Schizophrenia Diagnosis via FFT and Wavelet Convolutional Neural Networks utilizing EEG signals

Maryam Saeedi (Maryam.saeedi33@gmail.com)
Islamic Azad University Science and Research Branch https://orcid.org/0000-0002-4353-5334

Abdolkarim Saeedi
Islamic Azad University Science and Research Branch https://orcid.org/0000-0002-1308-1757

Pooya Mohammadi
K N Toosi University of Technology Faculty of Geodesy and Geomatics Engineering

Research Article

Keywords: Schizophrenia, Electroencephalography, Deep Learning, Fast Fourier, Transformer, Gramian Angular Field, Wavelets, Transformers, Recurrent Neural Networks

Posted Date: March 30th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1497490/v1

Schizophrenia Diagnosis via FFT and Wavelet Convolutional Neural Networks utilizing EEG signals

Maryam Saeedi[1], Pooya Mohammadi Kazaj[2], Abdolkarim Saeedi[1]

[1] Department of biomedical engineering, science and research branch, Islamic Azad University, Tehran, Iran

[2] Department of GIS, Faculty of Geodesy and Geomatics, K.N.Toosi University of Technology, Tehran, Iran

Abstract

Background: Schizophrenia is a chronic mental illness in which a person's perception of reality is distorted. Early diagnosis can help to manage symptoms and increase long-term treatment. The electroencephalogram (EEG) is now used to diagnose certain mental disorders.
Method: In this paper, we developed an artificial intelligence methodology built on deep convolutional neural networks with specialized layers. In the first phase, we used the Gramian Angular Field (GAF) including two methods: Gramian Angular Summation Field (GASF) and Gramian Angular Difference Field (GADF) to represent the EEG signals as various types of images. Then, well-known CNN architectures includes of Transformer CNN-LSTM and two new custom architectures which utilizing two-dimensional Fast Fourier transform layers (CNN-FFT) and wavelet transform layers (CNN-Wavelet) are preformed to extract useful information from the data. These layers allow automated feature extraction from EEG representation in the time and frequency domains.
Results: CNN-FFT and Transformer models derive most valuable features from signals based on the findings. CNN-FFT obtained the highest accuracy of 99.04 percent. Transformer, which has a 98.32 percent accuracy rate, also performs admirably.
Conclusion: This experiment outperformed other previous studies. Consequently, the strategy can aid medical practitioners in automated detection and early treatment of schizophrenia.
The code of the work is also publicly available on github: https://github.com/i1idan/schizophrenia-diagnosis-eeg-signals

Keywords

Schizophrenia, Electroencephalography, Deep Learning, Fast Fourier, Transformer, Gramian Angular Field, Wavelets, Transformers, Recurrent Neural Networks,

1. Introduction

Schizophrenia is a severe mental disorder in which a person instinctively uses an unusual interpretation of reality. Schizophrenia can be accompanied by hallucinations and mental behavioral disorders that disrupt the day-to-day and thus disable the patient. It's also defined as a brain disorder that affects a person's thinking, feeling, and perception. The main symptoms of schizophrenia are signs of insanity such as

auditory hallucinations and delusions. The exact cause of schizophrenia is not known, but researchers believe that a combination of genetic factors, chemicals in the brain, and environmental factors may play a role. For instance, evidence suggests that certain environmental factors, such as some viral infections, exposure to certain toxins, and highly stressful conditions may cause schizophrenia in people who are genetically predisposed to the disease. Schizophrenia is a long-term disease that has no cure but can be controlled. Therefore, people with schizophrenia need lifelong treatment. Additionally, early treatment may control symptoms before the onset of serious complications and may help improve the long-term prognosis. Today, EEG is used as a diagnosing tool for some psychiatric diseases.

Previous Work

Dvey-Aharon et al. [1] introduced a method “TFFO” (Time-Frequency transformation followed by Feature-Optimization) for schizophrenia detection operating on 75 subjects. The technique was utilized for single electrode recordings in order to make the procedure more practical. Akar et al. [2] used nonlinear approaches including approximate entropy (ApEn), Shannon entropy (ShEn), Kolmogorov complexity (KC), and Lempel-Ziv complexity (LZC) in their study and investigated the EEG signal complexity of 22 individuals. In schizophrenia patients, lower complexity values were reported especially in frontal and parietal regions. Dvey-Aharon et al. [3] introduced a new connectivity analysis method named 'Connectivity Maps'. Authors identified unique characteristics using these maps, signals were obtained from 3-5 electrodes and had a relatively short recording period. The experiment was done on a set of 50 subjects, where 25 were healthy and 25 were diagnosed with schizophrenia whom were also treated with anti-schizophrenia medications. Jahmunah et al. [4] studied 14 participants extracting 157 nonlinear features with methods consisting of activity entropy (ae), largest Lyapunov exponent (lx), Kolmogorov- Sinai (k-s) entropy, Hjorth complexity (hc) and mobility (hm), Rényi (re), ShEn, Tsallis (ts), KC, bispectrum (bs), cumulant(c), and permutation entropy(pe). In the mentioned work, significant features were selected and recognized with the t-test algorithm. Zhang et al. [5] analyzed 81 subjects including 32 controls and 49 patients. The features were extracted based on event related potentials (ERP) and were categorized using a random forest classifier. Buettner et al. [6] also built a random forest classifier based on the spectral analysis applied on 28 subjects (14 individuals in each group). Krishnan et al. [7] decomposed the EEG signals into Intrinsic Mode Functions (IMF) signals, performing Multivariate Empirical Mode Decomposition (MEMD). Then the IMF signal's complexity was determined by approximate, sample, permutation, spectral, and singular value decomposition entropies. Support Vector Machine with Radial Basis Function kernel (SVM-RBF) demonstrated the best results. Aslan et al. [8] transformed raw EEG signals of two separate datasets with a short-time Fourier Transform of images. A visual geometry group architecture with 16 layers (VGG16) was applied to classify two-dimensional time-frequency features into schizophrenia patients and healthy controls. Chandran et al. [9] with EEG recordings from 14 schizophrenia patients and 14 healthy controls, determined nonlinear properties such as Katz fractal dimension (KFD) and approximate entropy (ApEn). For diagnosis, they performed an LSTM architecture with four hidden LSTM layers while each layer had 32 neurons. Devia et al. [10] used a visual task procedure and features related to evoked potentials in order to identify patients with schizophrenia. The authors indicate that in the control group, photos with natural content cause later behavior and these differences are to be found in the occipital region. Lih oh et al. [11] studied 14 participants in each group (schizophrenia and healthy controls). They developed an eleven-layer convolutional neural network which was evaluated at 14-fold cross-validation. Shalbaf et al. [12] experimented 28 individuals introducing a methodology based on transfer learning with deep convolutional neural networks (CNNs) which uses the eighteen layered Residual Network (Res-Net18) to achieve a higher performance rate. All raw signals were turned into images by performing continuous wavelet transformation (CWT). Shu Lih, et al [13] with 28 subjects, 14 in each group performed an eleven-

