European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol.12, No.2, 2022

EEG based Schizophrenia Detection using SPWVD-ViT Model

Mesut Şeker[1*], Mehmet Siraç Özerdem[2]

[1*]Dicle University, Electrical-Electronics Engineering Department, Diyarbakir, Turkey. (e-mail: mesut.seker@dicle.edu.tr).
[2] Dicle University, Electrical-Electronics Engineering Department, Diyarbakir, Turkey. (e-mail: sozerdem@dicle.edu.tr).

ARTICLE INFO

Received: Oct., 20. 2022
Revised: Nov., 30. 2022
Accepted: Nov., 30. 2022

Keywords:
EEG
Schizophrenia
Neurological Disease
Vision Transformer
Diagnosis
Detection
Time-frequency image

Corresponding author: *Mesut Şeker*

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI: https://doi.org/10.36222/ejt.1192140

ABSTRACT

Schizophrenia is a typical neurological disease that affects mental state, and daily behaviours of patients. Combining image generation techniques with effective machine learning algorithms may accelerate treatment process. Moreover, possible early alert systems prevent diseases from reaching out crucial phase. The purpose of current study is to develop an automated EEG based schizophrenia detection with the Vision Transformer (ViT) model using Smoothed Pseudo Wigner Ville Distribution (SPWVD) time-frequency (TF) input images. EEG recordings from 35 schizophrenia (sch) patients and 35 healthy controls (hc) are analysed. We have used 5-fold cross validation for evaluation and testing of the method. Classification task is carried out via. subject-independent and subject-dependent method. We achieved overall accuracy of 87% for subject-independent and 100% for subject-dependent approach for binary classification. While ViT has been extensively used in Natural Language Processing (NLP) field, dividing input images within a sequence of embedded image patches via. transformer encoder is a practical way for medical image learning and developing diagnostic tools. SPWVD-ViT model is recommended as a disease detection tool not only for schizophrenia but other neurological symptoms.

1. INTRODUCTION

Neurological disorders may affect human's thinking ability and result general behavioral disorders. Schizophrenia is such neurological disease in same manner [1]. According to Word Health Organization (WHO), 24 million patients sufferer from schizophrenia has been reported in [2]. Within the early diagnosis of schizophrenia and the relevant treatment methods, the disease will be prevented from reaching out crucial phase, and the treatment process will be accelerated with the supply of required medications [3].

Most of mental disorders has been investigated by image and signal processing techniques. Electroencephalography (EEG) has been considered as a popular screening tool in Brain Computer Interfaces (BCI), neuroscience, engineering, and rehabilitation applications [4]. EEG provides muti-channel setup and higher time resolution , and has cheap, easy and practical aspect as a neuro-screening tool [5]. This method provides significant information process belongs to brain dynamics and has remained its prevalence for detection of neurological diseases in recent years.

Classification-based studies used in the diagnosis of schizophrenia show advanced feature engineering, but it is still observed that the desired level of qualified features haven't been investigated yet. Some of the recent studies based on disease detection techniques and diagnosis methods of schizophrenia can be summarized as follows: Kim et al. (2015) calculated spectral values of EEG sub-bands using Fast Fourier Transform (FFT) from 90 schizophrenia (SCH) patients with equal size of Healthy Controls (HC) and obtained 62.2% accuracy within Receiver Operating Characteristic (ROC) by delta sub-band frequency [5]. In another study including 25 healthy and 25 patients, time-frequency conversion was performed with EEG recordings and the best 5 electrodes for classification were determined [6]. In the related study, the highest performance with 93.9% accuracy was obtained from the F2 channel. Johannesen et al. (2016) extracted a total of 60 features from 40 schizophrenic and 12 healthy participants in their study and performed classification task with Multilayer Perceptron (MLP) and Support Vector Machine (SVM) [7]. Using theta and alpha frequencies in the frontal region, 87% accuracy was obtained for distinguishing patients from controls. In a complexity and entropy features based study [8], a total of 14 non-linear features has been proposed and highest accuracy of 92.91% is achieved using Radial Function based Support Vectors Machine (RFB-SVM).

