Classification of EEG signals using Transformer based deep learning and ensemble models

Mahsa Zeynali [a], Hadi Seyedarabi [a,*], Reza Afrouzian [b]

[a] Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran
[b] Faculty of Engineering, University of Tabriz, Miyaneh, Iran

인공지능응용학과
mariem kallel 23512051

BrAIn Lab.
Brain and Artificial Intelligence Lab.

Background and motivation

- Brain Computer interface (BCI): communication and control system provides interaction between a user and computer device.
- Previous studies used EEG signal in several BCI applications.
- A model which accurately classify EEG signal in BCI applications ?

Authors contribution:

- Self-collected dataset.
- 2 types of transformer were used for EEG classification task.
- Ensemble based model is used to enhance the performance of the previous model.

Materials and Methods

BrAIn Lab.
Brain and Artificial Intelligence Lab.

- Data was recorded from 8 subjects (4W , 4 M).
- 64 channel cap based on the international 10-20 system.
- Sampling rate = 250 Hz.
- Visual stimuli: Experiment has 5 runs:
  - run 1 : 50 images of Category 1(C1 = chair).
  - run 2 : 50 images of C2(headphone).
  - run 3 : 50 images of C3(mug).
  - run 4 : 50 images of C4(backpack).
  - run 5 : 50 images of C5(guitar).
- To keep subject attention → record their feedback during the experiment.
- Feedback task is : hit button when subject see a repeating image.

Experiment setup:

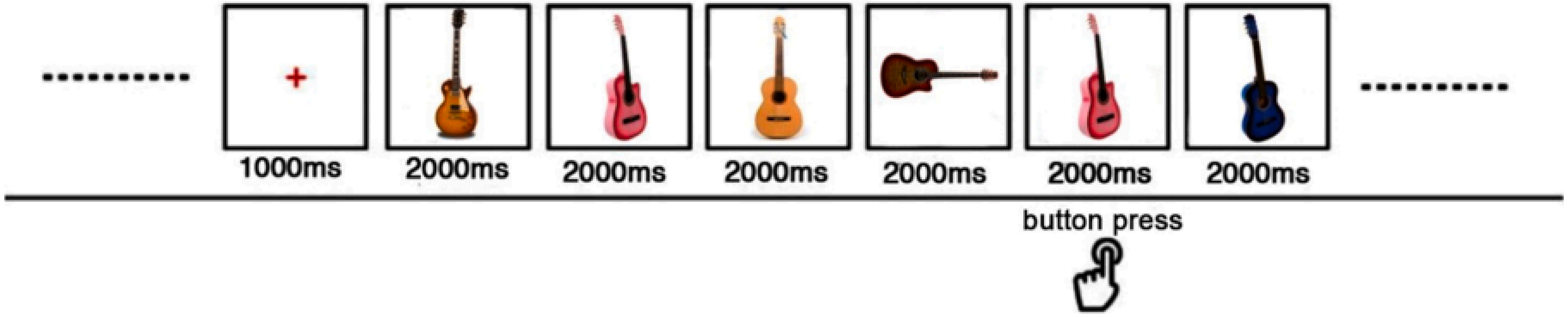


Fig. 1. Image presentation experiment.

- 50 images * 5 categories * 8 participants = 2000 samples.
- Dataset is divided following 8:1:1 ratios. (1600 + 200 + 200).

Preprocessing and Feature Extraction

BrAIn Lab.
Brain and Artificial Intelligence Lab.

- Preprocessing steps:
  - Epoch extraction using event markers.
  - Elimination of noise and artifacts using ICA.
  - 0.5 - 50 Hz low pass filter was applied.
  - Z-score normalization of data.
- Feature Extraction:
  - Power Spectral Density (was used later for spectral transformer).
  - Extraction of 5 EEG signal frequency bands and 3 statistical features for each band.(to compare results of ML model with transformer).
  - Statistical features are: avg absolute value/ entropy/ standard deviation.

Classification models:

- In this paper, authors implemented: SVM / MLP / GRU / 2 types of transformer / ensemble based model.
- SVM: RBF kernel function ($\gamma = 0.0009$ and $C = 0.8$).
- MLP: 1 hidden layer with 10 neurons.
- Gated recurrent unit:

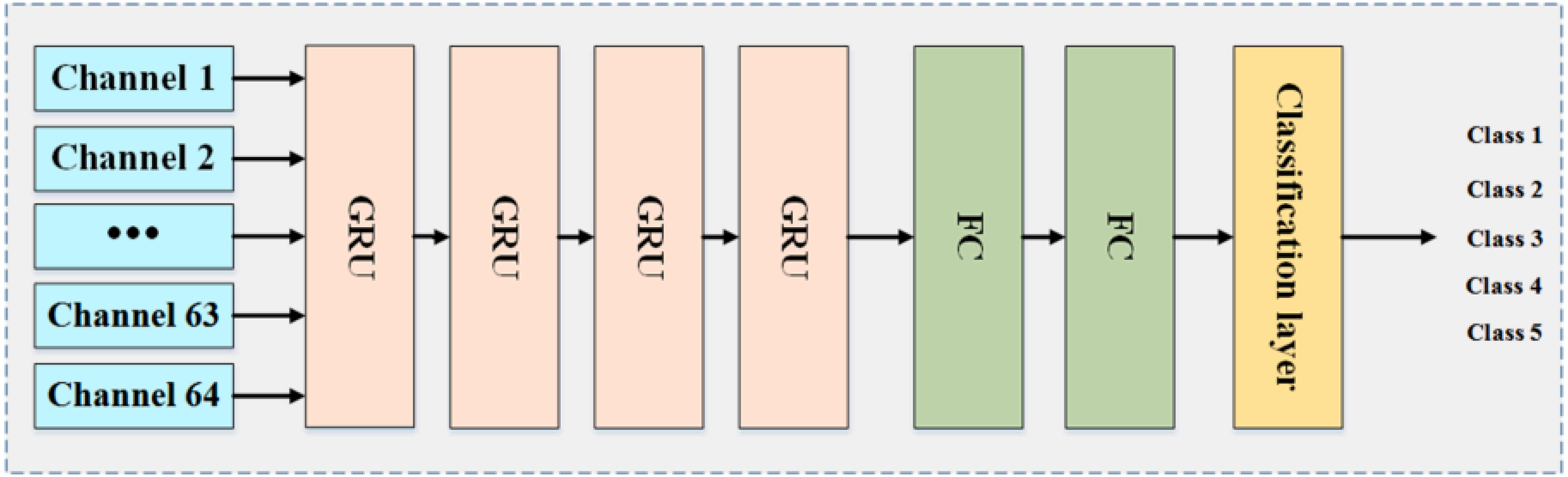


Fig. 3. The architecture of the proposed GRU-based network.

