Distribution Agreement

In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this dissertation. I retain all ownership rights to the copyright of the dissertation. I also retain the right to use in future works (such as articles or books) all or part of this dissertation.

Mahmoud Zeydabadinezhad Date

Biomarker Discovery from Sparse-Labeled Electrophysiological Datasets

By

Mahmoud Zeydabadinezhad
Doctor of Philosophy

Computer Science and Informatics

Babak Mahmoudi, PhD
Advisor

Ali Bahrami Rad, PhD
Committee Member

Kamran Paynabar, PhD
Committee Member

Matthew Reyna, PhD
Committee Member

Reza Sameni, PhD
Committee Member

Accepted:

Kimberly Jacob Arriola
Dean of the James T. Laney School of Graduate Studies

Date

Biomarker Discovery from Sparse-Labeled Electrophysiological Datasets

By

Mahmoud Zeydabadinezhad
M.S., Wayne State University , MI

Advisor: Babak Mahmoudi, PhD

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Computer Science and Informatics
2024

Abstract

Biomarker Discovery from Sparse-Labeled Electrophysiological Datasets

By Mahmoud Zeydabadinezhad

The widespread accessibility of diverse neuroimaging modalities—ranging from Local Field Potential (LFP) and Electroencephalography (EEG) signals to functional MRI (fMRI)— and the advances in Artificial Intelligence (AI) and machine learning (ML) methodologies have opened unprecedented avenues for investigating the neural patterns underlying sensorimotor and cognitive processes. These patterns of neural activities that are called physiomarkers or biomarkers are crucial for understanding the neural mechanisms of diseases, developing novel therapeutic interventions such as closed-loop neuromodulation, and studying the mechanisms of action of those treatments. Despite these technological advances, the applicability of contemporary AI/ML approaches is limited because there are often not enough examples in the labeled datasets. This sample size limitation arises from various factors such as ethical considerations in data collection, financial constraints, and the limited size of patient populations. Concurrently, there is an urgent need for models that not only perform well but are also explainable, particularly for the identification of neural biomarkers. While machine learning methodologies for biomarker identification in neural activity have been extensively studied, the focus has predominantly been on large labeled datasets. The issue of explainability is often relegated to a secondary concern. Research in other settings, such as computer vision, has ventured into methods tailored for small sample sizes, but these approaches seldom offer a balance between performance and explainability. Moreover, the applicability of these methods to neural activity data is uncharted territory. Addressing this gap is of paramount importance for several compelling reasons: The challenge is endemic in neuroscience, affecting a multitude of studies that operate under the constraints of limited sample sizes. The current limitations hinder the application of advanced automated data representation learning methods to neuroscience and have far-reaching implications for clinical applications. The development of an explainable automated data representation framework, tailored for limited sample sizes, stands to make a seminal contribution to neuroscience. Such a framework would not only facilitate biomarker identification but also enrich our understanding of neural activity. Building on this premise, our research specifically targets EEG data, intending to develop an automated and explainable data representation method for the critical task of quantifying the physiological effects of electrical neuromodulation. We hypothesize that the development of an explainable foundation model, tailored for EEG data analysis, will significantly enhance the quantification of physiological effects from small and heterogeneous EEG datasets. This model will surpass the limitations of current machine learning and deep learning methodologies by providing a robust, generalizable solution that is capable of interpreting complex biological signals without the need for extensive labeled data. To test our hypothesis, we develop analysis pipelines for biomarker identification from small and heterogenous data using manual feature extraction and traditional machine learning models

within the context of electrogastrography (EGG), and electroencephalography (EEG). Subsequently, we adapt a foundation model to automatically generate EEG data representations, tailored for a memory classification task.

Biomarker Discovery from Sparse-Labeled Electrophysiological Datasets

By

Mahmoud Zeydabadinezhad
M.S., Wayne State University , MI

Advisor: Babak Mahmoudi, PhD

A dissertation submitted to the Faculty of the
Emory College of Arts and Sciences of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Computer Science and Informatics
2024

Acknowledgments

To all who have joined me on this adventure, I extend my deepest gratitude. First and foremost, I would like to thank my advisor, Babak Mahmoudi, for his support throughout this journey and for instilling in me the intricacies of research itself. His passion, mentorship, and confidence in me have contributed immeasurably to my growth into an interdisciplinary scientist. I am also profoundly grateful to my committee members—Matthew Reyna, Reza Sameni, Ali Bahrami Rad, and Kamran Paynabar—whose invaluable wisdom has significantly contributed to the development of this dissertation.

Above all, my deepest gratitude goes to my family. This dissertation is dedicated to my parents, Asghar and Fatemeh, and to Emily, my partner in crime and best friend, whose love and support carried me through every high and low of this PhD journey. Her belief in me never faltered, and I could not have done this without her by my side.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Biomarkers are essential components of the rational development of diagnostics and therapeutics in medicine. In recent years, ML/AI has found applications in the identification of biomarkers from neural activity data [1–8]. Nonetheless, its clinical adoption remains limited. Two predominant impediments—lack of explainability and large labeled datasets—hamper the clinical utility of many contemporary ML-based methodologies [9, 10]. While technological innovations have facilitated the collection of expansive datasets, particularly in the realm of electroencephalography (EEG), the labor-intensive and costly nature of manual data labeling continues to pose challenges. From a therapeutic standpoint, the data that can be collected to identify effective biomarkers and inform closed-loop interventions are limited in cases such as neural stimulation therapies. The objective of this dissertation is to critically examine the limitations inherent in manual feature extraction and conventional machine learning methodologies for biomarker identification within neuromodulation applications, particularly in contexts where the available labeled data are limited. Subsequently, this dissertation aims to adopt a framework based on self-supervised learning strategies designed to improve automatic EEG data representation.

In the following subsections, we examine prior research focused on addressing the issue of representation learning with small sample sizes. Although machine learning models, especially deep learning algorithms, are generally considered to need large datasets for effective training, various methods have been proposed to handle the constraints of limited sample sizes. The broad field of Artificial Intelligence and machine learning research includes a wide range of methods, techniques, and theories, each designed to address specific problems [11]. The purpose of defining these categories is to provide a general understanding of the strategies that facilitate AI model training under the constraint of limited datasets. It is important to note that these categories are not isolated in practical applications; they are interconnected, and they are neither mutually exclusive nor collectively exhaustive.

1.1 Regularization/Model Complexity

A significant issue when dealing with small data size or small n, large P problems is model overfitting. Regularization methods are used to address overfitting by imposing constraints on the model's weights and biases during the training phase.

1.1.1 L1 Regularization (LASSO)

Originally introduced by Tibshirani [12], the least absolute shrinkage and selection operator (LASSO) minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. This method can be used for feature selection as it tends to drive the weights of unimportant features to zero.

$$L = \sum_{i=1}^{N}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

1.1.2 L2 Regularization (Ridge Regression)

Ridge regression adds the squared magnitude of the coefficient as the penalty term to the loss function. This method prevents the weights from growing too large, which can lead to overfitting.

$$L = \sum_{i=1}^{N}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

The key difference between Ridge and Lasso is that Lasso shrinks the less important feature's coefficient to zero thus, removing some features altogether.

1.1.3 Dropout regularization

This method was proposed by Hinton et al. [13] to address the overfitting in neural networks and works by randomly turning off learning neurons. Dropout controls neural network overfitting by randomly omitting subsets of features at each iteration of a training procedure [14]. It has been shown that feature dropout can be used as an adaptive regularization method that can be applied to any learning algorithm [14].

1.2 Lower Complexity Models

A recent study showed that model complexity is critical when only a few samples per class are available and that using low-complexity models can improve the state-of-the-art in deep learning with small data [15]. The rationale is that networks with fewer trainable parameters are less prone to overfitting and generalize better with small data; however, the results are still very far from the successes of high-capacity models with big data [15]. A limitation of this study is that the authors only focused on the image classification tasks on computer vision datasets like MNIST. Hence, it is unclear how this strategy would work in the case of biomarker identification.

Our preliminary studies showed that with careful feature engineering, we could train simple and interpretable models like Logistic Regression or Random Forest for seizure detection from EEG signals or stomach's state classification from electrogastrography (EGG) signals. Nonetheless, it should be noted that the seizure detection task was comparatively straightforward, and the identical pipeline may not exhibit equivalent efficacy for more intricate tasks, such as the quantification of neuromodulation in studies investigating memory enhancement.

1.3 Data Augmentation

Data augmentation is a technique to increase the size and diversity of a training dataset by generating new samples. Various methods exist for generating new samples, and selecting an appropriate one depends on the specific application and the inherent characteristics of the data.

1.3.1 Data Manipulation

Data Manipulation constitutes transforming the input data to create new samples. Techniques like random cropping and resizing [16], rotation and flipping [17], color jittering [18], cutout [19], adding different types of noise [20], elastic transformation [21], and linearly combining pairs of images and their labels during training (mixup) [22] are examples of transformation for creating new data samples. Some of these techniques have a limited scope of application and are only helpful in specific fields, such as computer vision. Data augmentation methods such as flipping, rotating, and scaling are insufficient to increase diversity and variance in medical data and specifically EEG [23, 24].

1.3.2 Generative Models

The utilization of generative models for data augmentation in medical datasets is on the rise [25–27]. Generative models could refer to various ML algorithms, such as generative adversarial networks (GANs) [28], variational autoencoders (VAEs) [29], or Generative Pre-trained Transformers (GPT) [30, 31] which are used to generate synthetic samples that can be added to an existing dataset to increase its size and diversity. This approach aims to improve the performance of ML models trained on such augmented datasets, which may lead to better accuracy and generalization. However, some challenges are associated with this approach. One such challenge is the lack of explainability of generative models. This means that it can be difficult to understand how the model generates new data samples, making it harder to validate the results and ensure that they are accurate. Another challenge is the computational complexity of generative models. These models are often extensive and require a lot of computational power to train and run, making them impractical for some applications[32]. Moreover, generative models often require a substantial amount of data to capture the underlying distribution effectively. This is particularly true for complex data types like images or medical data.

1.3.3 Biophysical Modeling

Biophysical modeling involves simulating biological systems using mathematical formalizations of their physical properties [33]. It can be used to fill data gaps where there is knowledge of how the underlying system works. Biophysical modeling has been applied to various fields, such as neuro-oncology, to characterize better the molecular, spatial, and temporal heterogeneity of tumors [34]. It also has been used to study the origins of Local Field Potential (LFP) and EEG [35, 36]. There are anticipated challenges with using biophysical modeling. The current knowledge of how a biological system works may not be enough to generate a meaningful representation of that system. The other challenge arises because most of our biophysical modeling knowledge stemmed from animal studies, and the inter-species differences may hinder accurate data generation using this approach [35].

1.3.4 Federated Learning

Federated learning (FL) is not a data augmentation method per se, but how it is designed may help to mitigate data scarcity problems in some cases [37]. In this design, decentralized data sources are used to train a central model without sharing the raw data itself [37]. However, a significant challenge with FL is data heterogeneity or non-identically distributed data. Researchers have proposed novel FL designs that use zero-shot data augmentation on under-represented data to mitigate statistical heterogeneity and improve accuracy performance [38].

Because in this dissertation our focus pivots towards neuromodulation, we looked into prior research on data augmentation for EEG [24, 39, 40]. A recent study conducted a review of 27 distinct studies that utilized generative algorithms for EEG data augmentation [41]. The findings showed an enhancement in classification accuracy, ranging from 1% to 40%, upon employing synthetic EEG data. This variability implies that the effectiveness of these DA techniques is not straightforward and depends on various factors, such as the specific EEG task and dataset being used which also cast a shadow on their generalizability. The generation of realistic synthetic EEG data may open up the possibility of data augmentation [42], however, the transition towards its clinical application, notably in augmenting EEG data for two distinct cohorts—patients and healthy control subjects—for the training of diagnostic classifiers, remains tentative [41]. A significant impediment to the widespread incorporation of EEG data augmentation lies in the lack of robust evaluation metrics and a well-defined optimal threshold for the number of synthetic data that needs to be generated [41].

1.4 Transfer Learning

Transfer learning is a machine learning technique in which a model that has been trained on one task is reused as a starting point for a different but related task[43]. Transfer learning can be used to improve the performance of a model on a new task by leveraging the knowledge that the model has already learned from the original task [44]. It is beneficial when the new task has a limited amount of labeled data, as the pre-trained model can provide a good starting point that can be fine-tuned with the new data. There are several different types of transfer learning. One common approach is to use a pre-trained model as a feature extractor, where the pre-trained model is used to extract features from the input data, and then a new classifier is trained on top of these features [45]. Another approach is to use a pre-trained model as an initialization for the new task, where the pre-trained model is used to initialize the weights of the new model, and then the new model is trained on the new task [45, 46]. The authors of [47] used transfer learning to develop object detection capabilities for novel biomarkers from microscopy imaging using previously trained models with fewer annotated images. They reduced the number of required annotated images and achieved comparable or improved detection accuracy, allowing for more efficient and accurate analysis of biomarkers and smaller datasets. Transfer learning has been used to mitigate the problem of individual differences in EEG data, which affects the reusability and generalization of models [48]. A recent study has shown the successful application of transfer learning in training a model to learn the EEG characteristics of an individual to achieve personalized and accurate seizure prediction [49]. However, an inherent risk in employing transfer learning methodologies for EEG signal analysis is the incidence of negative transfer [49]. This manifests when the incorporation of source domain knowledge detrimentally impacts the predictive performance within the target domain, a consequence often stemming from discrepancies between domains or suboptimal transfer techniques. Consequently, a careful evaluation of task transferability and domain similarity is imperative before the construction of transfer models.

1.5 Self Supervised Learning

Self-supervised learning (SSL), also known as predictive or pretext learning, is a category of ML algorithms intended to create data-efficient AI systems [50]. The idea behind SSL is learning to represent the data without relying on explicit human-labeled annotations before learning a task[51]. SSL has shown great potential in some medical applications like EEG classification [52–55]. It outperforms a purely supervised approach in low data regimes while capturing biomarkers without any access to labels [56]. SSL approaches can be classified into three main categories based on the type of task they use for representation learning: predictive, contrastive, and reconstruction-based methods [57].

1.5.1 Predictive Methods

These models learn to predict future samples or missing parts of the input data. Temporal sequence prediction, next-frame prediction in videos, and inpainting [58] are examples of predictive models in self-supervised learning [59].

1.5.2 Contrastive Methods

This approach focuses on learning representations by comparing similar and dissimilar samples from the input data. The model learns to distinguish between positive (similar) and negative (dissimilar) sample pairs. In recent developments, a specific category of contrastive SSL known as instance discrimination has emerged. This category, which includes methods like DINO [60], BYOL [61], and SimSiam[62], has further advanced the field by eliminating the requirement for negative samples.

1.5.3 Reconstructive Methods

Reconstructive methods in self-supervised learning are designed to learn meaningful data representations by reconstructing inputs, often by first corrupting or removing certain parts and then attempting to recreate the original version. These methods include various approaches such as autoencoders, denoising autoencoders, and masked language models, all of which have shown effectiveness in different domains.

Autoencoders form the foundational concept of reconstructive methods by encoding data into a compact latent representation and subsequently reconstructing it. This compression process ensures that the encoder extracts only the most relevant features, facilitating a useful representation for further tasks [63]. Variational Autoencoders (VAEs), an extension of standard autoencoders, learn a probabilistic distribution over the latent space, adding a generative capability that allows the model to generate new data samples [64].

Denoising autoencoders extend this concept by adding random noise to the input and training the model to reconstruct the noise-free version. This method forces the network to focus on the underlying structure of the input rather than surface-level features, thus increasing robustness to noise and improving feature extraction [65].

Another significant approach within reconstructive methods is masked language modeling (MLM), utilized by BERT [66], where words in a text sequence are randomly masked, and the model is tasked with predicting these masked words. This technique has proven to be a breakthrough in natural language processing, as it enables the model to learn bidirectional contextual relationships, ultimately improving performance across numerous NLP tasks. By predicting the missing content, the model effectively learns to represent linguistic patterns and dependencies, which are highly valuable for downstream applications, such as question answering and text classification.

There are some challenges associated with the application of SSL in biomarker identification. When designing a pretext task for a biomarker identification problem, it is essential to consider the unique properties of medical data to ensure optimal performance. For example, computer vision and medical image analysis fields both work with image data. However, medical images differ significantly from natural images regarding the number of channels, intensity, location, scale, and orientation. Therefore, using a pretext task designed based on natural images may not be optimal for medical images. Thus, it is crucial to carefully tailor the pretext task to the specific characteristics of the medical data to achieve the best possible results. One potential limitation of SSL in biomarker identification is its transferability to other tasks. While SSL may perform effectively in identifying a specific biomarker and disease, its performance may not generalize to other tasks without modification [57].

Chapter 2

Discovery of Electrogastrography Biomarkers Under Label Constraints

[1]

2.1 Background

Electrical vagus nerve stimulation (VNS) is emerging as a potential therapy for gastric motility disorders [68]. However, the VNS mechanisms of action on gastric motility regulation have yet to be fully understood. This understanding is crucial for developing more effective therapies. Gastric electric activity (GEA) is known to be a physiological signal that regulates gastric motility and can be recorded using electrogastrography (EGG) [69, 70]. EGG can be used as a feedback signal for closed-loop adaptive VNS interventions, however its adoption in clinical practice remains limited. The primary reasons for this underutilization include the intrinsic properties of GEA and the absence of standardized protocols for electrode placement, both of which constrain the clinical applicability when recorded non-invasively [71]. However, it is worth

---

[1] Parts of this chapter have been published in [67] under the terms of the Creative Commons Attribution License.

noting that recent studies have demonstrated the successful fabrication of high spatial resolution EGG and Magnetogastrogram systems [72]. These advancements hold the potential to establish a clinical standard for EGG recording, ultimately facilitating its widespread adoption in clinical settings. Pre-clinical and human studies have demonstrated that implanted electrodes can record GI myoelectric activities that contain significantly more information than skin surface electrodes [73]. Since the abdominal wall may have a low-pass filtering effect, the higher frequency information is attenuated in non-invasive EGG recordings [74]. These attenuated signal components are known to be associated with gastric contractions [75]. While invasive recordings provide physiological signals that contain more information, compared to non-invasive recordings, they pose several challenges. In studies with electrodes implanted in the GI system, the subjects are required to remain sedentary or under anesthesia during data recording [75]. Most EGG-based studies have been done in a controlled environment with the subject either instructed not to move or being anesthetized [76, 77]. Although collecting data in a controlled and sedentary manner is useful for some studies, it cannot synthesize realistic situations, such as recording data in ambulatory and non-clinical settings. One potential application of ambulatory EGG recordings is studying motion sickness [78] and the effectiveness of administered therapies to prevent emesis [79].

Our data collection approach was designed to synthesize realistic ambulatory settings. We opted for a rarely practiced approach that involved implanting the VNS cuffs and EGG electrodes around the vagus nerve and on the serosal layer of the ferret stomachs, respectively. To study the physiological effects of VNS on gastric function, our novel dataset was collected by recording EGG from the serosal layer of the ferret stomach in two different conditions, i.e., before applying the VNS (baseline) and during application of the VNS. The recording was done in vivo while the animals were awake and freely moving in their cage. To our knowledge, this is the first time that EGG signals were recorded in this manner. The artifacts caused by sudden movements of the ferrets and the relatively small sizes of the recorded data posed unique challenges, in this experiment design and data collection approach, as compared to that of recordings in a controlled and static environment. To prepare our dataset for analysis, we devised specific pre-processing steps. Our EGG recording conditions required pre-processing steps that may not be necessary when EGG is recorded sedentary. By drawing upon the knowledge of other biosignal processing fields, we have expanded the set of features that are conventionally employed for the analysis of EGGs. A voting-based algorithm was developed for feature selection and dimensionality reduction, aiming to mitigate the adverse consequences of limited sample size and an extensive set of features on feature relevance and predictive accuracy. This approach seeks to optimize the model's performance while addressing the challenges inherent in small-sample, high-dimensional datasets. Through cross-validation and statistical methods, we demonstrated that Random Forest is a powerful yet interpretable model for differentiating baseline state from VNS states. In this study, we formulated a machine learning approach to quantify the effects of VNS on GEA using EGG signals that were recorded invasively from the surface of the stomach. While previous research [80] has demonstrated the impact of VNS on alterations in gastrointestinal activity, these investigations have not explored EGG features beyond dominant frequency [69] (DF) and its derivatives. In this chapter, we have incorporated an expanded array of diverse features, drawing from those commonly employed in related fields such as electroencephalography (EEG) and electromyography (EMG), to represent a broad range of physiological properties beyond the standard DF. Using this approach, we aim to address two main questions: 1- Can we identify the electrophysiological effects of electrical VNS on EGG signals recorded in an invasive and non-sedentary manner? 2- Do the electrophysiological effects depend on the electrical VNS parameters?

