Extended Data

Graph Neural Networks for Neurological Diseases Detection in EEG Signals

Muhammad Awais[1], Samir Brahim Belhaouari[2] & Khelil Kassoul[3*]

[1]Department of Creative Technologies, Air University, Islamabad, Pakistan. [2]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar. [3]Geneva School of Business Administration, University of Applied Sciences Western Switzerland, HES-SO, Geneva, Switzerland. *e-mail: khelil.kassoul@hesge.ch

Extended Data Table 1| CHB-MIT-EEG. Patients with the longest and shortest recording durations are highlighted in bold.

Patient	Gender	Age	Number of seizure events (Tmax - Tmin in seconds)	Total duration of seizures (sec)	Total duration of non- seizures (sec)	Total duration (sec)
1	F	11	7 (28-102)	449	23476	23925
2	M	11	3 (10-83)	175	7984	8159
3	F	14	7 (48-70)	409	24791	25200
4	M	22	4 (50-117)	382	37977	38359
5	F	7	5 (97-121)	563	17437	18000
6	F	1.5	10 (13-21)	163	93053	93216
7	F	14.5	3 (87-144)	328	32209	32537
8	M	3.5	5 (135-265)	924	17076	18000
9	F	10	4 (63-80)	280	34219	34499
10	M	3	7 (36-90)	454	50010	50464
11	F	12	3 (23-753)	809	9250	10059
12	F	2	27 (14-98)	1016	33844	34860
13	F	3	10 (18-71)	450	24750	25200
14	F	9	8 (15-42)	177	25023	25200
16	F	7	8 (7-15)	77	17923	18000
17	F	12	3 (89-116)	296	10528	10824
18	F	18	6 (31-69)	323	19951	20274
19	F	19	3 (78-82)	239	10307	10546
20	F	6	8 (30-50)	302	19734	20036
21	F	13	4 (13-82)	203	13587	13790
22	F	9	3 (59-75)	207	10593	10800
23	F	6	7 (21-114)	431	31823	32254
24	Unknown	Unknown	16 (17-71)	539	42661	43200

Extended Data Table 2| Description of EEG Datasets: Number of Subjects, Gender Distribution, Age Range, Seizure Events, Seizure Duration, EEG Channels, and Sampling Rate.

Dataset	Number of Subjects	Gender Distribution	Age Range (years)	Seizure Events	Seizure Duration	EEG Channels	Sampling Rate (Hz)
CHB-MIT-EEG	22	10 male, 12 female	0.3-17.5	198	few seconds to several minutes	22	256
SSE-EEG	14	9 male, 5 female	20-71	47	few seconds to few minutes	Varies (Up to 34)	512
TUH-EEG	10,874	51% female	<1-90+	N/A (Ongoing collection)	N/A	Varies (Up to 31)	Most at 250

Extended Data Algorithm 1| Algorithm to convert EEG features set to graph nodes and edges.

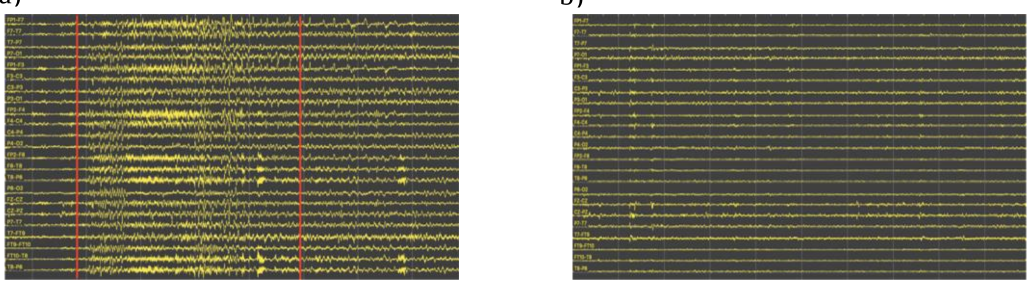
INPUT: X: a matrix, where each row represents the window features of a specific channel

y: a vector indicating the class label for each row in X.

