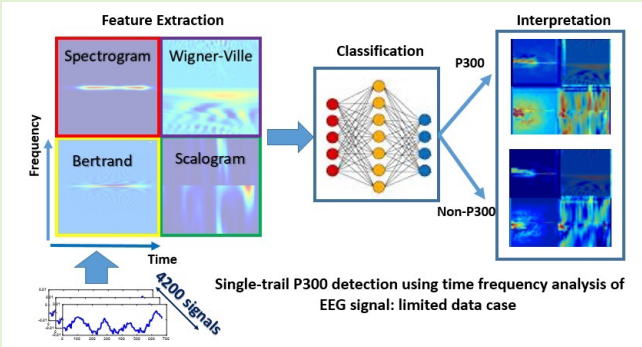
An Interpretable Convolutional Neural Network for P300 Detection: Analysis of Time Frequency Features for Limited Data

Mahnoosh Tajmirriahi, Zahra Amini, Hossein Rabbani, and Rahele Kafieh

***Abstract*—In this study, a new deep learning-based methodology is developed for P300 detection in brain computer interface (BCI) systems based on time-frequency (TF) features of EEG signals coupled to deep learning. The TF distributions can transform EEG signals to the TF images by simultaneous representation of time and frequency properties of the signal. However, they do not display the energy distribution of signals at different scales identically and their advantages may be greatly incorporated by using them together. Here, four TF images of single-trail EEG signal are computed and the concatenation of the TF (cTF) images of each signal is developed to be used as training data for a simple and lightweight deep learning-based classifier. The applied TF distributions are spectrogram, Wigner–Ville distribution, Morlet-scalogram, and Bertrand distribution. Performance of method is evaluated over limited data acquired from the normal and amyotrophic lateral sclerosis (ALS) datasets and accuracy of 96.56%, and 96.84% are achieved respectively, which is superior to the other comparing algorithms. Moreover, results of cross-subject classification indicate the promising ability of the method in eliminating calibration in BCI systems. Furthermore, the heat maps of the P300 and non-P300 classes are produced to explain important regions of cTF image for classifier decision and investigate which TF may help better classification. Results revealed the efficiency of cTF images for accurate P300 detection in simple structure classifiers having the advantage of fewer data and less memory requirement. This method can be employed in P300 speller BCI systems to improve the character recognition performance in.**



***Index Terms*—Time-frequency analysis, P300 detection, convolutional neural network, brain computer interface, amyotrophic lateral sclerosis (ALS).**

I. Introduction

ANY neurological disorder or damage in peripheral nerves and muscles such as amyotrophic lateral sclerosis (ALS), can reduce the interactions of a person with the environment [1]. Brain-computer interface (BCI) systems try to connect the brain of such patients to the environmental devices through translating brain signals into the computer commands. Recently, the P300 speller BCI systems are widely used in keyboard-display control software [2]. P300 wave is an event related potential (ERP) component, a large positive peak, that appears in a brain signal around 300 ms after infrequent stimuli and can be recorded through the electroencephalography (EEG) [3]. Correct performance of P300 speller BCI highly depends on the accurate detection of the P300 signals. However, due to low signal to noise ratio of EEG signal, single trial analysis of these signals is difficult and several trials are often carried out and then averaged before the signals being classified which slows down the classification process. In this work, we want to improve the single trial classification performance in P300 based BCIs by extracting proper features from EEG signals.

A. Related Works

Several methods have been proposed to detect P300 from single-trials or average of multi trials of EEG signals. These methods usually consist of feature extraction and classification

Manuscript received February 12, 2022; accepted February 25, 2022. Date of publication March 14, 2022; date of current version April 29, 2022. This work was supported in part by the Vice Chancellery for Research and Technology of Isfahan University of Medical Sciences under Grant 2400277 and Grant 2400276. The associate editor coordinating the review of this article and approving it for publication was Prof. Shih-Chia Huang. *(Corresponding author: Zahra Amini.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Wadsworth Center, NYS Department of Health, and Neuroelectrical Imaging and BCI Laboratory, IRCCS Fondazione Santa Lucia, Rome, Italy.

Mahnoosh Tajmirriahi, Zahra Amini, and Hossein Rabbani are with the Medical Image and Signal Processing Research Center, School of Advanced Technologies in Medicine, Isfahan University of Medical Sciences, Isfahan 81746734641, Iran (e-mail: mata.riahi@yahoo.com; zahraamini64@yahoo.com.au; rabbani.h@ieee.org).

Rahele Kafieh is with the Medical Image and Signal Processing Research Center, School of Advanced Technologies in Medicine, Isfahan University of Medical Sciences, Isfahan 81746734641, Iran, and also with the Biosciences Institute, Newcastle University, Newcastle upon Tyne NE24HH, U.K. (e-mail: rkafieh@gmail.com).

Digital Object Identifier 10.1109/JSEN.2022.3159475

1558-1748 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Indian Institute of Technology Gandhinagar. Downloaded on January 23,2025 at 12:20:13 UTC from IEEE Xplore. Restrictions apply.

TABLE I
RECENT P300-DETECTION APPROACHES (ACC.: ACCURACY, WS: WITHIN SUBJECT, CS: CROSS SUBJECT) SINGLE TRIAL PERFORMANCE OF STUDIES ARE REPORTED IF AVAILABLE

