Deep Feature Mining via Attention-based BiLSTM-GCN for Human Motor Imagery Recognition

Yimin Hou [1], Shuyue Jia [1, 2 *], Xiangmin Lun [1], Shu Zhang [3], Tao Chen [1], Fang Wang [1], and Jinglei Lv [4]

[1] School of Automation Engineering, Northeast Electric Power University

[2] Department of Computer Science, City University of Hong Kong

[3] School of Computer Science, Northwestern Polytechnical University

[4] School of Biomedical Engineering and Brain and Mind Center, The University of Sydney

EEG Deep Learning Library: https://github.com/SuperBruceJia/EEG-DL

Background

- **BCI**: establish connections between the brain and machines

    (1) **Acquire and analyze brain signals** while conducting **actual** or **imagery** tasks

    (2) **Control machines**

- **Significance**: help the disabled and understand the human brain
- **Types of BCI**:
    - Electroencephalography (EEG)
    - Magnetoencephalography (MEG)
    - Functional Magnetic Resonance Imaging (fMRI)
    - Invasive BCI Technologies (*e.g.*, Neuralink)

- **Reasons for using EEG for this project**:
    - Non-Invasiveness
    - High Temporal Resolution
    - Portability
    - Inexpensive Equipment

  A potential market

- **Specific Task**: EEG Motor Imagery (*e.g.*, control a wheelchair via imagery-based EEG signals)
- **Our Research**: develop EEG-based BCI technologies to improve current stroke rehabilitation strategies

Image Credit: in the public domain.

Key Points in dealing with EEG time series

- **Individual Variability → Lower Classification Accuracy**
  - ✓ Low SNR
  - ✓ Different brain electrical conductivity ← different anatomical structure of brain
  - ✓ Electrodes' positional error

Feature Extraction

EEG Electrodes' Structure Modeling

- **Slow Responding → Hard to develop Real-life Applications**
  - ✓ [**most literature**] Trial-level prediction (*e.g.*, 4 s)
  - ✓ Window/Slide-level prediction (*e.g.*, 0.4 s)
  - ✓ Time-resolved prediction (*e.g.*, 6.25 ms) (Our Work)

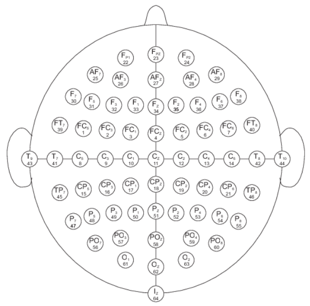
Time-resolved or Window-based Signal Sampling

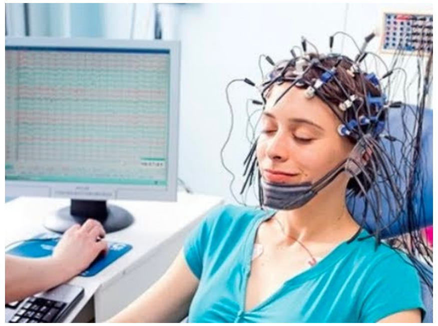
- **Lower Group-level Accuracy → Hard to develop Applications for a Group of People**
  - ✓ [**most literature**] Subject-level prediction (Our Work)
  - ✓ Group-level prediction (Our Work)

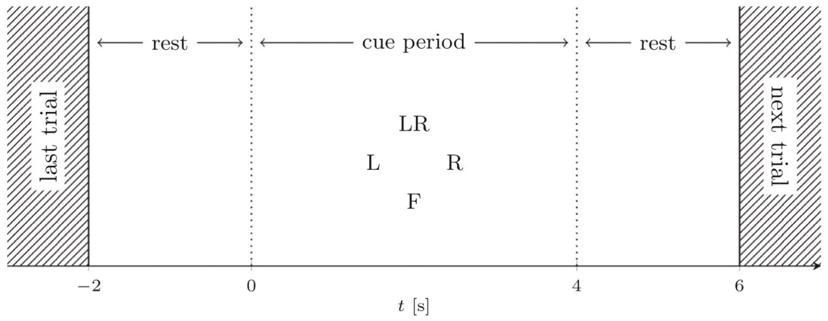
Benchmark Dataset

- The PhysioNet Dataset (EEG Motor Movement/Imagery Dataset)
- International 10-10 EEG System → **64 electrodes**

  (excluding electrodes Nz, F9, F10, FT9, FT10, A1, A2, TP9, TP10, P9, and P10)