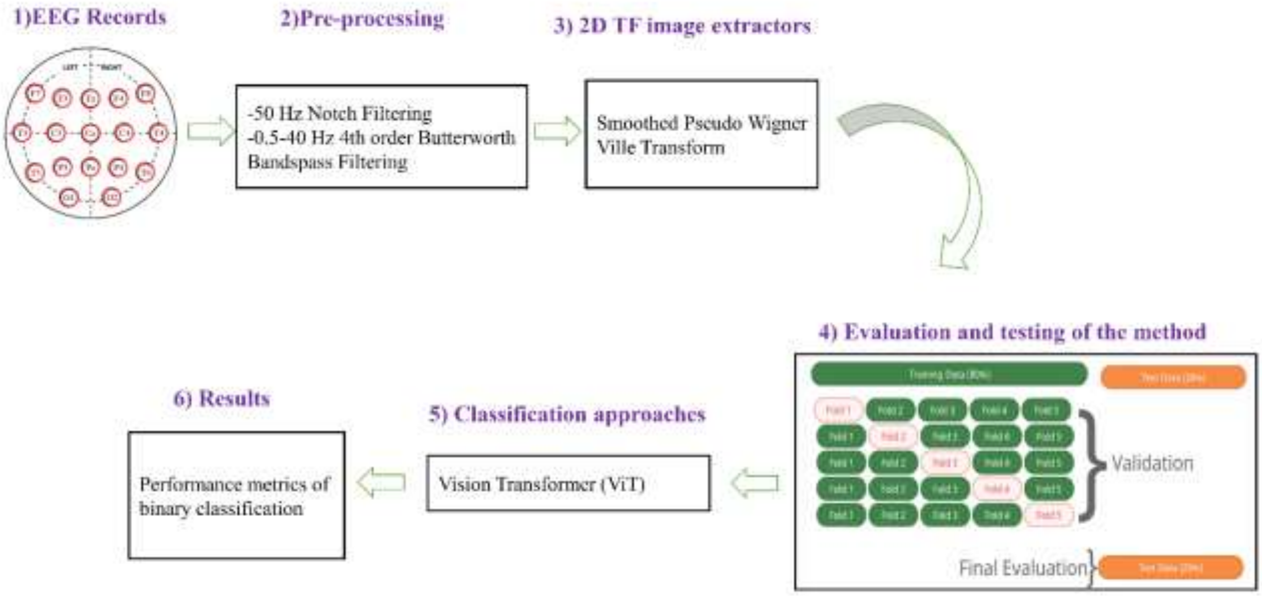


Figure 1. The followed methods in proposed study

Resting state EEG based deep learning algorithms for detection of schizophrenia have been yielded growing attention using in studies. Oh et al. [9] utilized from 14 sch patients and 14 controls with 19 channel EEG using time-domain representation. They proposed 11 layered Convolutional Neural Network (CNN) with SoftMax classifier. 10-fold cross validation is used for train(72%), validation(18%), and test (%10) split, and 81.26% accuracy is obtained by subject-independent classifications task. Phang et al. [10] acquired EEG from 39 HC and 45 SCH, proposed fusion of time, frequency, and brain topological connectivity as input data. RNN, CNN architectures are used with SoftMax classifier. 5-fold cross validation is used for train (60%), validation (20%), and test (20%) sets, and 92.87% accuracy is obtained. Calhas et al. [11] used time-frequency EEG representation using Siamese Neural Network with SVM, Random Forest (RF), Extreme Gradients Boosting (XGBoost), Naïve Bayes (NB), and k-Nearest Neighbors (k-NN) classifiers and leave-one-out validation. Highest accuracy of 83% is obtained from the experiment. Naira et. al. [12] proposed a CNN architecture with softmax classifier using matrices of correlation between EEG channels from 39 HC and 45 SCH participants. They obtained 90% of accuracy based on subject-dependent classification within 70%-30% train-test split evaluation method. A separate summary of previous studies for EEG based schizophrenia detection is given in Table I.

In current study, a time-frequency representation called Smoothed Pseudo Wigner Ville Distribution (SPWVD) is applied to EEG records of schizophrenic and healthy participants. Both subject dependent and independent evaluation are applied during train, validation, and test split. We propose a Vision Transformer (ViT) model to discriminate schizophrenic EEG from controls. Performance metrics of binary classification task will be given. The overall steps followed in current study is given in Fig. 1.

The rest of this paper is organized as follows: Section 2 explains experimental dataset, pre-processing steps, describes generated SPWVD time-frequency images, and architectural mechanism of ViT. Section 3 explains performance metrics of ViT using SPWVD images of SCH and HC. Finally, proposed method for detection of schizophrenic EEG records via. HC is concluded in Section 4.

2. DATASET, PRE-PROCESSING STAGE, TIME-FREQUENCY INPUTS, AND CLASSIFICATION TASK

2.1. EEG Data Acquisition

EEG datasets are acquired from 35 schizophrenia patients and 35 healthy controls aged between 11-13 years. All participants are male. Resting state EEGs are recorded using 16 channel 10-20 system of electrode placement during eyes closed. EEG channels are F7, F3, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1 and O2. Electrode replacement is shown in Fig2. Frequency range of records is 0.5-45 Hz, and electrode impedance is below 10 kΩ. Sampling frequency f_s is 128 Hz and length of each participant's record is 1 minute. Dataset is publicly available at http://brain.bio.msu.ru/eeg_schizophrenia.htm.