Classification approach	Method description	Single- trail Performance	Number of Subjects	Ref.
Time domain (TD)	PCA + Ensemble of SVM	Acc.= 30%	2 Normal	Kundu et al. [5]
	Histogram of Gradient Orientations + SVM	Channel-based Acc. = 20-85%	8 Normal, 8 ALS	Ramele et al. [6]
	Regularized Logistic Regression and CCA	Acc.= 86.3%	10 Normal	Diaz et al. [7]
Time-frequency domain (TF)	Wavelet Dictionaries + feature selection	Acc.=75%	18 Normal	Acevedo et al. [8]
	Wavelet transform + PCA +SVM	Acc.=98.53%	2 Normal	Hashmi et al. [9]
	Adaptive filtering + Spectrogram Graph	Acc. = 89.81%	2 Normal	Meng et al. [10]
	Stockwell Transform + statistical feature +SVM	Acc. = 84.1%	2 Normal	Vidal et al. [11]
	Cohen's class Feature + sparse autoencoder	Acc. = 82.70%	10 Normal	Ghazikhani et al. [12]
	Wavelet + Power spectral density + SVM	Acc. = 75%	2 Normal	Sahu et al. [13]
Deep learning-based	Sparse Autoencoder + Ensemble of SVM	Acc.=32%	12 Normal	Kundu et al. [14]
	Symbolized EEG + Autoencoded-(1D) CNN EEG	F1-score = 51.78, Acc.= 83.54	4 Normal	Venuto et al. [15]
	deep belief network (DBN)	Acc.= 40.3%	2 Normal	Lu et al. [16]
	DBN + Nesterov momentum	Acc. = 94.44%	9 Normal	Morabbi et al. [17]
	Deep learning (DL) + batch normalization	Acc. = 47%	3 Normal	Liu et al. [18]
	Convolutional neural networks (CNN)	Acc.= 39%	2 Normal	Cecotti et al. [19]
	Convolutional LSTM and DL ensembles	F1-score = 63%	4 Normal	Joshi et al. [20]
	Convolutional neural networks (CNN)	Acc.=90.55% for 5 trials	9 Normal	Shukla et al. [21]
	CNN with few filters	Acc. = 89%, 88%	26 Normal, 8 ALS	Gozalez et al. [22]
	3D input CNN	Acc. = 94.22%	2 Normal	Oralhan et al. [23]
	multi-scale CNN	WS Acc. = 60% CS Acc. = 30%	10 Normal	Wang et al. [24]
	EEG-Net (Depth wise and separable layers)	WS Acc. = 94% CS Acc. = 90%	15 Normal	Lawhern et al. [25]
	Multi-Task Autoencoder (ERPENET)	Acc. = 88.52%	139 Normal& Alcoholism	Ditthapron et al. [26]
	1D-CapsuleNet	Acc. = 84.2%	2 Normal	Liu et al. [27]

tasks to detect P300. Extraction of relevant features form EEG signals has an important role in efficient performance of subsequent classifier [4]. Till now, various features extracted in time domain (TD), time-frequency (TF) domain in particular Wavelet Transform (WT) have been used in P300 detection. Moreover, principal component analysis (PCA) and support vector machine (SVM) classifiers are widely used for dimensionality reduction of the features and classification.

Recent studies have tended to outsource the extraction and classification of features to deep networks. Most of the deep learning-based methods utilize raw data to train the network. However, these methods need large, diverse, and well-balanced training datasets in order to work effectively while availability of training data is limited in BCI applications. Therefore, methods using extracted features coupled to the deep learning network have also become very popular recently. We summarized some of the prominent classification approaches in P300 detection, their performance, and number of subjects of their dataset in Table I (mainly last three years).

B. Motivation

It has been investigated that, TF analysis of single trial of EEG epochs indicates event related changes in the magnitude and phase of the EEG signals at specific frequencies [28]. This is due to phase synchronized oscillation in ongoing EEG signal caused by ERPs [29]. Therefore, TF analysis can be applied as powerful tools to provide a more refined and detailed representation of the ERP signals such as P300 in the single trial of EEG epochs [30]. Historically, features extracted from TF analysis have been used to detect P300 in several studies. For instance, features extracted from the continuous wavelet transform and the t-value scalogram were used in [31] to detect ERP signals. Also, features extracted from pseudo Wigner-Ville distribution (SPWVD) of EEG signal were used in [32] for p300 detection. This method was the winner of the BCI Competition 2003. Moreover, according to Table I, recent studies also revealed the capability of features extracted from TF analysis in P300 detection.

On the other hand, CNN can automatically detects the important features of the images without any human supervision. They can be much more accurate, computationally efficient, and locally focused by converting non-image data into an image [33]. TF analysis produces TF matrices which can be easily converted to the TF images. As a result, by using TF images, indeed converting EEG signal to the image, CNN classifier can be well trained for P300 detection with fewer data and trials than current methods.

The above findings motivate us to find efficient TF images of EEG signals to develop an accurate CNN-based single trial P300 detection method utilizing limited training data.

Here, we utilize the concatenation of the TF (cTF) images obtained from STFT, wavelet transform (WT), Wigner–Ville distribution (WVD), and Bertrand distribution (BD) as a novel feature to train a simple convolutional neural classifier. We focus on the P300 detection part of the P300 speller as a binary classification task and demonstrate the efficiency of cTF images in P300 detection. To this end, we examine the proposed method on two datasets acquired from healthy and ALS subjects and compare the results with some state of the art methods. Results showed the superior performance of the proposed method in within-subject and cross-subject classification of P300 signals utilizing much fewer training data than comparing methods. Furthermore, we provide post-hoc interpretation by visualizing the proposed CNN classifier to explain which parts of the images are highlighted for this classification. This can help future studies select more efficient TF features of ERP signals.

The rest of this paper is organized as follows: In Section II, the dataset is introduced and the TF analysis, and the architecture of CNN classifier are described. In section III, we present evaluation results of the proposed method as well

Authorized licensed use limited to: Indian Institute of Technology Gandhinagar. Downloaded on January 23,2025 at 12:20:13 UTC from IEEE Xplore. Restrictions apply.

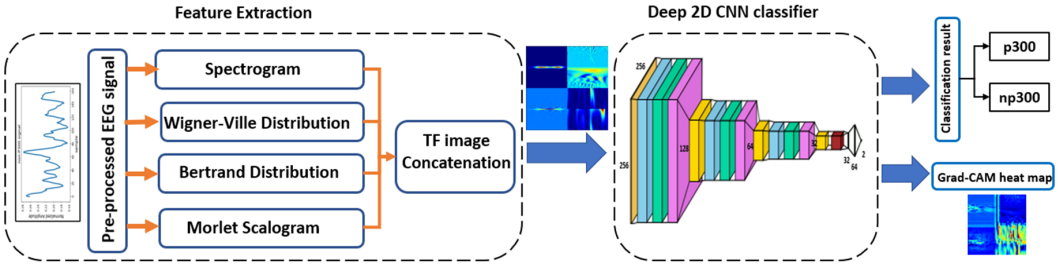


Fig. 1. Block diagram of proposed TF-based P300 classification method. Pre-processing consists of band pass filtering and normalization.

as interpretability. Finally, in section IV we provide a detailed discussion and conclusion.

II. Methods

A. Dataset

We used two benchmark datasets acquired from 6 × 6 character matrix P300 speller paradigm described in [2].

