# Retrieving Filter Spectra in CNN for Explainable Sleep Stage Classification

IEEE SPMB 2025

Stephan Goerttler[1,2], Yucheng Wang[2], Fei He[1] and Min Wu[2]

[1]Centre for Computational Science and Mathematical Modelling,
Coventry University, Coventry, UK

[2]Institute for Infocomm Research,
A*STAR, Singapore

6th December 2025

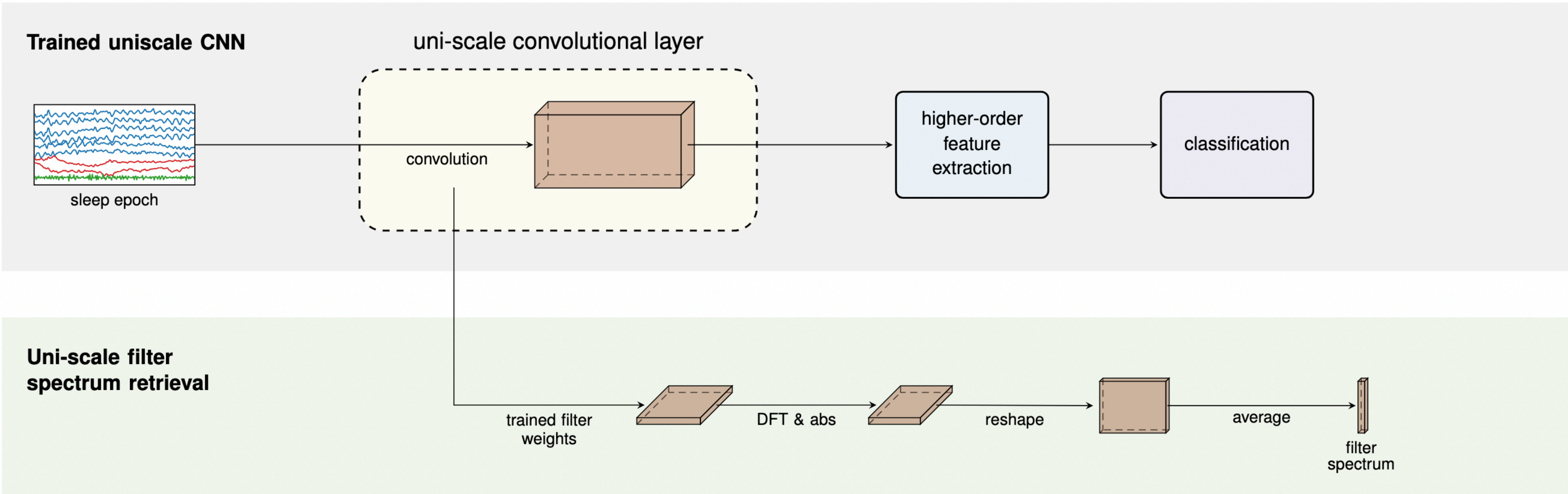
Coventry University

Agency for Science, Technology and Research
SINGAPORE

# Introduction

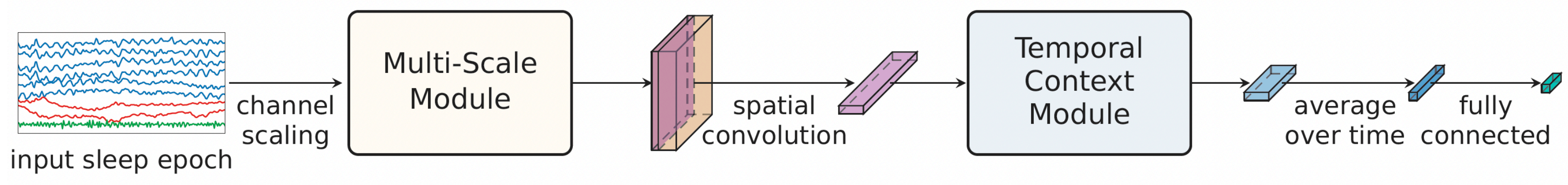
- *How does a convolutional layer see raw data?*
  - Spectral features and morphological features
  - Explainability tool for spectral processing
- Specific application in this study:
  - EEG-based polysomnography
  - Two models and two datasets
- Advantage:
  - Understand and improve model functioning