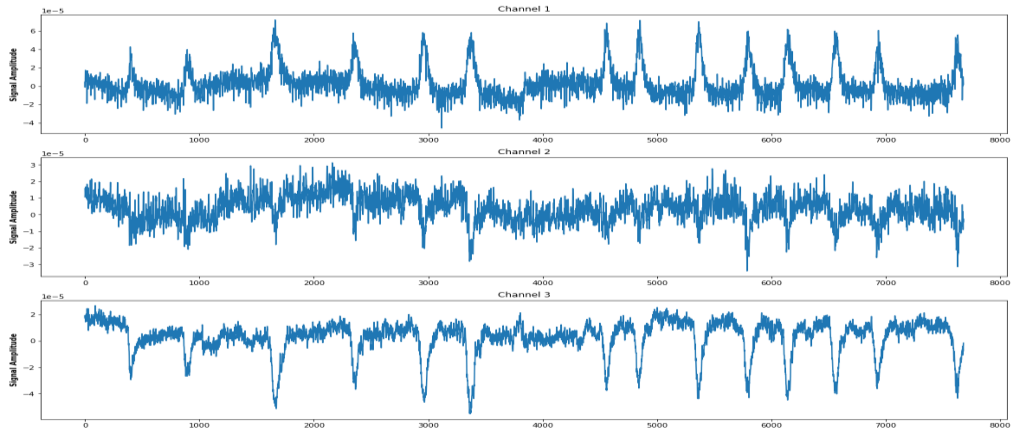
OUTPUT: G (graph structure)

1. Initialize an empty graph G
2. For i = 0 to i < number of rows in X, do:
   a. Add the i^{th} row as a node in the graph G
3. For i = 0 to i < length of X, do:
   a. For j = i+1 to j < length of X, do:
      i. If $i == j$, then continue
      ii. Calculate the Euclidean Distance between $X(i,:)$ and $X(j,:)$ and store it
      iii. If $d < 0.2$ and $d(X(i,:), X(j,:))$, then add an edge between nodes i and j in G
4. Return (G, y)

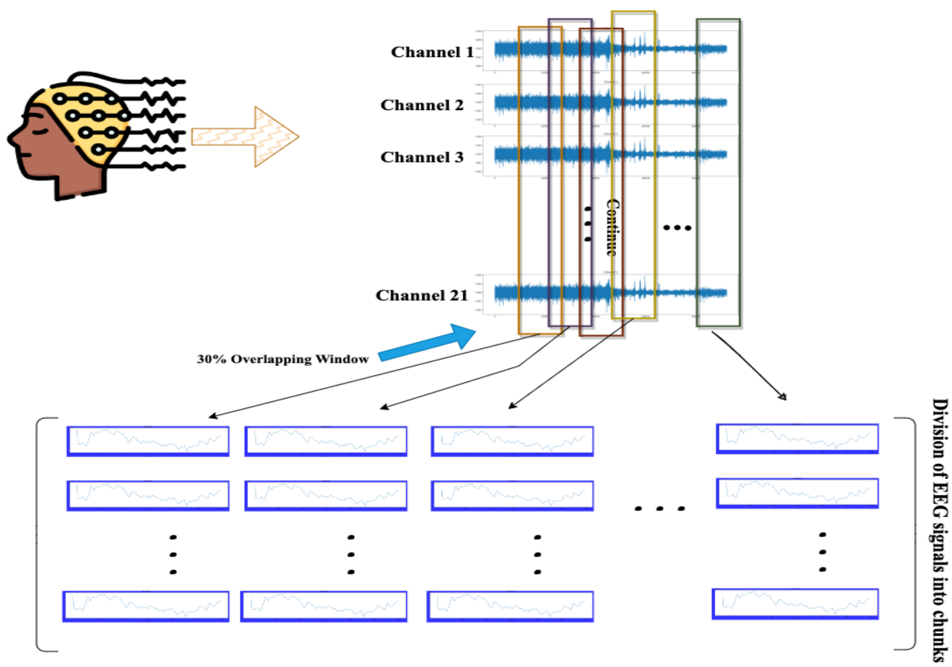
a) b)