The 1st dataset (dataset I) is a widely used public dataset provided by Wadsworth center for the BCI Competition III [34]. The data has been acquired from two subjects A, and B. For each subject a training and a test set are available. However, we applied the training sets of both subjects since they are labeled. EEG signal have been recorded in 64 channels. Here, we select seven suitable channels (Po7, Po8, Fz, C3, Cz, C4, and Pz based on 10-20 recording system) [35] to use in the proposed P300 detection method. This dataset consists of 2550 potentially P300 signals and 12,750 EEG signals without P300 (non-P300), for each channel of subject. The signals were digitized at a rate of 240 Hz and band-pass filtered with cutoff frequencies between 0.1 and 60 Hz. We extracted segments of 667 ms (160 samples) of EEG data after every stimulus.

The 2nd dataset (dataset II) was acquired by Riccio *et al.* [36], from P300 speller with eight subjects suffering from ALS. This dataset contains eight EEG channels (Fz, Cz, Pz, Oz, P4, P3, PO8, and PO7) and is based on the international 10–10 recording system. The signals were digitized at a rate of 256 Hz and band-pass filtered with cutoff frequencies between 0.1 and 30 Hz. During acquiring data, each rows and columns of the matrix flashed randomly 10 times at a rate of 4 Hz for 125 ms, having 125 ms inter stimulus interval. We extracted segments of 800 ms (204 samples) of EEG data after every stimulus. This dataset consists of 700 potentially P300 and 3,500 non-P300, for each channel of eight subjects.

The raw EEG signals are weak and mixed with non-EEG signals or background noise. In order to remove unwanted short and long term fluctuations, the EEG signals of each dataset were filtered applying the band pass filter of 0.1Hz to 20Hz implemented by 8th order Butterworth filter. This filter provides sharp transition band and well amplitude preserving in pass band, and caused satisfactory results in recent studies [15]. The band pass filtered signal was normalized to [0, 1] to provide same amplitude range for all selected channels. Then the pre-processed signals were used in the TF analysis. The overall procedures of the proposed TF feature extraction and P300 classification model is shown in Fig. 1.

B. TFD Methods

The EEG signal is a non-stationary data and has time varying instantaneous frequency. Therefore, it cannot be fully described using temporal and frequency domain features separately. Time-frequency distributions (TFDs) are useful to characterize such non-stationary signals. Indeed, TFD finds out how the energy density of the signal is simultaneously distributed in time and frequency. However, TFDs obtained by different transform methods are not the same, and they represent different underlying frequency information of data. They can be included in two important classes: Cohen's class and the affine class [37]. Although several TFDs have been used separately for describing EEG signals, the well-describing TFD is still unknown and simultaneous application of TFDs can help the network learn much due to advantages of each distribution. Well-known distributions of Cohen's class are STFT, and Wigner–Ville distribution (WVD) which are covariant to time-frequency shift of the signal. STFT was derived from the discrete Fourier transform of temporal windowed sections of signal of finite support and defined as follows [37].

$$STFT\{x[n]\} = \sum_{n=-\infty}^{n=\infty} x[n]\, w[n-m] e^{-j\omega n} \tag{1}$$

where $x[n]$ is the signal, and $w[n]$ is the window function. Here, we used Hamming window for section division of the signal and the squared magnitude of STFT, so called spectrogram, is used as the first TFD image.

Another member of Cohen's class, WVD, can provide the highest possible temporal versus frequency resolution in accordance to limitations of the uncertainty principle. Thereby we used it as second TFD image. A discrete-time continuous frequency expression of WVD is defined as (2) [38] where f_s is sampling rate of continuous signal $x(t)$, and $x[n]$ is the signal defined in the $0, 1, \ldots, N$ where $N$is even.

$$WVD_x\left(\frac{n}{f_s}, f\right) = \frac{2}{f_s} \sum_{m=\frac{-N}{2}+1}^{\frac{N}{2}-1} x[n+m] x^*[n-m]\, e^{-j4\pi f\left(\frac{m}{f_s}\right)} \tag{2}$$

Two prominent members of the affine class are the scalogram (squared magnitude of wavelet transform) and Bertrand distribution. The main property of affine class is that it satisfies time shifts and scale changes rather than time and frequency shifts. This means that, members of this class have time-scale covariance [39]. We used Morlet wavelet, which is very popular in TF analysis of EEG data, to produce scalogram of EEG signal.

The Bertrand distribution (BD) is a member of bilinear class that has close relation to a hyperbolic time frequency geometry. This property makes it suitable for the analysis of

Authorized licensed use limited to: Indian Institute of Technology Gandhinagar. Downloaded on January 23,2025 at 12:20:13 UTC from IEEE Xplore. Restrictions apply.

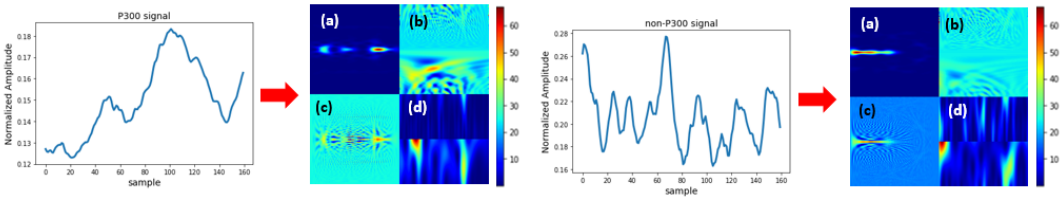


Fig. 2. Sample P300 (left panel) and non-P300 segments (right panel) of EEG signal averaged over segments and their corresponding concatenated TF (cTF) images (a) spectrogram, (b) Wigner-Ville distribution, (c) Bertran distribution, and (d) Morlet scalogram.

self-similar signals such as EEG signals [40], [41]. A discrete-time continuous frequency expression of BD can be defined as (3), in which f_s is sampling rate of continuous signal $x(t)$ and N is the length of the signal. In BD, group delay corresponds to a hyperbola and the instantaneous frequency also has a hyperbolic shape. These properties may well represent the complex fractal nature of EEG signals.

