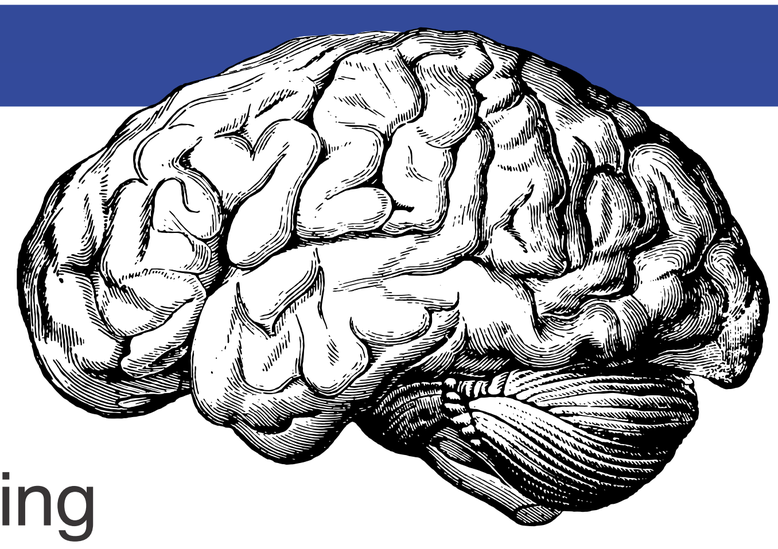
# Deep EEG



## Deep Learning for Infant-Brain EEG Modelling

Electroencephalography (EEG) is an electrical measure of brain activity. It can be used to monitor and predict neurodevelopmental trajectories. In this project, we present the first evaluation of using transformer- and state-space-based models with infant EEG data to predict neurodevelopmental outcomes, as measured by Bayley Scales of Infant Development scores.

## 1 Research questions

1. How well can deep learning models predict neurodevelopmental outcomes as measured by the Bayley Scales subtests?

2. How do pretrained models perform compared to models trained from scratch in predicting neurodevelopmental outcomes?

Assessing the models' predictive capabilities is the primary focus, but we also extensively investigate model performance to provide useful insights for future considerations.

## 2 Dataset and Spectral Analysis

### Bayley Scales Scores

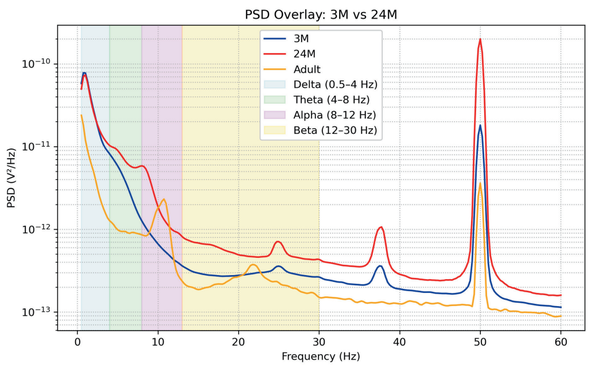
- The Bayley Scales are a diagnostic tool for assessing early childhood neurodevelopment.
- Part of the test assesses cognitive, language, and motor function in infants.
- Standardised scores are on a scale with a mean of 100 and standard deviation of 15, with scores less than 85 on any test considered "developmentally delayed", and "developmentally typical" otherwise.

### Khula EEG Dataset

- Part of a more general study to chart the cognitive development of infants from African populations.
- Consists of **1041 EEG recordings collected from 321 South African infants at 3, 6, 12 and 24 months of age.**

### Spectral Analysis of EEG data

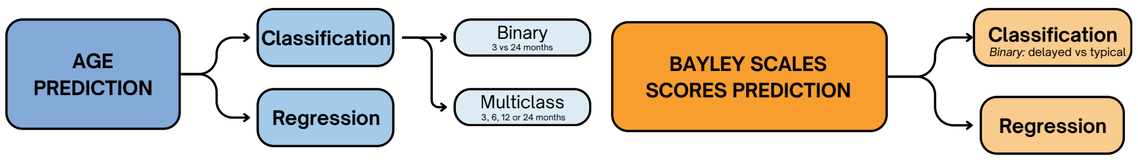
- We computed the power spectral density (PSD) of the EEG samples to analyse changes across age groups which a discriminative machine learning model may detect.
- We found changes in the alpha, delta and gamma frequency bands over time, consistent with current literature.