2.2 Methods

2.2.1 Data collection

Under isoflurane anesthesia (1 to 3%), seven adult male ferrets were chronically implanted with vagus nerve cuffs and GI serosal surface electrodes. Surface electrodes were placed at four locations along the stomach axis (named gastric1 to gastric4) and two locations at the duodenum (See Figure 4.1). Surgical implantation procedures were similarly designed to past studies [73, 81]. Leads were subcutaneously connected to a head connector (See Figure 2.1) and there were at least 10 days of recovery from surgery before the first data acquisition. All surgical and testing procedures were approved by the University of Pittsburgh Institutional Animal Care and Use Committee and were conducted following approved guidelines. All animal studies reported also followed the recommendations in the ARRIVE guidelines.

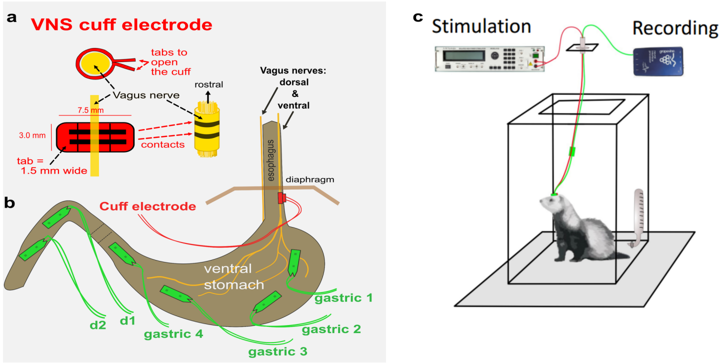


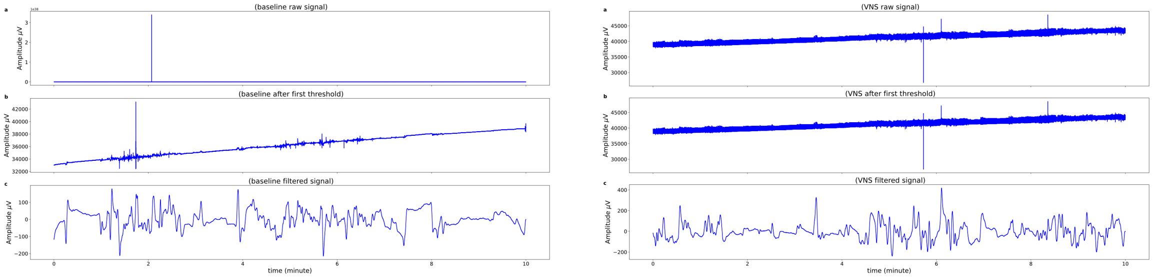
Figure 2.1: Electrode placements. a) show the structure of the nerve cuff, b) the position of surface electrodes (only ‘gastric 2’ was used in this study) and nerve cuff. c) A ferret’s head connection to the vagus nerve stimulator and EGG recording device.

At each data acquisition session, a within-subject design included 10 minutes of baseline recording (baseline) followed by 10 minutes of VNS. The VNS was a biphasic/bipolar signal, and its pulse amplitude and pulse width were set at 0.5 mA and 0.1 ms, respectively. Each animal received VNS at 10 and 30 Hz stimulation frequency on two different days resulting in a total of 14 data acquisition sessions.

EGG was recorded with a 2 KHz sample rate using a Grapevine NIP system (Ripple Neuro, Salt Lake City, UT USA). VNS was applied using an AM4100 stimulator (A-M Systems, Carlsborg, WA USA).

2.2.2 Pre-processing

Data were acquired from planar electrodes implanted on the serosal layer of the stomach in each ferret. Owing to the uncontrolled locomotion of the animals, which led to sporadic electrode or cable disconnections, the data acquired from each electrode exhibited varying durations and quality. Consequently, we decided to pick the signal from the electrode 'gastric2' that was assumed to be closest to the pacemaker area of the proximal stomach [82, 83] (See Figure 2.1). Before feature engineering, we developed an in-house pre-processing pipeline, written in Python, to prepare raw signals for downstream analysis. Raw signals occasionally contain spikes with large amplitudes that are not physiologically plausible. This could have been a result of the animal's sudden movements or an electrode loose connection. We empirically found that a threshold level of 1e8 µV can remove all these spikes (See Figure 2.2).



(a) (b)

Figure 2.2: Left panel: Effect of pre-processing steps. a) Raw baseline signal, b) Baseline signal after thresholding, c) Baseline signal after band-pass filtering. Right Panel: Effect of pre-processing steps. a) Raw VNS at 10 Hz signal, b) VNS signal after thresholding, c) VNS signal after band-pass filtering.

A sampling frequency of 2 kHz is several orders of magnitude greater than the slow wave and spike potential responses typically observed. Nevertheless, given that this domain remains relatively unexplored, we elected to include frequencies up to 1 kHz in our sampling procedure. This decision was made to encompass both known and potentially novel, higher-frequency signals that may be present within the data. In addition to this, by using a higher sampling rate, the quantization noise is spread over a wider frequency range, and then a low-pass filter can be applied to remove high-frequency noise, resulting in a cleaner signal. Finally, a higher sampling rate can help mitigate aliasing issues [84] caused by interference from other signals or noise sources. This can be particularly useful in environments, like the lab where ferrets were kept, with much electromagnetic interference or other signal disturbances. As the main frequency of EGG is in a narrow frequency band that is near DC (0.01-0.5Hz), we decided to use a digital filtering approach called Index Blocked Discrete Cosine Transform Filtering Method(IB-DCTFM) [85]. This method removes unwanted frequency range signals on the time domain by blocking specific DCT index on the DCT domain. Although like IIR filters, IB-DCTFM may cause signal distortion, such as the Gibbs phenomenon, but in comparison to FIR and IIR counterparts, IB-DCTFM provides several advantages including superior SNR and correlation coefficient to clean signal, stability, linear phase, and zero delay. IB-DCTFM has been used as a filtering method for EGG signals [86]. After bandpass filtering, we applied another thresholding but this time with the threshold level set to 2000 µV to keep signal amplitude in a physiologically plausible range. The thresholding procedure was done by substituting the values surpassing the designated threshold with the mean value of the signal amplitude. This method ensures that the signal is effectively constrained within the established bounds while maintaining its overall statistical properties. Our proposed analysis pipeline is built upon four modules (1) pre-processing and time and frequency domain feature extraction, (2) Feature selection using our proposed voting algorithm, (3) training and validating classifiers for two different classification scenarios, and (4) reporting the feature importance and classification metrics (See

Figure 2.3).

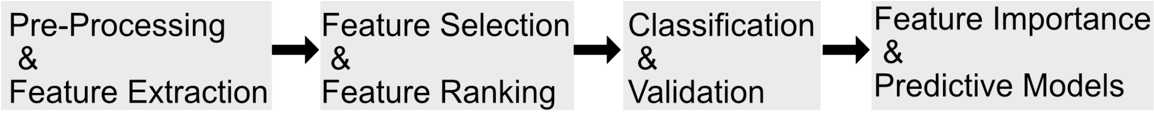


Figure 2.3: The analytic pipeline for EGG analysis and biomarker identification.

2.2.3 Feature engineering

For feature engineering, we used windows of 1-minute length with 20 seconds overlap for each 10-minute segment of our EGG signal. The choice of 1-minute length is to capture frequencies as low as 3 cycles per minute (CPM). This lower limit of 3 CPM is the reported bottom range for ferret gastric slow wave signals [87]. As the EGG signal is not stationary and has a chaotic nature, we hypothesized that dominant frequency (DF) and other features that are derived from DF, may not accurately describe the effects of VNS. Furthermore, existing research demonstrates that under conditions of dynamic and noisy EGG, particularly during rapid and unexpected movements, the selection of suitable EGG biomarkers (features) assumes heightened significance for maintaining the validity of the analysis [88]. Other fields of biosignal analysis, such as EEG or Electromyography (EMG) analysis [89], have developed features from both time and frequency domains that could be more suitable to extract information from non-stationary signals. In the following two sub-sections, we introduce the features that we adopted from the literature to represent 1-minute segments of the EGG signals.

Time domain features

Time-domain features (TDFs) are derived from the amplitude of EGG signals, capturing various characteristics that reflect the underlying dynamics of the data. Previous research has demonstrated that the amplitude of EGG signals is influenced by factors such as the ingestion of food or pharmaceutical substances [90], as well as the presence of nausea [88]. Consequently, we hypothesized that the statistical distributions of EGG signals during baseline and VNS periods would exhibit differences. We calculated a group of statistical features including mean, variance, mode, median, skewness (third moment describing data asymmetry), and kurtosis (fourth-moment determining tailedness of the distribution). Root mean squared value (RMS) and Line length are TDFs pertinent to signal amplitudes. Notably, Line length serves as an approximation of Katz's fractional dimension, as described in previous literature [91]. RMS offers insights into a signal's overall energy [92], which can facilitate differentiation between distinct signal classes or detection of particular events. For instance, research findings presented in [93] demonstrate a higher mean value or RMS during fasting as opposed to the postprandial state. The application of RMS in EGG analysis, therefore, may provide an additional perspective for understanding and interpreting data. Both RMS and Line length have been employed in EEG [94] and EGG studies [95, 96], attesting to their relevance and applicability in the analysis of such signals. Fractal dimensions, including PFD [97], have been extensively employed in EEG and ECG literature [98–101], indicating that they may offer valuable insights into the complexity and self-similarity of physiological signals. We direct readers with an interest in comparing various fractal dimension methodologies to consult reference [102] for a detailed examination and comparative analysis. Entropy is a measure of the unpredictability, complexity, or randomness of a signal or dataset [103]. Different entropy measures are related in the sense that they all quantify the complexity or randomness of a signal. Still, they do so using different approaches and algorithms. Some measures are more suitable for specific types of signals or applications. For example, approximate and sample entropy is more suitable for analyzing the regularity of time-series data, and permutation entropy is particularly useful for non-stationary signals [104]. It can be applied to study the dynamics and interactions of complex systems, such as biological systems. Entropy-based measurements serve as valuable tools for quantifying uncertainty and disorder in time series signals [105], including EGG signals [88]. Among the various entropy measures, approximate entropy and sample entropy are particularly useful for assessing the regularity and fluctuation in a time series [92]. Sample entropy has been demonstrated to be a robust feature for analyzing noisy EGG signals [88]. In addition to sample entropy, permutation entropy [104] and Singular Value Decomposition entropy (SVDEn) are employed to evaluate the local complexity and temporal-spatial complexity of a process [91], respectively. SVDEn has been employed in the examination of heart rate variability, owing to its straightforward implementation, and reduced computational complexity, particularly when analyzing short, nonstationary data series [106]. Signal variance, also known as the Hjorth activity parameter, is another time-domain feature. It indicates the surface of the power spectrum in the frequency domain [107]. Mobility and complexity, the other two Hjorth parameters [108], were also selected as time-domain features for their unique contributions. The mobility parameter is defined as the square root of the ratio of the variance of the first derivative of the signal to that of the signal itself. This parameter offers insights into the signal's dynamic characteristics. Meanwhile, the complexity parameter reveals how similar the shape of a signal is to a pure sine wave, providing information about the signal's waveform morphology. The value of Complexity converges to 1 as the shape of the signal becomes more similar to a pure sine wave [107]. In the context of signal analysis, certain features such as RMS and entropy measures may not exhibit a direct mathematical relationship. Nevertheless, these features can be employed in conjunction to provide a comprehensive understanding of a signal's characteristics. For instance, a high RMS value coupled with elevated entropy may be indicative of a signal characterized by significant noise and an abundance of random variations. Conversely, a high RMS value accompanied by diminished entropy could suggest the presence of a robust, periodic signal exhibiting a regular pattern.

Frequency domain features

Analysis of frequency domain features (FDFs) is important because FDFs can provide information regarding the rhythmic patterns of signals. In the field of EGG, DF, or peak frequency, is an FDF that has been widely used by researchers for EGG-related analysis [109]. Dominant power (DP) or the power associated with DF is another feature often used along with DF [88]. Spectral entropy (SE) is a measure of the random process uncertainty from the frequency distribution. SE has been used to measure the depth of anesthesia using EEG [110]. A low SE value means the frequency distribution is intense in some frequency bands. Its calculation is similar to that used for the Shannon entropy, but it replaces the probability distribution with the normalized power spectral density [111] (PSD). We calculated the mean value of signal power for 3-8 CPM, 8-11 CPM, and 11-15 CPM bands equivalent to bradygastria, normogastria, and tachygastria bands [112]. There is no broad consensus on what these ranges should be in ferrets, however, we chose these ranges based on methodology and findings of past studies [112, 113] related to ferrets EGG. In order to calculate the PSD, we set the desired frequency resolution to 0.1 CPM and used the Welch method. Table 2.1 lists the time- and frequency domain features used in this chapter.

2.2.4 Feature selection

As the number of features, also known as predictors (p), increases, the domain that they span increases at rates that the available data become sparse. This, in turn, requires more samples (n) to provide effective coverage of the domain for a predictive modeling problem such as classification. This concept is known as the "curse of dimensionality” [121]. As samples in high dimensional space may become equidistant, machine learning algorithms that use distance measures or other local models (in feature space) often

Table 2.1: List of time- and frequency-domain features used for EGG analysis. PMMP: Percentage of PSD that has higher value than DP/4. BP1-3: Relative band power between 3-8cpm, 8-11 cpm, and 11-15 cpm respectively.

Time domain	Frequency domain
Mean value	Dominant frequency (DF)[114]
Variance	Dominant Power (PD)[76]
Mode	PWMD[76]
Median	Spectral Entropy[115]
Skewness	BP1[114]
Kurtosis	BP2[114]
RMS	BP3[114]
Line Length[116]	Crest factor of PSD[76]
Approximate Entropy[88]	Median frequency
Sample Entropy[88]	Mean power frequency[76]
Permutation Entropy[117]	
SVD Entropy[91]	
Lempel-Ziv Complexity[118]	
Hjorth Mobility & Complexity[84]	
Petrosian Fractal Dimension (PFD)[119]	
Hurst Index[120]	

degrade in performance as the number of features is increased [114]. In this study, we had a total of 223 samples and 29 features. For the first scenario, VNS at 10 Hz, we had 114 samples (baseline=61, during VNS=53). For the second scenario, VNS at 30 Hz, we had 109 samples (baseline=51, during VNS=58). Drawing upon features utilized in other fields, such as EEG, ECG, or EMG, the present study incorporated infrequently employed features, such as PFD 27, into EGG signal analysis. Due to the lack of prior exploration of these features in the EGG domain, it was challenging to ascertain their informativeness a priori. Consequently, the inclusion of potentially redundant or non-informative features may have an adverse effect on the performance of the classifier [76], necessitating further investigation and potential refinement of the feature set. To demonstrate the presence of redundancy and correlation among the features, a three-step process was undertaken. First, the Spearman correlation coefficient was computed for the features, resulting in a symmetric correlation matrix. Second, this matrix was transformed into a distance matrix. Finally, hierarchical clustering 64 was employed to group and organize the features based on their similarity. A correlation heatmap, generated using the ordered features, illustrated the extent of correlation among the engineered features by exhibiting distinct hot and cold clusters (Figure 2.4).

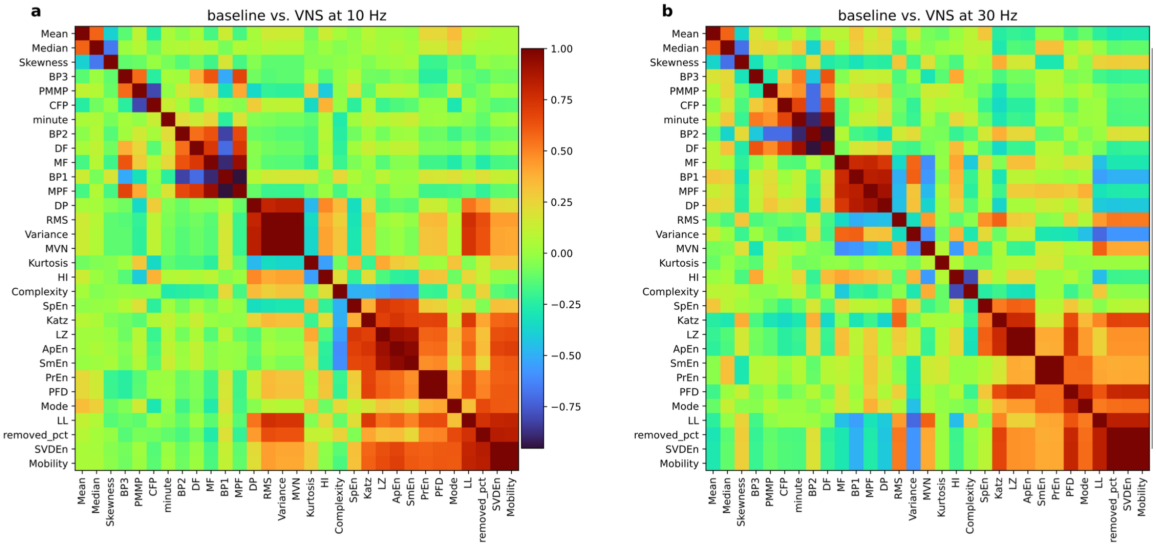


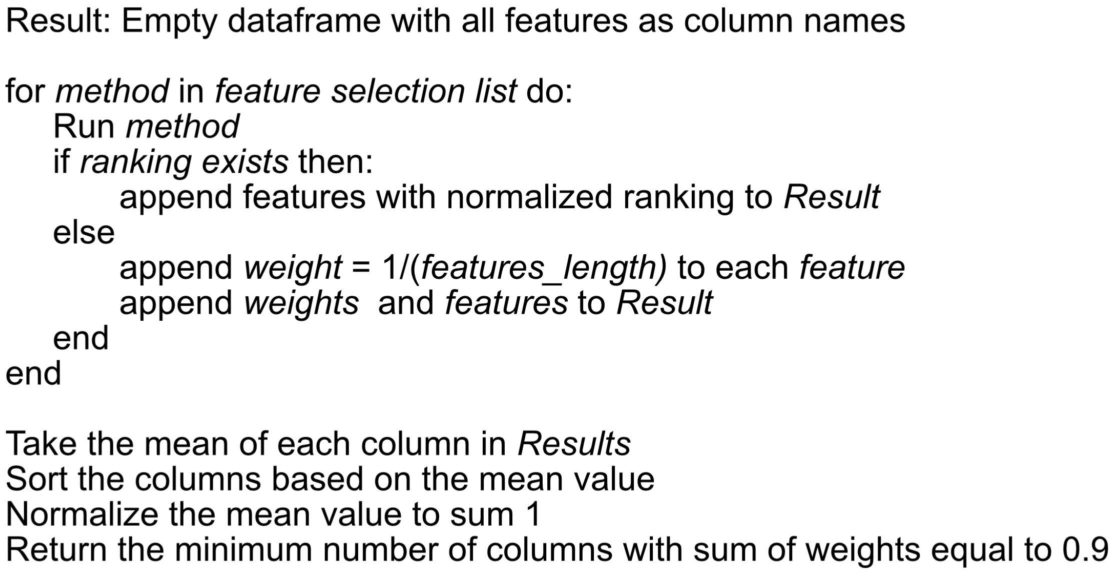
Figure 2.4: Correlation heatmap of the engineered features (See Table 2.1) a) baseline vs. VNS at 10 Hz b) baseline vs. VNS at 30 Hz. Each cell's color shows to what extent features are correlated.

As each feature selection method (Table 2.2) may select a different set of features with different orders (importance) [122], we proposed a voting algorithm to assign a

Table 2.2: List of feature selection methods used in the voting algorithm. Except for Variance Thresholding that is independent from the target variable and hence was applied to the whole data, all the other methods were implemented using 5-fold cross validation.

Feature selection with feature importance
ANOVA F-value Mutual Information[39]
L1-based Linear Support Vector Classifier (LSVC) [40,65]
L2-based LSVC[40,65]
Recursive Feature Elimination (RFE) Random Forest [39,40]
Backward Sequential Feature Selection (SFS) [40,65]
Forward SFS [40,65]
Permutation Importance (PI) Random Forest [23,65]
Feature selection without feature importance
Variance Threshold [39]

weight to each feature. These weights are scaled to add up to one. Feature importance is a by-product of some feature selection methods such as linear regression or decision trees [123]. Additionally, we used variance thresholding that removes all low-variance features. In this case, we had no feature importance, so we assigned an equal weight to each feature, the weight being 1/ (number of selected features). Next, we calculated the average weights for all features and sorted them based on their normalized rank. Ultimately, the optimal subset of features was determined by selecting features in descending order based on their respective normalized rank, with the cumulative sum of the ranks reaching a threshold of 0.9. A threshold of 0.9 for cumulative feature importance is based on a heuristic approach to retain a majority of the information while reducing the overall dimensionality of the dataset 66 (See Figure 2.5).



```
Result: Empty dataframe with all features as column names

for method in feature selection list do:
    Run method
    if ranking exists then:
        append features with normalized ranking to Result
    else
        append weight = 1/(features_length) to each feature
        append weights  and features to Result
    end
end

Take the mean of each column in Results
Sort the columns based on the mean value
Normalize the mean value to sum 1
Return the minimum number of columns with sum of weights equal to 0.9
```

Figure 2.5: Feature selection algorithm.

