ETH zürich   ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

📦 Code is available!

NEURAL INFORMATION PROCESSING SYSTEMS

PhysioWave: A Multi-Scale Wavelet-Transformer for Physiological Signal Representation

Integrated Systems Laboratory (ETH Zürich)

Yanlong Chen yanlchen@student-ethz.ch

Supervisors

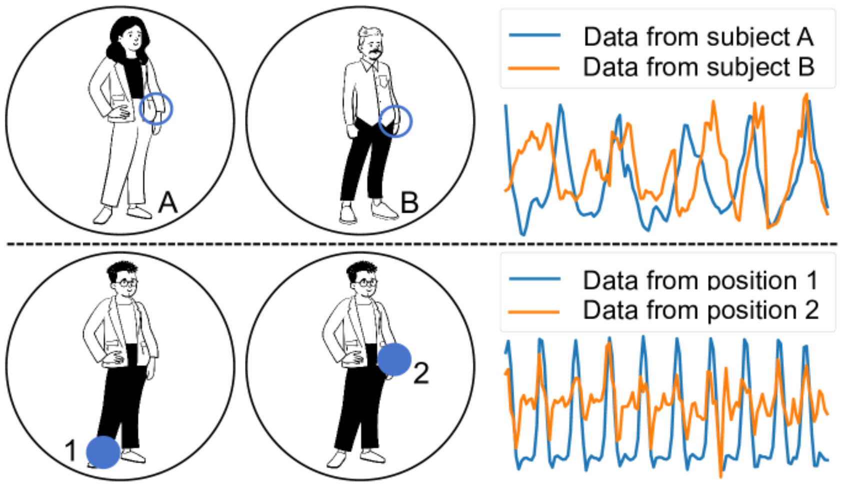
Dr. Yawei Li , Prof. Dr. Simone Benatti, Prof. Dr. Luca Benini, Mattia Orlandi, Pierangelo Maria Rapa

PULP Platform
Open Source Hardware, the way it should be!

@pulp_platform
pulp-platform.org
youtube.com/pulp_platform

Introduction & Motivation

- **Core Challenges in Biosignal Processing**
  - **Low Signal-to-Noise Ratio (SNR)**: Electrode–skin impedance, power-line interference, and motion artifacts obscure weak signals
  - **Strong Non-stationarity & Inter-subject Variability**: Statistical properties of biosignals change over time and differ significantly from one person to another,
  - **Cross-device / Cross-lab Heterogeneity**: Different sampling rates, channel counts, hardware frequency responses ⇒ domain shift