In addition, BD is invariant to hyperbolic time-shift and time frequency scaling, operations which are important in analysis of fractal signals.

$$BD_x\left(\frac{n}{f_s}, f\right) = \frac{2}{f_s}\sum_{m=(-N)/2+1}^{N/2-1} U_x\left(nF_0\left(\frac{m}{n}\right), m\right) G_0\left(\frac{m}{n}\right) e^{-j4\pi f\left(\frac{m}{f_s}\right)} \tag{3}$$

The hyperbolic functions $U_x(.)$ $F_0(.)$, $G_0(.)$ are defined as:

$$U_x(m,n) = x[n+m]x^*[n+m] \tag{4}$$

$$F_0(k) = \frac{k}{2}coth\left(\frac{k}{2}\right), G_0(k) = \frac{k/2}{sinh\left(\frac{k}{2}\right)} \tag{5}$$

The four mentioned TFDs are applied on the EEG signals (P300, and non-P300) and the zero-frequency component of the resultant TFDs are shifted to the center of the spectrum. In this way, for each EEG segment four TF images of size 160 × 160 for dataset I, and 204 × 204 for dataset II are obtained. The intensity range of all four TF images of each segment will be normalized to equalize their effect. At last, for each segment, the produced TF images are concatenated to obtain a single image of size 320 × 320 for dataset I, and 408 × 408 for dataset II. These images, namely cTF images, will be resized to 256 × 256 to feed to the CNN classifier. As it will be demonstrated, these cTF images greatly improve the performance of the CNN due to locality-based nature of image-processing CNNs. Figure 2 illustrates the result of sample averaged P300 and non-P300 segments and their corresponding cTF images of dataset I.

C. Classifier Network

The architecture of proposed classifier network is depicted in Fig. 3. The cTF images of size 256 × 256 are applied to the input layer of classifier. This network consists of three 2D convolutional layers (Conv2D) and three dense layers (dense part). Conv2D layers can determine correlations of adjacent pixels by applying multiple filters and can extract efficient local features. All the Conv2D layers are of the size

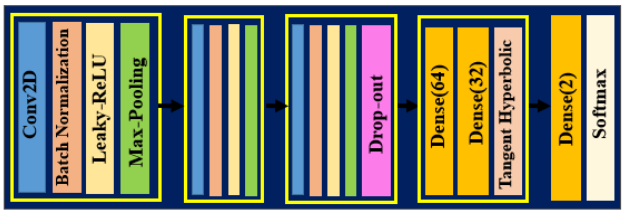


Fig. 3. Artichecture of the proposed classifier network. The flatten layer before the first dense layer is not represened. More details are provided in text.

32 × 5 × 5. The batch normalization layers are used after each Conv2D layer to stabilize the learning process and reduce the required training epochs. For the last Conv2D layer, we use leaky rectified linear unit (LeakyRelu of α =0) and for other Conv2D layers we choose LeakyRelu (α =0.2) as an activation functions. This selection adds the controllable nonlinearity to the network for better tracking the nonlinear pattern of the input image. Each activation function is followed by the Max-pooling layer to reduce computation complexity and extract more sharp and effective features. All the Max-pooling layers have pooling size and stride of size (2, 2) except the last one which has pooling size of (4, 4), and stride size of (6, 6). At the end of convolution part, we applied a dropout layer with dropout rate of 0.3 to prevent classifier from over fitting. The dense part of the classifier consists of two dense layers of size 64 and 32 followed by the tangent hyperbolic layer as an activation function. Afterwards, a dense layer of size 2 with a Softmax activation function completes the classifier network to specify labels to the images.

III. Results

A. Evaluation Metrics

In order to evaluate the performance of the classifier, we used different standard metrics: accuracy, precision, recall, and F1-score. These standard metrics are usually used to assess the performance of P300 detection algorithms. However, the most correlated metric to the accuracy of subsequent character prediction step is still unclear. Liu *et al.* [18] investigated this correlation and specified the F1-score as the metric having the highest correlation values for both subjects. This metric can be used to compare binary classifiers to determine which one can provide more reliable character prediction.

B. Results of Within Subject Classification of Dataset I

In order to train and evaluate within subject performance of the classifier, for each subject, 300 P300 and 300 non-P300 preprocessed signals were randomly selected from seven

Authorized licensed use limited to: Indian Institute of Technology Gandhinagar. Downloaded on January 23,2025 at 12:20:13 UTC from IEEE Xplore. Restrictions apply.

TABLE II
NUMBER OF TFD IMAGES APPLIED AS TRAIN AND TEST SETS

	Train set		Test set	
	P300	Non-P300	P300	Non-P300
A	2100 (300 × 7)	2100 (300 × 7)	350 (50 × 7)	350 (50 × 7)
B	2100 (300 × 7)	2100 (300 × 7)	350 (50 × 7)	350 (50 × 7)

selected channels (Po7, Po8, Fz, C3, Cz, C4, and Pz) for training set and 50 P300 and 50 non-P300 signals were randomly selected from aforementioned channels for test set. During training, 20% of training data was dedicated as validation set. Details of the number of training and test images of these dataset are reported in Table II. It should be noted that a limited amount of data is purposely used for training set to show that the proposed TF images are discriminative enough to help the network achieve acceptable performance. This advantage can be applied in situations where there is not enough data to train the classifier.

To evaluate the proposed method, the proposed network described in section II. C was implemented and trained for 300 iterations. We used the Binary Cross Entropy as the loss function and Adadelta as optimizer. Other values of hyper-parameters of the network were tuned as follows:

- Learning rate of the Adadelta optimizer is initialized to 10^{-3} and decayed with the rate of 0.95. Adadelta is a variant of gradient descent algorithm which have learning rate adapting over time.
- The dropout is set to 0.3 due to the highest classification accuracy in cross validation experiments.
- The batch size is set to 8. It was empirically observed that smaller batch sizes provide faster convergence, and more generalization to the test data.
- All parameters are initialized using random uniform method.

In order to produce cTF images, common TFD images (spectrogram, WVD, BD, Morlet scalogram) were computed first. Moreover, Stockwell transform (S-transform) is a TF analysis that has been recently used to extract features for P300 detection [11]. Here, the S-transform is also applied to the EEG data to compare its results with other TFD images. TFD images were separately used to train the classifier network to investigate their ability in discrimination of P300 and non-P300 signals individually. Results of calculated evaluation metrics of classifier for mentioned TFD images of test dataset is reported in Table III. It can be seen from the table that none of these methods could provide performance greater than 90.13% in terms of evaluation metrics. This may indicate the fact that TFDs obtained by different transform methods cannot merely represent all underlying frequency information of data and can highlight the motivation of concatenating the TFD images.