layered convolutional neural network (CNN) model in order to classify and extract features for signals. Shoeibi et al [14] investigated 14 subjects in each schizophrenia and control group. They applied different conventional machine learning methods and deep learning architectures which CNN-LSTM combination achieved the most promising result. Siuly et al [15] assesses 81 subjects, which includes 49 schizophrenia patients and 32 healthy controls, this research proposes a particular technique incorporating the empirical mode decomposition (EMD) approach, in which each EEG signal is decomposed into intrinsic mode functions (IMFs) by the EMD algorithm. Among the classifiers that authors considered, the ensemble bagged tree outperformed the others.

The aim of the current study is to determine EEG indices as a predictor for patients diagnosed with schizophrenia. In this study, several models were implemented one of which was an automated diagnosis approach based on the deep convolutional neural networks with custom layers. Gramian Angular Field (GAF) is used to encode the EEG signals as different types of images. Then two-dimensional Fast Fourier Transform layers and wavelet transform layers are used in different CNN architectures. These layers allow automatic feature extraction from EEG signals in the time, frequency, and time-frequency domain, which increases the feasibility and applicability of the method in reality. Figure 1 shows the outline of the study.

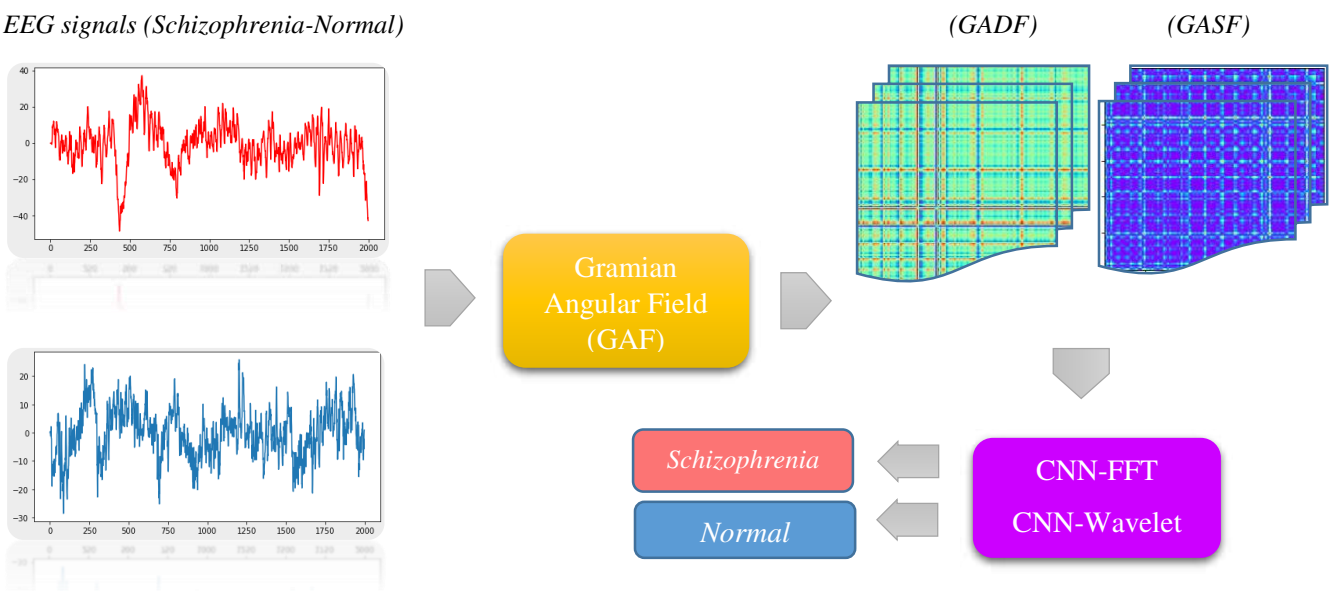


Figure 1 Block diagram of the proposed method

2. Materials and Methods

2.1 Dataset

The dataset used in this work was provided by Olejarckyz et al in 2017 [16] which is publicly available. Recordings include 14 paranoid schizophrenia patients (7 females) with age ranging from 27 to 32 and 14 normal subjects (7 females) with age ranging from 26 to 32. EEG data was recorded with eyes closed for

fifteen minutes. Recordings were obtained from 19 electrodes that were placed on the scalp according to 10-20 international standard electrode position classification system. The sampling frequency was 250Hz.