Transformers

- Positional encoding

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d}}}\right)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i+1}{d}}}\right)$$

- Encoding block: The **p** **alization** differs from the vanilla transformer.
- Anything else remains the same(multi head attention concept).

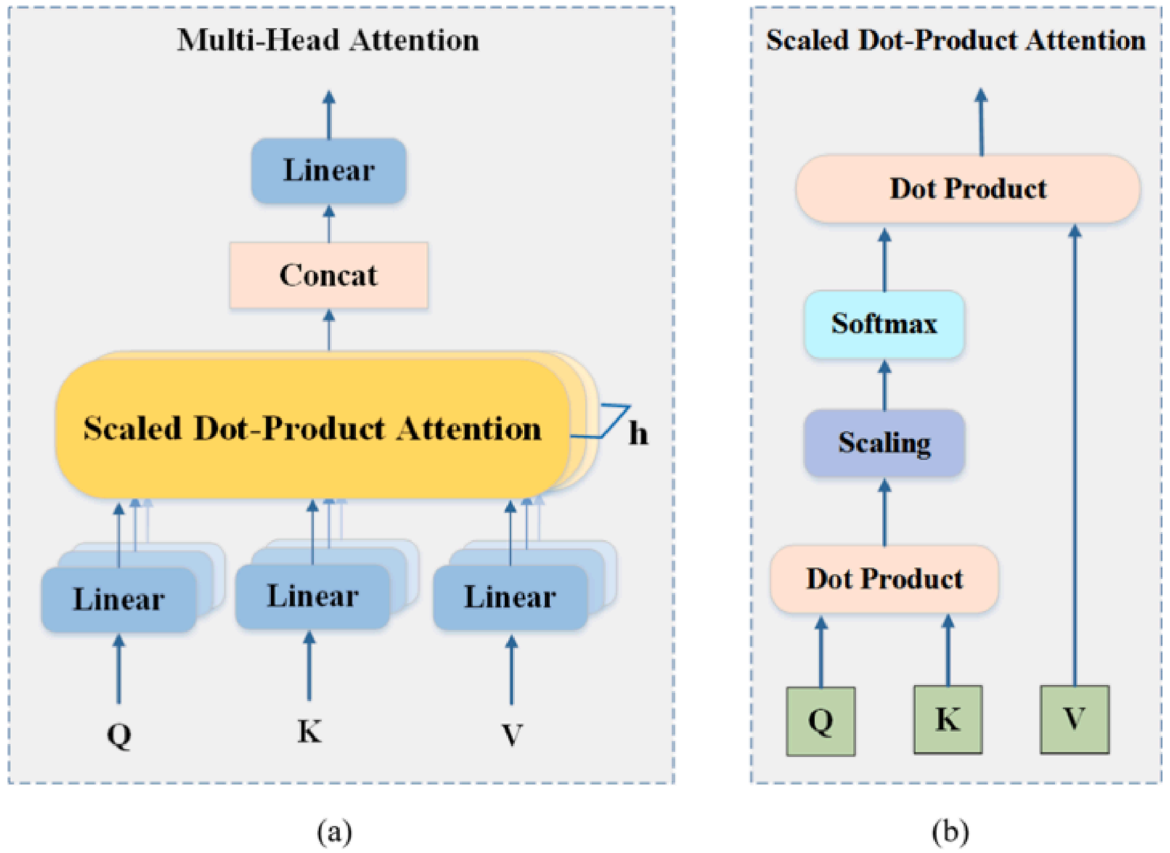
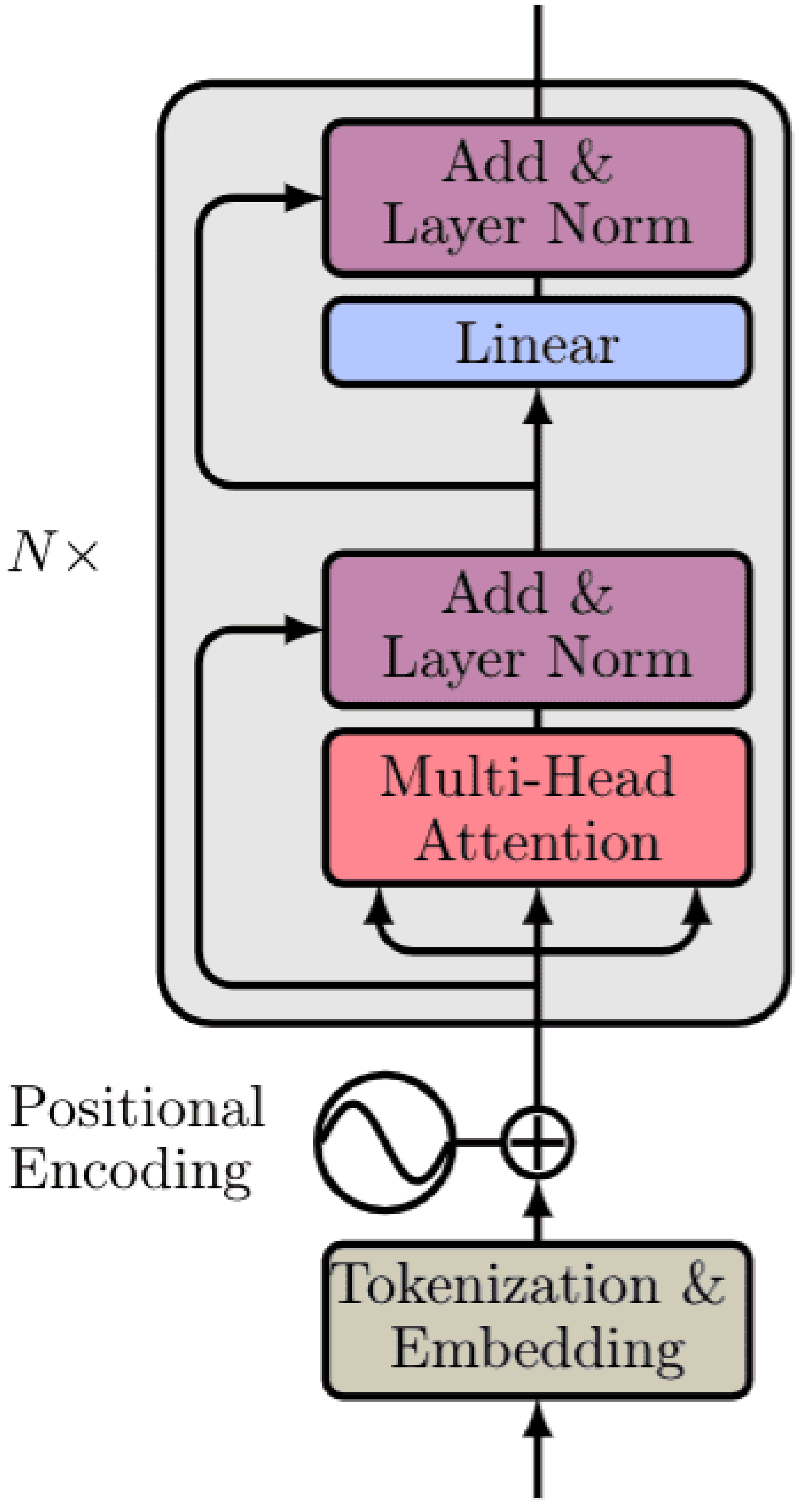
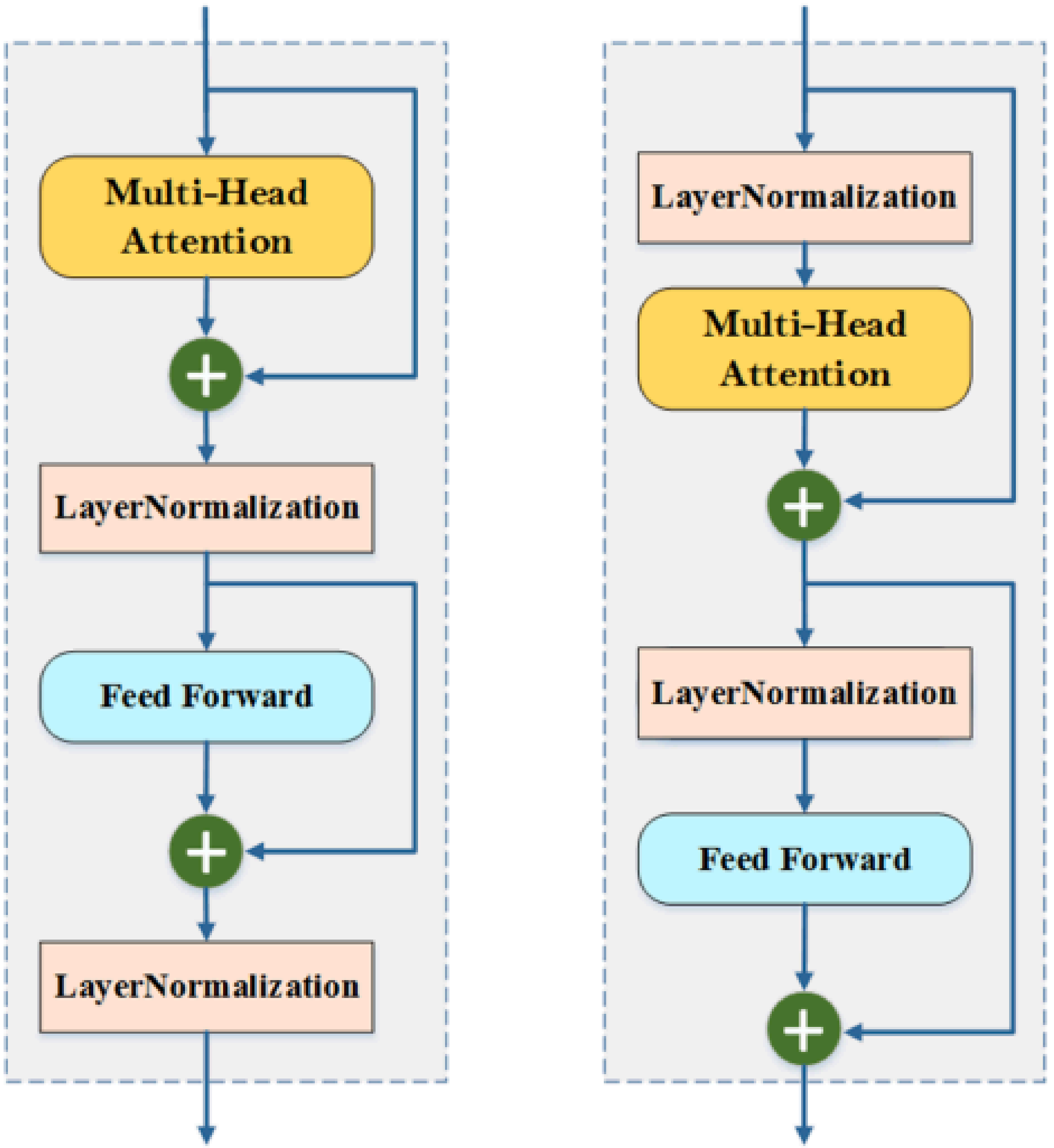


