

Deep Learning-Based EEG Seizure Classification Using Comparative Time–Frequency Spectral Representations.

Praveen Mewada¹, Siddhartha K M¹, Sriram Kumar P^{2*}, Ryan Gabriel¹,
VVSN Sainidhi², Rajamanickam YUVARAJ³, John THOMAS⁴, Karthik PA⁵,
Jac Fredo Agastinose Ronickom¹

¹Computational Neuroscience and Biology Lab, School of Biomedical Engineering, Indian Institute of Technology (BHU) Varanasi, Varanasi, 221005, Uttar Pradesh, India.

^{2*}Symbiosis Institute of Technology Hyderabad Campus, Symbiosis International (Deemed University), Pune, Hyderabad, 509217, Telangana, India.

³National Institute of Education (NIE), Nanyang Technological University (NTU), Singapore, 637616, , Singapore.

⁴Montreal Neurological Institute, McGill University, Montreal, 637616, Quebec, Canada.

⁵Department of Sensor and Biomedical Technology, National Institute of Technology Tiruchirappalli, Tiruchirappalli, 620015, Tamilnadu, India.

*Corresponding author(s). E-mail(s): sriramkumar@sithyd.siu.edu.in;

Contributing authors: pravenmewada.rs.bme23@itbhu.ac.in;

km.siddhartha.bme19@itbhu.ac.in; bismuthlover1404@gmail.com;

Vinnakota.sainidhi.phd2025@sitpune.edu.in; yuvaraj.rajamanickam@nie.edu.sg;

johnthomasnitc@gmail.com; sriramkumar@sithyd.siu.edu.in; jack.bme@iitbhu.ac.in;

Supplementary file

S.1 Statistical Significance Analysis

We performed pairwise Wilcoxon signed-rank tests using the fold-wise performance values of all evaluated deep learning models for the three time-frequency representations (TFRs), namely CWT, Mel-spectrogram, and STFT. For each TFR representation, nine deep learning models were evaluated, resulting in 36 unique pairwise comparisons. Thus, the statistical analysis was independently performed for CWT, Mel-spectrogram, and STFT. To account for the multiple pairwise comparisons, the resulting p -values were adjusted using the Holm correction to control the family-wise error rate. Across the three TFR representations, several model comparisons yielded raw p -values below the conventional significance level of 0.05. For the CWT representation, significant differences were observed in several pairwise comparisons before correction, particularly involving AlexNet vs cLSTM, cRNN, Swin Transformer and ResBiLSTM; similarly for cLSTM Vs cCNN, CNN-BiLSTM, CNN-LSTM, cRNN, Swin Transformer; cCNN vs cRNN, Swin Transformer, ResBiLSTM; CNN-BiLSTM vs cRNN, SwinTransformer, ResBiLSTM; CNN-LSTM vs cRNN, ResBiLSTM; Swin Transformer vs ResBiLSTM. Similarly, for the Mel-spectrogram representation, several pairwise comparisons yielded raw p -values below 0.05 like AlexNet vs cLSTM, cCNN, CNN-BiLSTM, CNN-LSTM, ResBiLSTM; cLSTM vs VGGNet, cCNN, cRNN, Swin Transformer; cCNN vs CNN-BiLSTM, CNN-LSTM, ResBiLSTM; CNN-BiLSTM vs Swin Transformer; CNN-LSTM vs Swin Transformer; Swin Transformer vs ResBiLSTM. In the STFT representation, a smaller number of comparisons showed raw p -values below 0.05 like AlexNet vs cLSTM, VGGNet, cCNN, CNN-BiLSTM, CNN-LSTM, ResBiLSTM; cLSTM vs cCNN, cRNN. However, none of the pairwise comparisons remained statistically significant after Holm correction at $\alpha = 0.05$ across any of the three TFR representations. Therefore, although differences in model performance were observed across CWT, Mel-spectrogram, and STFT, these differences cannot be considered statistically significant after accounting for multiple comparisons.

