RESEARCH ARTICLE

Transfer Learning Enabled Imagined Speech Interpretation Using Phase-Based Brain Functional Connectivity and Power Analysis

MEENAKSHI BISLA AND RADHEY SHYAM ANAND

Indian Institute of Technology Roorkee, Roorkee 247667, India

Corresponding author: Meenakshi Bisla (meenakshi_b@ee.iitr.ac.in)

ABSTRACT We propose a Transfer learning-enabled electroencephalography-based intuitive brain-computer interface system by utilizing phase-based brain functional connectivity methods such as phase lag index (PLI) and Intersite phase clustering (ISPC) along with power features to explore both phase and power-based information from electroencephalography (EEG) signals. Time-frequency decomposition using a complex morlet wavelet is applied to analyze the signal components in both the time and frequency domains and extract phase connectivity and power features. Functional connectivity methods aim to recognize functional interactions and statistical mutuality among signals acquired across various brain areas. The phase-based connectivity features are extracted simultaneously for multiple channels to investigate the phase synchronization among EEG signals across the entire brain. Next, Graph theory is adopted to trace connectivity between brain regions by calculating the connectivity degree of extracted PLI and ISPC features with other electrodes. In Parallel, Discrete wavelet convolution is performed to calculate the time variable frequency band's specific power from the imagined speech EEG data. Finally, Time-frequency images of the above-mentioned PLI, ISPC, and EEG power features are fed as input to DenseNet-121 architecture for classification. Dense Net architecture overcomes the problem of 'vanishing gradient' by connecting each layer directly with other layers, making the network densely connected. The maximum classification accuracy achieved is 100%, 99.14%, and 98.72% for binary, three-class, and four-class classifications, respectively. The experimental results indicate that the proposed phase-based connectivity features, EEG power, and the DenseNet-121 model have achieved excellent accuracy for two public datasets, outperforming the state-of-the-art methods. The outstanding results strengthen the possibility of real-time EEG-based intuitive brain-computer interface communication.

INDEX TERMS Brain-computer interface, electroencephalography, deep learning, medical signal processing, speech imagery, time-frequency analysis, phase connectivity.

I. INTRODUCTION

With the advancement in computer-aided data analysis methods, researchers are motivated to engage in machine learning, deep learning, and transfer learning technology for evolving brain-machine interfaces [1]. Brain-computer interface (BCI)

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

bonds the gap between one's inner world and physical world by translating their intentions onto computer commands in the form of messages to the outer world, providing non-motor-based interactions.

As a type of mental imagery, imagined speech is one of the most direct ways to elicit a pattern of brain activity since the user visualizes what they wish to say rather than using indirect commands, such as motor imagery, to express

their message [2], [3]. It has the facility of using a more significant number of commands as words or sentences or objects, etc., which in turn facilitates the design of BCI systems with extra degrees of freedom than what is imaginable with motor imagery (MI) based BCI systems, where expanding the class number entirely relies on body part movement that can overlap when a large number of classes are to employed [4]. Imagery-based BCI systems can potentially facilitate people with automation in daily life activities, especially people with psychological/physical diseases or disabilities, enabling individuals to control the environment through mental thoughts [5].

Popular paradigms with great accuracy include P300 [6], Steady State Visual Evoked Potentials (SSVEP) [7], and MI [8]based BCI system. However, their real-time application in daily life is restricted due to their lack of convenience, which necessitates continuous visual attention, extra-ocular muscle motor control, and external stimuli to produce the correct EEG signals [9]. One of the most essential non-invasive methods for obtaining brain data corresponding to imagined speech activity from the brain is electroencephalography (EEG). EEG acquisition plays a primary role compared to other non-invasive techniques for acquiring brain signals, such as MEG, FNIRS, FMRI, PET, etc. The increasing demand for comfortable, long-term, continuous brain signal monitoring has been rekindled by the development of wearable EEG technology for both BCI and medical applications [10]. EEG-based imagined speech detection can significantly boost the performance of brain-computer interface devices. Successful Interpretation of thoughts is the only possible way to provide communication aid to patients with locked-in syndrome who are physically incapacitated but psychologically conscious [2], thereby enlightening the quality of rehabilitation and clinical neurology [11]. The brain signals acquired using EEG electrodes are sent to the computer-aided analysis process and converted into meaningful commands for operating the BCI system [12]. Computer-aided analysis methods (Pre-processing, artifact removal, feature engineering, machine learning, deep learning, etc.) are essential as imagined speech EEG signals have complex and random characteristics, making them laborious and draining to interpret even for experts. These methods enable the faster and more accurate analysis of EEG signals without any alteration caused by subjective uncertainty and human error. We propose an innovative transfer learning-based architecture that addresses the problem of limited multi-class accuracy and poor spatial resolution. Neural correlates of speech imagery EEG signals are variable, and weak as compared to the vocal state; hence, it is challenging to interpret them using machine learning (ML) based classifiers. The applicability of modern deep learning and transfer learning methods, along with fine feature engineering methods like brain connectivity analysis, have seen substantial advances in complex EEG signal analysis as compared to ML-based methods. Transfer learning methods have achieved high accuracy in EEG signal classification with robust capability to extract characteristic

features [13]. Results in literature are not so astonishing when it comes to the simultaneous recognition of multiple classes. Considering the downsides in the aforementioned literature, an innovative transfer learning-based BCI system is proposed to classify imagined speech from EEG data. Surface Laplacian helps to identify the components specific to speech imagination by attenuating low spatial frequency components from the EEG data.

The awareness of imagined speech dates back to 1992 when Hans Berger invented the electroencephalogram (EEG) as a tool for synthetic telepathy [14]. Research in speech-imagery decoding mainly focuses on binary or multi-class classification of discrete vowels, syllables, words, and sentences. Such works typically encompass traditional signal processing, feature extraction techniques, and machine learning-based classifiers.

DaSalla et al. invented an EEG-based imagined speech decoding system based on binary classification of vowel categories like /a/, /u/, and rest with common spatial pattern (CSP) based features and nonlinear SVM for classification with accuracy of 68-79% [15]. Brigham and Kumar derived an algorithm for classifying two imagined syllables, /ba/ and /ku/, with the help of autoregressive coefficients and a k-nearest neighbor classifier, achieving classification accuracy between 46% and 88%, respectively [16].

Nguyen et al. obtained an average classification accuracy of 73.3%, 49.0 %, 50.1%, and 66.2% for short vs. long word vowels, short words, and long words classification, respectively, with the help of relevance vector machine classifier [17]. Cooney et al. used information from statistics of all remaining subjects to transfer knowledge to the target subject. The highest overall accuracy of 35.68% for 5-class classification is achieved when input layers of CNN are fine-tuned. In contrast, an accuracy of 34.41 is achieved when the final convolution layers are fine-tuned [18].

Saha et al. [19] used a Convolutional neural network along with a long short-term memory network (LSTM) for the classification of ±nasal, ±iy, ±bilab, ±uw, C/V and attained accuracies of 73.45%, 73.30%, 75.55%, 81.99%, and 85.23%, respectively. Panachakel and Ganesan [13] used a ResNet-50-based transfer learning model along with data augmentation for decoding imagined speech prompts. They attained a maximum classification accuracy of 95.5% for short vs long words and 79.7% for vowels.

Kamble et al. [20] achieved maximum classification accuracy of $89.6 \pm 4.6\%$ and $61.1 \pm 5.1\%$ in binary and multiclass (seven classes) signals, respectively, using Machine-learning-enabled adaptive signal decomposition for a brain-computer interface using EEG. Bisla and Anand [21]employed an amalgamation of the Two-dimensional convolutional neural network and LSTM and achieved an average accuracy of 43.76% on 4 class classifications of words 'pat,' 'pot,' 'knew,' and 'gnaw.' The latter performed phase-based connectivity analysis to identify the dominant neurophysiological dynamics of the speech imagery paradigm [2].

The real-time decoding of speech imagery from EEG signals is an essential issue that must be solved in the design of the BCI system. Despite the expanding interests and efforts in designing real-time imagined speech recognition systems since 2010, limited accuracy and poor signal-to-noise ratio of imagined speech signals hinder the development of BCI systems based on imagined speech decoding.

Considering the literature review, some binary classification works have archived appreciable accuracy, reaching +85%. Multi-class classification would be more viable for real-life applications and was demonstrated to have much lower classification rates than binary tasks.

Considering the randomness, low signal-to-noise ratio, and non-linearity of EEG data, EEG data analysis is a very challenging task. Especially in decoding Imagined speech-related components, poor SNR causes the imagery components of interest to be difficult to recognize from the numerous background neural activations.

Because of these issues, basic machine learning-based classification has proven to be effective in P300 [5], Steady State Visual Evoked Potentials (SSVEP) [6], and MI [7] based BCI systems but has not obtained satisfactory results when implemented for speech imagery task. The applicability of modern transfer learning methods, along with advanced filtering and feature extraction techniques, can substantially advance the interpretation of imagined speech EEG data.

EEG data offers excellent temporal resolution but lacks spatial resolution, leading to low accuracy on the source of information in decoding feeble components corresponding specifically to the imagination of speech from all the mixed neural activation during that time. To address this issue, we have applied surface Laplacian transform to the data to attain robustness against the potential confound of volume conduction by successful filtration of spatially broad assemblies of data.

We propose a novel deep learning architecture that addresses the problem of limited multi-class accuracy and poor signal-to-noise ratio 1) By applying Laplacian filtering to the EEG data for attenuating low-spatial frequency features and will help to isolate topographical features within several centimeters of range. 2) By exploring the important speech imagery-oriented features in both phase (PLI, ISPC) and magnitude domain (EEG power), and 3) by using the fine-tuned Dense transfer learning architecture for high-accuracy multi-class classification.