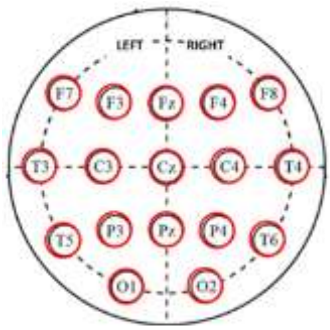


Figure 2. 10-20 electrode replacement of 16 EEG channels

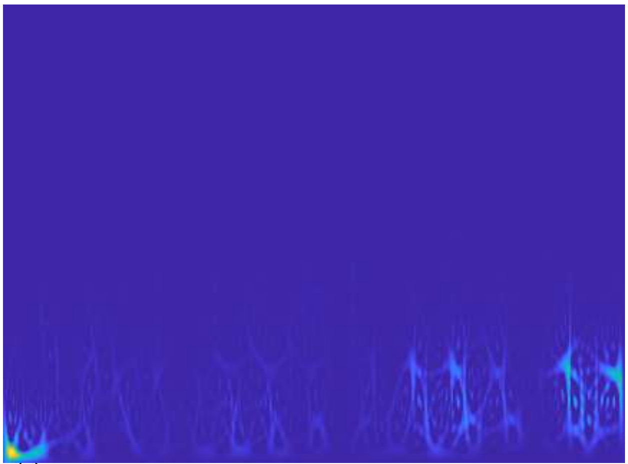
2.2. Pre-processing Step

In order to process with specific band-range and obtain clean EEG, we applied 4th order Butterworh band-pass filtering due to its linear response [13]. Moreover, a 50 Hz band-rejected Notch filter is applied to remove power line inference [14]. 60s length of signal is divided into 1s epoch, and TF image inputs are generated from each epoch.

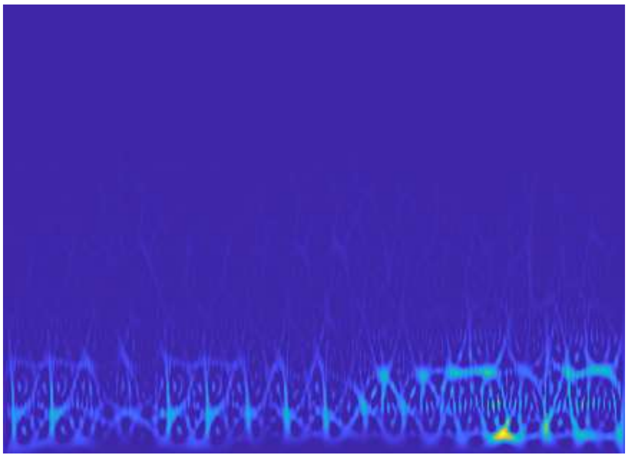
2.3. Time-Frequency (TF) Inputs

We converted 1D time series EEG signals to 2D time-frequency domain. There are many algorithms such as continuous wavelet transform (CWT) based scalogram [15], Short Time Fourier Transform based spectrogram [16], Hilbert Spectrum [17], Fourier and Wavelet based Synchro-squeezing transform [18]. According to our knowledge, there is only single study that achieves best results (acc. 93.36%) with Smoothed Pseudo Wigner Ville Distribution (SPWVD) with CNN for automated detection of schizophrenia using EEG [19].

The EEG records obtained from 16 channels of each participant is divided into 1s epochs. We obtain one-dimensional vector with combining epochs belongs to each channel. Dimension of vector within combined channels is 128x16. We obtained 60 SPWVD images from each participant. Dimension of each generated TF images is 434x343x3. We totally have 2100 images from SCH and 2100 images from HC. We resized the images as 224x224x3 with inter_area interpolation provided by OpenCV. Sample images are given in Fig. 3. All images are generated in MATLAB R2021a environment.



(a)



(b)

Figure 3. Some sample images using SPWVD for (a) HC and (b) SCH with original size of 434x343x3

SPWVD is an effective conversion technique that includes content of time-frequency localization of signal energy. This TF representation methods evaluates higher harmonics and time-domain localization. It reduces cross-term inference using TF windows. Formula for calculation of SPWVD can be written as follows:

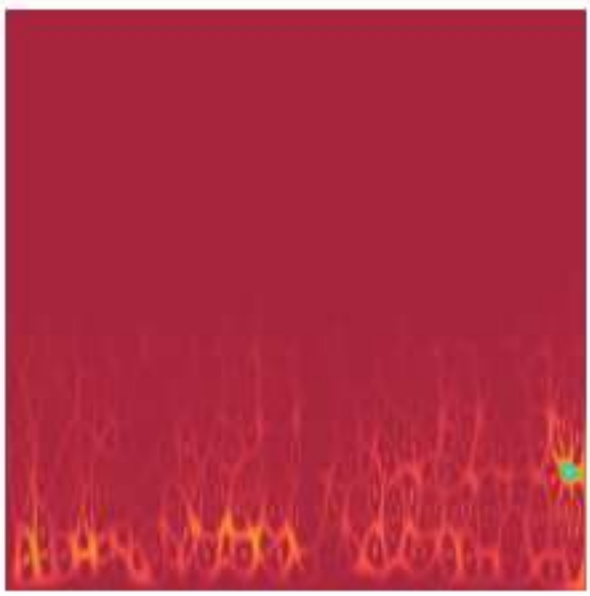
$$\mathrm{SPWVD}(t,f)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v\left(\frac{\tau}{2}\right)v^{*}\left(-\frac{\tau}{2}\right)u(t-t')\ldots \times z\left(t'+\frac{\tau}{2}\right)z^{*}\left(t'-\frac{\tau}{2}\right)\exp^{-j2\pi f t}dt'd\tau$$

