Master Thesis Proposal

Topic: Self-Supervised Transformer Model for EEG Data

Biomedical signals, particularly electroencephalogram (EEG) signals, have gained significant attention in artificial intelligence (AI)-powered models, including conventional machine learning (ML) and deep learning (DL), demonstrating promising results [1]. However, these models often struggle with challenges such as limited data availability, improper distribution, and storage constraints [1]. Labeled EEG data is especially scarce due to the high cost of acquisition and annotation, which limits the performance of deep learning models trained on EEG data.

To address this, we explore self-supervised pretraining for EEG data in the context of depth of anesthesia detection. Self-supervised learning (SSL) has shown potential in various EEG applications. Kostas et al. [2] applied it to brain-computer interaction (BCI) classification, while Das et al. [4] demonstrated the effectiveness of self-supervised pretraining for epileptic seizure detection from EEG data.

This thesis examines whether self-supervised pretraining can enhance biosignal classification, specifically for EEG-based depth of anesthesia monitoring. Using the open-source VitalDB dataset [5], the research will focus on selecting suitable pretraining strategies and model architectures.

Goals

- Literature review on self-supervised transformer models for EEG, with a focus on low channels setups
- Training of a self-supervised transformer model on EEG signals of the Vitaldb database
- Evaluation of the model on internal dataset depth of anesthesia classification dataset

The proposed work consists of the following parts:

- **Work Package 0 – Getting Started:**
  - Download Zotero (tool for literature research)
  - Understand the Prisma Method (used for literature research)
- **Work Package 1 – Get Familiar with VitalDB and Time Series:**
  - Read and analyze:
    - EEG Based Monitoring of General Anesthesia: Taking the Next Steps [12]
    - Deep learning for time series classification: a review [7]
  - Examine the Vitaldb [1] dataset and identify relevant data types for anesthesia monitoring.
- **Work Package 2 – Literature Review:**
  - Investigate recent pre-trained foundational time-series models (e.g. Chronos [9], [10]) and self-supervised Transformer Models for EEG and choose at least two for further investigation.
  - Compile bibliography on self-supervised pretraining strategies for biomedical signals/time series in general (e.g. [8]) and select at least three combination of pre-training tasks to investigate further: they can

  - be self-supervised or use other available signals from the VitalDB dataset (BIS score, EMG…)
  - Select at least 2 preprocessing/encoding techniques for the EEG data

- **Work Package 3: Model Training and Evaluation**

  Objective: Train or fine-tune a self-supervised Transformer model on EEG data.

  - Prepare the pre-training dataset from **VitalDB** (Choose relevant patients, create train/val/test splits, segment EEG signals into meaningful time windows suitable for training, and extract and preprocess EEG signals and relevant other data)

  - Select at least two appropriate **self-supervised transformer model architectures** (e.g., different encoding techniques).
  - Train or finetune the model on **VitalDB EEG signals** using the selected pre-training tasks.
  - Evaluate performance on the **internal depth of anesthesia classification dataset**.

- **Work Package 4: Model Analysis & Interpretation**

  - Perform quantitative analysis: F1-Score, accuracy, sensitivity, specificity, AUC-ROC, etc.
  - Conduct qualitative analysis
  - Compare self-supervised vs. supervised model performance

The thesis must contain a detailed description of all developed and used algorithms as well as a profound result evaluation and discussion. The implemented code has to be documented and provided. An extended research on literature, existing patents and related work in the corresponding areas has to be performed.

Advisors: Dr. Emmanuelle Salin, Arijana Bohr M.Sc., Prof. Dr. Bjoern Eskofier
Student: Zeeshan Ahmad
Start—End: 30.06.2025—30.12.2025

References

[1] VitalDB: https://vitaldb.net/
[2] Kostas, D., Aroca-Ouellette, S. and Rudzicz, F., 2021. BENDR: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data. *Frontiers in Human Neuroscience*, *15*, p.653659.
[3] Rafiei, M.H., Gauthier, L.V., Adeli, H. and Takabi, D., 2022. Self-supervised learning for electroencephalography. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(2), pp.1457-1471.
[4] Das, S., Pandey, P. and Miyapuram, K.P., 2022. Improving self-supervised pretraining models for epileptic seizure detection from EEG data. *arXiv preprint arXiv:2207.06911*.

[5] Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L. and Muller, P.A., 2019. Deep learning for time series classification: a review. *Data mining and knowledge discovery*, *33*(4), pp.917-963.

[6] Prabhakar Kamarthi, H. and Prakash, B.A., 2024. Large Pre-trained time series models for cross-domain Time series analysis tasks. *Advances in Neural Information Processing Systems*, *37*, pp.56190-56214.

[7] Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L. and Muller, P.A., 2019. Deep learning for time series classification: a review. *Data mining and knowledge discovery*, *33*(4), pp.917-963.

[8] Zhang, X., Zhao, Z., Tsiligkaridis, T. and Zitnik, M., 2022. Self-supervised contrastive pre-training for time series via time-frequency consistency. *Advances in neural information processing systems*, *35*, pp.3988-4003.

[9] Ansari, A.F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., Shchur, O., Rangapuram, S., Arango, S.P., Kapoor, S. and Zschiegner, J., 2024. Chronos: Learning the language of time series.

[10]Liang, Y., Wen, H., Nie, Y., Jiang, Y., Jin, M., Song, D., Pan, S. and Wen, Q., 2024, August. Foundation models for time series analysis: A tutorial and survey. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 6555-6565).

[11] Tipirneni, S. and Reddy, C.K., 2022. Self-supervised transformer for sparse and irregularly sampled multivariate clinical time-series. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *16*(6), pp.1-17.

[12]Kreuzer, M., 2017. EEG based monitoring of general anesthesia: taking the next steps. *Frontiers in computational neuroscience*, *11*, p.56.