Then, cTF images were produced by concatenation of the TFD images (spectrogram, WVD, BD, Morlet scalogram) having highest performance in their individual tests (Table III).

Results of calculated evaluation metrics of classifier for the cTF images of the test dataset is reported in Table IV. Furthermore, the proposed method was compared with some state of the art methods in this table. According to this table, the proposed method has higher performance than other methods. Since the F1-score has been proposed to compare

TABLE III
THE CALCULATED EVALUATION METRICS FOR DATASET I IN (%) USING TFD IMAGES SEPARATELY (PRES.: PRECISION, ACC.: ACCURACY)

SUBJECTS/ METHOD		PRES.	RECALL	ACC.	F1-SCORE
A	**Spectrogram**	83.06	86.76	84.57	84.86
	WVD	81.76	85.21	83.43	83.50
	BD	85.55	87.31	86.29	86.42
	Scalogram	86.36	87.21	87.86	86.78
	S-Transform	79.67	82.34	81.43	80.54
B	**Spectrogram**	85.45	88.65	86.29	87.02
	WVD	84.18	86.25	84.71	85.20
	BD	87.67	89.25	88.43	88.45
	Scalogram	89.42	90.87	89.43	90.13
	S-Transform	83.12	84.05	83.57	83.58

TABLE IV
THE CALCULATED EVALUATION METRICS FOR DATASET I IN (%) (PRES.: PRECISION, ACC.: ACCURACY)

SUBJECTS/ METHOD		PRES.	RECALL	ACC.	F1-SCORE
A	**MCNN1** [19]	30.80	69.03	68.99	42.60
	MCNN3 [19]	31.72	67.43	70.38	43.14
	CNN3D+ConvLSTM [20]	36.55	65.43	75.31	46.90
	BN3 [18]	33.62	63.67	72.99	44.00
	CONV-LSTM [20]	36.06	64.26	75.05	46.20
	Auto-encoded CNN [15]	36.57	67.23	75.10	47.37
	Proposed method	**96.04**	**95.50**	**95.71**	**95.76**
B	**MCNN1** [19]	38.30	73.40	68.99	50.34
	MCNN3 [19]	40.89	69.23	70.38	51.41
	CNN3D+ConvLSTM [20]	44.58	70.50	80.48	54.62
	BN3 [18]	42.93	66.97	79.66	52.32
	CONV-LSTM [20]	44.01	69.97	80.16	54.04
	Auto-encoded CNN [15]	48.64	66.53	82.71	56.19
	Proposed method	**97.10**	**95.83**	**97.06**	**96.46**

performance of P300 classifiers for further character prediction [18], paired t-test was down for F1-score to investigate that the proposed method achieves significantly higher P300 classification F1-score than other methods. Figure 4 shows the boxplots of F1-score annotated with statistical test (all paired p-values $< 1e{-}4$). F1-score improvement indicates promising ability of the classifier in providing much accurate character prediction especially in paradigms having fewer repetitions. As a result, the proposed features considerably improved the performance of the CNN classifier in comparison to other deep learning-based methods that use raw data.

In order to examine the complexity of the proposed model, it was realized in Python by using the Keras library with Tensorflow backend. It took approximately an hour to train the classifier (For a 12 GB, NVIDIA Tesla K80 GPU, and cTF images of size 256×256). The proposed CNN classifier has approximately 105000 trainable parameters which is pretty small number compared to other popular convolutional models. However, this number of parameters can be more reduced by adding extra Max-pooling layers exactly before the dense part of the model. In addition, the multiplicative complexity of N points TFD calculation is of order N^2 for STFT, and WVD and of order N for WT (where N is the sample length of the signal). For BD this complexity is higher but less than the upper band of communication complexity of TF analysis which is about $N^2 log^2(N)$ [42]. This point indicates that, in addition to the generality, the proposed P300 detection

Authorized licensed use limited to: Indian Institute of Technology Gandhinagar. Downloaded on January 23,2025 at 12:20:13 UTC from IEEE Xplore. Restrictions apply.

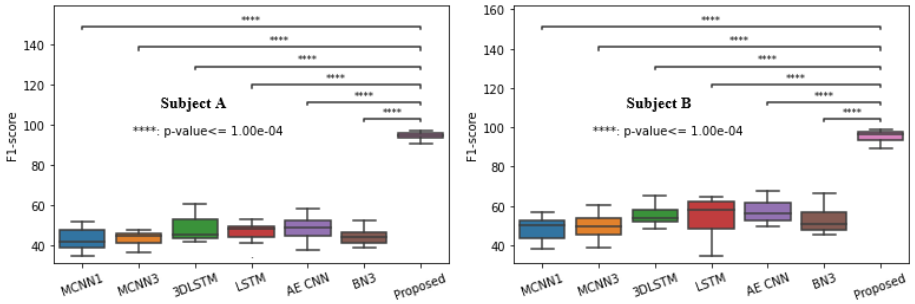


Fig. 4. Boxplot of the F1-score of the proposed method and comparing ones (Table IV) for both subjects. Paired t-test statistical annotation is added to the boxplot as well. ***** indicates that F1-scores of the comparing pairs have significant difference with p-value less than 1e-4.

method is simple enough to be efficiently implemented as a hardware BCI system.

C. Results of Within Subject Classification of Dataset II

In order to survey the generality of the proposed features for P300 detection, the method was examined utilizing dataset II. To this end, P300, non-P300 signals were randomly selected from dataset II (as detailed in Table II), and their cTF images were produced. Results of calculated evaluation metrics of classifier for the test data of dataset II is reported in Table V and compared with available studies on this dataset (in some studies, only the accuracy was reported). It can be seen that the proposed method outperforms other methods in terms of P300 detection.

Recently, Alvarado-Gonzalez *et al.* [22] compared the mean AUC values of some of the existing CNN architectures for dataset II and proposed a SepConv1D which performs better than compared methods. The mean AUC value of SepConv1D is reported 86.00% while the mean AUC value of our proposed method is 98.83%. This indicates the superiority of the proposed features in differentiation of P300 and non-P300 signals for ALS subjects as well.

