Jointly **decoding** multiple subjects with a **deep learning** model aided by **subject embeddings**

Uncovering **neuroscientifically interpretable** features from **nonlinear decoding** models

Paper: `arxiv.org/abs/2205.14102`

Code: `github.com/ricsinaruto/MEG-group-decode`

Decoding Across Subjects with Deep Transfer Learning

Richard Csaky[1,2,5], Mats W.J. van Es[1,2], Oiwi Parker Jones[2,3,4], Mark Woolrich[1,2]

University of Oxford; [1]Oxford Centre for Human Brain Activity, Department of Psychiatry; [2]Wellcome Centre for Integrative Neuroimaging; [3]Department of Engineering Science; [4]Jesus College; [5]Christ Church

BACKGROUND

- **Between-subject variability** of neuroimaging data **limits** the application of a single, **shared decoding model** across subjects [Olivetti et al., 2014] (Figure 1, right).
- **Previous approaches** include learnable affine transformations between subjects [Elango et al., 2017] and finetuning on target subjects [Cooney et al., 2019].

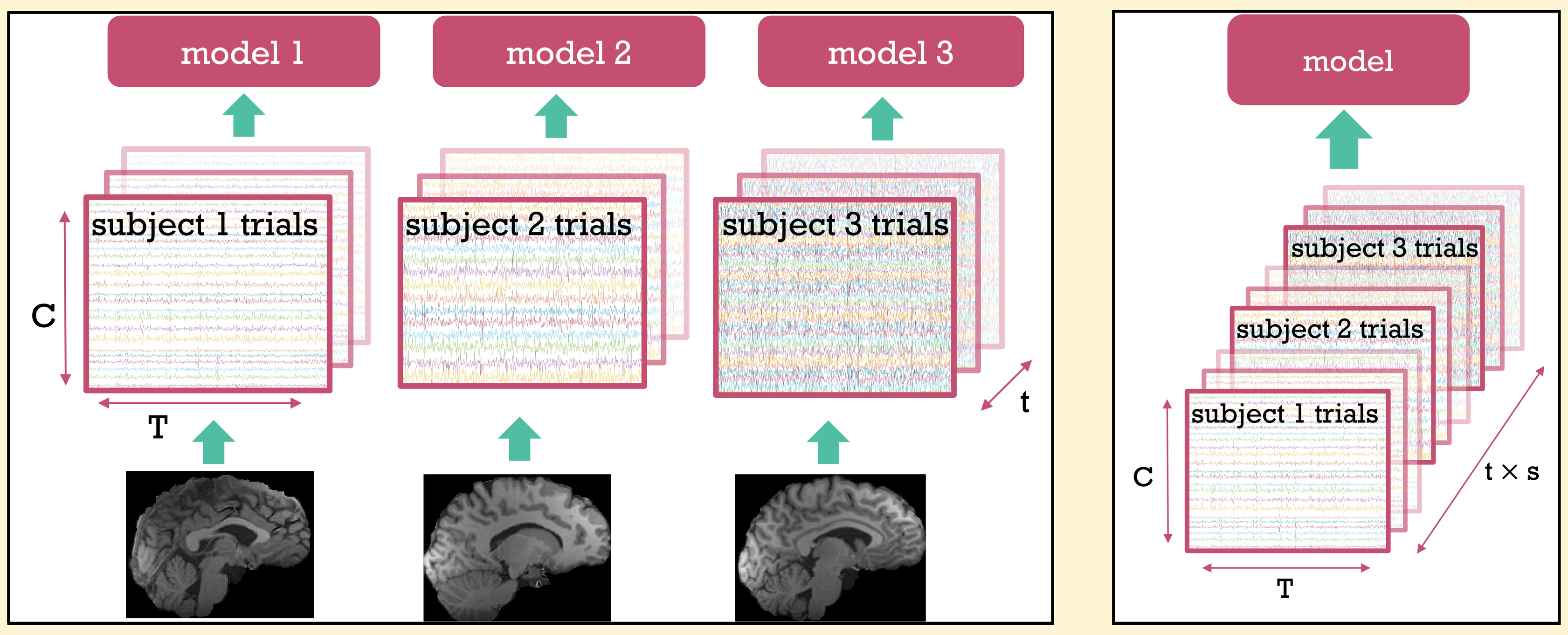

Figure 1: **Subject-level** (SL) and **group-level** (GL) modelling. **Left:** A **separate model** is trained on **each subject**. **Right:** A single, **shared model** is trained on the trials (**t**) of **all subjects (s)**.