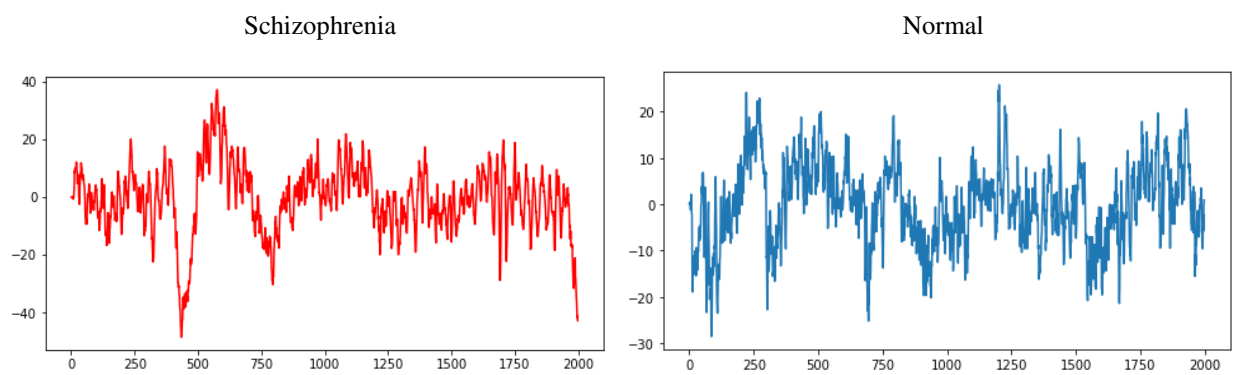


Figure 2 Examples of normal and schizophrenia EEG signals

2.2 Gramian Angular Field (GAF)

Gambian Angular Field (GAF) is a time-series encoding method first introduced by Wang et Al [17-20] for classification of EEG signals using deep convolutional neural networks. In this approach, initially signals are normalized into the [0, 1] interval using the following formula:

$$X = \{x_1 , \ x_2 , \ldots . x_n \} \tag{1}$$

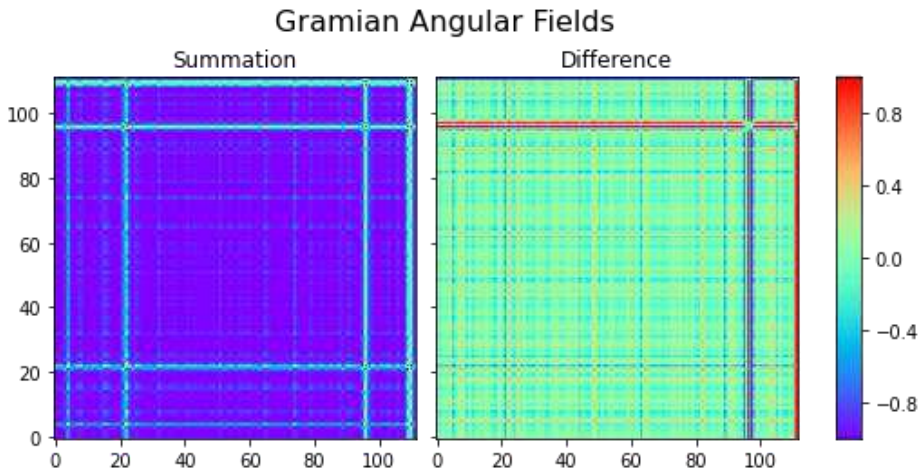
$$x^i = \frac{x_i - (X)}{(X) \ - min(X)}$$

Further, they are transformed into the polar coordinate system using the following equation:

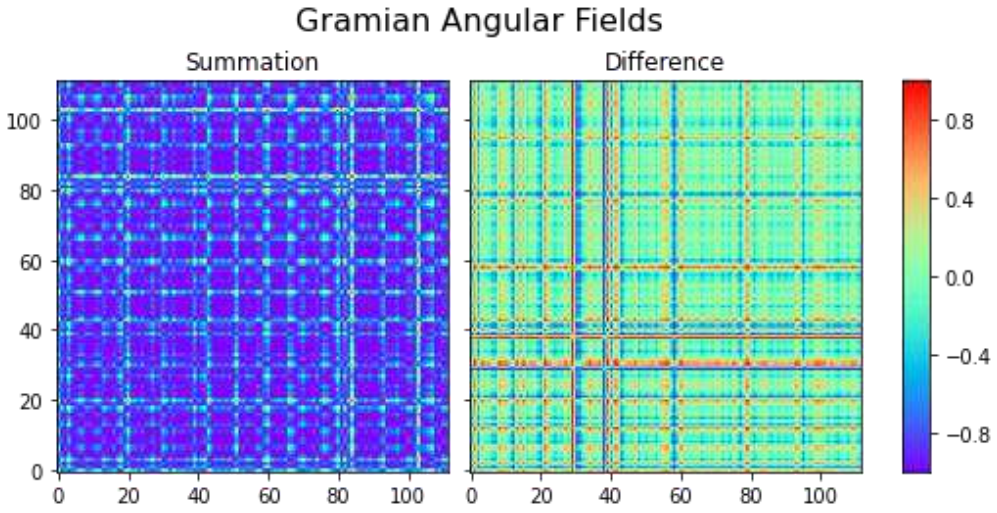
$$\emptyset = arccos \ arccos \left(x^i\right) \tag{2}$$

$$r = \frac{t^i}{N} \tag{3}$$

Lastly, the GAF Matrix (G) obtained from rescaled time-series resulted in images which were fed to the convolutional neural network. Two techniques implemented in this study were Gramian Angular Summation Field (GASF) and Gramian Angular Difference Field (GADF).



Schizophrenia



Normal

Figure 3 Comparison of normal and schizophrenia Gramian Angular Fields

2.3. Convolutional neural networks (CNN)

CNN is a special form of neural network that is commonly used in the tasks of image processing and classification [21]. CNN utilizes local information by using filters, so-called kernels, which is far more efficient in the number of parameters with respect to Multi-Layer Perceptron (MLP) [22]. It is the state-of-the-art technique of deep learning to analyze two-dimensional data. A typical setting consists of

convolutional layers for extracting features, followed by pooling layers to reduce the input dimension and lastly dense layers to map the features into a more distinct space for classification.