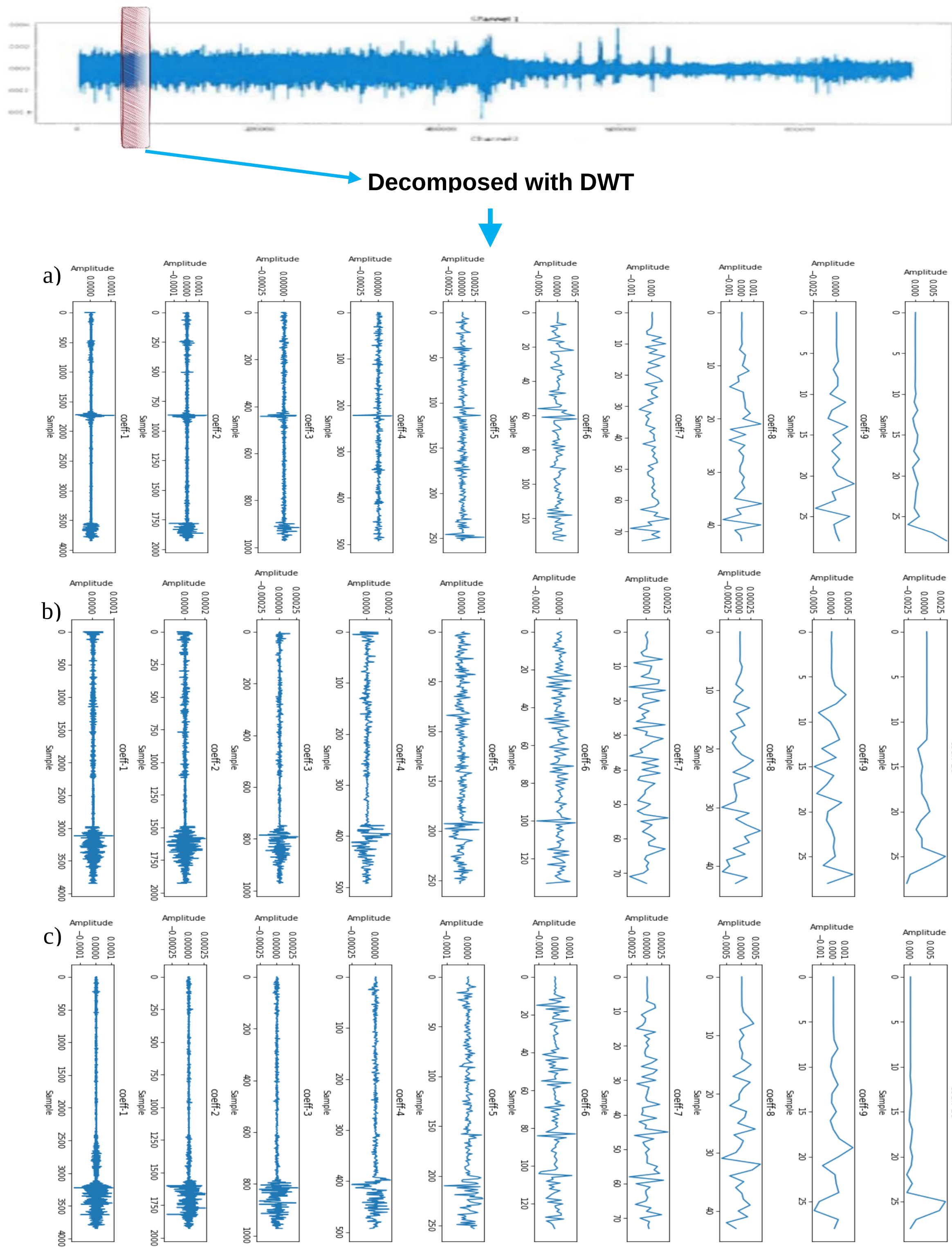
Extended Data Fig. 1 | CHB01_03 file with seizure and non-seizure signals: (a) shows abnormal electrical activity in the brain, consistent with a seizure. The figure shows a spike and wave pattern, which is a characteristic of a seizure. The spike and wave pattern are caused by abnormal synchronization of electrical activity in the brain. This synchronization can lead to a seizure, which is a sudden burst of electrical activity in the brain that can cause a variety of symptoms, including loss of consciousness, convulsions, and sensory disturbances. **(b)** shows normal electrical activity in the brain. The figure does not show any spike and wave patterns, which is consistent with normal brain activity. Normal brain activity is characterized by a variety of different electrical patterns, including alpha waves, beta waves, theta waves, and delta waves. These waves are associated with different levels of consciousness and brain activity. For example, alpha waves are associated with a relaxed state of wakefulness, while beta waves are associated with a state of alertness.



