LUNA: Efficient and Topology-Agnostic Foundation Model for EEG Signal Analysis

Berkay Döner[1] **Thorir Mar Ingolfsson[1]** Luca Benini[1] Yawei Li[1]

[1]Integrated Systems Laboratory, ETH Zürich, Zürich, Switzerland

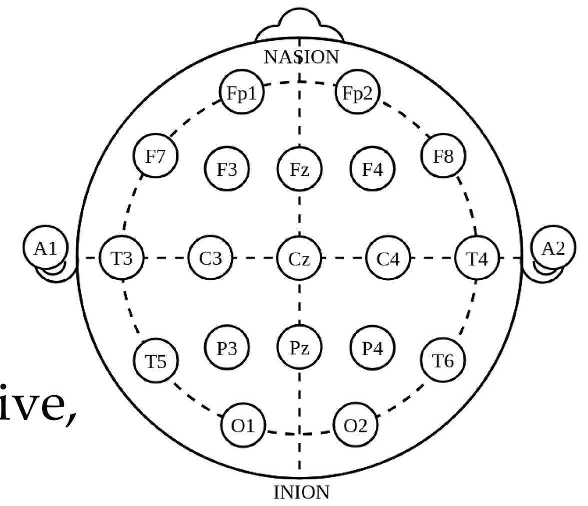
ETH zürich

NEURAL INFORMATION PROCESSING SYSTEMS

See our paper and code!

Paper Code

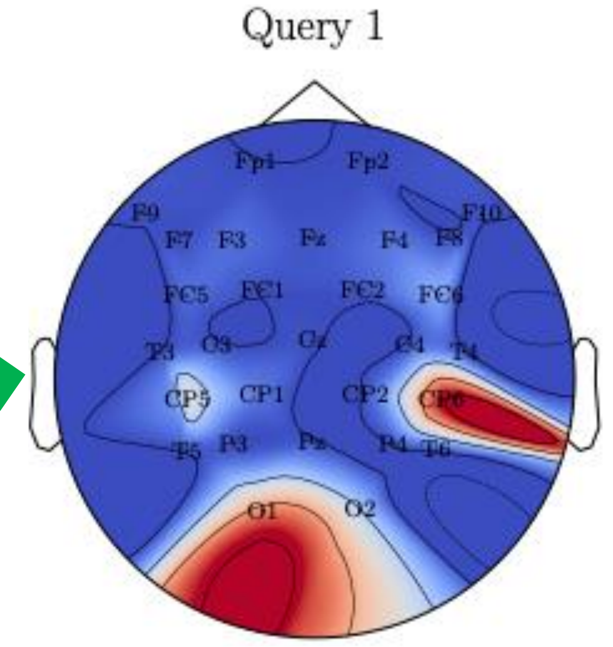
The Challenge With EEG Data



- **Topological Heterogeneity:** Electroencephalography (EEG) datasets vary significantly in the number and placement of electrodes, which limits the generalization of traditional analysis models.
- **Computational Bottlenecks:** Existing methods are often computationally intensive, with complexity that can increase quadratically with the number of channels, making large-scale analysis difficult.
- **Variable Data Structures:** EEG recordings also differ in duration and sampling rates, adding another layer of complexity.

Our Solution: LUNA

LUNA is a self-supervised foundation model designed to overcome these challenges. It projects multi-channel EEG data into a fixed-size, topology-agnostic latent space.



- **Channel Unification**: LUNA employs **learned queries** that use cross-attention to interact with EEG features from a variable number of channels.
- **Dual-Loss Pre-training**: The model is pre-trained using a combination of a masked-patch reconstruction loss and an auxiliary query specialization loss to encourage diverse and informative representations.