# Filter spectrum retrieval

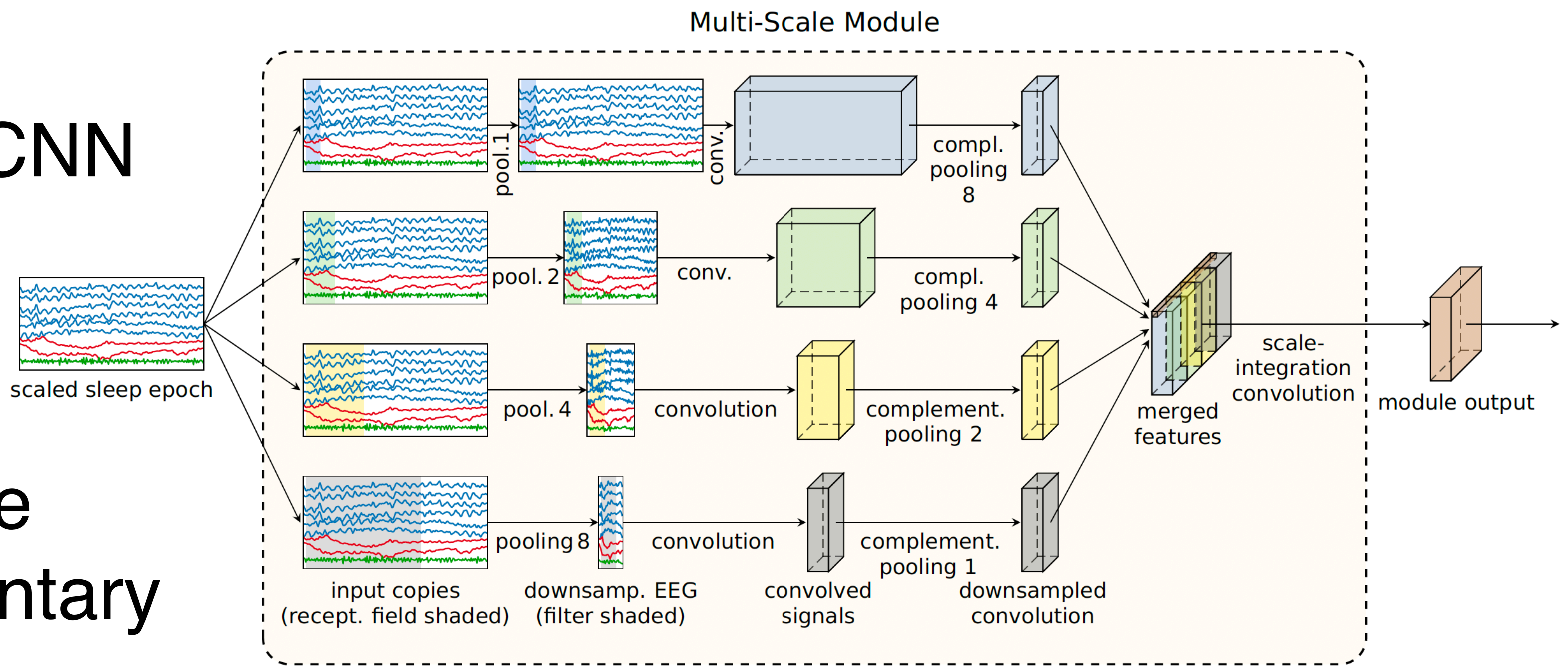



- Raw feature extraction: CNN-based
- Model I:
  - EEGNet (Lawhern *et al.*, [1])
- Filter spectrum retrieval:
  - Discrete Fourier transform, followed by averaging magnitudes