This study investigates a competent approach for designing a real-time brain-computer interface system based on inner thoughts (imagined/Covert speech) with +90% accuracy for both binary and multi-class classification. There are works in the literature where phase synchronization [22] based features like ISPC [2], PLI [23], and EEG power-based [24] features are extracted from EEG data; this is the first work to utilize them for the classification of imagined speech EEG data, along with graph theory for analysis inter-electrode

phase connectivity and thereafter converting the extracted phase connectivity and power features into time-frequency images for the classification process.

Imaginary speech decoding using phase-based connectivity and EEG power using EEG signals is one such area that has not been studied and worked upon extensively. Considering the non-stationary nature of EEG signals, time-frequency decomposition is utilized to analyze the signal components in both the time and frequency domains to extract phase connectivity and power features.

This paper proposes Transfer learning-enabled phase-based brain functional connectivity and power features for an electroencephalography-based intuitive brain-computer interface system. The proposed model embellishes the gigantic potential to develop a practical EEG-based speech imagery interpretation technology that can revolutionize the BCI domain.

The essential contributions of this research are listed below:

I. Although there are works in the literature where ISPC, PLI, and EEG power features are extracted from EEG data, this is the first work to utilize them for the classification of imagined speech EEG data by converting the extracted phase connectivity and power features into time-frequency images.

II. This is the first work to make use of a fine-tuned DenseNet-121 network as the base classifier for classifying imagined speech from time-frequency spectrogram images of extracted features.

III. Surface Laplacian transform is applied to combat the effect of poor spatial resolution of EEG signal; it helps identify the components specific to speech imagination by attenuating low spatial frequency components from the EEG data.

IV. Our methodology has greatly improved the multi-class classification performance compared to the state-of-the-art.

V. The paper presents a qualitative comparison of the classification performance across different deep learning, machine learning, and transfer learning classifiers.

The remaining portion of the research paper is systematized as follows: the second section contains information about datasets and methods implemented in the paper and provides a brief overview of features and classifiers employed. The proposed algorithm is presented in section three. The fourth section contains information about results and comparisons. The conclusion is given in a later section.

II. DATASETS AND METHODS
A. DATASETS
This study utilizes two publicly available imagined speech datasets. Dataset 1 belongs to Nguyen et al. [17]. It was recorded at human-oriented robotics and controlled labs., Dataset 2 belongs to Zhao and Rudzicz [24]. It was acquired by the Toronto Rehabilitation Institute.

1) DATASET 1

This dataset contains EEG recordings corresponding to four categories of speech imagination: long words, short words, vowels, and short vs long words. The Long words consist of two words, 'cooperate' and 'independent' and the short words consist of three words, 'in', 'out' and 'up,' while the set of three vowels contained '/a/,' '/i/' and '/u/.' A Brain Products ACTi Champ amplifier device with 64 electrode sensors was placed in accordance with the 10–20 international system to record the data with a sampling frequency of 1000 Hz, which was later down-sampled to 256 Hz. Also, a 5th-order Butterworth bandpass filter with the pass band from 8 - 70 Hz was used to eliminate any low-frequency disturbances and electromyogram (EMG) artifacts from the acquired EEG data. A notch filter was applied to eradicate the 60-Hz line noise. The dataset consists of 100 trials per prompt, except for a few subjects with 80 trials. More details about the dataset can be found in [17].

2) DATASET 2

It comprises EEG, audio, and face tracking data recorded during the imagery and articulation of four words taken from Kent's list of phonetically alike pairs (knew, gnaw, pot, pat) and seven syllabic prompts (/m/, /n/, /iy/, /uw/, /piy/, /diy/, /tiy/). Each prompt was displayed 12 times for an overall 132 trials per subject, and the trials were randomly permuted, except for a few subjects with 165 trials. EEG signals from 8 male and four female subjects (average age = 27.4, range = 14) were obtained with the help of a 64-channel Neuroscan Quick-cap utilizing a 10-20 electrode placement system. Every subject was right-handed with no visual, motor, or hearing weakness and had no antiquity of drug abuse or any neurological problem. EEG data was acquired with a sampling rate of 1000 Hz using a SynAmps RT amplifier.

Each trial comprises four distinct states of action:

I. **Resting state**, where the subject is told to rest for 5 seconds so that the subject starts the recording procedure with a relaxed state of mind.

II. **Stimulus state**, where both auditory and visual stimulus is presented.

III. **Speech imagery state**, in which the subject performed imagination of the prompt for 5 seconds without making any kind of articulatory movement.

IV. **Articulation state**, where the subject spoke the prompt audibly. A Kinect sensor is employed to record facial features and audio when the subject is articulating the prompt.

Data from 4 of the 12 subjects were omitted because of detached ground wires and two subjects falling asleep during data acquisition. 5-second speech imagery state is utilized for analysis. More detailed information about the dataset can be found in [24].

B. TIME-FREQUENCY ANALYSIS

Simultaneous mapping of both the time and frequency domain information is an ideal technique for extracting relevant patterns from complex and highly random imagined speech EEG signals, as spectral activity of the brain varies with respect to time. Time-frequency feature extraction methods allow the spectral activity to be mapped relative to the temporal variations. Solely, Time or frequency-based feature extraction researches have satisfactory results but have inadequacies in mapping frequency and time domains simultaneously.

Complex Morlet wavelets were employed to obtain estimations of the time variable frequency band-specific amplitude and phase information from the EEG data and to extract phase and power information from the non-stationary EEG data [25], [26]. Time-frequency decomposition using Complex morlet wavelets is a powerful and prevalent technique that provides balanced time-frequency localization. These wavelets are sinusoidal in shape and are weighted by a Gaussian kernel, as shown in "(1)". Where f_b refers to the bandwidth of the Gaussian window and f_c refers to the central frequency of the wavelet.

$$\Psi\left(t\right) = \frac{1}{\sqrt{\pi f_b}} e^{2\pi f_c t} e^{-\frac{t^2}{f_b}} \tag{1}$$

Time-frequency representation is obtained by performing convolution of input EEG signal with wavelet function ($\Psi(t)$) of complex morlet wavelet. The wavelet cycles vary from four cycles in the lowermost frequency to ten cycles in the higher most frequency.

Wavelet transform obtained as a result of convolution provides time-frequency information of the inputted EEG signal. Magnitude information is obtained by extracting the modulus of wavelet transform coefficients. Phase information is contained with the help of phase of the wavelet transform coefficients.

Time-frequency decomposition led to an [electrode(E) × frequency(F) × time(T) × trial(N)] matrix of size [64 × 30×1280 × 100] for dataset 1 and [62 × 30×5000 × 12] for dataset 2 with values in the complex domain. The resultant data matrix is utilized as input for phase-based connectivity analysis and power analysis, as shown in the paper's outline in Figure 5.

C. GRAPH THEORY

Graph theory is utilized in the research to analyze the brain's structural and functional connectivity by engaging connectivity degree as a network measure to study the interactions between different neuronal units [25]. The graph's nodes refer to the electrodes, and the graph's edges refer to the association among electrodes. Phase lag index and intersite phase clustering were used in this work as phase connectivity measures to build brain networks for EEG classification. Figure. 1 represents the all-to-all PLI and ISPC connectivity matrix at a frequency and time of 45hz and 2.5 seconds, respectively, for dataset 2. The upper triangle of Figure 1. a corresponds to PLI connectivity, and the lower triangle corresponds to ISPC connectivity. Both the axes (X-axis, Y-axis) refer to

electrodes, and the intensity of color at each point signifies the extent of the connectivity between the electrodes.

The thresholding technique filters the feeble connections from the connectivity matrix that are not significant to the underlying network structure. This step is beneficial in simplifying complex network associations by focusing on the most critical connections and filtering out spurious trends. Statistical thresholding technique is utilized in this study where threshold (T) is set to a value of one standard deviation above the median connectivity as shown in "(2)". The connectivity values above the calculated threshold value are set to '1', implying strong associations, and values lower than the calculated threshold value are set to zero, representing weak and feeble connections.

$$T = M(\text{connectivity}) + SD(\text{connectivity}) \qquad (2)$$

where M signifies median connectivity and SD signifies standard deviation connectivity. The median and SD connectivity-based binary thresholding technique is considered very simple and advantageous for connectivity analyses as it delivers the threshold valve found in the data statistics, which is rarely prone to outliers. Figures 1b and 1. c represent the ISPC and PLI connectivity matrices after applying the thresholding technique.

D. PHASE-BASED BRAIN FUNCTIONAL CONNECTIVITY ANALYSIS

Brain functional connectivity analysis deals with the associations amongst similar patterns of activations within the units of the nervous system. Functional connectivity refers to the study of functional interactions and aims to recognize statistical mutuality among signals acquired across various brain areas [27], [28]. The proposed study utilized two distinct phase-based non-linear brain functional connectivity entities to evaluate the proposed approach across the connectivity matrix of a graph. The graph's edges refer to the connectivity between the electrode pairs generated with the help of ISPC and PLI features. Phase-based connectivity investigates the synchronization of phases between two signals. There is synchronization between two electrodes when their neuronal units are functionally connected.

The idea of phase synchronization was broadly discussed by Rosenblum and coworkers [29]. In brief, the phase difference between two signals is constant when there is severe phase locking between them. In contrast, the easier concept introduced by Rosenblum et al. [29] in the context of phase entertainment only needs the difference of phase between two signals to be bounded ($\leq 2\pi$).

