B[FM]²: Brain Foundation Model via Flow Matching with SplitUNet

Jaedong Hwang[1] Kathleen Zhang[1] Wei Dai[1] Konstantinos Kontras[1,2] Maarten De Vos[2] Ila Fiete[1] Paul Pu Liang[1]

[1] MIT [2] KU LEUVEN

Brain Foundation Model

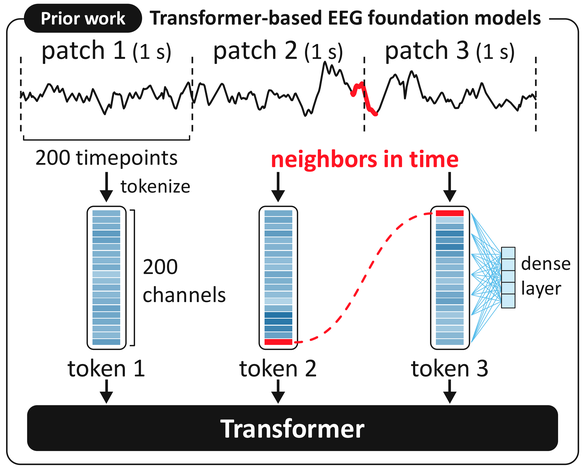


Brain signals are continuous, multi-channel time series.
B[FM]² is designed for both.

NEURAL INFORMATION PROCESSING SYSTEMS

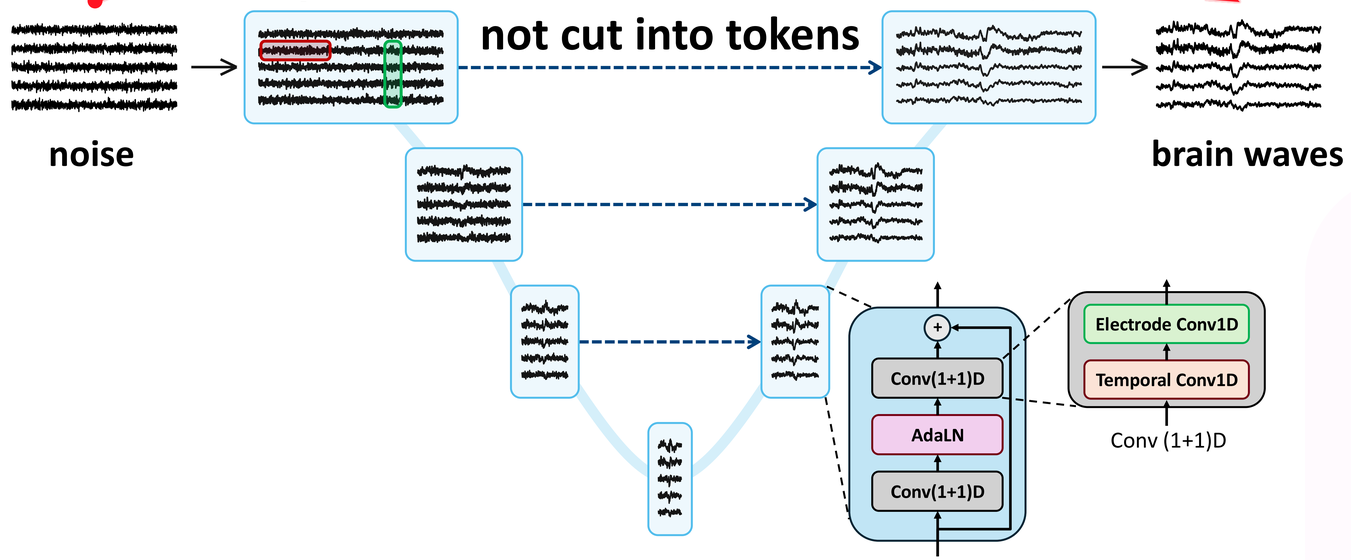
Tokenization breaks continuous EEG

Timepoints mix as channels, ignoring temporal correlation.
Neighbors in time (red) land in different tokens.



