



## Exploration of face-perceptual ability by EEG induced deep learning algorithm

Lidia Ghosh <sup>\*</sup>, Dipayan Dewan, Abir Chowdhury, Amit Konar

Jadavpur University, Electronics and Telecommunication Department, 188, Raja S.C. Mallick Rd, Kolkata, India



### ARTICLE INFO

**Keywords:**

Face perception-ability  
Electroencephalography  
Deep learning  
Prosopagnosia

### ABSTRACT

Face perception essentially refers to an individual's ability to understand and interpret a familiar face. This paper attempts to quantify the face perceptual ability of human subjects using their memory responses, acquired by an electroencephalographic (EEG) device, during engagement of the human subjects in a face recognition task. The entire experimental protocol is designed in the settings of pattern recognition, comprising five main steps, namely artifact removal, feature extraction, classifier training and testing, and face-perception ability measurement. The primary objective of the paper is to design a deep neural network model that utilizes both the temporal and spatial EEG features to categorize the familiar face (FF) and unfamiliar face (UF) responses of human subjects from their brain signals. To extract the spatial information, the acquired raw EEG data is transformed into multi-spectral 2-dimensional images by a deep learning approach. The spatial information, along with the EEG time-frequency domain (temporal) features are then utilized to train the proposed deep neural network, which can preserve the spatial as well as temporal EEG features that are more robust and less sensitive to the variations along each dimension. The proposed deep neural network paradigm demonstrates promising results in classifying the EEG responses for FF and UF recognition with high classification accuracy.

Biological underpinning of the cortical mechanism using sLORETA and event related potential (ERP) analysis in perceiving FFs and UFs by both the healthy controls and prosopagnosic patients is also a significant inclusion in the paper. The first and foremost advantage of the proposed work is that it provides a technological means to assess the variations in face perceptual ability of individuals. It can also be used as a biomarker for early prosopagnosia diagnosis.

### 1. Introduction

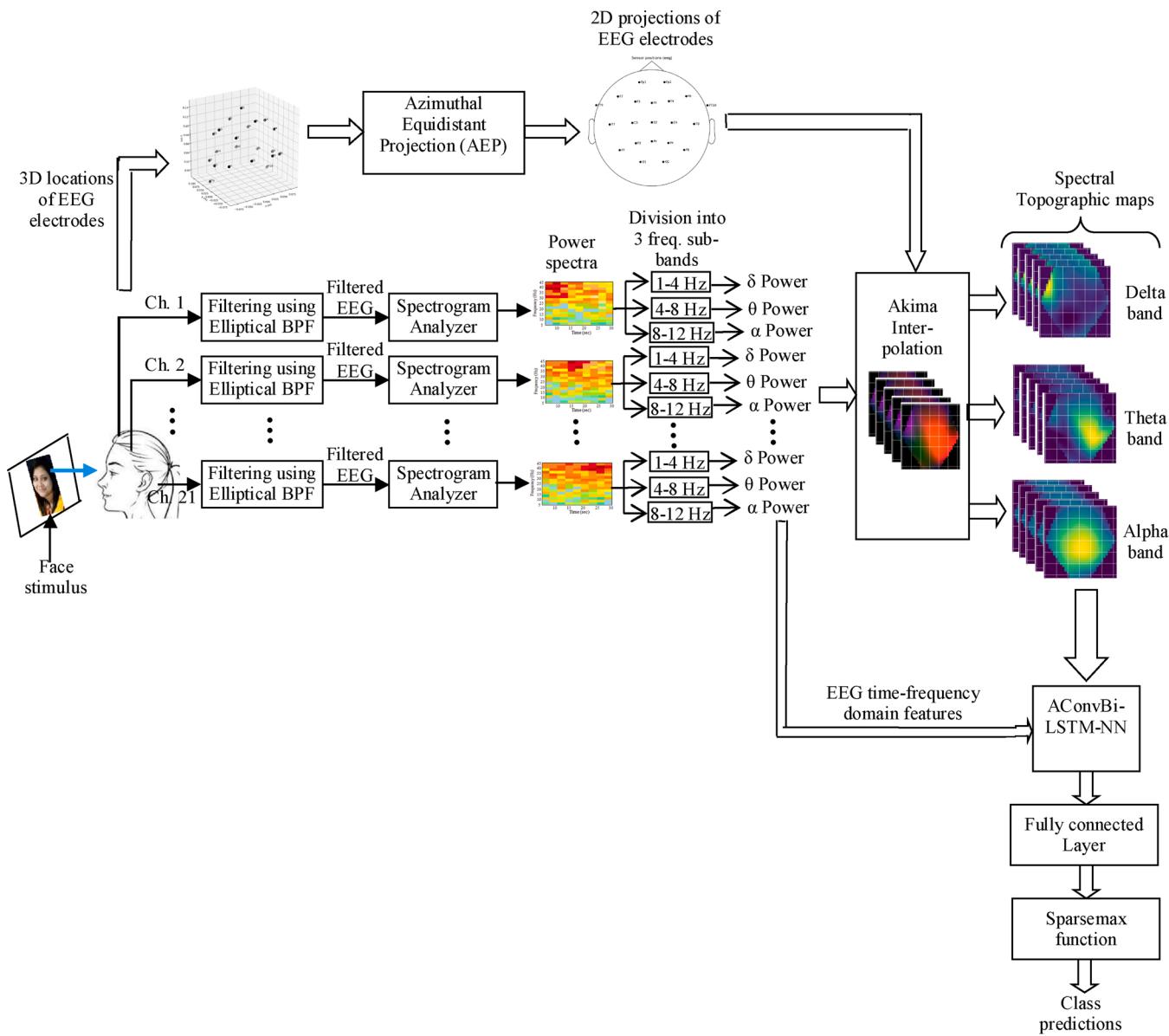
Human beings usually recognize familiar persons by analyzing the facial features of the old and new faces, they have already stored in their memory. In search of the method how brain uses the information stored into the memory to analyze and recognize a familiar face, it is still a big mystery in the history of neuro-cognitive science [1,2]. Human face perception and recognition process always remains a significant research arena as it is an important skill for successful social interaction. Determining the individual ability to categorize familiar faces (FFs) and unfamiliar faces (UFs) has high social impacts covering the analysis of psychological impairment [3], forensic inspections [4], identity authentication [5] and criminal investigations [6]. There exists a fairly large number of interesting literature [7–13] offering different ways of explaining how human brain processes FF and UF stimuli during face recognition. Among them, only a handful of literature [7,8] report on

the activation of different brain regions in responses to the FF and UF stimuli. There also exist works [10,11] confirming the appearance of face-sensitive event related potential (ERP) components, that are evoked due to the stimulations with FFs and UFs.

Extensive literature review on human memory involvement in face recognition task portrays that there is a small region in the Long term memory (LTM) [12,13], called fusiform face area (FFA) [20], located in the inferior temporal cortex of fusiform gyrus (Brodmann area 37), which is specified for storing facial memories. Whenever a subject tries to recognize a given face, a face matching process is performed into the working memory (WM) [12]. Face matching in WM requires the retrieval of stored facial information from FFA of LTM to the WM. Thereby, FFA plays a vital role in categorizing FFs and UFs. Prosopagnosia (face blindness) is a kind of cognitive disorder wherein a person loses the ability to recognize known faces, including one's own self (self-recognition) due to the impairment of FFA [20]. The present

\* Corresponding author.

E-mail address: [lidiaghosh.bits@gmail.com](mailto:lidiaghosh.bits@gmail.com) (L. Ghosh).

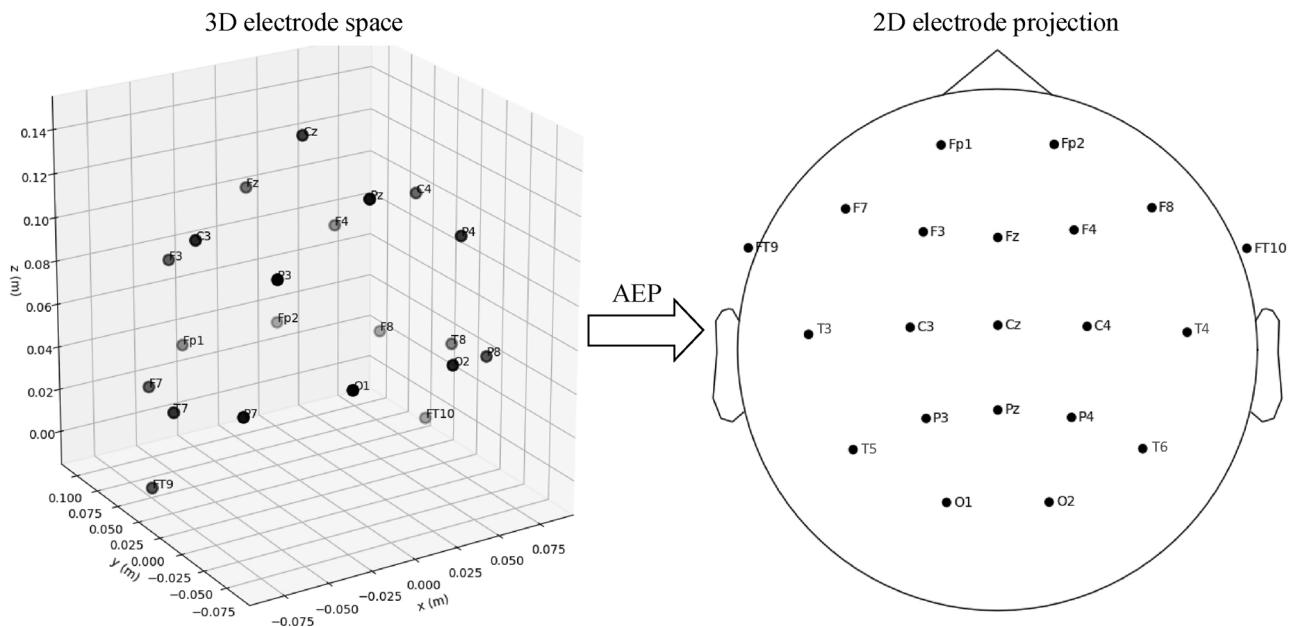


**Fig. 1.** Schematic block-diagram of the overall framework.

paper aims at assessing the face-perceptual ability of individuals by analyzing the electroencephalographic (EEG) signals captured during face recognition task.