## 3 Methodology

### Prediction tasks

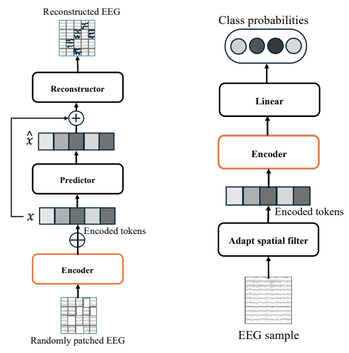
We attempted the easier task of age prediction in addition to Bayley Scales score prediction:




### Model Architectures

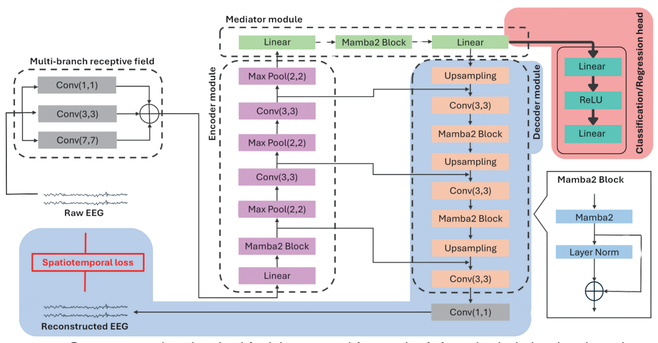
We adapted an existing transformer-based model and state-space model:

#### Transformer model: EEGPT




Pretraining (left) and finetuning (right) architectural setup.

#### State-space model: EEGM2

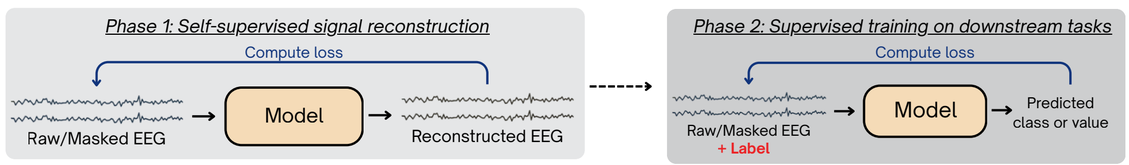



Components shaded in blue used in pretraining, but detached and replaced with components shaded in red when doing downstream tasks.

Logistic regression and linear regression baselines (for classification and regression tasks, respectively) were implemented, using **manual feature extraction** with the PSD as features.

### Training strategies

For each architecture, we train models **from scratch** and use **pretrained models** in two phases:




## 4 Results

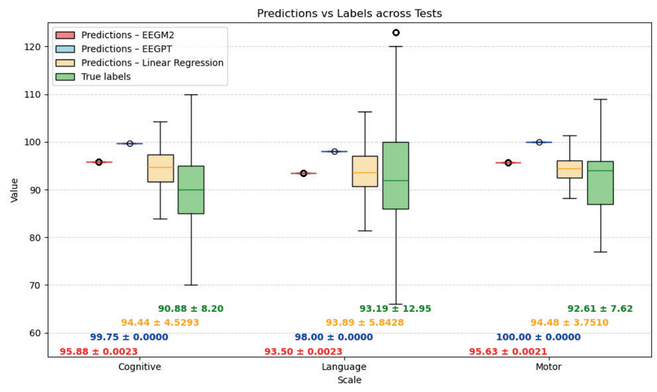
### Classification tasks

| Task | Model | Accuracy | Balanced Accuracy |
|---|---|---|---|
| Multi-class age | EEGPT (FT) | $0.3328 \pm 0.1070$ | $0.3123 \pm 0.1211$ |
| | EEGM2 (scratch) | $0.2829 \pm 0.0032$ | $0.2500 \pm 0.0000$ |
| | Logistic Regression | $\mathbf{0.6099 \pm 0.0000}$ | $\mathbf{0.6159 \pm 0.0000}$ |
| Binary age | EEGPT (FT) | $0.5828 \pm 0.0754$ | $0.5525 \pm 0.1049$ |
| | EEGM2 (FT) | $0.5081 \pm 0.0030$ | $0.5000 \pm 0.0000$ |
| | Logistic Regression | $\mathbf{0.9325 \pm 0.0000}$ | $\mathbf{0.9325 \pm 0.0000}$ |
| Bayley Scale scores | EEGM2 (FT) | $0.5472 \pm 0.2313$ | $0.5000 \pm 0.0000$ |
| | Logistic Regression | $\mathbf{0.6708 \pm 0.0000}$ | $\mathbf{0.5499 \pm 0.0000}$ |

Table: Best results in **boldface.** "FT" is the finetuned model, and "scratch" is the model trained from scratch.