D. Cross-Subject Performance of the Proposed Method

The P300 classification based on the information extracted by other subjects is known as cross-subject classification. To calculate the AUC of the cross-subject classification, as in within-subject P300 detection (section III. B, C), we carried out the experiment utilizing training set contained cTF images of P300, non-P300 signals randomly selected from the subjects of dataset I to train a model for testing on subjects of dataset II. The pairwise AUC values are computed over 700 test cTF images of each subjects of dataset II and are illustrated in Fig. 5. The mean ± std of cross-subject AUC values is 96.83± 1.57 which is lower than that of within-subject but is still high enough to provide a reasonable performance for the cross-subject classification. It can hopefully promise a great advance in the generalization of BCI systems by eliminating the subject calibration.

E. Model Interpretation

In order to explain how the proposed classifier makes its decision of P300 and non-P300 signals, we provide a post hoc interpretation by visualizing the regions of input that are important for predictions of output classes. In this way we can interpret what the model learns from proposed method. Here, we used a gradient-weighted class activation mapping

TABLE V
THE CALCULATED EVALUATION METRICS FOR DATASET II IN (%)
(PRES.: PRECISION, ACC.: ACCURACY, F1: F1-SCORE)

	SUBJECTS/ METHOD	PRES.	RECALL	F1	ACC.
Proposed Method	**Subject 1**	94.40	97.11	95.73	95.70
	Subject 2	94.92	96.00	95.44	95.41
	Subject 3	96.56	97.72	97.00	97.10
	Subject 4	96.60	97.12	96.82	97.53
	Subject 5	95.21	96.32	95.74	95.70
	Subject 6	97.12	98.50	97.79	97.82
	Subject 7	96.18	97.62	96.89	97.00
	Subject 8	98.50	97.72	98.09	96.71
	Mean Values	**96.18**	**97.26**	**96.68**	**96.84**
	Random committee [13]	83.30	75.00	78.93	75.00
	Temporal Features + Linear Discriminant analysis [36]				87.4
	Temporal Features + Linear classifier [43]				89.5

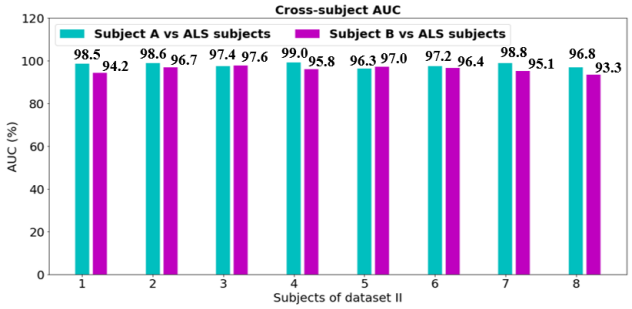


Fig. 5. Bar plot of cross-subject AUC calculated for subjects A, B of dataset I versus ALS subjects of dataset II.

(Grad-CAM) technique which uses the class-specific gradient information of the last convolutional layer to produce a localization map of important regions of the input image [44]. Grad-CAM produces the heat maps that, for a particular class, indicates the regions of input image which help the classifier make the decision for prediction. It can produce a separate visualization for every class present in the image. To this end, a test image is applied to the trained network and based on the prediction of the network, Grad-CAM produces the heat map for predicted class. Here, the input images consist of four TFDs, hence, it is expected that the heat maps indicate the TFDs which classifier may use for each class. Sample heat map images for the P300, and non-P300 classes are portrayed in Fig. 6, respectively. The colors in the heat map images vary from red zones (highest contribution of the region in the prediction) to dark blue zones (lowest contribution of the region in the prediction).

Authorized licensed use limited to: Indian Institute of Technology Gandhinagar. Downloaded on January 23,2025 at 12:20:13 UTC from IEEE Xplore. Restrictions apply.

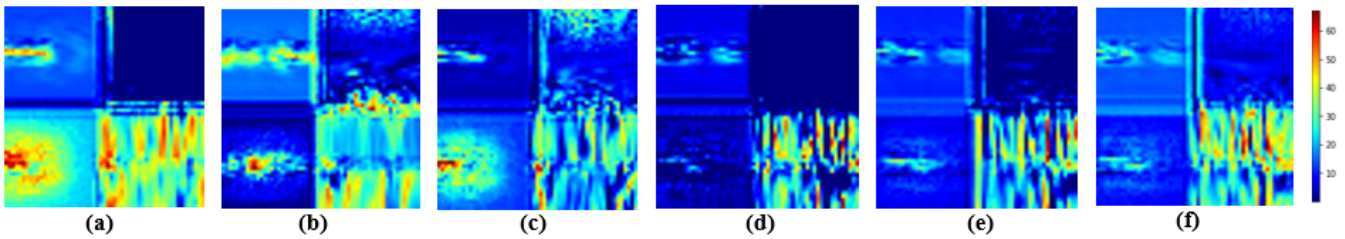


Fig. 6. (a), (b), (c) Sample P300 class-specified Heat maps. (d), (e), (f) sample non-P300 class-specified Heat maps. The heatmaps are correspondant to cTF images which a sample example of them is depicted in Fig. 2.

According to heat map results, the classifier may explain the P300 signal by mostly considering the spectrogram, Morlet scalogram, and BD part of cTF image (refer to Fig. 6(a-c)). Moreover, it may explain the non-P300 signal by highlighting the Morlet scalogram part of the cTF image (refer to Fig. 6 (d-f)). However, WVD part of cTF image has the least effect on classification decision. Perhaps this point is due to complicated interference terms in the TF plane caused by WVD which reduce differentiation of P300 and non-P300 signals.

IV. Discussion and Conclusion

Current P300 detection methods suffer from several inherent drawbacks. Computational complexity of the methods are often high and real-time performance is limited. Several trials must be carried out and averaged to increase the performance of classification which slows down the process. Individualized calibration is necessary due to unique EEG patterns of users. Data acquisition is hard in particular for disabled persons who are usually the target society. Obviously, a P300 detection method which can address these issues will be crucial for practical BCI systems. The purpose of this study is to develop a novel P300 detection method which has low complexity, needs less training data and individualized calibration, and at the same time, to further increase the accuracy and robustness compared to previous studies. For this purpose, we proposed a novel TF based feature to represent various dynamic aspects of the EEG signal, and then, we used a deep learning based method to distinguish the P300 signals.

