vehicle remote control using brain–computer interface," IEEE Transactions on Human-Machine Systems, vol. 48, pp. 337-348, 2018.

[3] A. Al-Saegh, et al., "Deep learning for motor imagery EEG-based classification: A review," Biomedical Signal Processing and Control, vol. 63, p. 102172, 2021.

[4] R. Zhang, et al., "A new convolutional neural network for motor imagery classification," in 2019 Chinese Control Conference (CCC), 2019, pp. 8428-8432.

[5] S. U. Amin, et al., "Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion," Future Generation computer systems, vol. 101, pp. 542-554, 2019.

[6] S. Aggarwal and N. Chugh, "Signal processing techniques for motor imagery brain computer interface: A review," Array, vol. 1, p. 100003, 2019.

[7] X. Tang, et al., "Motor imagery EEG recognition with KNN-based smooth auto-encoder," Artificial intelligence in medicine, vol. 101, p. 101747, 2019.
[8] G. C. Jana, et al., "Enhancing the performance of motor imagery classification to design a robust brain computer interface using feed forward back-propagation neural network," Ain Shams Engineering Journal, vol. 9, pp. 2871-2878, 2018.
[9] Z. Tang, et al., "Single-trial EEG classification of motor imagery using deep convolutional neural networks," Optik, vol. 130, pp. 11-18, 2017.
[10] R. Liu, et al., "Identification of Anisomerous Motor Imagery EEG Signals Based on Complex Algorithms," Computational intelligence and neuroscience, vol. 2017, 2017.
[11] M. J. Ferdous and M. S. Ali, "Time-Frequency Analysis of EEG Signal processing for Artifact Detection."
[12] S. Chaudhary, et al., "Convolutional neural network based approach towards motor imagery tasks EEG signals classification," IEEE Sensors Journal, vol. 19, pp. 4494-4500, 2019.
[13] Y. Du, et al., "Classification of seizure EEGs based on short-time fourier transform and hidden markov model," in 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2020, pp. 875-880.
[14] N. Sharma, et al., "Epileptic seizure detection using STFT based peak mean feature and support vector machine," in 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN), 2021, pp. 1131-1136.
[15] W. Qiao and X. Bi, "Ternary-task convolutional bidirectional neural turing machine for assessment of EEG-based cognitive workload," Biomedical Signal Processing and Control, vol. 57, p. 101745, 2020.
[16] D. Freer and G.-Z. Yang, "Data augmentation for self-paced motor imagery classification with C-LSTM," Journal of neural engineering, vol. 17, p. 016041, 2020.
[17] A. Al-Saegh, et al., "CutCat: An augmentation method for EEG classification," Neural Networks, vol. 141, pp. 433-443, 2021.
[18] X. Wang, et al., "A Hybrid Transfer Learning Approach for Motor Imagery Classification in Brain-Computer Interface," in 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech), 2021, pp. 496-500.
[19] W. Wei, et al., "A transfer learning framework for RSVP-based brain computer interface," in 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 2963-2968.
[20] X. Wei, et al., "Inter-subject deep transfer learning for motor imagery eeg decoding," in 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER), 2021, pp. 21-24.
[21] D.-K. Kim, et al., "Sequential Transfer Learning via Segment After Cue Enhances the Motor Imagery-based Brain-Computer Interface," in 2021 9th International Winter Conference on Brain-Computer Interface (BCI), 2021, pp. 1-5.
[22] K.-T. Kim, et al., "Subject-Transfer Approach based on Convolutional Neural Network for the SSSEP-BCIs," in 2021 9th International Winter Conference on Brain-Computer Interface (BCI), 2021, pp. 1-3.
[23] T. Kaur and T. K. Gandhi, "Automated brain image classification based on VGG-16 and transfer learning," in 2019 International Conference on Information Technology (ICIT), 2019, pp. 94-98.
[24] Y. Mangalmurti and N. Wattanapongsakorn, "COVID-19 and Other Lung Disease Detection Using VGG19 Pretrained Features and Support Vector Machine," in 2021 25th International Computer Science and Engineering Conference (ICSEC), 2021, pp. 51-56.
[25] M. Ahmed, et al., "Automatic COVID-19 pneumonia diagnosis from x-ray lung image: A Deep Feature and Machine Learning Solution," in Journal of Physics: Conference Series, 2021, p. 012099.
[26] A. Saidi, et al., "A novel epileptic seizure detection system using scalp EEG signals based on hybrid CNN-SVM classifier," in 2021 IEEE Symposium on Industrial Electronics & Applications (ISIEA), 2021, pp. 1-6.
[27] N. S. Holm and S. Puthusserypady, "An improved five class mi based BCI scheme for drone control using filter bank CSP," in 2019 7th International Winter Conference on Brain-Computer Interface (BCI), 2019, pp. 1-6.
[28] S. S. Poorna, et al., "EEG based control using spectral features," in 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2018 2nd International Conference on, 2018, pp. 788-794.
[29] A. Chiuzbaian, et al., "Mind controlled drone: An innovative multiclass SSVEP based brain computer interface," in 2019 7th International Winter Conference on Brain-Computer Interface (BCI), 2019, pp. 1-5.
[30] F. A. Al-Nuaimi, et al., "Mind drone chasing using EEG-based brain computer interface," in 2020 16th International Conference on Intelligent Environments (IE), 2020, pp. 74-79.
[31] K. LaFleur, et al., "Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain–computer interface," Journal of neural engineering, vol. 10, p. 046003, 2013.
[32] J.-H. Jeong, et al., "Towards brain-computer interfaces for drone swarm control," in 2020 8th International Winter Conference on Brain-Computer Interface (BCI), 2020, pp. 1-4.
[33] Y. An, et al., "UAV control in 2D space based on brain computer interface," in 2017 4th International Conference on Systems and Informatics (ICSAI), 2017, pp. 594-598.
[34] J. Mei, et al., "Using SSVEP-BCI to continuous control a Quadcopter with 4-DOF motions," in 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 4745-4748.
[35] B. H. Kim, et al., "Quadcopter flight control using a low-cost hybrid interface with EEG-based classification and eye tracking," Computers in biology and medicine, vol. 51, pp. 82-92, 2014.
[36] K. Ishizuka, et al., "High accuracy and short delay 1ch-ssvep quadcopter-bmi using deep learning," Journal of Robotics and Mechatronics, vol. 32, pp. 738-744, 2020.