If ϕ_1 and ϕ_2 refers to the phase of signals and $\Delta\phi$ Refers to the phase difference between two signals, the general phase synchronization can be calculated as (n and m are constant integers) shown in "(3)":

$$\Delta\phi_{nm} \mid = |n\phi_1 - m\phi_2| \leq const \qquad (3)$$

Instantaneous phase differences are projected on the unit circle, and the length R of the average is given by "(4)":

$$R = |\frac{1}{N}\sum_{k=0}^{N-1}e^{i\Delta}\Phi(t_k)| \qquad (4)$$

Here (t_k) are discrete time steps, and N refers to the number of samples. The value of R is '1' for ideal phase synchronization, whereas R is zero when there is a random distribution of phases. R is oblivious to signal amplitude and depends solely on their phase correlations.

1) PHASE LAG INDEX

PLI is introduced as a novel measure of statistical interdependencies between the time series to obtain a consistent evaluation of invariant phase synchronization against the presence of familiar sources and amplitude effects like volume conduction or active reference electrodes, etc. [23]. PLI is a phase-based functional connectivity measure that ignores zero-phase lag connectivity to prevent spurious correlations. Distribution of phase angles primarily on the positive or negative side of the imaginary axis will result from the non-volume directed connectivity [2].

EEG data may display bogus associations between electrodes sensing the identical source with phase lags of 0 or π radians. The phase lag will be π radians if the nodes are on the opposite poles of the dipole. Therefore, such nonzero and constant phase lag indicates upfront linkages between the underlying systems and cannot be affected by volume conduction from a single powerful source.

PLI quantifies the extent to which the complex plane's phase angle difference distribution is oriented toward the positive or negative sides of the imaging axis, as indicated "(5)".

$$PLI_{mn} = \left|x^{-1}\sum_{t=1}^{x}sgn\left(imag\left(X_{mnt}\right)\right)\right| \qquad (5)$$

where sgn represents the -1 for negative values, $+1$ for positive values, and 0 for zero values [1], and imag (X) indicates the imaginary fraction of the cross-spectral density at trial t.

The central idea utilized in the PLI method is to abandon the phase differences centered around 0 and π. The likelihood that the distribution of phase angle differences $\Delta\phi$ are in the range $-\pi < \Delta\phi < 0$ is different from the likelihood they are between the range $0 < \Delta\phi < \pi$. This asymmetry indicates a constant, nonzero phase difference between the two signals of the time domain. The symmetric distribution exists when the median phase difference is centered around a value of 0 or π (influenced by volume conduction).

The value of PLI lies in the range of 0 to 1. A PLI value of '0' either specifies no synchronization or synchronization with a phase angle difference centered at 0 and π. A PLI of '1' specifies perfect lockup of phase angles with $\Delta\phi$ having value dissimilar from 0 and π. This paper employed the PLI to measure functional connectivity among all electrode pairs (64 for Dataset 1 and 62 for Dataset 2) at 30 frequencies (spaced between 4 Hz to 80 Hz for Dataset 1) and
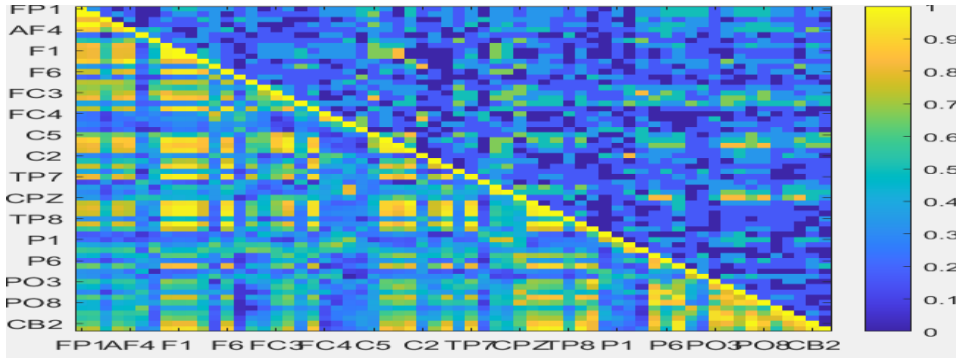
(a) **PLI and ISPC connectivity matrix before thresholding. The upper triangle corresponds to the PLI connectivity and the lower triangle corresponds to ISPC connectivity before thresholding.**
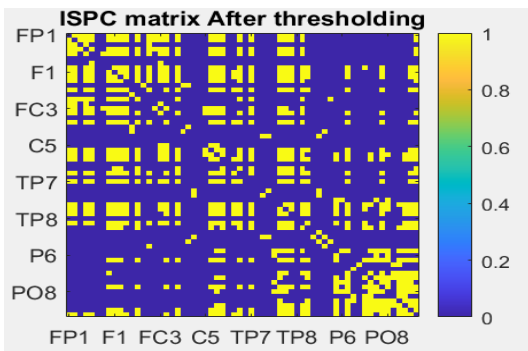
(b)**ISPC connectivity matrix after thresholding.**
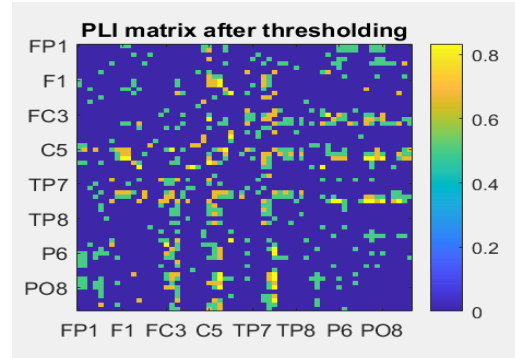
(c)**PLI connectivity matrix after thresholding**

FIGURE 1. Represents the all-to-all PLI and ISPC connectivity matrices at a frequency and time of 45hz and 2.5 second respectively for dataset 2. Both the axes(X-axis, Y-axis) refers to electrodes, and the intensity of colour at each point signifies the extent of the connectivity between electrode. Figure 1.a represents the joint connectivity matrix before thresholding and figure 1.b and figure 1.crepresents threshold connectivity matrix.

between 1 Hz to 50 Hz for Dataset 2) and all time points (1280 for dataset 1 and 5000 for dataset 2). It resulted in a matrix of size(E*E*F*T), i.e., $64 \times 64 \times 30 \times 1280$ for dataset 1 and a matrix of size $62 \times 62 \times 30 \times 5000$ for dataset 2, as shown in figure 5. This method addressed a common spurious and fake connectivity issue with phase-based connectivity measurements. Refer to for more information on PLI [23].

2) INTERSITE PHASE CLUSTERING

Interstice phase clustering (ISPC): ISPC measures phase synchronization between two signal sources in the brain. It can detect associations between brain regions that are not powerfully coupled in amplitude or power but have functional connectivity via phase synchronization. In this study, electrode connectivity was assessed for specific frequencies and time points, qualifying the recognition of substantial dynamics in electrode connectivity.
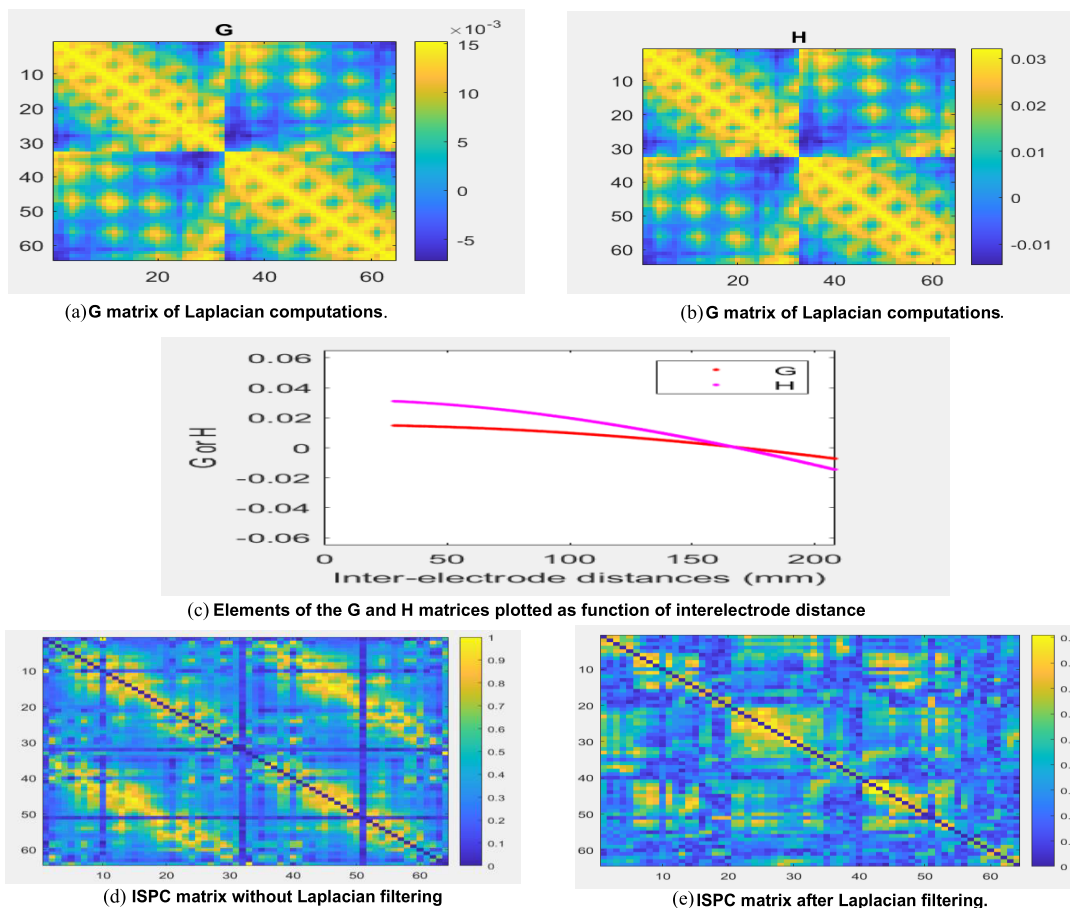
The term "ISPC" describes the clustering of phase angle differences between electrodes at various time points represented in polar space. Instead of relying on phase values, ISPC relies on clustering phase angle differences. Phase angle discrepancies between electrodes are averaged throughout time to calculate ISPC [2]. The phase angles of time-series signals are calculated using Discrete wavelet transform with complex morlet wavelet as mother wavelet. The expression "(6)" summarizes the procedure of calculating ISPC between two signals.