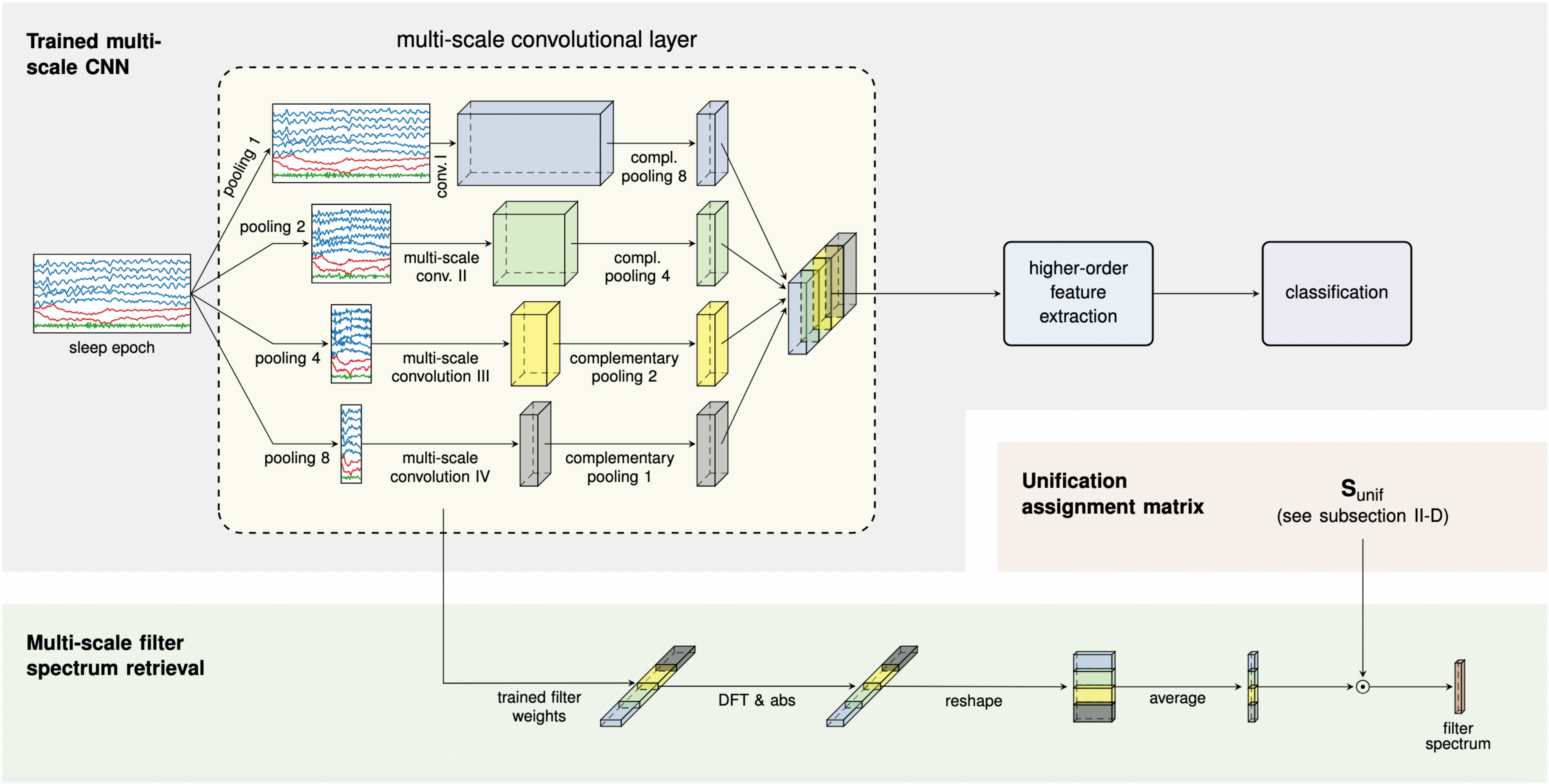
# Filter spectrum retrieval II




- Model II:
  - Multi-scale & attention CNN (MSA-CNN) [2]
- Special case:
  - Convolutions on multiple scales using complementary pooling
  - Requires combining the scales



# Multi-scale filter spectrum retrieval




- DFT & averaging magnitudes *per scale*
- Fourier frequencies determined by sampling rate and downsampling factor
- Use assignment matrix to combine the scales

