General-Purpose Brain Foundation Models for Time-Series Neuroimaging Data

Mohammad Javad Darvishi Bayazi
Mila, Québec AI Institute
Université de Montréal
mj.darvishi92@gmail.com

Hena Ghonia
Mila, Québec AI Institute

Roland Riachi
Mila, Québec AI Institute

Bruno Aristimunha
Inria TAU, LISN-CNRS
Université Paris-Saclay

Arian Khorasani
Mila, Québec AI Institute
Université de Montréal

Md Rifat Arefin
Mila, Québec AI Institute
Université de Montréal

Amin Darabi
Mila, Québec AI Institute
Université de Montréal

Guillaume Dumas
Mila, Québec AI Institute
Université de Montréal

Irina Rish
Mila, Québec AI Institute
Université de Montréal
irina.rish@umontreal.ca

Abstract

Brain function represents one of the most complex systems driving our world. Decoding its signals poses significant challenges, particularly due to the limited availability of data and the high cost of recordings. The existence of large hospital datasets and laboratory collections partially mitigates this issue. However, the lack of standardized recording protocols, varying numbers of channels, diverse setups, scenarios, and devices further complicate the task. This work addresses these challenges by introducing the **Brain Foundation Model (BFM)**, a suite of open-source models trained on brain signals. These models serve as foundational tools for various types of time-series neuroimaging tasks. This work presents the first model of the BFM series, which is trained on electroencephalogram (EEG) and functional Magnetic Resonance Imaging (fMRI) signal data. Our results demonstrate that BFM can generate signals more accurately than baseline models. Model weights and pipelines are available at https://bit.ly/3CCI0HW.

1 Introduction

Foundation models trained on large-scale datasets have revolutionized the field of artificial intelligence, demonstrating emergent capabilities across various tasks beyond their original objectives. Their adaptability and transferability make them valuable as a base for a wide range of applications (Bommasani *et al.* [2021]). However, time series analysis has lagged due to the scarcity of well-curated data and its inherent complexities, making it challenging to create comprehensive datasets (Rasul *et al.* [2024b]).

Time series analysis is vital across numerous fields, from finance to healthcare. In healthcare, one critical application is analyzing human physiological functions over time. This provides insights into our mechanisms and aids in diagnosing and treating dysfunctions and disorders. The brain is among the most complicated and essential systems governing human behavior and perception (Buzsaki [2019]). However, decoding brain activity is particularly challenging due to its extreme complexity, requiring vast amounts of data to capture its dynamics. Recording such brain signals is not only

NeurIPS 2024 Workshop on Time Series in the Age of Large Models.

costly but also resource-intensive (Rashid *et al.* [2020]; Roy *et al.* [2019]). A promising approach to mitigate these challenges is leveraging the transferability of foundation models to enhance brain activity analysis.

This work aims to develop a general-purpose brain activity foundation model capable of leveraging knowledge from large language models (LLMs) and general time series data. Due to its domain-agnostic nature, this model can transfer knowledge across various biosignals. We focus here on electroencephalography (EEG) and functional Magnetic Resonance Imaging (fMRI) as non-invasive methods of measuring brain activity. This framework is not limited to these two modalities and can be used for all other time series. We focused on these signals because of their wide applications ranging from medical diagnosis to brain-computer interfaces (BCI) (Hossain *et al.* [2023]; Safayari and Bolhasani [2021]; Popa *et al.* [2020]; Siuly *et al.* [2016]). We introduce BFM, a model that learns a robust representation of brain data capable of generating realistic brain signals. We expect its potential effectiveness in several applications.