Table 1: Pairwise Wilcoxon Signed-Rank Test Results for CWT (Raw p -value / Holm-corrected p -value)

Model	Alex Net	cLSTM	VGG NET	cCNN	CNN- BiLSTM	CNN- LSTM	cRNN	Swin Transformer	Res BiLSTM
AlexNet	–	0.001953 / 0.087891	0.083984 / 1.000000	0.980469 / 1.000000	0.845703 / 1.000000	0.476562 / 1.000000	0.003906 / 0.140625	0.001953 / 0.087891	0.001953 / 0.087891
cLSTM	–	–	0.164062 / 1.000000	0.001953 / 0.087891	0.001953 / 0.087891	0.013672 / 0.423828	0.037109 / 0.927734	0.007812 / 0.250000	0.845703 / 1.000000
VGGNET	–	–	–	0.193359 / 1.000000	0.275391 / 1.000000	0.431641 / 1.000000	0.757812 / 1.000000	0.910156 / 1.000000	0.193359 / 1.000000
cCNN	–	–	–	–	0.734375 / 1.000000	0.652344 / 1.000000	0.003906 / 0.140625	0.027344 / 0.738281	0.001953 / 0.087891
CNN- BiLSTM	–	–	–	–	–	0.941406 / 1.000000	0.019531 / 0.585938	0.019531 / 0.585938	0.003906 / 0.140625
CNN-LSTM	–	–	–	–	–	–	0.048828 / 1.000000	0.193359 / 1.000000	0.003906 / 0.140625
cRNN	–	–	–	–	–	–	–	0.431641 / 1.000000	0.083984 / 1.000000
Swin Transformer	–	–	–	–	–	–	–	–	0.048828 / 1.000000
ResBiLSTM	–	–	–	–	–	–	–	–	–

Table 2: Pairwise Wilcoxon Signed-Rank Test Results for MEL (Raw p -value / Holm-Corrected p -value)

Model	Alex Net	cLSTM	VGG NET	cCNN	CNN- BiLSTM	CNN- LSTM	cRNN	Swin Transformer	Res BiLSTM
AlexNet	–	0.007812 / 0.320312	0.300781 / 1	0.013672 / 0.505859	0.003906 / 0.171875	0.001953 / 0.087891	0.193359 / 1	0.492188 / 1	0.001953 / 0.087891
cLSTM	–	–	0.037109 / 1	0.001953 / 0.087891	0.769531 / 1	0.845703 / 1	0.048828 / 1	0.013672 / 0.505859	0.625000 / 1
VGGNET	–	–	–	0.492188 / 1	0.193359 / 1	0.083984 / 1	0.910156 / 1	0.625000 / 1	0.232422 / 1
cCNN	–	–	–	–	0.003906 / 0.171875	0.001953 / 0.087891	0.322266 / 1	0.375000 / 1	0.005859 / 0.251953
CNN- BiLSTM	–	–	–	–	–	0.769531 / 1	0.193359 / 1	0.019531 / 0.703125	0.695312 / 1
CNN-LSTM	–	–	–	–	–	–	0.130859 / 1	0.009766 / 0.390625	0.845703 / 1
cRNN	–	–	–	–	–	–	–	0.492188 / 1	0.275391 / 1
Swin Transformer	–	–	–	–	–	–	–	–	0.009766 / 0.390625
ResBiLSTM	–	–	–	–	–	–	–	–	–

Table 3: Pairwise Wilcoxon Signed-Rank Test Results for STFT (Raw p -value / Holm-Corrected p -value)

Model	Alex Net	cLSTM	VGG NET	cCNN	CNN- BiLSTM	CNN- LSTM	cRNN	Swin Transformer	Res BiLSTM
AlexNet	–	0.001953 / 0.087891	0.001953 / 0.087891	0.048828 / 1	0.001953 / 0.087891	0.027344 / 0.929688	0.130859 / 1	0.083984 / 1	0.019531 / 0.703125
cLSTM	–	–	0.695312 / 1	0.048828 / 1	0.910156 / 1	0.845703 / 1	0.027344 / 0.929688	0.130859 / 1	0.193359 / 1
VGGNET	–	–	–	0.492188 / 1	0.695312 / 1	0.910156 / 1	0.322266 / 1	0.431641 / 1	0.625 / 1
cCNN	–	–	–	–	0.193359 / 1	0.769531 / 1	0.625 / 1	0.845703 / 1	0.625 / 1
CNN- BiLSTM	–	–	–	–	–	0.769531 / 1	0.130859 / 1	0.130859 / 1	0.232422 / 1
CNN-LSTM	–	–	–	–	–	–	0.431641 / 1	0.492188 / 1	0.625 / 1
cRNN	–	–	–	–	–	–	–	0.910156 / 1	0.492188 / 1
Swin Transformer	–	–	–	–	–	–	–	–	0.625 / 1
ResBiLSTM	–	–	–	–	–	–	–	–	–