,where $v(t)$ is cross-term reducing in time and $u(t)$ is cross-term reducing in frequency domains. In order to adjust the frequency and temporal resolution, length of the time and frequency windows are set individually. The superiority of SPWVD can be explained as follows: STFT assumes a signal to be stationary ovel all windows, and CWT generates cross-term inference in frequency domain. There is a tradeoff between spectrogram and scalogram due to TF localization and TF cross term. SPWVD overcomes this tradeoff and serves practical solutions for TF representation [19].

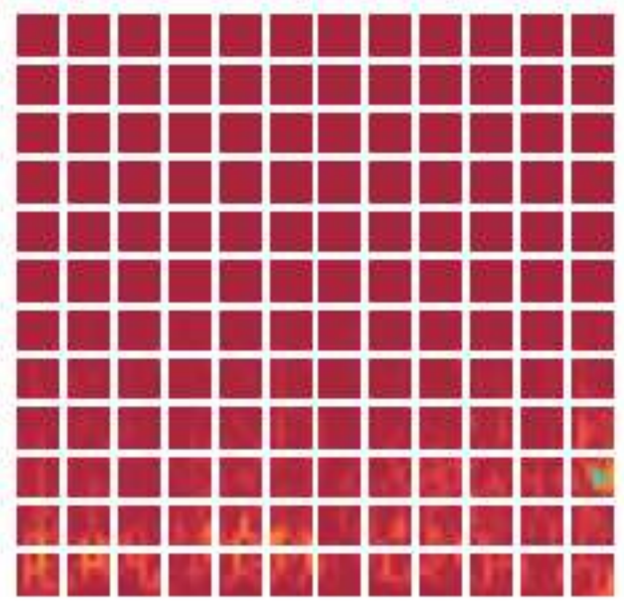
2.4. Vision Transformer (ViT) Model

Vaswani et al. added another dimension to previous attention models mainly rely on recurrence and convolutions. They have used proposed model to a specific approach of natural language processing [20]. From the inspiration of this success, Dosovitskiy et al. attempted to apply the standard Transformer model to images for image classification application [21].

To convert an input image as a structure in a sense of having a sequence of words, we divide input image into smaller N number of 2-dimensional patches via. $N = HW/P^2$, where H: height of image, W: width of image and P: resolution pixel. Each image patch is flattened into a vector length of P^2xC, where C is number of channels. A sample input of before/after dividing patches is given in Fig. 4.



(a)



(b)

Figure 4. (a) A sample input image before dividing patches (b) divided patches in input images

We resized (224,224,3) input images to (72x72x3). Size of the patches to be extracted from input images is 6. Number of patches per images is calculates as 144 (12*12). Other hyperparameters belongs to ViT are as follows: learning rate= 0.001, weight decay= 0.0001, batch size=32, number of epochs= 500, projection dimension= 64, number of heads= 4, projection dimension in transformer unit=2, size of transformer layer=8, size of the head layer of the final classifier Multi-Layer Perceptron (MLP)= [2048,1024], and dropout rate=0.5. We have normalized all image data in range of [0-1] to avoid high range of pixels and improve convergence speed.

We have also used data augmentation to increase number of images with following operation: random flip with "horizontal", random rotation with factor=0.2, and random zoom with height and width factor= 0.2. The proposed architecture of ViT can be seen in Fig. 5. ViT employs the encoder feature of the original Transformer design. The encoder receives the sequence of embedded image patches as input. Values of the learnable class embeddings are received by the classification head that is attached to the output to produce a classification output. This implementation is carried out by a MLP with one hidden layer and Gaussian Error Linear Unit (GELU) non-linearity. We developed the code given publicly available at https://keras.io/examples/vision/image_classification_with_vision_transformer/ and classification task is carried out as subject dependent and subject independent approaches in a workstation with AMD Ryzen 7 3700X 8-Core Processor 3.60 GHz, 32 GB RAM, and NVIDIA GeForce 3060 with 12 GB VRAM using TensorFlow library on Spyder (Python 3.9) environment.