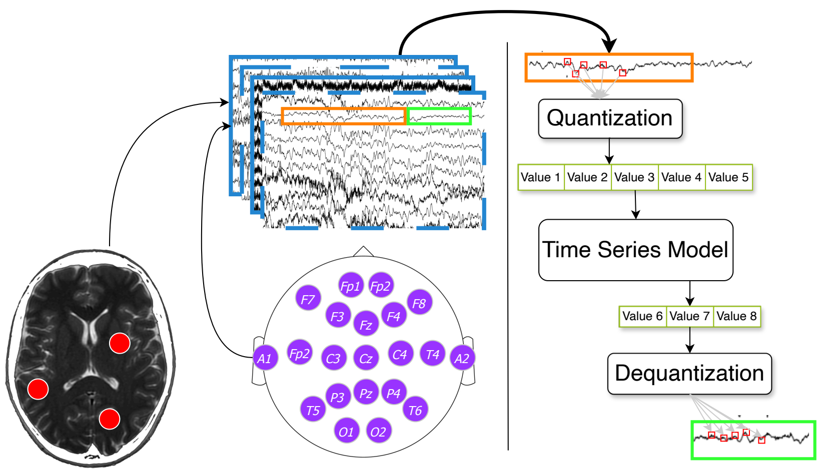

Figure 1: Overview of BFM. (Left) Schematic of EEG electrodes on the scalp and fMRI ROI capturing signals from various brain regions. (Right) Training of Time-series model.

2 Related Work

Time-series (TS) forecasting: Time series forecasting is essential in many domains, ranging from finance to healthcare (Zhang *et al.* [2017]; Jin *et al.* [2018]). Accurate forecasts are critical in informing decision-making processes and strategic planning [Wu *et al.*, 2022; Lai *et al.*, 2018]. Traditional approaches to time series forecasting include statistical methods such as Autoregressive Integrated Moving Average (ARIMA) and ETS models, which use autocorrelations and decompositions into explicit fundamental components, respectively, to predict future values. Despite the success of these models, they share common inherent limitations in their assumptions of linear relationships and stationary distributions - both of which are often not the case in real-world data (Liu *et al.* [2023]).

Modern research in time series forecasting has also seen the rise of deep learning-based methods (Benidis *et al.* [2022]) focusing on the use of multi-layer perceptrons (MLPs), recurrent neural networks (RNNs) and Transformers (Vaswani *et al.* [2017]; Nie *et al.* [2023]; Wu *et al.* [2020]; Salinas *et al.* [2020a]). These new developments seek to address the challenges faced by statistical models by utilizing non-linear functions and training on diverse, complex datasets. In particular, recent work has been done for foundational time series forecasting models analogous to large language models (LLMs). Time series foundation models leverage similar techniques such as self-supervised learning and scale to achieve state-of-the-art performance across a variety of domains and datasets (Woo *et al.* [2024]; Rasul *et al.* [2024a]; Ansari *et al.* [2024]).

Biosignals foundation models: In recent years, several foundation models have been developed to advance the analysis of diverse biosignals. Abbaspourazad *et al.* [2023] developed foundation models using extensive PPG and ECG data collected via Apple Watch. Ortega Caro *et al.* [2023] introduced the Brain Language Model (BrainLM), which serves as a foundation model for fMRI recordings. Azabou *et al.* [2024] presented POYO-1, a unified, scalable framework for neural population decoding focused on invasive neural activities. Zhang *et al.* [2022] applied self-supervised contrastive pre-training for time series through time-frequency consistency. Cui *et al.* [2023] proposed Neuro-GPT, integrating an EEG encoder with a GPT model. Chen *et al.* [2024] introduced EEGFormer, a pre-trained model leveraging large-scale compound EEG data. Jiang *et al.* [2024] introduced the Large Brain Model (LaBraM), which is trained on EEG data from BCI using vector-quantization for tokenization and masked patches for learning representations.

These models face several limitations. Primarily, their tokenization methods are *often specific to the type of signal or the number of channels*, which hinders their scalability and generalizability. This limitation has led to evaluations on tasks that are sometimes saturated [Kiessner *et al.*, 2024; Darvishi-Bayazi *et al.*, 2024]. In this work, we aim to develop a versatile model that can leverage various types of time series data to learn robust representations and facilitate transfer learning from LLMs and general time series models to biosignals.