Inter-subject Variability & Domain shift

Framework of our Foundation Model

- **PhysioWave Pretraining Strategy:**

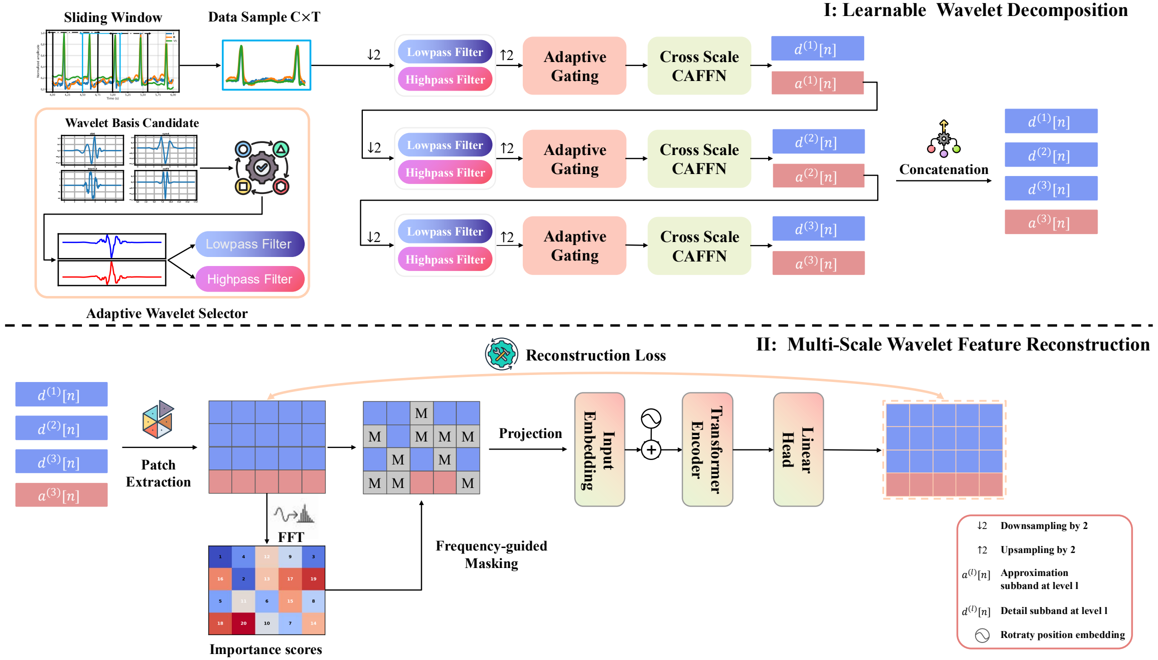
Wavelet Initialization
- Initialize a set of standard wavelet functions (db4,sym4) → low-pass and high-pass filters.

Wavelet decomposition
- Obtain multi-scale frequency-band representations

Frequency-guided Masking
- Patches are masked and passed through Transformer layers for patch reconstruction.

Wavelet Initialization

Wavelet decomposition



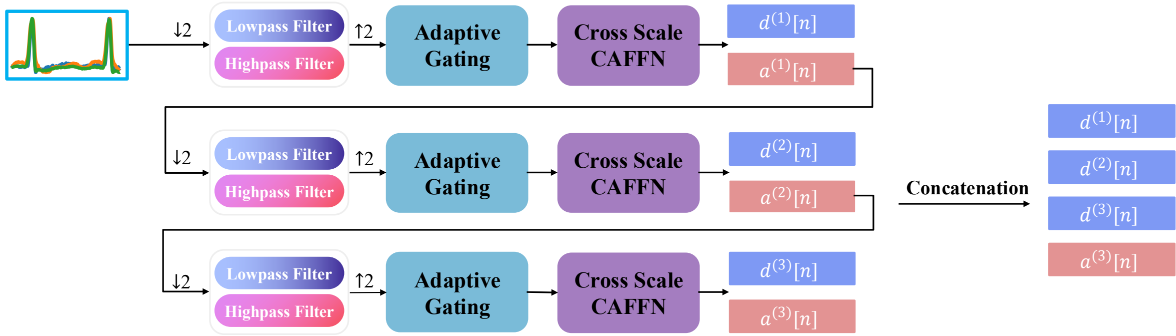
Frequency-guided Masking

Framework of our Foundation Model

- **I: Learnable Wavelet Decomposition:**

Wavelet decomposition breaks down signals into different frequency components at multiple time scales

$a^{(l)}[n]$	Approximation Subband: Keeps the slow, coarse components of the signal. (contour, baseline drift, low-frequency rhythms)
$d^{(l)}[n]$	Detail Subband: Isolates the fast, fine components. (spikes, sharp edges, sudden transients, and high-frequency noise)



multi-scale feature

❌ **Problem: Low SNR**

✅ **Solution:** High pass bands isolate transient spikes from baseline noise, improving denoising.