2.2.5 Model Selection/Training

We calculated all TDFs and FDFs listed in Table 2.1 for each 1-minute of the EGG signal. Following Algorithm 1, we selected the most important features for classification scenarios and used them to train the classifiers. The choice of classification method depends on the data and the context in which the classifier's output will be used. Finding a classification method with the highest predictive accuracy and interpretability is challenging in practical settings, especially in datasets with small sample sizes. Moreover, the desired trade-off between interpretability, accuracy, and computational efficiency also plays a crucial role in determining the appropriate method for a given task 67. There are advantages and disadvantages to each classification model under different circumstances. Decision trees are relatively fast and useful if one needs to share the results with an audience interested in how a conclusion was reached, however, they tend to overfit [124]. Support vector machines (SVM) are another choice for binary classification. They often provide high accuracy and tend not to overfit the data. Linear SVMs, as opposed to their non-linear counterparts, produce a linear decision boundary that can be easily understood and visualized [125]. However, the practitioners need to spend time training and tuning SVMs up front. Artificial neural networks (ANNs) are powerful for modeling nonlinear data with a high number of input features. However, ANNs can become computationally expensive. As the number of nodes and layers increases, it is difficult to interpret how an ANN has reached its solution, and fine-tuning an ANN may involve multiple steps and hyperparameters. In our case, with a rather small number of observations and a large number of features, it is crucial to select models that can effectively handle high-dimensional data and mitigate the risk of overfitting. To identify the best model, we took an empirical approach to test and discover which classifier achieves the best classification performance [126]. Considering the limitations of our dataset (small sample size, large number of features) and the research questions we aimed to answer, we used Random Forest classifiers [123], SVM [125] (with linear and radial kernel), Naïve Bayes classifier [127], and linear regression classifier for our binary classification tasks. The model selection procedure was as follows: Data were divided into training and test sets in a stratified manner to keep the ratio of baseline (class 0) to VNS (class 1) the same for both training and test data. 80% of the whole data was used for training, and the rest was used at the inference step. All models were trained with their default parameters and evaluated using stratified 5-fold cross-validation (CV). The shuffle parameter of cross-validation was set to False, to maintain the original sequence of the samples. After the initial training, we selected the best classifier based on its performance and tuned its hyperparameters. Tuning was implemented by utilizing a hyperparameter optimization framework named Optuna [128]. Optuna allows for dynamic construction of the search space and provides a combination of an efficient searching and pruning algorithm to speed up optimization.

2.2.6 Model evaluation

We utilized several evaluation metrics, including accuracy, ROC-AUC, f1-score, and f2-score, to assess the performance of our models. The ROC-AUC has been demonstrated to be a measure of choice for assessing the performance of a classifier for imbalanced data [129]. Consequently, we have chosen to highlight AUC as the primary metric in our results, while providing additional metrics such as accuracy, f1-score, and f2-score in the supplementary materials for further reference. To prove that our trained classifier has an ROC-AUC score significantly higher than the chance level (0.5), we conducted a permutation test [97, 130]. We first trained the classifier on the original dataset and computed its ROC-AUC scores using a 5-fold CV. Following this, we performed a permutation test by randomly shuffling the true labels of the dataset, retrained the classifier on this permuted dataset, and obtained the ROC-AUC scores of 5-fold CV for each shuffle. This procedure was repeated 1000 times to generate a distribution of permuted ROC-AUC scores. Next, we compared the ROC-AUC scores of the trained classifier on the original dataset to the distribution of permuted ROC-AUC scores by calculating the p-value, using a two-sample Kolmogorov-Smirnov (KS) test [97]. KS test is a non-parametric test that is sensitive to variations in both the location and shape of the empirical cumulative distribution functions pertaining to the two samples under consideration. If the p-value was found to be below 0.05 significance level, we could reject the null hypothesis and conclude that the trained classifier exhibited an ROC-AUC score significantly higher than 0.5, indicating its performance surpasses random guessing.

2.3 Results

In this section, we demonstrate our feature selection and EGG state classification results. We considered two scenarios, i.e., baseline vs. VNS at 10 Hz and baseline vs. VNS at 30 Hz. One motivation for feature selection is first to find features that are correlated with each other and second, to remove those with high correlation from the analysis. Figure 1 presents the clustered correlation heatmaps of all features for two distinct scenarios: baseline versus VNS at 10 Hz and VNS at 30 Hz. This visual representation facilitates a comprehensive understanding of the relationships among features. In addition to the engineered features, we included 'removed-pct' which is the percentage of removed signal after applying the pre-processing steps. We only used samples with 'removed-pct', less than 30% for model training. From these heatmaps, it is evident that there are clusters of features that are positively or negatively correlated with each other. The correlation heatmaps in Figure 2.4 suggest a positive correlation between 'removed-pct' and features such as variance and RMS. This can be an indication of higher error in signal measurements due to physiologically implausible high signal values or abrupt changes in recorded signals, reflected in RMS and variance, respectively. Figure 2.6 exhibits the features chosen for the first scenario, as identified by our feature selection algorithm and organized in a descending sequence of significance, as established by the Random Forest classifier (See Methods). The Sample Entropy [88] (SmEn) of the signal emerged as the most significant feature, while Root Mean Square (RMS) was identified as the second most important attribute. RMS is a measure of the signal's overall energy or amplitude [92], while Sample Entropy is a measure of the complexity or irregularity of a time series signal [92]. In the context of the first scenario, the differing Sample Entropy values between the baseline and VNS at 10 Hz suggest that the underlying dynamics or patterns of the EGG signals change because of VNS. Also, the variation in RMS could be indicative of the effect of VNS with 10 Hz frequency on the overall energy of EGG signals being analyzed. In general, most of the features selected in this scenario pertain to the signal's amplitude. (See Figure A.1)

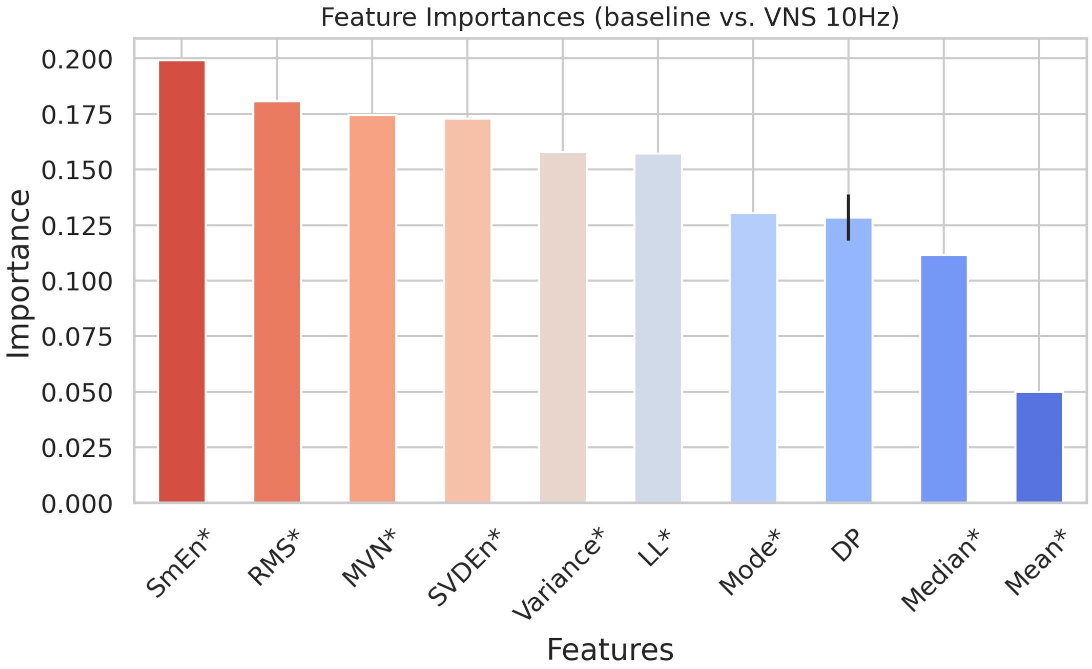


Figure 2.6: Selected features via our feature selection algorithm for baseline vs. VNS at 10 Hz (first scenario) and organized in a descending sequence of significance, as established by the Random Forest classifier. Error bars represent standard deviation.

Utilizing the chosen features, as illustrated in Figure 2.6, a Random Forest classifier [123] was trained. Figure 2.7 demonstrates the Receiver Operating Characteristic - Area Under the Curve (ROC-AUC or AUC [131]) of this classifier, encompassing three distinct cases: training the classifier using the selected features, employing randomly chosen features for training, and utilizing randomly shuffled labels for the training process. We conducted a two-sample Kolmogorov-Smirnov (KS) test [97] to assess the null hypothesis that the AUC values, derived from 5-fold cross-validation (CV) of the classifier trained with the selected features from Figure 2.6, originate from the same distribution as the AUC values for the other two cases presented in Figure 2.7. The obtained p-values were 0.0 (with test statistic of 1.0 and 0.872, respectively), enabling us to accept the alternative hypothesis that the AUC values for each case stem from distinct distributions (See Figure A.2 for metrics other than AUC). These findings indicate that our feature selection algorithm played an important role in augmenting the performance of the classifier, thereby demonstrating its effectiveness. (See Methods for details)

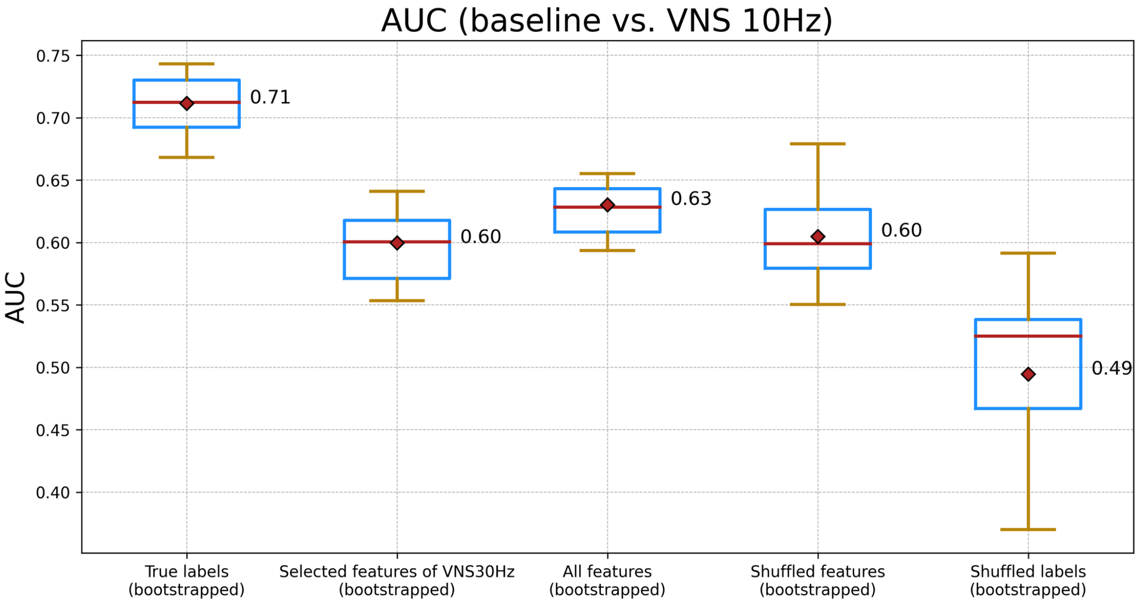


Figure 2.7: A comparison of the AUC for a Random Forest classifier trained with features from Figure 2.6, randomly chosen features, and randomly shuffled labels, examining the scenario of baseline vs. VNS at 10 Hz.

Contrary to the first scenario, Figure 2.8 reveals that the majority of the selected features for the second scenario, baseline vs. VNS at 30 Hz, are entropy and frequency-based. This observation suggests that, in the case of VNS at 30 Hz, the changes induced by the VNS are more prominently reflected in the signal's pattern or frequency content rather than its amplitude or energy (See Figure A.2 and Figure A.3). This distinction highlights the potential differences in the underlying mechanisms and effects of VNS at various frequencies, which may provide valuable insights into the physiological responses to stimulation. It is worth mentioning that the most important feature in the second scenario, Petrosian Fractal Dimension [119] (PFD), was originally introduced for the quantitative interpretation of epileptic EEG recordings [102, 119] (See A.1). Additionally, the presence of dominant frequency (DF), Dominant Power (DP), and normogastric band power (BP2) (See Table 2.1) in Figure 2.8, aligns with the insights provided by entropy-based features, further emphasizing that VNS at 30 Hz demonstrates a greater impact on signal pattern and frequency shifts, rather than on signal amplitude or energy.

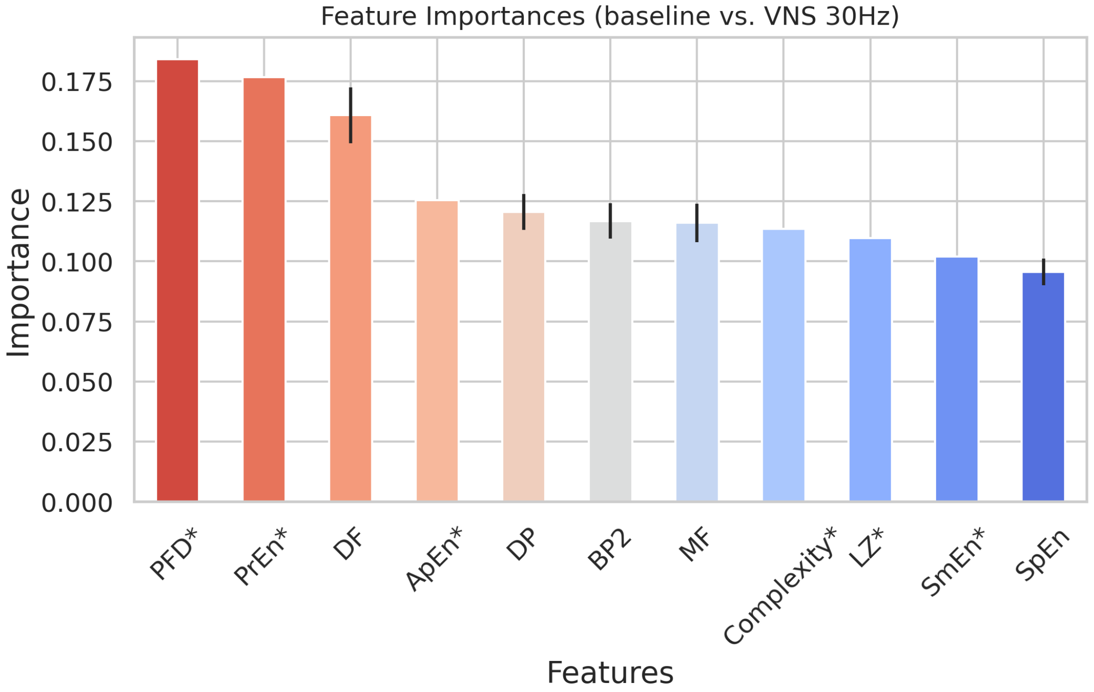


Figure 2.8: Selected features via our feature selection algorithm for baseline vs. VNS at 30 Hz (second scenario) and organized in a descending sequence of significance, as established by the Random Forest classifier. Error bars represent standard deviation.

Figure 2.9 demonstrates the AUC values for a Random Forest classifier trained for the second scenario (See Figure A.4 for metrics other than AUC). Similar to the first scenario, the AUC values derived from a 5-fold CV of the classifier trained with the selected features in Figure 2.8 were statistically significantly different from the other two cases showed in Figure 2.9 (two-sample Kolmogorov-Smirnov test, test statistic: 0.948 and p-value: 0.0) Taken together, the findings from Figure 2.6 to Figure 2.9 indicate that, within the framework of our study, it is feasible to distinguish the effects of VNS on the EGG signals. Furthermore, the frequency of VNS may modulate the alterations observed in EGG, manifesting either as changes in signal amplitude and energy or as shifts in signal complexity and frequency contents.

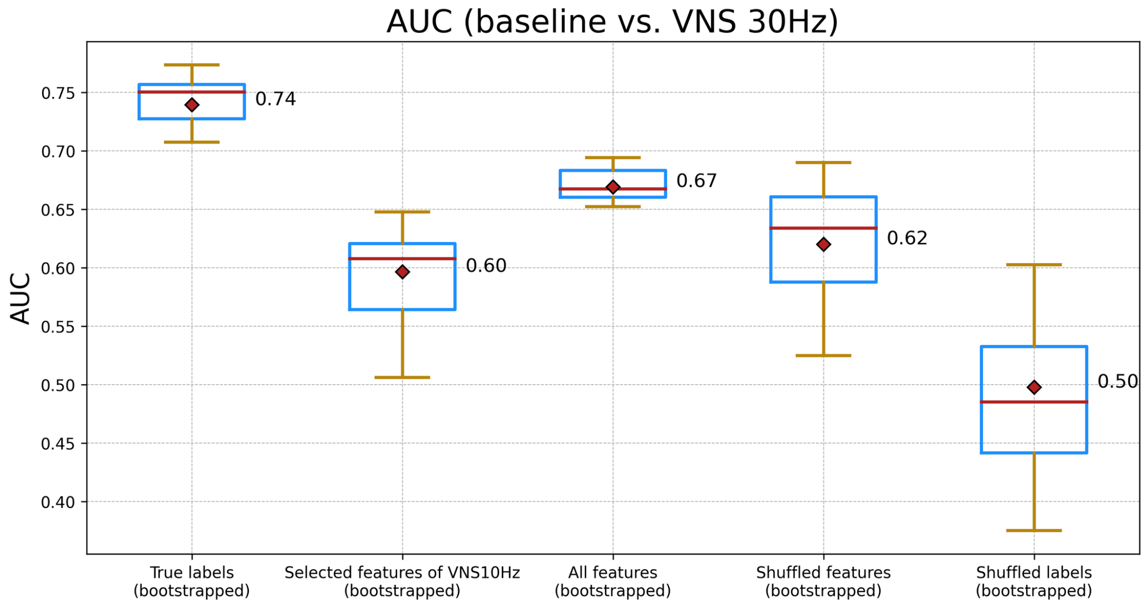


Figure 2.9: A comparison of the AUC for a Random Forest classifier trained with features from Figure 2.8, randomly chosen features, and randomly shuffled labels, examining the scenario of baseline vs. VNS at 30 Hz.

2.4 Conclusion

The main goal of this chapter was to introduce a novel dataset and analysis pipeline to determine the effect of electrically induced VNS on EGG signals. Characterizing this effect is essential for better understanding the underlying physiological mechanisms of VNS for regulating the GI function and inform designing closed-loop GI-VNS systems. Our study utilized a data acquisition schema that is rarely used due to the difficulties associated with surgery, electrode implantation, and long-term data collection. The advantages of this schema compared to cutaneous EGG were recording data that are not filtered by the abdominal wall and synthesizing a more realistic ambulatory setting. Disadvantages of the schema were loose electrode connections or abrupt movements of the ferrets causing artifacts in the recorded data. These artifacts were successfully removed by utilizing appropriate preprocessing steps. However, after removing the corrupted data portions, we were left with a small sample size, whereby 19.2% of the total collected data was discarded. Although dominant frequency (DF) is known to be the most widely used feature in EGG-related studies (See Methods), our preliminary analysis showed that DF alone was not informative enough to classify between baseline and VNS states in our data. To address this problem, we leveraged a machine learning approach. We employed a broader range of engineered features beyond DF to extract more information from our data. In exchange for these added features, we had to resolve another problem: the ratio between the number of samples (n) and the number of predictors or features (p). In our study, this ratio was close to 3. This can become problematic as most machine learning algorithms assume that there are many more samples than predictors 28 or p << n. In our case, the condition was exacerbated because the data was noisy, and heterogeneous, and could result in overfitting. There are multiple approaches such as filter, wrapper, and embedded methods [76] to handle datasets with too many features and a small number of observations. However, it is important to note that no single method is suitable for all datasets and situations 30. To harness the power of each feature selection method, we devised a voting algorithm to rank the features selected by each single feature selection method. We made a final decision based on majority voting. Our experiments demonstrated the efficacy of the voting algorithm, as evidenced by the enhancement of the AUC value of the trained classifier. Moreover, the interpretability of the model was improved by reducing the total number of features utilized for classification and organizing them according to their importance. A model is considered interpretable if a human can comprehend the rationale behind its predictions [116]. Reducing the number of features and organizing them according to their importance contributes to interpretability, as it makes it easier to understand how each feature influences the prediction. Our analysis pipeline helped to demonstrate that we could programmatically distinguish between EGG signals recorded during baseline and VNS indicating that the electrophysiological effect of VNS on EGG signal can be identifiable. This finding is in accordance with previous research that investigated the effect of VNS on GI function and whether it changes the EGG signal [132]. To examine the influence of VNS frequency on the alterations observed in EGG, we employed the selected features for VNS at 10 Hz and 30 Hz. The selected features revealed that the impact of VNS at 10 Hz was predominantly noticeable in time domain features associated with signal amplitude and energy. Conversely, for VNS at 30 Hz, features pertaining to frequency content and entropy of signal were of greater significance. There are limitations to our study in terms of data and methodology. The number of discarded samples was 33 and 44 for VNS at 10 Hz and 30 Hz, respectively. This 33% increment in discarded values in VNS at 30 Hz could be due to more loose electrode connections or a change in ferret movement patterns. Nevertheless, this difference in the quality of recorded signals during VNS at 10Hz and VNS at 30Hz could be a confounding factor and should be accounted for in future work. Another confounding factor is the state-dependency nature of EGG. For instance, two baseline recordings made of the same animal may differ if recorded in different sessions, based on how long before the recording session a ferret has ingested food. The same holds for signals recorded during the application of electrical VNS. A limitation associated with our methodology was that frequency domain features employed in this study are based on Fast Fourier Transform (FFT), however one assumption in FFT is that the input signal is periodic, but EGG is a non-stationary signal with chaotic properties. The future directions of our research will be focused on addressing the limitations of our data acquisition and analysis. The quality of collected data will be improved by more robust implantation of electrodes and employing wireless recording equipment. This will lead to a reduction in the number of invalid samples and more consistency in recording from different sessions. Regarding data analysis, we will adopt spectral analysis tools better suited for non-stationary signals such as wavelets and empirical mode decomposition [98] (EMD). These methods may provide a more accurate representation of EGG spectral information. Modern machine learning methods designed to generalize to Out-of-Distribution (OOD) data would offer an avenue to explore the state-dependent character of the recorded EGG and its inter- and intra-variability.