3 Background and Method

Time-series Forecasting: Consider a dataset $\{\mathcal{X}_i\}_{i=1}^{S}$ where each $\mathcal{X}_i = [x_1, \ldots, x_T] \in \mathbb{R}^{T_i \times N}$ is a multivariate time series with T_i time steps and N channels. Given an input time series window $x_{t:t+C+P} = [x_t, \ldots, x_{t+C+P}]$ of length $C + P \geq 2$ for $t \in \{1, \ldots, T - C - P\}$, we look to forecast the $P \geq 1$ future values. In this work, we adopt a probabilistic and channel-independent approach. This means that we individually treat each channel as a univariate time series, and we do not explicitly model the dependencies between each channel. Moreover, given $x_{t:t+C}$ we output logits ϕ and prediction $\hat{y}_t \sim \mathbb{P}_\phi(\cdot|x_{t:t+C})$ such that $\hat{y}_t \approx y_t = x_{t+C+1:t+C+1+P}$.

Large Language Models (LLMs): have demonstrated remarkable performance by leveraging massive datasets and learning billions of parameters ([Dubey *et al.*, 2024; Brown, 2020]). These models are largely based on the Transformer architecture ([Vaswani, 2017]). One prominent example is **T5: Text-to-Text Transfer Transformer** [Raffel *et al.*, 2020; Chung *et al.*, 2024], an encoder-decoder, sequence-to-sequence model that exemplifies transfer learning in natural language processing tasks. As general pattern recognizers ([Mirchandani *et al.*, 2023]), LLMs can also effectively process time-series data [Zhou *et al.*, 2023; Jin *et al.*, 2023]. In this work, we use T5 as the backbone for our approach, though it can seamlessly be replaced with other LLMs.

Chronos: is a model that uses T5 (Raffel *et al.* [2020]) architecture as a backbone and were trained on publicly available diverse time series datasets. BFM uses Chronos tokenizer (Ansari *et al.* [2024]) and pre-trained Chronos-T5 based (Raffel *et al.* [2020]) models. The implementation of the Chronos tokenizer was motivated by the fact that in language tasks, tokens are derived from a finite dictionary. In contrast, with time series data, values are from an unbounded, typically continuous domain. The tokenizer uses mean scaling (Salinas *et al.* [2020b]) to normalize context window $x_{1:C}$ to $[(x_1 - m)/s, ..(x_C - m)/s]$ where $m = 0$ and $s = \frac{1}{C}\sum_{i=1}^{C}|x_i|$. After mean scaling, the tokenizer applies quantization to convert them into discrete tokens. The quantization function chooses B centers, $c_1 < ... < c_B$ uniformly and $B - 1$ edges b_i separating them, $c_i < b_i < c_{i+1}$, for $i \in \{1, ..., B - 1\}$. Dequantization function can be defined as $d(j) = c_j$, where $j \in \{1, ..., B - 1\}$. As mentioned in Figure 1, we consider each location series as independent time series, which is then passed to the tokenizer.

Brain Foundation models: BFM is univariate probabilistic forecasting model, based on Chronos-t5-large (700M) architecture. BFM is trained using categorical cross entropy objective function between ground truth label distribution and categorical distribution predicted by the model. We use Continuous Ranked probability score (CRPS) (Gneiting and Raftery [2007]; Matheson and Winkler [1976]) to evaluate model performance, which is commonly used to evaluate probabilistic forecasts. We report the CRPS averaged across all the time series of a dataset and over the prediction horizon using 20 empirical samples.

Datasets In this work, our objective is to learn robust representations of brain signals recorded from the scalp or different regions of interest (ROIs), as illustrated in Figure 1 (left). These multivariate signals reflect the underlying electrical and bold activity of the brain. This study uses the NMT EEG dataset, the MOABB benchmark, and resting state fMRI from the Adolescent Brain and Cognitive Development (ABCD) study.