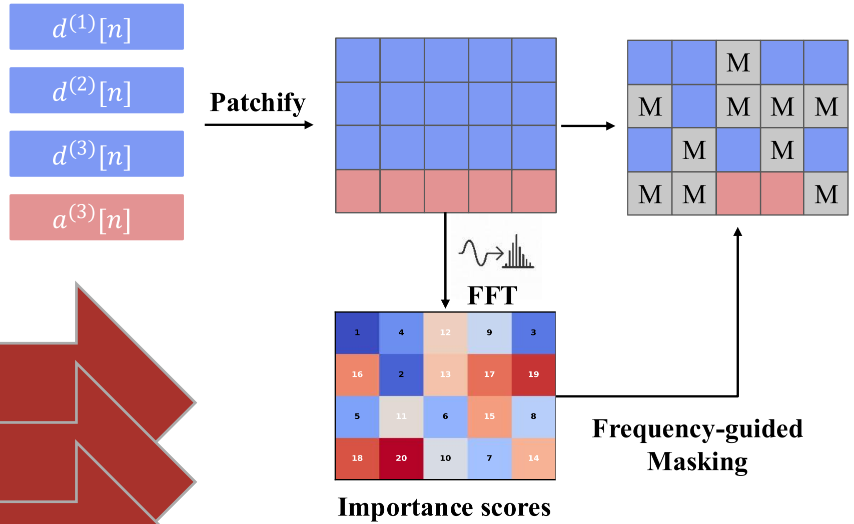
❌ **Problem: Strong Non stationarity**

✅ **Solution:** multi-scale analysis captures both fast and slow dynamics, so the model sees transient bursts and long-term trends separately

Framework of our Foundation Model

- **II: Multi-Scale Wavelet Feature Reconstruction:**

- We mask a subset of time-frequency patches across all bands, then train the model to reconstruct the missing coefficients.



🔄 Frequency-Guided Masking Flow

🔍 Spectral Analysis

The energy of each patch in the frequency domain is computed.

Analyzes the spectral importance of different patches using FFT.

🎭 Strategic Masking

Preferentially mask patches with higher frequency importance (the most informative regions).

💪 Robust Learning

FgM turns reconstruction into a harder, context-driven puzzle, forces the model to learn robust, scale-aware features.

Framework of our Foundation Model

Datasets used for pretraining

- PhysioWave-ecg is trained on approximately 182 GB of 12-lead ECG recordings, while PhysioWave-emg utilizes about 823 GB of EMG data. For each modality, we provide three parameter configurations: Small (5M), Base(15M), and Large (37M).
- Pretraining: 50 epochs on 16 × NVIDIA A100 60 GB (global batch 64; AdamW + cosine LR).

Table 7: 12-lead ECG corpora used for pretraining.

Dataset	Subjects	Records	Dur. (s)	f_s (Hz)	Size
MIMIC-IV-ECG [43]	∼160 000	∼800 000	10	500	90.4 GB
MedalCare-XL [44]	13	16 900	10	500	26.2 GB
CODE-15% [45]	233 770	345 779	10	400	63.3 GB
Norwegian Athlete [46]	28	28	10	500	52.8 MB
Georgia Cohort [47]	10 344	10 344	10	500	1.2 GB

Table 8: Surface-EMG corpora used for pretraining.

Dataset	Subjects	Records	Dur. (s)	f_s (Hz)	Channels	Size
NinaPro DB6 [48]	10	∼8.4 k	4	2 000	14	20.3 GB
NinaPro DB7 [49]	22	∼5.4 k	5	2 000	12	30.9 GB
NinaPro DB8 [50]	12	∼2.4 k	7.5	1 111	16	23.6 GB
EMG2Pose [51]	193	25 253	60	2 000	16	431 GB
EMG2Qwerty [52]	108	1 135	1 080	2 000	16	317 GB

