1ST WORKSHOP ON ARTIFICIAL INTELLIGENCE
FOR BIOMEDICAL DATA - AIBIO

Self-Attention as a Predictor of EEG Anomalies

Natalia Koliou (presenter),
Maria Sierra, Christoforos Romesis,
Stasinos Konstantopoulos, and Luis Montesano

MANOLO
CLOUD-EDGE EFFICIENT & TRUSTWORTHY AI

Funded by
the European Union

25 Oct 2025

Overview

1. Introduction & Background
2. Research Methodology
3. Experiments
4. Results & Discussion
5. Conclusion & Future Work

Introduction & Background

Introduction & Background

- EEG signals are highly sensitive, low-amplitude recordings easily contaminated by **artifacts** (noise not generated by brain activity).
- Artifacts may arise from physiological (muscle tension, sweating) or technical sources (electrode detachment).
- What counts as "noise" is often **task-dependent**: a feature irrelevant for one application may be meaningful for another.
- Traditional denoising via **reconstruction/prediction errors** (Autoencoders, LSTMs) ignores contextual relevance.
- **Our idea:** Train attention-based models on a downstream task (e.g., sleep stage classification) and use its attention patterns to infer anomalies without explicit artifact labels.

Research Methodology

Self-Attention as a Predictor of EEG Anomalies

Research Methodology

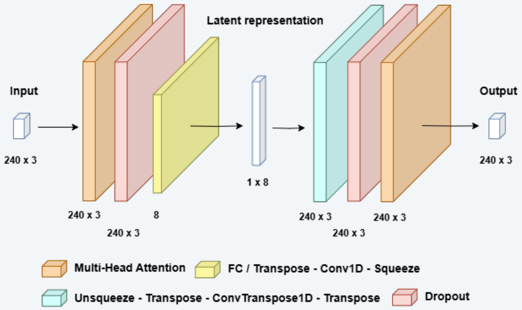
- Traditional anomaly detection uses **reconstruction or prediction errors**, but these often miss **context-specific anomalies**.
- **Reconstruction error** can fail because models may learn to reproduce artifacts if trained on contaminated data (false negatives).
- **Prediction error** can incorrectly flag unpredictable but normal variations as anomalies (false positives).
- The **attention-based approach** assumes that **low-attention regions** correspond to noisy or irrelevant segments.

Hypothesis: during a task (e.g., reconstruction or prediction), the model learns to ignore artifacts that hinder performance.

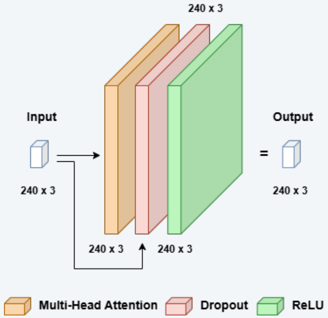
Research Methodology

Acronym	*Architecture*	*Detection*
LSTM *C-LSTM* *AE_err*	LSTM Autoencoder Convolutional LSTM Autoencoder Attention-based Autoencoder	Reconstruction error
TP_err	Transformer Predictor	Prediction error
AE_att *TP_att*	Attention-based Autoencoder Transformer Predictor	Attention
MNE	IIR filter	

Research Methodology



AE_err/AE_att architecture



Left:

TP_err/TP_att architecture

Experiments

Data Preprocessing

Our approach was evaluated using the **BOAS dataset**:

- EEG signals recorded from a **two-channel headband** at **128 Hz**.
- Signals segmented into **30s epochs**, each labeled into five sleep stages (Wake, N1, N2, N3, REM).
- **56 overnight recordings** from different users in total; 43 train, 3 validation, 10 test.
- Data normalized using **median and IQR** for robustness (skipped for MNE to preserve signal shape).
- Ground truth from Bitbrain's **proprietary artifact estimation algorithm**.

Note: 6 test recordings have minimal noise; comparisons focus on the remaining 4.

Training Process & Hyperparameters

- EEG segments (30s, 3840 samples/channel) split into **16 chunks of 240 samples** for AE and TP methods.
- Custom **BlendedLoss** balances trend (median) and local pattern (mean) errors:

$$\text{Loss} = (1-b)\cdot \text{median}(|\hat{x}-x|^{p}) + b \cdot \text{mean}(|\hat{x}-x|^{p})$$

- Methods compared:
  1. **MNE**: expert-designed EEG noise filter.
  2. **AE_att / TP_att**: anomalies inferred via attention weights.
  3. **LSTM, C-LSTM, AE_err, TP_err**: anomalies inferred via reconstruction/prediction errors.
- **Training setup:** batch size: 512, epochs: 1000, patience: 30, optimizer: Adam, lr: 1e-4, scheduler: ReduceLROnPlateau.

Results & Discussion

Results & Discussion

- Three evaluation setups: **30s, 5min, 10min windows**.
- Window marked as **noisy** if any sample is annotated as noise.
- Predictions binarized with **threshold**: 1% of training data marked as noise.
- **F2-score ($\beta = 2$)** emphasizes recall to account for the strong class imbalance.
- **Longer windows** improve results due to denser positives and more context.
- Attention-based methods (AE_att, TP_att) outperform error-based counterparts (AE_err, TP_err).
- TP_att outperforms MNE on some recordings, likely due to better handling of **uncorrelated channel segments**.

Attention is a promising indicator for anomaly detection in EEG signals.

Conclusion & Future Work

Conclusion

- We extract **anomaly indicators** from intermediate layers of deep networks trained on sequence tasks, without needing explicit anomaly labels.
- Validated on **EEG signals** from wearable sleep monitoring devices.
- Attention layers show promising performance for anomaly detection.

Future Work

- Analyze **attention components** (Q, K, V) to understand which parts focus on ignoring noisy patterns.
- Investigate model robustness to **uncorrelated/noisy** segments.
- Extend experiments to **diverse** datasets and tasks.
- Explore **multi-task training** to better distinguish between noise and task-irrelevant segments.
- Use **explainable ML** tools to interpret attention-based anomaly detection more transparently.

Thank you for your attention!

Thank you for your attention! Any questions?

Natalia Kolliou, Christoforos Romesis, Stasinos Konstantopoulos

Maria Sierra, Luis Montesano

This research was co-funded by the European Union under GA no. 101135782 (MANOLO project). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or CNECT. Neither the European Union nor CNECT can be held responsible for them.

Funded by the European Union

MANOLO

CLOUD-EDGE EFFICIENT & TRUSTWORTHY AI

https://manolo-project.eu