Self-attentive Bidirectional LSTM Networks for Temporal Decoding of EEG Motor States

Sara Kamali, Fabiano Baroni, Pablo Varona

GNB lab, Escuela Politecnica Superior

Autonomous University of Madrid, Spain

Sara.kamali@uam.es

17.06.2025

IWANN 2025, Coruña, Spain

Motivation

Classification of brain signals to movement and non-movement with explainable NN

- BCI and neurotechnology application
- Biomarker detection applications
- Highlighting the key intervals for motor intention detection

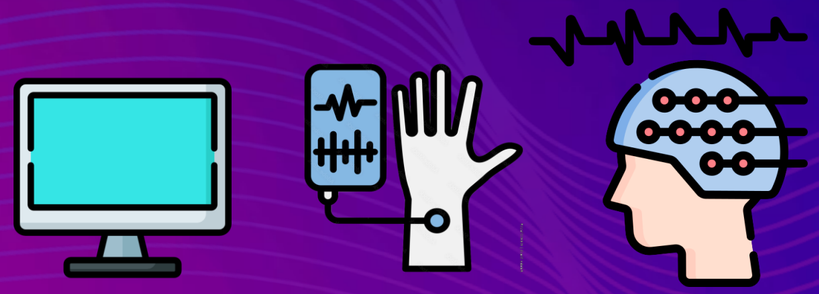
Challenges

- Temporal dynamics
- Inter-subject variability
- Inter-trial variability

EEG & Task Description

Data characteristics

Simultaneous EEG & EMG recording
Sampling rate: 512 Hz
Initial number of subjects: 38
Number of trials per hand: 20
(Cho et al. 2017)



Task

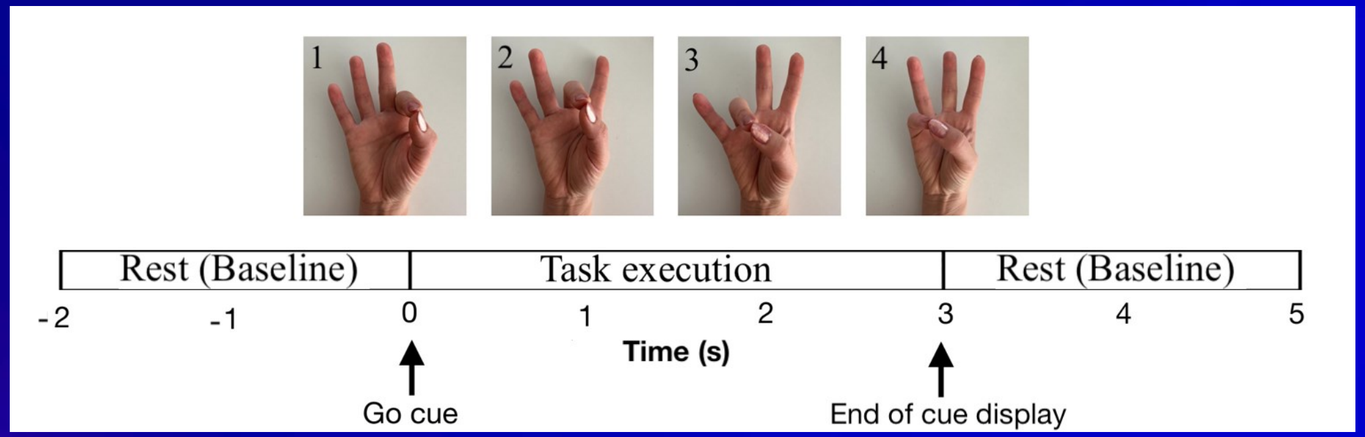


Figure 1. Stereotyped finger-pinch task. A visual cue randomly instructed participants to pinch with the left or right hand; rest (baseline) intervals were of randomly varying duration *(Kamali et al., 2024)*

EEG & EMG Preprocessing

1. Application of EEG/EMG preprocessing pipelines
2. Extraction of brain dipoles (Fig. 2)
3. Extraction of the EMG onset times