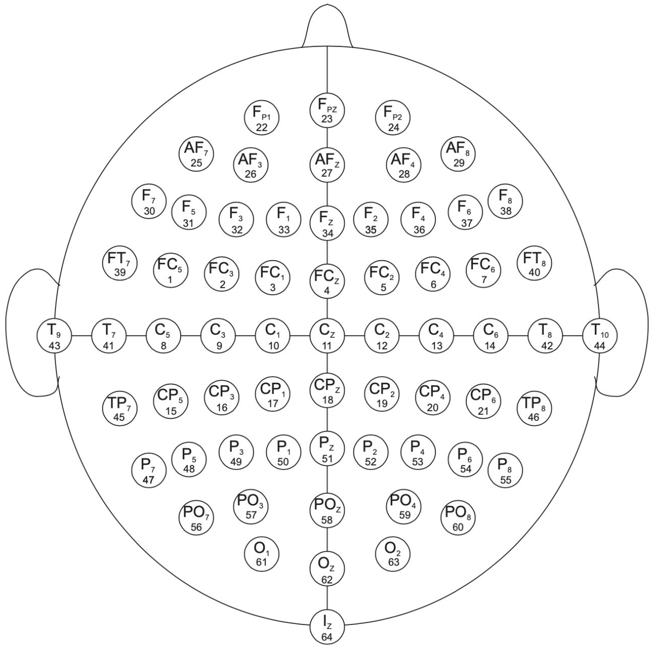




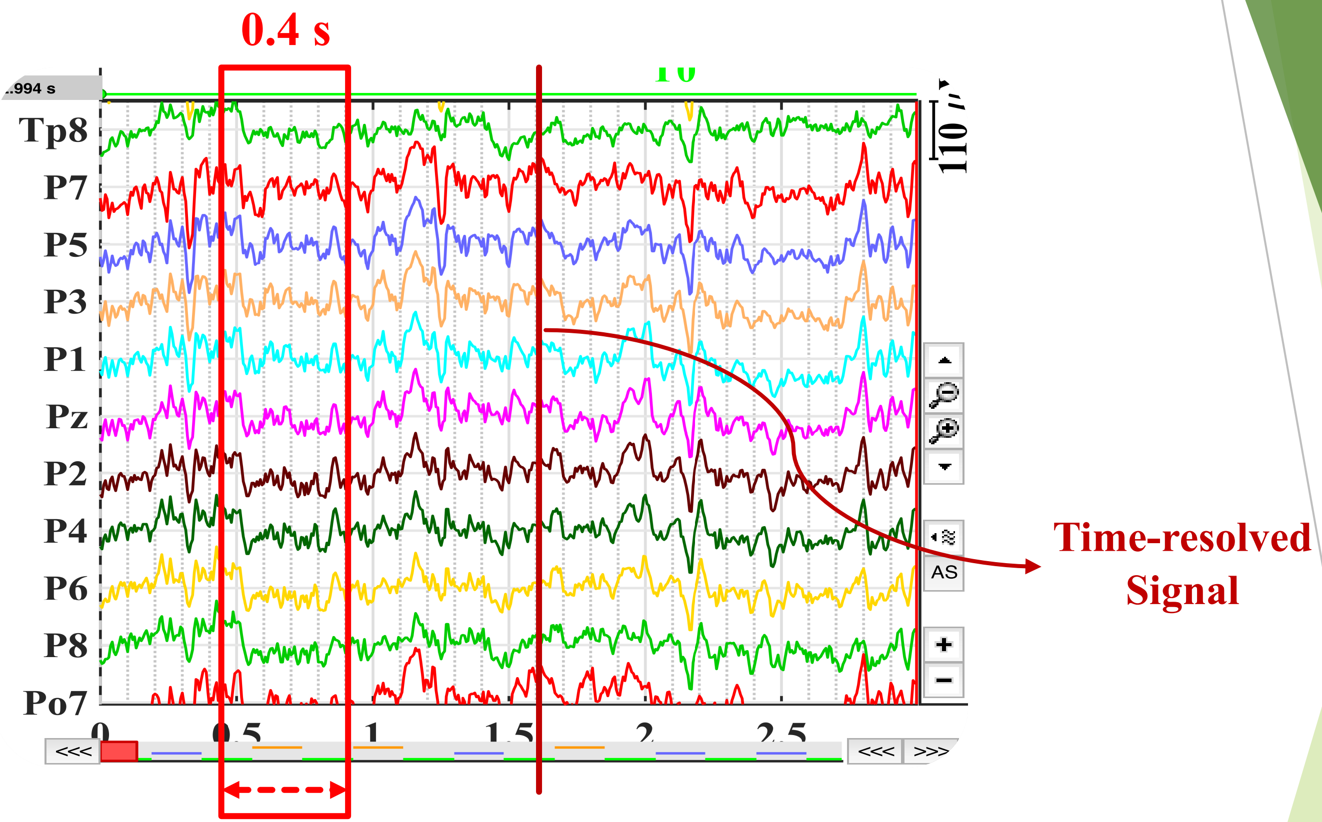
- **109 subjects (the largest number of participants in the field of EEG Motor Imagery)**
- Task: **4-class EEG Motor Imagery Classification**
  - ✓ Imagining (Task 1) left fist, (Task 2) right fist, (Task 3) both fists, (Task 4) both feet
- Each subject → 3 runs, 7 trials, 4 classes → 84 trials in total
- Each trial → 4 seconds experimental duration, 160 Hz Sampling Rate → 640 Time Points
- We apply the **Time-resolved Sampling Method**
  - ✓ Total samples per subject: 3 runs × 7 trials × 4 classes × 4 seconds × 160 Hz = 53,760 samples
  - ✓ Experimental Setting: 90% as the training set and the left 10% as the test set

Image Credit: The PhysioNet Dataset and the middle image is in the public domain.

One Problem of the GCNs-Net



Spatial information



Temporal information

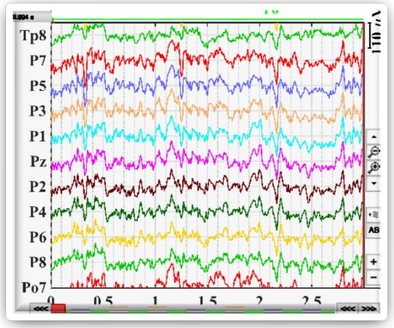
- ✓ GCNs-Net is based on **Time-resolved Signal** → doesn't consider **Temporal Information**

Motivation:

- ✓ **[Spatial-Temporal Analysis] Consider Temporal and Spatial Information from EEG signals**
- ✓ **[Responsive] Maintain High Responding Time**

Image Credit: The PhysioNet Dataset.

