Proceedings of the 2019 5th International Conference on Advances in Electrical Engineering (ICAEE), 26-28 September, Dhaka, Bangladesh

DenseNet Based Speech Imagery EEG Signal Classification using Gramian Angular Field

Md. Monirul Islam[1], Md. Maruf Hossain Shuvo[2]
Department of Electronics and Communication Engineering
Khulna University of Engineering & Technology, Khulna - 9203, Bangladesh
im.monirul755@gmail.com[1], maruf.shuvo1@yahoo.com[2]

***Abstract*— One of the most challenging tasks in the Brain-Computer Interface (BCI) system is to classify the speech imagery electroencephalography (EEG) signals. In this work, we addressed the existing low classification accuracy problem with deep learning and improved beta band selection method. When the subject imagines, uttering a word rather saying it directly, there are changes in electrical stimulation in the brain. These electrical stimulations of the brain are recorded using EEG signal recording device. The recorded EEG data is then processed using the Dual-Tree Complex Wavelet Transform (DTCWT) for beta band selection which is responsible for activity related to imagery. To take advantage of Deep Convolutional Neural Networks (DCNN), we converted the time series EEG data into images. We generated images using two versions of Gramian Angular Field (GAF): Gramian Summation Angular Filed (GASF) and Gramian Difference Angular Field (GADF). Then these images were fed to DenseNet for image classification. DenseNet is an improved version of DCNN that minimizes the vanishing gradient problem. Between two different image generation techniques, GADF has the best average classification accuracy rate of 90.68%. The dataset used in this study named 'The KARA ONE Database' is collected from Computational Linguistics Lab, University of Toronto, Canada.**

***Index Terms*— Brain Computer Interface, Electroencephalogram, Dual Tree Complex Wavelet Transform, Angular Field, DenseNet.**

I. Introduction

In non-invasive BCI system, electrodes are placed on the scalp of head. These electrodes are used to record electrical activity generated by brain to perform different tasks. Recorded signals are then processed according to the needs and desired operations are carried out using it. EEG based BCI system is popular for translating brain's electrical activity into actions via different classification methods [1].

In case of imagery task classification, such as imagery limb movements or imagery words classification, the user never actually performs the task in practice rather only imagines to perform the task. In imagined word classification, the user is asked to imagine to utter the word rather saying it practically. Then corresponding EEG signal is recorded and processed for identifying the word [2].

Different types of approaches have been taken to classify speech imagery EEG signals. These approaches can be divided into two major groups, one is word classification and another is vowel classification. Regularized neural network was used in [3] to classify words. The authors reported average classification accuracy of $75\% \pm 9.6$. They tried to classify 'yes' vs. 'no'. Castanẽda et al. [4] tried a new approach to classify EEG signal: sonification and textification to classify speech imagery EEG signal. Jahangiri et al. [5] classified four words, 'left', 'right', 'back' and 'forward'. But in the experiment, these words were shortened to 'le', 'ry', 'ba' and 'fo' respectively. Their maximum classification accuracy ranged from 72% to 88% by using pseudo-linear LDA classification. Nguyen et al. [6] proposed covariance matrix descriptors and Relevance Vector Machines classifier. They used these methods to classify speech imagery words. A group of short words consisting 'in', 'out', 'up' and another group of long words consisting of 'cooperate' and 'independent' were used. Their method achieved 70% and 95% classification accuracy for three and two words respectively.

The data used in this work is collected from The KARA ONE Database: Phonological Categories in imagined and articulated speech, Computational Linguistics, Department of Computer Science, University of Toronto [13]. The dataset contained EEG recording of imagery speeches of 7 phonemic/syllabic prompts and 4 words. A hybrid dataset was constructed by using all the images from every subjects in order to make the system bias free of any particular subject.

In next section, a brief theoretical background of this work is presented. Materials and methods used in this work are described in details the section after that. Results obtained from this study is analyzed in 'Result Analysis' section and finally the conclusion of this work is drawn.