NMT is a public, annotated dataset comprising healthy and pathological EEG recordings (Khan *et al.* [2022]). It consists of 2,417 recordings from unique participants, providing multichannel EEG data and labels indicating the participants' pathological state, classified as normal or abnormal. In addition, demographic information such as gender and age is included. We leverage the predefined training and testing splits based on subjects for model pretraining and signal generation. As shown in Figure 1 (Left), each EEG channel is treated as an independent time series, which is further divided into two segments: a context window for conditioning and a prediction target window.

MOABB is a comprehensive BCI library (Aristimunha *et al.* [2023]) that aggregates several EEG datasets. In this work, we selected large datasets—either in terms of the number of trials or the number of subjects—to avoid the bias often present in BCI studies that rely on small, single datasets (Jayaram and Barachant [2018]). Specifically, we used the BCI Competition IV 2a dataset [Tangermann *et al.*, 2012], Cho2015 (Cho *et al.* [2017]), Weibo2014 (Yi *et al.* [2014]), and Liu2024 (Liu *et al.* [2024]). These datasets vary in recording protocols, number of channels, trial lengths, and classification tasks, providing a diverse testing ground for our model.

ABCD rs-fMRI: We utilized resting-state fMRI data from the ABCD study. The voxel-wise fMRI data were reduced to the activity of 100 brain regions using dimensionality reduction based on the Schaefer-Yeo atlas (Schaefer *et al.* [2018]). The preprocessing steps included removing recording artifacts and subtracting the mean signal to enhance data quality.

4 Empirical Evaluation

BFM aims to predict future signal values based on previous time series samples. The following section compares our model's performance quantitatively against several baseline models. A qualitative analysis of the forecasting results can be found in the appendix in Figure 6.2.

Forecasting/Generation performance: We evaluated the performance of the BFM against several baseline models. The first baseline is a Naive Model, which forecasts the next value using the last observed value. The final baseline is Chronos-Original, which leverages the pre-trained Chronos model trained on a general time series. Table 4 shows that BFM improves the performance of other models in distribution on the unseen subjects of the EEG and fMRI datasets and out-of-distribution zero-shot performance on the Moabb datasets.

Table 1: Performance of BFM on various datasets (CRPS). Lower CRPS values indicate better performance (↓). CRPS for BFM is reported as mean $\pm$ std over three seeds.

Dataset	Naive	Chronos	BFM
NMT-EEG(unseen-subjects)	1.2531	0.8306	0.7675+-0.0039
ABCD-fMRI (unseen-subjects)	0.0093	0.00567	0.00543+-0.0009
BNCI2014_001	1.5478	0.9275	0.9005$\pm$0.0006
BNCI2014_004	1.4118	0.9293	0.8722$\pm$0.0016
BNCI2015_001	1.5038	0.9327	0.8219$\pm$0.0051
Weibo2014	1.1461	0.9882	0.8721$\pm$ 0.0199
Cho2017	1.4568	0.9176	0.8684$\pm$0.0048
Liu2024	1.3151	0.9169	0.8614$\pm$0.0155

Impact of Model Size and Transfer Learning from Language and Time-Series Models: To identify the optimal model to initialize the BFM, we examine the effects of model scaling and transfer learning from pre-trained language and time-series models on forecasting performance. Specifically, we evaluate models of varying sizes to assess how scaling impacts the validation loss. As shown in Figure 2 (Left), larger models consistently achieve lower validation loss, leading us to select the largest model for subsequent experiments. Furthermore, we explore the utility of transfer learning by initializing BFM with weights from pre-trained language models and time-series models. Figure 2 (Right) demonstrates a positive transfer from language model weights, with even lower loss observed when initializing from the Chronos time-series model weights. This highlights the potential of transfer learning from other modalities in enhancing neuroimaging tasks.