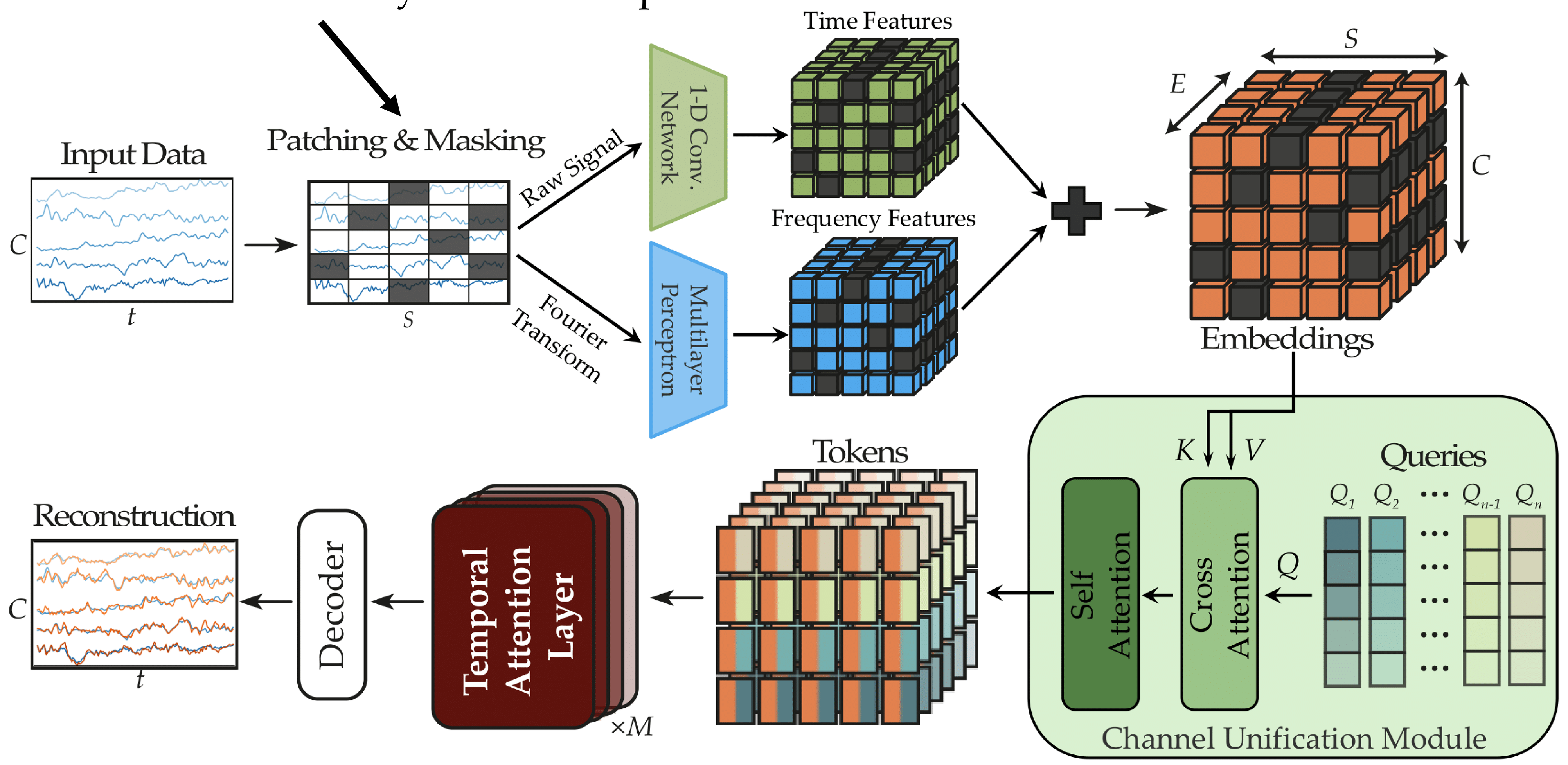
$$\mathcal{L}_{rec} = \frac{1}{N_{masked}} \sum_{i \in M} \text{SmoothL1}(x_{orig_i}, x_{recons_i}) + \alpha \cdot \frac{1}{N_{visible}} \sum_{i \notin M} \text{SmoothL1}(x_{orig_i}, x_{recons_i})$$

$$\mathcal{L}_{total} = \mathcal{L}_{rec} + \lambda_{spec} \mathcal{L}_{spec}$$

Pre-training Data

TUEG (20/22 Channels), **Siena** (29 Channels) >21,000 hours of raw EEG

Masked patch reconstruction with an auxiliary **Query Specialization Loss** to force diversity in learned spatial filters.