$$ISPC_f = |x^{-1} \sum_{t=1}^{x} e^{i(\varnothing mt - \varnothing nt)}| \qquad (6)$$

where $\varnothing m$ and $\varnothing n$ are the phase angle of electrodes 'm' and 'n' at frequencies 'f' and 'x' is the number of trials. To evaluate phase-based connectivity related to tasks, we have calculated ISPC across trials. The distribution of phase angle difference is computed at each trial over the time points with ISPC-time. While it is generated at each time point over trials with ISPC trials. Compared to ISPC-time, ISPC trials offer compelling evidence of task-related modulations in connectivity. Furthermore, ISPC-trials are computed independently at each time point, resulting in no loss of temporal precision [2].

The ISPC ranges between 0 and 1. An ISPC of zero specifies no phase synchronization. An ISPC of 1 specifies perfect phase synchronization between the time domain signals. This paper employed the ISPC to find functional connectivity between all electrode pairs (64 for Dataset 1 and 62 for Dataset 2) at 30 frequencies (spaced between 4 Hz to 80 Hz for Dataset 1 and between 1 Hz to 50 Hz for Dataset 2) and all time points (1280 for dataset 1 and 5000 for dataset 2). Resulting in a matrix of size $64 \times 64 \times 30 \times 1280$ for dataset 1 and a matrix of size $62 \times 62 \times 30 \times 5000$ for dataset 2, as is shown in the outline of the paper in Figure 5. A detailed description of ISPC can be found in [25] and [27].

(a) G matrix of Laplacian computations.

(b) G matrix of Laplacian computations.

(c) Elements of the G and H matrices plotted as function of interelectrode distance

(d) ISPC matrix without Laplacian filtering

(e) ISPC matrix after Laplacian filtering.

FIGURE 2. Effect of Laplacian filtering of EEG data before calculating ISPC to combats volume conduction. Figure 2.a and figure 2.b eletrode * eletrode weighting matrices 'G' and 'h' that were utilized in the Laplacian computations Figure 2.c represents Elements of the G and H matrices plotted as function of interelectrode distance, Figure 2.d shows ISPC matrix without applying Laplacian filtering to EEG data and Figure 2.e shows ISPC matrix after applying. Laplacian filtering. represents threshold connectivity matrix.

ISPC was selected because it delivers a high degree of temporal precision and is capable of offering relatively solid indications for variations in connectivity. It is perfect for EEG signal analysis due to its robustness to time lag and frequency non-stationarities. However, ISPC is prone to volume conduction, which restricts the precise identification of the source of the mind's electrical activity, sometimes resulting in inaccurate or ambiguous connectivity information. To address this issue, we have applied surface Laplacian transform to the data to attain robustness against the potential confound of volume conduction by successful filtration of spatially broad assemblies of data. The Laplacian is applied to the EEG data before time-frequency analysis. It attenuates low-spatial frequency features and will help isolate topographical features within several centimeters of range.

The spherical spline method by Perrin and colleagues is utilized in the study for surface Laplacian transform [30]. The computation of G and H matrices is the initial step in calculating Laplacian. The weighting matrices 'G' and 'H' are of size "electrode *electrode" and are used in the Laplacian computations as shown in Figure 2. a and 2. b, respectively. These matrices are utilized to attenuate low

spatial frequency components from the data. After that, the cosine distance among all pairs of electrodes is computed, and then the weighted sum of activity at all electrodes is subtracted from the activity of each electrode [25]. In the study, The Laplacian is calculated concurrently with the help of matrix algebra for all electrodes and all time points. Figure 2. c shows the elements of the matrices 'G' and 'H' plotted as a function of interelectrode distance. Figure 2.d shows the ISPC matrix without applying Laplacian filtering to EEG data, and Figure 2. e shows the ISPC matrix after the application of Laplacian filtering. Laplacian filtering has significantly attenuated the impact of volume conduction, as evident in Figure 2. e.

E. CONNECTIVITY DEGREE
The connectivity degree is a simple but practical measure of the extent to which a vertex acts as the intersection of many connections [25]. It is denoted by the sum of all the suprathreshold connections, except for the vertex itself, regardless of the location of these connections. After counting the number of subthreshold connections for each electrode, the input connectivity matrix of size

(a) PLI connectivity   (b) ISPC connectivity   (c) EEG power
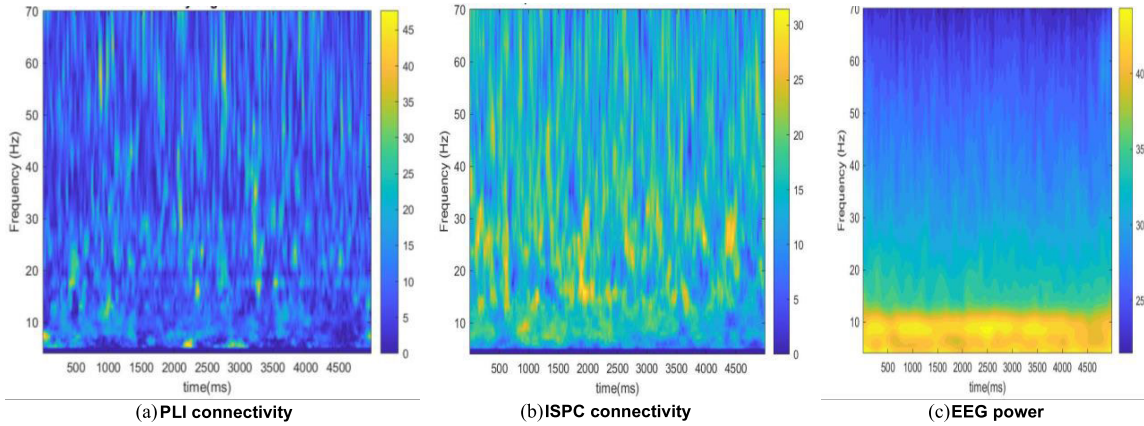
FIGURE 3. Represents the time-frequency image of extracted PLI, ISPC, and EEG power features. Where Figure 3. a represents the generated time-frequency image corresponding to PLI connectivity, Figure 3. b shows the generated time-frequency image corresponding to ISPC connectivity, and Figure 3. c corresponds to EEG power.

[electrode(E) × electrode(E) × frequency(F) × time(T)] is transformed into a new matrix of size [electrode(E) × frequency(F) × time(T)] as shown in outline of the paper in figure 5. The first dimension represents the connectivity degree at each electrode. It refers to the strength of association of that particular electrode with other electrodes. The size of the connectivity matrix for dataset 1 and dataset 2 is [64 × 30×1280] and [64 × 30×1280], respectively.

F. EEG POWER ANALYSIS

The power spectrum reflects the distribution of signal power over frequency. EEG power analysis is a well-established method of examination of EEG signals [25]. This study used discrete wavelet convolution to calculate the time variable frequency band's specific power from the imagined speech EEG data. The power is averaged across all trails of a particular prompt. To analyze the power spectrum characteristics of an imagined speech EEG data, power for 30 linearly spaced frequency components ranging from 4 Hz to 70Hz for dataset 1 and 1 to 50 Hz for dataset 2 was calculated at each time point, resulting in a power matrix of size 64 × 30×1280 for dataset 1 and 62 × 30×5000 for dataset 2. The units of power are db/Hz.

G. CONSTRUCTION OF TIME-FREQUENCY IMAGES

In this step, the signal classification task is converted into an image classification task by converting the extracted phase and power features into a time-frequency image. Time-frequency images facilitate simultaneous time-frequency representation of data and have become an effective means for analyzing and characterizing signal features like power and phase at different time and frequency points [31]. PLI connectivity, ISPC connectivity, and EEG power at each electrode are portrayed on a time-frequency image, leading to 192 images per prompt for Dataset 1 and 186 images per prompt for Dataset 2. MATLAB-generated images were of size 875 ×656. All images were resized to a standard size of 100 × 100 before feeding as input to the classifier section.

Figures 3. a, 3. b, and 3. c show the generated time-frequency image corresponding to PLI connectivity, ISPC connectivity, and EEG power, respectively.

Once the time-frequency images of all the extracted PLI, ISPC, and power features have been obtained for each subject, these images are provided as input into the neural networks for classifications and predictions. Most pre-trained neural networks, like VGG, Dense Net, Resnet, etc., are trained on image datasets and are valid for image classification problems. Hence, extracted features are converted into time-frequency images to utilize the classification ability of this network.

H. TRANSFER LEARNING ALGORITHM

Transfer learning is an influential deep learning technique utilized in the study to apply existing models and their knowledge to help the classification process [32], [33]. It has proven very effective in training deep networks even with limited data. In this study, fine-tuned DenseNet121 architecture is employed to classify the imagined speech datasets. The task is to transfer the learning of a DenseNet121 trained with ImageNet to a model that classifies time-frequency images of PLI connectivity, ISPC connectivity, and EEG power features extracted from the dataset.