Features extracted from dipoles' time series

Time-resolved power at the mu (7–14 Hz) and beta (15–28 Hz) bands for single trials

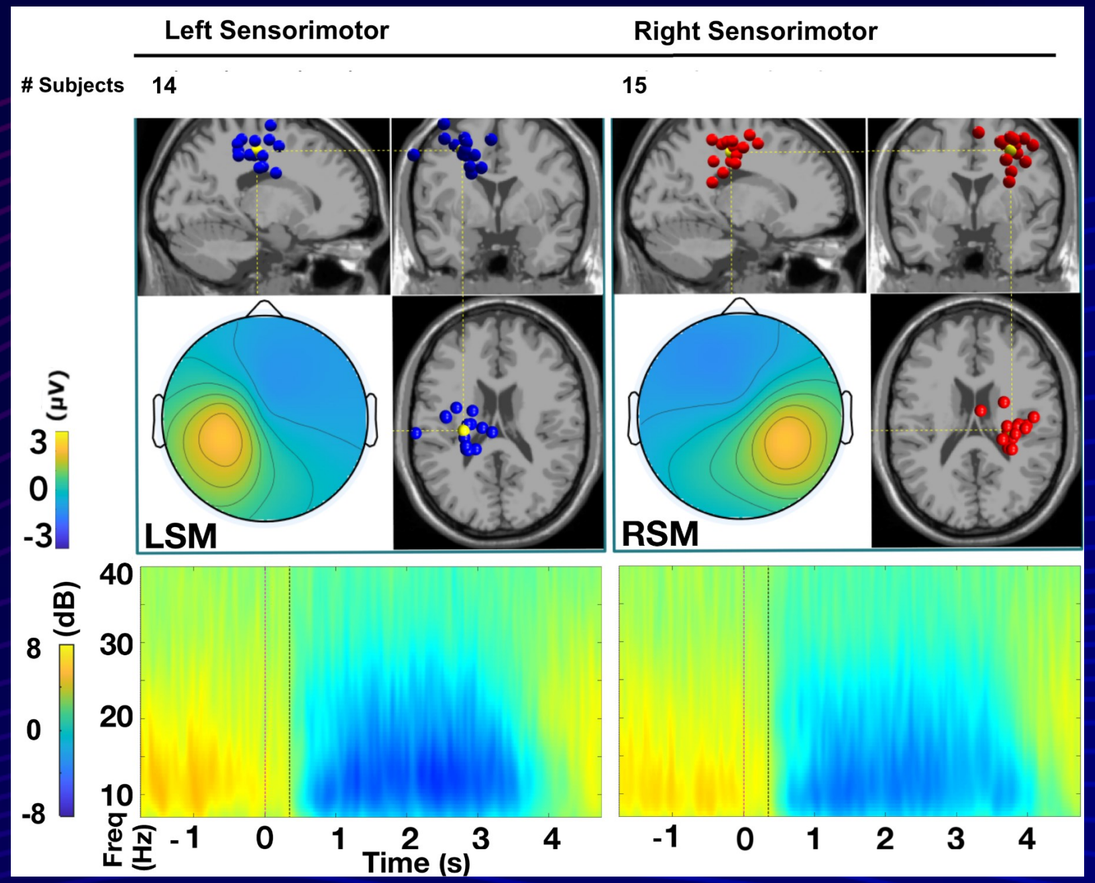


Figure 2. Extracted brain sources (dipoles) after the preprocessing of EEG data and ICA, at the group-level clustering. *(Kamali et al., 2024)*

Data Preprocessing

Extracted a 2-second window around the average latency of all the subjects
Down-sampled the data to 128 Hz

Data Labeling

Binary labeled the samples

- Class 0: non-movement interval
- Class 1: movement interval

Target

Time-resolved classification of power

Network Architecture

- Input: Batch-size x 256 x 2
- Models: Two-layer BiLSTM + Self-attention layer
- Output: Time-distributed output, size: Batch-size x 256 x 2

.