Framework of our Foundation Model

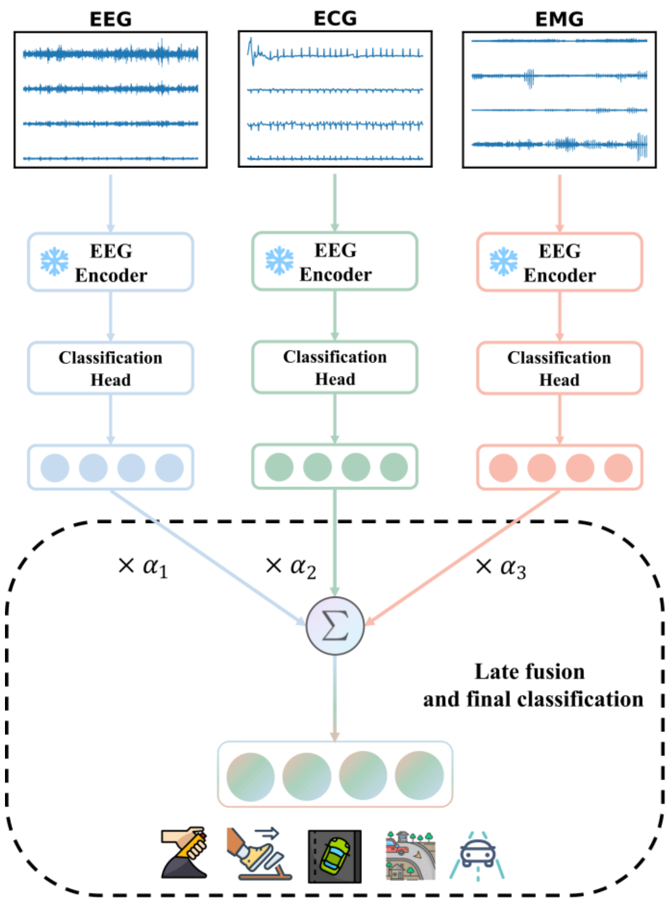
Methods for Downstream Tasks

🎯 Single-modal tasks.

- The entire pretrained encoder is finetuned end-to-end.
- All embeddings are mean-pooled connected with a lightweight classification head.

🔗 Multi-modal tasks

- For each modality, keep the pretrained encoder frozen and train only small classification head together with a set of learnable fusion weights α
- The final prediction is obtained by taking a weighted average of the per-modality logits using these α values.



Multi-modal framework: Classification of driving behaviors in the MPDB dataset.