II. Theoretical Background

EEG rhythms are mainly classified into five different classes: Alpha (α) $(8-13Hz)$, Beta (β) $(13-30Hz)$, Gamma (γ) $(>30Hz)$, Theta (θ) $(4-8Hz)$ and Delta (δ) $(0.5-4Hz)$. For better classification accuracy, it is required to work with right EEG rhythms as different EEG rhythms are related to different physiological activity. Beta rhythm is associated with active thinking, solving problems and attentions. The amplitude of beta rhythm is normally below $30\mu V$. When people in panic state, beta rhythm

978-1-7281-4934-9/19/$31.00 ©2019 IEEE

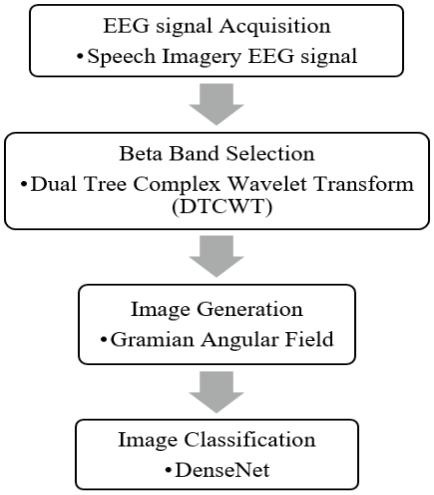


Fig. 1: Proposed work flow for Speech Imagery EEG signal Classification

with higher peak is recorded [7]. Therefore, decomposing EEG signals into sub bands and extracting beta rhythm is important. This task is performed using wavelet based decomposing. For a non stationary signal like EEG, wavelet based signal decomposition method such as Discrete Wavelet Transformation (DWT) is effective because it gives good time-frequency resolution [8]. But DWT has four major drawbacks: oscillations, shift invariance, aliasing and lack of directionality. The solution of these problems is complex wavelet based decomposition. Therefore, to have better time frequency resolution for beta wave extraction and avoiding the limitations of DWT, Dual Tree Complex Wavelet Transformation (DTCWT) [9] was used in this study as DTCWT addresses the limitations of DWT efficiently.

After the signal is decomposed into sub-bands, trivially, time series features are extracted from the signal [4]. But in recent years, Deep Convolutional Neural Networks (DCNN) has revolutionized image classifications [10] with higher classification accuracy compared with traditional methods. To take this advantage, we converted our time series data into images, then these images were used for classification. A classical method to convert time series data into image is time frequency resolution map. In this type of image representation, important information of signal is lost, thus it not suitable for EEG signal classification as information get lost in the process of time domain to frequency domain conversion which ultimately may degrade the total performance of the model. In our approach, it was desired to avoid the loss of information as much as possible. To do so, Gramian Angular Field (GAF) [11] was used for converting time series data into images. After that, Densely Connected Convolutional Networks or DenseNet [12] was used to make classification. In DenseNet, feature maps of all preceding layers are used as input for all subsequent layers. DenseNet has following advantages: reducing vanishing gradient problem, feature reusing and stronger feature propagation. As DenseNet encourages feature reusing, number of parameters are also substantially reduced. In the next section, a detailed description of methods used in this work is presented.

III. Materials and Methods

In this paper, we are proposing image based EEG signal classification using DenseNet and Gramian Angular Field. The work-flow of this work is given in Fig. 1. In the EEG signal acquisition phase, data were recorded for speech imagery actions. Beta rhythm were extracted in the signal processing stage. Dual Tree Complex Wavelet Transform (DTCWT) was used for beta-band selection. Gramian Summation Angular Filed (GASF) and Gramian Difference Angular Field (GADF) were used to generate images from selected sub-bands. Finally classification was made using DenseNet. This simulation work was performed using Python 3.6 and Keras 2.2.4. Keras is an open source high-level neural network API famous for its user friendly nature. Images were generated using python's 'pyts' (version 0.7.5) package.

A. EEG signal acquisition

The KARA ONE Database: Phonological Categories in imagined and articulated speech, Computational Linguistics, Department of Computer Science, University of Toronto, [13] dataset is combined with three modalities namely EEG, face tracking and audio. EEG signals recorded with all 64 channels were used in this study. SynAmps RT amplifier were used for EEG signal recording. The signal was sampled at $1kHz$. This dataset is based on imagined and vocalized phonemic and single-word prompts and again we only used the imagined phonemic and single-word prompts. There were total 7 phonemic / syllabic prompts ('iy', 'uw', 'piy', 'tiy', 'diy', 'm' and 'n') and 4 words ('pat', 'pot', 'knew' and 'gnaw') used in the data recording. The trials were segmented into 4 parts, i.e., rest state, stimulus state, imagined state and speaking state. In the rest state, the participants were told to clear their mind and relax for 5 seconds. In the stimulus state, a word prompt was shown on the computer screen and corresponding auditory utterance was played. Then came the imagined state, where the participants were told to imagine. EEG recording of imagined part was chosen for the study.