Can we use deep learning to improve performance by using a shared model that generalises across subjects?

METHODS

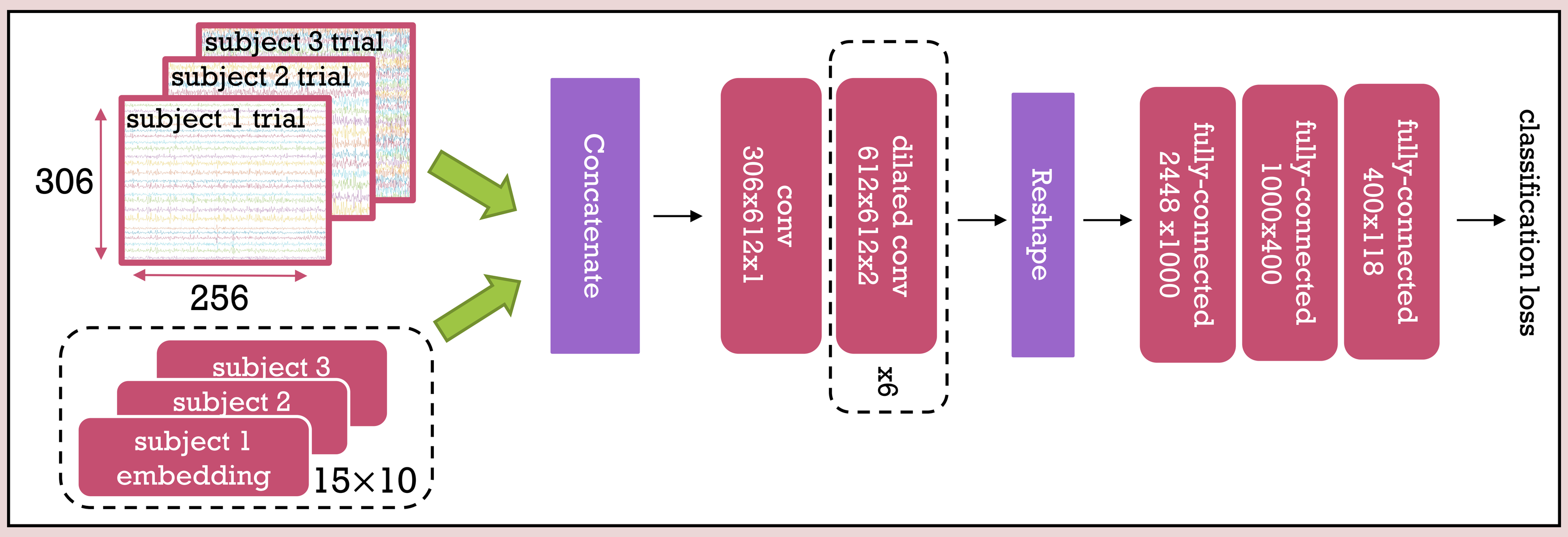

Figure 2: **Group-level WaveNet** [van den Oord et al., 2016] **Classifier with subject embeddings**. Dashes show differences between subject and group-level models. **Embeddings** of size 10 are concatenated with input trials to **provide information** about which trial is coming from which of the 15 **subjects**, tackling **between-subject variability**. The model should **learn** general features **across subjects**, and **adapt** its internal representations for **each subject**.

Model analysis

- **Spatial** and **temporal information** analysed with **permutation feature importance** (PFI), by permuting across timesteps (for each channel) and across channels (for each timestep).
- **PFI** with kernel output deviation as the measure to uncover **spatial**, **temporal**, and **spectral sensitivity** of individual **kernels**.
- In **spectral PFI**, **frequency** content is **disrupted** in specific bands by shuffling Fourier coefficients of Fourier-transformed input examples.