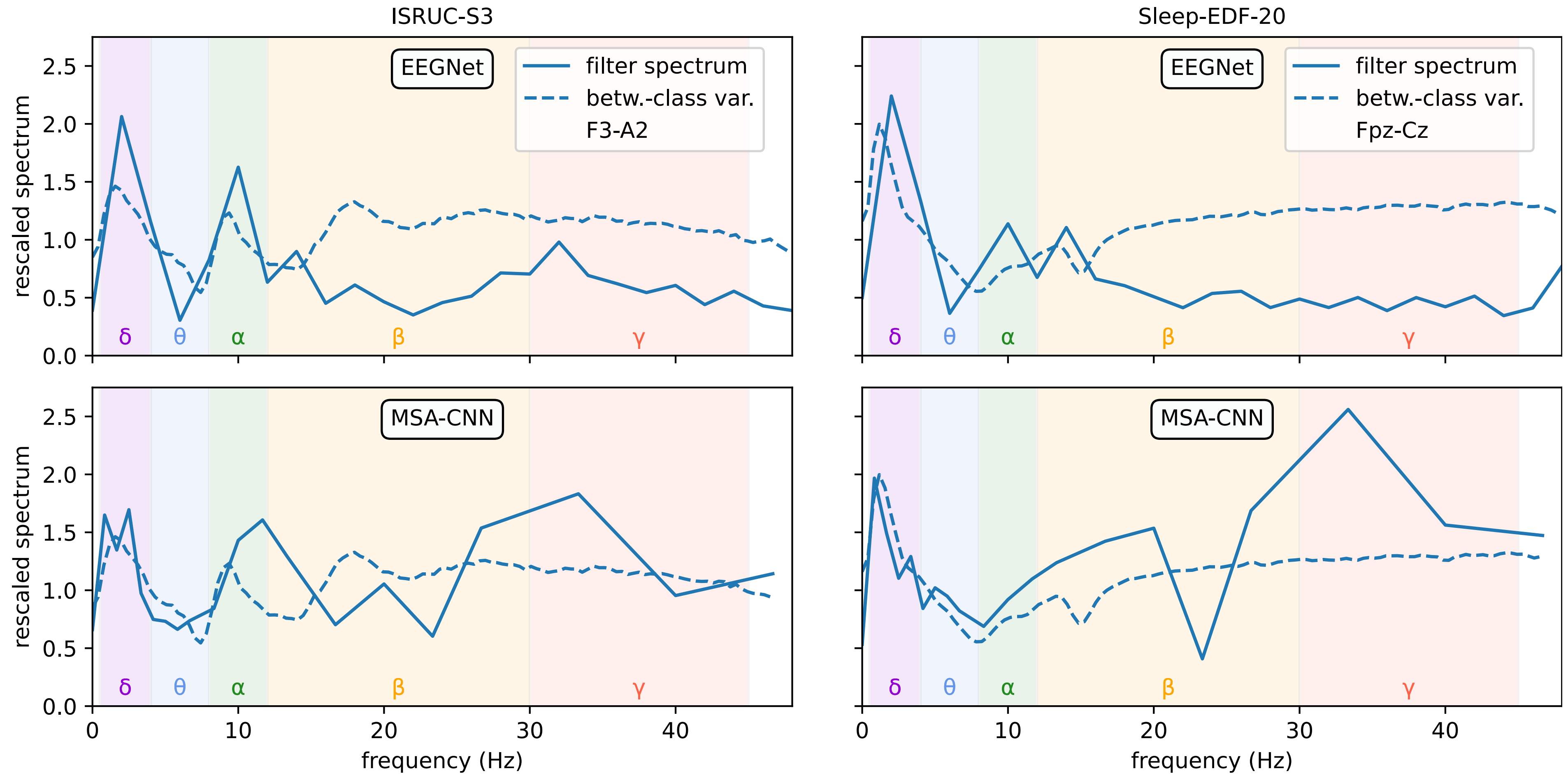
# Between-class spectral variation

- *Which frequencies are useful for classification?*
  - Use measure directly computed from data
- Measure for variation of spectral power *between* classes:
  1. average spectral density across all samples for each class
  2. compute standard deviation between the classes
  3. normalise by *within-class* spectral variation
- Within-class spectral variation:
  1. compute standard deviation for each class
  2. average across classes

# Polysomnography datasets

- ISRUC-S3 [3]:
  - 10 healthy participants, 8,589 annotated samples
  - Used channels: 6 EEG, 2 EOG, 1 EMG
- Sleep-EDF-20 [4]:
  - 20 healthy participants, 42,308 annotated samples
  - Used channels: 2 EEG, 1 EOG
- 5 sleep stages
- 30 second samples @ 100Hz
- cutoff frequency 40Hz