B. Beta band selection

Dual Tree Complex Wavelet Transform (DTCWT) [9] is called dual tree because it has two trees, one is complex in nature and another one is real. The name complex also comes from the complex tree. The decomposition and reconstruction process in DTCWT is give in Fig 2. The DTCWT is composed of high pass [for real tree, $h_1(m)$ and for complex tree, $g_1(m)$] and low pass [for real tree, $h_0(m)$ and for complex tree, $g_0(m)$] filters for both real and complex tree. The low pass filter output passes through another low pass and high pass filters, and this process goes on recursively. After passing

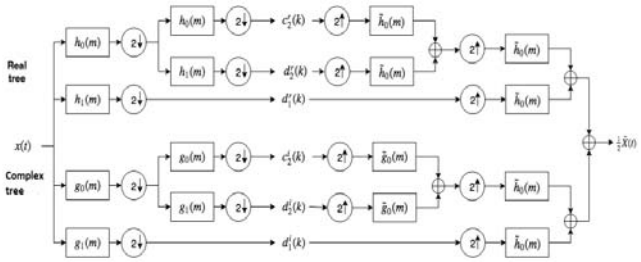


Fig. 2: Decomposition and reconstruction process in DTCWT

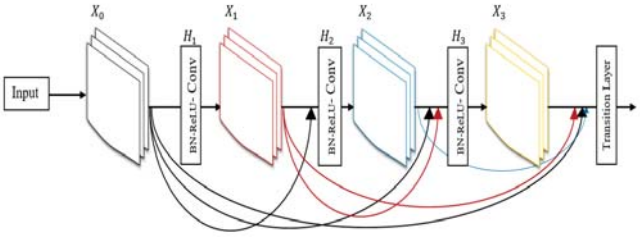


Fig. 3: A three layer Dense block

through every filter a down sampling is used. Whereas in the reconstruction process, up sampling is used. In reconstruction process, the signals from the output of same level filters are recombined.

C. EEG time series signal to image generation

After the beta band has been extracted from the raw EEG signal, now these time series data needs to be converted into images. We have conducted experiment with two different versions of Gramian Angular Field (GAF) for time series to image conversion namely: Gramian Summation Angular Filed (GASF) and Gramian Difference Angular Field (GADF). A brief explanation about these methods are given below: Gramian Angular Field (GAF) is simply mapping the time series data into polar coordinates [11]. GAF has two advantages for converting time series data into image. It preserves temporal dependency and contains temporal correlations. GAF starts with rescaling the time series data in the range of $[-1, 1]$. Let, $X = (x_1, x_2, \ldots, x_i)$ of i real world values be the time series data. Then the rescaled data will be:

$$\tilde{X}_i = \frac{(x_i - max(X) + (x_i + min(X))}{max(X) - min(X)} \tag{1}$$

Where $\tilde{X}_i$ is the rescaled data for every time stamp data i. As the data is rescaled, no information is lost. Then we map the data into polar coordinate using following equation:

$$\psi_i = \begin{cases} arc\ cos\ (\tilde{X}_i) & \text{for } -1 \leq \tilde{X}_i < l \\ r = \frac{i}{N} \end{cases} \tag{2}$$

Here, N is a constant factor that regularize the span of polar coordinates and ψ_i and r are the polar co-ordinate's angle and amplitude of original time series data respectively for time stamp i.

As the data is mapped into polar coordinate system, not in Cartesian coordinate, it has two advantages: firstly, polar coordinates preserves the absolute temporal relations which Cartesian coordinates does not. Secondly, this equation produces a unique map, so in case of inversing the data, it produces a loss less transformation. After the data is being mapped into polar coordinate, now we can exploit the angular perspective. To do this we can use trigonometric sum or difference between two points to obtain temporal relation. This trigonometric sum or difference generates two types of GAF, namely Gramian Summation Angular Filed (GASF) and Gramian Difference Angular Field (GADF).