In this study, we computed the TFD images of limited number of P300 and nonP300 signals and used them to train a lightweight CNN. These signals were acquired from normal subjects. Evaluation results indicated that none of these features could provide performance greater than 90% alone (Table III). Then, the concatenation of the most effective TFDs (cTF), was applied to the CNN to characterize P300 and non-P300 EEG signals. Evaluation results indicated outstanding performance of these concatenation rather than using TFD images lonely (Table IV). The accuracy was increased to 97.06% for subject B. Statistical results also revealed significant improvement of proposed method comparing to previous ones (Fig. 4). We repeated the training and evaluation of the CNN by cTF images of 8 ALS subjects as well. Results indicated the outstanding performance of our method (Table V). Here, mean accuracy of the proposed method was enhanced 96.84%.

Results of within-subject classification illustrated the reliability and reproducibility of the proposed P300 detection method provided by the cTF images through their compliance with dynamic brain source of ERP signals. More detailed survey of Table IV indicates that in contrast to other comparing methods, the difference between the calculated metrics of two subjects is small in the proposed method. This indicates that the proposed model can hopefully be applied in cross-subject classifications with greater data variability as a general framework

In addition, outstanding results of the cross subject classification of healthy and ALS dataset indicated that the proposed method can eliminate the subject calibration and can provide a reasonable performance when subject calibration is not possible (Fig. 5).

We applied the Grad-CAM method to infer and interpret from the trained network and visualize the most important TFDs contribute in network decision (Fig. 6). Post hoc interpretation revealed that the scalogram, plays important role in the explanation of the P300 and non-P300 signals. This is consistent with previous studies which tried to extract discriminative features from wavelet transform of the EEG signals. Another important point in the heat maps of the P300 class is the notable role of BD part of cTF image in the decision of the classifier. It indicates that BD may be hopefully used in representation of P300 signals. It may overcome the limitations of scalogram due to its bi-frequency kernel distribution which localizes perfectly the hyperbolic group delay signals which corresponds to the self-similar nature of EEG signals.

The advantages of the proposed method can be summarized as follows. This method has high performance and relatively low complexity, thus, can enhance the classification performance in P300 speller based BCIs and can generally be applied in real-time BCI applications in particular for ALS people. The simple structure of proposed CNN, having the advantage of fewer data and less memory requirement, makes it promising for mobile, and hardware applications as well. It may remove the individualized calibration which can hopefully overcome the data acquisition challenges especially in the ALS patients.

However, the method has some challenging issues which should be acknowledged. The procedure for tuning the hyperparameters in the present study is done manually. Using automatic parameter optimization would provide more generalization and can be investigated in future work. In addition, there is a trade-off between the parameters of CNN (through adding Max-pooling layers) and the performance of the network which must be optimized for hardware implementation. The length of extracted P300 and nonP300 segments, the resolution of the TFD images, and the size of cTF images

Authorized licensed use limited to: Indian Institute of Technology Gandhinagar. Downloaded on January 23,2025 at 12:20:13 UTC from IEEE Xplore. Restrictions apply.

can also affect the performance of the network and should be engineered carefully.

Overall, this study introduced novel efficient discriminative features for P300 characterization. In this way, it would help develop new methodology for reliable, fast, real-time, and reproducible P300 detection.

REFERENCES

[1] L. C. Wijesekera and P. N. Leigh, "Amyotrophic lateral sclerosis," *Orphanet J. Rare Diseases*, vol. 4, no. 1, pp. 1–22, 2009.

[2] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.

[3] W. S. Pritchard, "Psychophysiology of P300," *Psychol. Bull.*, vol. 89, no. 3, p. 506, 1981.

[4] T. Fang *et al.*, "Recent advances of P300 speller paradigms and algorithms," in *Proc. 9th Int. Winter Conf. Brain-Comput. Interface (BCI)*, Feb. 2021, pp. 1–6.

[5] S. Kundu and S. Ari, "P300 detection with brain–computer interface application using PCA and ensemble of weighted SVMs," *IETE J. Res.*, vol. 64, no. 3, pp. 406–414, May 2018.

[6] R. Ramele, A. J. Villar, and J. M. Santos, "Histogram of gradient orientations of signal plots applied to P300 detection," *Frontiers Comput. Neurosci.*, vol. 13, p. 43, Jul. 2019.

[7] C. F. B. Diaz and A. F. R. Olaya, "A novel method based on regularized logistic regression and CCA for P300 detection using a reduced number of EEG trials," *IEEE Latin Amer. Trans.*, vol. 18, no. 12, pp. 2147–2154, Dec. 2020, doi: 10.1109/TLA.2020.9400443.

[8] R. Acevedo, Y. Atum, I. Gareis, J. B. Manresa, V. M. Bañuelos, and L. Rufiner, "A comparison of feature extraction strategies using wavelet dictionaries and feature selection methods for single trial P300-based BCI," *Med. Biol. Eng. Comput.*, vol. 57, no. 3, pp. 589–600, Mar. 2019.

[9] M. F. Hashmi, J. D. Kene, D. M. Kotambkar, P. Matte, and A. G. Keskar, "An efficient P300 detection algorithm based on kernel principal component analysis-support vector machine," *Comput. Electr. Eng.*, vol. 97, Jan. 2022, Art. no. 107608, doi: 10.1016/j.compeleceng.2021.107608.

[10] H. Meng, H. Wei, T. Yan, and W. Zhou, "P300 detection with adaptive filtering and EEG spectrogram graph," in *Proc. IEEE Int. Conf. Mechtron. Autom. (ICMA)*, Aug. 2019, pp. 1570–1575.

[11] A. Pérez-Vidal, C. Garcia-Beltran, A. Martínez-Sibaja, and R. Posada-Gómez, "Use of the stockwell transform in the detection of P300 evoked potentials with low-cost brain sensors," *Sensors*, vol. 18, no. 5, p. 1483, May 2018.

[12] H. Ghazikhani and M. Rouhani, "A deep neural network classifier for P300 BCI speller based on Cohen's class time-frequency distribution," *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 29, no. 2, pp. 1226–1240, Mar. 2021.

[13] M. Sahu, S. Verma, N. K. Nagwani, and S. Shukla, "EEG signal analysis and classification on P300 speller-based BCI performance in ALS patients," *Int. J. Med. Eng. Inform.*, vol. 12, no. 4, pp. 375–400, 2020.

[14] S. Kundu and S. Ari, "P300 based character recognition using sparse autoencoder with ensemble of SVMs," *Biocybern. Biomed. Eng.*, vol. 39, no. 4, pp. 956–966, Oct. 2019.