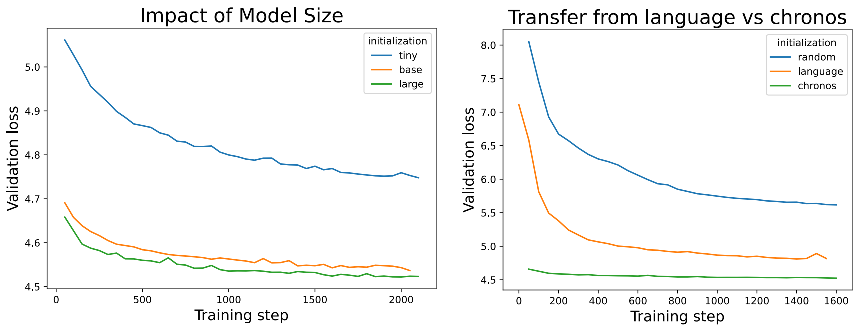

Figure 2: Scaling and Transfer Behavior. (Left) Larger models show smaller validation loss. (Right) Positive transfer from language and general time series models to EEG signals.

5 Discussion and Conclusion

We introduced and evaluated the Brain Foundation Model (BFM), a framework capable of learning the dynamics of brain signals by leveraging knowledge from both language models and general time series data. Despite the mixed findings regarding the effectiveness of transferring knowledge from LLMs to time series (Tan *et al.* [2024]; Zhou *et al.* [2023]), our results demonstrate a positive transfer from LLMs to EEG data, with an even stronger transfer observed from general time series models to the brain. BFM exhibited strong generalization across datasets beyond its training set. We believe that this model possesses the core characteristics of a foundation model and has the potential to improve multimodal analysis of simultaneous EEG-fMRI analysis (Lioi *et al.* [2020]; Ciccarelli *et al.* [2023]), brain-body signal analysis for human state assessments (Darvishi-Bayazi *et al.* [2023]) or decoding speech (Défossez *et al.* [2023]; Millet *et al.* [2022]). We believe this unified framework significantly advances BCI applications, diagnostic tools, and neuroscience research through the analysis of brain signals.

Acknowledgements

We acknowledge the support from the Canada CIFAR AI Chair Program and from the Canada Excellence Research Chairs Program. This research was made possible thanks to the computing resources on the Summit and Frontier supercomputers provided by the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. We sincerely appreciate Danilo Bzdok and Shambhavi Aggarwal valuable comments, which greatly contributed to the improvment of this paper.

References

Salar Abbaspourazad, Oussama Elachqar, Andrew C Miller, Saba Emrani, Udhyakumar Nallasamy, and Ian Shapiro. Large-scale training of foundation models for wearable biosignals. *arXiv preprint arXiv:2312.05409*, 2023.

Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, et al. Chronos: Learning the language of time series. *arXiv preprint arXiv:2403.07815*, 2024.

Bruno Aristimunha, Igor Carrara, Pierre Guetschel, Sara Sedlar, Pedro Rodrigues, Jan Sosulski, Divyesh Narayanan, Erik Bjareholt, Quentin Barthelemy, Robin Tibor Schirrmeister, Emmanuel Kalunga, Ludovic Darmet, Cattan Gregoire, Ali Abdul Hussain, Ramiro Gatti, Vladislav Goncharenko, Jordy Thielen, Thomas Moreau, Yannick Roy, Vinay Jayaram, Alexandre Barachant, and Sylvain Chevallier. Mother of all bci benchmarks, 2023. Version 1.1.0. DOI: `10.5281/zenodo.10034223`.

Mehdi Azabou, Vinam Arora, Venkataramana Ganesh, Ximeng Mao, Santosh Nachimuthu, Michael Mendelson, Blake Richards, Matthew Perich, Guillaume Lajoie, and Eva Dyer. A unified, scalable framework for neural population decoding. *Advances in Neural Information Processing Systems*, 36, 2024.