$$GASF = \begin{bmatrix} cos(\phi_1 + \phi_1) & \cdots & cos(\phi_1 + \phi_i) \\ cos(\phi_2 + \phi_1) & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ cos(\phi_i + \phi_1) & \cdots & cos(\phi_i + \phi_i) \end{bmatrix} \tag{3}$$

and

$$GADF = \begin{bmatrix} sin(\phi_1 - \phi_1) & \cdots & sin(\phi_1 - \phi_i) \\ sin(\phi_2 - \phi_1) & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ sin(\phi_i - \phi_1) & \cdots & sin(\phi_i - \phi_i) \end{bmatrix} \tag{4}$$

From equation (3) and (4), GAF matrices are trigonometric representation of previously time series data that had been converted into polar coordinate. Each elements of GAF matrix is a sinusoidal output of either summation or difference of two time stamp data. In next section, we are going to discuss DenseNet for classifying these images.

D. Image classification with DenseNet

Deep Convolutional Neural Network (DCNN) is the combined approach of convolutional operations and neural networks. DCNN has revolutionized the visual object classification tasks. But in DCNN, adding more layers have created some problems such as: with deeper networks, information about the input vanishes or washes out as it goes through many linear and nonlinear operations. Also deeper networks takes longer time to train the networks. Many authors has tried to address this issue, among them ResNet [14] is prominent one. ResNet allows better information and gradient flow by adding the input to the output directly. In this way a shortcut path is created from earlier layer to current layers. But directly summing the features means loss of valuable information. This problem is more accurately addressed in Densely Connected Convolutional Network or DenseNet [12]. Fig. 3 shows a dense block in DenseNet. From the Fig. 3 it is clear that, X_i-th layer has i inputs. And in between each layer there are nonlinear transformation H_i. Batch Normalization (BN), Rectified Linear Unit (ReLU), Convolutions (Conv) are possible functions of H_i. Input output

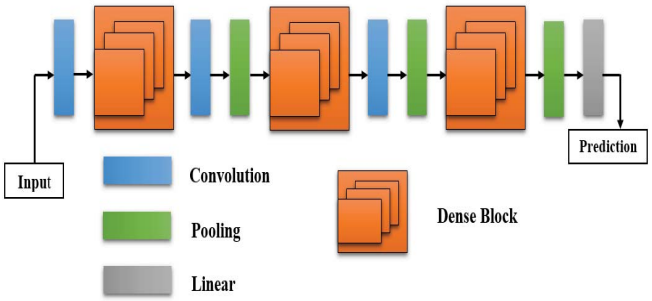


Fig. 4: A DenseNet consisting three dense blocks

relationship of any CNN is $X_i = H_i(X_{(i-1)})$. As mentioned earlier, this relationship changes with concatenation. Therefore $X_i = H_i([X_0, X_1, \cdots, X_{i-1}])$. Where, $[X_0, X_1, \cdots, X_{i-1}]$ refers to concatenation of features maps which were produced in previous stages. But this concatenation process in not viable as the shape changes after every layer. To solve this problem, whole network is divided into small blocks and to match the dimensions, down sampling is used. This property is illustrated in Fig. 4. To keep the dimension consistent, transition layer in introduced between two dense blocks. In the dense block, feature map's dimension remains constant but total number of filters in between them change. Convolution and pooling operation are performed in transition layer. DenseNet introduces a new parameter named growth rate, k. If K_0 is the number of channel in the input, if the i-th layer has $K_0 + k(l-1)$ input feature maps then H_i has k feature maps which is termed as growth rate. Growth rate, k plays important role to obtain state of the art classification result. In our work, $k = 32$ was used to avoid the unnecessary deep network. Input image dimension of the network was $(224, 224, 3)$ i.e., channel last and the output was prediction of word label. After image generation, a hybrid dataset was formed by combining all the subjects. This dataset was then used for classification. For classification, 'one label' vs 'rest' was used, e.g., 'piy' vs. rest of the ten labels. In the next section, results of this study is discussed.