3. PRACTICAL RESULTS

We first divided TF images of SCH and HC class as 80% train and 20% test set. Stratified 5-fold cross validation is applied to the train set to achieve validation set. Classification task is divided 2 sub-sections: subject dependent and subject independent classification. In subject-dependent classification task, we randomly shuffled TF images to the train, validation, and test set. In subject independent classification task, TF images of each participant are only in train, validation, or test set. Accuracy and loss values in each epoch is drawn to prove there is no overfitting. Accuracy needs to be increased and loss may be decreased along 500 epochs. Confusion matrices are given to indicate true and predicted label for each class. Finally, performance metrics of accuracy, precision, recall and f1-score are given. The procedure to calculate related metrics are given in Table I. In an ideal classifier, FP and FN should be zero. Moreover, precision and recall values needs to be one. F1-score is a metric that takes precision and recall into account and show trustworthy results. Calculation of f1-score is given as follows:

$$F1 - core = 2 * \frac{Precision*Recall}{Precision+Recall} \qquad (1)$$

3.1. Subject-Dependent Classification

Performance metrics after subject-dependent classification task is given in Fig 4. (a). 373 TF images out of 420 are classified as sch, and 354 TF images out of 420 are recognized as hc. Discrimination of sch patients TF patterns is greater than discrimination of hc. Overall accuracy is achieved as 87%.

3.2. Subject-Independent Classification

Performance metrics after subject-dependent classification task is given in Fig 4. (b). All sch patients are correctly classified, and only one TF images of hc is misclassified as schizophrenia. Overall accuracy is obtained approximately as 100%.

Considering the studies in literature, researchers have preferred spectral features, non-linear measurements, functional connectivity, correlation between EEG channels, and TF EEG conversion techniques. ROC analysis, conventional classifiers (i.e. k-NN, SVM), CNN with softmax, and hybrid architectures (i.e. feature extraction with deep neural networks, and classification with conventional classifiers) are considered as machine learning methods. There are many methodological differences, namely datasets (lengths of records, number of participants, demographics of participants, etc), pre-processing approaches, length of EEG epochs, different evaluation methods, extracted features and classification architectures between current work and other studies. Therefore, it is not easy to make a direct comparison and evaluation. On the other hand, none of previous studies have included ViT for classification stage, and current study has both subject-dependent, and independent classification task. We believe that it is more convenient to validate a network with a previously unseen test set. Even if all extracted TF images from successive epochs for a participant are not same due to non-linear behavior of EEG , epochs from same participant may include some common patterns, and this will result with overfitting in performance. If the test set has totally separate patterns considering train set, or train set doesn't include enough diversity to enable classifiers to capture common pattern with test set, lower accuracy in test set will be obtained. This trade-off can be taken into consideration during evaluation and testing of method. In a nutshell, 87% of accuracy for subject-independent method still provides sufficient diagnostic rate for clinical purposes.