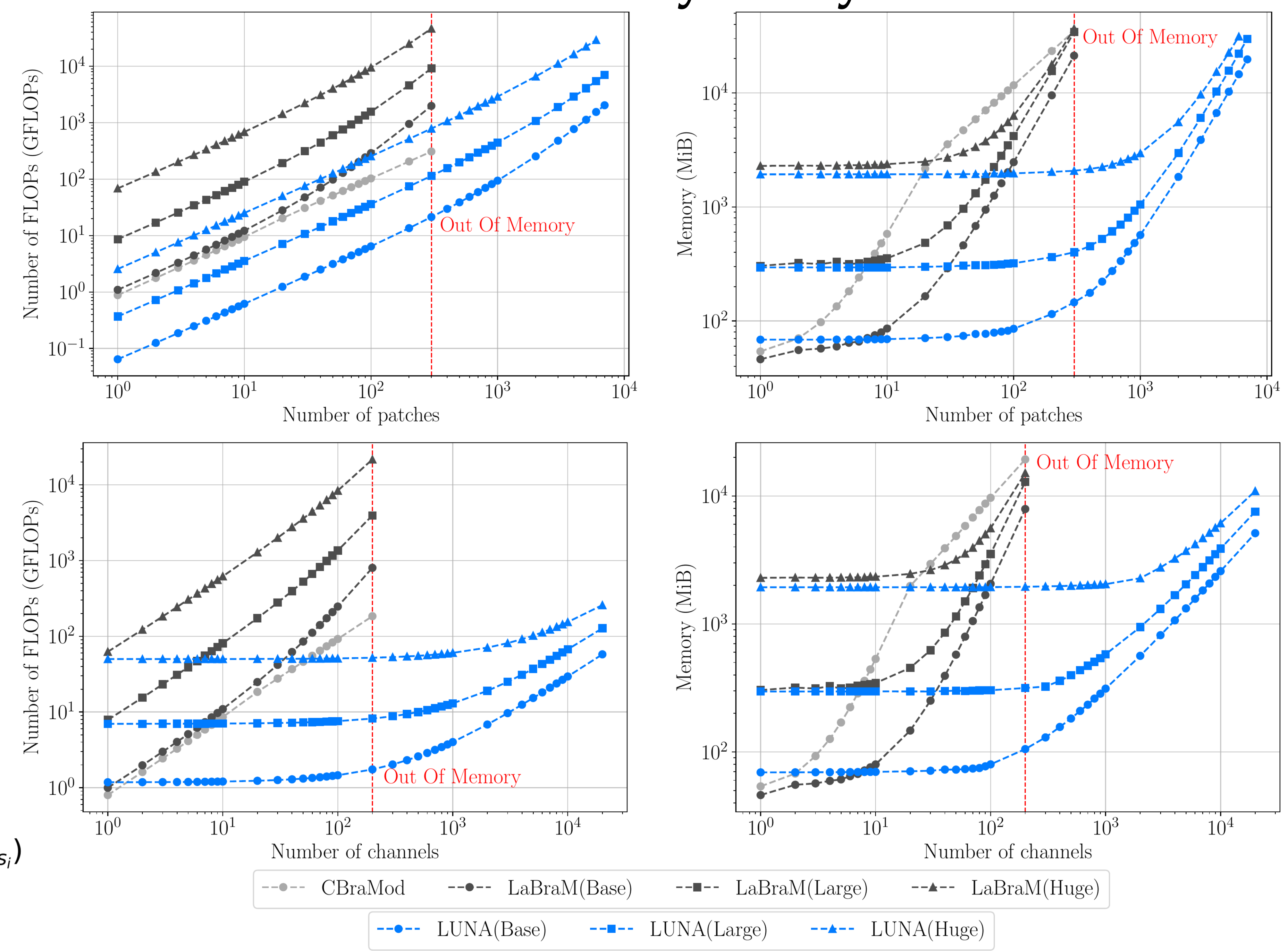
3 model sizes of LUNA

Base	Large	Huge
7M	43M	311M

The Core Innovation:
Topology-Invariant Encoder: Standard Transformers scale quadratically $O\left((S \times C)^2\right)$ with channels **Solution:** We project variable channels into a **fixed latent space Q** using learned queries.

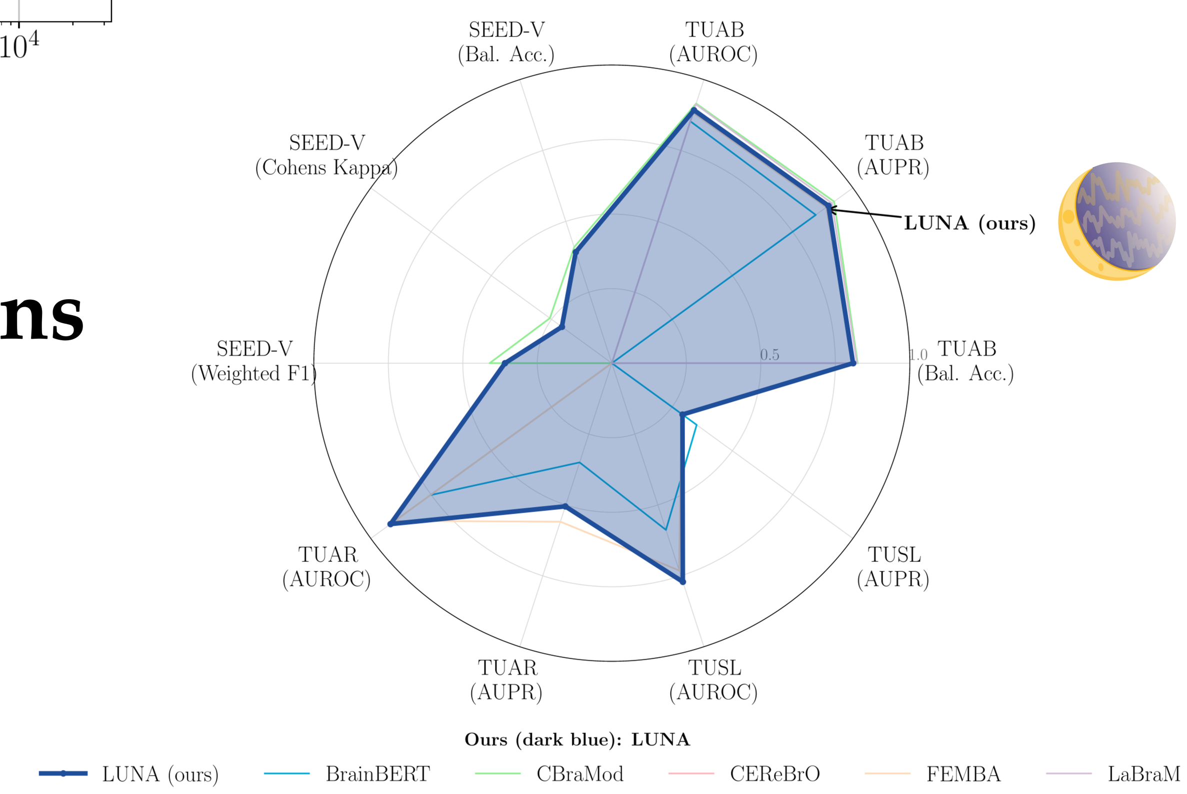
Result: This decouples computational cost from electrode count, enabling linear scaling.