Chapter 3

Seizure Detection from In-the-Ear EEG Recordings

[1]

3.1 Background

Epilepsy is a neurological disorder characterized by recurrent seizures [134]. Seizures are distinguished by their temporal variability, wherein the frequency, location, and manifestation of seizures may fluctuate over time [135]. A notable example of this temporal variability is the seizure laterality shift, in which an individual's seizures transition from originating in one cerebral hemisphere to the other. Further, the timing of seizures may also be affected by external factors, such as changes in medication, sleep patterns, stress levels, and hormonal fluctuations [134]. Electroencephalography (EEG) has long been the standard of care for monitoring cerebral activity, and is critical in clinical practice for detecting seizures [136]. However, given their paroxysmal nature, seizures and interictal epileptiform discharges often prove elusive to record

---

[1] Parts of this chapter have been published in [133] under the terms of the Creative Commons Attribution 4.0 license.

with EEG, even under continuous patient monitoring [137, 138]. Scalp EEG is a widely used diagnostic tool for epilepsy, allowing for the detection and characterization of abnormal electrical activity in the brain. Unfortunately, as important as it is, current outpatient EEG monitoring is limited to approximately three days [139, 140], as traditional scalp EEG electrodes cannot be maintained for longer intervals. Longer-term monitoring requires inpatient hospitalization, where EEG technologists can monitor the EEG tracings and repair electrodes when needed in this setting. However, inpatient monitoring is expensive, inconvenient for patients and families, and not readily available outside of major metropolitan areas. Additionally, patients frequently do not have seizures or interictal activity during inpatient monitoring, prompting future hospital readmissions [141]. The need to expand our capability for long-term EEG monitoring is critical not only because seizures are paroxysmal by nature, but also because the occurrence of seizures with different foci of onset (such as bitemporal onset) does not follow a normal distribution [142]. Seizures may occur from a single focus for several days or even weeks before any are seen from a distinct independent focus. Increasing the duration of EEG monitoring may identify additional epileptogenic foci and may thereby contribute to our understanding of why certain patients who are thought to have clearly delineated seizure foci may at times not obtain seizure freedom after the resection or ablation of the purported epileptogenic zone. Ambulatory EEG, which is the use of scalp EEG in a naturalistic, home setting over longer periods of time, has been shown to be more sensitive in detecting seizures than traditional, in-office 20-minute recordings, particularly in persons with epilepsy whose seizures are difficult to predict or occur infrequently [139]. However, this traditional method has several limitations. The results obtained from traditional EEG tests might not fully capture or accurately represent the patient's neurological state during their normal daily activities [143]. In addition to the unwieldy bundle of wires and data logger hardware used in scalp recordings, one of the most significant limitations is the duration of recording, which is limited by skin breakdown that can occur from chronically adhered wet electrodes [144]. Recording integrity also reduces over time as the fidelity of scalp EEG recordings cannot be maintained without frequent, sometimes daily, repairs. The limitations of scalp EEG have led scientists to pursue lower-profile, long-wear ambulatory EEG technologies, such as in-ear electrode devices. Prior studies have successfully shown the effectiveness of capturing in-ear EEG data using a variety of sensor interfaces. Some examples include sensors pressed into viscoelastic ear plugs ("foamies") [145], custom ear shells with embedded sensors worn within the ear canal and concha made from 3D scanned silicone ear impressions [146], and multiple-night sleep studies using similar custom shells with embedded sensors [147]. In [148], a multi-modal system encompassing a behind-the-ear EEG device, along with ECG and accelerometry, was investigated for its potential in seizure detection. This integrated approach showed promise and [149] utilized behind-the-ear EEG in addition to ECG and photoplethysmogram (PPG) to propose a monitoring system for epileptic users, however, they observed the loss of EEG information in ambulatory settings, a shortfall further compounded by the absence of seizure detection outcomes. Turning to the specific realm of behind-the-ear EEG, the study in [150] considered seizure detection based on behind-the-ear EEG and reported that the obtained sensitivities were too low for practical use, however, this work illuminated the complementary role of ECG in enhancing the seizure detection process. Meanwhile, [151] showcased the capacity of behind-the-ear EEG for visual recognition of ictal EEG patterns as well as being used in a seizure detection algorithm. However, they reported that the ictal EEG data used in their study were recorded with the hospital system using Ag/AgCl electrodes. Further, [152] presented the use of a commercial EEG device to record behind-the-ear EEG to evaluate seizure detection algorithms in hospitalized patients. Finally, the authors of [153] conducted a feasibility study to demonstrate the safety of recording long-term ear-EEG in patients with Alzheimer's disease, stopping short of extending these findings to the development of a seizure detection algorithm. The above studies show successful capture of EEG data from within the ear or behind the ear which in some cases compared favorably to scalp EEG monitoring. However, these studies were built around bench-top electronic prototypes which are not scalable solutions for personalized long-term ambulatory EEG monitoring. In contrast to these lab-based in-ear EEG studies, this study demonstrates the design and manufacture of an ambulatory in-ear EEG monitor using scalable, commercially available 3D scanning, computer-aided design (CAD) modeling, and 3D printing processes and techniques used for designing and manufacturing hearing aids and custom high-end in-ear monitors thus, demonstrating the potential for long-term comfort and wearability needed for in-ear EEG devices. The Aware hearable is designed to conduct continuous monitoring of brain activity, record data on the device for long-term analysis, and provide valuable information pertaining to seizure patterns, triggers, and the efficacy of treatment, all in a non-invasive, ambulatory form factor. The long-term comfort and wearability of the device are particularly useful for individuals who experience infrequent or hard-to-predict seizures. Aware also offers the advantage of being applicable to a wider range of patients, including those who do not meet the criteria for costly surgically implantable monitoring and stimulation devices such as Responsive Neurostimulation (RNS) [154]. The Aware in-ear hearable is built using the same methods and practices as hearing aids and custom in-ear headphones, allowing for a sleek custom pro-consumer product design. Through discussions with subject matter experts and test subjects, it is believed that a future integrated system within a small wearable form-factor that unobtrusively blends into a user's daily wear and activities, making an EEG recording device more accessible both physically and in appearance compared to traditional scalp-worn electrodes, may lead to reduced social stigma of wearing a medical device and increase user acceptance for long-term data collection. This design approach aligns with current trends in wearable technology, ensuring the device blends seamlessly into the user's daily life and social roles without attracting unwanted attention [155]. It is worth noting that other wearable devices are used for seizure detection, such as wearable sensors and smartwatches [156–160]. These technologies leverage the various sensors found in these devices, such as accelerometers and gyroscopes, to detect seizures and alert the wearer as well as the caregiver. However, despite their convenience and non-invasive nature, they are worn at the extremities, making them prone to motion artifact, and are not as effective as EEG signals in detecting seizures that do not trigger substantial motor activity [161]. The Aware hearable, due to its dry electrodes placed within the ear canal with proximity to the brain, has an edge over these wearable devices in capturing seizures as EEG signals are considered the most effective signal for seizure detection, providing a direct measure of the electrical activity of the brain. While EEG is known to be highly susceptible to motion artifacts, which is a major concern for mobile EEG research, utilizing a 3D stereolithography (STL) digital model generated from United Sciences' "eFit" 3D ear scanner allows placement of the dry electrode sensors at the bony region of the ear canal known as the second bend, where the auditory canal passes through the skull wall. Utilizing the second bend allows the Aware hearable to fit and "lock" into place with the subject's unique ear anatomy, thereby reducing motion artifact and ensuring the best possible signal integrity. The primary objective of this study is to advance the field of chronic EEG recordings by examining the viability and interpretability of EEG data obtained from sensors within the ear canal, particularly due to their proximity to the temporal lobes—the most epileptogenic regions of the brain [162]. This is conducted in a clinical setting using the Aware in-ear hearable with embedded dry electrodes, manufactured through commercial processes. Supporting this, [163] indicates that ear-EEG can enhance source localization in temporal brain regions. Furthermore, [164] demonstrates that ear-EEG is especially sensitive to sources in the temporal cortex, owing to the proximity of the ear electrodes to these regions. During this study, the

Aware earbuds were worn by subjects who had just undergone invasive, intracranial EEG implantation, and had both ictal and interictal EEG patterns observable on previously obtained scalp EEG. These subjects had an array of up to 20 electrode probes implanted into various regions of their brains based on their pre-implantation hypothesis, including the temporal lobe. The subjects were asked to wear the Aware hearable earbuds beginning 24-48 hours after their implant procedure. For consistency, the earbuds were inspected daily and placed into the subjects' ears by a technician.

3.2 Methods

3.2.1 Earbud

To manufacture each set of Aware hearables, a proprietary non-contact eFit 3D ear scanner was utilized (United Sciences, Atlanta, GA) to 3D scan the unique anatomy of each subject's ear canal and concha where the hearable would be inserted. Originally developed for custom hearing protection and later commercialized into the hearing aid and custom headphone industries to replace industry-standard silicone impressions, the eFit scanner utilizes a patented [165] ring-laser scanner to scan the subject's ear, creating a near-perfect 3D scan of their ear, without any of the imperfections or pressure-induced distortion to the ear canal caused by the silicone impression process. Through stringent testing, the eFit scanner has shown greater volumetric accuracy and repeatability between ear scans than traditional silicone impression practices, with volumetric accuracy within $90\mu m$, allowing for a comfortable custom-fit with complete consistent contact with the inner surface of the ear canal, resulting in high-quality data recordings. Each customized device housing along with dry electrode contact points were modeled in a CAD software specific for earmold modeling (www.cyfex.com), 3D printed using a biocompatible photo-reactive acrylate thermoset photopolymer (https://etec.desktopmetal.com/), coated with a hypo-allergenic light-polymerizing single-component lacquer (www.otoplastik.dreve.de), and electroless nickel plated with silver/silver chloride (Ag/AgCl) for the dry electrodes, allowing the device to precisely and comfortably fit to the individual surfaces of the subject's ear and optimize the electrophysiologic signal quality (Figures 3.1 - 3.3). The ear shells were designed with pass-through vent ports to allow for normal hearing. 3D CAD modeling (traditionally from a 3D scan of a physical silicone ear-mold impression) and 3D printing have been utilized for decades for the design and manufacture of custom hearing aids. The new technique of 3D scanning the ear in its natural state with the non-contact eFit scanner enables full digital efficiency to this workflow. Novel to this study is the comparison of in-ear EEG to intracranial recordings, under the premise of demonstrating the use of 3D scanning and modeling techniques to design the custom fit dry electrode shape allowing for control of contact pressure for lower artifact with high comfort wearability. The goal of this research is to verify in-ear EEG as a potential means of capturing long-term ambulatory EEG data in a non-clinical environment with future use in consumer and medical devices.

Figure 3.1: Data logger box with hard-wired earbuds.

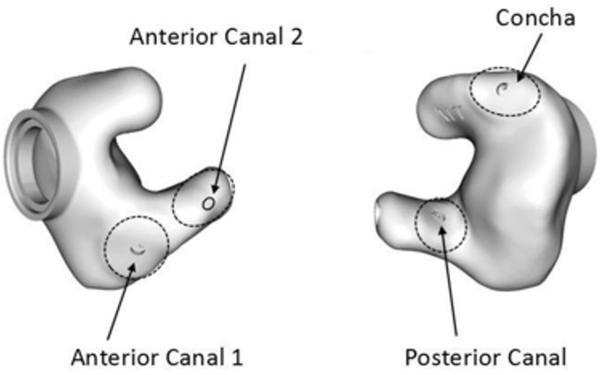


Figure 3.2: Detail of outside (on left) and inside (on right) of right earbud showing dry electrode placement.

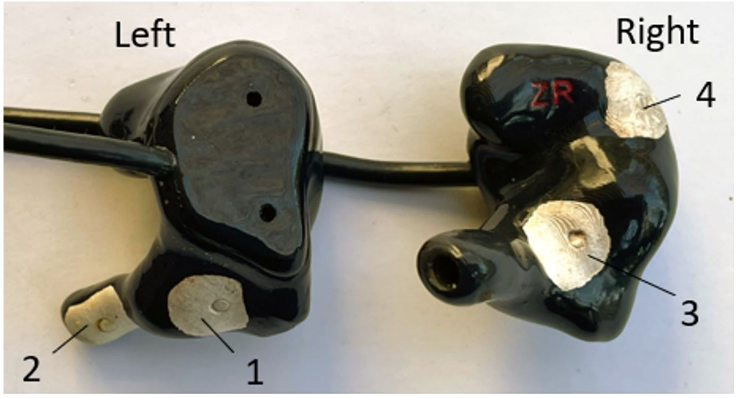


Figure 3.3: Detail of outside of left earbud showing anterior canal 1 (1) and anterior canal 2 (2) electrodes (posterior canal and concha electrodes not visible); detail of inside of right earbud showing posterior canal (3) and concha (4) electrodes (anterior canal 1 & 2 electrodes not visible).

3.2.2 Hardware

The earbuds were hardwired to a small datalogger that was located adjacent to the subject's main in-room datalogger, with wires running along the electrode bundle leading from the subjects' implant sight. The Aware hearable datalogger was built with an Arduino-compatible, 8-channel interface utilizing a 32-bit PIC32MX250F128B microcontroller with ChipKIT UDB32-MX2-DIP bootloader (www.microchip.com), with an ADS1299 digitizer (www.ti.com), sampling at 250Hz. Battery life was not a focus of this study. To ensure uninterrupted data collection, a freshly charged external 10,000mAh rechargeable battery pack was swapped daily during the morning skin inspection routine Raw data files are collected on a removable microSD card in .TXT format in 24-hour increments. The TXT files were downloaded to a secure laptop and converted to European Data Format (EDF) containing one uninterrupted digitized polygraphic recording for visual analysis using EDFbrowser (www.teuniz.net).

3.2.3 Electrodes

The earbuds feature eight electrodes combined, four in the left ear and four in the right ear as shown in Table 3.1. Each electrode is connected to an OPA2378 amplifier (www.ti.com) within the ear shell and is hardwired to the inputs on the data logger. The system is actively grounded using a conventional driven right leg (DRL) contact to the body with the right posterior canal electrode connected as a reference and the right anterior canal 1 electrode connected as the non-amplified BIAS input. Similar to its use in ECG, DRL is sometimes employed in EEG systems to reduce common-mode interference. By actively driving the right leg electrode, the system aims to create a common reference point that helps cancel out interference common to both the active electrodes and the reference electrode. This technique contributes to the overall noise reduction in the EEG signal, allowing for a more accurate representation of brain electrical activity. With the seven electrode inputs and one BIAS input, the remaining open channel was connected to a manual trigger switch used to send a digital signal

Table 3.1: Electrode names and their corresponding locations/descriptions

Electrode name	Electrode location/description
CH1	Left Anterior Canal 1
CH2	Left Anterior Canal 2
CH3	Left Posterior Canal
CH4	Left Concha
BIAS/DRL	Right Anterior Canal 1
CH5	Right Anterior Canal 2
CH6/Reference	Right Posterior Canal
CH7	Right Concha
CH8/TTL	Trigger sync between datalogger systems

pulse between the Aware data logger and Natus™bedside data logger system. This pulse served as a timestamp to align the Aware and Intracranial Montage (ICM) recordings for post-analysis. Each morning during rounds, the earbuds were removed for skin inspection, the rechargeable battery was replaced, and the system restarted.

3.2.4 Subjects

In accordance with Emory University institutional review board (IRB) protocols, informed consent was obtained from all participants in this study, and the study was conducted in accordance with the guidelines for ethical research. The enrollment criteria for the use of Aware for seizure detection included subjects 18 years of age and older who were admitted to the Epilepsy Monitoring Unit at Emory University Hospital for invasive, iEEG monitoring, and had both ictal and interictal EEG patterns observable on previously obtained scalp EEG. The study was initially limited to subjects with temporal lobe epilepsy, and recruitment was later opened to include those with non-temporal lobe epilepsy. However, there were certain exclusion criteria that needed to be considered, which included the inability to safely tolerate the use of Aware due to conditions such as antecedent skin breakdown or recent injury to the ear, participation in any other device trial that would preclude the use of Aware, and prior scalp EEG study with sufficient background abnormalities as to prevent observation of a posterior dominant rhythm or sleep spindles. Subjects were instructed to document their subjective experience of tolerability in a comfort diary that utilized a visual face scale, with accompanying descriptions, to report any discomfort or inconvenience associated with using Aware. Additionally, a daily skin inspection log was meticulously kept assessing for any adverse skin reactions or breakdowns. The subjects were asked to wear the Aware device for a minimum of 20 hours per day and were given the discretion to remove the device at any time, as well as instructed to remove the device during any participation in other research studies. Seven subjects were enrolled, with four undergoing data collection and analysis (Two male, mean age=41.5 years, ± 9.47). Subject 101 expressed greater post-op discomfort, and while initially wearing the hearable for several multi-hour increments at the start of the study, chose to discontinue the study before any usable seizure data was collected. Subject 103 had their iEEG cancelled and therefore no data was collected. All seizures necessary for clinical purposes for subject 105 were captured and the subject was subsequently explanted before their Aware hearable data collection study could be implemented.

3.2.5 Data Acquisition

The Aware earbuds captured EEG data over 413 hours of wear-time across four subjects, recorded seizures lasting in duration between 30 seconds to over five minutes, and several subjects wore the hearable for consecutive 24-hour intervals throughout the study including during sleep. While the high-quality EEG recordings from Aware allowed the interpreting epileptologist to detect most electrographic seizures, traditional visual analysis methods are inadequate for the sheer volume of EEG data generated by devices such as Aware. This necessitates the implementation of a quantitative approach to data analysis, specifically machine learning techniques, which have been demonstrated to be effective in identifying patterns in EEG data indicative of seizures and sleep states [165–169].

3.2.6 Manual Data Annotation

The EEG signal captured by Aware, in addition to the concordant iEEG signal obtained via Natus$^{\text{TM}}$ equipment, was reviewed by a board-certified Epileptologist. EDFbrowser was used to review the Aware data. Awake and asleep epochs were identified on iEEG via visual analysis. A minimum of 30 minutes of awake and 30 minutes of asleep EEG data was analyzed to determine the presence of epileptiform discharges, as well as to evaluate the normal sleep architecture. All seizures documented in the electronic medical record during the iEEG recording coinciding with Aware use were reviewed by the Epileptologist to determine if they were detected by the hearable. Seizures analyzed were anatomically correlated with either a post-operative MRI, post-operative CT or 3D reconstruction utilizing a pre-operative MRI and post-operative CT to explore differences in detection between the iEEG and Aware.

3.2.7 Data Analysis

To validate the EEG recordings from the Aware hearable, we developed a machine learning classification pipeline to classify between different physiological states using the Aware hearable EEG recordings. We formulated a seizure detection problem where the classification task was to differentiate between ictal and non-ictal states using the EEG recordings. We utilized intracranial EEG data obtained from clinical data acquisition systems (Natus$^{\mathrm{TM}}$) to benchmark our classification pipeline. We applied our pipeline to classify between sleep and awake states for additional validations of the Aware hearable EEG recordings. The classification pipeline was comprised of 4 stages: Preprocessing, feature extraction, and model training and evaluation.