DenseNet-121 provides exceptional model performance and offers various benefits, such as effective parameter utilization, enhanced gradient flow, and adaptableness. Dense Net architecture overcomes the problem of 'vanishing gradient' by connecting each layer directly with other layers, making the network densely connected. There are X(X+1)/2 direct associations for 'X' layers. This approach endorses the resourceful usage of features, leading to a more efficient network for feature propagation and reuse. After performing a qualitative comparison of the classification performance of DenseNet-121 with different deep learning (CNN), machine learning (SVM, RF), and transfer learning classifiers (ResNet-50), It is concluded that DenseNet-121architecture tends to provide higher classification accuracy with few
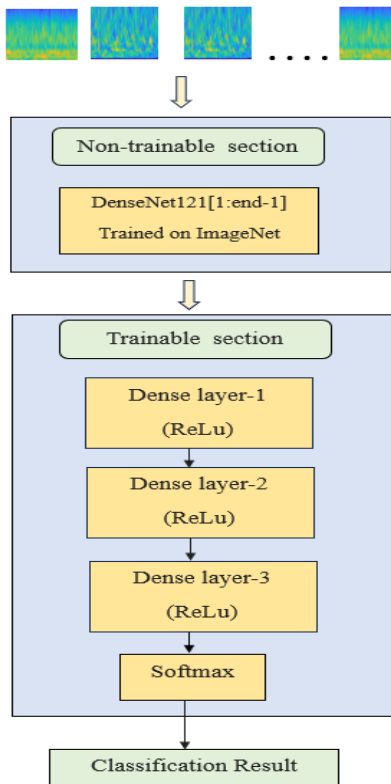
FIGURE 4. Architecture of proposed DenseNet-121-based transfer learning model for classifying imagined speech prompts. The input time-frequency images are of size 100*100*3. The size of Dense layer-1, Dense layer-2, and Dense layer-3 was set to 256, 128, and 64 units, respectively. The parameters of DenseNet121[1: end-1] were frozen, and only the dense layers appended after the DenseNet-121 model were trained with Dataset 1 and Dataset 2.

parameters and low computation load. More information on DenseNet-121 is available at [34]. Figure 4 represents the basic design of the classifying architecture DenseNet-121 employed in this work. The Dense Net 121 neural network was pre-trained on the images of a large dataset named ImageNet [35]. It was utilized as the base architecture of the classification process. Since this base architecture network is initially trained for classifying 1000 entities of the ImageNet dataset, the last fully connected layer has a size of 1000 with a Softmax activation function. This layer was replaced by three dense layers with the ReLU activation function. The size of Dense layer-1, Dense layer-2, and Dense layer-3 is set to 256, 128, and 64, respectively. The number of neurons in the output layer with Softmax activation function was set according to the number of classes in the imagined speech classification task (binary or multi-class). Dropout layers of rate 0.25 after each dense layer are not shown in Figure 4.

During the training on dataset 1 and dataset 2, All the layers of the Dense Net-121 model were frozen, and only the added dense layers (Dense layer-1, Dense layer-2, Dense layer-3) were trained. Thus, the Dense Net layers act as feature extraction layers. Adam optimizer [36] was employed to optimize the cross-entropy loss function, and a learning rate of 0.001 was chosen for optimization.

III. PROPOSED ARCHITECTURE

The applicability of modern transfer learning and feature engineering methods have seen substantial advances in EEG signal analysis [13], [37]. The outline of the paper is provided in Figure. 5. The initial step of the designed model is to extract phase-based brain functional connectivity and power features from both datasets. The proposed model utilized two distinct phase-based brain functional connectivity quantities (ISPC and PLI) to evaluate our approach across the connectivity matrix of the graph. Extracting time and frequency components of the EEG signal with the help of a complex market wavelet is the primary step for extracting power and phase features, as it extracts the estimations of time variable frequency band-specific amplitude and phase from the imagined speech EEG signal. Discrete wavelet transform with Complex Morlet wavelets as mother wavelet was used for time-frequency decomposition. Surface Laplacian was applied to the data before extracting ISPC connectivity to overcome the volume conduction effects. The connectivity degree is used as a network measure to study the interactions between different neuronal units. The nodes of the graph refer to the electrodes, and the edges of the graph refer to the association among paired electrodes. Time-varying frequency band-specific power was calculated to analyze the power spectrum characteristics of imagined speech EEG data.

In the second step, the Signal classification task is converted into an image classification task by converting extracted phase and power features in a time-frequency image. This paper applies the transfer learning technique to classify imagined speech EEG data. Time-frequency images of PLI and ISPC connectivity and power features were classified with the help of the Dense Net 121 model pre-trained on a large dataset with 1000 different categories of images.

In this paper, two open-access datasets were utilized to validate the effectiveness of the proposed architecture. The designed transfer learning-based model performed the classification of imagined prompts. For a classification task of N classes, the chance level accuracy (C-Accuracy) is defined in "(7)":

$$\text{C-Accuracy}(\%) = (1/N) \times 100 \qquad (7)$$

10-fold cross-validation is performed to validate the acquired classification results. The data was divided into ten folds, and during each cross-validation iteration, nine folds were employed for the training process and one for the testing process. This procedure is repeated for ten iterations to test data from all the folds once. During each cross-validation step, 80% of the training data was utilized for the training purpose and the remaining 20% for the validation purpose. The number of epochs for the training process was set to 100, and the batch size was set to 16. ADAM optimizer is used with an initial learning rate of 0.001 to tune the parameters of the neural network and in order to minimize the cost function. Hyperparameters including the number and size of Dense feature extraction layers in the trainable section, optimizer type, cost function, learning rate, batch size, number of
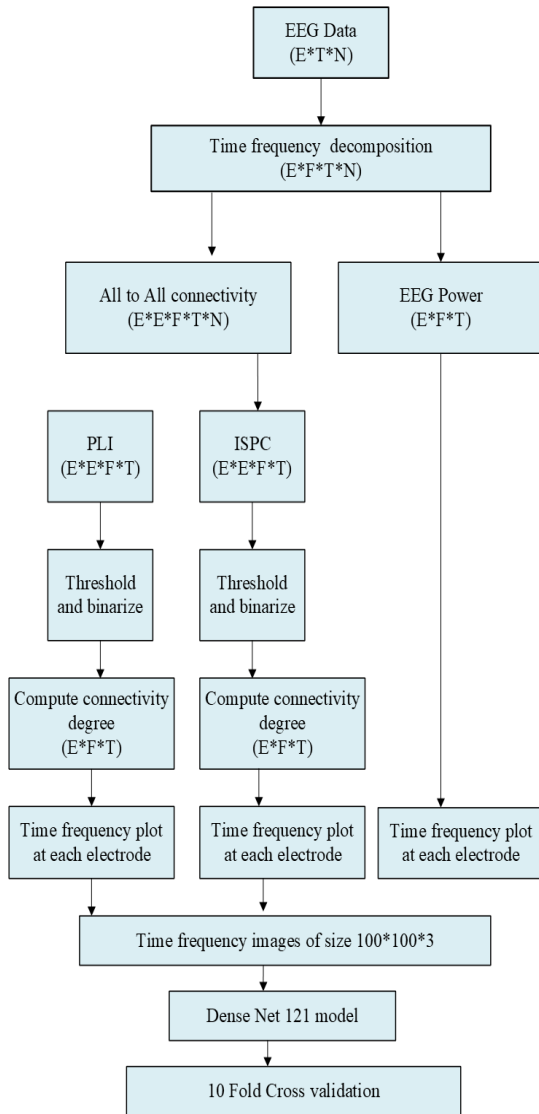
FIGURE 5. Outline of the paper. 'E' represents Number of electrodes (64 for dataset 1 and 62 for dataset 2), 'F' represents number of frequency points (30 points between 8Hz to 70Hz for dataset 1 and 30 points between 2Hz to 50Hz for dataset 2), 'T' represents Number of time points (1280 for dataset 1 and 5000 for dataset 2) and 'N' represents number of trails (100 for dataset 1, 12 or 15 for dataset 2).

epochs, dropout rate, etc., have a great influence on network performance, but there is no detailed strategy to choose their values, doing a large number of iterations is the only way to determine the values of these hyperparameters. Regularization techniques such as dropout have been adopted in the fine-tuning network to diminish model overfitting as it tends to advance the algorithm's performance even on the statistics that are outside the training set. Early stopping was applied in our experiment to prevent overfitting and imposing smoothness constraints. Early stopping was implemented based on validation accuracy with a tolerance of 30 epochs to overcome the effects of overfitting. Table 1 contains the training parameters of the proposed DenseNet-121-based transfer learning

TABLE 1. Training parameters of proposed DenseNet-121 based transfer learning mode.

Optimizer	Initial learning rate	Loss function	Epoch	Batch size
Adam	0.001	Categorical cross-entropy	100	16

TABLE 2. Specification of layers used in trainable section of proposed DenseNet-121 based transfer learning model.

Layer Name	Layer type	size	Activation function
Dense layer_1	Fully connected Dense layer	256	Relu
Dense layer_2	Fully connected Dense layer	128	Relu
Dense layer_3	Fully connected Dense layer	64	Relu
Softmax	Classification layer	2/3/4	Softmax
Layer Name	Layer type	Dropout rate	
Dropout_1	Drop out layer	0.25	
Dropout_2	Drop out layer	0.25	
Dropout_3	Drop out layer	0.25	

model. Table 2 contains specifications of layers used in the trainable section of the proposed Architecture.