Experiments & Results

ECG Rhythm Classification Performance

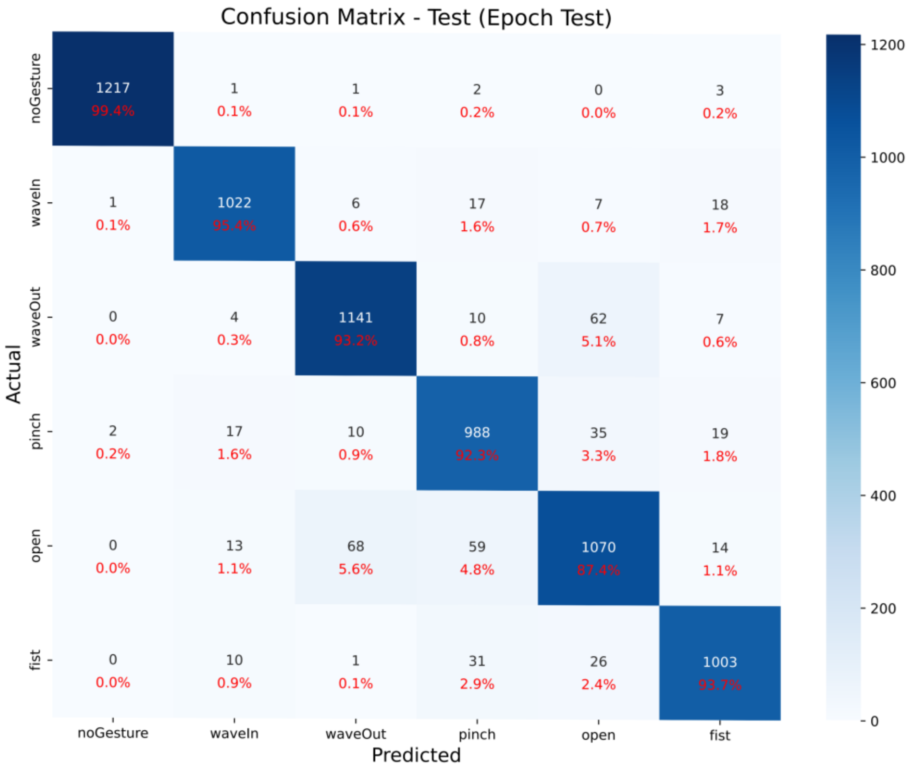
Method	Parameters	PTB-XL		CPSC 2018		Chapman–Shaoxing	
		F1	AUROC	F1	AUROC	F1	AUROC
ECG-Chat 2024	13B	55.9	94.1	**80.1**	95.7	—	—
MERL 2024	11M	48.1	91.9	72.8	92.6	—	87.9
MaeFE 2023	9M	64.7	88.6	71.6	94.5	—	—
OpenECG-SimCLR	11M	46.9	91.5	73.1	92.4	52.3	95.1
OpenECG-BYOL	11M	47.7	91.1	72.8	92.6	51.5	94.8
OpenECG-MAE	11M	48.1	90.9	74.5	93.2	50.8	94.2
Ours–Small	5M	65.8	92.7	71.6	95.5	52.1	96.4
Ours–Base	15M	64.5	93.4	72.5	95.9	53.8	97.2
Ours–Large	37M	**66.7**	**94.6**	73.1	**96.1**	**54.8**	**98.3**