3.2.8 Pre-processing

The raw hearable data, comprising eight electrodes (as detailed in Table 3.1), was utilized in the analysis. CH1 – CH7 were EEG data signals, and CH8 was a TTL-sync signal and was excluded from further processing. Based on the design of the hardware, CH6 was selected as the reference electrode and subtracted from the other electrodes. We chose the right posterior canal electrode as the reference electrode because we found in our experiments that it is less prone to motion artifact potentially due to its location at the first bend within the ear. The remaining six channels were first filtered to remove the line noise and its harmonics (at 60 and 120 Hz) and then filtered by a bandpass filter (0.5- 100 Hz) using a one-pass, zero-phase, non-causal bandpass

Table 3.2: Aware hearable montage description

Vector type	Comprising electrodes
Long vector	CH1-CH5, CH3-CH5
Short vector	CH1-CH2, CH1-CH3, CH2-CH3

filter. Following the visual examination of the filtered signals, we saw that one subject (Subject107) had excessive artifacts present on electrodes CH4 and CH7, likely due to deformation of the exterior ear pinna caused by pressure from the patient's head bandage overwrap. We decided to exclude data from these specific electrodes for all participants to maintain uniformity and integrity in the data analysis process. To emphasize the disparities in cerebral activity between various locations, we created a montage utilizing the remaining electrodes, namely CH1, CH2, CH3, and CH5 (Table 3.2). The idea of short and long vectors is based on unilateral and bilateral (cross-head) channel derivation, respectively [151, 152]. In prior studies, it has been observed that in addition to unilateral (short) channels, including bilateral cross-head (long) channels exhibit a higher significance in detecting epileptic activities. This observation is attributed to the inherent asymmetry characteristic of focal seizures [166]. Our neurologist annotated the ictal segments on both the Aware data and the concurrent iEEG signals (See Manual Data Annotation). For each ictal segment, we selected a corresponding window of equal duration and one hour preceding the onset of the seizure as the non-ictal segment. We labeled ictal segments with 1 and non-ictal segments with 0. Subsequently, we concatenated the non-ictal and ictal segments and standardized the entire signal. The signals were then segmented into one-second intervals with no overlap.

3.2.9 Feature Extraction

In the evolving field of seizure detection, the selection of appropriate features for classification purposes is pivotal. The features listed in Table 3.3 are rooted in the

Table 3.3: Summary of extracted EEG features

Feature	Description
Line Length [116]	A simplification of Katz's fractal dimension and a measure of complexity or simplicity in a signal.
delta	Signal power in 1-4 Hz range
theta	Signal power in 5-8 Hz range
alpha	Signal power in 9-12 Hz range
sigma	Signal power in 13-16 Hz range
beta	Signal power in 17-30 Hz range
low_gamma	Signal power in 31-45 Hz range
high_gamma	Signal power in 55-90 Hz range
Complexity [170]	Hjorth complexity, A measure to understand the complexity and structure of a signal.

historical context of EEG analysis and have been validated by numerous studies for their effectiveness in distinguishing seizure activity from normal brain activity. The Line Length (LL), an operational simplification of Katz's fractal dimension, has been shown to be an efficient feature for seizure onset detection [116, 171]. Complementing the time-domain analysis, the frequency-domain features — delta, theta, alpha, sigma, beta, and gamma — represent power spectral densities within their respective frequency bands. These bands are integral to EEG interpretation. Delta waves, for instance, are known to be prominent during deep sleep stages and have been observed to change during seizure episodes, particularly in temporal seizures [172]. Theta and alpha waves, associated with drowsiness and relaxed wakefulness respectively, also exhibit alterations during seizures [173]. Sigma waves, though less commonly emphasized, can offer additional insights into seizure dynamics, especially considering their normal presence during sleep spindles [174]. Beta waves, linked with active cognitive engagement, have been reported to increase phase-amplitude coupling with gamma waves during seizures, offering a potential biomarker for seizure detection [175]. Alterations in gamma activity have been correlated with the onset and spread of seizure activity, making it a potential feature for seizure detection algorithms [176]. Lastly, Hjorth complexity extends the analysis by offering a measure of the signal's

Table 3.4: Summary of sample size per subject

Subject	# ictal samples	# sleep samples	Hours worn	# seizures captured
Subject102	665	3100	216	5
Subject104	452	400	98.5	5
Subject106	107	720	29	1
Subject107	170	1740	70	2

overall volatility and unpredictability, which is inherently higher during seizures. This parameter adds depth to the feature set by encapsulating the dynamic nature of EEG signal changes during epileptic events [177]. In this study, we utilized each 1-second segment to extract time and frequency domain features as outlined in Table 3.3.

3.2.10 Model Training and Evaluation

In this study, the epileptologist conducted a comprehensive review of the Aware dataset, encompassing the recordings from all four subjects. Each instance of seizure activity was identified and labeled within the dataset. Subsequently, the validity of these identified seizure episodes was corroborated by reviewing iEEG data. Table 3.4 summarizes the total recording hours, number of samples in ictal(interictal) and sleep(awake) states, alongside the seizure counts for each subject. In the process of preparing our dataset for model training and evaluation, we partitioned it into two subsets: 80% allocated for training and 20% reserved for testing. To ensure that the distribution of classes remains consistent across both subsets, we employed stratified sampling. We used Logistic Regression and Random Forest classification models in our classification pipeline. Prior to inputting the features into the classifiers, we performed normalization to ensure the data was of uniform scale. Subsequently, we utilized the test data to assess the efficacy of the trained models. We employed accuracy, sensitivity, false positive rate (FPR), and receiver operating characteristic (ROC) curve as evaluation metrics [178]. Our seizure detection pipeline is depicted in figure 3.4.

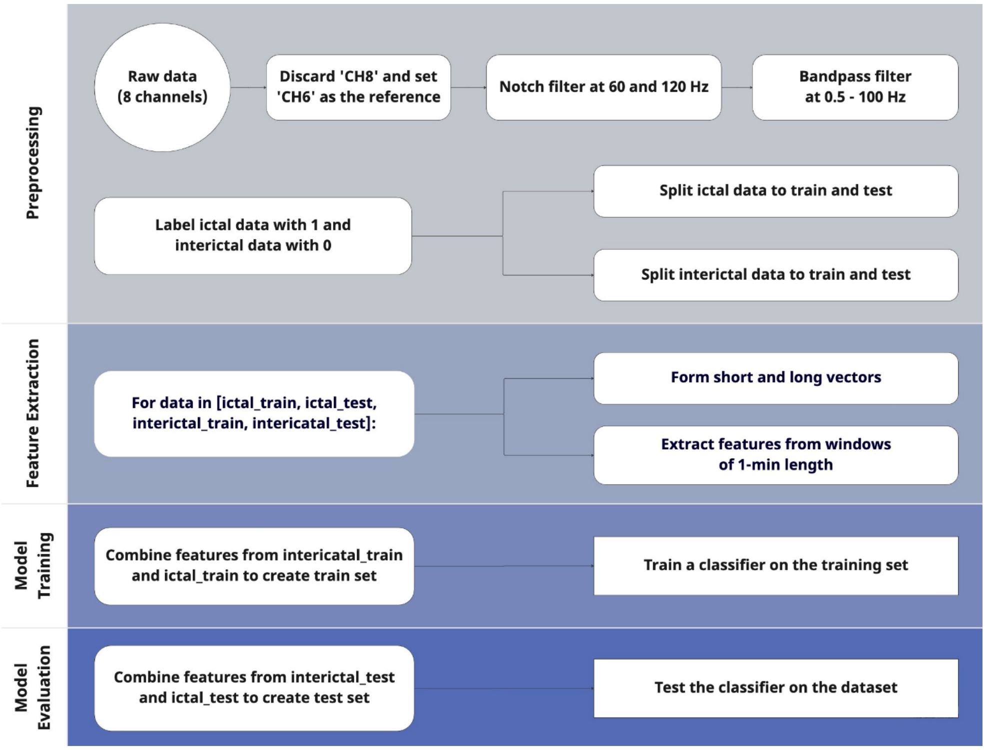


Figure 3.4: Flowchart illustrating the seizure detection pipeline, including preprocessing, feature extraction, model training, and evaluation steps.

3.2.11 Results

We tested the physiological validity of the Aware hearable EEG recordings, in the context of seizure detection and sleep classification applications. The inclusion of sleep detection results in our seizure detection study is founded on the premise that discerning the state of wakefulness or sleep during seizure events can enhance diagnostic accuracy and inform tailored treatment strategies. For instance, identifying whether seizures predominantly occur during sleep can guide more targeted medication regimens, such as administering doses primarily before bedtime and minimizing unnecessary medication exposure during wakefulness [179]. Figure 3.5 provides a detailed illustration contrasting seizure versus interictal conditions, as well as sleep versus awake states. In this illustrative example, the signals are derived from Subject106, selected randomly for demonstration purposes.

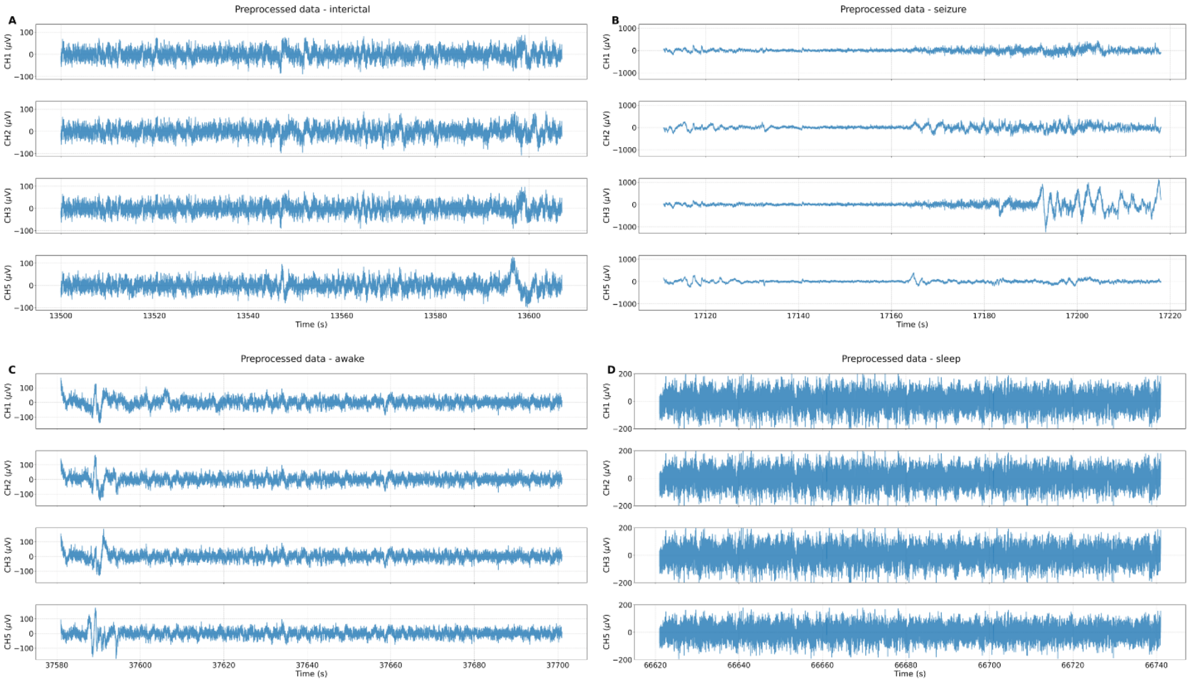


Figure 3.5: Comparative visualizations of preprocessed EEG data of subject106 captured using the Aware system. Panels (A) and (B) depict EEG activity during interictal and seizure states, respectively, while panels (C) and (D) exhibit recordings obtained during periods of wakefulness and sleep.

This depiction encompasses preprocessed data only from the electrodes used for the classification tasks. To demonstrate the efficacy of our seizure detection pipeline, we used the data pertaining to eight patients from the Kaggle seizure detection competition [180], as a benchmark. This provided us with a rigorous test of the performance of our approach. The Kaggle seizure competition data consists of training and testing data for both human and canine subjects. The training data consists of 1-second clips of EEG recordings labeled as either "ictal" for seizure data or "interictal" for non-seizure data. Data is described in more detail in [180]. In order to address the high dimensionality of the Kaggle data, which exceeded 6 channels, an additional preprocessing step was implemented involving the use of principal component analysis (PCA) [181]. Specifically, the first 6 components of PCA were selected for further analysis. We curated a balanced dataset by selecting an equal number of samples for ictal and interictal classes. The efficacy of our seizure detection pipeline is demonstrated by the performance of a trained Random Forest algorithm on the Kaggle test data, as depicted in figure 3.6. The mean accuracy among all subjects was found to be 0.79, with a standard deviation of 0.13. This indicates that the seizure detection accuracy of the studied subjects is consistently above the chance level [182], with a relatively low level of variance. The area under the ROC curve (AUC) was determined to be 0.93.

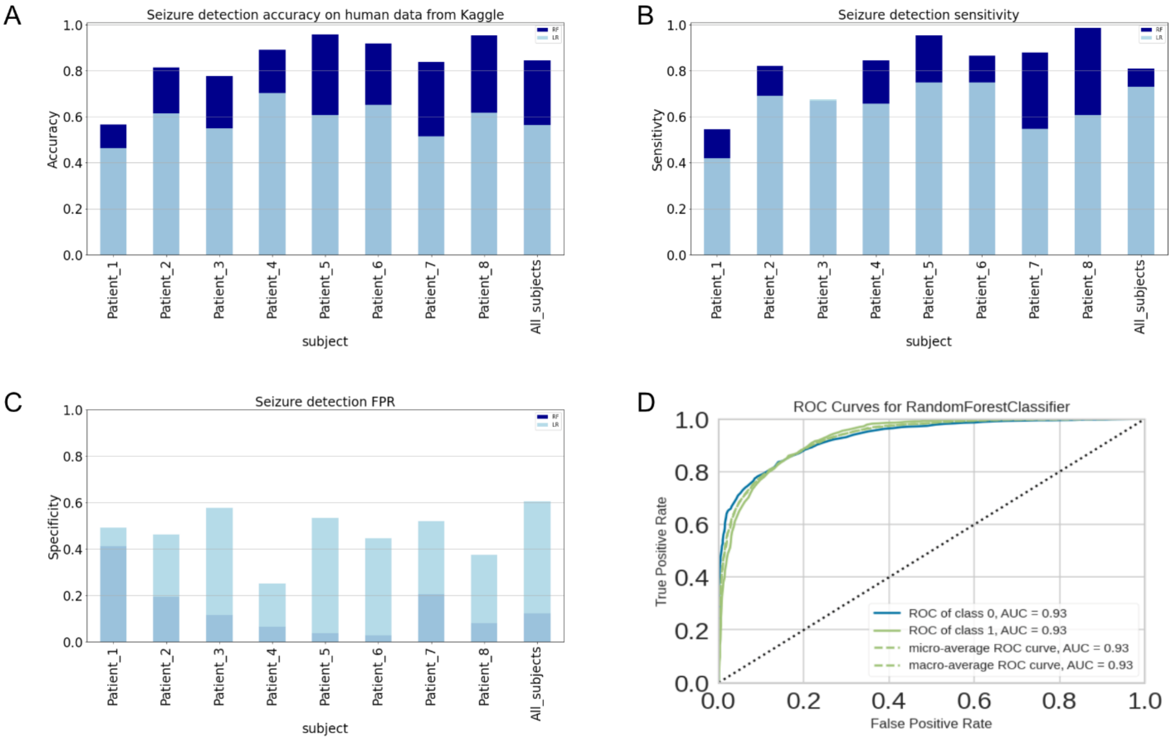


Figure 3.6: Results of seizure detection on Kaggle iEEG-PCA data. (A) Accuracy: logistic regression (LR) results were used as the baseline. Random Forest (RF) showed superior results compared to LR in all cases (B) sensitivity (C) false positive rate (FPR): the trained RF had a lower FPR than the baseline. (D) ROC curve for aggregated Kaggle iEEG-PCA data. (for readability only results of RF model are shown)

Additionally, in order to provide a reliable benchmark for our method, we utilized the iEEG data that was collected concurrently with Aware signals. As with the Kaggle data, we employed principal component analysis to address the high dimensionality of the iEEG data. The results of seizure detection, including accuracy, sensitivity, the false positive rate for each recording session, and ROC curve on aggregated data are depicted in figure 3.7. On average, the accuracy of our random forest and logistic

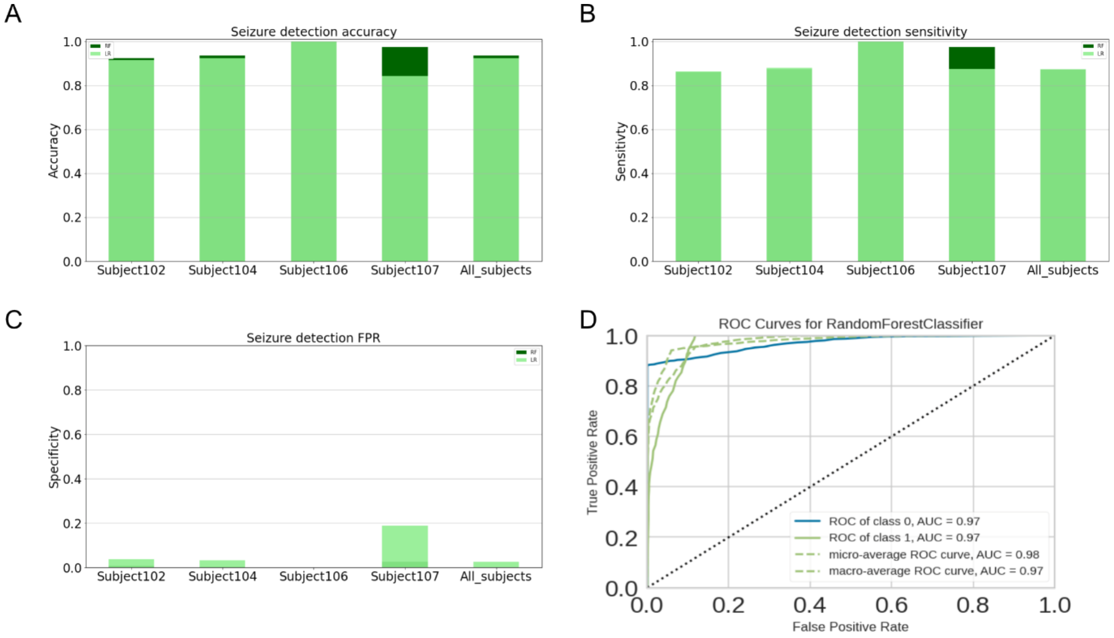


Figure 3.7: Results of seizure detection on iEEG-PCA data. (A) Accuracy: logistic regression (LR) results were used as the baseline. Random forest (RF) showed superior results compared to LR, in 4 out of 5 cases. (B) Sensitivity: LR had higher sensitivity than the trained RF in 4 out of 5 cases. (C) False positive rate (FPR): except for one case, the trained RF had a lower FPR than the baseline. (D) ROC curve for aggregated iEEG-PCA data. (for readability only results of RF model are shown)

regression models was 0.88 and 0.80, respectively, with standard deviations of 0.1 and 0.17. Similarly, the mean sensitivity of these models was 0.88 and 0.81, with standard deviations of 0.1 and 0.17. The mean false positive rate of the logistic regression and random forest models was 0.11 and 0.21, with standard deviations of 0.12 and 0.18. The ROC curve for the seizure detection results was obtained by combining data from all sessions, with an AUC of 0.97. The results of our seizure detection analysis using the Aware hearable data are depicted in figure 3.8, including accuracy, sensitivity, false positive rate for each recording session, and ROC curve on combined data. The average accuracy of our random forest and logistic regression models was found to be 0.86 and 0.80, respectively, with standard deviations of 0.13 and 0.20. The mean sensitivity of these models was 0.91 and 0.83, with standard deviations of 0.12 and 0.26. The mean false positive rate of the logistic regression