The proposed framework is formulated in the settings of traditional machine learning problem. Here, the most challenging task is to extract relevant signal features and to design a suitable classifier for classifying the EEG activities of human subjects in response to the FF and UF stimuli. There is a plethora of classification techniques in the literature wherein the classifier take in EEG signals as input [14–19,50–53]. In Tuncer et al. [50], the authors designed an ensemble local graph structure induced feature extraction (FE) technique for EEG data classification in epilepsy detection. Tuncer et al. introduced a novel FE algorithm for EEG signal classification utilizing the key concept of multi-threshold based ternary pattern and dynamic center based binary pattern [51]. On the other hand, Aydemir et al., in [52], proposed an intelligent system for EEG classification which consists of a pre-processing module based on Tunable Q wavelet transform, FE module based on quadruple symmetric pattern, a neighborhood component analysis based nonparametric feature selector module and a kNN classifier. They have employed a multilevel learning paradigm for

the proposed method and tested using seven cases for a five-class classification problem. Their proposed framework yields a very high accuracy over 98% with a low computational complexity. The other interesting work offered by Tuncer et al. [53] illustrates a local senary pattern based texture descriptor for extracting relevant EEG features which is then followed by feature combining, feature selection and classification phases for designing an high accuracy epilepsy diagnostic system. Due to the availability of huge amounts of data and advancements in computational intelligence, now-a-days deep neural network (DNN) is gaining immense popularity in various research fields. Along with the developments in Brain-Computer Interfacing (BCI) technology, DNN models have been applied in diverse fields of BCI which outperform the state-of-the-art machine learning algorithms [15]. Most of the researchers and scientists are taking a keen interest to apply DNN models like Convolutional Neural Network (CNN) [16], Generative Adversarial Network [56], Recurrent Neural Network [17], and Long Short-term Memory (LSTM) [18] in image, audio and/or video signal processing as well as time-series prediction.. The present approach attempts to modify the existing parallel architecture of deep recurrent CNN, proposed by [28] in several aspects for the present classification



**Fig. 2.** 3D to 2D projection of EEG electrodes using AEP.

task.

There are plenty of papers in the literature where DNN algorithms are employed to analyze EEG time-series data [17,19,28]). Alizadeh and Fazel [57] developed a CNN model (HOG-CNN) which is trained by the combination of raw pixel data and histogram of oriented gradient features. As the method was suffering from overfitting, thus dropout, batch normalization and L2 regularization were used. Liu, Zhang and Pan [58] proposed an ensemble CNN approach (EN-CNN), where each subnet is a compact version of CNN trained separately. There are also evidences of recent successes in deep learning algorithms in the domain of EEG response analysis of human face categorization [57–62]. However, there is hardly any work where a DNN classifier is designed to take both the spatial and temporal EEG features as its input with an aim to categorize FF and UF responses from EEG signals. With respect to [28], three major modifications are undertaken in this paper: first, during the conversion of the EEG time-sample values into 2-dimensional images by applying appropriate projection and interpolation methods, Akima interpolation [29] is used instead of bicubic interpolation. The use of Akima interpolation is to help estimating the intermediate power values between pair of electrodes resulting in a smoother topographical map, and eventually producing the best spatial feature-vector after feeding it to the CNN. Secondly, instead of using the traditional Rectified Linear Unit (ReLU) activation function (AF), we examine the usability of two new AFs: Exponential Linear Sigmoid SquashHing (ELiSH) [34] and HardELiSH [34], as a strategy to boost the convergence of the proposed model. The utilization of Sparsemax function [35] instead of the primitive Softmax function in the dense layer of the classifier module to compute the class probabilities is the next contribution here. The particular choice is made as the sparsemax yields sparse probabilities unlike the softmax making the model output more interpretable. When compared with the existing machine learning and deep learning classifier algorithms, the proposed approach shows the supremacy for the present face recognition problem.

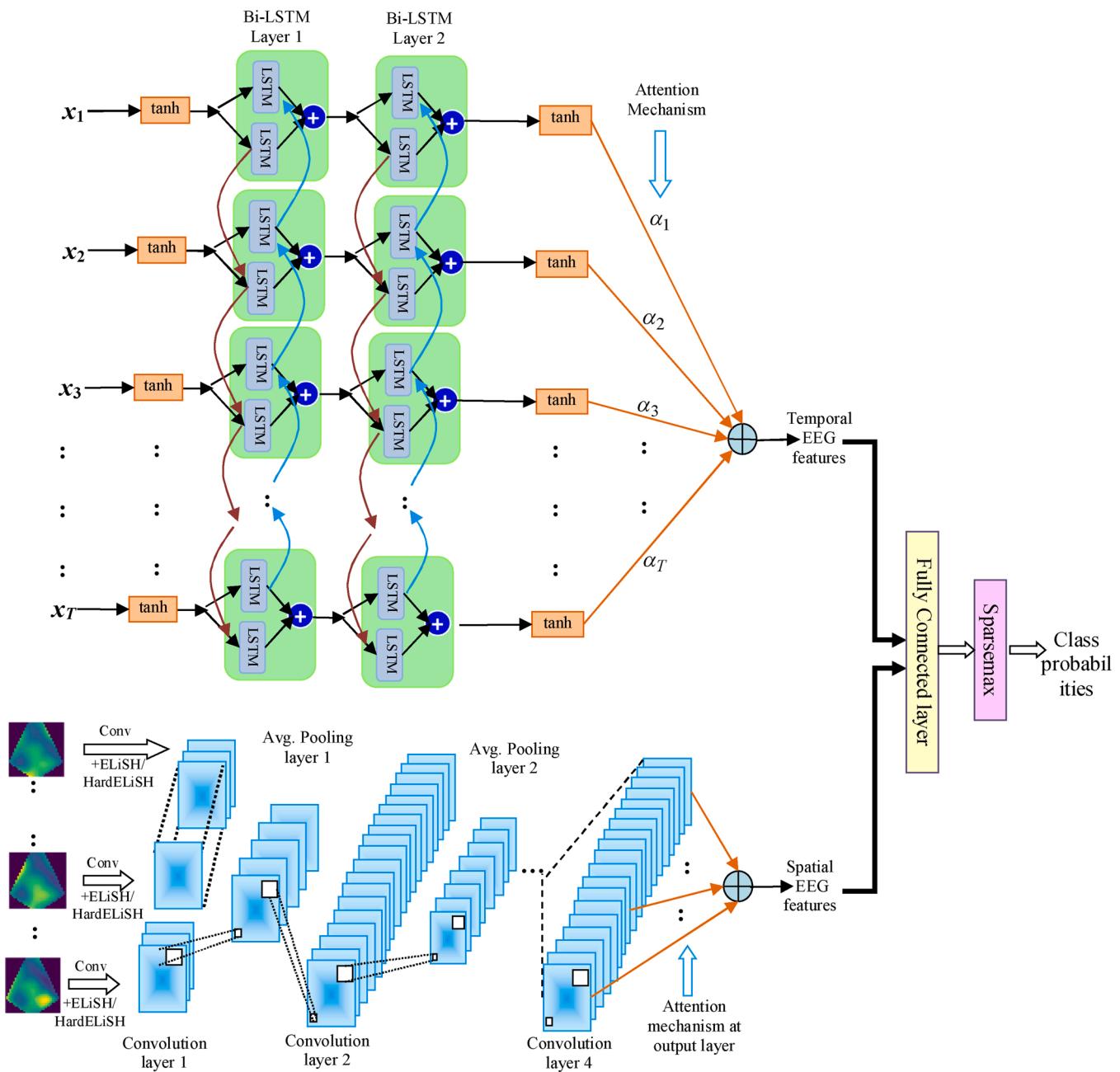
The organization of the remainder of the paper is as follows. In section II, the principles & methodologies adopted to design proposed algorithm of this paper are discussed from the viewpoint of mathematics. Section III gives an outline of the technique adopted for assessing the face recognition ability of individuals. Section IV gives an overview of the experimental framework, EEG data acquisition and the subjective details. The biological implications and the other experimental results are presented in Section V and VI. The advantages and disadvantages of

the proposed system are summarized in the discussion section VII and the future directions and concluding remarks are listed in Section VIII and IX respectively.

## 2. Proposed network architecture

The primary objective of the present approach is to design a DNN, which can efficiently discriminate the EEG responses for FF and UF categorization by the human subject with the highest possible accuracy. We have attempted to design an attention-based parallel architecture of convolutional Bi-LSTM Neural Network (A-ConvBi-LSTM-NN) that utilizes the spatial as well as temporal features of the acquired EEG data to categorize the desired class labels. A-ConvBi-LSTM-NN consists of a CNN [21] to extract spatial information and a Bi-LSTM [22] to deal with the temporal features. Fig. 1 delivers a schematic block-diagram of the overall framework.