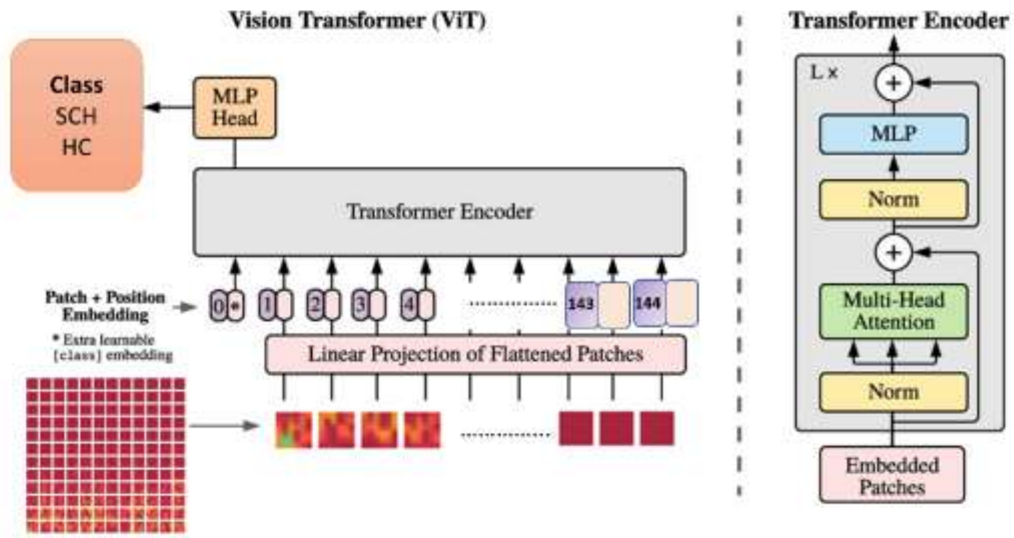


Figure 3. The proposed architecture of Vision Transformer (ViT) Model

TABLE I
CALCULATION OF PERFORMANCE METRICS FROM CONFUSION MATRICES

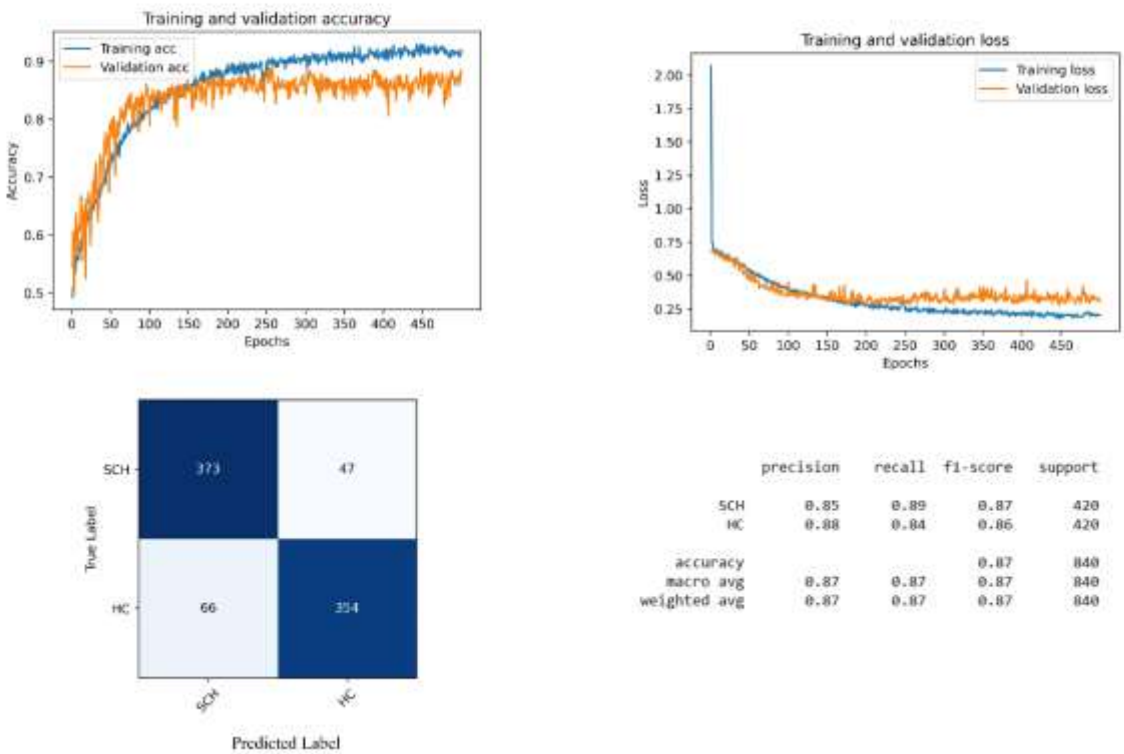
		Predicted Label		
		Positive	**Negative**	
True Label	**Positive**	TP	FN	Sensitivity $\frac{TP}{TP+FN}$
	Negative	FP	TN	Specificity $\frac{TN}{TN+FP}$
		Precision $\frac{TP}{TP+FP}$	Recall $\frac{TP}{TP+FN}$	Accuracy $\frac{TP+TN}{TP+FP+FN+TN}$