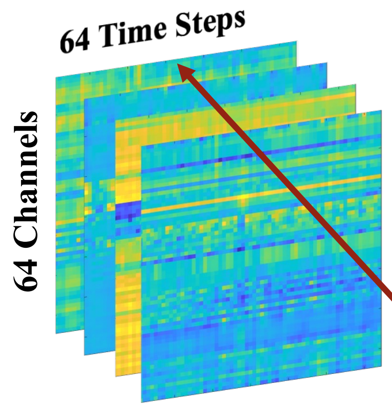
64-channel Raw EEG Signals Acquisition



Data over experimental Duration (4 seconds)

Slice

Spatial Information



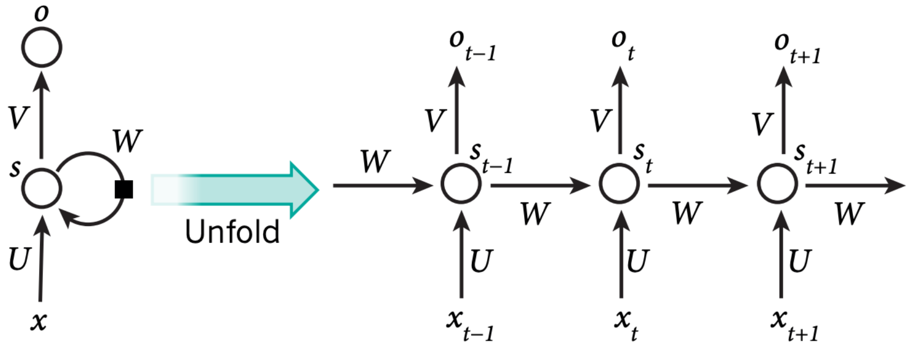
0.4-second Segments

Temporal Information

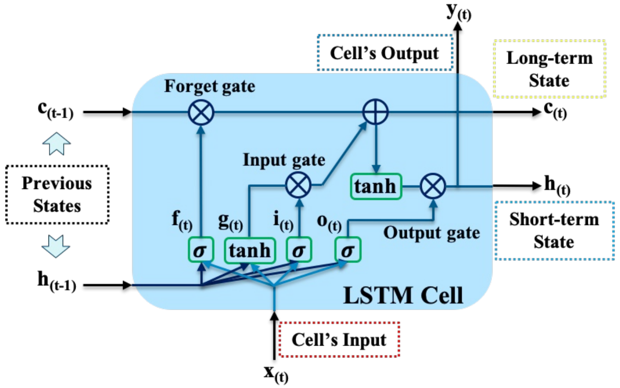
- ✓ 4-s Signals (experimental duration): **0.4-s segments** over time
- ✓ Each Segment: **64 channels × 64 time steps**
- ✓ Pre-processed Data: **Temporal** Information + **Spatial** Information

Temporal Information Extraction

unrolling the network through time



Recurrent Neural Network (RNN)

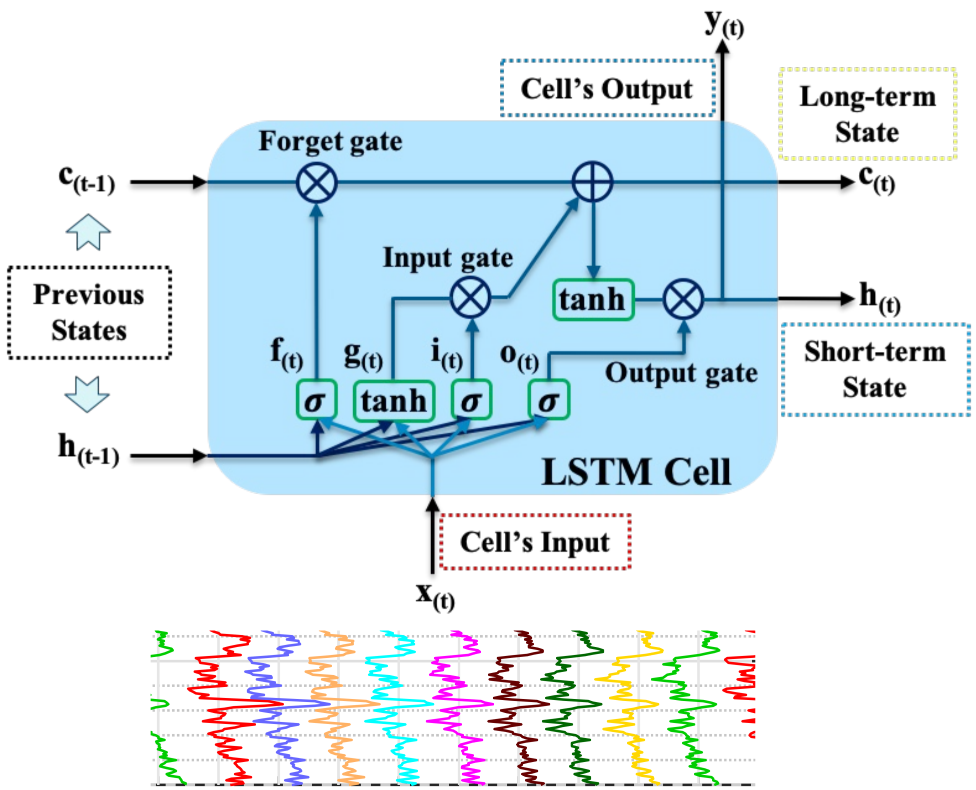


Long Short-Term Memory (LSTM)

- ✓ Designed for **order-mattered sequential data**, *e.g.*, time series
- ✓ The learned features at *time step t* are affected by $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$ → **continuly learn** from time series
- ✓ LSTM: better capture **long-range sequence dependencies**
- ✓ Gated Recurrent Units (GRU): **lightweight** architecture with comparable performance

Image Credit: LeCun *et al.*, Deep Learning, In Nature.