2.3.1 Convolutional layers

Convolutional layers apply a linear transformation using a specific kernel. Filters are a way of extracting features from input patterns. The procedure may be described as:

$$x_{ij}^{l} = \sum_{a=0}^{k-1} \sum_{b=0}^{k-1} w_{ab}\, y^{l-1}\,(i+a)(j+b) \tag{4}$$

$$y_{ij}^{l} = \sigma\left(x_{ij}^{l}\right) \tag{5}$$

The kernel has a size of $K \times K$, in the case of a symmetrical trainable kernel and the weights are denoted as w_{ab}. x_{ij}^{l} represents an output from the current convolution layer l and y^{l-1} shows the last layers output which is the input to the next layer. Finally the output y_{ij}^{l} is calculated after applying a non-linearity through an activation function which was chosen to be Rectified Linear Unit (ReLU). ReLU is superior to Sigmoid and Tanh due to faster calculation time and better back propagation in extreme values.

2.3.2 Pooling layers

Pooling layers are a technique to reduce the input size while extracting features with convolutional layers, the most conventional form is called MaxPooling. Where the pixel with the maximum Intensity is chosen over each window, reducing the image size to $\frac{N}{K} \times \frac{N}{K}$ with a pooling kernel size of $K \times K$. Pooling layers improve the computations in neural networks and provide a method to deal with overfitting by constraining the feature space.

2.3.3 Fully-connected layers

Fully-connected layers, also called Dense Layers, use linear regression like transformations to give more importance to features that can improve the decision boundary. The procedure is shown below:

$$v_j(n) = \sum_{i=0}^{m} w_{ji}(n) y_j(n) \tag{6}$$

$$y_i(n) = \varphi_j\left(v_j(n)\right) \tag{7}$$

Where y_i is the output of the layer i, computed by multiplying previous layers' weights w_{ji} and output y_j, to get $v_j(n)$, then $v_j(n)$ is activated (φ_j), resulting in $y_i(n)$.

2.4 Fourier Transform

Decomposes the signal into a series of sine-based functions, the absolute values of the Fourier Transform represent the signals' frequency behavior [23-25]. In this work, Fourier Layers were introduced as a starting point for the convolutional neural networks so that the features can be extracted from a different representation that of the data which appear to be more useful compared to the time domain features.

2.5 Wavelet Transform

The wavelet transform [26-27] represents an input signal in multiple resolutions using bandpass filters. In a discrete wavelet transform (DWT) signals are categorized into high and low frequency components known as detail and approximation coefficients respectively. Then the approximation coefficients are split into next-level approximate and detailed coefficients and the process continues to depend on the depth of the wavelet tree.

2.6 FFT-CNN2D

A 19-layer 2D-FFT-CNN model was used in this analysis, as shown in Figure 4, with a number of parameters in each layer. A two-dimensional FFT operation is implemented in the input layer. Then, in the next layers, two convolution operations followed by a max-pooling process were conducted in order to extract features from images and reduce their size. A dropout layer also was applied after the mentioned procedures to reduce overfitting possibilities. Following that, for the next round also two subsequent convolutions were used to construct the following layers (layer 6-7). After the convolutions, the max-pooling operation is applied once more to obtain layer 8. To generate layers (9-15), three more convolution and dropout operations were performed alternately. After that, a Flattening layer was used in order to convert the data into a 1-dimensional array for inputting it to the next dropout layer which was connected to a dense layer with 32 units. Finally, the last dense layer (classification layer) with two output neurons presented normal and schizophrenia.

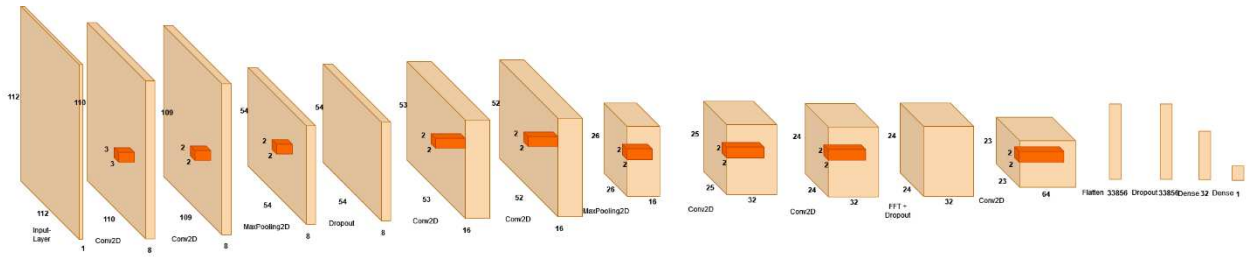


Figure 4 FFT-CNN2D architecture

2.7 Wavelet-CNN2D

A 2D-Wavelet-CNN model was introduced in this study, as it shown in Figure 5, with the number of parameters in each layer, filter sizes, and strides are summarized in Table 1. First, to achieve the output of the first layer, a two-dimensional Gabor wavelet transform operation was implemented. The rest of the structure is identical to the 2D-FFT-CNN. A Gabor filter, introduced by Dennis Gabor [28-29], is a linear texture analysis filter in image processing. It examines whether the image contains any specific frequency content in specific directions in a localized region around the point or region of the assessment. A two-dimensional Gabor filter is a Gaussian kernel function induced by a sinusoidal plane wave in the spatial domain.

2.8 Long Short Term Memory (LSTM)

LSTM, a variant of RNN [30-31], preserves the long-term memory of what has been gone through its network (Hochreiter and Schmidhu 1997). These networks have been introduced to mitigate vanishing gradients from which vanilla RNN always suffered (Hochreite 1998). The issue is that the network gradient becomes close to or almost equal to zero before reaching the first layers; therefore, backpropagation becomes less effective. Each LSTM has a cell state vector apart from a hidden state vector to overcome this issue. At each time step, the next LSTM cell can choose to read from its previous cell state, write to it or modify it using four same-shape gates that allow easy information transition and gradient flow through long sequences. These gates are explained in the following sections.