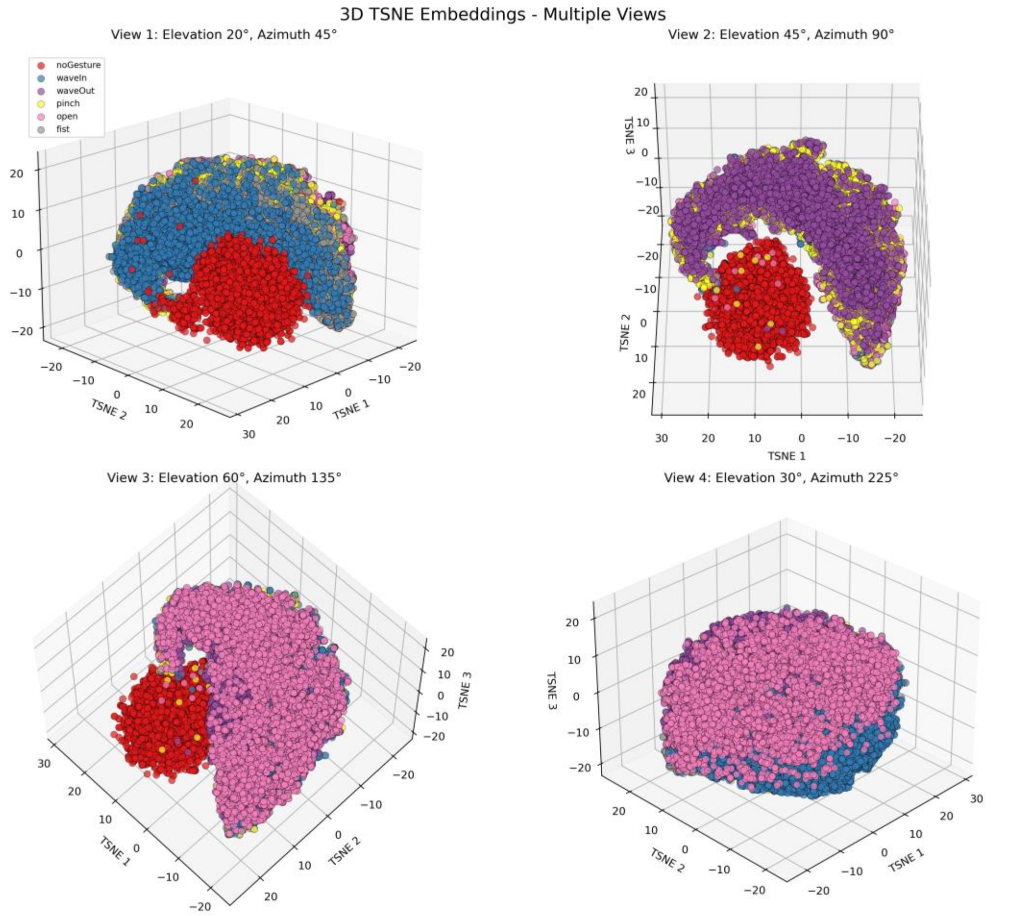
Best Performance | Our Methods | Small Model (≤15M) | Medium Model (15-50M) | Large Model (>50M)