Long Short-term Memory (LSTM)



64-channel Raw EEG Signals
at Time Step $x_{(t)}$

- ✓ **RNN**: **Vanishing Gradient problem**
- ✓ **LSTM**: Capture **Long-range Dependencies**

  by the long-term state path $\mathbf{c}_{t-1} \rightarrow \mathbf{c}_t$ (improve the gradient flow)
- ✓ **Gate**: **control information flow**
- ✓ **Input Gate**: store $\mathbf{x}_t$ and control $\mathbf{c}_t$'s input
- ✓ **Forget Gate**: control $\mathbf{c}_{t-1}$
- ✓ **Output Gate**: control $\mathbf{c}_t$'s output

  $\rightarrow$ short-term state $\mathbf{h}_t$ (**Cell's Output**)
- ✓ More parameters to store information
- ✓ Bidirectional:

  (1) $\mathbf{x}_1 \rightarrow \mathbf{x}_t$

  (2) $\mathbf{x}_t \rightarrow \mathbf{x}_1$
- ✓ **GRU**: Update Gate, Reset Gate; hidden state

Attention Mechanism

- ✓ Signals or Outputs

<u>Equally</u> treated/contributed

vs.

<u>Differently</u> treated/contributed <u>with preference/importance</u>

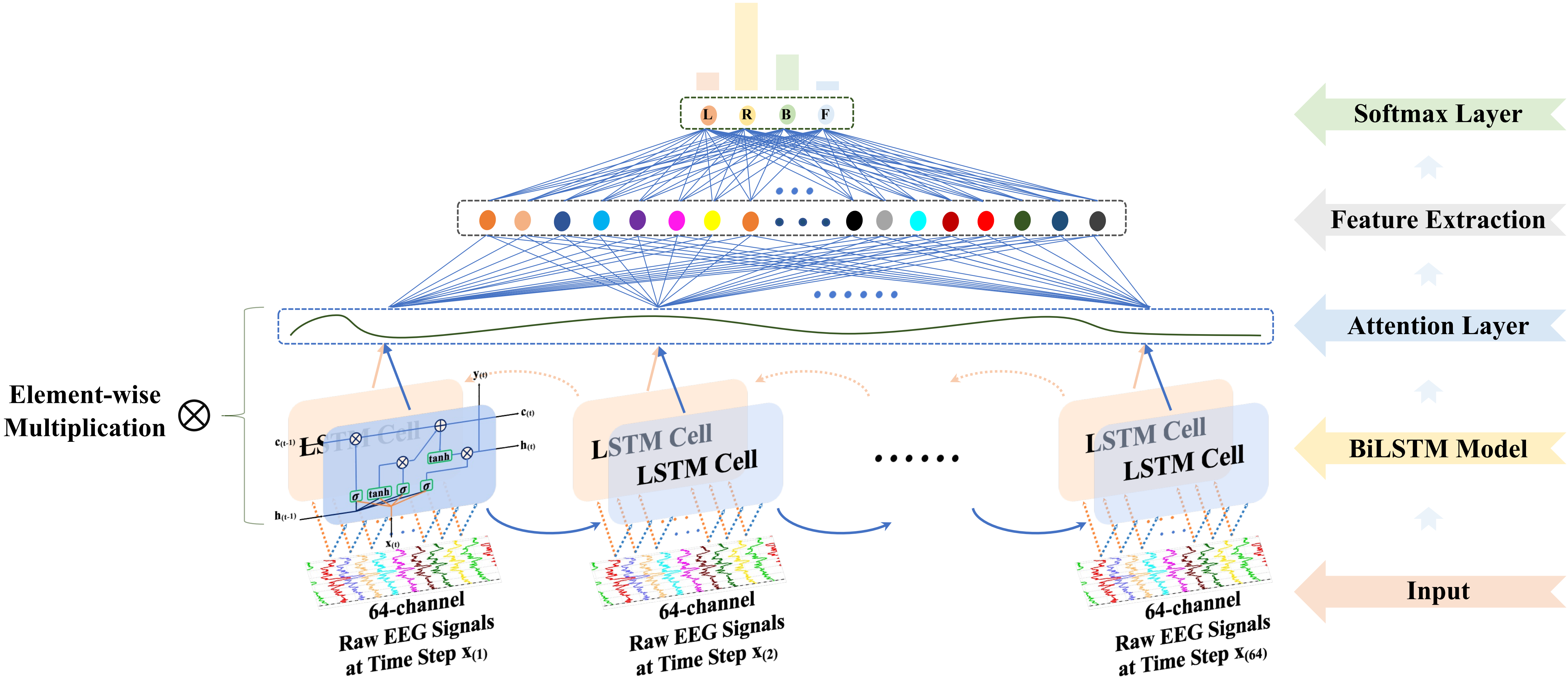
FC Layer $\mathbf{U}_t = \tanh(\mathbf{W}_w \mathbf{y}_t + \mathbf{b}_w)$

Attentional Weights $\boldsymbol{\alpha}_t = \dfrac{\exp(\mathbf{U}_t^T \mathbf{W}_U)}{\sum_t \exp(\mathbf{U}_t^T \mathbf{W}_U)}$

Weighted Sum $\hat{\mathbf{U}}_t = \sum_t \boldsymbol{\alpha}_t \mathbf{y}_t$

Image Credit: Prof. Bolei Zhou.