S.2 Computational Time Analysis

We conducted computational-time analysis for the three TFRs considered in this study, namely CWT, Mel Spectrogram, and STFT, in combination with the evaluated deep learning architectures. For each TFR-model combination, the training time, testing time, and average inference time per sample were recorded using the Nested 10-fold cross-validation (10 outer folds and 5 inner folds) strategy. The results are presented in Table 4. The results demonstrate substantial differences in computational requirements among the evaluated architectures. It is observed that, for the CWT representation, ResBiLSTM requires the least training time of 162.14 s, whereas cLSTM requires the highest training time of 20,809.6 s. In terms of test time, cRNN achieves the lowest time of 0.2082 s, while VGGNet records the highest test time of 65.598 s. Similarly, cRNN has the lowest average inference time of 0.465 ms/sample, whereas VGGNet has the highest inference time of 145.77 ms/sample. For the Mel representation, cRNN requires the least training time of 82.34 s, while Swin Transformer requires the highest training time of 879.58 s. Regarding test time, cRNN achieves the lowest value of 0.1915 s, whereas VGGNet records the highest value of 69.683 s. In terms of average inference time, cRNN exhibits the lowest value of 0.403 ms/sample, while VGGNet has the highest value of 154.85 ms/sample. For the STFT representation, AlexNet requires the least training time of 116.9 s, whereas VGGNet requires the highest training time of 23,792.6 s. The lowest test time is achieved by cRNN at 0.088 s, while VGGNet records the highest test time of 38.602 s. Similarly, cRNN has the lowest average inference time of 0.511 ms/sample, whereas VGGNet exhibits the highest inference time of 85.78 ms/sample. From The results, we conclude that cRNN followed by cCNN and ResBiLSTM provides a favorable balance between classification capability and computational efficiency among the evaluated architectures. Its low inference latency is particularly relevant to potential real-time seizure monitoring applications. This information has been incorporated into the supplementary material of the revised manuscript.

Table 4: Computational time analysis of the evaluated deep learning models using different time–frequency representations (TFRs).

TFR	Model	Training Time (s)	Test Time (s)	Avg Time/Sample (ms)
CWT	AlexNet	6050.6	17.757	39.46
	cLSTM	20809.6	25.713	57.14
	VGGNET	17306	65.598	145.77
	cRNN	124.29	0.2082	0.465
	cCNN	252.81	0.4864	1.085
	CNN-LSTM	535.08	1.2481	2.789
	CNN-BiLSTM	477.48	1.2731	2.842
	ResBiLSTM	162.14	0.434	0.969
	Swin-Transformer	885.66	2.028	4.531
Mel	AlexNet	559.6	18.852	41.89
	cLSTM	1668.2	27.297	60.66
	VGGNET	1642	69.683	154.85
	cRNN	82.34	0.1915	0.403
	cCNN	278.73	0.5072	1.066
	CNN-LSTM	506.29	1.1881	2.499
	CNN-BiLSTM	497.42	1.226	2.579
	ResBiLSTM	217.13	0.4655	0.979
	Swin-Transformer	879.58	2.0235	4.256
STFT	AlexNet	116.9	4.693	10.43
	cLSTM	7407.3	5.555	12.34
	VGGNET	23792.6	38.602	85.78
	cRNN	203.94	0.088	0.511
	cCNN	390.28	0.2129	1.233
	CNN-LSTM	963.38	0.4503	2.607
	CNN-BiLSTM	1012.92	0.4678	2.707
	ResBiLSTM	394.1	0.1999	1.159
	Swin-Transformer	1468.24	0.7581	4.392