Konstantinos Benidis, Syama Sundar Rangapuram, Valentin Flunkert, Yuyang Wang, Danielle Maddix, Caner Turkmen, Jan Gasthaus, Michael Bohlke-Schneider, David Salinas, Lorenzo Stella, François-Xavier Aubet, Laurent Callot, and Tim Januschowski. Deep learning for time series forecasting: Tutorial and literature survey. *ACM Comput. Surv.*, 55(6), dec 2022.

Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

Tom B Brown. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020.

Gyorgy Buzsaki. *The brain from inside out*. Oxford University Press, USA, 2019.

Yuqi Chen, Kan Ren, Kaitao Song, Yansen Wang, Yifan Wang, Dongsheng Li, and Lili Qiu. Eegformer: Towards transferable and interpretable large-scale eeg foundation model. *arXiv preprint arXiv:2401.10278*, 2024.

Hohyun Cho, Minkyu Ahn, Sangtae Ahn, Moonyoung Kwon, and Sung Chan Jun. Eeg datasets for motor imagery brain–computer interface. *GigaScience*, 6(7):gix034, 2017.

Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, et al. Scaling instruction-finetuned language models. *Journal of Machine Learning Research*, 25(70):1–53, 2024.

Giuseppina Ciccarelli, Giovanni Federico, Giulia Mele, Angelica Di Cecca, Miriana Migliaccio, Ciro Rosario Ilardi, Vincenzo Alfano, Marco Salvatore, and Carlo Cavaliere. Simultaneous real-time eeg-fmri neurofeedback: A systematic review. *Frontiers in Human Neuroscience*, 17:1123014, 2023.

Wenhui Cui, Woojae Jeong, Philipp Thölke, Takfarinas Medani, Karim Jerbi, Anand A Joshi, and Richard M Leahy. Neuro-gpt: Developing a foundation model for eeg. *arXiv preprint arXiv:2311.03764*, 2023.

Mohammad-Javad Darvishi-Bayazi, Andrew Law, Sergio Mejia Romero, Sion Jennings, Irina Rish, and Jocelyn Faubert. Beyond performance: the role of task demand, effort, and individual differences in ab initio pilots. *Scientific Reports*, 13(1):14035, 2023.

Mohammad-Javad Darvishi-Bayazi, Mohammad Sajjad Ghaemi, Timothee Lesort, Md Rifat Arefin, Jocelyn Faubert, and Irina Rish. Amplifying pathological detection in eeg signaling pathways through cross-dataset transfer learning. *Computers in Biology and Medicine*, 169:107893, 2024.

Alexandre Défossez, Charlotte Caucheteux, Jérémy Rapin, Ori Kabeli, and Jean-Rémi King. Decoding speech perception from non-invasive brain recordings. *Nature Machine Intelligence*, 5(10):1097–1107, 2023.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.

Khondoker Murad Hossain, Md Ariful Islam, Shahera Hossain, Anton Nijholt, and Md Atiqur Rahman Ahad. Status of deep learning for eeg-based brain–computer interface applications. *Frontiers in computational neuroscience*, 16:1006763, 2023.

Vinay Jayaram and Alexandre Barachant. Moabb: trustworthy algorithm benchmarking for bcis. *Journal of neural engineering*, 15(6):066011, 2018.

Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous eeg data in bci. *arXiv preprint arXiv:2405.18765*, 2024.

Bo Jin, Haoyu Yang, Leilei Sun, Chuanren Liu, Yue Qu, and Jianing Tong. A treatment engine by predicting next-period prescriptions. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, page 1608–1616, New York, NY, USA, 2018. Association for Computing Machinery.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, et al. Time-llm: Time series forecasting by reprogramming large language models. *arXiv preprint arXiv:2310.01728*, 2023.

Hassan Aqeel Khan, Rahat Ul Ain, Awais Mehmood Kamboh, Hammad Tanveer Butt, Saima Shafait, Wasim Alamgir, Didier Stricker, and Faisal Shafait. The nmt scalp eeg dataset: an open-source annotated dataset of healthy and pathological eeg recordings for predictive modeling. *Frontiers in neuroscience*, 15:755817, 2022.