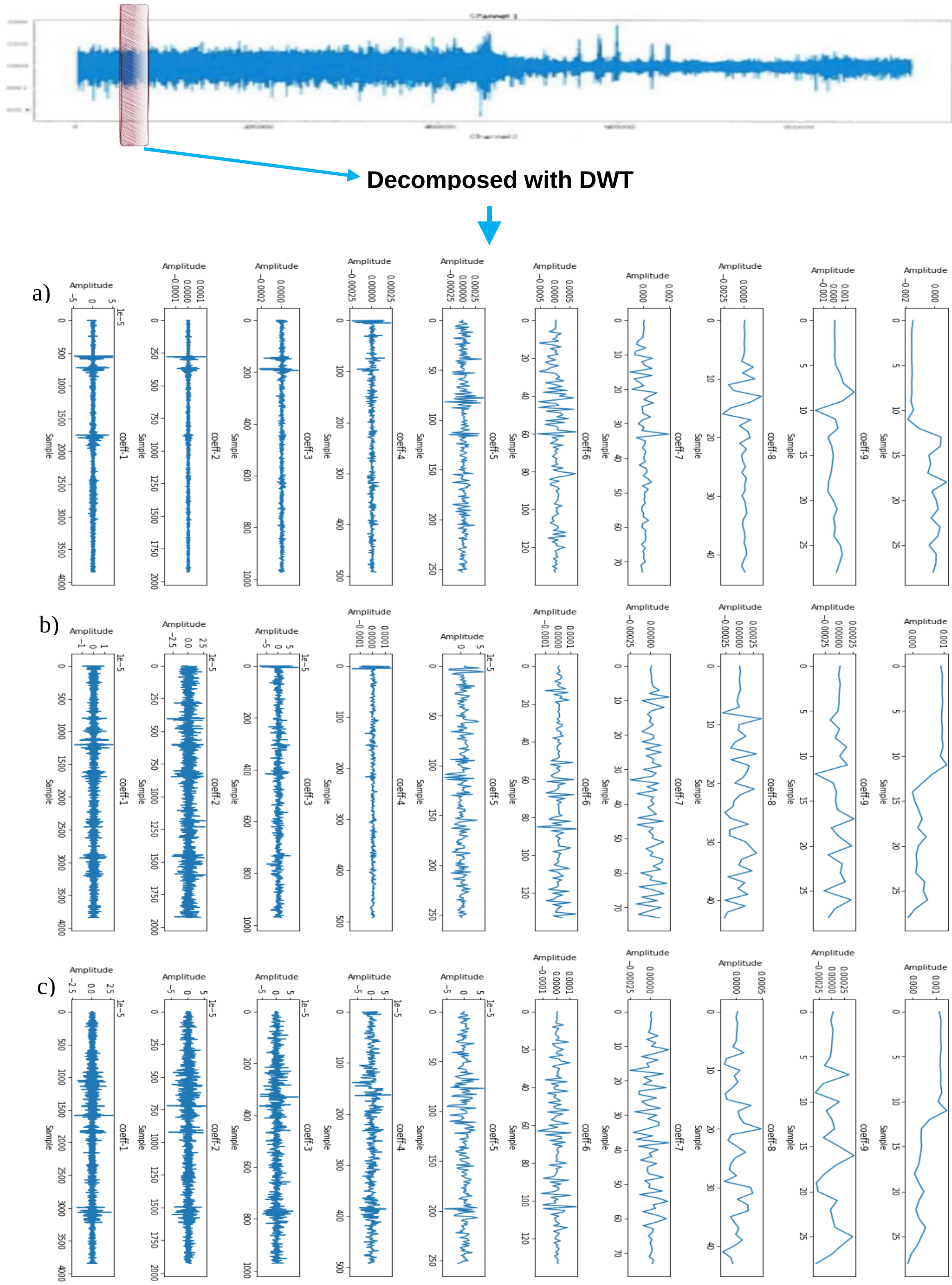
Extended Data Fig. 2 | First three channels from PN00-1 file: We have displayed the EEG file PN00-1 from the Siena Scalp EEG Database. This particular file consists of 34 channels, capturing electrical activity from various regions of the scalp. For visualization purposes, we have chosen to display the first three channels of the EEG file.