Fig. 5. The structure of the encoding block. a) Multi-head Attention; b) Scaled Dot-Product Attention.

Transformers: 2 types





(a) (b)

Architecture of the proposed model:

BrAIn Lab.
Brain and Artificial Intelligence Lab.

- To extract temporal dependencies of EEG:
  → temporal transformer was used.
- To extract spatial dependencies of EEG:
  → Spectral transformer was used.

In total, authors experimented with 4 transform
TempTrans pre / post LN.
SpaTrans pre / post LN.

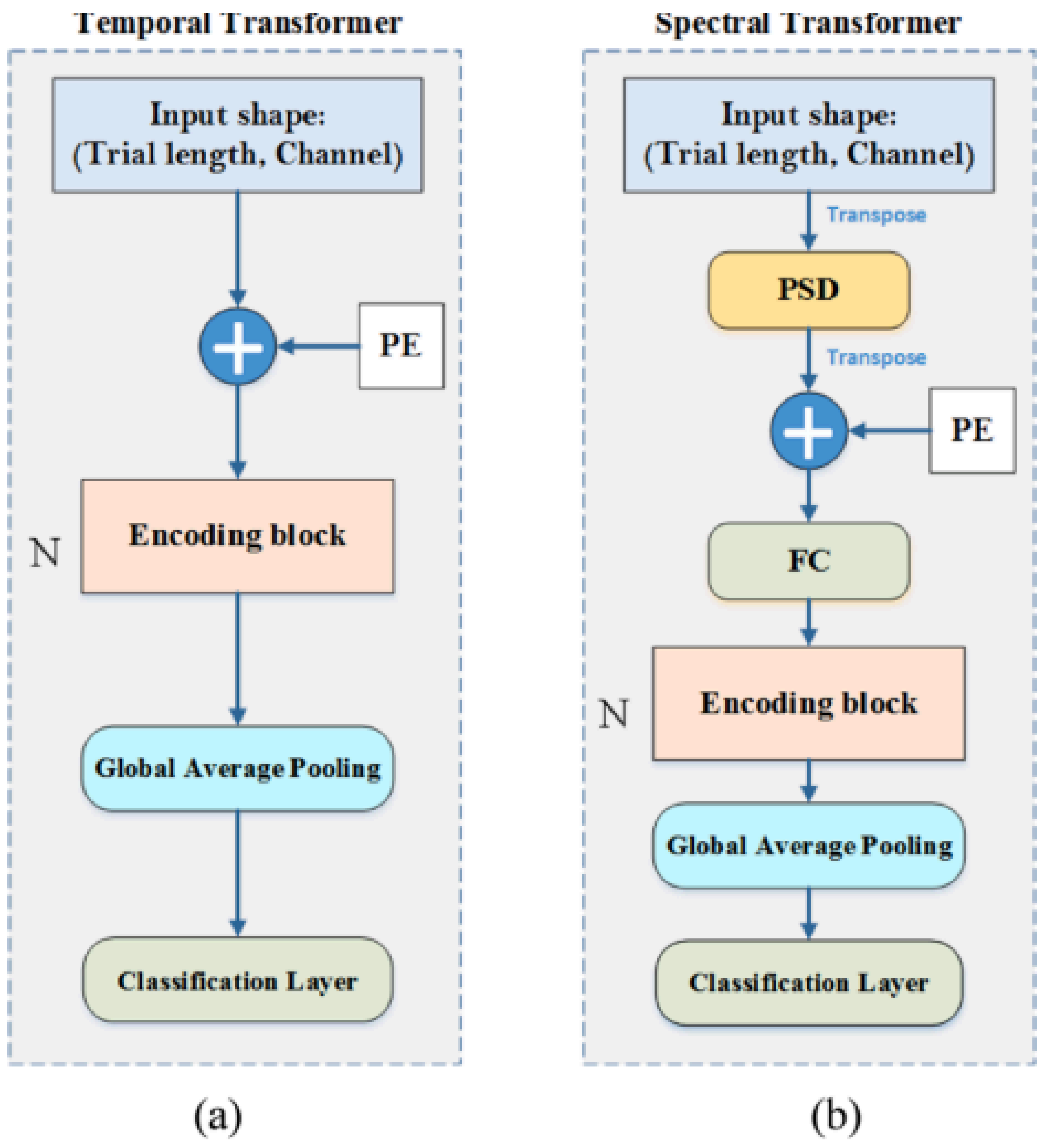
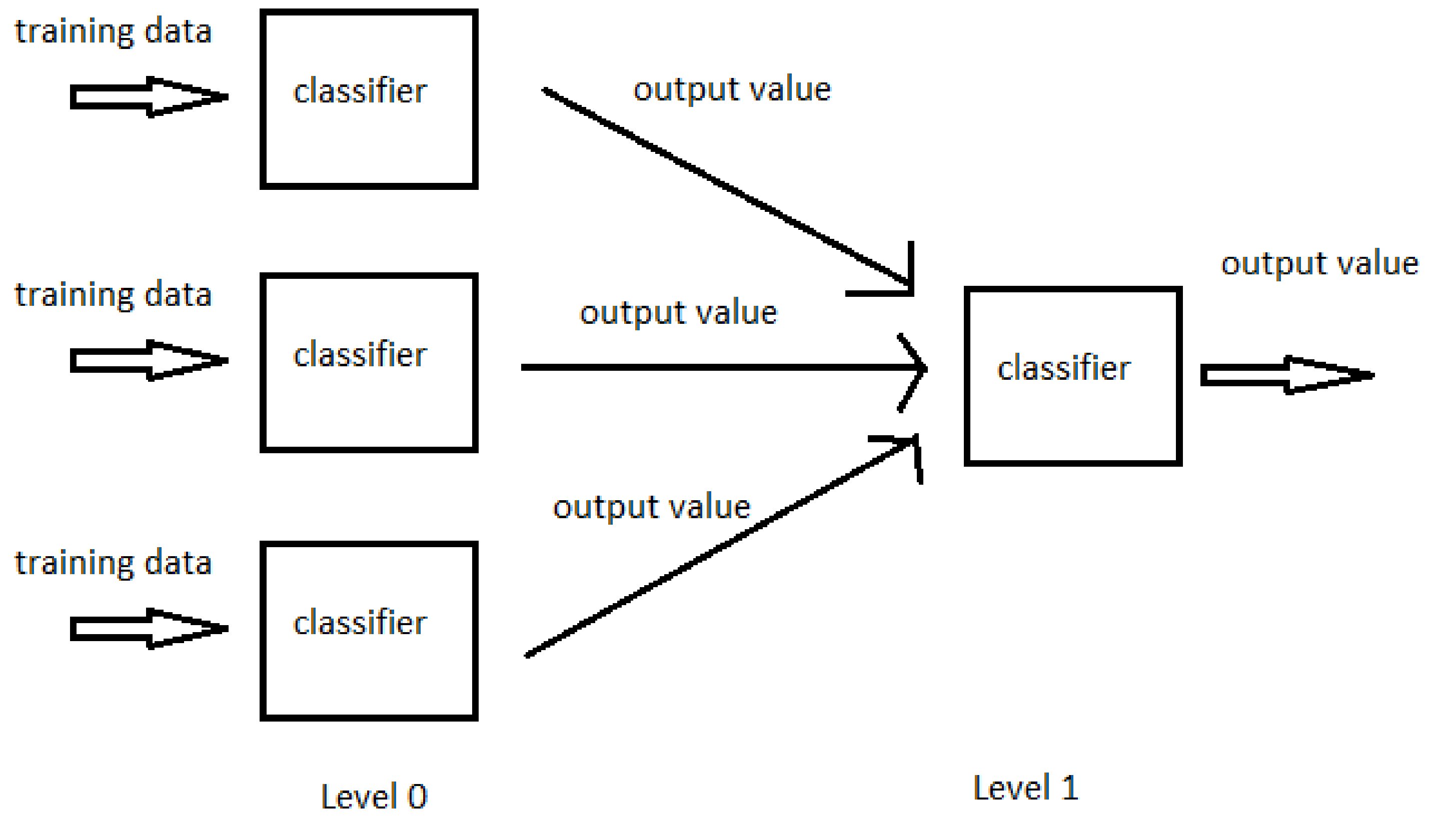


Fig. 6. Proposed Transformer based model. a) Temporal Transformer; b) Spectral Transformer.

Architecture of the proposed model:

BrAIn Lab.
Brain and Artificial Intelligence Lab.

- To enhance the performance of EEG classification task using Transformers, ensemble model is proposed.
- In this paper, 2 methods were employed:
    a. Unweighted model averaging majority voting.
    b. Stacked Generalization method.