2.8.1 Forget gate

This gate decides whether to preserve information from the previous cell state or forget it. Cell states are the primary components for transmitting information through LSTM cells. This operation can be calculated as equation (8):

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \tag{8}$$

Similar to the input gate formula, x_t as input data and h_{t-1} as the previous layer's hidden state are concatenated and fed to a linear transition with w_f and b_f as weights and biases, respectively. Eventually, the result is activated using a sigmoid function. The output closer to zero means to forget, and one means to retain.

2.8.2 Input gate

The input gate controls whether the cell state is updated by calculating the importance of the current input data to the network. This procedure can be calculated as equation (9):

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \tag{9}$$

where x_t is the current input data, and h_{t-1} is the previous layer's hidden state which both are concatenated and passed through a linear transformation with w_i as weights and b_i as biases. Finally, a sigmoid function is applied to the previous step, bounding the output between 0 and 1 where the output closer to zero means no update and closer to one means to update.

2.8.3 Candidate gate

Like the input gate, the candidate gate is also responsible for updating the cell state. However, unlike the input gate, this gate performs the Tanh activation function over the linear transformation. The expression is simplified into equation (10):

$$c_t = tanh(w_c[h_{t-1}, x_t] + b_c) \tag{10}$$

Like the other gates, this gate also feeds concatenated x_t and h_{t-1} to a linear transformation with w_c and b_c presenting unique weights and biases. Finally, the cell state is calculated as in equation (11):

$$c_t = f_t * c_{t-1} + i_t * ca_t \tag{11}$$

2.8.4 Output gate

This gate controls what percentage of the current cell state's information should be considered as the final output and transferred as the hidden state to the next cell. This gate can be expressed as equation (12):

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \tag{12}$$

Similar to other gates, this gate also conveys concatenated input x_t and h_{t-1} and feeds them through a linear transformation and subsequently to a sigmoid activation function while w_o and b_o featuring unique weights and biases of the linear layer. It is worth mentioning that the current cell state's information is passed through the $Tanh$ activation function before multiplying to the output gate. Thus, the current hidden state and the current cell's final output is calculated as in equation (13):

$$h_t = o_t * tanh(c_t) \tag{13}$$

2.9 Convolutional neural network-long short-term memory (CNN-LSTM)

CNN-LSTM has been gaining much attention in the past few years. It is used on variant tasks like finding spatiotemporal relations in input text for text classification [32]or depressive disorder diagnosis on EGG signals [33] CNN-LSTM results from a series of convolutional layers followed by some LSTM layers. First, it extracts rich features using convolutional layers from input data. Then feeds these features to LSTM layers responsible for extracting temporal information. Finally, the classification is done by applying fully-connected layers to the temporal information obtained by LSTM layers. The CNN-LSTM

used in this study consisted of three one-dimensional convolutional layers with a kernel size of 5 and filter sizes of 8, 4, and 2, respectively. Then a dropout layer with a 50% chance of dropping input neurons is applied to avoid overfitting. Following that, a one-dimensional max-pooling layer with a pool size of 2 is applied to reduce the size of the features and, consequently, the required computation power. After that, an LSTM layer with a filter size of 512, which is then followed by the second dropout layer with a 25% drop chance, was used. Following that, a fully connected layer with 128 units pursued by the third dropout layer with a 25% drop chance was used. Eventually, a fully connected layer with one neuron unit is used to perform classification. All Activation functions were ReLU, except for the classification layer which was the sigmoid function.

2.10 Transformer

Machine translation was the first application of transformers [34]. In contrast to text data, the signal in the time series is separated into n identical fragments, which is comparable to the embedding procedure on text characteristics. The input data is specified theoretically as:

$$X = X_1, X_2, \dots X_N \tag{14}$$

Positional Embeddings have been used in transformers to maintain the order of the series after the separation occurs. This layer creates an embedding depending on the maximum length of the segment and the total number of segments, which is subsequently appended to the feature vector. Keys, Values, and Queries are other crucial components of the transformer that aid in the calculation of attention weights; the formula is as follows:

$$ATTENTION = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{15}$$

The scaling factor is known as the parameter d_k. The Multi Head component of the transformer also refers to the fact that the attention's result is made up of several attention weights, also known as heads that are merged to learn many representations of the input in the end. It's also worth noting that for a classification problem, only the encoder portion of a transformer is required, with the decoder section being omitted entirely.

$$Q = XW_Q \tag{16}$$

$$K = XW_K \tag{17}$$

$$V = XW_V \tag{18}$$

Parameters	CNN-FFT	CNN-Wavelet	CNN-LSTM	Transformer
Batch Size	4	4	4	4
Loss Function	crossentropy	crossentropy	crossentropy	crossentropy
Optimizer	Adam	Adam	Adam	Adam
Learning rate	0.001	0.001	0.001	0.001
reduce lr	50	50	50	50
Epochs	200	200	200	200

Table 1 models tuning parameters

2.11 Evaluation

Independently, 2 versions of the proposed models were trained on 80% of the data and then evaluated on the rest of t. This procedure was then repeated ten times and the results were averaged to derive the mean and standard deviation of the metrics for each model. The accuracy, sensitivity and specificity measures are computed in this study as follows:

$$Sensitivity = \frac{TP}{TP+FN} \tag{19}$$

$$Specificity = \frac{TN}{TN+FP} \tag{20}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{21}$$