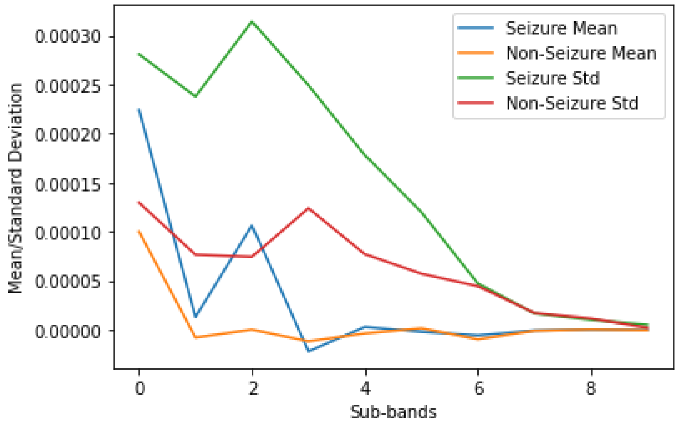
Extended Data Fig. 3 | Division of EEG signals into sequences of overlapping windows of 30 seconds with a 30% overlapping window: The use of a 30-second window with a 30% overlap on multi-channel EEG signals has been chosen to strike a balance between preserving temporal information and minimizing the impact of windowing artifacts. The 30% overlapping ratio ensures that the signal chunks retain continuity, thereby reducing the likelihood of information loss at the edges of the windows.