L0: 2 transformers L1: Multi class Logistic Regression

Results and Discussion:

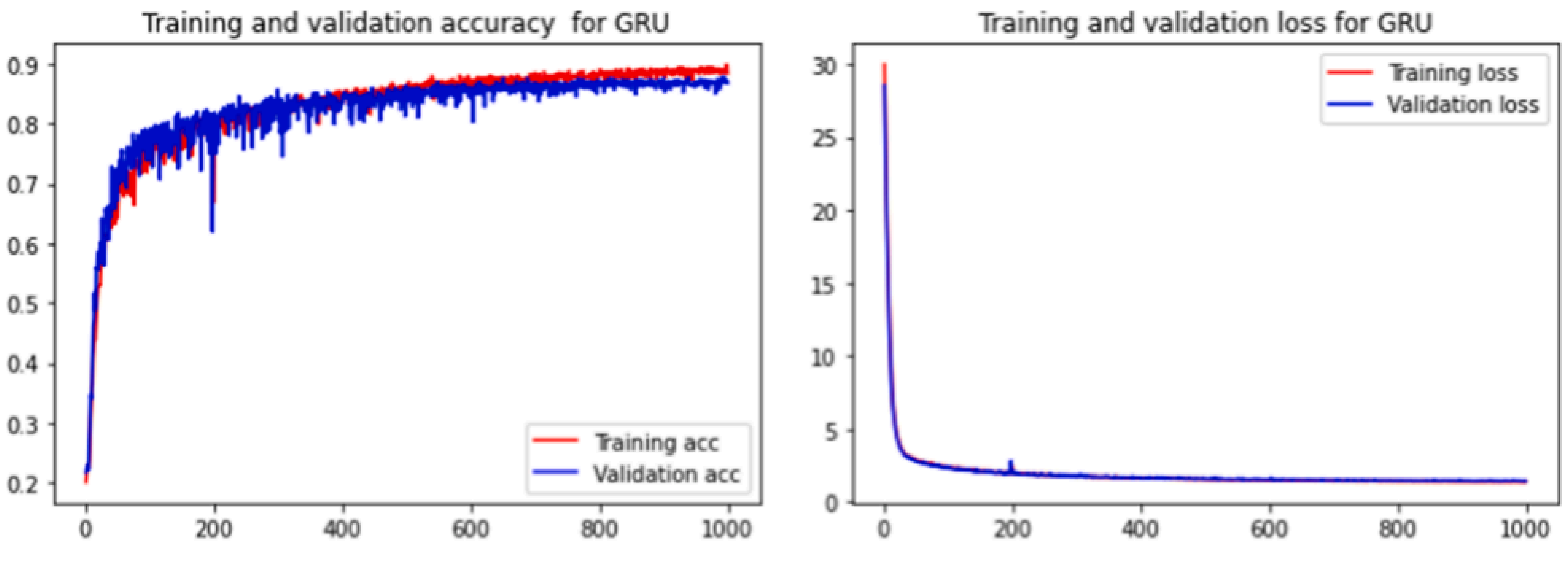
BrAIn Lab.
Brain and Artificial Intelligence Lab.

1. Hand crafted feature extraction + SVM / MLP.

Table 1
The results obtained for the MLP and SVM classification algorithm.

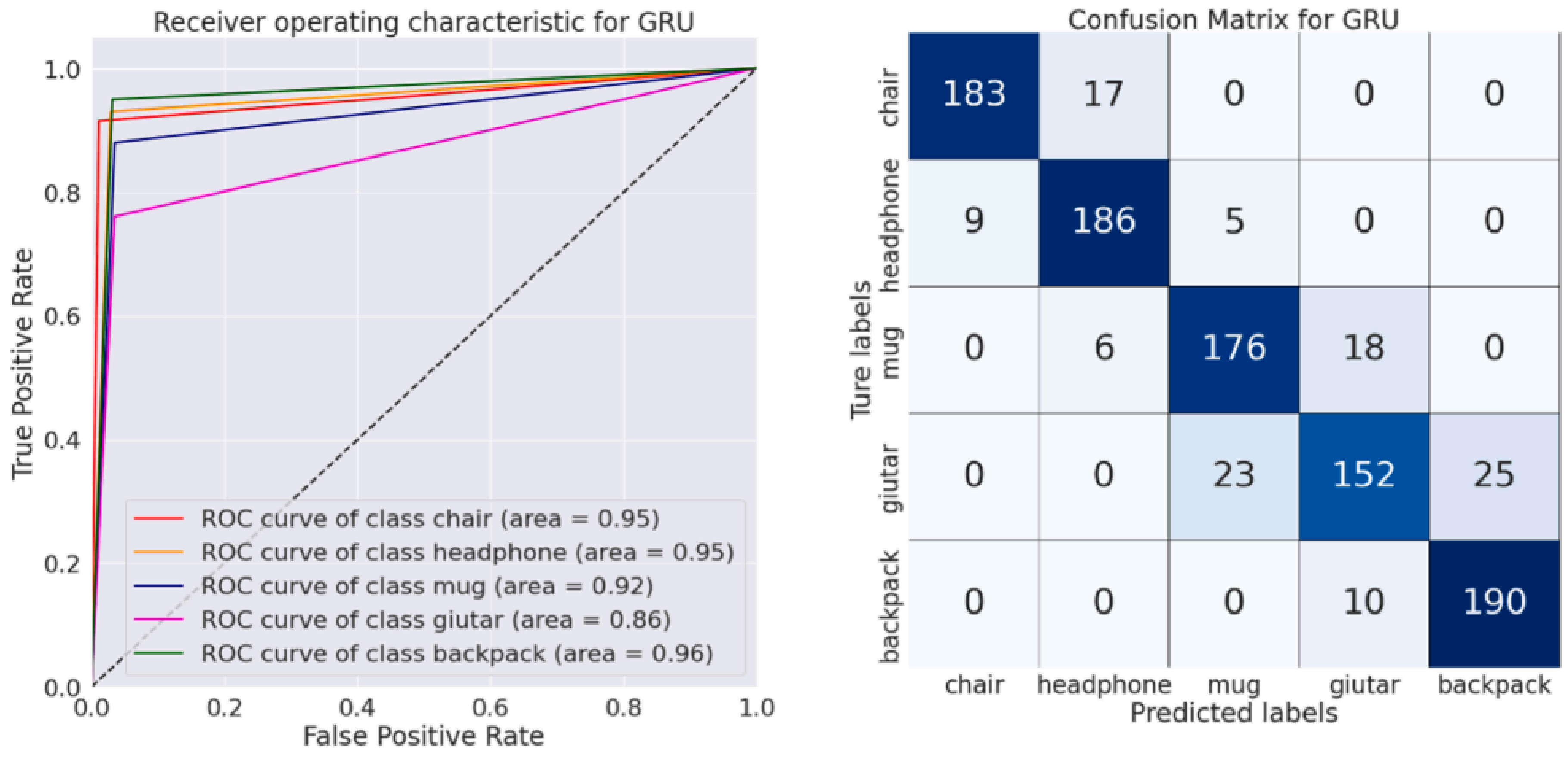
Classifier	Accuracy%	AUC	F-score
SVM	69	0.80	0.69
MLP	81	0.88	0.81

2. GRU: 88.70% acc and 0.88 F1-score.



(a) (b)

Fig. 8. Training results of the proposed GRU network using training and validation dataset. a) Accuracy curve; b) loss curve.