Experimental setup

- **Data**: task-Magnetoencephalography (**MEG**), where **15 subjects** view **118** different **images** with each image viewed **30 times** [Cichy et al., 2016].
- **Multiclass decoding** done on the **1-second epoch** post-stimulus using all **306 channels**.
- **Linear** (identity activation function) and **nonlinear** versions of subject-level and group-level **models** are **compared**.

RESULTS

Group models with subject embedding achieve similar accuracy to subject-level models

- At the **subject level** (SL), **linear** models are **better** than nonlinear.
- A **group-level** model **without** subject **embeddings** is much **worse** than SL models.
- Subject **embeddings help** a lot, closing the gap with SL models. **Nonlinearity** is crucial.
- **Finetuning** the **group model** on each subject separately **surpasses SL** models.

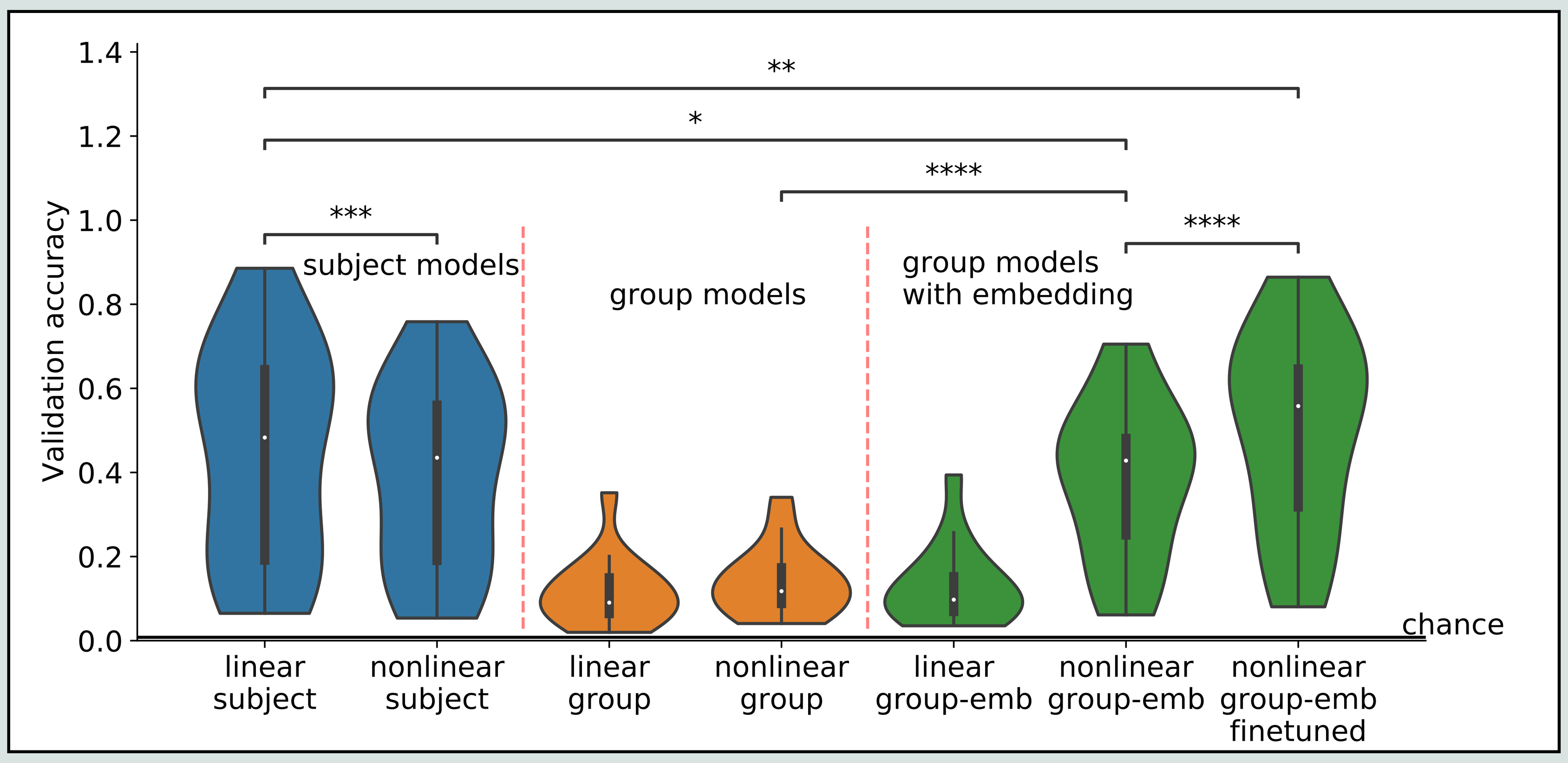

Figure 3: **Subject-level** and **group-level WaveNet Classifier** on the validation set of each subject. Train to validation ratio is 4:1 for each subject and class. `nonlinear group-emb finetuned` separately on each subject.

Group models generalise much better to new subjects than subject-level models

- **Group-level** (GL) models trained on 14 subjects are **above chance** on the **left-out subject**.
- **Finetuned GL** models **better** than training from **scratch** when using **<70%** of the left-out subject's **train set**.