IV. RESULTS AND COMPARISON

Algorithm Evaluation: A tenfold cross-validation method was employed to confirm the legitimacy of classification outcomes. All the designated trials were subjectively alienated into ten equally sized subdivisions, where 80% of data was used for training and 20% was used for validation purpose. Accuracy, Precision, recall, and F1 score were utilized as measures for classifier evaluation. The average accuracy, precision, F1 score and recall for N class classification was calculated using "(8)", "(9)", "(10)" and "(11)", respectively.

$$\text{Average Accuracy} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \quad (8)$$

$$\text{Precision} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i} \quad (9)$$

$$\text{Recall} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FN_i} \quad (10)$$

$$\text{F1 Score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (11)$$

where TP denotes the true positive rate, FP denotes the false positive rate, TN denotes the true negative rate, and FN denotes the false negative rate. All these quantities (TP, FP, TN, FN) refer to the numeral of predicted positives or negatives that were accurate or inaccurate.

TABLE 3. Percentage classification accuracy, precision, recall, and F1-score obtained with proposed DenseNet-121 model for long words "independent" and "cooperate" of dataset 1.

Subject	Independent	Cooperate	Average Accuracy	Average Precision	Average Recall	Average F1-score
2	92.31	97.37	94.84	94.9	95.12	94.89
6	98.45	98.95	98.7	98.7	98.95	98.73
7	89.74	100	94.87	94.91	95.2	94.98
9	94.87	86.84	90.85	90.91	91.29	91.1
11	100	100	100	100	100	100
Average	95.07	96.63	95.85	95.88	96.11	95.94

TABLE 4. Percentage classification accuracy, precision, recall, and f1-score obtained with proposed DenseNet-121 model for long vs short words "in" and "cooperate" of dataset 1.

Subject	In	Cooperate	Average Accuracy	Average Precision	Average Recall	Average F1-score
1	97.44	100	98.72	98.76	98.91	98.83
5	92.31	92.11	92.21	92.22	92.61	92.19
8	92.31	97.37	94.84	94.91	95.02	94.93
9	97.44	100	98.72	98.72	98.84	98.79
10	92.34	97.37	94.85	98.87	99.25	99.14
14	82.05	89.47	85.76	85.8	86.14	85.81
Average	92.39	96.05	94.18	94.88	95.12	94.94

TABLE 5. Percentage classification accuracy, precision, recall, and f1-score obtained with proposed DenseNet-121 model for short words "in", "out" and "up" of dataset 1.

Subject	In	Out	Up	Average Accuracy	Average Precision	Average Recall	Average F1-score
1	92.31	100	100	97.43	97.44	97.71	97.39
3	84.62	97.44	100	94.02	94	94.2	94.09
5	84.62	87.18	94.74	88.84	88.91	89.17	89
8	94.87	94.87	97.37	95.70	95.70	95.97	95.8
12	97.44	100	100	99.14	99.2	99.76	99.38
Average	96.91	95.89	98.42	95.02	95.05	95.36	95.13

Results: The investigational results show that the proposed method can classify imagined speech with high accuracy and efficiency as compared to the state of the arts. High multi-class classification accuracy and good generalization ability of the designed algorithm can deliver assurance for the implementation of the designed network architecture in real-time BCI system designs. The classification of imagined speech EEG signals was carried out using various deep learning and machine learning methods, namely, convolutional neural networks, support vector machine, random forest, ResNet-50, and DenseNet-121. After testing and comparing with lots of deep learning and machine learning methods, the DenseNet-121 architecture with Phase connectivity and power features provided the best multi-class classification accuracy, while others have failed badly, providing poor multi-class classification results except for the ResNet-50-based classifier. All of these models used almost the same training and testing procedure, yet the accuracy of the proposed model is appreciably higher. The employed feature engineering methods and fine-tuned Dense-Net-121 architecture are responsible for this enormous upsurge in performance.

A. DATASET 1

Two binary (long words, long vs short words) and two multiclass (Short words, vowel) classifications were performed on dataset 1. Percentage classification accuracy, precision, recall, and F1-score obtained with the proposed DenseNet-121-based transfer learning model with long words, long vs. short words, short words, and vowel classes are shown in TABLE 3, TABLE 4, TABLE 5, TABLE 6, respectively. After feature extraction, the foremost step is choosing the finest classifier to complete the classification process. Machine learning (Support vector machine, random forest (RF)), deep learning (Convolution neural network), and transfer learning methods (ResNet-50, DenseNet-121) were explored. Classification accuracies have excelled the chance level accuracy for all the adopted classifiers, proving the efficiency of ISPC, PLI connectivity, and EEG power features in classifying imagined speech. The transfer learning approach using fine-tuned DenseNet-121 architecture achieved the best classification results among all the classifiers. The classification results obtained for dataset 1 using convolution neural network (CNN), ResNet-50, and DenseNet-121classifiers are shown in TABLE 7, TABLE 8, TABLE 9, and Table 10,

TABLE 6. Percentage classification accuracy, precision, recall, and f1-score obtained with proposed DenseNet-121 model for vowels "/A/", "/I/" and "/U/" of dataset 1.

Subject	/a/	/i/	/u/	Average Accuracy	Average Precision	Average Recall	Average F1-score
8	86.62	97.44	97.37	93.81	93.9	94.1	94
9	94.87	97.44	100	97.43	97.47	97.6	97.4
11	92.31	89.74	97.37	93.14	93.14	93.32	93.2
12	94.87	100	100	98.29	98.31	98.4	98.2
13	100	94.87	100	98.29	98.3	98.67	98.3
15	92.31	76.92	89.47	86.23	86.24	86.48	86.4
Average	93.49	92.73	97.36	94.53	94.56	94.76	94.58

TABLE 7. Classification accuracy (%) obtained for dataset 1 considering long words classification with CNN, ResNet-50, AND Dense net -121 classifiers.

Long words (Independent Vs Cooperate)	Subject	2	6	7	9	11	Average
	Accuracy (CNN)	53.12	67.27	45.92	64.73	59.36	58.08
	Accuracy (Resnet-50)	82.39	84.38	79.97	84.64	76.63	81.60
	Accuracy (Dense Net -121)	94.84	98.7	94.87	90.85	100	95.85

TABLE 8. Classification accuracy (%) obtained for dataset 1 considering long vs short words classification with CNN, ResNet-50, and Dense net -121 classifiers.

Long vs short (In vs Cooperate)	Subject	1	5	8	9	10	14	Average
	Accuracy (CNN)	54.86	58.84	49.40	53.58	57.39	53.92	54.66
	Accuracy (ResNet-50)	91.56	89.69	88.11	84.48	81.09	89.41	87.39
	Accuracy (Dense Net -121)	98.72	92.21	94.84	97.22	94.84	85.76	93.94

TABLE 9. Classification accuracy (%) obtained for dataset 1 considering vowel classification with CNN ResNet-50, and Dense net -121 classifiers.

Vowel ('a', 'i', 'u')	Subject	8	9	11	12	13	15	Average
	Accuracy (CNN)	43.38	49.49	48	39.37	42.77	45.05	44.67
	Accuracy (ResNet-50)	72.85	74.60	71.17	75.93	79.04	82.62	76.03
	Accuracy (Dense Net -121)	93.14	97.14	93.2	98.29	98.3	86.23	94.35

TABLE 10. Classification accuracy (%) obtained for dataset 1 considering short word classification with CNN, ResNet-50, and dense net - 121 classifiers.

Short Words (In, out, up)	Subject	1	3	5	8	12	Average
	Accuracy (CNN)	51.62	56.9	62.73	51.28	58.37	56.18
	Accuracy (Resnet-50)	78.62	82.59	84.39	71.26	84.41	80.24
	Accuracy (Dense Net -121)	97.436	94.02	88.84	95.70	99.146	95.02

considering long words vs. short words, vowels, and short words classification, respectively.

The overall two-class classification accuracies achieved with the DenseNet121 classifier were 95.85% and 93.94% for long words and long vs. short words, respectively. The overall three-class classification accuracies achieved with the DenseNet-121 classifier were 94.35% and 95.02% for vowels and short words, respectively. Based on the accuracies obtained from the classification of time-frequency images of PLI connectivity, ISPC connectivity, and EEG power, Figure 6 shows classification results obtained for dataset 1 concerning the Machine learning classifiers (SVM and random forest), 7-layer convolution neural network, and transfer learning methods (ResNet-50, DenseNet-121) with extracted phase and power features. The DenseNet-121 algorithm outperforms extracted features by bringing the accuracy above 90% for most cases. The Maximum accuracy obtained by DenseNet-121classifier is 100% for long-word classification from subject 11. Also, after comparison of the accuracy values attained by SVM, random forest, CNN, and ResNet-50 classifiers, it is strongly evident that the DenseNet-121 classifier, in amalgamation with time-frequency images of

phase-based connectivity and power feature, offers better classification accuracy among all chosen classification methods. All the results were obtained after 10-fold cross-validation.

B. DATASET 2
The designed transfer learning enabled phase-based brain functional connectivity and power for an electroencephalography-based intuitive brain-computer interface system tested and created with dataset 1. Dataset 2 is utilized to validate the results of the developed model. Four class classification of the words ('pat,' 'pot,' 'knew,' 'gnaw') of dataset 2 was performed with the help of the DenseNet-121 classifier. Classification accuracies have excelled the chance level accuracy for all the classification methods, proving the efficiency of ISPC and PLI connectivity and EEG power features in classifying imagined speech. The multiclass (4 class) classification accuracy, precision, recall, and F1-score obtained for dataset 2 using DenseNet-121 are given in TABLE 11. The average four-class classification accuracy obtained using the DenseNet-121 classifier was 92.69%. TABLE 11 shows that the DenseNet-121 algorithm performs

TABLE 11. Percentage classification accuracy, precision, recall, f1-score obtained for dataset 2 considering words ("PAT", "POT", "KNEW", "GNAW") classification using dense net -121 classifier.