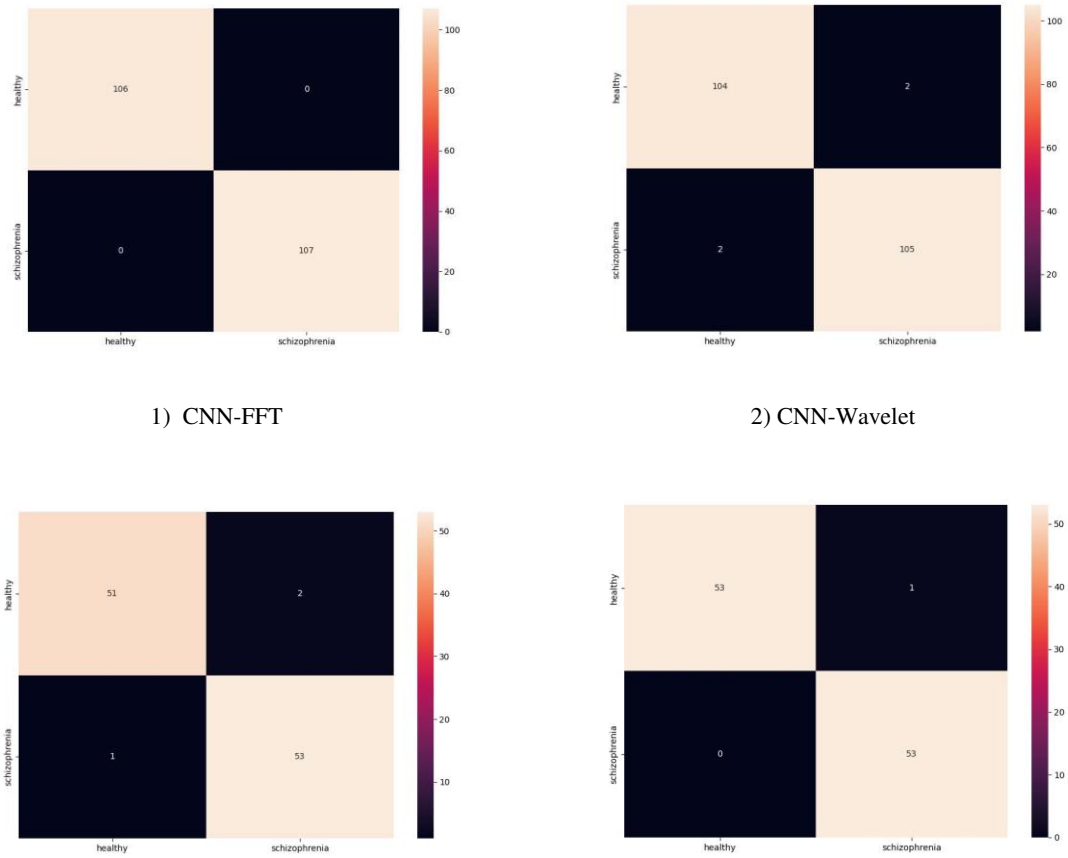
3 Results

In this section, the results of applying the proposed approaches on the dataset are provided for four proposed models. Each signal is converted into an image using GASF and GADF methods. Since there are 2 measures (GASF and GADF) and 19 different channels, 38 images are obtained for each signal. Figure 2 shows a sample image for healthy subjects and schizophrenia patients. These images used as inputs of introduced

convolutional neural networks. The data preparation is the same for Transformer-based and CNN-LSTM models except that the signals are not converted to tow-dimensional images and are fed directly to the models. For the model training section, a batch size of 4 was selected and each network was trained for 200 hundred epochs. Additionally, the initial learning rate was set to 0.001 and if there were no improvements after 50 epochs, learning rate was decreased with a factor of 0.1. Binary Cross Entropy was chosen as the loss function and in the optimization phase, ADAM algorithm was chosen due to its superior results and shorter run-time. Classifier tuning parameters are shown in table 1. Training was performed on 80% of the data for the classification of schizophrenia patients and healthy subjects and the rest of the data was used for evaluating the performance of the classifier using various metrics (accuracy, sensitivity, and specificity). Table 2 shows the best-obtained accuracy over different deep neural network architectures. Based on this table, the best result is reached by the proposed CNN-FFT. Without custom layers, like FFT and Gabor filter, models suffered from overfitting or a high variance. The classification results of models without these layers are also provided. Based on these results, both CNN-FFT and Transformer-based models extract useful information from the signals. Highest Accuracy 99.04% achieved by CNN-FFT and Transformer-based model with 98.32% accuracy also presented a remarkable performance.

Model – Metrics	Accuracy	Sensitivity	Specificity	F1 Score
CNN-FFT	99.04% ± 0.0852	0.9906± 0.0065	0.9859± 0.0047	99.04% ± 0.0852
CNN-Wavelet	92.80% ± 0.0169	0.9332± 0.0575	0.9251± 0.0573	92.80% ± 0.0169
CNN-LSTM	0.9598% ± 0.1124	0.9591± 0.03047	0.9065± 0.053	0.9598% ± 0.1124
Transformer	98.32% ± 1.208	0.9823± 0.0187	0.9666± 0.0396	98.32% ± 1.208

Table 2 comparison of the proposed models



1) CNN-FFT

2) CNN-Wavelet

3) CNN-LSTM

4) Transformer

Figure 5 Confusion Matrix

4 Discussion

In this experiment, we employed custom deep learning and Gramian angular field (GAF) approaches to automate the identification of schizophrenia patients and healthy controls. The accuracy of the CNN-FFT architecture in images of GADF and GASF methods on 19 channels of EEG signals is 99 percent. The use of GADF and GASF approaches to transform a 1-D EEG signal into a 2-D representation that can be directly fed to CNN architecture is one of the work's key innovations. There are a variety of methods for converting a 1-D signal to a 2-D image, including traditional methods based on time-frequency distribution (STFT, wavelets, and so on) [8]. Another innovation of the research is the use of FFT and Wavelet layers in the deep model which processes the input images of the network and creates a time and frequency representation of the EEG signals, and instead of using manual time-frequency features, this method accelerates and automates the process of detection. In table 2, that the CNN-FFT model outperformed other approaches in terms of accuracy, the findings of this analysis are associated with the new best similar research that used EEG signals from the same database ([11], [8], [9], [13] and [14]) and a separate database ([1] ,[10] and [35]) As can be seen, the accuracy obtained in this experiment is higher than that reported in the previous studies using conventional machine learning methods for extracting linear and non-linear