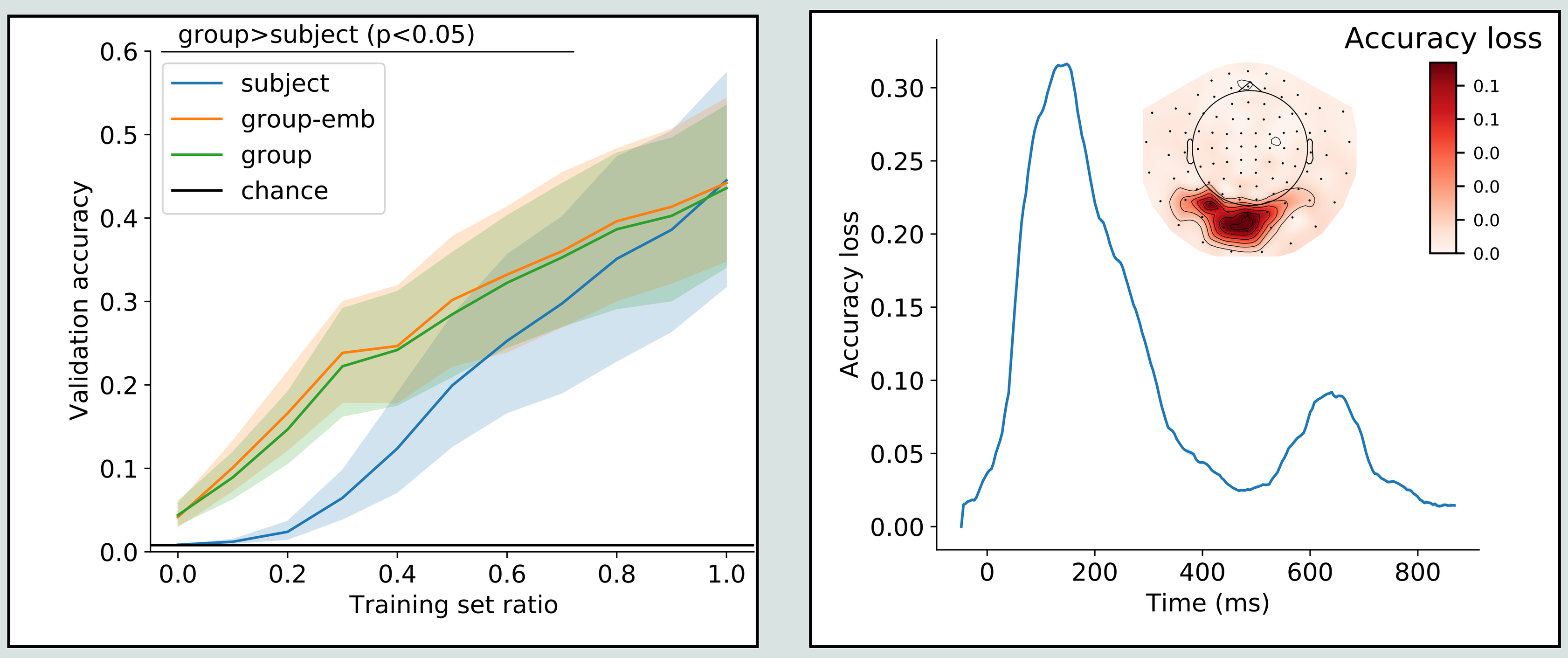

Figure 4. **Left: Generalisation and finetuning** (amount on horizontal axis) on **left-out subjects** (repeated across all subjects). `subject` is trained from scratch, while `group-emb` and `group` are initialised with a nonlinear GL model trained on the 14 other subjects.
Right: Temporal (accuracy loss w.r.t. trial timing) and **spatial** (accuracy loss w.r.t. channels) **PFI** for the `nonlinear group-emb` model.

Neuroscientific insights can be gained from deep learning models