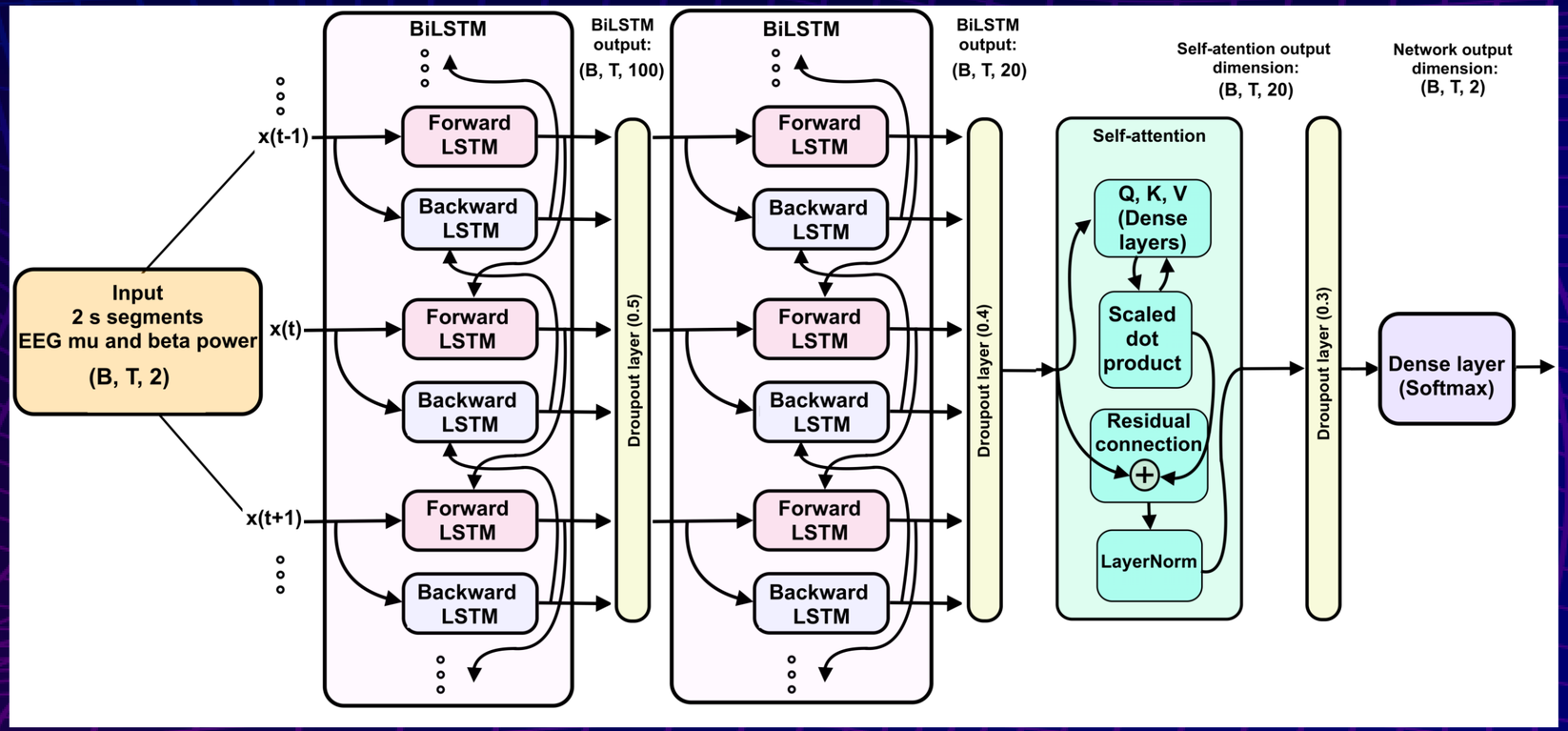


Figure 3. Block diagram of the neural network architecture. Dimension of each layer is shown as (B, T, feature-dimension), where B is the batch size and T is the number of samples, equal to 256.