IV. Result Analysis

In the KARA ONE database, MM05, MM08, MM09, MM10, MM11, MM12, MM15, MM16, MM18, MM19, MM20, MM21 were used. As mentioned earlier in the proposed method, beta sub-band was selected using DTCWT from raw EEG signal. This decomposition is shown in Fig. 5. Change of electrical activity due to imagined speech is illustrated in Fig. 5(a). Fig. 5(b) and 5(c) show the output of imagery and real tree respectively of level 2 decomposition. Due to down sampling by a factor of 2, sampling points after second level is reduced to one fourth of raw EEG signal. To use both real and imagery components of DTCWT, absolute value is used (Fig. 5(d)). Two classes of image were generated, namely GADF and GASF. Sample image for these two methods are given in Fig. 6. For every single trial, one image is generated. That means, as subject MM08 has 132 trials, for MM08 subject, 132 images generated for GASF and GADF each. For each label, classification accuracy of images generated by GASF and GADF is listed in Table I. The average classification accuracy of all 11 labels for GASF images and GADF images has been found as 90.54% and 90.68% respectively. From Table II, it is clear that our proposed method has worked well on KARA ONE dataset. This is due to the fact that using DTCWT for beta band selection and using DenseNet as classifier. DTCWT helps to extract beta band from raw EEG signal and leaving the limitations that are generally introduced by DWT. Also, DenseNet allows us to build a deep neural network with proper feature reuse and stronger information flow, which allows better classification accuracy comparing with other architectures.

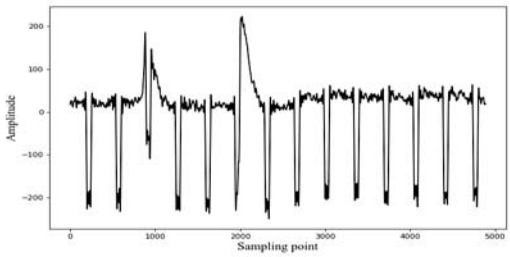
TABLE I: Classification Accuracy of GADF and GASF Images

Labels	**GADF Image**	**GASF Image**
	Classification accuracy (%)	
/diy/	90.32	90.00
/tiy/	90.69	90.92
/m/	90.31	90.62
/uw/	91.13	90.73
pat	91.31	90.45
knew	90.81	90.64
gnaw	91.81	90.66
/n/	90.54	90.75
/iy/	90.94	90.77
pot	89.88	90.48
/piy/	90.75	91.01
Average	90.68	90.54

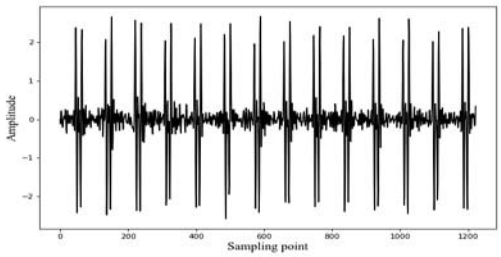
TABLE II: Comparative Study Between Different Works Performed on KARA ONE Dataset

Authors	**Methods**	**Accuracy (%)**
[13]	SVM-quad classifier	79.16
[15]	NES	74.00
This work	DenseNet	90.68

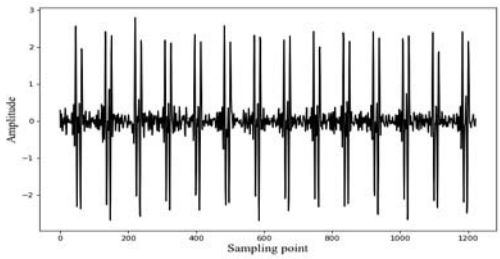
Table III shows the performance analysis of different types of approaches. These approaches include state of the art SVM algorithm to recent DCNN algorithms because, although some methods has achieved higher accuracy than our proposed method but their work is limited by the number of labels. For example, authors in [16] achieved perfect classification accuracy. But they only used the vowels. On the contrary, there were 11 different labels were used in our study. Also,