Ann-Kathrin Kiessner, Robin T Schirrmeister, Joschka Boedecker, and Tonio Ball. Reaching the ceiling? empirical scaling behaviour for deep eeg pathology classification. *Computers in Biology and Medicine*, page 108681, 2024.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, page 95–104, New York, NY, USA, 2018. Association for Computing Machinery.

Giulia Lioi, Claire Cury, Lorraine Perronnet, Marsel Mano, Elise Bannier, Anatole Lécuyer, and Christian Barillot. Simultaneous eeg-fmri during a neurofeedback task, a brain imaging dataset for multimodal data integration. *Scientific data*, 7(1):173, 2020.

Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting, 2023.

Haijie Liu, Penghu Wei, Haochong Wang, Xiaodong Lv, Wei Duan, Meijie Li, Yan Zhao, Qingmei Wang, Xinyuan Chen, Gaige Shi, et al. An eeg motor imagery dataset for brain computer interface in acute stroke patients. *Scientific Data*, 11(1):131, 2024.

James E Matheson and Robert L Winkler. Scoring rules for continuous probability distributions. *Management science*, 22(10):1087–1096, 1976.

Juliette Millet, Charlotte Caucheteux, Yves Boubenec, Alexandre Gramfort, Ewan Dunbar, Christophe Pallier, Jean-Remi King, et al. Toward a realistic model of speech processing in the brain with self-supervised learning. *Advances in Neural Information Processing Systems*, 35:33428–33443, 2022.

Suvir Mirchandani, Fei Xia, Pete Florence, Brian Ichter, Danny Driess, Montserrat Gonzalez Arenas, Kanishka Rao, Dorsa Sadigh, and Andy Zeng. Large language models as general pattern machines. *arXiv preprint arXiv:2307.04721*, 2023.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers, 2023.

Josue Ortega Caro, Antonio Henrique Oliveira Fonseca, Christopher Averill, Syed A Rizvi, Matteo Rosati, James L Cross, Prateek Mittal, Emanuele Zappala, Daniel Levine, Rahul M Dhodapkar, et al. Brainlm: A foundation model for brain activity recordings. *bioRxiv*, pages 2023–09, 2023.

Livia Livint Popa, Hanna Dragos, Cristina Pantelemon, Olivia Verisezan Rosu, and Stefan Strilciuc. The role of quantitative eeg in the diagnosis of neuropsychiatric disorders. *Journal of medicine and life*, 13(1):8, 2020.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67, 2020.

Mamunur Rashid, Norizam Sulaiman, Anwar P. P. Abdul Majeed, Rabiu Muazu Musa, Ahmad Fakhri Ab. Nasir, Bifta Sama Bari, and Sabira Khatun. Current status, challenges, and possible solutions of eeg-based brain-computer interface: A comprehensive review. *Frontiers in Neurorobotics*, 14, June 2020.

Kashif Rasul, Arjun Ashok, Andrew Robert Williams, Hena Ghonia, Rishika Bhagwatkar, Arian Khorasani, Mohammad Javad Darvishi Bayazi, George Adamopoulos, Roland Riachi, Nadhir Hassen, Marin Biloš, Sahil Garg, Anderson Schneider, Nicolas Chapados, Alexandre Drouin, Valentina Zantedeschi, Yuriy Nevmyvaka, and Irina Rish. Lag-llama: Towards foundation models for probabilistic time series forecasting, 2024.

Kashif Rasul, Arjun Ashok, Andrew Robert Williams, Hena Ghonia, Rishika Bhagwatkar, Arian Khorasani, Mohammad Javad Darvishi Bayazi, George Adamopoulos, Roland Riachi, Nadhir Hassen, et al. Lag-llama: Towards foundation models for probabilistic time series forecasting. *Preprint*, 2024.

Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering*, 16(5):051001, 2019.

Atefeh Safayari and Hamidreza Bolhasani. Depression diagnosis by deep learning using eeg signals: A systematic review. *Medicine in Novel Technology and Devices*, 12:100102, 2021.