- **Information** content (Figure 4, right) and **kernel sensitivity** (Figure 5, middle and left) peak at **100-150 ms** post-stimulus, and within channels over **visual areas**.
- **Kernel spectral sensitivity** (Figure 5, right) coincides with MEG PSD (**1/f, 10 Hz peak**).

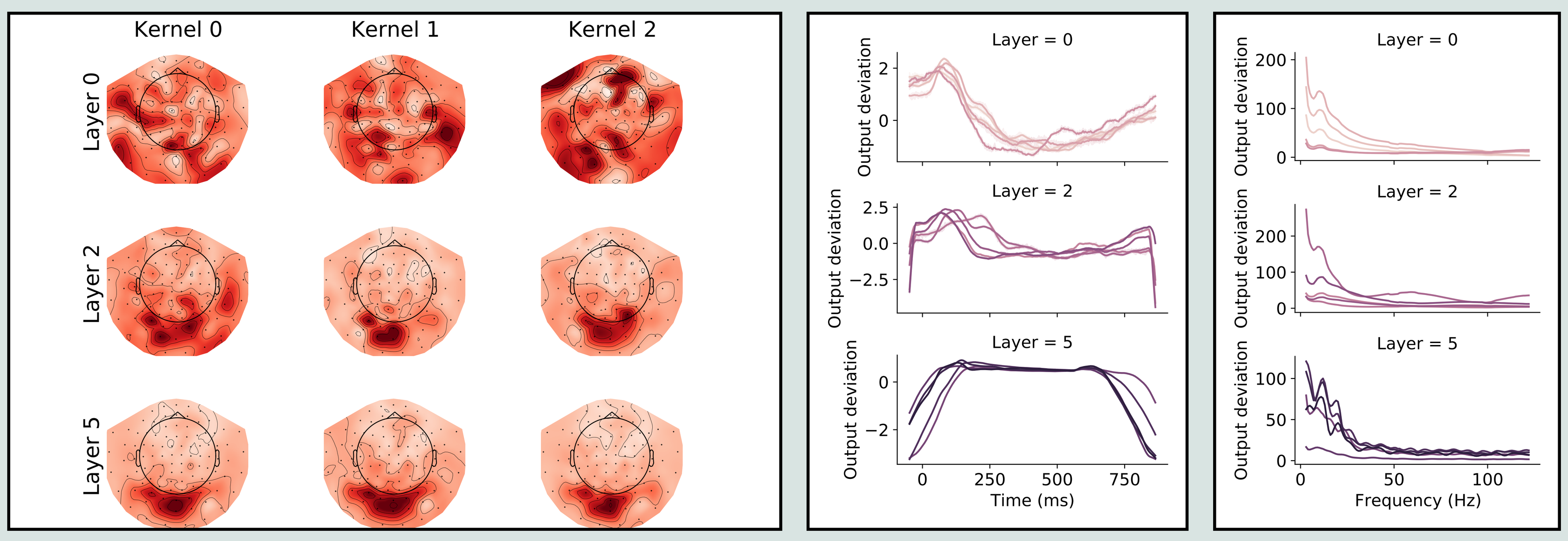

Figure 5. **Spatial** (left), **temporal** (middle), and **spectral** (right) **PFI** across `nonlinear group-emb` **kernels** within 3 layers (rows). **Kernel output** deviation w.r.t. spatial **location** (left, red shading), trial **timing** (middle), **frequency** (right). For temporal and spectral PFI individual kernels are lines.