Training and Validation

Training

- Batch size: 4
- Epochs: Max 20
- Early stopping
- LR scheduler

Loss

- Sparse categorical cross-entropy

True label

Predicted label

$$\mathcal{L} = -\sum_{t}\sum_{c} y_{t,c}\log(\hat{y}_{t,c})$$

Validation

- Leave-One-Subject-Out (LOSO) Cross-Validation

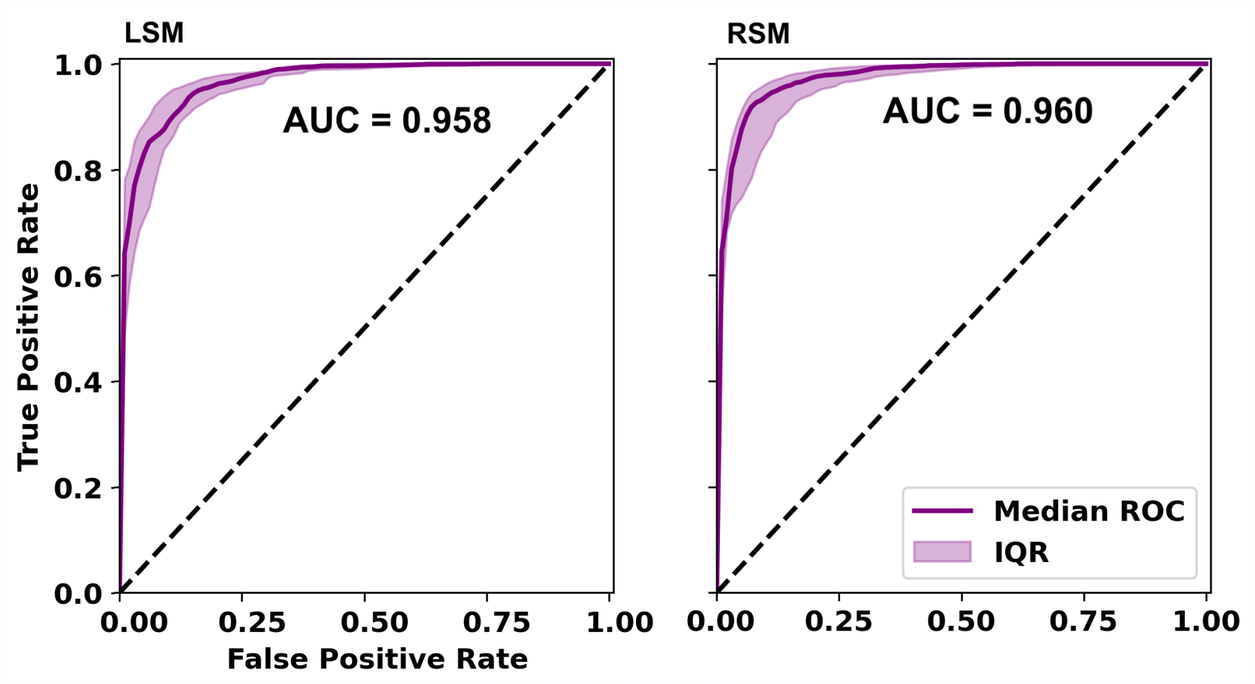
Comparison of RNNs

Table 1. Classification performance metrics (mean ± std) over LOSO folds, for LSM and RSM. Results presents the performance of three network variants that differ only in the configuration of the first two recurrent layers; all subsequent layers and hyper-parameters were held constant.

RNN model	Brain Area	Accuracy	Precision
2×BiLSTM	LSM	0.88 ± 0.05	0.88 ± 0.05
	RSM	0.88 ± 0.04	0.89 ± 0.04
2×BiGRU	LSM	0.76 ± 0.11	0.77 ± 0.09
	RSM	0.80 ± 0.09	0.81 ± 0.08
2×BiRNN	LSM	0.65 ± 0.10	0.66 ± 0.10
	RSM	0.64 ± 0.08	0.65 ± 0.08