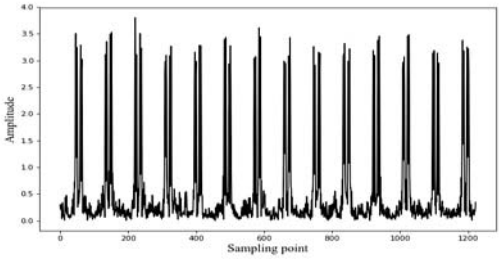
(a) Recorded EEG signal of subject MM15



(b) Extracted imagery part of beta band using DTCWT

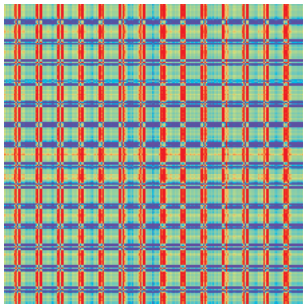


(c) Extracted real part of beta band using DTCWT

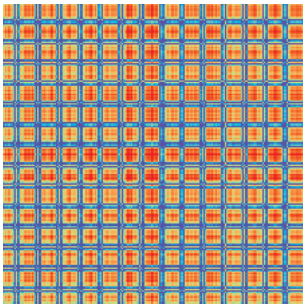


(d) Absolute value of extracted beta band

Fig. 5: Extracted beta band from recored EEG signal using DTCWT



(a) GADF plot



(b) GASF plot

Fig. 6: Converted image from time series data for Subject MM15 Trial 16

TABLE III: Comparison Between Proposed Method and Other Works

Authors	**Classification Models**	**Labels**	**Accuracy (%)**
[3]	Regularized neural network	yes vs. no	75
[4]	Random Forest	6 Spanish words	83.34
[5]	pseudo-linear LDA	left, right, back, forward	88
[6]	Relevance Vector Machines	in, out, up	75
[16]	Deep Belief Network	a, e, i, o, u	100
This work	DenseNet	4 words and 7 phonemic prompts	90.68

Table III indicates that, for better performance, deep convolution neural network is the only viable option. Because, all the proposed methods in Table III using neural network performed comparatively well. Therefore, considering with the introduction of DCNN in the field of Speech Imagery EEG signal classification for BCI system, performance of these BCI system has increased. Further extending these works with more diverse dataset will unlock practical BCI system.

V. Conclusion

The goal of this work was to improve speech imagery EEG signal classifications. The existing problems of Discrete Wavelet Transform were solved by using the Dual Tree Complex Waveform Transform. Furthermore, to improve the classification accuracy, DenseNet was chosen. Also the image generation techniques were studied to find the best method to convert time series data into images. Also comparing with different datasets, our work's performance was satisfactory while considering the number of labels that are used in the training set. As in a regular humans daily conversation, thousands of words are used, this work can be further extended to a larger word collections which one day will pave the way for a Speech Imagery BCI system.

References

[1] K. J. Panoulas, L. J. Hadjileontiadis, and S. M. Panas, "Brain-computer interface (bci): Types, processing perspectives and applications," in

Multimedia Services in Intelligent Environments, pp. 299–321, Springer, 2010.

[2] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, no. 7753, p. 493, 2019.

[3] A. R. Sereshkeh, R. Trott, A. Bricout, and T. Chau, "Eeg classification of covert speech using regularized neural networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 12, pp. 2292–2300, 2017.

[4] E. F. González-Castañeda, A. A. Torres-García, C. A. Reyes-García, and L. Villaseñor-Pineda, "Sonification and textification: Proposing methods for classifying unspoken words from eeg signals," *Biomedical Signal Processing and Control*, vol. 37, pp. 82–91, 2017.

[5] A. Jahangiri and F. Sepulveda, "The contribution of different frequency bands in class separability of covert speech tasks for bcis," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2093–2096, IEEE, 2017.

[6] C. H. Nguyen, G. K. Karavas, and P. Artemiadis, "Inferring imagined speech using eeg signals: a new approach using riemannian manifold features," *Journal of neural engineering*, vol. 15, no. 1, p. 016002, 2017.

[7] S. Sanei, *Adaptive processing of brain signals*. John Wiley & Sons, 2013.

[8] M. Madhusmita, B. Mousumi, P. D. Narayan, and M. S. Kumar, "A novel method for epileptic eeg classification using dwt, mga, and anfis: A real time application to cardiac patients with epilepsy," in *Cognitive Informatics and Soft Computing*, pp. 525–534, Springer, 2019.

[9] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE signal processing magazine*, vol. 22, no. 6, pp. 123–151, 2005.

[10] M. Shahbazi and H. Aghajan, "A generalizable model for seizure prediction based on deep learning using cnn-lstm architecture," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 469–473, IEEE, 2018.

[11] Z. Wang and T. Oates, "Encoding time series as images for visual inspection and classification using tiled convolutional neural networks," in *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[12] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[13] S. Zhao and F. Rudzicz, "Classifying phonological categories in imagined and articulated speech," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 992–996, IEEE, 2015.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[15] P. Sun and J. Qin, "Neural networks based eeg-speech models," *arXiv preprint arXiv:1612.05369*, 2016.

[16] R. A. Sree and A. Kavitha, "Vowel classification from imagined speech using sub-band eeg frequencies and deep belief networks," in *2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN)*, pp. 1–4, IEEE, 2017.