Attention-based Bidirectional Long Short-term Memory (Bi-LSTM)



Model Design Ablation Study

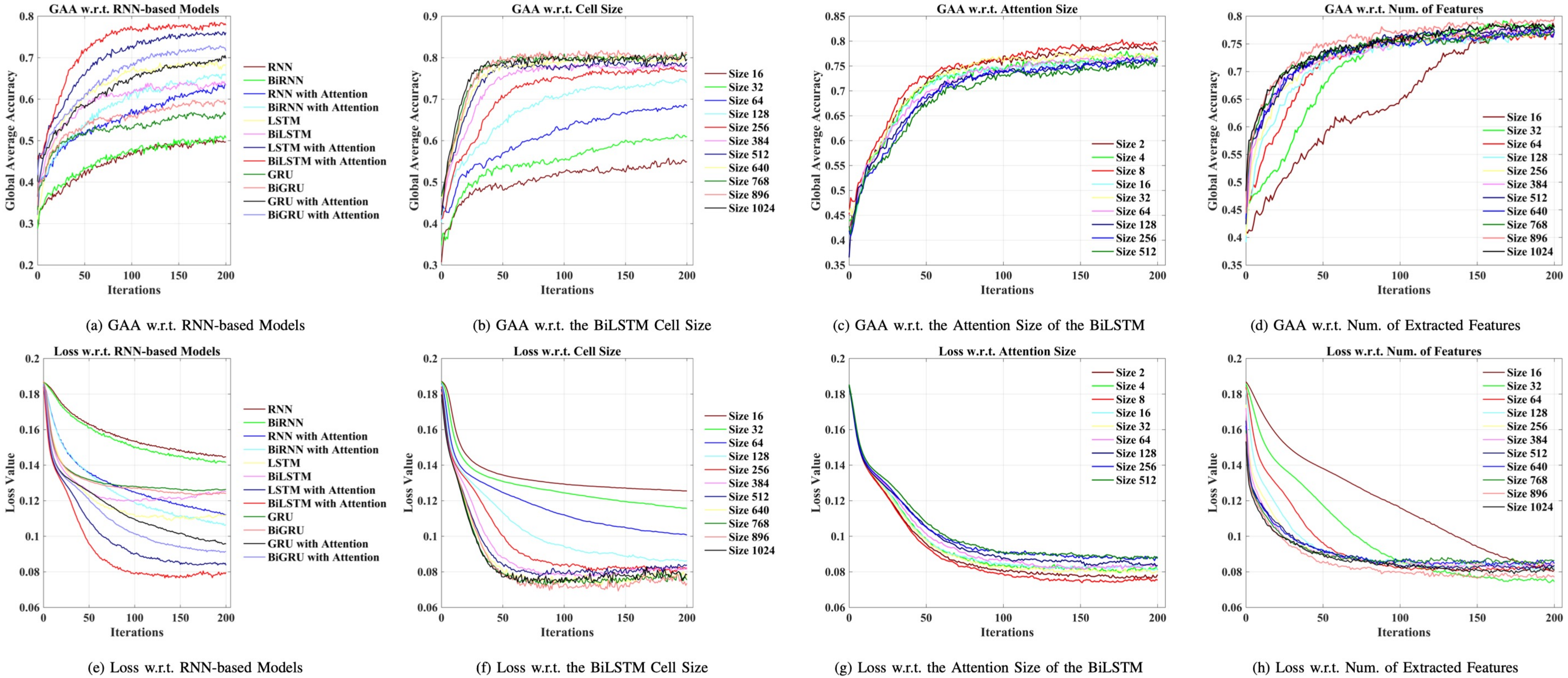
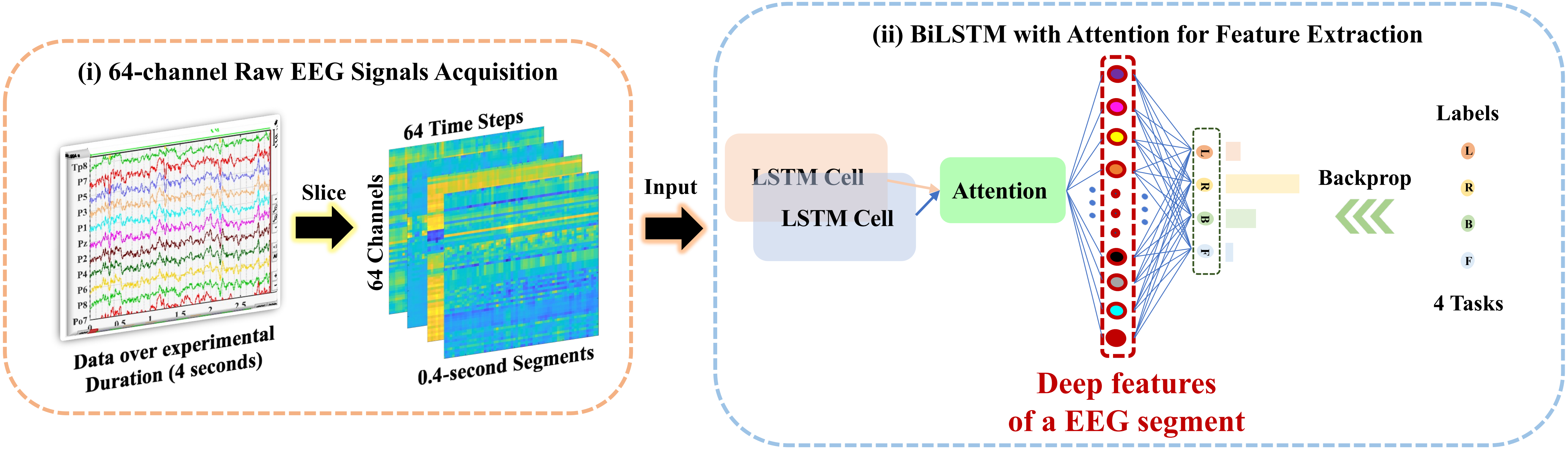