Efficiency Analysis



Linear Scaling: LUNA (blue lines) maintains constant compute cost even as channel density increases, unlike baselines (grey lines).

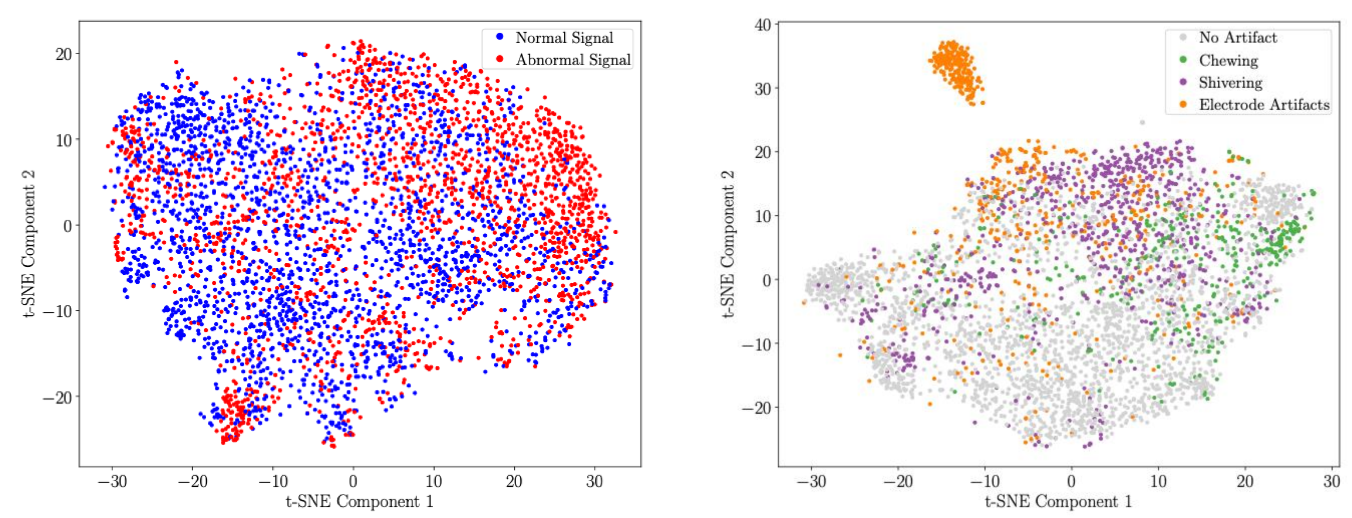
Resource Savings: Reduces FLOPs by **300×** and GPU memory by **10×** compared to LaBraM-Huge. * **Impact:** Enables training on high-density caps where other models run out of memory.



Key Results & Contributions

High Performance Across Benchmarks:
- Achieves **state-of-the-art performance** with a **0.921 AUROC** on the **TUAR** benchmark.
- Demonstrates highly competitive results in tasks such as **artifact detection**, **slowing classification**, and **emotion recognition**.
- Allow **query interpretability** by looking at learned queries.

Unprecedented Efficiency:
- Reduces computational load by **300x in FLOPs**
- Saves up to **10x in GPU memory usage**, enabling scalable analysis of large datasets.



t-SNE of **LUNA embeddings** (**before** finetuning)
Clustering by signal type already visible