To perform the experiments, the EEG signals are first acquired from the subjects' scalp using a 21-channel EEG device when the subjects are involved in recognizing FFs and UFs. Once the raw EEG signals are obtained, the data of each of the 21 channels: Ch. 1, Ch.2, ..., Ch. 21 (as mentioned in Fig. 1) are pre-processed and filtered by employing an Elliptical Band pass filter (EBPF) in order to eliminate the effects of various physiological and environmental noises. The artifact-free EEG signals of individual channel are then fed to a spectrogram analyzer for extracting the time-frequency domain features. We adopt Short-time Fourier Transform (STFT) [25] for this purpose as STFT is a powerful tool to obtain the power spectrum of the filtered EEG. There exist literatures [24], validating the existence of delta (1–4 Hz), theta (4–8 Hz) and alpha (8–12 Hz) activation for face recognition. Therefore, the entire frequency band is divided into the three sub-bands and the corresponding mean spectral power in delta ( $\delta$ ), theta ( $\theta$ ) and alpha ( $\alpha$ ) bands (see Fig. 1) has been calculated to obtain the respective frequency spectrum. To calculate the STFT of the signals, we choose window length as 100 and get an output dimension ( $51 \times 10$ ) for a 1 s window size. The EEG spectral power values are then fed to the two different modules: i) image data processing module and ii) time-series prediction module to extract spatial and temporal EEG features respectively. For extracting spatial features, the filtered EEG data obtained from individual channels are first converted to 2-dimensional multi-spectral images by applying an appropriate method described below and then are fed to an attention-based CNN (ABCNN). On the other hand, the time-frequency



**Fig. 3.** The proposed AConvBi-LSTM-NN Architecture.

domain STFT spectral values obtained for delta, theta and alpha frequency bands are parallelly used to train an attention-based Bi-LSTM network to obtain the temporal features of the acquired EEG. The spatial and temporal features thus obtained are then fed to a fully connected layer followed by a sparsemax function to obtain the final class probability. Therefore, here we have two most challenging tasks. First is to generate 2D multispectral images from the sampled EEG data and the second is to design the DNN such that it can utilize both the temporal and spatial information of the EEG signals to classify the desired classes with high efficacy. The detail description of the principles and methodologies adopted for the present problem are described below.

#### 2.1. Generation of 2-dimensional multi-spectral topographic images from EEG time-series data

As mentioned above, EEG signal is captured using 21 Ag/AgCl

electrodes/channels, placed around the scalp of the subject, to acquire the electrical response from different brain regions. Having superior temporal resolution [23], EEG activities can be predominantly analyzed in time-frequency domain feature space. One approach to extract the spatial information from the EEG data is to integrate the spatial dimension of the EEG signals with its time-frequency feature space [28, 32] by the following steps:

1. **Projection of EEG electrodes from 3D space to 2D space:** Since the electrodes of EEG device are placed over the 3D space of human scalp, therefore a suitable projection method is required to convert the EEG electrode representation from 3D space to its corresponding 2D space [28] by preserving their in-between relative distances. To serve the purpose, we employ a well-known planar (azimuthal) projection technique, referred to as *Azimuthal Equidistance Projection* (AEP) [26], which can preserve both the distance and direction

- between the electrode pairs from the central point in the 2D space in a proportional way as it was in the 3D space. The 3D EEG electrode space and its 2D representation after AEP are depicted in Fig. 2.
- 2. Interpolation of the EEG power with the intermediate electrode values:** To finally represent the temporal information of the EEG data in the form of successive images, we then employ *Akima Interpolation* [29] technique by interpolating each band power, obtained by performing STFT, with the 2D EEG electrode space (Fig. 1). Akima interpolation produces a smooth topographic image (also called topographic map) by estimating power values among each electrode-pair. The procedure is repeated for the three frequency bands: delta, theta and alpha individually to obtain 3 different topographic maps. These topographic images are then merged together and used as input to the ABCNN module.

## 2.2. Classification using AConvBi-LSTM-NN

The proposed AConvBi-LSTM-NN architecture is shown in Fig. 3. As shown in the figure, the proposed system consists of 2 major modules: i) EEG topographic image data processing by attention based CNN (ABCNN) [30] and ii) EEG time-series prediction by attention based Bi-LSTM. To classify the FF and UF responses from EEG signals of the human subject, the proposed system takes both the temporal (filtered EEG time-series) and spatial information (spectral images of topographic maps) parallelly as input. The topographic multispectral images are fed to the ABCNN module and the corresponding filtered EEG samples are fed to the Bi-LSTM module simultaneously. Subsequently, the spatial and temporal feature space obtained from the ABCNN and Bi-LSTM modules are fused together and fed to a fully connected layer followed by a sparsemax function with an ultimate aim of obtaining the class probabilities. Both the modules are discussed briefly in the following sub-sections.

**B.1. Attention-based CNN (ABCNN) for spatial feature extraction of EEG data:** CNN is an efficient tool of extracting spatial features from its input when the input data contains both the information about local connectivity as well as spatial arrangement between EEG electrodes [33]. In the present paper, we employ ABCNN [30] to extract the spatial features from the multispectral images of topographic maps. The classical CNN here contains 4 convolution and 3 average (avg.) pooling layers. The convolution operation in convolution layer on its input  $s_m$  can be mathematically expressed as

$$y_n = f((s_m * w_{mn}) + b_n) \quad (1)$$

Here,  $y_n \in \mathbb{R}^{Q \times Q \times P}$  denotes the  $n^{\text{th}}$  output feature map of dimension  $Q \times Q$ ,  $f(\cdot)$  is an activation function (AF),  $w_{mn}$  represents the weights of the convolution filter and  $b_n$  is the bias term.

Choice of an appropriate AF is an important factor in CNN as it directly affects the performance rate. Generally, sigmoid function is the common choice for the AF in CNN. However, here we examine the effect of introducing two new AFs, namely Exponential Linear Sigmoid Squashing (ELiSH) and Hard Exponential Linear Sigmoid Squashing (HardELiSH) [34], defined by Eqs. (2) and (3) respectively, to define the non-linear operation in the convolution layer.

$$f(a) = \begin{cases} \left( \frac{a}{1 + e^{-a}} \right), & a \geq 0 \\ \left( \frac{e^a - 1}{1 + e^{-a}} \right), & a < 0 \end{cases} \quad (2)$$

$$f(a) = \begin{cases} a \times \max(0, \min(1, (\frac{a+1}{2}))), & a \geq 0 \\ (e^a - 1) \times \max(0, \min(1, (\frac{a+1}{2}))), & a < 0 \end{cases} \quad (3)$$

The ELiSH function has the similar property of Swish function [37], which combines the Exponential Linear Unit (ELU) [35] and sigmoid

function [36] as defined by (2). The sigmoid function helps in better information propagation [34] while the ELU function benefits in removing the vanishing gradient problem. The hard version of ELiSH, HardELiSH involves the combination of HardSigmoid [34] and ELU in the negative portion while the positive portion comprises of the combination of HardSigmoid and linear as depicted by Eq. (3). The above-mentioned factors inspired us utilizing these two AFs in the present context. Furthermore, the attention mechanism is induced in the output of each convolution layer [30], as depicted in Fig. 3, to reweight the output of the last convolution layer by inducing avg. pooling through direct attention weighting.

The topographic maps of 3 freq. bands obtained for all the experimental trials of two different classes are used as the input images of the ABCNN. The first convolution layer of the ABCNN contains 32 filters of dimension  $3 \times 3$  with a stride of 1 pixel. To make consistency in the input image dimension after each convolution operation, zero-padding is used. The length of the convolution filter increases in the succeeding layers but the filter size remains same. There is an attention-based avg. pooling layer after each convolution layer which has window size of  $2 \times 2$  with a stride of 1 pixel. The last convolution layer is followed by a fully connected layer and this layer finally produces the spatial EEG feature-vector.

**B.2 Attention-based Bi-LSTM for temporal feature extraction of EEG data:** Designing a time-series prediction network that can accurately model the temporal dependencies of EEG signals is another important concern in BCI field because of the non-stationary characteristics of EEG signals. Here, we employ a Bi-LSTM network for extracting the temporal EEG features. After pre-processing and artifact removal, the filtered EEG samples are used here as the direct input to the Bi-LSTM module.

A Bi-LSTM network, consisting of two LSTM units [38], enables processing the input time-series in forward and backward directions simultaneously, as shown in Fig. 3. Each of the two LSTM units has four major elements: input gate (IG), output gate (OG), forget gate (FG) and memory cell [38]. Let,  $X = [x_1, x_2, \dots, x_{t-1}, x_t, x_{t+1}, \dots, x_T]$  and  $Y = [y_1, y_2, \dots, y_{t-1}, y_t, y_{t+1}, \dots, y_T]$  be the input and output time-series of the Bi-LSTM network at time-steps  $t = [1, T]$ . Each LSTM unit does the following operations on the input  $S$ .

At time-step  $t$ , the memory cell of an LSTM takes  $x_t$  (input time-series at time  $t$ ) and  $h_{t-1}$  (output of preceding hidden layer, i.e., at time  $t-1$ ) as its inputs and selects which information needs to be kept and which information needs to be forgotten depending on the FG information  $g_t$ , defined as

$$g_t = \text{sig}(U_{gx} \cdot x_t + U_{gh} \cdot h_{t-1} + b_g). \quad (4)$$

Similarly, the IG information  $i_t$ , current state information  $s_t$ , OG information  $o_t$  and hidden state information  $h_t$  can be represented as

$$i_t = \text{sig}(U_{ix} \cdot x_t + U_{ih} \cdot h_{t-1} + b_i) \quad (5)$$

$$\hat{s}_t = \tanh(U_{sx} \cdot x_t + U_{sh} \cdot h_{t-1} + b_s) \quad (6)$$

$$s_t = i_t \cdot \hat{s}_t + g_t \cdot s_{t-1} \quad (7)$$

$$o_t = \text{sig}(U_{ox} \cdot x_t + U_{oh} \cdot h_{t-1} + b_o), \quad (8)$$

$$h_t = o_t \cdot \tanh(s_t). \quad (9)$$

In the above Eqs. (4)–(9),  $U_{ab}$  represents the weight matrix from layer  $a$  to layer  $b$  and  $b_g$ ,  $b_i$ ,  $b_s$ ,  $b_o$  are the bias terms. The optimal values of the weight matrices and bias terms are obtained while the Bi-LSTM is trained using back-propagation through time algorithm. The AFs here used are sigmoid ( $\text{sig}$ ) and tanh.