TABLE II

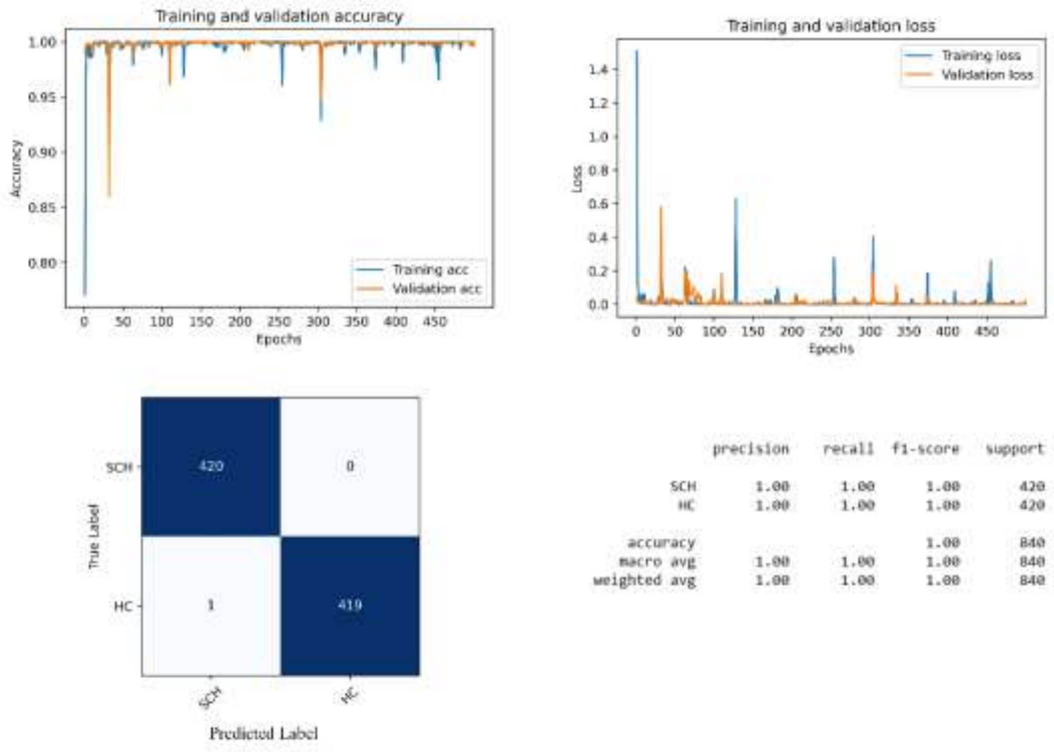
SUMMARY OF STUDIES FOR EEG BASED AUTOMATED DETECTION OF SCHIZOPHRENIA

Authors	**Input Data**	**Architecture/ Classifier**	**Accuracy (%)**
Kim et al. [5] +	FFT based spectral values of 5 EEG sub-bands	ROC Analysis	62.20
Dvey- Aharon et al. [6] +	Time-frequency conversion	k-NN	93.9
Johannesen et al. [7] +	Theta1, theta 2, alpha, beta, And gamma frequency features	SVM	87.0
V. Jahmunah et al. [8] +	Non-linear features, t-test for feature selection	RBF-SVM	92.91
Oh et al. [9]*	Time-domain representation	CNN with softmax	81.26
Phang et al. [10] +	TF and topological connectivity	RNN, CNN with softmax	92.87
Calhas et al. [11] +	TF EEG representation	Siamese Neural Network With SVM, RF, XGBoost, NB, k-NN	83
Nairan et al. [12]**	Matrices of correlation between channels	CNN with softmax	90
Proposed work	Smoothed Pseudo Wigner Ville Transform (SPWVD)	Vision Transformer (ViT)	87 * 100**

*:subject-independent evaluation, **:subject-dependent evaluation, +:not attended



(a)



	precision	recall	f1-score	support
SCH	1.00	1.00	1.00	420
HC	1.00	1.00	1.00	420
accuracy			1.00	840
macro avg	1.00	1.00	1.00	840
weighted avg	1.00	1.00	1.00	840

(b)

Figure 4. Training and validation accuracy/loss, confusion matrices and performance metrics for **(a)** Subject-Independent and **(b)** Subject-Dependent Classification Approach

2. SUMMARY AND FUTURE WORK

In current study, ViT architecture is proposed for EEG based schizophrenia detection for image classification application. Analyzing and recognizing images within a sequence of embedded image patches via. transformer encoder gives rise to medical image classification for disease detection. We extracted 2D SPWVD TF images from 1D schizophrenic and healthy EEG time series. SPWVD is an efficient alternative TF image extractor over widely used spectrogram and scalogram images by providing solutions for cross-term inference and TF localization. Unlike previous studies for schizophrenia detection with resting state EEG, we combined subject dependent, and independent classification methods in SPWVD-ViT model. We obtained 87% overall accuracy for subject-independent, and 100% overall accuracy for subject-dependent approach for automated schizophrenia detection.

It would be also interesting to add more participants and increase number of EEG records to validate current results. Including sub-types or different stages of schizophrenia in dataset and performing a multi-class method will also be another potential work for future. We can even apply same disease detection method to other patients with epilepsy, dementia, or Parkinson to put forward robustness of developed tool. ViT architectures need more powerful hardware basement to reach out better performance. We aim to increase the rate of subject-independent schizophrenia detection by using high performance resources. As an alternative work, image sets generated from an advanced method are fed into a CNN, and then feature maps of CNN are passed onto the Transformer encoder in future investigations. Finally, performance of ViT model with pre-trained or trained from scratch deep learning pipelines is also targeted future study.

ACKNOWLEDGEMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. A sincere appropriation to Julius Bamwenda for his diligent proofreading of the manuscript.

REFERENCES

[1] V. Rajinikanth, S. C. Satapathy, S. L. Fernandes, and S. Nachiappan, “Entropy based segmentation of tumor from brain MR images – a study with teaching learning based optimization,” Pattern Recognit. Lett., vol. 94, pp. 87–95, 2017.

[2] “Schizophrenia.” [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/schizophrenia. [Accessed: 10-Jan-2022].

[3] Z. Wang and T. Oates, “Imaging time-series to improve classification and imputation,” IJCAI Int. Jt. Conf. Artif. Intell., vol. 2015-Janua, no. Ijcai, pp. 3939–3945, 2015.

[4] M. Seker and M. S. Ozerdem, “EEG Coherence as a Neuro-marker for Diagnosis of Schizophrenia,” in 2020 28th Signal Processing and Communications Applications Conference, SIU 2020 - Proceedings, 2020.

[5] J. W. Kim, Y. S. Lee, D. H. Han, K. J. Min, J. Lee, and K. Lee, “Diagnostic utility of quantitative EEG in un-medicated schizophrenia,” Neurosci. Lett., vol. 589, pp. 126–131, 2015.