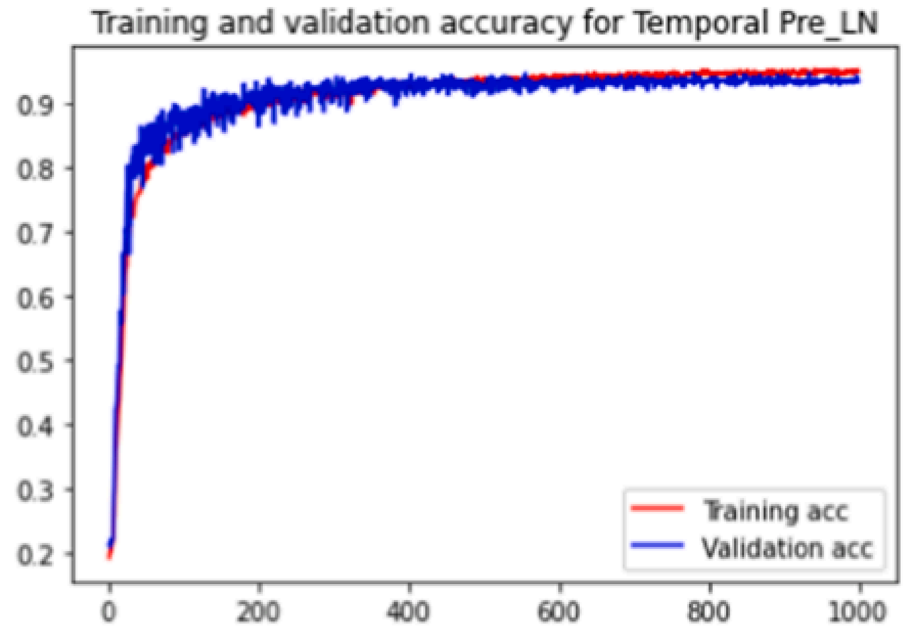


(a) (b)

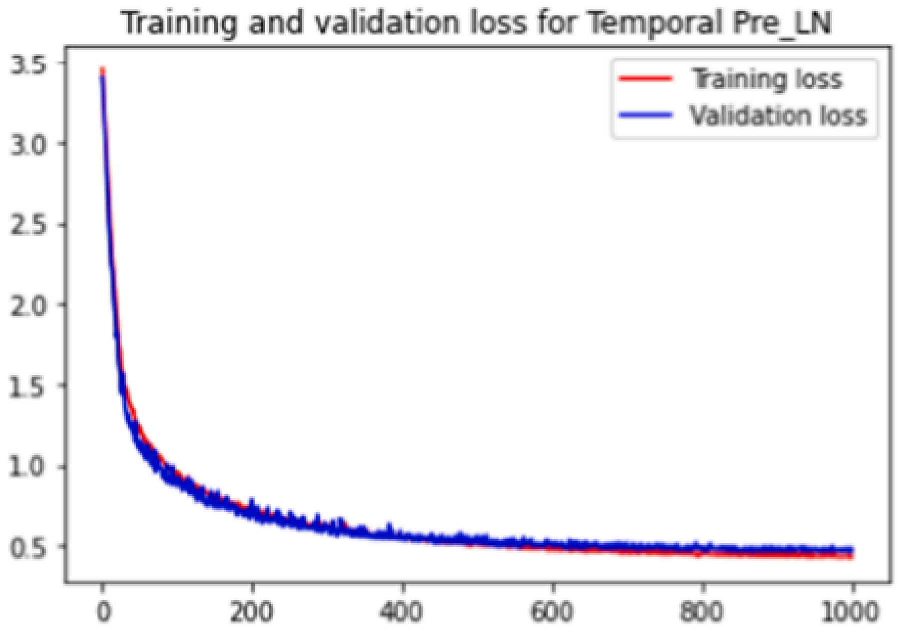
Fig. 9. Confusion matrix and ROC curve using GRU for the test dataset. a) ROC curve; b) Confusion matrix.

Results and Discussion:

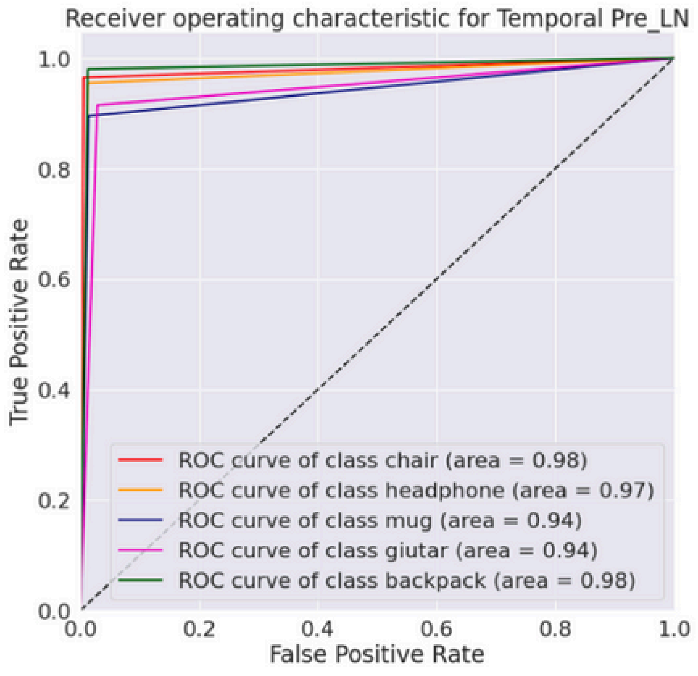
3. PreLN Transformers: Temporal and Spectral



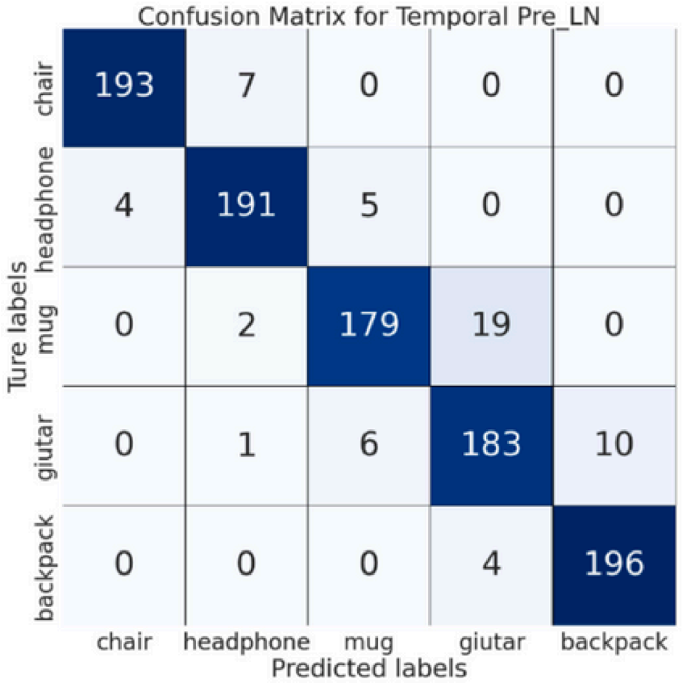
(a)