[37] N. K. Al-Qazzaz, et al., "EEG Feature Fusion for Motor Imagery: A New Robust Framework Towards Stroke Patients Rehabilitation," Computers in biology and medicine, p. 104799, 2021.

[38] G. Xu, et al., "A deep transfer convolutional neural network framework for EEG signal classification," IEEE Access, vol. 7, pp. 112767-112776, 2019.

[39] R. Mahamune and S. H. Laskar, "Classification of the four-class motor imagery signals using continuous wavelet transform filter bank-based two-dimensional images," International Journal of Imaging Systems and Technology, 2021.

[40] A. M. Álvarez-Meza, et al., "Time-series discrimination using feature relevance analysis in motor imagery classification," Neurocomputing, vol. 151, pp. 122-129, 2015.

[41] J. Wang, et al., "Toward optimal feature and time segment selection by divergence method for EEG signals classification," Computers in biology and medicine, vol. 97, pp. 161-170, 2018.

[42] C. Xu, et al., "Two-level multi-domain feature extraction on sparse representation for motor imagery classification," Biomedical Signal Processing and Control, vol. 62, p. 102160, 2020.

[43] S.-H. Oh, et al., "A novel EEG feature extraction method using Hjorth parameter," International Journal of Electronics and Electrical Engineering, vol. 2, pp. 106-110, 2014.

[44] N. Bagh and M. R. Reddy, "Hilbert transform-based event-related patterns for motor imagery brain computer interface," Biomedical Signal Processing and Control, vol. 62, p. 102020, 2020.

[45] X. Zhu, et al., "Separated channel convolutional neural network to realize the training free motor imagery BCI systems," Biomedical Signal Processing and Control, vol. 49, pp. 396-403, 2019.