- **Transformer and SSM models both perform poorly**, collapsing to majority class predictions on all classification tasks.
- Feature-based logistic regression model performs well on age prediction (but also poorly on all other tasks). This indicates that **simpler models with manual feature extraction may be more appropriate in the context of smaller datasets.**

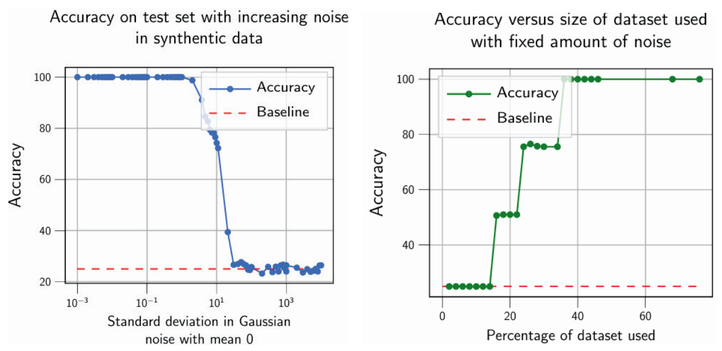
### Regression tasks




#### Key findings

- Deep models fail to capture the true distribution of scores, instead collapsing to values within the mean.
- Linear regression better captures the score distribution, which suggests that simpler, feature-engineered implementations are more suited for modestly sized datasets.

### Synthetic data task



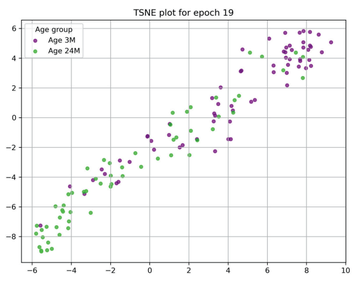

#### Motivation and findings

- We used synthetic data to investigate the poor model performance in a controlled setting.
- We find that **both transformer and SSM models achieve perfect classification accuracy provided sufficient data and a high enough signal-to-noise ratio.**
- This suggests that the dataset is too small for these complex models.

### Intrinsic Evaluations

#### EEGPT

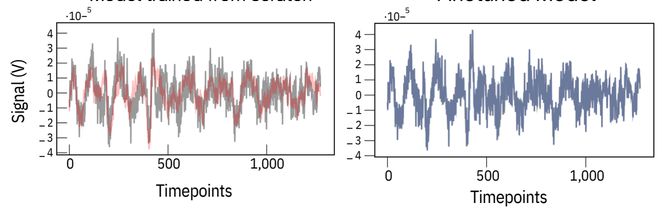
The intermediate feature representations reveal that the model does not recognise similarities amongst data samples.




However, the **fine-tuned** model on **binary classification** reveals clearer clustering of same-age data samples.

#### EEGM2

We analyse the model's performance on signal reconstruction.




- The model performs well on signal reconstruction.
- **Transfer learning improves performance on signal reconstruction.**
- Poor performance on downstream tasks may be due to not using intermediate encoder information, which is used during pretraining.

## 5 Conclusions and future work

This work provides an **initial evaluation** of and **clarifies the current limitations** of deep learning models for brain developmental research using infant EEG.

- The transformer-based EEGPT and state-space model-based EEGM2 do not perform well on downstream tasks.
- Poor performance is due to insufficient data for these models or a low signal-to-noise ratio in the data.
- EEGM2 should combine features at multiple temporal and spectral resolutions.
- Use longer segments to provide models with more context for predictions.
- Prioritise using larger infant EEG datasets, preserving channel selection and montage consistency.


Department of Computer Science
University of Cape Town
Cape Town, South Africa

**Project Team**

**Caleb Bessit**
bsscal002@myuct.ac.za

**Tziyona Cohen**
chntzi001@myuct.ac.za

**Supervisors**
Dr Francois Meyer
Associate Professor Jonathan Shock