David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020.

David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International journal of forecasting*, 36(3):1181–1191, 2020.

Alexander Schaefer, Ru Kong, Evan M Gordon, Timothy O Laumann, Xi-Nian Zuo, Avram J Holmes, Simon B Eickhoff, and BT Thomas Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral cortex*, 28(9):3095–3114, 2018.

Siuly Siuly, Yan Li, and Yanchun Zhang. Eeg signal analysis and classification. *IEEE Trans Neural Syst Rehabilit Eng*, 11:141–144, 2016.

Mingtian Tan, Mike A Merrill, Vinayak Gupta, Tim Althoff, and Thomas Hartvigsen. Are language models actually useful for time series forecasting? *arXiv preprint arXiv:2406.16964*, 2024.

Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J Miller, Gernot R Müller-Putz, et al. Review of the bci competition iv. *Frontiers in neuroscience*, 6:55, 2012.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc., 2017.

A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers, 2024.

Neo Wu, Bradley Green, Xue Ben, and Shawn O'Banion. Deep transformer models for time series forecasting: The influenza prevalence case, 2020.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting, 2022.

Weibo Yi, Shuang Qiu, Kun Wang, Hongzhi Qi, Lixin Zhang, Peng Zhou, Feng He, and Dong Ming. Evaluation of eeg oscillatory patterns and cognitive process during simple and compound limb motor imagery. *PloS one*, 9(12):e114853, 2014.

Liheng Zhang, Charu Aggarwal, and Guo-Jun Qi. Stock price prediction via discovering multi-frequency trading patterns. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, page 2141–2149, New York, NY, USA, 2017. Association for Computing Machinery.

Xiang Zhang, Ziyuan Zhao, Theodoros Tsiligkaridis, and Marinka Zitnik. Self-supervised contrastive pre-training for time series via time-frequency consistency. *Advances in Neural Information Processing Systems*, 35:3988–4003, 2022.

Tian Zhou, Peisong Niu, Liang Sun, Rong Jin, et al. One fits all: Power general time series analysis by pretrained lm. *Advances in neural information processing systems*, 36:43322–43355, 2023.

6 Appendix

6.1 Training Details

We trained model of 3 sizes initialized from chronos weights, tiny(8M), base(200M), and large(700M) for 2K steps with effective batch size 2048, to study the effect of model size on validation loss performance. We use context length 512 for BFM and prediction length 64, with linear scheduler for learning rate starting from 0.001. BFM large was trained for 6K steps with an effective batch size of 2048 on eight nodes(4 AMD MI250X or 8 separate GPUs, each having 64 GB of high-bandwidth memory) using DDP (Data Distributed parallelization).

6.2 Forecasting visualization: qualitative analysis

Figure 3 presents several examples of the predicted signals, including the observed signals, the median of the predicted signals, and the 80% prediction interval. The results qualitatively demonstrate that our models effectively learn the underlying patterns in EEG signals, generating meaningful and realistic samples.

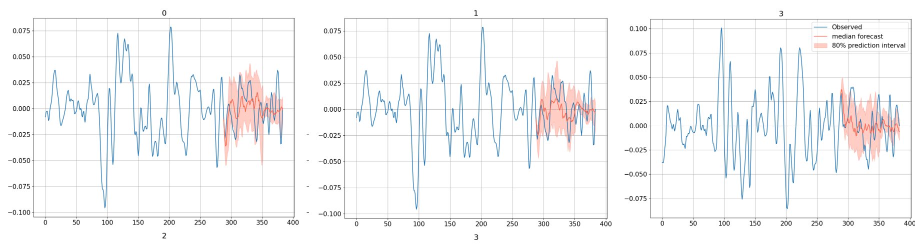

Figure 3: Forecasting Results. Three examples of EEG signals generated by the proposed time-series model. The predicted signals are compared to the original EEG recordings to evaluate the accuracy of the model's predictions.