At time-step  $t$ , the forward sequences  $\vec{h}_t$  and backward sequences  $\bar{h}_t$  of the Bi-LSTM can be computed as,

$$\overrightarrow{h}_t = \tanh(G_{\overrightarrow{h}} \cdot x_t + U_{\overrightarrow{h}} \cdot \overrightarrow{h}_{t-1} + b_{\overrightarrow{h}}) \quad (10)$$

$$\overleftarrow{h}_t = \tanh(G_{\overleftarrow{h}} \cdot x_t + U_{\overleftarrow{h}} \cdot \overleftarrow{h}_{t-1} + b_{\overleftarrow{h}}) \quad (11)$$

In (10) and (11),  $G_{\overrightarrow{h}}$  and  $G_{\overleftarrow{h}}$  represent the weight matrices, connecting the input to the hidden layer (HL), respectively in the forward and the backward direction. Let,  $h_t$  be the composite effect of  $\overrightarrow{h}_t$  and  $\overleftarrow{h}_t$ , evaluated as

$$h_t = \tanh(W_h \cdot \overrightarrow{h}_t + W_{\overleftarrow{h}} \cdot \overleftarrow{h}_t + b_y) \quad (12)$$

Here, we use an attention-based Bi-LSTM in which the attention weights,  $\gamma_t$  and the output of the attention,  $y$ , are computed as

$$\gamma_t = \frac{\exp(h_t)}{\sum_{t=1}^T \exp(h_t)} \quad (13)$$

$$y = \sum_{t=1}^T \gamma_t \cdot h_t \quad (14)$$

The attention output  $y$  is finally treated as the temporal EEG feature-vector, extracted by the Bi-LSTM network.

**B.3 Class Prediction using the integrated spatiotemporal features:** Finally the spatial and temporal EEG feature-vectors obtained from the parallel architecture of CNN and Bi-LSTMs are concatenated to produce the Spatio-temporal feature-vector. This Spatio-temporal feature-vector is then sent to a fully connected layer followed by a Sparsemax function [39], instead of the standard softmax function to obtain the respective class probabilities. Sparsemax function provides a sparse probability distribution from the output layer, where very small probabilities are reduced to zero values. Thus, it provides an optimal discrimination among class outputs in adjacent layers.

### 3. Face recogniton ability measurement

Face recognition-ability is defined by the ability of a person to discriminate distinct faces, in particular familiar and unfamiliar faces. Prosopagnosic patient has the kind of cognitive disorder, where the patient suffers from impairment to recognize familiar faces, sometimes the ability of self-recognition can also be lost. The motivation here is to assess the face perceptual-ability of individuals from their EEG responses towards face stimuli.

Let,  $\bar{z}^{1,k}, \bar{z}^{2,k}, \dots, \bar{z}^{m,k}$  be the average spatio-temporal feature of training data for a given electrode for class  $k$ , where the averaging is performed over  $n \times l$  instances of the  $j$  th feature and mathematically can be expressed as

$$\bar{z}^{j,k} = \frac{1}{n \times l} \sum_{r=1}^{n \times l} z^{r,j,k} \quad (15)$$

where,  $z^{r,j,k}$  denotes the  $j$  th feature of  $r$  th instance for class  $k$ . Here,  $n \times l$  instances are obtained by considering  $l$  learning epochs (session) each comprised of  $n$  trials.

Now let,  $z_{j,s}^*$  be the  $j$  th feature of a test data of subject  $s$ . We compute an Euclidean distance metric ( $d_{k,s}$ ) between  $\bar{z}^{j,k}$  and  $z_{j,s}^*$  for all the features  $j = 1 \dots m$ . The distance metric is defined as

$$d_{k,s} = \sqrt{\sum_{j=1}^m (\bar{z}_{j,k} - z_{j,s}^*)^2} \text{ for } \forall k \in [1, K] \quad (16)$$

Now if  $d_{k,s} \leq d_{k',s}$  for  $\forall k' \neq k$ , we declare that the test data falls into the class  $k$ . Then, we define the ability of face recognition of subject  $s$  as

$$p_s = \left( 1 - \frac{d_{k,s}}{\sum_{\substack{k'=1 \\ k' \neq k}}^K d_{k',s}} \right) \quad (17)$$

Then we compute  $p_s$  values for all the  $s \in [1, S]$  subjects and sort them in the descending order of the  $p_s$  values. The higher is the  $p_s$  value, the greater is the face recognition ability.

### 4. Experimental framework

The section describes the experimental protocol undertaken to perform the experiments.

#### 4.1. Acquisition of EEG signal

The data collection process has been carried out in the Artificial Intelligence (AI) Laboratory of Jadavpur University using a 21-channel EEG device manufactured by Nihon Kohden. For the current study EEG data is collected from pre-frontal, frontal, motor cortex, parietal, occipital and temporal regions by following the norms of international 10/20 electrode placement system.

#### 4.2. Subjective details

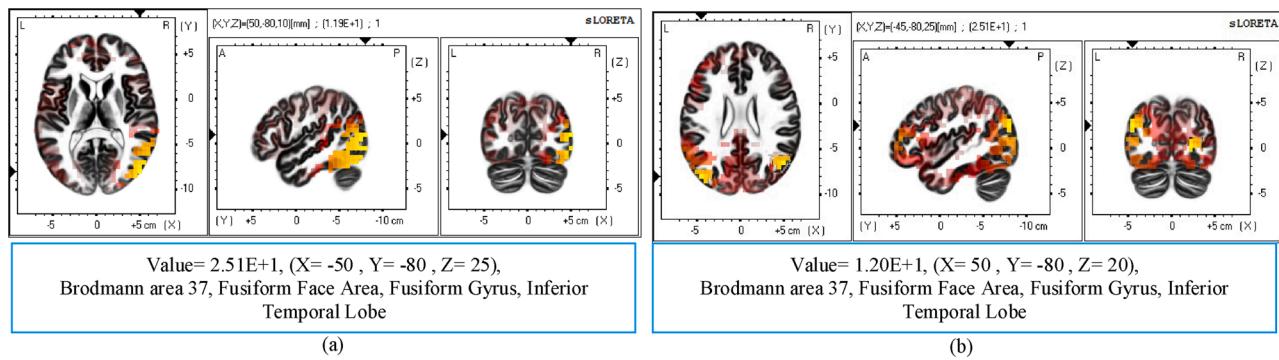
The participants include 35 healthy volunteers (20 male and 15 female in 20–26 years age-group), with no physical disabilities. Additionally, 3 subjects suffering from prosopagnosia also participate in the experiment. A written consent is taken from each subject before their participation.

#### 4.3. Stimulus preparation

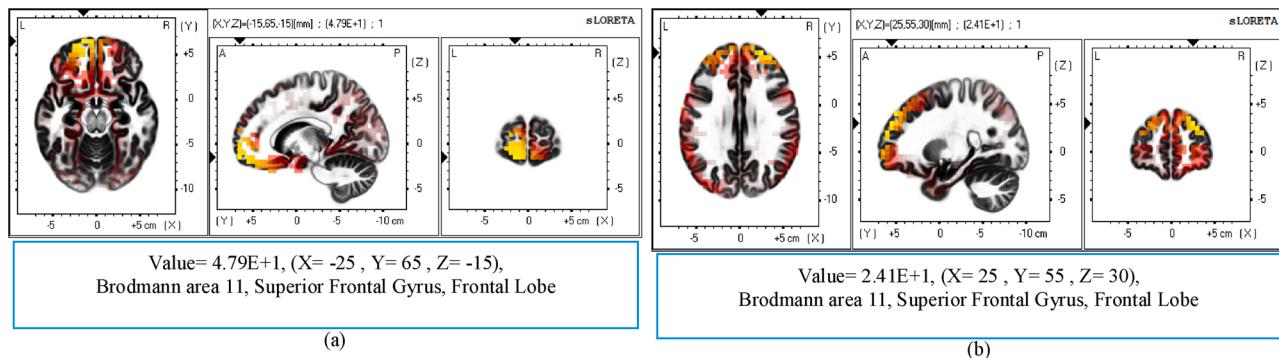
The experiment is carried out over 40 days for each subject. On each day, 5 experimental trials are conducted, where in each trial the subject is provided with the visual stimuli of a sequence of 2-dimensional images of 10 familiar (famous) faces and 10 unfamiliar faces, which are presented in a random manner for 10 s duration each and asked to recognize the faces, and the corresponding EEG signals are acquired from the subject's scalp. A relaxation interval of 10 s is given between two successive face stimuli to remove the effect of the previous stimulus in the brain signal from the successive one. We prepare a set of 15 famous persons' faces as FF stimuli and any 10 of them are presented to the subject during each trial. On the other hand, for the UF stimuli preparation, a subset of 10 unique images is prepared for each trial for each subject. Thus, we have altogether  $(40 \text{ days} \times 5 \text{ trials/day} \times 10 \text{ face stimuli/trial}) = 2000$  experimental instances per subject for each of the familiar and unfamiliar classes. As we have 35 healthy subjects and 3 prosopagnosics, so altogether the data dimension becomes  $2000 \times 35 = 70,000$  for healthy subject and  $2000 \times 3 = 6000$  for prosopagnosics. Now the total database is divided into 70 : 15 : 15 ratio for Training, Validation and Testing database.

#### 4.4. Artifact removal

An EBPF [41] of order 10 with pass-band frequencies 0.5–50 Hz has been used to get rid of the interference of physiological and environmental artifacts from the EEG signals. Eye-artifacts are removed by Independent Component Analysis (ICA) [42].



**Fig. 4.** sLORETA solutions for healthy subjects during (a) familiar and (b) unfamiliar face recognition.



**Fig. 5.** sLORETA solutions for Prosopagnosia patients during (a) familiar and (b) unfamiliar face recognition.

## 5. Biological implications

In order to understand the biological underpinnings of the discriminative behavior between the healthy subjects and the patients suffering from prosopagnosia while discriminating FFs and UFs, the following experiments are carried out.