1. In the ECG **multi-label** task, every ECG sample may exhibit several abnormalities;
2. The model must output all relevant diagnoses from a set of SNOMED-CT labels.
3. Performance is reported with macro-averaged F1 and AUROC over this full label set.

Experiments & Results

Visualization of EMG downstream tasks

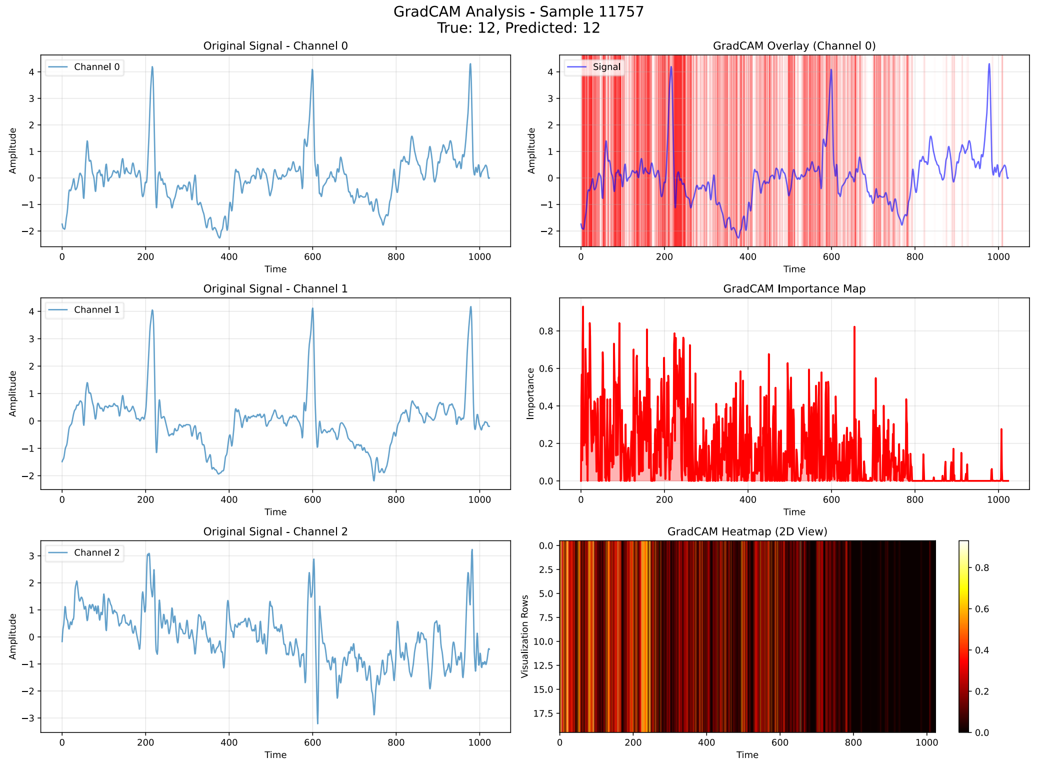




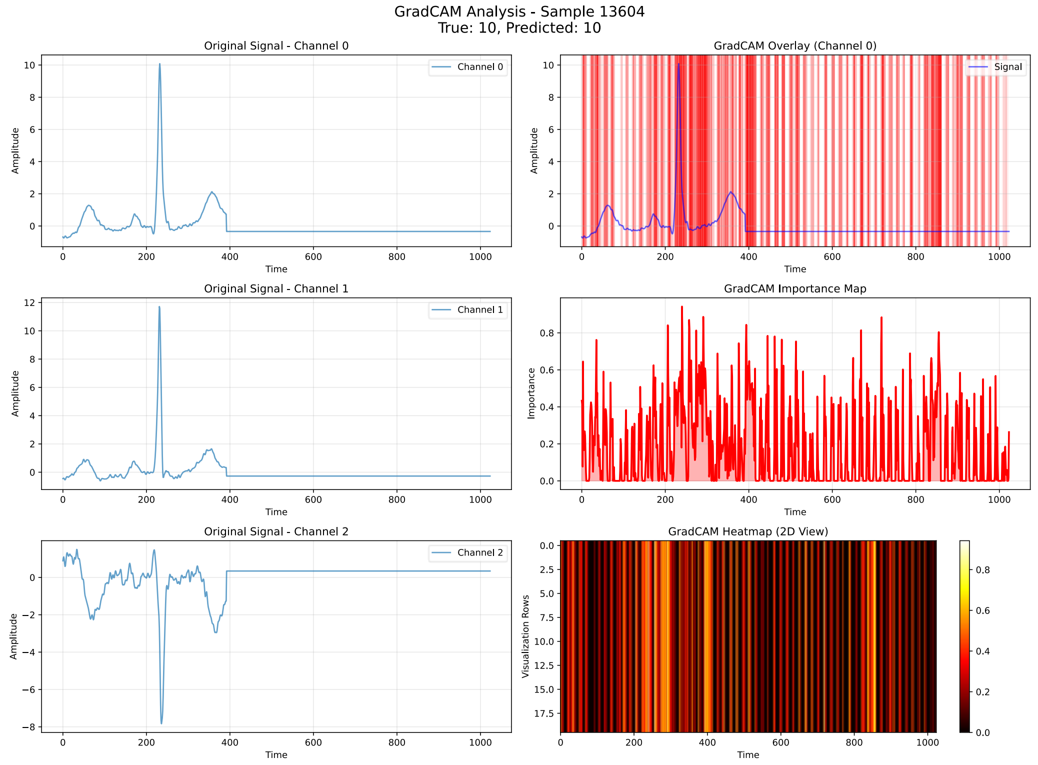
3D t-SNE embeddings visualization from multiple viewing