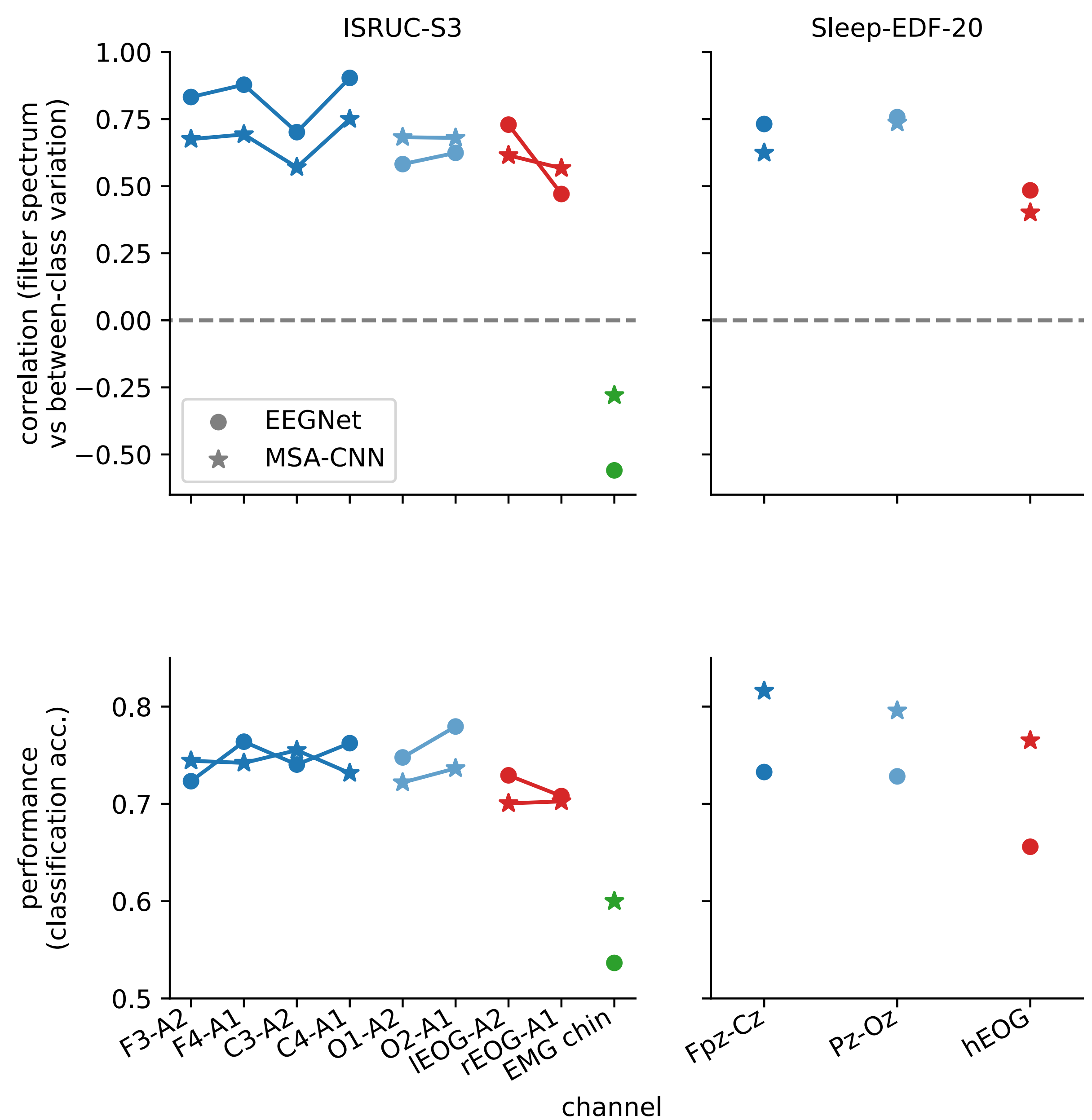
# Filter spectrum correlates with data characteristics




- Dashed: between-class spectral variation (frontal channel)
- Low frequencies (δ, θ, α):
  - trained EEGNet & MSA-CNN filters correlate with data-based variation

# Spectrum-data correlation aligns with performance

- Abstract previous results:
  - Low-frequency correlation now one data point
- Correlation depends on channel
- Similar pattern across datasets




- Single-channel performance:
  - Same pattern as correlation!
- Spectral information used for classification

# Conclusion

- Explainability tool:
  - Understand spectral processing in CNNs for EEG
- Convolution filter spectrum correlates with variation in data
- Higher single-channel performance for higher correlation
  - Helps gain insight into model performance
- Limitations:
  - Correlation estimates may be unreliable due to limited number of data points
  - Comparison between correlation estimates and performance for channel importance analysis was only assessed visually
  - Relatively small size of the datasets used in this study may limit the transferability of our model

# References

[1] Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. Journal of neural engineering, 15(5), 056013.

[2] Goerttler, S., Wang, Y., Eldele, E., Wu, M., & He, F. (2025). MSA-CNN: A Lightweight Multi-Scale CNN with Attention for Sleep Stage Classification. arXiv preprint arXiv:2501.02949.

[3] Khalighi, S., Sousa, T., Santos, J. M., & Nunes, U. (2016). ISRUC-Sleep: A comprehensive public dataset for sleep researchers. Computer methods and programs in biomedicine, 124, 180-192.

[4] Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. circulation, 101(23), e215-e220.