### 5.1. Intra-cortical distribution of electrical response

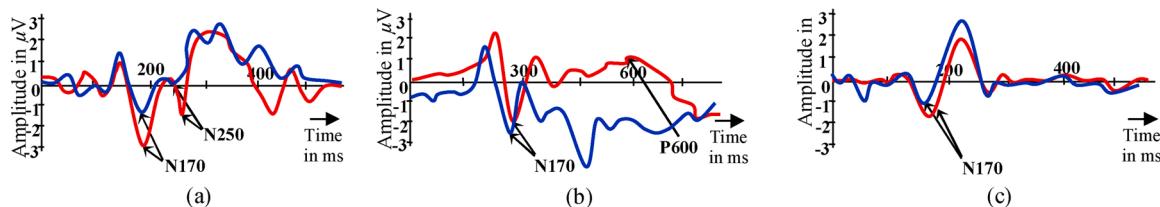
The experiment is performed by assessing the electrical current density of scalp EEG measurements obtained during recognition of FFs and UFs using sLORETA software [43]. Figs. 4 and 5 represent sLORETA solutions during FF and UF perceptions for healthy subjects and prosopagnosia patients respectively. For healthy subjects, it is found that the fusiform face area (FFA), located at inferior temporal region of fusiform gyrus (Brodmann area 37), has the highest activation during face recognition. Moreover, it is noticeable that the FFA of right hemisphere has a greater activation than that of left during FF recognition. On the other hand, the prosopagnosia patients do not show even a minimum activation in neither side of the FFA.

### 5.2. ERP signal analysis towards the discrimination FF and UF responses

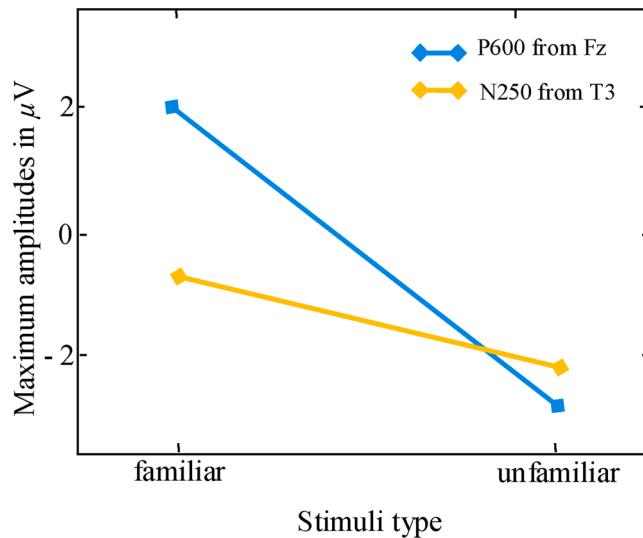
The ERP signals associated with face recognition are obtained by performing the following steps:

1. The raw EEG signals acquired from the subjects' scalp are cropped offline into epochs of 1800 ms duration including 500 ms pre-stimulus baseline.
2. These epochs are base-line corrected and pre-processed.
3. The filtered epochs are then averaged over each condition and each participant separately to obtain the grand averaged ERP values.

The ERPs obtained for FF (marked by red line) and UF (marked by blue line) responses for both the healthy subjects and prosopagnosia patients are plotted in Fig. 6. For healthy controls, it is observed that the two ERP signals, namely N250 and P600, show discriminating response for FF and UF stimulations. N250 is a negative amplitude ERP, appeared within 230–330 ms after the onset of the stimulus from the medial temporal brain regions ( $T_3$  and  $T_4$  electrodes) of FFA (Fig. 6(a)) and P600 is an ERP with positive amplitude which appears within 450–750 ms after the stimulus-onset (Fig. 6(b)), from medial frontal electrode ( $F_2$ ). It is evident from Fig. 6(a) and (b) that both the N250 and P600 have higher amplitudes for FF recognition than that in the UF



**Fig. 6.** Grand averaged ERPs (red line for FF response and blue line for UF response) of (a) medial temporal region and (b) mid-frontal region for healthy controls, and (c) medial temporal region of Prosopagnosia patients.



**Fig. 7.** Maximum P600 amplitudes at Fz electrode and minimum N250 amplitudes at T3 electrode as functions of FF and UF responses.

recognition for healthy controls. On the other hand, neither of the two ERPs are found for prosopagnosia patients.

Additionally, it is observed that for healthy controls, there is a large negative deflection, namely N170, within 160–180 ms after the onset of the stimulus for both the FF and UF responses [44]. N170 is also known as “face-selective ERP” as it is elicited by face stimuli only rather than non-face stimuli. It is interesting to note that Prosopagnosic patients also produce N170 signal but with relatively smaller amplitude than that of the healthy control, which in turn reveals that although their impairment is confined to the face discrimination, yet their ability to distinguish face and non-face stimuli is intact.

A statistical test using a  $2 \times 2$  mixed model ANOVA [40] has been undertaken to quantitatively validate the existence of P600 and N250 amplitude differences to discriminate the FF and UF responses. The result of the analysis performed on the P600 maximum amplitudes obtained from Fz electrode and N250 minimum amplitudes obtained from

T3 electrode are depicted in Fig. 7. The figure reveals that P600 shows the most significant differences in FF versus UF responses with  $F(1, 28) = 771.915$ ,  $p$  value  $< 0.0005$ . Thus it confirms that the differences in P600 amplitudes can be used as relevant features to discriminate FF and UF responses.

## 6. Results

The detailed experimental results of the proposed approach are discussed in this section. The source code of the algorithm is available in [55].

### 6.1. Selecting relevant EEG frequency-bands for categorizing FF and UF responses

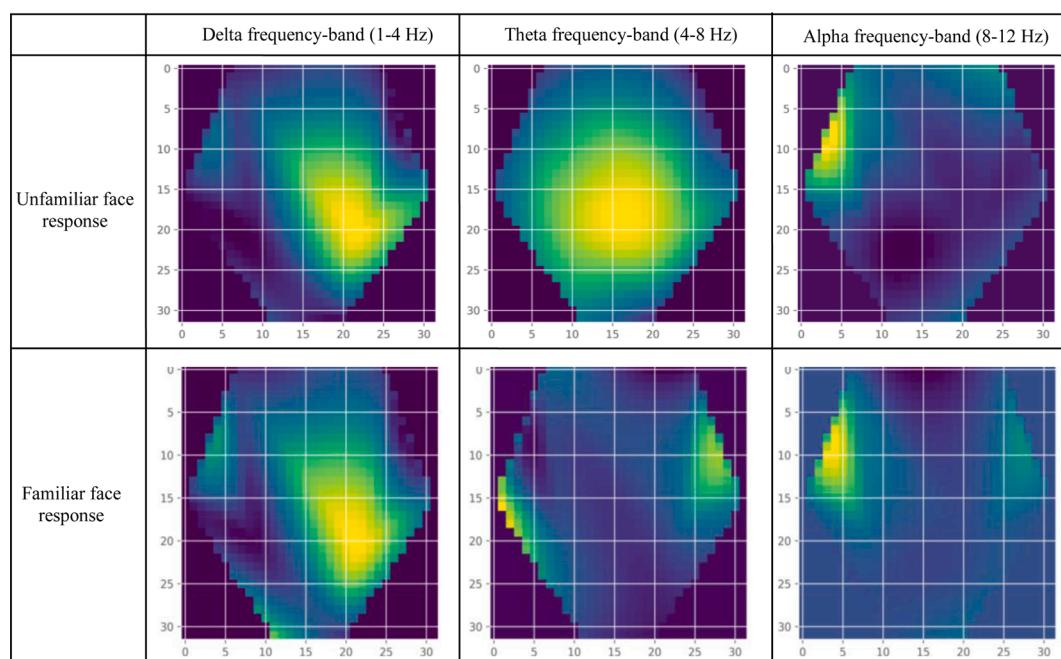
To identify the relevant EEG frequency-bands that can offer significant discrimination between FF and UF responses, we compare the EEG activities obtained in terms of topographic maps for the 3 different frequency-bands: delta (1–4 Hz), theta (4–8 Hz) and alpha (8–12 Hz). The topographic maps for FF and UF responses are presented in Table 1. In color bar, yellow color specifies the maximum activation; blue defines the minimum activation and green indicates moderate activation. The table portrays that, the theta band demonstrates the maximum activation as well as the best discrimination for the FF and UF responses.

**Table 2**  
Relative performance evaluation of AConvBi-LSTM-NN.

Classifiers	Percentage Classifier Accuracy	True Positive Rate	False Positive Rate
AConvBi-LSTM-NN	91.339	0.98	0.02
Bi-LSTM	81.94	0.91	0.09
LSTM	78.36	0.80	0.77
RNN	75.56	0.82	0.74
CNN	73.02	0.81	0.76
kSVM	66.59	0.76	0.71
LSVM	62.22	0.78	0.71

**Table 1**

Activations in the three freq. bands: Delta, Theta and Alpha Bands for FF and UF responses: A Comparative Study.



**Table 3**

Percentage classification accuracy with varying activation functions.

AF	CA (%)
Exponential Linear Sigmoid SquashHing (HardELISH)	91.339
Exponential Linear Sigmoid SquashHing (ELISH)	90.916
Rectified Linear Unit (RELU)	88.95
Leaky RELU	89.93
Scaled Exponential Linear Unit (SELU)	88.16
Exponential Linear Unit (ELU)	89.21
Hyperbolic Tangent (Tanh)	87.27
Sigmoid	84.53

**Table 4**

Parameters of AConvBi-LSTM-NN and their values.

Modules of AConvBi-LSTM-NN	Module Parameters	Parameter Values
CNN	Total no. of convolution layers	4
	No. of avg. pooling layers	3
	Pooling size	2 × 2
	no. of filters in each convolution layer	32
	Size of each filter	3 × 3
Bi-LSTM	Total no. of LSTM layers	400
	Size of HLS	120

## 6.2. Relative performance analysis of the AConvBi-LSTM-NN

Performance evaluation of the proposed DNN algorithm against the other state-of-the-art classifier algorithms has been done with respect to the 3 metrics: percentage Classifier Accuracy (%CA), True Positive Rate (TPR) and False Positive Rate (FPR) [45,46]. The parametric values obtained by each of the 7 classifier algorithms are compared in Table 2, which shows that AConvBi-LSTM-NN yields best results.