[6] Z. Dvey-Aharon, N. Fogelson, A. Peled, and N. Intrator, “Schizophrenia detection and classification by advanced analysis of EEG recordings using a single electrode approach,” PLoS One, vol. 10, no. 4, pp. 1–12, 2015.

[7] J. K. Johannesen, J. Bi, R. Jiang, J. G. Kenney, and C.-M. A. Chen, “Machine learning identification of EEG features predicting working memory performance in schizophrenia and healthy adults,” Neuropsychiatr. Electrophysiol., vol. 2, no. 1, pp. 1–21, 2016.

[8] V. Jahmunah et al., “Automated detection of schizophrenia using nonlinear signal processing methods,” Artif. Intell. Med., vol. 100, no. June, p. 101698, 2019.

[9] S. L. Oh, J. Vicnesh, E. J. Ciaccio, R. Yuvaraj, and U. R. Acharya, “Deep convolutional neural network model for automated diagnosis of Schizophrenia using EEG signals,” Appl. Sci., vol. 9, no. 14, 2019.

 ISSN 2536-5010 | **e-ISSN** 2536-5134

[10] C. R. Phang, F. Noman, H. Hussain, C. M. Ting, and H. Ombao, “A Multi-Domain Connectome Convolutional Neural Network for Identifying Schizophrenia from EEG Connectivity Patterns,” IEEE J. Biomed. Heal. Informatics, vol. 24, no. 5, pp. 1333–1343, 2020.

[11] J. Tudela, M. Martínez, R. Valdivia, J. Romo, M. Portillo, and R. Rangel, “On the use of pairwise distance learning for brain Signal classification with limited observations,” Nature, vol. 388. pp. 539–547, 2010.

[12] C. A. T. Naira and C. J. L. Del Alamo, “Classification of people who suffer schizophrenia and healthy people by EEG signals using deep learning,” Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 10, pp. 511–516, 2019.

[13] S. S. Daud and R. Sudirman, “Butterworth Bandpass and Stationary Wavelet Transform Filter Comparison for Electroencephalography Signal,” Proc. - Int. Conf. Intell. Syst. Model. Simulation, ISMS, vol. 2015-Octob, pp. 123–126, 2015.

[14] X. Jiang, G. Bin Bian, and Z. Tian, “Removal of artifacts from EEG signals: A review,” Sensors (Switzerland), vol. 19, no. 5, pp. 1–18, 2019.

[15] Z. Aslan and M. Akin, “A deep learning approach in automated detection of schizophrenia using scalogram images of EEG signals,” Physical and Engineering Sciences in Medicine, vol. 45, no. 1. pp. 83–96, 2022.

[16] Z. Aslan and M. Akin, “Automatic detection of schizophrenia by applying deep learning over spectrogram images of EEG signals,” Trait. du Signal, vol. 37, no. 2, pp. 235–244, 2020.

[17] A. Zülfikar and A. Mehmet, “Empirical mode decomposition and convolutional neural network-based approach for diagnosing psychotic disorders from eeg signals,” Appl. Intell., vol. 52, no. 11, pp. 12103–12115, 2022.

[18] K. Jindal, R. Upadhyay, P. K. Padhy, and L. Longo, “6 - Bi-LSTM-deep CNN for schizophrenia detection using MSST-spectral images of EEG signals,” in Artificial Intelligence-Based Brain-Computer Interface, V. Bajaj and G. R. Sinha, Eds. Academic Press, 2022, pp. 145–162.

[19] S. K. Khare, G. S. Member, V. Bajaj, S. Member, U. R. Acharya, and S. Member, “SPWVD-CNN for Automated Detection of Schizophrenia Patients Using EEG Signals,” vol. 70, 2021.

[20] A. Vaswani, N. Shazeer, and N. Parmar, “Attention is All You Need,” in 31st Conference on Neural Information Processing Systems (NIPS), 2015.

[21] A. Dosovitskiy, L. Beyer, A. Kolesnikov, and D. Weissenborn, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” in ICLR 2021, 2021.

BIOGRAPHIES

Mesut Şeker received the B.Sc. degree in 2014 from Zirve University, Gaziantep, Turkey. He received his M.Sc degree in Electrical-Electronics Engineering in 2017 from Dicle University and he is currently studying for Ph.D in Electrical-Electronics Engineering at Dicle University. He is also Research Assistant in the same department. His research interest include neural time series analysis, EEG signal processing, neuroscience, pattern recognition and machine learning.

Mehmet Siraç Özerdem received the B.Sc. degree in Electrical-Electronics Engineering in 1994 from Eastern Mediterranean University in Northern Cyprus. He received his M.Sc degree in Electrical Engineering in 1998 from Yıldız Technical University. He received the Ph.D. degree in Computer Engineering in 2003 from Istanbul Technical University. Currently, he works as Professor and researcher at Dicle University, Diyarbakır, Turkey. His research focuses on machine learning, biomedical signal processing, bioinformatics and embedded system design.