Subject	Pat	Pot	Gnaw	Knew	Average accuracy	Average precision	Average recall	Average F-1 score
MM05	80.77	96	100	100	94.192	94.2	94.7	94.1
MM08	80.77	100	100	100	95.20	95.2	95.6	95.1
MM09	76.92	100	92	100	92.23	92.3	92.4	92.2
MM10	76.92	96	96	95.83	91.187	92.2	92.7	91.2
MM11	73.08	88	88	95.83	86.20	86.3	86.5	86.1
MM12	80.77	100	96	100	94.2	94.3	94.4	94.1
MM14	76.92	100	100	100	94.23	94.2	94.7	94
P02	80.77	100	100	95.83	94.15	94.2	94.2	94
Average	78.36	97.5	96.5	98.43	92.68	92.86	93.15	92.6

TABLE 12. Comparison of the average accuracy (%) of the proposed algorithm with the state of art methods from the literature for dataset 1.

Long words	Paper	Proposed	[16]	[19]	[13]	[38]	[39]
	Accuracy	95.85	66.31	88.81	73.74	79.98	62.99
Short Words	Paper	Proposed	[16]	[19]		[13]	[38]
	Accuracy	95.02	50.06	83.95		71.54	73.66
Long vs Short	Paper	Proposed	[16]		[19]		[13]
	Accuracy	93.94	80.05		92.8		73.57
Vowels	Paper	Proposed	[16]	[19]		[13]	[38]
	Accuracy	94.35	48.96	86.27		66.16	74.25
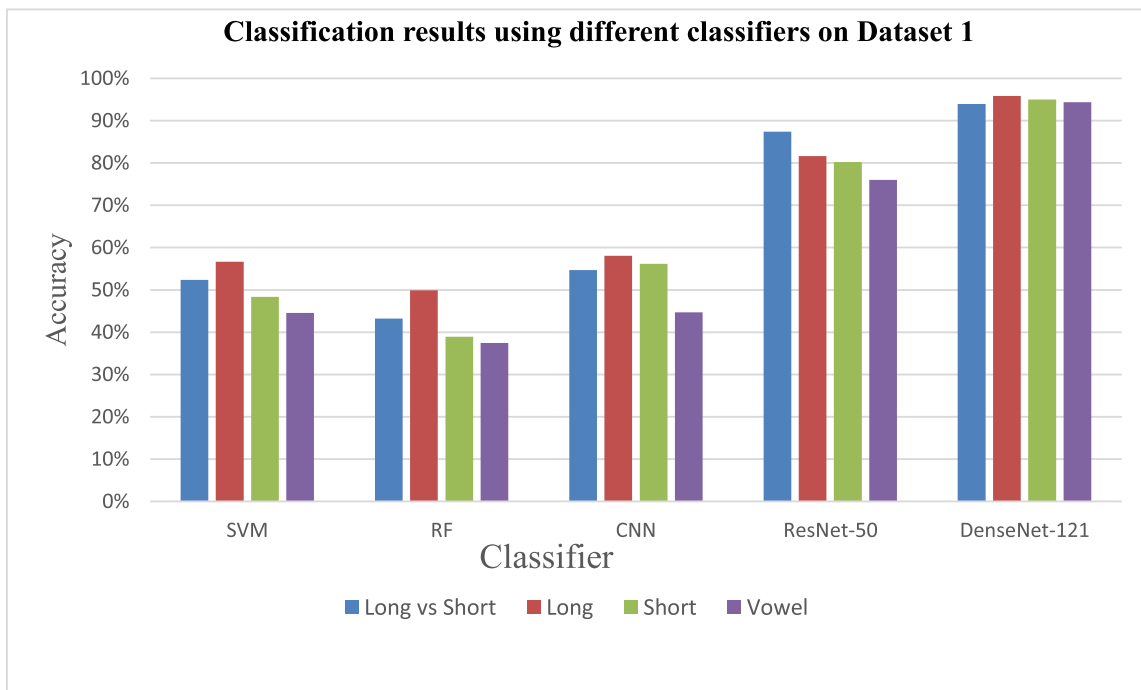
FIGURE 6. Average accuracy (%) values obtained for dataset 1 for different machine learning classifiers like SVM (support vector machine), RF (random forest), and deep learning classifiers like CNN (convolution neural networks), ResNet-50, and DenseNet-121.

excellently with extracted features by bringing the accuracy above 85% for all the cases. The Maximum accuracy obtained by DenseNet-121classifier is 95.2% for four-word classification from subject MM08. The accomplished classification results make it clear that DenseNet-121, combined with time-frequency images of phase-based connectivity and power features, provides better classification results than the state of the arts. All the results were obtained after 10-fold cross-validation.

C. COMPARISION

In order to appraise the efficiency of the proposed method, the accuracies obtained with the proposed transfer learning-enabled phase-based brain functional connectivity and power features architecture were compared with the other imagined speech decoding methods in the literature. The effectiveness of the proposed algorithm is clearly evident in TABLE 12 and TABLE 13, which compares the average accuracy of the proposed method with the state-of-the-art methods from

TABLE 13. Comparison of the average accuracy (%) of the proposed algorithm with the state of art methods from the literature for dataset 2.

Words	Paper	Proposed	[21]	[20]
'pat,' 'pot,' 'knew,' 'gnaw.'	Accuracy	92.68	43.76	60.63

the literature of imagined speech decoding with the help of Dataset 1 and Dataset 2, respectively. Classification accuracies have excelled among all the literature results in the arena of imagined speech decoding, proving the efficiency of ISPC and PLI connectivity and EEG power features in classifying imagined speech. The excellent results strengthen the possibility of real-time EEG-based intuitive brain-computer interface communication.

V. CONCLUSION

This study proposes a novel method to design an imagined speech based brain-computer interface system. This work has achieved the highest classification accuracies in imagined speech decoding compared to the state of the arts. The phase-based connectivity features, EEG power features, and transfer learning technology have proven significant excellence in classifying speech imagination EEG data compared to other techniques in the literature of speech imagery decoding. PLI connectivity, ISPC connectivity, and EEG power features were selected after excessive exploration and testing of classification accuracy for numerous time domain and frequency domain features. The DenseNet-121 classifier was selected after qualitatively comparing results with other classifiers such as ResNet-50, 7-layer CNN, SVM, random forest, etc. A comparison of the proposed study with literature studies from the speech imagination field verified that the proposed technique offers outstanding accuracies as compared to the state of the arts. The Effectiveness of the proposed algorithm is evaluated on two publicly available datasets, ensuring the potential of real-time imagined speech interpretation system design. The availability of an exhaustive research-grade dataset is one major hindrance in the design of an imagined speech-based real-time brain-computer interface system. Also, there are other major issues like poor signal-to-noise ratio, poor spatial resolution, and a limited amount of data to train deep learning architecture. In future works, we would like to acquire our datasets and validate the performance of the proposed methodology on the acquired datasets. We also plan to build more of the latest feature extraction and classification methodologies for efficiently decoding imagined speech by overcoming above mentioned limitations. It is also planned to explore more phase-based approaches like instantaneous phase difference sequence, phase coherence, phase coupling, etc., to enhance the decoding performance. We would also like to deeply explore some other graph theory-based connectivity measures like local efficiency, betweenness centrality, participation coefficient, clustering coefficient etc. The future aim is to design a real-time brain-computer interface system for interpreting imagined speech using an EEG data acquisition

device and real-time classification algorithm. The long-term vision of our research is to provide communication aid to individuals who are physically incapacitated but psychologically conscious by the successful real-time implementation of imagined speech EEG data.

REFERENCES

[1] K. Barnova, M. Mikolasova, R. V. Kahankova, R. Jaros, A. Kawala-Sterniuk, V. Snasel, S. Mirjalili, M. Pelc, and R. Martinek, "Implementation of artificial intelligence and machine learning-based methods in brain–computer interaction," *Comput. Biol. Med.*, vol. 163, Sep. 2023, Art. no. 107135, doi: 10.1016/j.compbiomed.2023.107135.

[2] M. Bisla and R. S. Anand, "Analysis of imagined speech characteristics using phase-based connectivity measures," in *Proc. IEEE 13th Int. Conf. Control Syst., Comput. Eng. (ICCSCE)*, Aug. 2023, pp. 68–73, doi: 10.1109/iccsce58721.2023.10237139.

[3] F. H. Guenther, J. S. Brumberg, E. J. Wright, A. Nieto-Castanon, J. A. Tourville, M. Panko, R. Law, S. A. Siebert, J. L. Bartels, D. S. Andreasen, P. Ehirim, H. Mao, and P. R. Kennedy, "A wireless brain–machine interface for real-time speech synthesis," *PLoS ONE*, vol. 4, no. 12, p. e8218, Dec. 2009, doi: 10.1371/journal.pone.0008218.

[4] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain–computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, Jul. 2001, doi: 10.1109/5.939829.

[5] N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, "EEG-based brain–computer interfaces using motor-imagery: Techniques and challenges," *Sensors*, vol. 19, no. 6, p. 1423, Mar. 2019, doi: 10.3390/s19061423.