## 6.3. Impact of activation functions on classifier accuracy

The proposed classifier accuracy is further analysed to examine the effect of different activation functions used in the first Convolution layer of the CNN module. Table 3 depicts the results of the analysis, which clearly infers that the percentage classifier accuracy for ELISH and HardELISH activation functions is the highest.

## 6.4. Hyper-parameters of the proposed model

In order to ensure the robustness of the proposed model, the optimally selected values of the hyper-parameters in accordance to achieve the highest classification accuracy are presented in Table 4. Since the

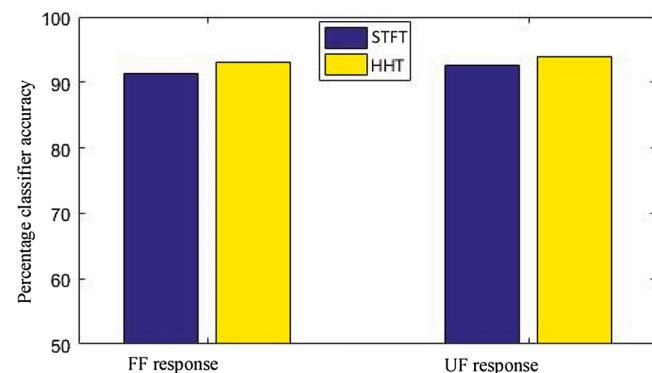


Fig. 9. Comparison between the performances of HHT and STFT in terms of percentage classifier accuracies.

model is an integrated version of two modules: ABCNN and attention based Bi-LSTM, therefore the table enlists the parameters of individual modules such as no. of convolution layers, no. of avg. pooling layers, total no. of filters, size of a filter and pooling layer size for a CNN, and total no. of LSTM layers and size of hidden layers (HLS) for a Bi-LSTM. Apart from these, the dropout rate and overall learning rate of the model are 0.2 and 0.0001 respectively. Furthermore, we use the callback method to select the best possible model weights for the training phase when validation accuracy reached maximum.

## 6.5. The effect of Hilbert-Huang transform as a spectrogram analyzer

In the proposed approach, we have used STFT for extracting the time-frequency domain features of the EEG signals. Although, high classifier accuracy (above 91%) is achieved, but it is difficult to avoid the major disadvantage of resolution trade-off between time and frequency, offered by the STFT. To deal with this problem, we have undertaken a comparative analysis between STFT and an empirical mode decomposition (EMD) based spectrogram analyser, named as Hilbert-Huang transform (HHT). In recent years, many of the researchers are using HHT to analyse non-stationary and non-linear signals such as EEG, ECG and the like [63–65]. The idea here is to iteratively extract the intrinsic mode functions [65] by EMD and then to employ the Hilbert-Huang Transform for calculating the power spectrum of the EEG signal. The results of HHT for FF and UF responses are shown in Fig. 8.

To compare the relative performance of the STFT and HHT, we use HHT as a spectrogram analyzer instead of STFT and examine the corresponding classifier results and plotted it in Fig. 9. The figure shows a marginal improve in classifier accuracy when HHT is used as a spectrogram analyser instead of STFT.

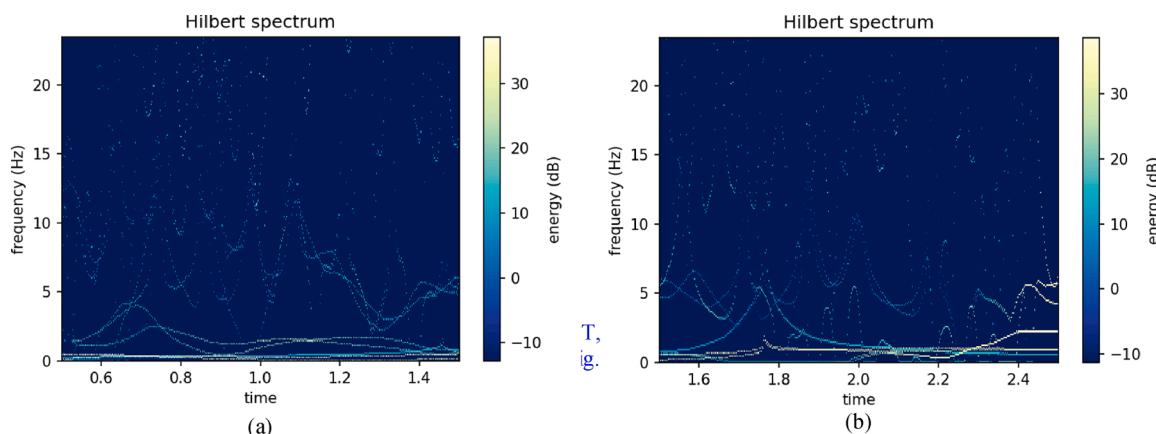


Fig. 8. Time-frequency plots using HHT for (a) FF response and (b) UF response.

**Table 5**

Comparisons with existing deep neural networks.

Deep Neural Networks	Number of layers	Input	CA (%)
AConvBi-LSTM-NN	4 CNN and 2 Bi-LSTM layers	Topographic Images + Time-frequency domain features (Spatio-temporal EEG features)	91.339
DMVST-NET [47]	4 CNN and 2 LSTM layers	Spatio-temporal EEG features	86.72
CNN-BiLSTM Cascade architecture [33]	4 CNN and 2 Bi-LSTM layers	Spatio-temporal EEG features	87.96
CNN-BiLSTM Parallel architecture [33]	4 CNN and 2 Bi-LSTM layers	Spatio-temporal EEG features	88.10
ABCNN [30]	2000 HLs	Spatio-temporal EEG features	80.75
DCRNN [48]	2000 HLs	Spatio-temporal EEG features	85.11
CNN	2000 HLs	Spatio-temporal EEG features	73.02
Bi-LSTM [38]	2000 HLs	Time-series (Temporal EEG features)	81.94
LSTM	2000 HLs	Temporal EEG features	78.36
RNN	2000 HLs	Temporal EEG features	75.76

**Table 6**

Validating the effect of EEG electrodes on AConv-Bi-LSTM-NN performance.

Brain Regions (Electrode specifications)	CA (%)	TPR	TFR
Prefrontal only (Fp1 and Fp2)	74.96	0.80	0.20
Frontal only (F3, F4, Fz, F7 and F8)	74.15	0.82	0.18
Temporal only (T3, T4, T5 and T6)	68.61	0.77	0.23
Parietal only (P3, P4 and Pz)	65.96	0.78	0.22
Motor cortex only (C3, C4 and Cz)	50.43	0.61	0.39
Occipital only (O1 and O2)	53.07	0.59	0.41
Fronto-temporal only (FT9 and FT10)	41.16	0.55	0.45
Pre-frontal + Frontal	69.60	0.77	0.23
Pre-frontal + Frontal + Parietal	64.47	0.72	0.30
Pre-frontal + Frontal + Occipital	59.02	0.65	0.35
Pre-frontal + Parietal + Temporal	83.84	0.96	0.04
Prefrontal + Frontal + Parietal	78.03	0.87	0.13
<b>Pre-frontal + Frontal + Temporal</b>	<b>91.339</b>	<b>0.98</b>	<b>0.02</b>
Pre-frontal + Parietal + Occipital	80.95	0.97	0.03
Pre-frontal + Parietal + Occipital + Temporal + Motor Cortex	83.51	0.96	0.04

### 6.6. Relative performance analysis with the other deep learning networks

The proposed classifier performance is also compared with the existing deep learning algorithms in [Table 5](#). It is quite evident from the table that the accuracy obtained by the AConvBi-LSTM-NN surpasses the others by an enhanced marginal level. Therefore, AConvBi-LSTM-NN is able to offer an accurate categorization of the two distinct class-labels.

### 6.7. Effect of different EEG channels on classifier accuracy

In this experiment, we examine the impact of different EEG channels on the classifier accuracy. To perform the experiment, we select the EEG electrodes of a specific brain region at a time, i.e., for example F<sub>3</sub>, F<sub>4</sub>, F<sub>z</sub>, F<sub>7</sub> and F<sub>8</sub> electrodes for frontal region, P<sub>3</sub>, P<sub>4</sub> and P<sub>z</sub> for parietal region and the like. The spatiotemporal features extracted from the selected EEG electrodes are then fed to the proposed classifier network and it is checked whether the classifier correctly distinguishes the desired class-label with best accuracy. The results are provided in [Table 6](#). The same experiment is also carried out for the different combinations of two or more brain regions as depicted in [Table 6](#). It reveals that the classifier accuracy enhances by approximately 9–11.5% when the EEG features extracted from prefrontal, frontal and temporal regions are jointly used instead of the other combinations.

**Table 7**

Validation using McNemar's statistical test.

Reference Algorithm (X): AConvBi-LSTM-NN		
Algorithm to be compared (Y)	z	Acceptance/Rejection of the hypothesis $H_0$
CNN-BiLSTM Parallel architecture [33]	3.721	<b>Accepted</b>
CNN-BiLSTM Cascade architecture [33]	3.96	Rejected
ABCNN [30]	5.94	Rejected
DCRNN [48]	4.61	Rejected
DMVST-NET [47]	8.37	Rejected
Bi-LSTM [38]	6.78	Rejected
LSTM	6.99	Rejected
CNN	7.15	Rejected
SVM	7.87	Rejected
QDA	9.16	Rejected
LDA	9.02	Rejected
kNN	10.67	Rejected
RNN	5.90	Rejected

**Table 8**

Face recognition ability analysis for prosopagnosia patients.