[46] G. Dai, et al., "HS-CNN: a CNN with hybrid convolution scale for EEG motor imagery classification," Journal of neural engineering, vol. 17, p. 016025, 2020.

[47] C. Kim, et al., "An effective feature extraction method by power spectral density of EEG signal for 2-class motor imagery-based BCI," Medical & biological engineering & computing, vol. 56, pp. 1645-1658, 2018.

[48] L. Sun, et al., "A contralateral channel guided model for EEG based motor imagery classification," Biomedical Signal Processing and Control, vol. 41, pp. 1-9, 2018.

[49] P. Kant, et al., "CWT Based transfer learning for motor imagery classification for brain computer interfaces," Journal of Neuroscience Methods, vol. 345, p. 108886, 2020.

[50] M. Z. Baig, et al., "Filtering techniques for channel selection in motor imagery EEG applications: a survey," Artificial intelligence review, pp. 1-26, 2019.

[51] M. Tangermann, et al., "Review of the BCI competition IV," Frontiers in neuroscience, p. 55, 2012.

[52] N. Begum and M. K. Hazarika, "Maturity detection of tomatoes using Transfer Learning," Measurement: Food, p. 100038, 2022.

[53] Y. Xia, et al., "A Switch State Recognition Method based on Improved VGG19 network," in 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2019, pp. 1658-1662.

[54] S. Kavitha, et al., "Neural Style Transfer Using VGG19 and Alexnet," in 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2021, pp. 1-6.

[55] N. Abuared, et al., "Skin Cancer Classification Model Based on VGG 19 and Transfer Learning," in 2020 3rd International Conference on Signal Processing and Information Security (ICSPIS), 2020, pp. 1-4.

[56] S. Mascarenhas and M. Agarwal, "A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification," in 2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON), 2021, pp. 96-99.

[57] C. Huang, et al., "Predicting human intention-behavior through EEG signal analysis using multi-scale CNN," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 18, pp. 1722-1729, 2020.

[58] T. H. Shovon, et al., "Classification of motor imagery EEG signals with multi-input convolutional neural network by augmenting STFT," in 2019 5th International Conference on Advances in Electrical Engineering (ICAEE), 2019, pp. 398-403.

[59] W. Mumtaz, et al., "Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD)," Biomedical Signal Processing and Control, vol. 31, pp. 108-115, 2017.

[60] B. S. Nanthini and B. Santhi, "Electroencephalogram signal classification for automated epileptic seizure detection using genetic algorithm," Journal of natural science, biology, and medicine, vol. 8, p. 159, 2017.

[61] R. Masoomi and A. Khadem, "Enhancing LDA-based discrimination of left and right hand motor imagery: Outperforming the winner of BCI Competition II," in 2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI), 2015, pp. 392-398.

[62] T.-U. Jang, et al., "Motor-imagery EEG signal classification using position matching and vector quantisation," International Journal of Telemedicine and Clinical Practices, vol. 1, pp. 306-313, 2016.

[63] A. B. Das and M. I. H. Bhuiyan, "Discrimination and classification of focal and non-focal EEG signals using entropy-based features in the EMD-DWT domain," Biomedical Signal Processing and Control, vol. 29, pp. 11-21, 2016.

[64] S. K. Bashar and M. I. H. Bhuiyan, "Classification of motor imagery movements using multivariate empirical mode decomposition and short time Fourier transform based hybrid method," Engineering science and technology, an international journal, vol. 19, pp. 1457-1464, 2016.

[65] R. Chatterjee, et al., "Dimensionality reduction of EEG signal using fuzzy discernibility matrix," in 2017 10th International Conference on Human System Interactions (HSI), 2017, pp. 131-136.

[66] S. V. Eslahi, et al., "A GA-based feature selection of the EEG signals by classification evaluation: Application in BCI systems," arXiv preprint arXiv:1903.02081, 2019.

[67] H. K. Lee and Y.-S. Choi, "Application of continuous wavelet transform and convolutional neural network in decoding motor imagery brain-computer interface," Entropy, vol. 21, p. 1199, 2019.

[68] M. Wei, et al., "Motor Imagery EEG Signal Classification based on Deep Transfer Learning," in 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), 2021, pp. 85-90.