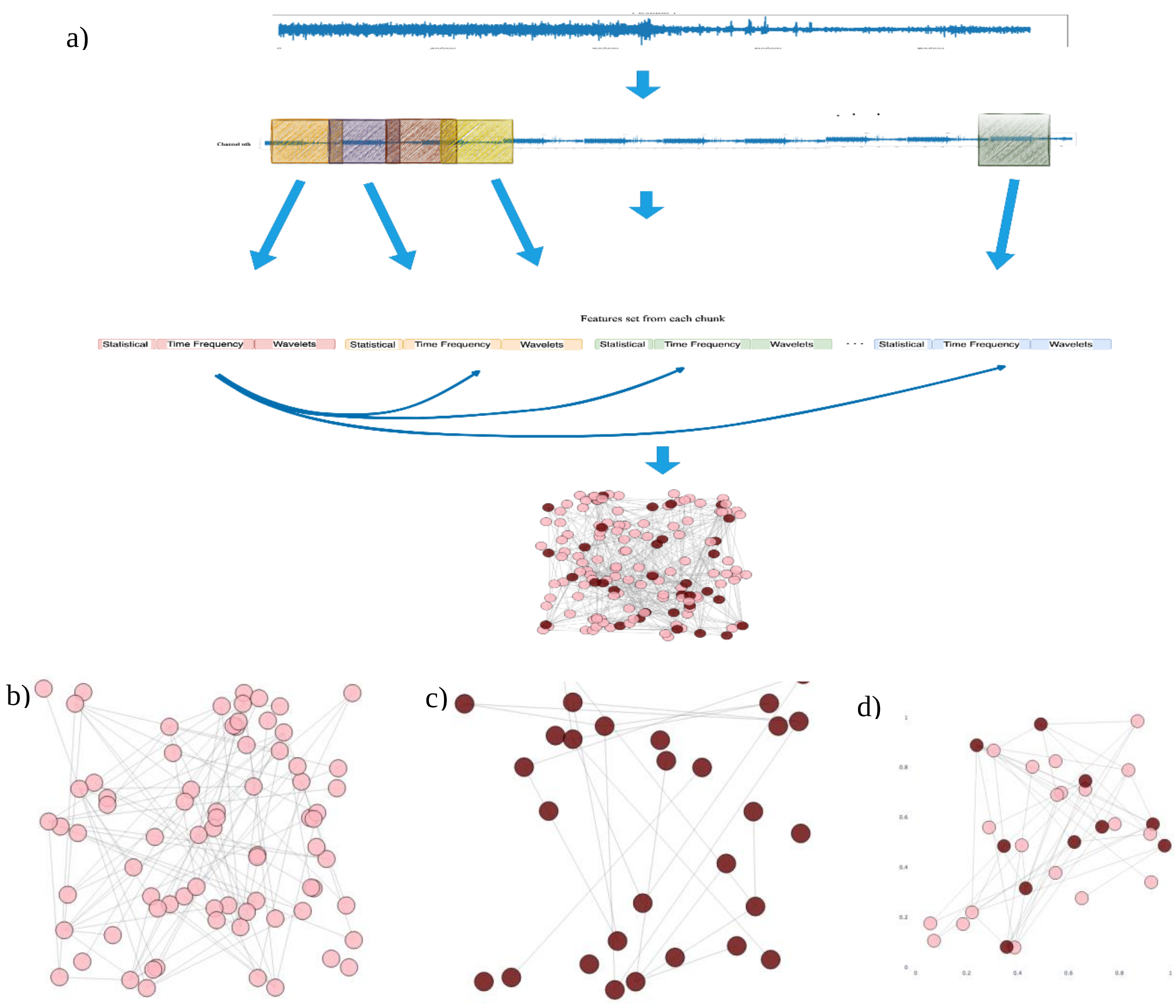
Extended Data Fig. 4 | Coefficient Visualization with normal Signal. (a) CHB-MIT-EEG. **(b)** SSE-EEG. **(c)** TUH-EEG. We conducted an analysis on three datasets using the Daubechies wavelet 8db. The figure shows the results for the non-seizure windows. For each frequency sub-band obtained through the Daubechies wavelet, we created separate subplots arranged in descending order of frequency content. Each subplot is labeled with the number of coefficients being plotted, while the x-axis is labeled as "Sample" and the y-axis as "Amplitude."