Subject ID	p <sub>s</sub>	Rank
S1	0.844	5
S2	0.749	10
S3	0.765	9
S4	0.872	4
S5	0.823	8
S6	0.910	2
<b>S7</b>	<b>0.921</b>	<b>1</b>
S8	0.876	3
S9	0.827	6
S10	0.815	7
<b>PS1</b>	<b>0.387</b>	<b>11</b>
<b>PS2</b>	<b>0.231</b>	<b>13</b>
<b>PS3</b>	<b>0.317</b>	<b>12</b>

### 6.8. Statistical validation using McNemar's test

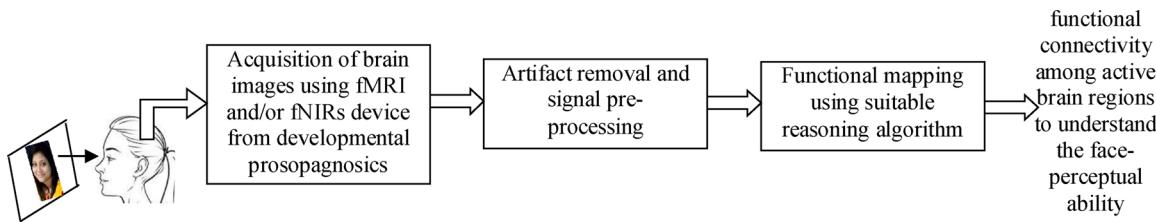
The performance of the proposed DNN has been statistically validated using the well-known McNemar's test [49] as a single database prepared by the AI lab. of Jadavpur University has been utilized for the present study. The z score evaluated for the test has been illustrated by

$$z = \frac{(|m_{01} - m_{10}| - 1)^2}{m_{01} + m_{10}} \quad (18)$$

where all the terms in the equation convey the same meaning as indicated in [49]. If we consider two DNN algorithms A and B, where A denotes the proposed AConvBi-LSTM-NN and B indicates any one of the other DNN algorithms to be compared as listed in [Table 7](#). The result confirms that the null hypothesis ( $H_0$ ) for the existing DNNs except Parallel CNN-BLSTM [33] have been rejected since the z-score for all of them is beyond  $\chi^2_{1,0.95} = 3.84$ , i.e., the Chi-square value at 95% confidence level and 1 degree of freedom.

### 6.9. Ranking of subjects according to Face-perceptual ability

To rank subjects based on face recognition-ability, we compute p<sub>s</sub> values using Eq. (17) for all the subjects including the 3 patients suffering from prosopagnosia. Then we sort the subjects based on the descending order of their face recognition-ability. We here present the results for only 10 healthy subjects (S) and 3 prosopagnosia patients (PS) [Table 8](#) due to space economy. It is evident from the table that the prosopagnosics have the least p<sub>s</sub> value and thus showing poor face recognition ability.



**Fig. 10.** Work-plan for determining the functional connectivity among active brain regions to understand the developmental prosopagnosia.

## 7. Discussion

The proposed work is dedicated to detect face-perceptual ability of individuals by classifying the EEG responses for FF and UF recognition. The advantages and limitations of the described work can be stated based on the below mentioned parameters.

- Robustness analysis:** The classification of EEG responses towards FF and UF recognition of individuals is done using an attention-based convolutional Bi-LSTM network. If any classifier suffers from overfitting and/or underfitting, then that reflects in the performance of the classification accuracy. But, in the proposed work, 91.79% classification accuracy is obtained which negates the occurrence of overfitting and underfitting and validates the robustness of the proposed method.
- Convenience:** The proposed architecture has minimal hardware setup (a 21-channel EEG sensor is required). Further, the work is independent of the gender and age of the subjects. So, this makes the system very much convenient to deal with the assessment of the inter-individual variances in face-recognition skills and thus can be used as a biomarker in prosopagnosic patient diagnosis.
- Benefits:** The main inspiration behind this work is to rank individuals with varying degree of face-perceptual ability. If the subjects with a below threshold face-perceptual ability can be identified, then they may have the possibility to be prosopagnosic and can be advised for an early check up to start the necessary treatments as soon as possible.
- Productivity:** The system proposed can able to correctly identify the face-perceptual ability of individuals based on their EEG responses
- Convergence:** The computation time of 5 s on Intel Core i9 10920x (12 core, 24 threads) processor @ 4.6 GHz and 64GB RAM running on GPU-NvidiaQuadro RTX 5000 16 GB platform is quite small respect to using in real-time too. Tensorflow and keras [31] software is used to implement the model.
- Flexibility:** The proposed algorithm is applied for a large number of participants who have different face-perceptual ability.
- Feasibility:** The work can be easily applicable to a larger mass of people, so feasibility is not an issue.
- Efficiency:** The efficiency of the proposed method has been compared with a wide range of already available algorithms and every time the work proves its worth.
- Diversity:** The proposed system has been implemented on a diverse group of subjects.
- Reliability:** The proposed approach is already realized on the real-world scenario.

## 8. Future directions

Determining the individual ability to recognize FFs and UFs has tremendous applications in eyewitness testimony while appointing efficient persons in law enforcements, the courts, police work and national and homeland security [54]. Prosopagnosics, those who have poor face-recognition ability, can broadly be distinguished into two classes: Developmental and Acquired. In developmental prosopagnosia (DP), most of the sufferers have their genetic history of prosopagnosia

unlike due to any structural brain lesions as in the case of acquired prosopagnosia (AP). AP is caused by the damage of the brain regions (may be due to accidents or any other brain impairment) that contribute in recognizing faces. Therefore, the future studies should endeavor to understand the biological underpinnings of developing prosopagnosia genetically in the terms of functional connectivity analysis of the active brain regions. Brain imaging devices like functional magnetic resonance imaging (fMRI) and functional near infra-red spectroscopy (fNIRs) can be the suitable means of acquiring the spatial information of the brain regions activated during face recognition. The acquired responses can then be analyzed to determine the functional mapping from the spectroscopic analysis to the perceptual-ability assessment using a suitable reasoning algorithm. The intended work-plan is summarized in Fig. 10.

## 9. Conclusion

The paper investigates the performance of a novel deep neural network architecture to distinguish EEG signals of human subjects for FF and UF responses. Most classification techniques in the literature deal with the extraction of temporal features of EEG signals. However, both the spatial and temporal features of brain EEG data are taken into account by our model, AConvBi-LSTM-NN, making it one of the more accurate and robust deep neural network for EEG analysis. The classification accuracy of 91.339% for our network makes it all the more promising when compared to the state-of-the-art.

The brain understanding of face-selectivity is also obtained by standardized low resolution electromagnetic topographic analysis (sLORETA) software. From sLORETA activations it is evident that the right hemisphere of the fusiform face area (FFA) shows higher activations for FF stimulation rather than unfamiliar for healthy controls. On the other hand, the prosopagnosic patients do not show any significant activation in FFA due to the impairment in occipito-temporal region. Additionally, the ERP analysis shows a remarkable discrimination for familiar and UF responses, which can be useful in EEG-based identity authentication trait. By analysing the face-perception ability, the proposed model can also be used as a bio-marker to detect the early diagnosis of prosopagnosia.

## CRediT authorship contribution statement

**Lidia Ghosh:** Methodology, Writing - original draft, Validation, Data curation, Software. **Dipayan Dewan:** Visualization, Software, Data curation, Investigation. **Abir Chowdhury:** Writing - original draft, Validation, Investigation. **Amit Konar:** Conceptualization.

## Acknowledgment

The fund delivered by MHRD for RUSA-II project of Jadavpur University, India is whole-heartedly acknowledged.

## Declaration of Competing Interest

The authors report no declarations of interest.