Flow Matching

pretrains on continuous brain waves, not cut into tokens



SplitUNet

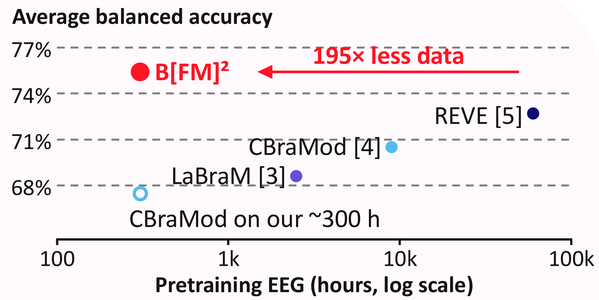
splits time from electrodes and keeps every electrode

195× less data

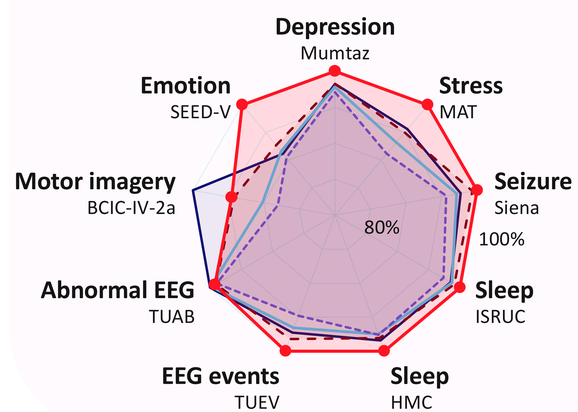
~300 hours of EEG and one GPU.
Previous best used 60,000 hours.

More accurate

75.4% average balanced accuracy on nine tasks. Previous best 72.7%.

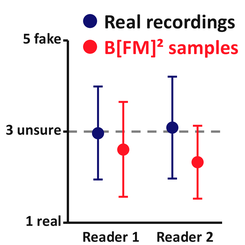


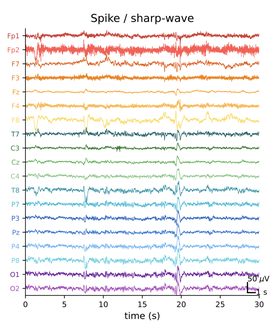




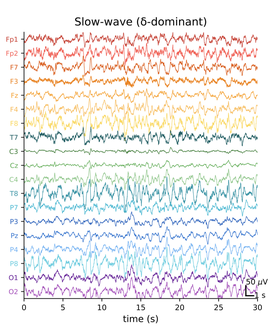
Each axis shows balanced accuracy as a percentage of the best model on that task, from 60% at the center to 100% at the outer ring.

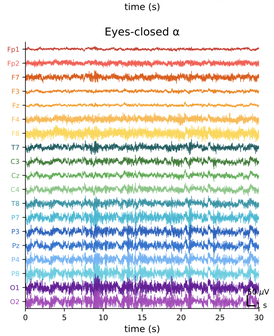
Neurologists could not differentiate real and generated EEGs



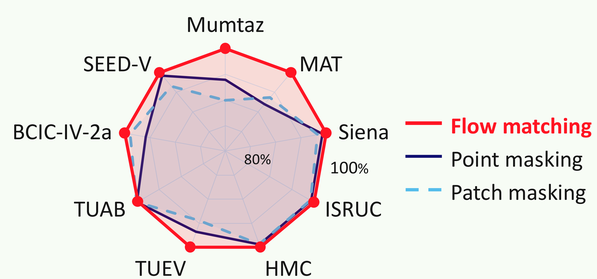


Mean ± std. per neurologist, 25 real and 25 gen. segments

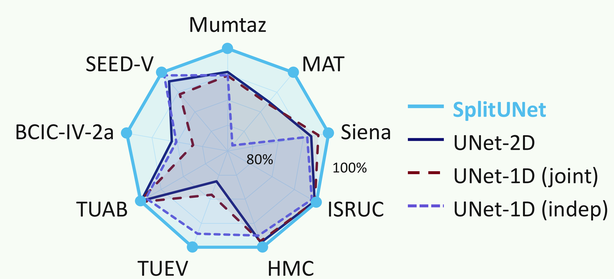




Flow matching outperforms masking



SplitUNet outperforms other UNets



❑ Reference [1] Lipman et al. Flow matching for generative modeling. ICLR 2023. [2] Tong et al. Minibatch optimal transport flow matching (OT-CFM). TMLR 2024. [3] Jiang et al. LaBraM. ICLR 2024. [4] Wang et al. CBraMod. ICLR 2025. [5] El Ouahidi et al. REVE. NeurIPS 2025. [6] Zhou et al. CSBrain. NeurIPS 2025. [7] Obeid & Picone. TUH EEG corpus. Front. Neurosci. 2016.