[15] D. De Venuto and G. Mezzina, "A single-trial P300 detector based on symbolized EEG and autoencoded-(1D)CNN to improve ITR performance in BCIs," *Sensors*, vol. 21, no. 12, p. 3961, Jun. 2021.

[16] Z. Lu, N. Gao, Y. Liu, and Q. Li, "The detection of P300 potential based on deep belief network," in *Proc. 11th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2018, pp. 1–5.

[17] S. Morabbi, M. Keyvanpour, and S. V. Shojaedini, "A new method for P300 detection in deep belief networks: Nesterov momentum and drop based learning rate," *Health Technol.*, vol. 9, no. 4, pp. 615–630, Aug. 2019.

[18] M. Liu, W. Wu, Z. Gu, Z. Yu, F. F. Qi, and Y. Li, "Deep learning based on batch normalization for P300 signal detection," *Neurocomputing*, vol. 275, pp. 288–297, Jan. 2018.

[19] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 433–445, Mar. 2011.

[20] R. Joshi, P. Goel, M. Sur, and H. A. Murthy, "Single trial P300 classification using convolutional LSTM and deep learning ensembles method," in *Proc. Int. Conf. Intell. Hum. Comput. Interact.*, 2018, pp. 3–15.

[21] P. K. Shukla, R. K. Chaurasiya, and S. Verma, "Performance improvement of P300-based home appliances control classification using convolution neural network," *Biomed. Signal Process. Control*, vol. 63, Jan. 2021, Art. no. 102220.

[22] M. Alvarado-González, G. Fuentes-Pineda, and J. Cervantes-Ojeda, "A few filters are enough: Convolutional neural network for P300 detection," *Neurocomputing*, vol. 425, pp. 37–52, Feb. 2021.

[23] Z. Oralhan, "3D input convolutional neural networks for P300 signal detection," *IEEE Access*, vol. 8, pp. 19521–19529, 2020.

[24] H. Wang *et al.*, "Performance enhancement of P300 detection by multiscale-CNN," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.

[25] V. Lawhern, A. Solon, R. Waytowich, M. Gordon, P. Hung, and B. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, p. 56013, 2018.

[26] A. Ditthapron, N. Banluesombatkul, S. Ketrat, E. Chuangsuwanich, and T. Wilaiprasitporn, "Universal joint feature extraction for P300 EEG classification using multi-task autoencoder," *IEEE Access*, vol. 7, pp. 68415–68428, 2019.

[27] X. Liu, Q. Xie, J. Lv, H. Huang, and W. Wang, "P300 event-related potential detection using one-dimensional convolutional capsule networks," *Expert Syst. Appl.*, vol. 174, Jul. 2021, Art. no. 114701.

[28] S. Makeig, S. Debener, J. Onton, and A. Delorme, "Mining event-related brain dynamics," *Trends Cognit. Sci.*, vol. 8, no. 5, pp. 204–210, May 2004.

[29] S. Makeig *et al.*, "Dynamic brain sources of visual evoked responses," *Science*, vol. 295, no. 5555, pp. 690–694, Jan. 2002.

[30] B. J. Roach and D. H. Mathalon, "Event-related EEG time-frequency analysis: An overview of measures and an analysis of early gamma band phase locking in schizophrenia," *Schizophr. Bull.*, vol. 34, no. 5, pp. 907–926, Aug. 2008.

[31] V. Bostanov, "BCI competition 2003-data sets Ib and IIb: Feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1057–1061, Jun. 2004.

[32] E. Ebrahimzadeh, S. M. Alavi, A. Bijar, and A. Pakkhesal, "A novel approach for detection of deception using smoothed pseudo wigner-ville distribution (SPWVD)," *J. Biomed. Sci. Eng.*, vol. 6, no. 1, pp. 8–18, 2013.

[33] A. Sharma, E. Vans, D. Shigemizu, K. A. Boroevich, and T. Tsunoda, "DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture," *Sci. Rep.*, vol. 9, no. 1, pp. 1–7, Dec. 2019.

[34] B. Blankertz *et al.*, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, Jun. 2006.

[35] A. Rakotomamonjy and V. Guigue, "BCI competition III: Dataset II-Ensemble of SVMs for BCI P300 speller," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1147–1154, Mar. 2008.

[36] A. Riccio *et al.*, "Attention and P300-based BCI performance in people with amyotrophic lateral sclerosis," *Frontiers Hum. Neurosci.*, vol. 7, p. 732, Nov. 2013.

[37] L. Cohen, *Time-Frequency Analysis*, vol. 778. Upper Saddle River, NJ, USA: Prentice-Hall, 1995.

[38] J. O'Toole, M. Mesbah, and B. Boashash, "A discrete time and frequency Wigner–Ville distribution: Properties and implementation," in *Proc. Int. Conf. Digit. Signal Process. Commun. Syst.*, vol. 400, 2005, pp. 19–21.

[39] F. Hlawatsch and R. L. Urbanke, "Bilinear time-frequency representations of signals: The shift-scale invariant class," *IEEE Trans. Signal Process.*, vol. 42, no. 2, pp. 357–366, Feb. 1994.

[40] A. Papandreou, F. Hlawatsch, and G. F. Boudreaux-Bartels, "A unified framework for the Bertrand distribution and the Altes distribution: The new hyperbolic class of quadratic time-frequency distributions," in *Proc. IEEE-SP Int. Symp. Time-Frequency Time-Scale Anal.*, 1992, pp. 27–30.

[41] M. Tajmirriahi and Z. Amini, "Modeling of seizure and seizure-free EEG signals based on stochastic differential equations," *Chaos, Solitons Fractals*, vol. 150, Sep. 2021, Art. no. 111104.

[42] M. Vishwanath, R. M. Owens, and M. J. Irwin, "The computational complexity of time-frequency distributions," in *Proc. IEEE 6th SP Workshop Stat. Signal Array Process.*, 1992, pp. 444–447.

[43] S. Iqbal, B. A. Rizvi, P. P. M. Shanir, Y. U. Khan, and O. Farooq, "Detecting P300 potential for speller BCI," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Apr. 2017, pp. 295–298.

[44] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

Authorized licensed use limited to: Indian Institute of Technology Gandhinagar. Downloaded on January 23,2025 at 12:20:13 UTC from IEEE Xplore. Restrictions apply.