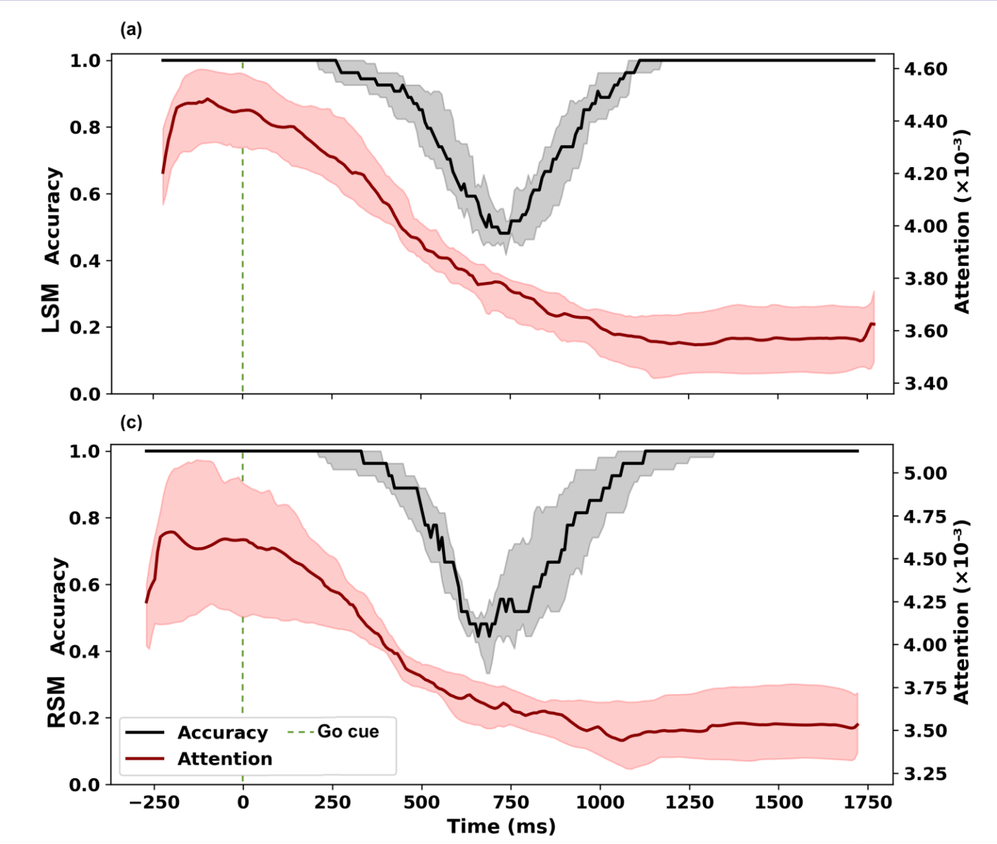
Performance Metrics

Figure 4. The median ROC curves, along with the median AUC values, for the LSM and RSM. Shaded areas in all panels represent the IQR. The values are computed across all LOSO folds.



Attention Weights & Accuracy

Figure 5. Attention and accuracy weights. The median sample-wise classification accuracy across subjects, along the median of the average attention weights per time point for the LSM and RSM. IQR range shown in shaded areas. Results include all the LOSO folds.



Discussion

- ✓ LSTM captures long-range dependencies, yet reaches better performance compared to GRU and RNN
- ✓ Attention layer improves interpretability
- ✓ Model emphasizes anticipatory brain activity

Conclusion

- BiLSTM + attention: robust & interpretable EEG decoder
- Applicable to BCI, neural prosthetics, motor control research

Open code and credits

Pablo Varona

Fabiano Baroni

Sara Kamali

Access codes here
https://github.com/GNB-UAM/Kamali-BiLSTMs-Self-attention-for-motor-EEG

Acknowledgment
Funded by PID2024-155923NB-I00, CPP2023-010818, PID2021-122347NB-I00

Thank you!

Questions?

References

Sara Kamali, Fabiano Baroni, and Pablo Varona. "Mu and beta power effects of fast response trait double dissociate during precue and movement execution in the sensorimotor cortex." *bioRxiv* (2024): 2024-11.

Hohyun Cho, Minkyu Ahn, Sangtae Ahn, Moonyoung Kwon, and Sung Chan Jun. "EEG datasets for motor imagery brain–computer interface." *GigaScience* 6, no. 7 (2017): gix034.