[6] C. Guan, M. Thulasidas, and J. Wu, "High performance p300 speller for brain–computer interface," in *Proc. IEEE Int. Workshop Biomed. Circuits Syst.*, Dec. 2004, pp. 293–296, doi: 10.1109/BIOCAS.2004.1454155.

[7] B. Allison, T. Luth, D. Valbuena, A. Teymourian, I. Volosyak, and A. Graser, "BCI demographics: How many (and what kinds of) people can use an SSVEP BCI?" *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 2, pp. 107–116, Apr. 2010, doi: 10.1109/TNSRE.2009.2039495.

[8] C. Park, D. Looney, N. U. Rehman, A. Ahrabian, and D. P. Mandic, "Classification of motor imagery BCI using multivariate empirical mode decomposition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 1, pp. 10–22, Jan. 2013, doi: 10.1109/TNSRE.2012.2229296.

[9] Z. Işcan and V. V. Nikulin, "Steady state visual evoked potential (SSVEP) based brain–computer interface (BCI) performance under different perturbations," *PLoS ONE*, vol. 13, no. 1, Jan. 2018, Art. no. e0191673, doi: 10.1371/journal.pone.0191673.

[10] M. Bisla and R. S. Anand, "Wearable EEG technology for the brain–computer interface," in *Computational Intelligence in Healthcare Applications*. New York, NY, USA: Academic, 2022, pp. 137–155, doi: 10.1016/B978-0-323-99031-8.00005-3.

[11] P. Saha, M. Abdul-Mageed, and S. Fels, "SPEAK YOUR MIND! Towards imagined speech recognition with hierarchical deep learning," in *Proc. Interspeech*, Sep. 2019, pp. 141–145, doi: 10.21437/interspeech.2019-3041.

[12] K. Mohanchandra and S. Saha, "A communication paradigm using subvocalized speech: Translating brain signals into speech," *Augmented Human Res.*, vol. 1, no. 1, p. 3, Dec. 2016, doi: 10.1007/s41133-016-0001-z.

[13] J. T. Panachakel and R. A. Ganesan, "Decoding imagined speech from EEG using transfer learning," *IEEE Access*, vol. 9, pp. 135371–135383, 2021, doi: 10.1109/ACCESS.2021.3116196.

[14] T. J. La Vaque, "The history of EEG Hans Berger," *J. Neurotherapy*, vol. 3, no. 2, pp. 1–9, Apr. 1999, doi: 10.1300/j184v03n02_01.

[15] C. S. DaSalla, H. Kambara, M. Sato, and Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns," *Neural Netw.*, vol. 22, no. 9, pp. 1334–1339, Nov. 2009, doi: 10.1016/j.neunet.2009.05.008.

[16] K. Brigham and B. V. K. V. Kumar, "Imagined speech classification with EEG signals for silent communication: A preliminary investigation into synthetic telepathy," in *Proc. 4th Int. Conf. Bioinf. Biomed. Eng.*, Jun. 2010, pp. 1–4, doi: 10.1109/ICBBE.2010.5515807.

[17] C. H. Nguyen, G. K. Karavas, and P. Artemiadis, "Inferring imagined speech using EEG signals: A new approach using Riemannian manifold features," *J. Neural Eng.*, vol. 15, no. 1, Feb. 2018, Art. no. 016002, doi: 10.1088/1741-2552/aa8235.

[18] C. Cooney, R. Folli, and D. Coyle, "Optimizing layers improves CNN generalization and transfer learning for imagined speech decoding from EEG," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Oct. 2019, pp. 1311–1316, doi: 10.1109/SMC.2019.8914246.

[19] P. Saha, S. Fels, and M. Abdul-Mageed, "Deep learning the EEG manifold for phonological categorization from active thoughts," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2762–2766, doi: 10.1109/ICASSP.2019.8682330.

[20] A. Kamble, P. Ghare, and V. Kumar, "Machine-learning-enabled adaptive signal decomposition for a brain–computer interface using EEG," *Biomed. Signal Process. Control*, vol. 74, Apr. 2022, Art. no. 103526, doi: 10.1016/j.bspc.2022.103526.

[21] M. Bisla and R. S. Anand, "EEG based brain computer interface system for decoding covert speech using deep neural networks," in *Proc. IEEE 12th Int. Conf. Commun. Syst. Netw. Technol. (CSNT)*, Apr. 2023, pp. 414–419, doi: 10.1109/CSNT57126.2023.10134633.

[22] K. Raeisi, M. Mohebbi, M. Khazaei, M. Seraji, and A. Yoonessi, "Phase-synchrony evaluation of EEG signals for multiple sclerosis diagnosis based on bivariate empirical mode decomposition during a visual task," *Comput. Biol. Med.*, vol. 117, Feb. 2020, Art. no. 103596, doi: 10.1016/j.compbiomed.2019.103596.

[23] C. J. Stam, G. Nolte, and A. Daffertshofer, "Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources," *Human Brain Mapping*, vol. 28, no. 11, pp. 1178–1193, Nov. 2007, doi: 10.1002/hbm.20346.

[24] S. Zhao and F. Rudzicz, "Classifying phonological categories in imagined and articulated speech," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 992–996, doi: 10.1109/ICASSP.2015.7178118.

[25] M. X. Cohen, *Analyzing Neural Time Series Data*. Cambridge, MA, USA: MIT Press, 2014, doi: 10.7551/mitpress/9609.001.0001.

[26] G. Liu, L. Tian, and W. Zhou, "Multiscale time-frequency method for multiclass motor imagery brain computer interface," *Comput. Biol. Med.*, vol. 143, Apr. 2022, Art. no. 105299, doi: 10.1016/j.compbiomed.2022.105299.

[27] L. Lee, L. M. Harrison, and A. Mechelli, "A report of the functional connectivity workshop, Dusseldorf 2002," *NeuroImage*, vol. 19, no. 2, pp. 457–465, Jun. 2003, doi: 10.1016/s1053-8119(03)00062-4.

[28] J. Li, H. Hua, Z. Xu, L. Shu, X. Xu, F. Kuang, and S. Wu, "Cross-subject EEG emotion recognition combined with connectivity features and meta-transfer learning," *Comput. Biol. Med.*, vol. 145, Jun. 2022, Art. no. 105519, doi: 10.1016/j.compbiomed.2022.105519.

[29] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, "Phase synchronization of chaotic oscillators," *Phys. Rev. Lett.*, vol. 76, no. 11, pp. 1804–1807, Mar. 1996, doi: 10.1103/physrevlett.76.1804.

[30] F. Perrin, O. Bertrand, and J. Pernier, "Scalp current density mapping: Value and estimation from potential data," *IEEE Trans. Biomed. Eng.*, vol. BME-34, no. 4, pp. 283–288, Apr. 1987, doi: 10.1109/TBME.1987.326089.

[31] I. Shafi, A. Aziz, S. Din, and I. Ashraf, "Reduced features set neural network approach based on high-resolution time-frequency images for cardiac abnormality detection," *Comput. Biol. Med.*, vol. 145, Jun. 2022, Art. no. 105425, doi: 10.1016/j.compbiomed.2022.105425.

[32] M. A. Morid, A. Borjali, and G. Del Fiol, "A scoping review of transfer learning research on medical image analysis using ImageNet," *Comput. Biol. Med.*, vol. 128, Jan. 2021, Art. no. 104115, doi: 10.1016/j.compbiomed.2020.104115.

[33] L.-J. Lo, C.-T. Yang, W.-C. Chiang, and H.-H. Lin, "A quantitative method for the assessment of facial attractiveness based on transfer learning with fine-grained image classification," *Pattern Recognit.*, vol. 145, Jan. 2024, Art. no. 109970, doi: 10.1016/j.patcog.2023.109970.

[34] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.

[35] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[37] Z. Khademi, F. Ebrahimi, and H. M. Kordy, "A transfer learning-based CNN and LSTM hybrid deep learning model to classify motor imagery EEG signals," *Comput. Biol. Med.*, vol. 143, Apr. 2022, Art. no. 105288, doi: 10.1016/j.compbiomed.2022.105288.

[38] P. Saha and S. Fels, "Hierarchical deep feature learning for decoding imagined speech from EEG," in *Proc. AAAI Conf. Artif. Intell.*, Jul. 2019, vol. 33, no. 1, pp. 10019–10020, doi: 10.1609/aaai.v33i01.330110019.

[39] M. Jiménez-Guarneros and P. Gómez-Gil, "Standardization-refinement domain adaptation method for cross-subject EEG-based classification in imagined speech recognition," *Pattern Recognit. Lett.*, vol. 141, pp. 54–60, Jan. 2021, doi: 10.1016/j.patrec.2020.11.013.

MEENAKSHI BISLA received the B.Tech. degree in electronics and communication engineering from Lovely Professional University, Jalandhar, India, in 2015, and the M.Tech. degree in signal processing from Guru Govind Singh Indraprastha, Delhi, India, in 2018. She is currently pursuing the Ph.D. degree in electrical engineering with Indian Institute of Technology Roorkee, India. From 2018 to 2019, she was an Assistant Professor with Chandigarh Group of Colleges, India. Her research interests include deep learning, medical signal processing, neuroscience, and rehabilitation. She was a recipient of the Ministry of Human Resource and Development (MHRD) Research Fellowship, since 2019.

RADHEY SHYAM ANAND received the B.E. degree in electronics and communication engineering and the M.Tech. and Ph.D. degrees in electrical engineering from Indian Institute of Technology Roorkee, India. Currently, he is a Professor with the Department of Electrical Engineering, Indian Institute of Technology Roorkee. His research interests include medical image processing, signal processing, machine learning, deep learning, ECG, EEG, and ultrasound signal analysis.

• • •