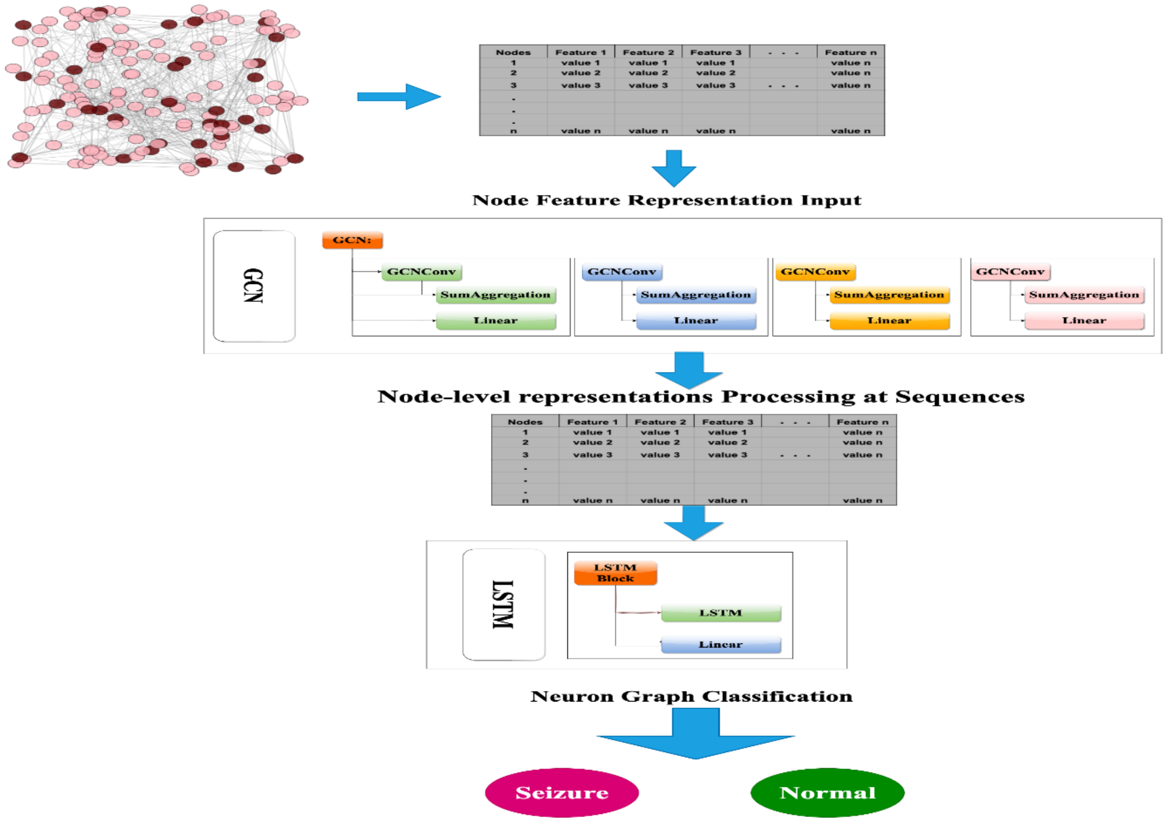
Extended Data Fig. 5 | Coefficient Visualization with Seizure Signal. (a) CHB-MIT-EEG. **(b)** SSE-EEG. **(c)** TUH-EEG. We conducted an analysis on three datasets using the Daubechies wavelet 8db. The figure shows the results for the seizure windows. For each frequency sub-band obtained through the Daubechies wavelet, we created separate subplots arranged in descending order of frequency content. Each subplot is labeled with the number of coefficients being plotted, while the x-axis is labeled as "Sample" and the y-axis as "Amplitude."