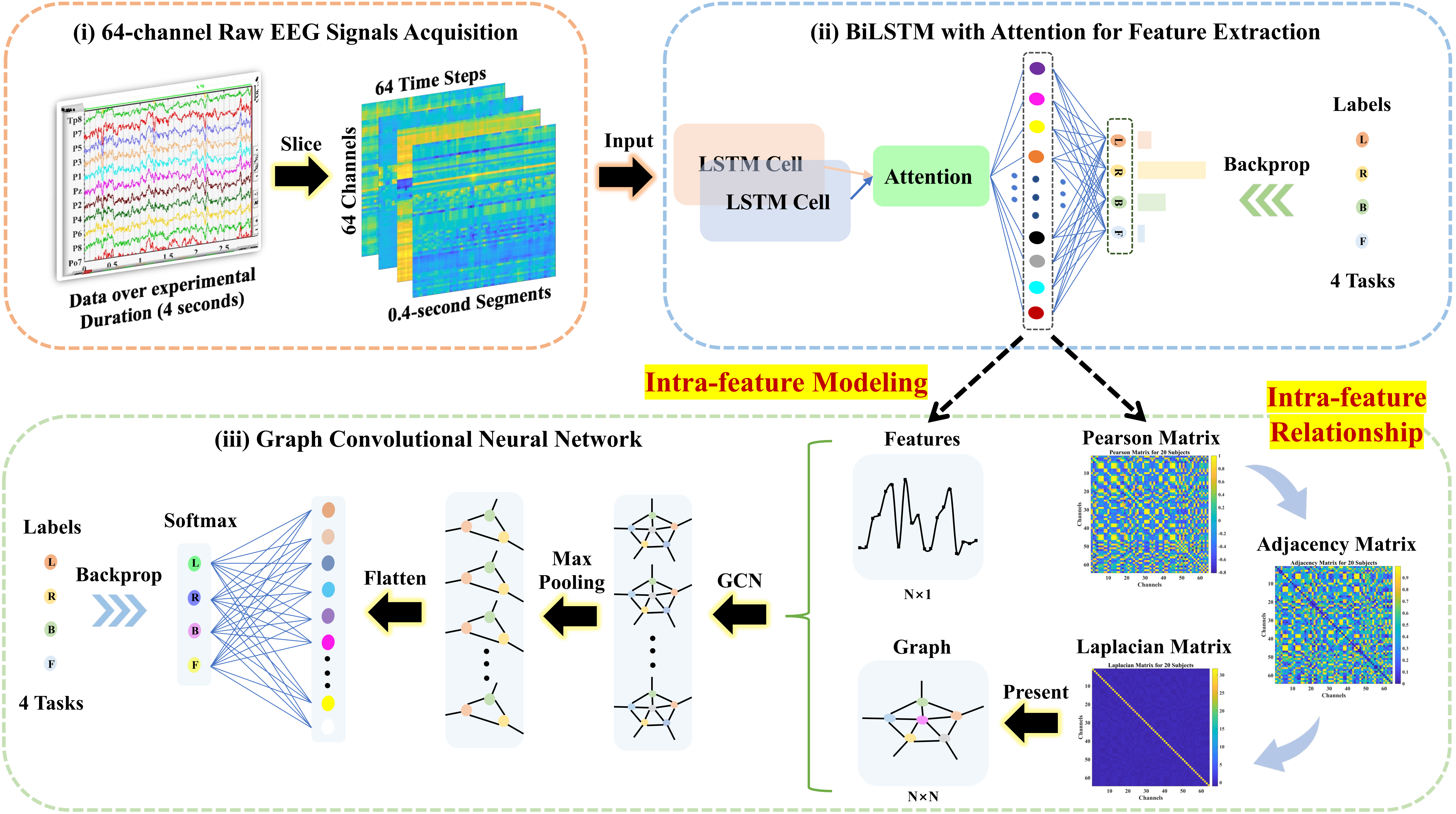


Fig. 3: Models and Hyperparameters Comparison w.r.t. the RNN-based Methods for Feature Extraction