## References

- [1] B.R.P. Walton, P.J. Hills, Face distortion aftereffects in personally familiar, famous, and unfamiliar faces, *Front. Psychol.* 3 (2012) 258.
- [2] H. Ito, A. Sakurai, Familiar and unfamiliar face recognition in a crowd, *Psychology* 5 (2014) 1011.
- [3] M. Eimer, A. Gosling, B. Duchaine, Electrophysiological markers of covert face recognition in developmental prosopagnosia, *Brain* 135 (2012) 542–554.
- [4] A.K. Jain, B. Klare, U. Park, Face matching and retrieval in forensics applications, *IEEE Multimed.* 1 (2012) 20–28.
- [5] Y. Zeng, Q.W.K. Yang, L. Tong, B. Yan, J. Shu, D. Yao, EEG-based identity authentication framework using face rapid serial visual presentation with optimized channels, *Sensors* 19 (2019) 6–19.
- [6] D. Muramatsu, Y. Makihara, H. Iwama, T. Tanoue, Y. Yagi, Gait verification system for supporting criminal investigation, in: 2013 2nd IAPR Asian Conference on Pattern Recognition, IEEE, 2013, pp. 747–748.
- [7] B. Rossion, R. Caldara, M. Seghier, A.M. Schuller, F. Lazeyras, E. Mayer, A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing, *Brain* 126 (2003) 2381–2395.
- [8] E. Kaan, T.Y. Swaab, Repair, revision, and complexity in syntactic analysis: an electrophysiological differentiation, *J. Cogn. Neurosci.* 15 (2003) 98–110.
- [9] R.N. Haber, The impending demise of the icon: a critique of the concept of iconic storage in visual information processing, *Behav. Brain Sci.* 6 (1983) 1–11.
- [10] A. Gosling, M. Eimer, An event-related brain potential study of explicit face recognition, *Neuropsychologia* 49 (2011) 2736–2745.
- [11] J.W. Tanaka, T. Curran, A.L. Porterfield, D. Collins, Activation of preexisting and acquired face representations: the N250 event-related potential as an index of face familiarity, *J. Cogn. Neurosci.* 18 (2006) 1488–1497.
- [12] A.D. Baddeley, *The Psychology of Memory*, 1976, New York.
- [13] R.G.M. Morris, E.I. Moser, G. Riedel, S.J. Martin, J. Sandin, M. Day, C. O'Carroll, Elements of a neurobiological theory of the hippocampus: the role of activity-dependent synaptic plasticity in memory, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358 (2003) 773–786.
- [14] A. Craik, Y. He, J.L. Contreras-Vidal, Deep learning for electroencephalogram (EEG) classification tasks: a review, *J. Neural Eng.* 16 (2019), 031001.
- [15] X. Zhang, L. Yao, X. Wang, J. Monaghan, D. Mcalpine, A survey on deep learning based brain computer interface: recent advances and new frontiers, arXiv preprint arXiv 1905 (2019).
- [16] A. Supratak, H. Dong, C. Wu, Y. Guo, DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG, *IEEE Trans. Neural Syst. Rehabil. Eng.* 25 (2017) 1998–2008.
- [17] S. Kuanar, V. Athitsos, N. Pradhan, A. Mishra, K.R. Rao, Cognitive analysis of working memory load from EEG, by a deep recurrent neural network, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, pp. 2576–2580.
- [18] S. Alhagry, A.A. Fahmy, R.A. El-Khoribi, Emotion recognition based on EEG using LSTM recurrent neural network, *Emotion* 8 (2017) 355–358.
- [19] L. Ghosh, S. Ghosh, A. Konar, P. Rakshit, A.K. Nagar, Decoding of EEG signals using deep long short-term memory network in face recognition task, in: 2018 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2018, pp. 477–483.
- [20] J. Zhang, J. Liu, Y. Xu, Neural decoding reveals impaired face configural processing in the right fusiform face area of individuals with developmental prosopagnosia, *J. Neurosci.* 35 (2015) 1539–1548.
- [21] L.C. Yann, Y. Bengio, Convolutional networks for images, speech, and time series, in: *The Handbook of Brain Theory and Neural Networks*, vol. 3361, 1995.
- [22] X. Hu, Q. Yuan, Epileptic EEG identification based on deep Bi-LSTM network, in: 2019 IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT), IEEE, 2019, pp. 63–66.
- [23] A. Gevins, H. Leong, M.E. Smith, J. Le, R. Du, Mapping cognitive brain function with modern high-resolution electroencephalography, *Trends Neurosci.* 18 (1995) 429–436.
- [24] Y. Yang, Y. Qiu, A.C. Schouten, Dynamic functional brain connectivity for face perception, *Front. Hum. Neurosci.* 9 (2015) 662–673.
- [25] Daniel Griffin, Jae Lim, Signal estimation from modified short-time Fourier transform, *IEEE Trans. Acoust. Speech. Signal Process.* 32 (1984) 236–243.
- [26] Paul S. Anderson, An oblique, poly-cylindrical, orthographic azimuthal equidistant cartographic projection: its purpose, construction and theory, *Cartography* 8 (1974) 182–186.
- [27] Pouya Bashivan, Irina Rish, Mohammed Yeasin, Noel Codella, Learning representations from EEG with deep recurrent-convolutional neural networks, arXiv preprint arXiv:1511.06448 (2015).
- [28] Hiroshi Akima, A new method of interpolation and smooth curve fitting based on local procedures, *J. ACM* 17 (1970) 589–602.
- [29] Wenpeng Yin, Hinrich Schütze, Bing Xiang, Bowen Zhou, Abcnn: attention-based convolutional neural network for modeling sentence pairs, *Trans. Assoc. Comput. Ling.* 4 (2016) 259–272.
- [30] Antonio Gulli, Sujit Pal, *Deep Learning With Keras*, Packt Publishing Ltd, 2017.
- [31] Y.H. Kwon, S.B. Shin, S.D. Kim, Electroencephalography based fusion two-dimensional (2D)-convolution neural networks (CNN) model for emotion recognition system, *Sensors* 18 (2018) 1383.
- [32] X. Ma, J. Zhang, B. Du, C. Ding, L. Sun, Parallel architecture of convolutional bi-directional lstm neural networks for network-wide metro ridership prediction, *IEEE Trans. Intell. Transp. Syst.* 20 (2018) 2278–2288.
- [33] C. Nwankpa, W. Ijomah, A. Gachagan, S. Marshall, Activation functions: comparison of trends in practice and research for deep learning, arXiv preprint arXiv:1811.03378 (2018).
- [34] D.A. Clevert, T. Unterthiner, S. Hochreiter, Fast and accurate deep network learning by exponential linear units (elu), arXiv preprint arXiv:1511.07289 (2015).
- [35] X. Yin, J.A.N. Goudriaan, E.A. Lantinga, J.A.N. Vos, H.J. Spiertz, A flexible sigmoid function of determinate growth, *Ann. Bot.* 91 (2003) 361–371.
- [36] H.H. Chieng, N. Wahid, P. Ong, S. Raj Kishore Perla, Flatten-T Swish: a thresholded ReLU-Swish-like activation function for deep learning, arXiv preprint arXiv: 1812.06247 (2018).
- [37] S. Hochreiter, J. Schmidhuber, Long short-term memory, *J. Neural Comput.* 9 (November) (1997) 1735–1780. MIT Press.
- [38] A. Martins, R. Astudillo, From softmax to sparsemax: a sparse model of attention and multi-label classification, *International Conference on Machine Learning* (2016) 1614–1623.
- [39] A. Kuznetsova, R.H.B. Christensen, C. Bavay, P.B. Brockhoff, Automated mixed ANOVA modeling of sensory and consumer data, *Food Qual. Prefer.* 40 (2015) 31–38.
- [40] Lawrence J. Christiano, Terry J. Fitzgerald, The band pass filter, *Int. Econ. Rev.* 44 (2003) 435–465.
- [41] Aapo Hyvärinen, Erkki Oja, "Independent component analysis: algorithms and applications, *Neural Netw.* 13 (2000) 411–430.
- [42] R.D. Pascual-Marqui, Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details, *Methods Find. Exp. Clin. Pharmacol.* 24 (2002) 5–12.
- [43] S.R. Schweinberger, E.C. Pickering, I. Jentzsch, A.M. Burton, J.M. Kaufmann, Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions, *Cogn. Brain Res.* 14 (2002) 398–409.
- [44] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, *Inf. Process. Manage.* 45 (2009) 427–437.
- [45] C. Ferri, J. Hernández-Orallo, R. Modroño, An experimental comparison of performance measures for classification, *Pattern Recognit. Lett.* 30 (2009) 27–38.
- [46] Huaxiu Yao, Fei Wu, Jintao Ke, Xianfeng Tang, Yitian Jia, Siyu Lu, Pinghua Gong, Jieping Ye, Zhenhui Li, Deep multi-view spatial-temporal network for taxi demand prediction, *Thirty-Second AAAI Conference on Artificial Intelligence* (2018).
- [47] Yaguang Li, Rose Yu, Cyrus Shahabi, Yan Liu, Diffusion convolutional recurrent neural network: data-driven traffic forecasting, arXiv preprint arXiv:1707.01926 (2017).
- [48] J. Demšar, Statistical comparisons of classifiers over multiple data sets, *J. Mach. Learn. Res.* 7 (2006) 1–30.
- [49] T. Tuncer, S. Dogan, F. Ertam, A. Subasi, A novel ensemble local graph structure based feature extraction network for EEG signal analysis, *Biomed. Signal Process. Control* 61 (2020) 102006.
- [50] T. Tuncer, S. Dogan, F. Ertam, A. Subasi, A dynamic center and multi threshold point based stable feature extraction network for driver fatigue detection utilizing EEG signals, *Cogn. Neurodyn.* (2020).
- [51] E. Aydemir, T. Tuncer, S. Dogan, A Tunable-Q wavelet transform and quadruple symmetric pattern based EEG signal classification method, *Med. Hypotheses* 134 (2020) 109519.
- [52] T. Tuncer, S. Dogan, E. Akbal, A novel local senary pattern based epilepsy diagnosis system using EEG signals, *Australas. Phys. Eng. Sci. Med.* 42 (2019) 939–948.
- [53] B. Duchaine, Individual differences in face recognition ability, *Psychol. Sci. Agenda* (2015). Newsletter article, <https://www.apa.org/science/about/psa/2015/06/face-recognition>.
- [54] <https://github.com/dipayandewan94/AConvBi-LSTM-NN>.
- [55] Q. Zhang, Y. Liu, Improving brain computer interface performance by data augmentation with conditional Deep Convolutional Generative Adversarial Networks, arXiv preprint arXiv:1806.07108 (2018).
- [56] K. Liu, M. Zhang, Z. Pan, Facial expression recognition with CNN ensemble, *2016 International Conference on Cyberworlds (CW)* (2016) 163–166.
- [57] N.M. Blauch, M. Behrmann, D.C. Plaut, Deep learning of shared perceptual representations for familiar and unfamiliar faces: Reply to commentaries, *Cognition* (2020), 104484.
- [58] N.M. Blauch, M. Behrmann, D.C. Plaut, Computational insights into human perceptual expertise for familiar and unfamiliar face recognition, *Cognition* 104341 (2020).
- [59] N. Abudarham, L. Shkller, G. Yovel, Critical features for face recognition, *Cognition* 182 (2019) 73–83.
- [60] A. Özbeşay, S. Arica, Familiar/unfamiliar face classification from EEG signals by utilizing pairwise distant channels and distinctive time interval, *Signal Image Video Process.* 12 (2018) 1181–1188.
- [61] R.J. Oweis, E.W. Abdulhay, Seizure classification in EEG signals utilizing Hilbert-Huang transform, *Biomed. Eng. Online* 10 (2011) 38.
- [62] Z. Yuyi, L. Surui, S. Lijuan, L. Zhenxin, D. Bingchao, Motor imagery EEG discrimination using Hilbert-Huang entropy, *Biomed. Res.* 28 (2017) 727–733.
- [63] Q. Liu, L. Ma, S.Z. Fan, M.F. Abod, Q. Ai, K. Chen, J.S. Shieh, Frontal EEG temporal and spectral dynamics similarity analysis between propofol and desflurane induced anesthesia using Hilbert-Huang transform, *Biomed. Res. Int.* 2018 (2018) 1–16.