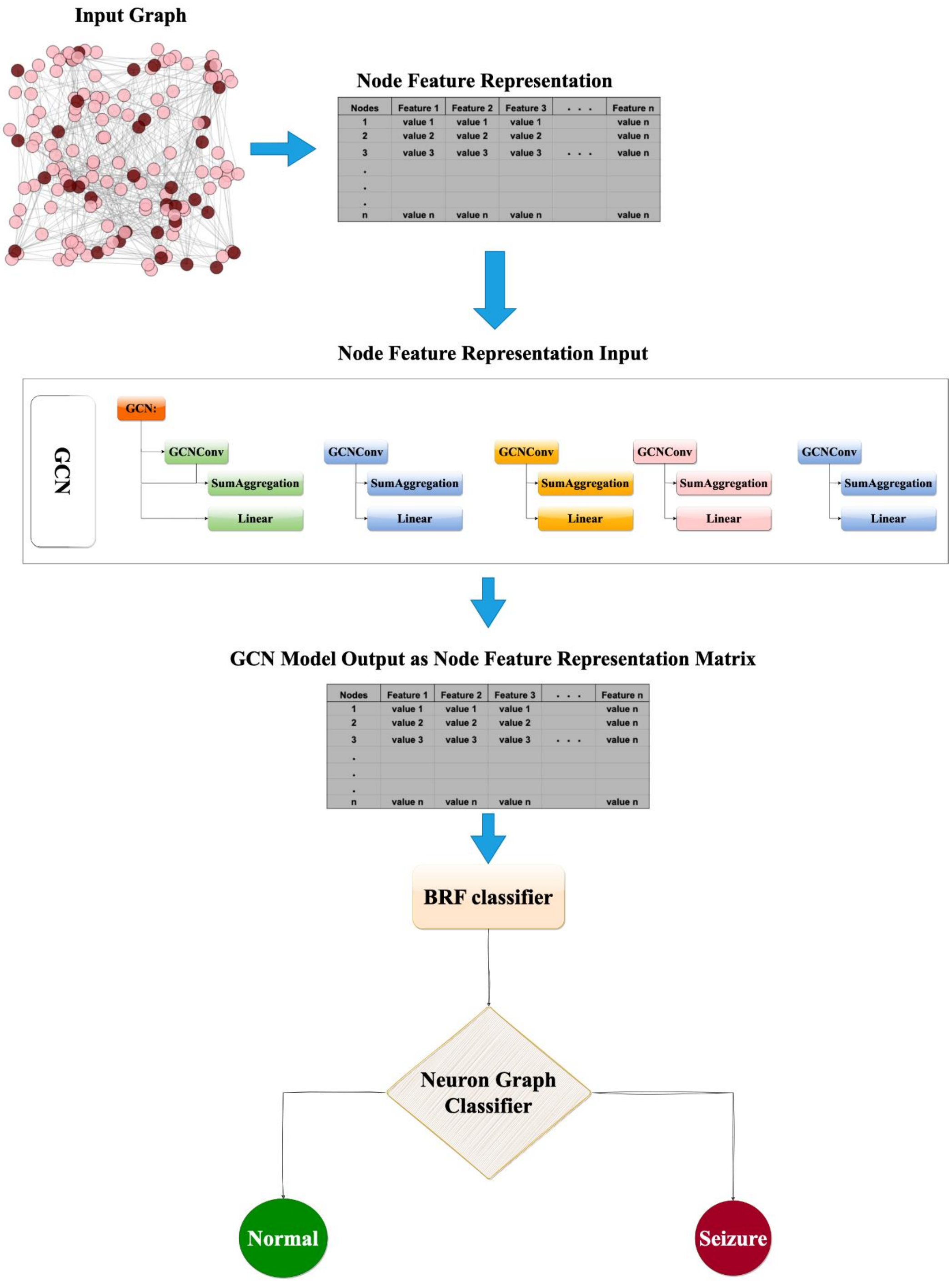
Extended Data Fig. 6 | Distinction of Seizure and Non-Seizure EEG Signals through the Mean and Standard Deviation of Energy from sub-bands using the Daubechies Wavelet Transform. The EEG signals were subjected to the wavelet transform to capture their distinctive features. The wavelet coefficients were analyzed in two separate windows, each representing seizure and non-seizure signals, to determine their mean and standard deviation.



Extended Data Fig. 7 | Graph representation. (a) Process of converting Features set into Graphs representation. **(b)** Graph nodes classification for normal signal. **(c)** Graph nodes classification for seizure signal. **(d)** Graph nodes classification for sample from Seizure and Normal.



Extended Data Fig. 8 | GCN-LSTM model diagram: The GCN component uses node features and edge connectivity to create node-level representations of the graph's structure. It applies non-linear activation functions through layers to capture both local and global information. The LSTM component takes these node-level representations and processes them over time, using gates to regulate information flow and remembering past representations. The final hidden state summarizes the sequence. A linear layer transforms the hidden state to generate the final prediction based on the graph data.



Extended Data Fig. 9 | GCN-BRF model diagram: The GCN-BRF model combines the strengths of GCN for learning graph representations and BRF for classification, resulting in enhanced performance on graph classification tasks. The model initially employs the GCN operation to capture temporal dependencies within the output features. Ultimately, the BRF classifier is utilized to generate the final prediction.