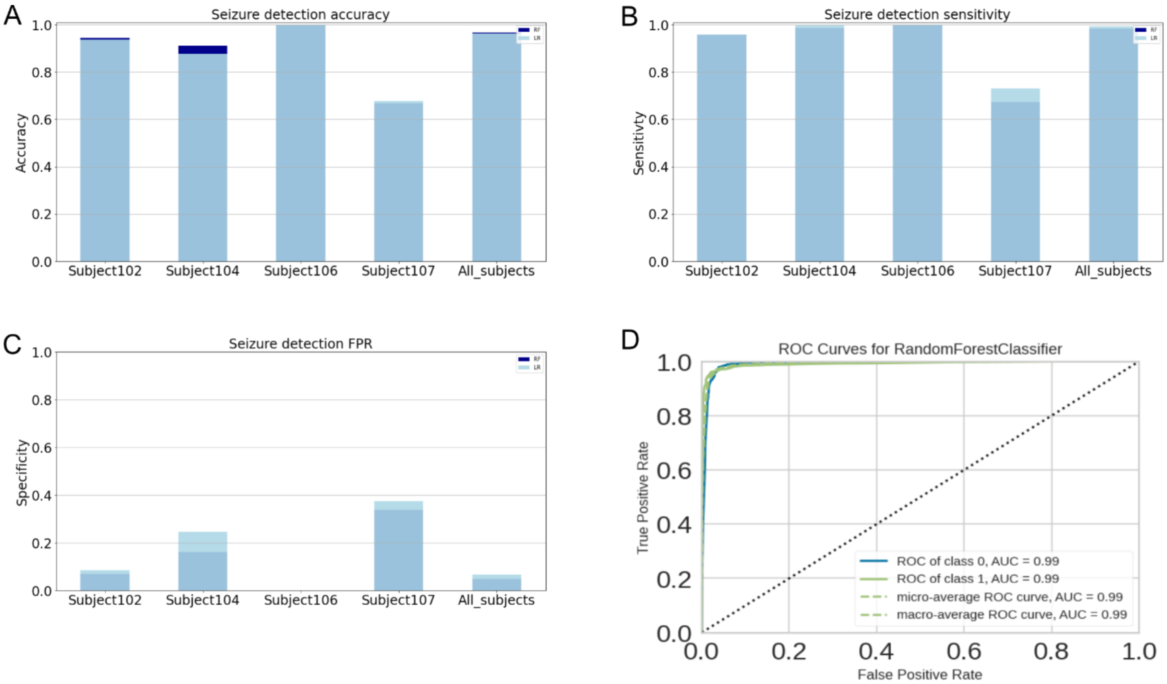


Figure 3.8: Results of seizure detection on Aware hearable data. (A) Accuracy: logistic regression (LR) results were used as the baseline. (B) Sensitivity (C) false positive rate (FPR) (D) ROC curve for aggregated Aware hearable data. (for readability only results of RF model are shown)

and random forest models was 0.18 and 0.21, with standard deviations of 0.15 and 0.18. The ROC curve for the seizure detection results was obtained by pooling data from all sessions, with an AUC of 0.99. In addition to the seizure detection test, we applied our machine learning classification pipeline to classify between sleep and awake states from the Aware data. Our epileptologist annotated sleep segments using a combination of iEEG data and video recordings. As with the seizure detection task, we selected a corresponding window of equal duration from the awake state for each sleep segment in order to facilitate comparison and analysis. The accuracy of sleep detection analysis utilizing Aware data is depicted in figure 3.9. Our random forest model outperforms the logistic regression model, with an average accuracy of 0.96 compared to 0.82. This is further supported by the smaller standard deviation of 0.03 for the random forest model compared to 0.16 for the logistic regression model. The AUC for sleep detection was 0.99. An independent samples t-test was conducted to evaluate the statistical distinction in classification accuracies derived from Aware

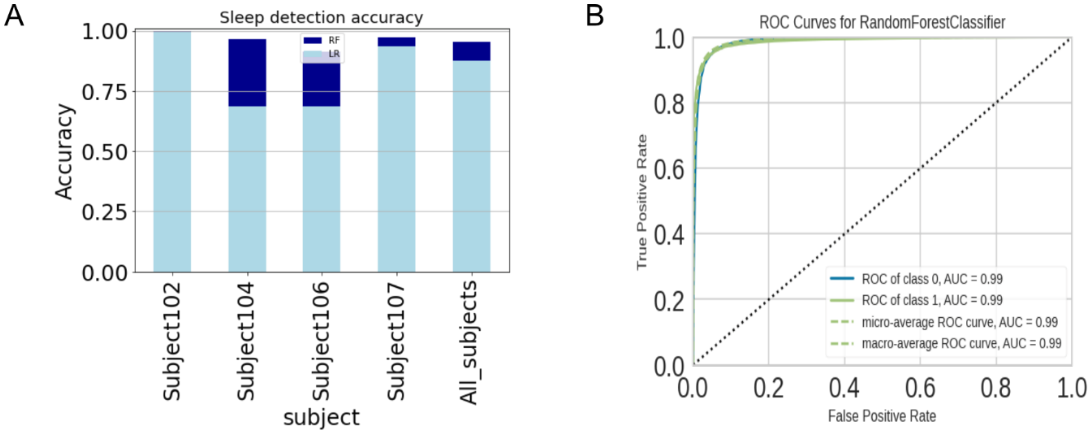


Figure 3.9: Sleep detection utilizing Aware hearable data (A) accuracy (B) ROC curve for the trained random forest.

data and iEEG data. The calculated t-statistic was 0.07, accompanied by a p-value of 0.93. These results indicate the absence of a statistically significant difference in the 'accuracy' metric between the two data sets. Consequently, we fail to reject the null hypothesis, which postulates equivalence in the performance of the classification accuracies across the two modalities. Regarding tolerability, most subjects reported varying degrees of comfort throughout their stay, which was confounded by the subjects having concurrent indwelling intracranial electrodes in place throughout the study, which are a universal source of discomfort. The overall mean tolerance was 3.98, suggesting that, on average, the subjects experienced a degree of discomfort marginally below the midpoint of the scale. It is imperative to consider that the participants were implanted with intracranial electrodes, a factor that might have affected the reported discomfort. The standard deviation was 1.12 which reflects a moderate variation in the tolerability scores across the subjects.

3.2.12 Conclusion

Previous research has established the potential of ear-EEG as a viable tool for seizure detection. This technology has demonstrated efficacy both as a standalone modality and as a component within a multi-modal system. In this study, we sought to determine the feasibility of using the Aware hearable manufactured from hearing aid industry standard materials and practices and equipped with embedded electrodes to record EEG signals from the ear canals. To this end, we collected EEG data from both left and right ears and compared it to data obtained using gold standard iEEG. Our results indicated that the ear canal recordings were suitable for detecting seizures and discriminating between wakefulness and sleep. The Aware hearable demonstrated the viability of using a non-invasive long-term electrophysiologic earbud device for extended periods of time in a clinical setting, thus potentially translating to ambulatory settings which would significantly reduce costs and burden to patients and families. Of note, the Aware hearable was not able to detect seizures that were deep and limited in propagation (Subject 104). This is a known limitation of scalp EEG [183], which the Aware hearable most closely resembles. A similar comment can be made about the instance where Aware failed to lateralize the seizure (Subject 102); propagation of ictal patterns is what is visible on scalp EEG, which can appear as bilateral signals. In contrast, iEEG electrodes have only limited spatial resolution, and if the implant does not cover the initiating nodes, it can give falsely lateralizing information; the discordance regarding lateralization in Subject 106 could be due to this or due to how the seizure propagated before being seen on the surface. The other limiting factor in this analysis is that epileptiform discharges observed on iEEG are often not observable on the scalp (and presumably Aware), however, there were no epileptiform spikes identified via the hearable despite all enrolled subjects demonstrating interictal epileptiform discharges on prior scalp studies (same for sleep spindles and posterior dominant rhythms). It is possible that this discrepancy is due to the limited spatial sampling of the hearable. Future trials of the Aware study might consider the adoption of scalp EEG instead of iEEG. This consideration stems from the notion that both scalp EEG and the Aware system utilize more comparable methods for EEG signal acquisition. Furthermore, the use of scalp EEG is generally less complex and more accessible compared to iEEG. Comfort level is highly personal, and several users showed no issues wearing the earbuds for 20+ hours a day for the length of their study. In our analysis of the EEG data obtained using Aware, we employed a combination of preprocessing, feature extraction, and classical machine learning techniques to classify the data. Our choice of classical machine learning algorithms was informed by the fact that they tend to perform better than deep learning algorithms when the available data is small. This is because classical machine learning algorithms make use of hand-crafted features and apply simpler models that are designed to be more robust with smaller datasets. As previously demonstrated in the literature, tree-based models often outperform deep learning models on tabular data [184]. There are several reasons for this phenomenon. For instance, tree-based models are more adept at handling high-dimensional datasets and interactions between features than deep learning models. This is because tree-based models can learn simple, interpretable decision rules based on a few key features, rather than attempting to learn complex, non-linear relationships across all features as deep learning models do. Additionally, tree-based models tend to be less sensitive to noise and outliers in the data, which can be problematic for deep learning models [185]. Furthermore, tree-based models are often easier to train and tune, particularly when the data is small or imbalanced. Given these considerations, we chose to employ random forest classifiers in our analysis. These classifiers are a type of ensemble learning algorithm that combines the predictions of multiple decision trees to achieve improved performance [186]. In the domain of seizure detection using ear-EEG, various studies have yielded promising results using machine learning classifiers. Reference [148] details the use of a Support Vector Machine (SVM) classifier applied to multi-modal data, achieving a sensitivity spectrum of 84% to 100%. Another study reported sensitivities between 77% and 82% when utilizing an SVM classifier with behind-the-ear EEG data [150]. Further research highlighted in [151] has echoed these findings, with sensitivities ranging from 83% to 100% using similar methods.

Diverging from SVM classifiers, [152] achieved a mean sensitivity of 90.4% with an autoencoder-based model. Our research contributes to this evolving field with a seizure detection system that not only aligns with the high sensitivity benchmarks previously reported but, in some instances, surpasses them. With a mean sensitivity of 91%, our system stands as a testament to the efficacy of using in-ear-EEG for long-term seizure monitoring. Overall, our clinical results corroborate with prior studies that ear canals should serve as a potentially viable alternative for recording EEG signals in ambulatory settings [187], and that the Aware hearable demonstrates this feasibility using scalable commercial manufacturing processes, with potential implications for the fields of neurology and brain-computer interface research. One possible application for this technology is the detection of sleep and wake states. To improve the detectability of motor seizures that are not visible as a detectable pattern from any scalp EEG system, accelerometers may be added to the Aware hearable in future designs. For future work, we will consider the incorporation of artifact removal algorithms within our preprocessing framework, with the objective of potentially improving the system's overall performance. We hope that using the Aware hearable device will provide patients with a much more comfortable experience with long-term seizure monitoring and detection while affording clinicians and researchers far greater amounts of data than existing non-invasive ambulatory methods allow.

Chapter 4

Neural Biomarkers of Memory: A Sparse-Label SEEG Analysis

This chapter investigates the predictive capability of feature engineering and classical machine learning in identifying neural biomarkers associated with memory recall, specifically focusing on the impact of basolateral amygdala (BLA) stimulation.

4.1 Background

The direct recording and stimulation of the human brain have significantly advanced our understanding of the physiological mechanisms underlying memory [188]. Stereoelectroencephalography (SEEG), with its exceptional temporal resolution, offers a unique capability to capture dynamic brain patterns in near real time. This technique allows for detailed observation of neural activity before, during, and after brain stimulation, providing invaluable insights into the temporal dynamics of memory processing. Recently, our research collaborators leveraged these techniques to determine whether electrical stimulation of the basolateral amygdala (BLA), a region with known connectivity to the hippocampus [189] and a key role in emotional memory [190], could affect memory processing. In doing so, we observed that BLA stimulation has the capacity to enhance declarative memory without evoking subjective awareness or an emotional response [191, 192]. Our subsequent work has sought to characterize which patterns of evoked activity are predictive of the memory-enhancing effects of BLA stimulation; in particular, we identified evoked slow-gamma (31-55 Hz) activity within the CA1 region of the hippocampus as a biomarker predictive of whether an image would subsequently be remembered vs. forgotten [193].

Despite the success of traditional methods such as linear models or manual feature extraction techniques in identifying key neural biomarkers, they remain inherently limited by their reliance on the manual selection of features. This approach often requires domain expertise and may overlook subtle but meaningful patterns in the data [194]. In the context of SEEG signal analysis, where the high dimensionality and complexity of brain activity present significant challenges, these limitations may become more pronounced.

In this chapter, we seek to investigate the efficacy of manual feature engineering and classical machine learning for predicting whether an image will subsequently be remembered or forgotten.

4.2 Methods

4.2.1 Study Participants

We recruited a cohort of 16 individuals (n = 31 experimental sessions) with medically refractory epilepsy undergoing a stereoelectroencephalography procedure for our study. A subset of these sessions (n = 17, 56%) were excluded from our analyses because of poor performance on the memory task (false alarm rate > 44%, a threshold expected based on our prior work [191]). The mean (± SD) age of the included cohort (n = 8 individuals, 14 experimental sessions) was 36.2 years (± 6.4 years), and 63% were female. Each study participant provided written informed consent prior to the experiment. No exclusions were made regarding the sex, gender, race, or ethnicity of participants. Study protocols and procedures were approved by Institutional Review Boards at Washington University in St. Louis (IRB 202104033) and the University of Utah (IRB 00144045).

4.2.2 Intracranial Electrophysiology

Intracranial depth electrodes were implanted to localize seizures in regions determined during a multidisciplinary case conference. Post-operative electrode placements were localized by co-registering pre- and post-operative images using open-source software developed by our research group [195]. Neurophysiological recordings were acquired using a neural signal processor (Blackrock Microsystems, Salt Lake City, UT; Nihon Koden USA, Irvine, CA) connected to the clinical EEG amplifier. (Cadwell Industries, Inc., Kennewick, WA). During a recording session, neural signals were sampled at 2 kHz and referenced to an electrode located within white matter, distant from the stimulation site.

4.2.3 Experimental Design and Stimulation

Patients performed a two-stage (i.e., encoding, retrieval) visual recognition memory task separated by approximately 24 hours. During the encoding session, patients were shown a series of 160 images consisting either of single objects (e.g., hammer) or scenes (e.g., city landscape); the proportion of objects and scenes shown were equal (50/50). Images were presented on screen for 3s and followed by a fixation cross for 6.5-7.5s. During half of the encoding trials, we delivered direct electrical theta-burst stimulation to the basolateral amygdala for the 1s immediately after the image was removed from the screen, similar to our previous work [191]. Stimulation was bipolar, charge-balanced, 0.5-1.0 mA, and delivered in 8 equally-spaced bursts of 50 Hz pulses. Memory was subsequently tested during a self-paced retrieval task in which patients were asked to identify images that were shown during the encoding session approximately 24 hours prior.

4.2.4 Signal Processing

We conducted a series of standard offline preprocessing techniques to enhance the signal-to-noise ratio of our SEEG recordings. Specifically, we implemented the following steps: (1) bipolar re-referencing of signals between adjacent contacts along the same lead, (2) use of a zero-phase band-pass filter at 0.1 Hz and 95 Hz, (3) application of a zero-phase notch filter to eliminate line noise at 60 Hz, (4) down-sampling of the recordings to 200 Hz, (5) segmentation of the data into 8-second peri-stimulation epochs, and (6) manual exclusion of electrode contacts and epochs exhibiting significant noise or epileptiform artifacts upon visual inspection.

4.2.5 Feature Engineering

We utilized multitaper analysis [196] to estimate the power spectral density (PSD) of the pre-processed SEEG recordings for each channel. Following this, we isolated the aperiodic (1/f) components of the SEEG power spectra by modeling the aperiodic signal L as an exponential function in semilog-power space, using linear frequencies and logarithmic PSD [197], as presented in the equation below.

$$L = a + \log(F^b)$$

Here, a represents the intercept, b is the slope, and F denotes the vector of input frequencies.

The average band power was computed for the canonical frequency bands and included as part of the input features. In an effort to enhance the predictive capacity of our models, we integrated non-linear features, specifically Permutation Entropy

(PE), Detrended Fluctuation Analysis (DFA), and Phase-Amplitude Coupling (PAC), into our analytical framework. These features were deliberately selected to capture complex neural dynamics that conventional band power measures may fail to detect. A comprehensive list of all input features is presented in Table 4.1.

Table 4.1: List of Features and Descriptions

Feature	**Description**
Canonical Frequency Bands	Average spectral power of theta, alpha, beta, slow-gamma, and high-gamma bands.
PE	Permutation Entropy, Measures the complexity of time series by analyzing the ordinal patterns.
DFA	Detrended Fluctuation Analysis, evaluates the long-term correlation and scaling properties.
PAC	Phase-Amplitude Coupling, measures interactions between phase and amplitude of different frequencies.

PE has been extensively employed to assess the complexity and unpredictability of neural signals by quantifying the inherent variability and disorder within these signals [198]. DFA is particularly suited for managing the non-stationary nature of SEEG signals, allowing the characterization of long-range temporal correlations in neural processes [198]. PAC is instrumental for understanding the modulation of faster oscillatory amplitudes by the phase of slower oscillations, a phenomenon linked to cognitive processes, sensory integration, and motor coordination [199].

We included all electrodes positioned within the hippocampal anatomical region, specifically those located ipsilateral to the stimulation site.

4.2.6 Predictive Models

We employed Logistic Regression with an Elastic Net regularization (LR-ENR), along with XGBoost, as our predictive models, utilizing the features outlined in Table 4.1. The Elastic Net method aims to address both multicollinearity and model sparsity by combining L1 (Lasso) and L2 (Ridge) penalties [200]. XGBoost is recognized for its robust capacity to model complex non-linear relationships within tabular data [201]. Unlike logistic regression, which assumes that there exists a linear relationship between each explanatory variable and the logit of the response variable, XGBoost employs gradient boosting on decision trees, allowing it to model intricate patterns and interactions among features that may not be apparent through linear methods. This algorithm can automatically account for feature importance and selection. However, we employed a permutation-based feature importance method [202] to determine the most important features for both LR-ENR and XGBoost classifiers. This model-agnostic approach assesses the significance of each feature by measuring the increase in prediction error when the feature's values are randomly shuffled, effectively disrupting any relationship between the feature and the target variable [202]. Specifically, the method calculates the baseline error of the model using cross-entropy, which measures the difference between the true labels and the predicted probabilities produced by the model. The cross-entropy is computed for the original feature set and then repeatedly permutes each feature's values to compute the cross-entropy loss again, with the difference indicating the feature's importance. For each feature, the permutation is repeated multiple times, and the mean and standard deviation of the error increase are recorded. The significance of each feature's importance is assessed using a t-test, comparing the importance score to zero. Pairwise t-tests are also performed between the most important feature and others to determine whether its importance is significantly greater. This ensures that we not only capture the feature's relevance but also determine if its contribution is statistically significant.

4.2.7 Model Evaluation

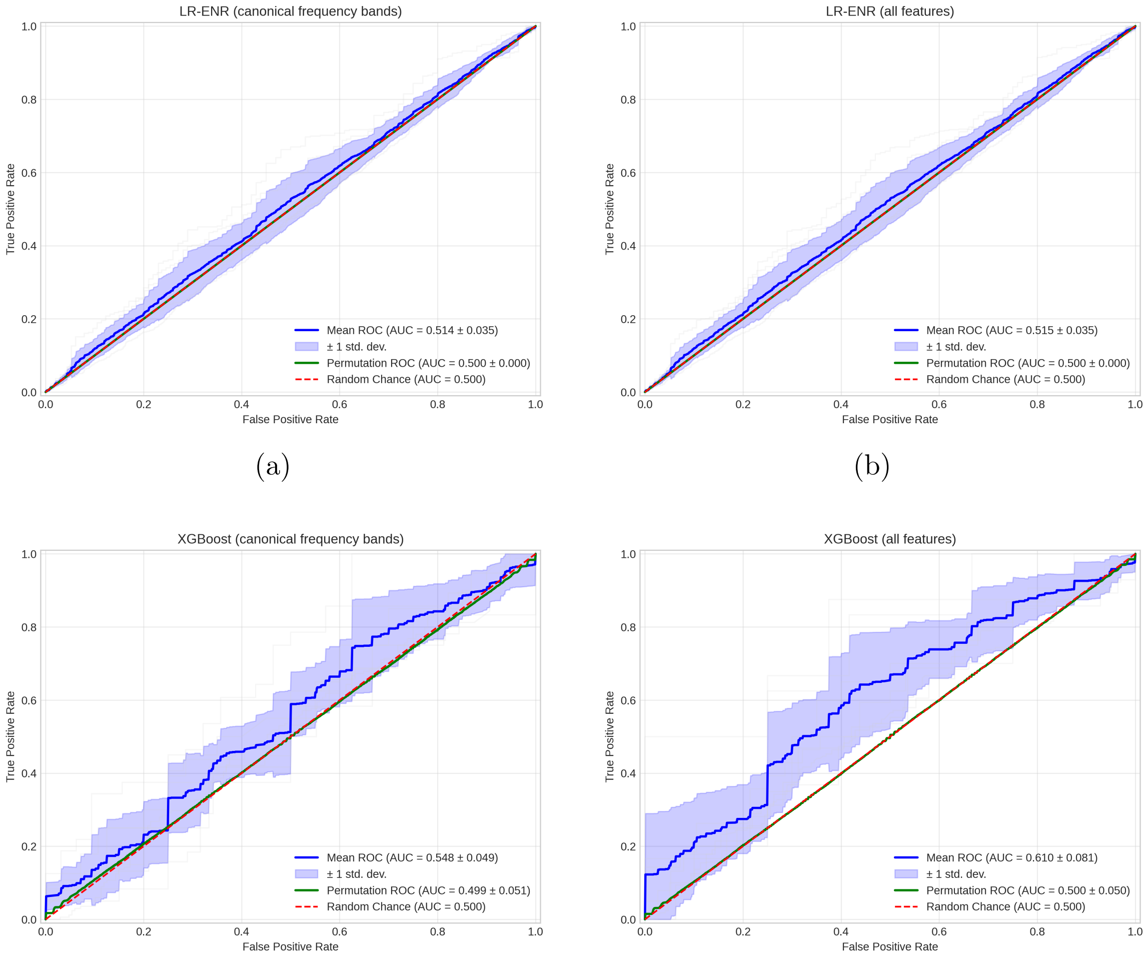
To assess the predictive performance of the trained logistic regression and XGBoost, we employed Leave-One-Subject-Out (LOSO) cross-validation, which involved training the models on data from all subjects except one, with the left-out subject used for testing.