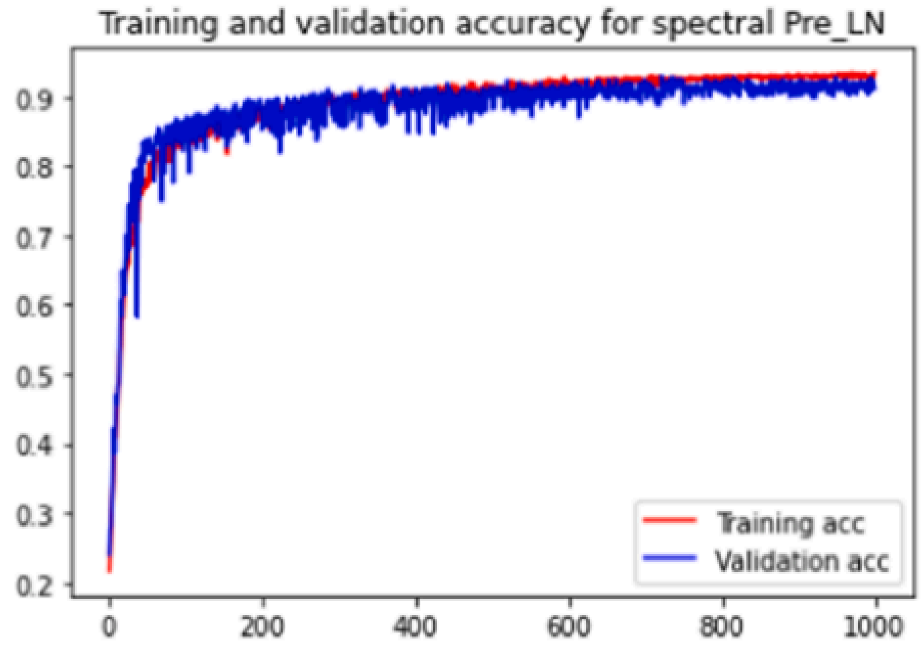
(b)



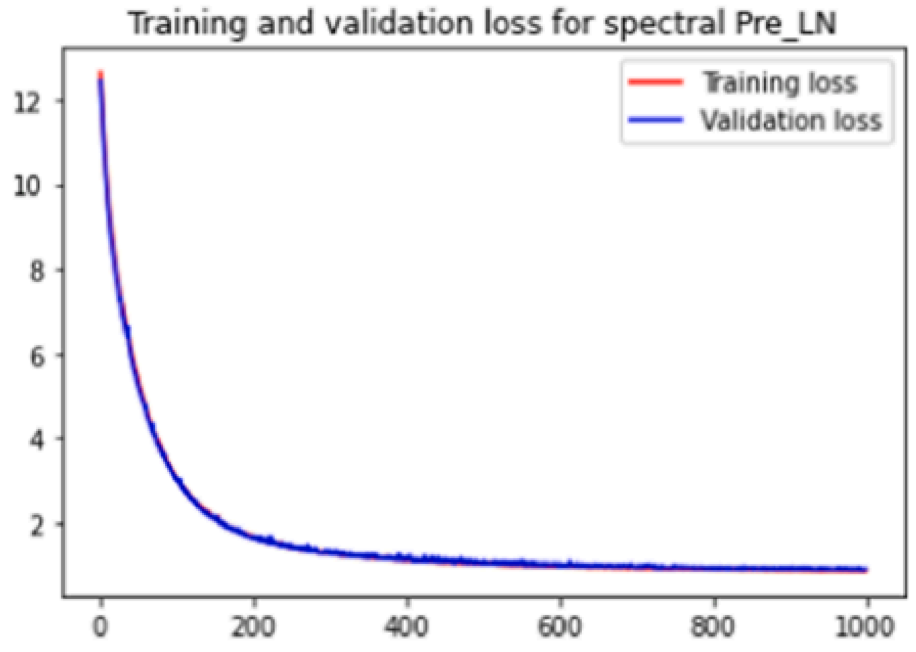
(a)



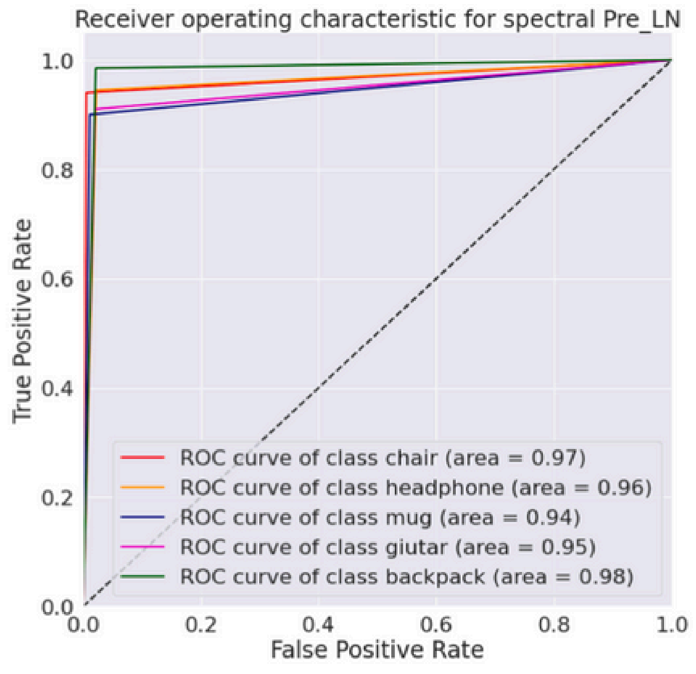
(b)

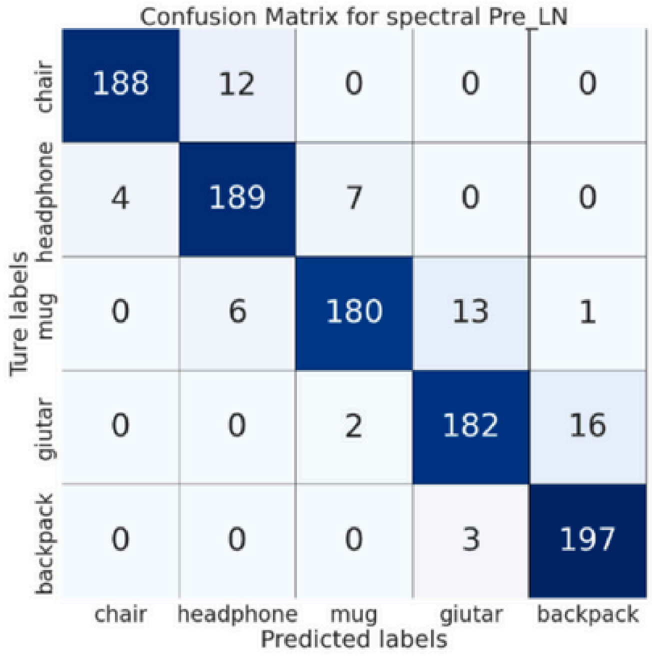


(c)



(d)





Results and Discussion:

BrAIn Lab.
Brain and Artificial Intelligence Lab.