Topological Structure of Features

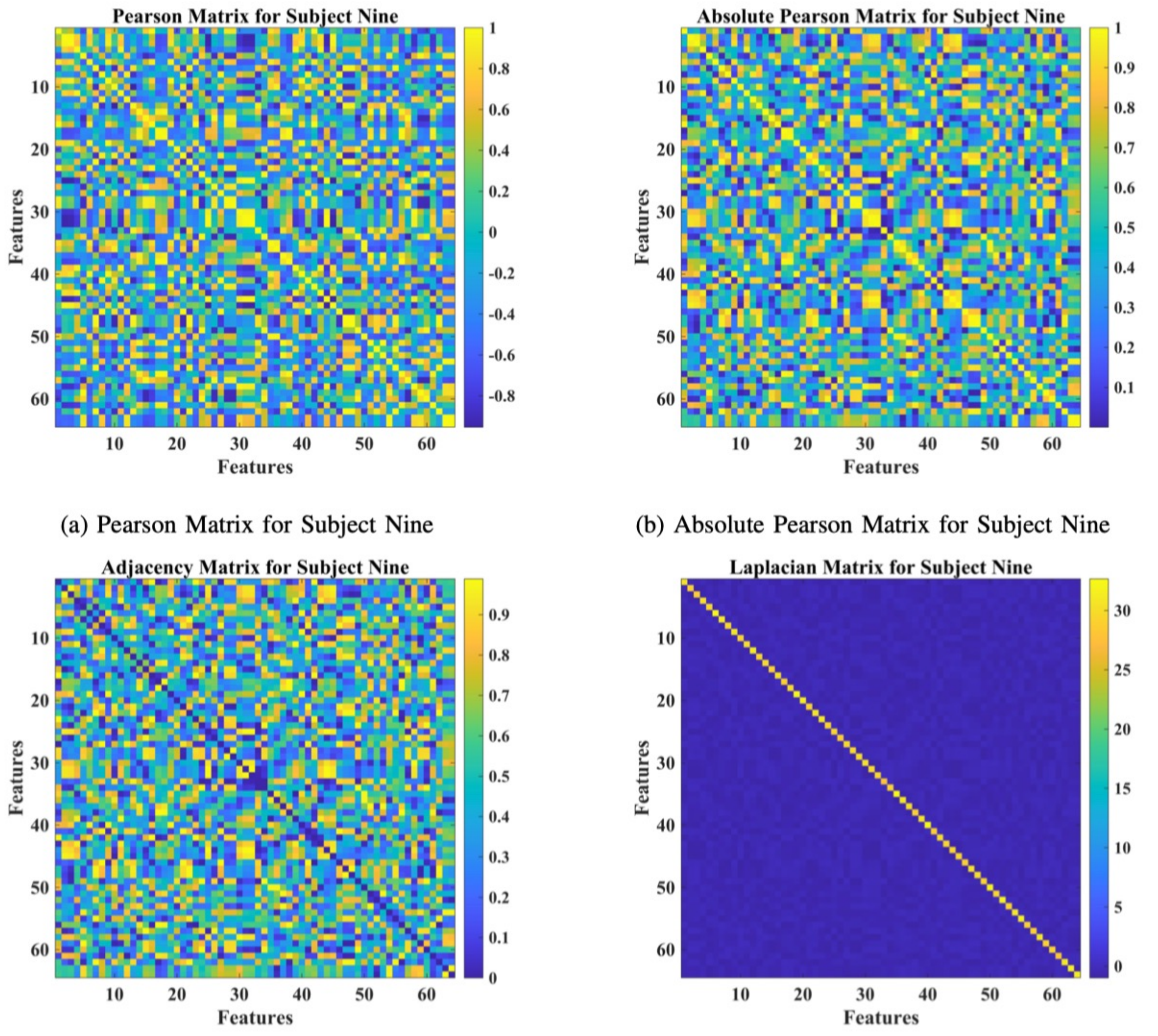


✓ **Deep Feature Mining → Intra-feature Relationship → Intra-feature Modeling**



(i) 64-channel Raw EEG Signals Acquisition
Slice
64 Time Steps
64 Channels
Data over experimental Duration (4 seconds)
0.4-second Segments
Input
(ii) BiLSTM with Attention for Feature Extraction
LSTM Cell
LSTM Cell
Attention
Labels
Backprop
4 Tasks
Intra-feature Modeling
Intra-feature Relationship
(iii) Graph Convolutional Neural Network
Features
N×1
Pearson Matrix
Adjacency Matrix
Graph
N×N
Laplacian Matrix
Present
GCN
Max Pooling
Flatten
Softmax
Labels
Backprop
4 Tasks