To establish a chance-level baseline, we randomized the labels in the training set and repeated the aforementioned procedures. This step provided a reference point against which to compare the models' performance, ensuring that any predictive power observed was not due to chance.

4.3 Results

In this section, we present the assessment of the predictive performance of the LR-ENR and XGBoost models for the memory classification task. This evaluation utilized the feature set described in Table 4.1. Figure 4.1 illustrates the performance of the LR-ENR and XGBoost classifiers based on the Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) metric. The LR-ENR model exhibited no discernible improvement over random guessing, regardless of whether it employed canonical frequency band power features alone or in combination with entropy and fractal features. In contrast, the XGBoost classifier, when leveraging the complete feature set, yielded a mean AUC of 0.610 (±0.081 SD), significantly outperforming the permutation chance baseline (AUC $= 0.500$) as determined by a Mann-Whitney U test ($U = 1247$, $p = 1.6 \times 10^{-4}$). This finding indicates a robust, albeit modest, predictive capability.

The feature importance rankings for this model are illustrated in Figure 4.2. Notably, Phase-Amplitude Coupling (PAC), which reflects the influence of theta phase modulation on the amplitude of slow-gamma oscillations, emerged as the most significant feature. This finding suggests that interaction between theta and slow-gamma rhythms may play a role in hippocampal function concerning processes such as memory consolidation.



(a) (b)

(c) (d)

Figure 4.1: ROC curves demonstrating the performance of LR-ENR and XGBoost classifiers for memory prediction. (a-b) LR-ENR performance using canonical frequency band power features only (AUC = 0.514 ± 0.035) and all engineered features (AUC = 0.515 ± 0.035), respectively. (c-d) XGBoost performance using canonical frequency band power features only (AUC = 0.548 ± 0.049) and all engineered features (AUC = 0.610 ± 0.081), respectively. While LR-ENR shows performance comparable to random chance regardless of feature set, XGBoost achieves statistically significant predictive performance when utilizing all features. Blue lines represent mean ROC curves with shaded regions indicating ±1 standard deviation, while red and green lines show random chance and permutation baselines, respectively.

4.4 Conclusion

Although canonical band powers have been widely utilized in EEG analysis and have demonstrated significant utility across various contexts [193, 203–205], our findings indicate that, for this specific dataset and task, they did not provide sufficient discriminatory power for the LR-ENR and XGBoost classifiers to effectively differentiate

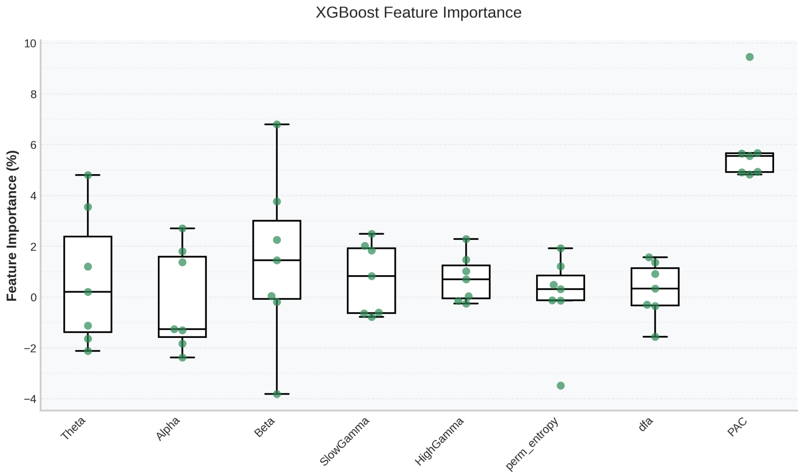


Figure 4.2: Biomarker importance graph for classification between "remembered" and "forgotten" trials using the XGBoost classifier. The PAC demonstrates notably higher importance, suggesting its effectiveness in memory prediction.

between trials where images were remembered versus forgotten. By integrating non-linear features, we sought to enhance the model's ability to detect intricate patterns associated with memory processes, thereby potentially improving the accuracy of our predictive modeling beyond what could be achieved using canonical band powers alone. The emergence of phase-amplitude coupling (PAC) as the most important feature in our analysis is notable. It suggests that stimulation of the amygdala during the encoding of images may enhance memory consolidation by increasing the phase-amplitude coupling between theta and slow-gamma oscillations in the hippocampus. This finding aligns with previous research identifying theta-gamma coupling as a key mechanism in working memory [206–209].

In chapter 5, we will look into the predictive capabilities of EEG foundation models, an approach grounded in self-supervised learning. The objective of this emerging field is generalizable, compressed representations of EEG data—those that go beyond the traditional methods in machine learning.

Chapter 5

Adapting Foundation Models for EEG Data Representation

5.1 Background

The limitations of manual feature extraction in electrophysiological signal analysis, particularly for electrogastrography (EGG), have underscored the need for more sophisticated methodologies, as discussed in Chapters 2 through 4. While classical machine learning techniques, supported by human-engineered features, produced encouraging results in analyzing electroencephalography (EEG) data, these same approaches proved insufficient for EGG data. The complexity, variability, and sparse labeling of EEG datasets require models capable of capturing nuanced relationships and representations that manual methods may fail to fully address [210, 211]. This chapter introduces the potential of foundation models as a transformative tool in addressing these challenges, specifically focusing on their application to EEG data representation, despite not yielding optimal results in all scenarios [212].

Foundation models are large-scale deep learning architectures that establish general-purpose capabilities through pre-training on vast datasets across diverse tasks [213].

These models offer a new paradigm for representation learning in domains characterized by data complexity and limited labeled examples. Despite their promise, foundation models for EEG are still in their infancy. They aim to develop generalizable embeddings that can be adapted for specific downstream analyses [214–216]. However, in the context of EEG data, these models are yet to consistently achieve optimal performance in capturing the intricacies of brain signals, particularly across different individuals, conditions, and experimental setups [216].

5.2 The Importance of Foundation Models for EEG Data

The unique attributes of EEG datasets, such as high dimensionality, temporal dependencies, and low signal-to-noise ratio (SNR) [216], present significant challenges for traditional machine learning approaches. In the case of EEG signals, there is a critical need to capture both local features, such as specific frequency band activities, and global temporal patterns that span entire recording sessions. Traditional manual feature extraction often focuses on predefined metrics, such as power spectral densities or wavelet coefficients [217], which limits the model's ability to adapt to the full range of variability inherent in EEG data. Foundation models address these challenges in two ways:

1. Generalizable Representations: Pre-training on large-scale EEG data allows the foundation models to develop generalized embeddings transferable to vast numbers of downstream tasks with limited labeled data. Further fine-tuning on domain-specific EEG data allows these models to adapt learned features to address specific tasks, such as biomarker discovery—supported perhaps in some contexts by manual feature extraction, though likely benefiting from further automated and adaptive techniques to improve performance.

2. Multi-Scale Feature Learning: By employing complex architectures like transformers or deep convolutional networks, foundation models can extract both low-level temporal features and high-level global patterns, offering a multi-scale perspective that is essential for comprehensive EEG analysis [218].

These advantages make foundation models particularly suitable for handling the sparse-labeling issue inherent in many EEG datasets. Unlike conventional methods that require careful manual labeling, foundation models can be pre-trained on large-scale, unlabeled datasets, using self-supervised learning techniques to learn meaningful representations. This capability reduces the need for extensive manual intervention and may facilitate the discovery of biomarkers with minimal labeled data [219].

5.3 Bridging the Gap: From Manual Extraction to Automated Representation

The transition of foundation models from manually extracted features to automatically learned ones indicates a critical turning point in the analysis of EEG data. Manual methods, most often dependent on domain knowledge in defining the features, are prone to potential biases and insufficient capability to capture the full complexity of the underlying physiological processes [220]. By contrast, foundation models provide an increase in predictive power through data-driven feature discovery, especially in situations where there is little labeled data and subtle interdependent patterns would otherwise escape traditional methods. On the flip side, it lacks an explainability tool to uncover what features really contribute to predictions.

Chapters 2 through 4 highlighted the strengths and limitations of classical machine learning with manual feature extraction. Although effective for seizure detection from EEG, these methods fell short when applied to EGG or SEEG, pointing towards a need for more scalable and robust approaches. FMs bridge this gap by allowing features to be learned directly from scarce annotated raw data, eliminating the need for human-engineered features and enabling the model to discover latent patterns that may correlate with physiological or pathological conditions.

In the following sections, we will explore two specific foundation models adapted for EEG data representation, illustrating their architectures, pre-training strategies, and the benefits they bring to biomarker discovery in sparse-labeled contexts.

5.4 BERT-inspired Neural Data Representations

This framework introduced in [221] represents a pioneering attempt at learning from massive amounts of unlabeled EEG data using transformer-based architectures and self-supervised learning techniques developed mainly for natural language processing. The following is a detailed explanation of this model and how we pre-trained and fine-tuned it for a classification task.

Model Training: The model is trained on the pretext task using a loss function that encourages the model to make correct predictions on the pretext task [222]. For example, if the pretext task is to predict the rotation of an image, and the model's prediction is denoted by $\hat{y}$, the true rotation by y, and the parameters of the model by θ, the training process could be formulated as minimizing the following loss function:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} L(y_i, \hat{y}_i)$$

where L is a suitable loss function (like cross-entropy loss for classification tasks), N is the number of training examples, and the sum is over all training examples.

Representation Learning: The goal of the training process is to learn a function f parameterized by θ that maps the input data x to a useful representation [222]. This function f is typically the part of a model up to the last layer or two. The learned representation for an input x can be denoted by $f_\theta(x)$. The hope is that this representation captures the important features of the data that are relevant for the pretext task and also for other related tasks.

Fine-tuning: The learned representation $f_\theta(x)$ can be used as a starting point for training a model on a downstream task. This involves adding a few new layers to the model, denoted by $g_\phi(.)$, and training these new layers on the downstream task while keeping the parameters θ fixed or allowing them to change slowly. The fine-tuning process can be formulated as minimizing the following loss function:

$$\min_\phi \frac{1}{M} \sum_{j=1}^{M} L'(y'_j, g_\phi(f_\theta(x'_j)))$$

where L' is a suitable loss function for the downstream task, M is the number of examples for the downstream task, y'_j and x'_j are the labels and inputs for the downstream task, and the sum is over all examples for the downstream task.

5.4.1 Application to EEG

A significant obstacle in the implementation of SSL for EEG representation learning lies in the generation of disparate yet semantically coherent views of EEG samples. As evidenced in existing literature [41, 223], the efficacy of generating meaningful views of EEG samples is contingent upon the specific application at hand. In other words, the success of various EEG augmentation techniques is intrinsically linked to the objectives of the downstream tasks they serve. This observation serves as a compelling rationale for transitioning from contrastive approaches to methodologies that are either predictive or reconstructive in nature. Drawing on the methodology presented in [221], we aim to adopt a strategy from the realms of Language Modeling (LM) and speech recognition. Our goal is to explore the potential application of this strategy to EEG data representation. In Masked Language Modeling (MLM), the model learns to reconstruct language token(s) given the surrounding context [224]. The task often used in MLM is as follows:

Consider a sequence composed of N tokens, denoted as $t_1, ..., t_N$. Also, take into account a subset of these token indexes, represented as I_m. For every token index i that is a member of I_m, a mask, denoted as M, is applied to the tokens. This mask serves to modify or conceal the tokens in a certain manner so that:

$$q_i = \begin{cases} M & \text{if } i \in I_m \\ t_i & \text{otherwise} \end{cases} \quad \forall i \in N$$

In the context of MLM, a sequence model, akin to a transformer [66], undertakes the task of reconstructing the original arrangement of tokens from a sequence that has been strategically masked.

Unfortunately, the intertwined nature of adjacent EEG samples does not support masking and recovering individual points. This would likely result in a model learning an interpolation method rather than EEG features. In simpler terms, the smoothness of the data poses a challenge in deriving meaningful features solely by recovering missing individual samples, as previously discussed in similar SSL studies involving speech [225]. On the other hand, if we were to mask a continuous range of tokens, we could avoid solely focusing on interpolating missing samples. However, reconstructing time-series data proves difficult due to various challenges, one being that it is complex to capture the amount of error over time within consecutive sequences [225, 226]. The commonly used loss functions for such reconstruction, such as mean squared error or mean absolute error, incorrectly assume independence in the error between elements in the series [227]. One way to avoid degradation to interpolation and issues with time series reconstruction is to employ contrastive predictive coding (CPC) as the sequence learning task [228]. CPC facilitates the learning process of acquiring a robust feature representation and comprehending the sequential structure of data. It accomplishes this by accurately tracing the evolution of these representations, all under the guidance of a singular loss function [228].

Wav2Vec 2.0 [229] is a self-supervised model developed for speech recognition tasks that employ parts of CPC and MLM. What makes Wav2Vec 2.0 interesting for EEG representation is that the model begins by processing the raw audio input and extracting relevant features, similar to how our auditory system focuses on important aspects of sound to understand speech. It then employs SSL, predicting parts of the audio based on other parts, akin to how a child learns language by predicting what comes next in a conversation. After this initial training, the model can be fine-tuned on a smaller dataset where the audio is paired with the correct transcription, allowing it to adjust its predictions to match the correct transcriptions [229].

With the fine-tuning capabilities of Wav2Vec 2.0 demonstrated through its refined speech recognition following initial training, we are keen to explore the analogous possibilities within EEG data. Transitioning from the domain of auditory processing to that of EEG signals, we will delineate the specificities of this approach. This will encompass the dataset employed for initial pre-training, the pre-processing methodologies applied, the detailed procedure of pre-training, and the subsequent outcomes when fine-tuning for a downstream EEG data that was not part of the pre-training data. The intent is to adapt the underpinnings of Wav2Vec 2.0's self-supervised learning framework to extract and learn from the complexities of EEG data, with the aspiration of achieving enhanced task-specific performance when the size of data is small. Illustrating this approach, Figure 5.1 depicts the overall schematic of the pre-train and fine-tune procedures employed in our study.

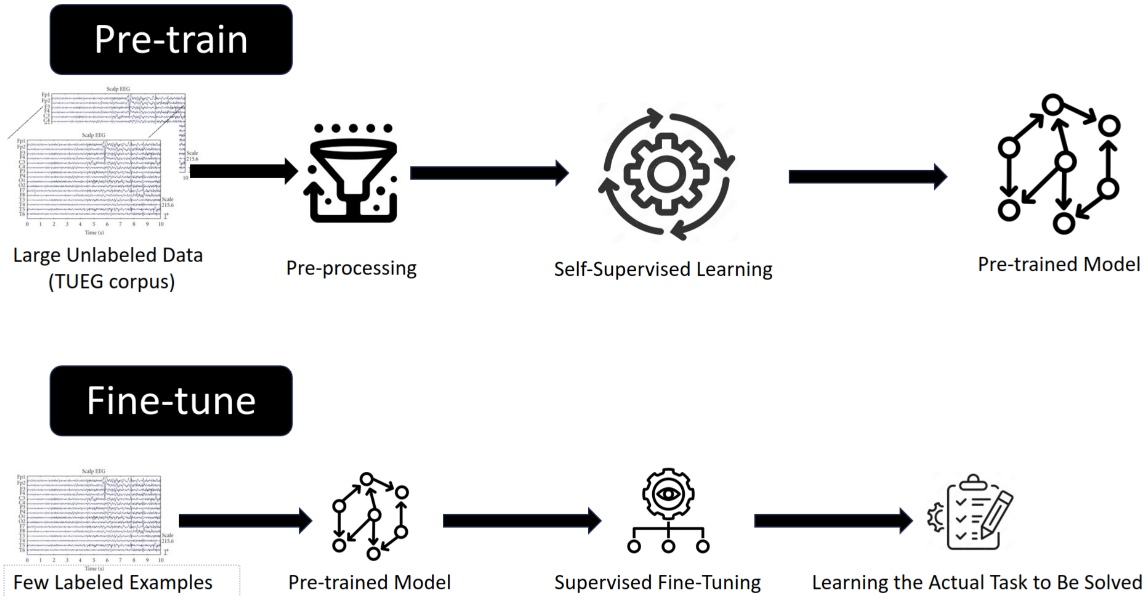


Figure 5.1: Schematic Overview of the Pre-training and Fine-tuning Procedures Applied to EEG Data. This diagram illustrates the initial large-scale self-supervised learning using the TUEG corpus, followed by the supervised fine-tuning stage employing a smaller set of labeled EEG examples.

5.4.2 Pre-training Dataset

The efficacy of SSL is closely linked to the variety and extent of the dataset used during the pre-training phase [230]. To the best of our knowledge, the Temple University EEG corpus (TUEG) [231] stands as the most expansive EEG dataset currently available to the public. In our research, we have employed the latest release of this corpus, version 2.0.0, which comprises EEG data spanning from 2002 through 2017. This dataset contains a vast collection of EEG recordings, with over 26,000 sessions and close to 15,000 patients represented. This dataset is continually updated and expanded, making it a rich resource for research in various fields, including neuroscience, and machine learning. The demographic details pertinent to this corpus are presented in Table 5.1.

Table 5.1: TUEG corpus description (version 2.0.0)

Characteristic	**Detail**
Number of Subjects	14987
Gender	51% Female
Age	1-month to 90 yrs old
Number of Sessions	26846
Number of Recordings	69652
Total Duration	27062 hours
Sampling Frequency	Between 250Hz and 1024Hz
Electrode Configuration	Mostly (95%) 10/20 system

5.4.3 Pre-processing

Pre-processing is designed to be minimal. Most of the recordings adhered to the conventional 10/20 electrode placement standard, thus, in accordance with the guidelines outlined in [232], a set of 19 channels was chosen to clearly represent the 10/20 system. In instances where the channel configuration of a recording did not align with the chosen set of 19 channels, we eliminated any additional channels and compensated for absent ones by filling them with zeroes. All recordings were scaled to [-1, 1] and to compensate for the lost amplitude information we added a 20th channel to each recording that we filled with a constant value derived from the equation below:

$$A = \frac{\max(s_i) - \min(s_i)}{\max(S_{\text{ds}}) - \min(S_{\text{ds}})}$$

where s_i is a subset of the total set of samples S_{ds}, referring to a single recording. The reference sampling frequency was standardized to 256Hz, to which all recordings were resampled for consistency [221].

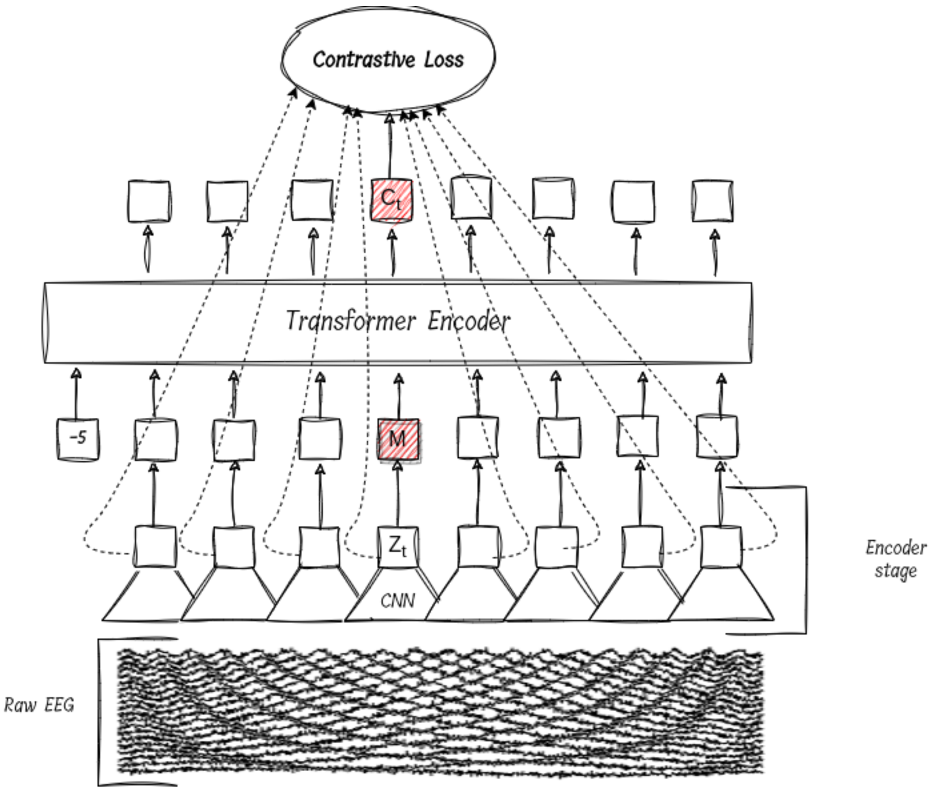


Figure 5.2: The overall architecture for the pre-training phase. The encoder stage compresses the raw EEG to a sequence of latent vectors. Then transformer encoder attempts to produce c_t to be more similar to the masked Z_t than it is to a random sampling over the latent vectors Z_i, where $i \neq t$.