Table 2

The result of post_LN and Pre_LN models for Temporal and Spectral Transformer.

Transformer model	Accuracy	AUC	F1-score
Temporal_Post_LN	85.60	0.91	0.85
Temporal_Pre_LN	94.20	0.96	0.94
Spectral_Post_LN	88	0.92	0.87
Spectral _Pre_LN	93.60	0.94	0.93

LayerNormalization is performed between the residual blocks in the Post-LN Transformer which causes large gradients in the parameters close to the output layer.
→ Solution: A learning rate warmup stage is needed to prevent this issue. Pre-LN Transformers which LayerNormalization would take place inside the residual blocks. The gradients of the Pre-LN Transformer behave properly at initialization. As a result, the input of the feed-forward layer and self-attention is normalized and functions as a regularization.

Results and Discussion:

4. Ensemble-Based Model:
2 PreLN Transformers: Temporal and Spectral

Table 3
The result of ensemble models.

Ensemble model	Accuracy%	AUC	F1-score
Stacked Generalization	**96.1**	**0.97**	**0.96**
Unweighted Model Averaging	95.1	0.96	0.95

Table 4
Comparing the results of different methods on the recorded database.

Methods	Accuracy (%)
SVM + DWT Features	69
MLP + DWT Features	81
Proposed GRU-based model	88.70
Temporal Transformer	94.2
Proposed Spectral Transformer	93.6
Unweighted Model Averaging ensemble (Temporal + Spectral Transformer)	95.1
Stacked Generalization ensemble (Temporal + Spectral Transformer)	**96.1**

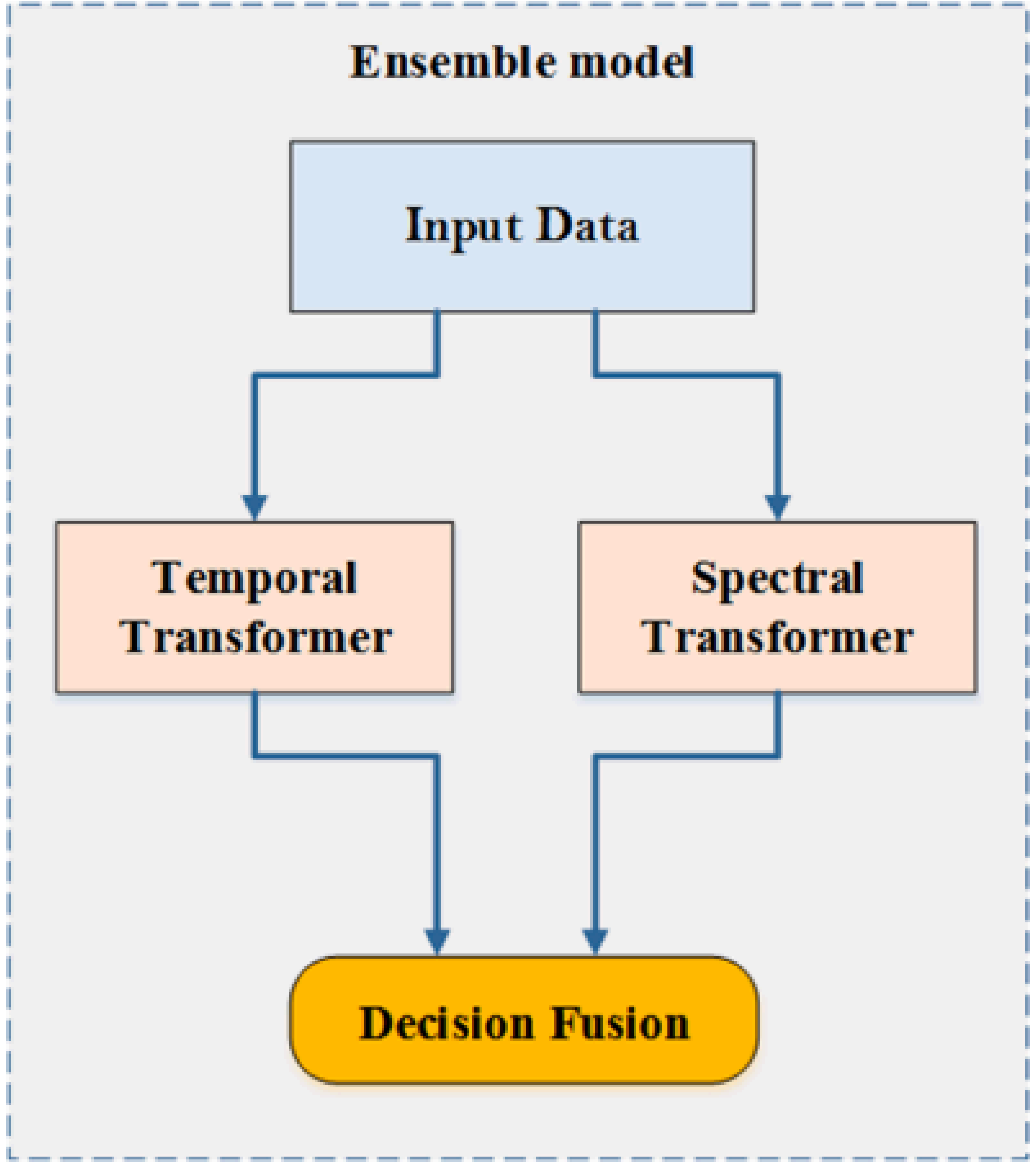


Fig. 7. Overall frame diagram of the ensemble model.

Conclusion

BrAIn Lab.
Brain and Artificial Intelligence Lab.

- Transformer based classification combined with ensemble models are proposed for EEG classification for BCI application.
- Spectral transformer use PSD to acquire frequency features.
- Placement of LN is important to get high performance, pre LN transformer achieved better results compared to post LN transformer.
- Transformers have acceptable results, however the performance can be enhanced using stacked generalization.
- Usage of larger dataset with more subjects, trials or tasks can improve the performance and generalization of the proposed method.

Resources

link for this paper
https://www.sciencedirect.com/science/article/pii/S1746809423005633?ref=pdf_download&fr=RR-2&rr=91095c311a4aeb8a