Experiments & Results

GradCAM visualization results

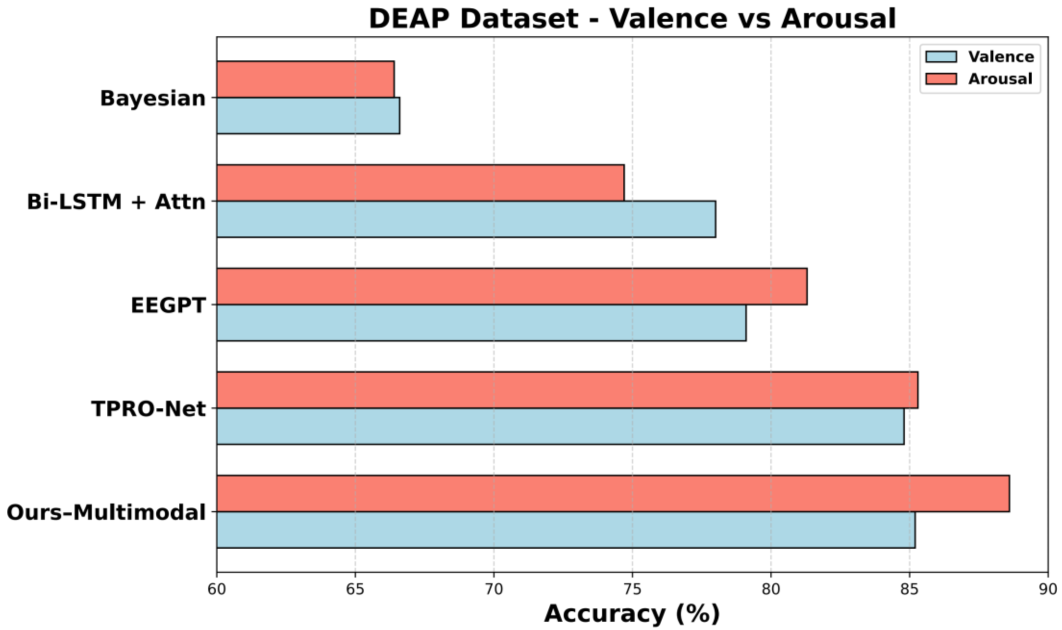


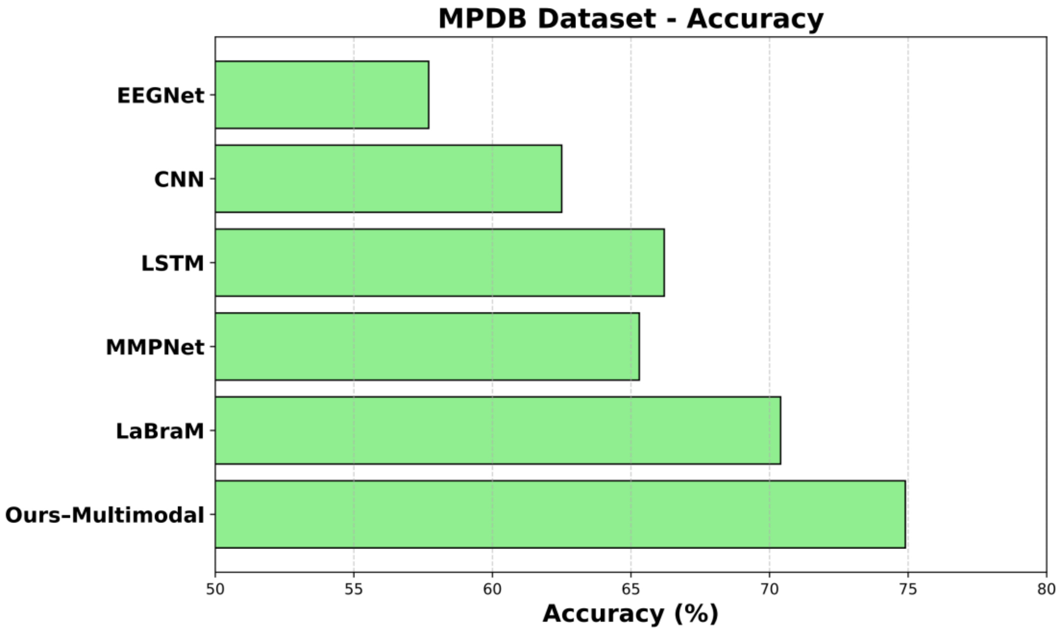
T-wave Inversion (TInv)



ST-segment Depression (STD)

Experiments & Results





😊DEAP gathers simultaneous EEG and EMG while subjects watch music videos designed to evoke specific emotions.
🚗MPDB studies driver behaviour using synchronized EEG, EMG, and ECG signals.(Smooth driving, Acceleration, Deceleration, Lane-change, Turning)

Conclusion & Future Work

🔬 Expanding signal modalities

- extending signals like EEG and PPG, creating comprehensive coverage of physiological data types.

🤖 Developing a universal multimodal model

- moving from separate models for different signals to a single, unified architecture that can seamlessly process all biosignal modalities simultaneously, functioning as a "physiological diagnostic system."

🔗 Enhanced integration capabilities

- improving the multi-modal fusion mechanisms to create a truly generalized framework that can handle diverse sensor inputs and cross-modal interactions for robust clinical applications.

Q&A

Yanlong Chen yanlchen@student-ethz.ch

Institut für Integrierte Systeme – ETH Zürich
Gloriastrasse 35
Zürich, Switzerland

DEI – Universitá di Bologna
Viale del Risorgimento 2
Bologna, Italy