Topological Structure of Features



(a) Pearson Matrix for Subject Nine

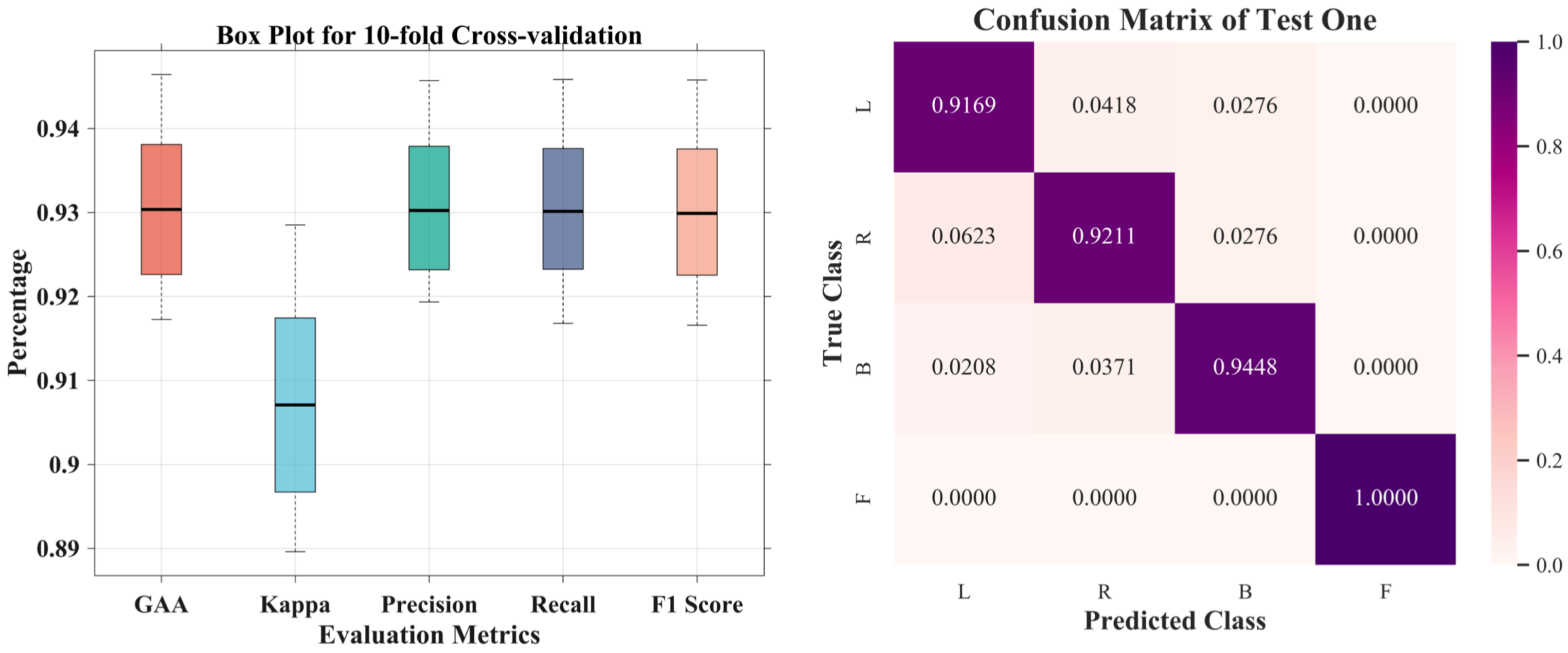
(b) Absolute Pearson Matrix for Subject Nine

(c) Adjacency Matrix for Subject Nine

(d) Laplacian Matrix for Subject Nine

Fig. 4: The Pearson, Absolute Pearson, Adjacency, and Laplacian Matrices for Subject Nine.

Experimental Results - Groupwise Prediction



(a) Box Plot for 10-fold cross validation

(b) Confusion Matrix of Test One

Fig. 5: Box plot and confusion matrix for 10-fold cross validation.

Note:

(1) Box Plot (Maximum Score, Upper Quartile, Median, Lower Quartile, and Minimum Score)

(2) Confusion Matrix: TP, TN, FP, and FN

Experimental Results - Subject-Specific Adaptation

TABLE II: Subject-level Evaluation

No. of Subject	GAA	Kappa	Precision	Recall	F1 Score
1	94.05%	92.06%	94.20%	94.32%	94.16%
2	96.43%	95.19%	96.06%	96.06%	96.06%
3	97.62%	96.79%	97.33%	97.08%	97.18%
4	90.48%	87.34%	91.30%	91.11%	90.42%
5	95.24%	93.61%	95.96%	95.06%	95.38%
6	94.05%	92.02%	93.40%	94.96%	93.66%
7	98.81%	98.40%	98.81%	99.07%	98.92%
8	95.24%	93.60%	95.39%	95.04%	95.19%
9	98.81%	98.39%	99.11%	98.68%	98.87%
10	94.05%	91.98%	93.39%	94.70%	93.61%
Average	**95.48%**	**93.94%**	**95.50%**	**95.61%**	**95.35%**

TABLE III: Current studies comparison on subject-level prediction

Related Work	Max. GAA	Approach	Database
Ortiz-Echeverri *et al.* (2019)	94.66%	Sorted-fast ICA-CWT + CNNs	
Sadiq *et al.* (2019)	95.20%	EWT + LS-SVM	BCI Competition IV-a Dataset
Taran *et al.* (2018)	96.89%	TQWT + LS-SVM	
Zhang *et al.* (2019)	83.00%	CNNs-LSTM	
Ji *et al.* (2019)	95.10%	SVM	BCI Competition IV-2a Dataset
Amin *et al.* (2019)	95.40%	MCNNs	
Dose *et al.* (2018)	68.51%	CNNs	
Hou *et al.* (2019)	96.00%	ESI + CNNs	Physionet Database
This work	**98.81%**	**Attention-based BiLSTM-GCN**	

Takeaways and Future Work

✓ Spatial-Temporal Analysis

(1) Converge to both **Subject-level and Groupwise Predictions** and handle **Individual Variability**.

(2) The 0.4-s sample size **Time-Resolved Solution** toward fast response.

✓ Deep Feature Mining

(1) ↑ **Highest Accuracy**

(2) Advance **Clinical Translation** of EEG-based BCI technology to meet diverse demands, such as those of paralyzed patients.

✓ Future Work

Long-range Dependencies among intra-subject or inter-subject EEG signals can be modeled via **Non-local Modeling**, **Self-attention Mechanism**, **Transformer**, and **AI foundation Models**.

Thank you!

Any question?

Acknowledgment: Dr. Zhijian Hou and Dr. Dingquan Li.