attributes, demonstrating the superiority of the proposed approach. Furthermore, as compared to other deep learning approaches on time-series data of EEG signals, our results with CNN-FFT and images that are built with Gramian Angular Field have higher accuracy in comparison to other researches of this kind. Additionally, this study has the benefit of comparing various deep learning models. Ultimately, this research performed better in automatic schizophrenia of depressive patients and healthy controls so far, as seen in Table 6. The research's biggest flaw is the scale of the dataset used to train the networks. We were able to solve this problem by using regularization terms and modifying deep models. Our long-term goal is to extend the experimental area by gathering more data.

Authors	Methods	samples	Classification Methods	Accuracy
Dvey-Aharon[1]	TFFO	25 normal 25 patients	KNN	88.7%
Devia[10]	ERP feature extraction	11 normal 9 patients	LDA	71%.
Phang et al[35]	connectivity	39 normal 45 patients	CNN	93.06%
Shu Lih Oh et al.[11]	-	14 normal 14 patients	CNN	98.07%
Aslan et al[8]	Short-time Fourier Transform (STFT)	39 normal 45 patients 14 normal 14 patients	VGG-16,	95% 97%
Chandran et al[9]	Katz fractal dimension (KFD) and approximate entropy (ApEn),	14 normal 14 patients	LSTM	99.0%.
Shu Lih, et al [13]	-	14 normal 14 patients	11-layered deep CNN model Subject base testing using 14-fold Non-subject base testing using 10-fold	Non-subject base testing: Acc: 98.07% Sen: 97.32% Spe: 98.17% Ppv: 98.45% Subject base testing: Acc: 81.26% Sen: 75.42% Spe: 87.59% Ppv: 87.59%
Shoeibi et al [14]	-	14 normal 14 patients	1D CNN-LSTM	99.25%

Siuly et al [15]	Approximate Entropies empirical mode decomposition (EMD) based characteristics	49 patients with schizophrenia and 32 healthy control subject	EBT (ensemble bagged trees)	89.59%
Current study	-	14 normal 14 patients	CNN-FFT CNN-Wavelet CNN-LSTM Transformer	99.0%.

Table 3 a summary of previous studies on automated EEG-based schizophrenia detection

5 Conclusion

By using Gramian Angular Field (GADF, GASF) and a variety of well-known deep learning algorithms, a thorough assessment was done in this article. The best model is CNN-FFT which has the highest accuracy of 99.04 percent in classifying schizophrenia patients and healthy controls. This deep learning model identifies the time and frequency properties of EEG signals. Compared to all recent researches, this study yielded the best performance. As a result, the new approach will assist health-care professionals in identifying schizophrenia patients for early detection and treatments.

6 References

[1] Dvey-Aharon Z, Fogelson N, Peled A, Intrator N (2015) Schizophrenia Detection and Classification by Advanced Analysis of EEG Recordings Using a Single Electrode Approach. PLOS ONE 10(4): e0123033. https://doi.org/10.1371/journal.pone.0123033
[2] Akar, S. A., Kara, S., Latifoğlu, F. A. T. M. A., & Bilgiç, V. (2016). Analysis of the complexity measures in the EEG of schizophrenia patients. International journal of neural systems, 26(02), 1650008.
[3] Dvey-Aharon Z, Fogelson N, Peled A, Intrator N (2017) Connectivity maps based analysis of EEG for the advanced diagnosis of schizophrenia attributes. PLoS ONE 12(10): e0185852. doi:10.1371/journal.pone.0185852
[4] Jahmunah, V., Oh, S. L., Rajinikanth, V., Ciaccio, E. J., Cheong, K. H., Arunkumar, N., & Acharya, U. R. (2019). Automated detection of schizophrenia using nonlinear signal processing methods. Artificial intelligence in medicine, 100, 101698.
[5] Zhang, L. (2019, July). EEG Signals Classification Using Machine Learning for The Identification and Diagnosis of Schizophrenia. In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 4521-4524). IEEE.
[6] Buettner, R., Beil, D., Scholtz, S., & Djemai, A. (2020, January). Development of a machine learning based algorithm to accurately detect schizophrenia based on one-minute EEG recordings. In Proceedings of the 53rd Hawaii International Conference on System Sciences.
[7] Krishnan, P. T., Raj, A. N. J., Balasubramanian, P., & Chen, Y. (2020).]]> Schizophrenia DSchizophrenia detection using Multivariate Empirical Mode Decomposition and Entropy Measures from

Multichannel EEG Sentropy measures from multichannel EEG signal. Biocybernetics and Biomedical Engineering.
[8] Aslan, Z., & Akin, M. (2020). Automatic Detection of Schizophrenia by Applying Deep Learning over Spectrogram Images of EEG Signals. Traitement du Signal, 37(2), 235-244.
[9] Chandran, A. N., Sreekumar, K., & Subha, D. P. (2020). EEG-Based Automated Detection of Schizophrenia Using Long Short-Term Memory (LSTM) Network. In Advances in Machine Learning and Computational Intelligence (pp. 229-236). Springer, Singapore.
[10] Devia, C., Mayol-Troncoso, R., Parrini, J., Orellana, G., Ruiz, A., Maldonado, P. E., & Egaña, J. I. (2019). Eeg classification during scene free-viewing for schizophrenia detection. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 27(6), 1193-1199.
[11] Shu Lih Oh et al. Oh, S.L., Vicnesh, J.,Ciaccio,E.J.,Yuvaraj, R.,Acharya, U.R.(2019). Deep convolutional neural network model for automated diagnosis of schizophrenia using EEG signals.AppliedScience, 9(14):2870
[12] Shalbaf, A., Bagherzadeh, S., & Maghsoudi, A. (2020). Transfer learning with deep convolutional neural network for automated detection of schizophrenia from EEG signals. Physical and Engineering Sciences in Medicine, 43(4), 1229-1239.
[13] Oh, S. L., Vicnesh, J., Ciaccio, E. J., Yuvaraj, R., & Acharya, U. R. (2019). Deep convolutional neural network model for automated diagnosis of schizophrenia using EEG signals. *Applied Sciences*, *9*(14), 2870.
[14] Shoeibi, A., Sadeghi, D., Moridian, P., Ghassemi, N., Heras, J., Alizadehsani, R., ... & Gorriz, J. M. (2021). Automatic Diagnosis of Schizophrenia in EEG Signals Using CNN-LSTM Models. *Frontiers in Neuroinformatics*, *15*.
[15] S. Siuly, S. K. Khare, V. Bajaj, H. Wang and Y. Zhang, "A Computerized Method for Automatic Detection of Schizophrenia Using EEG Signals," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 11, pp. 2390-2400, Nov. 2020, doi: 10.1109/TNSRE.2020.3022715.
[16] Olejarczyk E, Jernajczyk W. Graph-based analysis of brain connectivity in schizophrenia. PLoS One. 2017 Nov 30;12(11):e0188629. doi: 10.1371/journal.pone.0188629. PMID: 29190759; PMCID: PMC5708839.
[17] G. Zhang, Y. Si, D. Wang, W. Yang and Y. Sun, "Automated Detection of Myocardial Infarction Using a Gramian Angular Field and Principal Component Analysis Network," in *IEEE Access*, vol. 7, pp. 171570-171583, 2019, doi: 10.1109/ACCESS.2019.2955555.
[18] H. Xu et al., "Human Activity Recognition Based on Gramian Angular Field and Deep Convolutional Neural Network," in IEEE Access, vol. 8, pp. 199393-199405, 2020, doi: 10.1109/ACCESS.2020.3032699.
[19] G. Zhang, Y. Si, D. Wang, W. Yang and Y. Sun, "Automated Detection of Myocardial Infarction Using a Gramian Angular Field and Principal Component Analysis Network," in IEEE Access, vol. 7, pp. 171570-171583, 2019, doi: 10.1109/ACCESS.2019.2955555.
[20] M. M. Islam and M. M. H. Shuvo, "DenseNet Based Speech Imagery EEG Signal Classification using Gramian Angular Field," 2019 5th International Conference on Advances in Electrical Engineering (ICAEE), 2019, pp. 149-154, doi: 10.1109/ICAEE48663.2019.8975572.
[21] Saeedi, A., Saeedi, M., Maghsoudi, A. *et al.* Major depressive disorder diagnosis based on effective connectivity in EEG signals: a convolutional neural network and long short-term memory approach. *Cogn Neurodyn* **15,** 239–252 (2021). https://doi.org/10.1007/s11571-020-09619-0
[22] Saeedi, M., Saeedi, A. & Maghsoudi, A. Major depressive disorder assessment via enhanced k-nearest neighbor method and EEG signals. *Phys Eng Sci Med* **43,** 1007–1018 (2020). https://doi.org/10.1007/s13246-020-00897-w
[23] Delimayanti, Mera Kartika, et al. "Classification of brainwaves for sleep stages by high-dimensional FFT features from EEG signals." *Applied Sciences* 10.5 (2020): 1797.
[24] Sun, Jie, et al. "A hybrid deep neural network for classification of schizophrenia using EEG Data." *Scientific Reports* 11.1 (2021): 1-16.

[25] Aslan, Zülfikar, and Mehmet Akin. "Automatic Detection of Schizophrenia by Applying Deep Learning over Spectrogram Images of EEG Signals." *Traitement du Signal* 37.2 (2020).

[26] Saeedi, M., Saeedi, A. & Maghsoudi, A. Major depressive disorder assessment via enhanced k-nearest neighbor method and EEG signals. *Phys Eng Sci Med* **43,** 1007–1018 (2020).

[27] Sharma, M., & Acharya, U. R. (2021). Automated detection of schizophrenia using optimal wavelet-based $$ l_1 $$ l 1 norm features extracted from single-channel EEG. *Cognitive Neurodynamics*, 1-14.

[28] Gabor, D. (1946). Theory of communication. Journal of the Institute of Electrical Engineers, 93, 429–457

[29] R. Roslan and N. Jamil, "Texture feature extraction using 2-D Gabor Filters," 2012 International Symposium on Computer Applications and Industrial Electronics (ISCAIE), 2012, pp. 173-178, doi: 10.1109/ISCAIE.2012.6482091.

[30] Long short-term memory. S Hochreiter, J Schmidhuber. Neural computation, MIT Press, 1997.

[31] Nagabushanam, P., George, S. T., & Radha, S. (2019). EEG signal classification using LSTM and improved neural network algorithms. Soft Computing, 1-23.

[32] Luan, Y., & Lin, S. (2019). Research on Text Classification Based on CNN and LSTM. 2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 352-355.

[33] Saeedi, A., Saeedi, M., Maghsoudi, A. *et al.* Major depressive disorder diagnosis based on effective connectivity in EEG signals: a convolutional neural network and long short-term memory approach. *Cogn Neurodyn* 15, 239–252 (2021). https://doi.org/10.1007/s11571-020-09619-0

[34] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).

[35] Phang, C.-R., Noman, F., Hussain, H., Ting, C.-M., & Ombao, H. (2020). A Multi-Domain Connectome Convolutional Neural Network for Identifying Schizophrenia From EEG Connectivity Patterns. IEEE Journal of Biomedical and Health Informatics, 24(5), 1333–1343. doi:10.1109/jbhi.2019.2941222