5.4.4 Pre-training Procedure

The architecture of the pre-training phase is schematically illustrated in Figure 5.2. The architecture of the pre-training algorithm draws parallels to Wav2Vec 2.0, with two significant modifications [221]: 1- multi-channel inputs as opposed to Wav2Vec 2.0's single-channel approach, 2- removing the quantization module found in Wav2Vec 2.0. In Wav2vec 2.0 and during the unsupervised pre-training, the quantization module, specifically a Gumbel softmax operation or a product quantization mechanism, discretizes the continuous audio features extracted from the raw audio into a finite set of quantized representations, often referred to as "codebooks" or "quantized tokens" [229]. The reason for omitting the quantization module here is the lack of "codebooks" for EEG data. The Pre-training process unfolds in the following manner [221, 229]: Initially, a convolutional layer generates a series of compressed representations from the multi-dimensional input. In order to leverage the entire sequence for a singular task, such as classification, an initial token is appended to the start of the sequence. This token is represented by a vector filled with the value -5. Subsequent segments of this sequence are then masked. This altered sequence is then fed into the transformer layer [221, 233], which aims to produce outputs closely resembling the original, unmasked input at a specific position t. The self-supervised training involves a contrastive loss function where the model is trained to match the true representation of the masked segment against a set of incorrect ones (collected from other segments of the same sequence)[234]. The outputs of the transformer encoder, and outputs from the feature extractor (convolutional stage), serve as inputs to the contrastive loss below [221]:

$$\mathcal{L} = -\log \frac{\exp(\text{sim}(c_t, b_j))/\kappa}{\sum_{b_i \in B_D} \exp(\text{sim}(c_t, b_i))/\kappa}$$

where c_t denotes the output of the transformer encoder at position t, b_i represents the (original/un-masked) encoded EEG vector at some offset i, and B_D is a collection of 20 uniformly chosen distractors/negatives from the same sequence, in addition to b_t. sim stands for cosine similarity function where $sim(x, y) = \frac{x^T y}{\|x\|\|y\|}$ is utilized to evaluate the degree of similarity between vectors, and its sensitivity is modulated by a temperature coefficient κ, which is assigned a value of 0.1 [230]. This loss function is designed to refine the output of the transformer at position t to closely align with the encoded representation at t, notwithstanding the fact that this input to the transformer is masked. Consequently, the transformer is tasked with learning a sufficiently comprehensive model of EEG within its latent space, such that the complete latent sequence can effectively characterize position t. Throughout the training process, which included both pre-training and fine-tuning stages with downstream data, we utilized the Adam optimizer, with the weight decay parameter set at 0.01 [221]. In addition, we employed a cosine learning rate decay strategy, complemented by a linear warm-up phase that spanned 5% of the total training steps for pre-training and 10% for fine-tuning [221]. Some of the used hyperparameters are summarized in Table 5.2.

Table 5.2: Hyperparameters for pre-training BENDR and its downstream tasks

Dataset	**Batch Size**	**Epochs**	**Learning Rate**
TUEG [231]	64	50	5×10^{-5}
MDD [235]	16	10	5×10^{-5}

5.4.5 Fine-tunning Procedure

Fine-tunning was done by removing the transformer module and adding a linear layer on top of the convolutional stage initialized with weights from the pre-training [221]. The architecture of the fine-training phase is schematically illustrated in Figure 5.3.

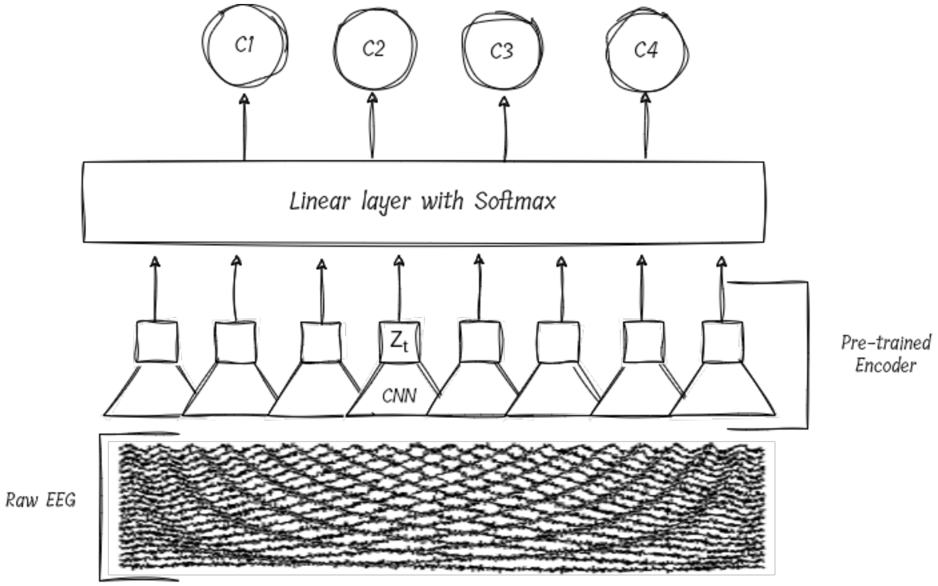


Figure 5.3: The overall architecture for fine-tuning phase. Initially, the convolutional encoder weights are established via pre-training. Subsequently, these weights are fine-tuned alongside the linear layer weights through supervised learning for each downstream task (here we assumed the task has 4 distinct classes).

5.4.6 Evaluation Procedure

To assess the viability of foundation models in EEG data analysis, we conducted an evaluation of the BENDR model, which we pre-trained on the TUEG corpus for 50 epochs. Subsequently, the model underwent fine-tuning for a classification task involving Major Depressive Disorder (MDD), utilizing data that was entirely distinct from the pre-training dataset [235]. In addition, to provide a comparison of model performance across different tasks, we incorporated the results from [221] for two additional classification tasks: Motor Imagery (MI)[236] and Sleep Staging (SS)[237], as outlined in Table 5.3.

Table 5.3: Performances of downstream tasks

Dataset	**Paradigm**	**SR (Hz)**	**Channel**	**Subjects**	**Targets**	**Metric**	**Result**
BCI IV 2a [236]	Motor Imagery	250	22	9	4	ACC	0.40
SSC [237]	Sleep Staging	100	2	78	5	BAC	0.65
MDD [235]	Major Depressive Disorder	256	19	52	2	ACC	0.86

It is worth noting that our validation approach diverges from past research that concentrated on within-subject classification [238–240]; instead, we implemented a one-subject-leave-out cross-validation method. This technique presents a more formidable challenge because it must contend with the variability that exists between different subjects. The binary classification of MDD dataset [235] showed promising performance in detecting Major Depressive Disorder with an accuracy of 0.86 among 52 subjects (26 healthy controls). Taken together, these results provide evidence of the adaptability of a pre-trained foundation model to different task-specific environments and so suggest that foundation models may be useful for memory classification tasks and with data described in Chapter 4.

5.5 Adapting a Foundation Model for SEEG Data Representation and Analysis

Although BENDR showed promising results for adapting to unseen EEG data and tasks, it suffers from a major restriction: the channel number and configuration. It only works with 19 channels of 10/20 standard EEG signal acquisition system. It would be impractical to adapt this model to the SEEG data from chapter 4 as each subject has a different number of implanted electrodes and in different brain regions.

Large Brain Model for EEG (LaBraM) introduced in [241] is a foundation model designed to overcome the limitations of other FMs such as BENDR. LaBraM's architecture leverages Transformer models, inspired by the success of LLMs such as BERT [66] and GPT [242]. These models apply the masked modeling paradigm and a neural tokenizer to learn from raw EEG signals, offering a more general and flexible representation of EEG data. Unlike most FMs that are constrained by the diverse configurations of EEG recording devices [215, 221], LaBraM introduces a patching mechanism to efficiently utilize diverse EEG data from approximately 20 datasets.

5.5.1 Pre-training Datasets

To facilitate a comparison between LaBraM [241] and BENDR [221] on the major depressive disorder (MDD) classification task (refer to 5.4.6), we selected the entire TUEG corpus [231] for pre-training LaBraM.

Moreover, to assess LaBraM's performance on the memory classification task presented in Chapter 4, we retrained the model using SEEG data from the "Delayed Free Recall of Word Lists" (DFR) dataset [243]. This retraining marks a pioneering application of SEEG data in the pre-training of a foundation model specifically aimed at memory prediction tasks. Table 5.4 provides a brief description of the DFR dataset.

Table 5.4: DFR data description

SR (Hz)	#Channels (min, max, median)	#Subjects	Recording (hours)	Age
1000	26, 224, 120	293 (158 Male)	371	18-65

5.5.2 Pre-processing

we implemented the following pre-processing steps for TUEG and DFR datasets: (1) use of a zero-phase band-pass filter at 0.1 Hz and 95 Hz, (2) application of a zero-phase notch filter to eliminate line noise at 60 Hz, and (3) down-sampling of the recordings to 200 Hz.

5.5.3 EEG Patching

EEG datasets often vary in terms of channel numbers, sampling rates, and experimental paradigms, which makes it difficult to develop models that utilize them in one setting. LaBraM's approach to segmenting raw EEG signals into channel patches, inspired by image patch embeddings used in computer vision [244], involves dividing the EEG signals into fixed-length segments for each channel. These segments, or 'channel patches', allow the model to manage varying numbers of channels and different data lengths effectively.

Formally, given EEG signals represented as a matrix $X \in \mathbb{R}^{C \times T}$, where C is the number of EEG electrodes (channels) and T is the total number of time samples, the channel patching process involves segmenting X into non-overlapping segments. Each EEG channel is divided into patches of a fixed length t, resulting in $\lfloor \frac{T}{t} \rfloor + 1$ patches per channel. Thus, for each channel c_i, the corresponding patches $x_{c_i,k} \in \mathbb{R}^w$ are formed, where w represents the window size for each patch, and k ranges over all possible patches [241].

5.5.4 Temporal Encoding

The temporal encoder in the LaBraM model is designed to capture temporal features from the segmented EEG patches. The key idea is that each EEG patch $x_{c,k}$ is passed through a temporal convolution block to extract local temporal patterns. This block consists of a 1-D convolution layer, followed by a group normalization layer, and then a GELU activation function [241]. The components are described below:

- **1D Convolution Layer**: The convolutional layer extracts temporal features from the patch. The 1D convolution for a given patch can be represented as:

  $$y_{c,k} = \mathrm{Conv1D}(x_{c,k}; W, b)$$

  where:

  - $y_{c,k} \in \mathbb{R}^{t'}$ is the output of the convolutional layer (which may have a different length t', depending on the kernel size and stride).
  - $W \in \mathbb{R}^{F \times t'}$ represents the convolution filter weights, where F is the number of filters used.
  - b represents the bias term for each filter.

  The convolution operation can be mathematically defined as:

  $$y_{c,k}[n] = \sum_{i=1}^{t'} W[i] x_{c,k}[n-i] + b, \quad \forall n \in [1, t']$$

  Here, $W[i]$ represents the weight of the i-th position of the filter, and $x_{c,k}[n-i]$ represents the signal value of the $(n-i)$-th position.

- **Group Normalization [245]**: The output of the convolution layer, $y_{c,k}$, is passed through a group normalization layer for stable training. The group normalization operation is defined as:

$$\mathrm{GN}(y_{c,k}) = \frac{y_{c,k} - \mu_g}{\sqrt{\sigma_g^2 + \epsilon}}$$

where:

- μ_g and σ_g are the mean and standard deviation computed for a specific group of channels.
- ϵ is a small constant added to prevent division by zero.

- **GELU Activation Function [246]**: Finally, a Gaussian Error Linear Unit (GELU) activation is applied to introduce non-linearity into the temporal representation:

$$z_{c,k} = \mathrm{GELU}(\mathrm{GN}(y_{c,k})) = 0.5 y_{c,k} \left(1 + \mathrm{erf}\left(\frac{y_{c,k}}{\sqrt{2}}\right)\right)$$

where $\mathrm{erf}(\cdot)$ represents the error function, which smooths the activation, unlike the standard ReLU.

After the temporal convolution block, each EEG patch is transformed into an embedding vector $e_{c,k} \in \mathbb{R}^d$, where d is the embedding dimension to be used as the input to the Transformer encoder. Thus:

$$e_{c,k} = f_{\text{Temporal Block}}(x_{c,k}) \in \mathbb{R}^d$$

Finally, the embeddings from all patches and channels are concatenated to form the input sequence for the Transformer encoder [233] to produce the output embeddings.

5.5.5 Neural Tokenizer

Similar to BENDR's framework, LaBraM implements the masking and prediction paradigms inherent to Masked Language Modeling (MLM). However, LaBraM distinguishes itself through the incorporation of a quantization mechanism that transforms temporal EEG representations into discrete embeddings, which are systematically derived from a predetermined codebook architecture [241]. This approach allows EEG data to be represented as discrete tokens inspired by the Vector Quantized Variational Autoencoder (VQ-VAE) technique, commonly used in image representation learning [247]. The methodology begins by defining a neural codebook

$$V = \{v_i \mid i = 1, \ldots, K\} \in \mathbb{R}^{K \times D},$$

where K represents the number of discrete neural embeddings and D is the dimension of each embedding. The EEG signal sample x is segmented into patches, denoted as

$$p = \{p_i \mid i = 1, \ldots, N\},$$

with N representing the total number of patches, which is computed from the number of channels and segment length. Each patch p_i is quantized by finding the closest match within the neural codebook V, achieved by minimizing the Euclidean distance between the ℓ_2-normalized patch and embedding, as formulated in:

$$z_i = \arg\min_j \|\ell_2(p_i) - \ell_2(v_j)\|^2,$$

where ℓ_2 indicates ℓ_2 normalization. This helps in maximizing codebook utilization, effectively distributing the patch representations among all available neural embeddings.

Fourier spectrum prediction is used to train the neural tokenizer to capture the frequency components of EEG signals [241]. Due to the noisy, nonstationary nature of EEG data, reconstructing raw signals directly is impractical. Instead, the approach focuses on reconstructing the Fourier spectrum of the EEG, which reveals underlying brain activities. Specifically, the amplitude A_m and phase ϕ_m are calculated from the Discrete Fourier Transform (DFT) of an EEG patch ($\tilde{x}_{m,c,k}$), allowing training to focus on reconstructing these features rather than the raw signal:

$$A_m = \sqrt{\mathrm{Re}(\tilde{x}_{m,c,k})^2 + \mathrm{Im}(\tilde{x}_{m,c,k})^2},$$

$$\phi_m = \arctan\left(\frac{\mathrm{Im}(\tilde{x}_{m,c,k})}{\mathrm{Re}(\tilde{x}_{m,c,k})}\right).$$

The neural decoder then regresses the predicted spectrum, optimizing the total loss using Mean Squared Error (MSE) for both amplitude and phase [241]. The final loss function includes a codebook loss term to ensure consistency between the embeddings and their corresponding neural tokens. This total training loss L_T is defined as follows [241]:

$$L_T = \sum_{x\in\mathbb{D}} \sum_{i=1}^{N} \left(\|o_{A_i} - A_i\|_2^2 + \|o_{\phi_i} - \phi_i\|_2^2 + \|\mathrm{sg}(\ell_2(p_i)) - \ell_2(v_{z_i})\|_2^2 + \|\ell_2(p_i) - \mathrm{sg}(\ell_2(v_{z_i}))\|_2^2\right),$$

where:

- $\mathbb{D}$ represents all EEG data samples.
- o_{A_i} and o_{ϕ_i} are the predicted amplitude and phase values for the i-th patch.
- The notation $\mathrm{sg}(\cdot)$ denotes the stop-gradient operation, which acts as an identity in the forward pass but prevents gradients from flowing during backpropagation.
- The term $\|\mathrm{sg}(\ell_2(p_i)) - \ell_2(v_{z_i})\|_2^2$ encourages the neural codebook embeddings to match the corresponding tokenized representations, thereby promoting stable embedding updates.

- Conversely, $\|\ell_2(p_i) - \text{sg}(\ell_2(v_{z_i}))\|_2^2$ ensures that the representation for each patch p_i remains close to the nearest codebook embedding, effectively reducing quantization error.

The codebook loss terms play a significant role in balancing between the encoder and codebook representations, thereby ensuring that the neural tokenizer can effectively learn to discretize and reconstruct meaningful patterns from the noisy EEG data [241].

5.5.6 Pre-training Procedure

The process begins by segmenting a given EEG sample x into patch embeddings [241]

$$e = \{e_i \mid i = 1, \ldots, N\}.$$

A random mask M is generated, with each element $m_i \in \{0, 1\}$ representing the mask status for the i-th patch, and a proportion r of m being equal to 1. The masked patches of x are then replaced by a learnable mask token, resulting in corrupted EEG patches e^M, which are then processed using temporal and spatial embeddings, as defined in 5.5.4. These embeddings are subsequently fed into a Transformer encoder, yielding hidden vectors

$$h = \{h_i \mid i = 1, \ldots, N\},$$

which are ultimately used to predict the corresponding neural tokens via a linear classifier:

$$p(v' \mid e^M) = \text{softmax}(\text{Linear}(h)).$$

The training loss for the masked EEG modeling is defined as:

$$L_M = -\sum_{x \in \mathbb{D}} \sum_{m_i = 1} \log p(v_i \mid e^M),$$

where $\mathbb{D}$ represents the entire dataset, and $m_i = 1$ indicates the patches to be reconstructed [241].

To further enhance training efficiency, a symmetric masking strategy is employed. This involves calculating the inverse of the mask M, denoted as $\tilde{M}$, and performing a second masked EEG modeling step using $\tilde{M}$ [241]. This generates an additional prediction loss, L_{symM}. The motivation for symmetric masking is twofold: it reuses the same discrete representations while reducing computational overhead and it increases data diversity, which improves downstream performance. The overall pre-training objective for LaBraM is given by [241]:

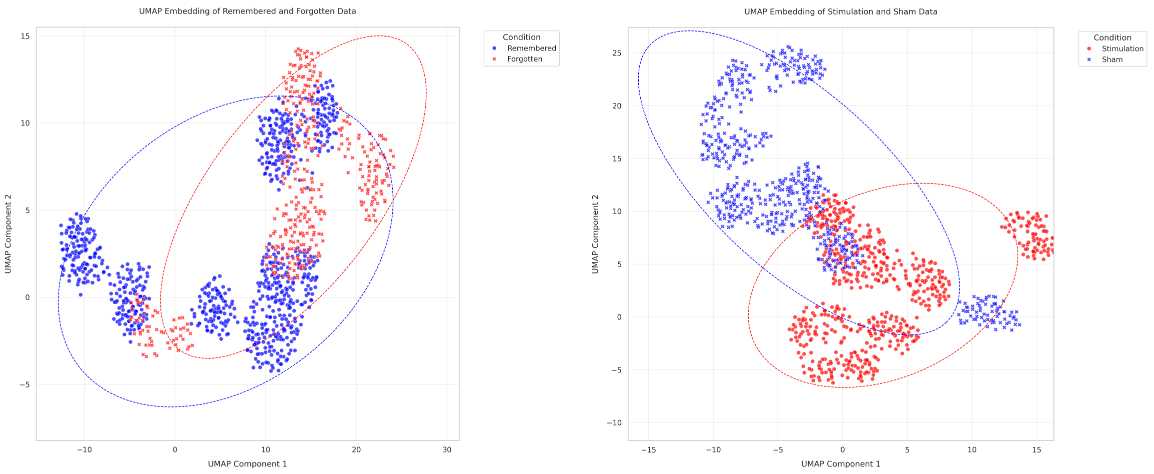
$$L = L_M + L_{symM}.$$

5.5.7 Fine-tuning Procedure

For fine-tuning the pre-trained model, a final classification layer is added to the architecture and is trained, while the pre-trained Transformer layers remain frozen [241]. In linear probing, the weights learned during pre-training are retained without any updates, and only the linear classifier head is adapted to address the specific downstream task. To execute linear probing, the pre-trained Transformer blocks are first employed as feature extractors, which means that input EEG data passes through these layers to generate fixed representations for each input.

Importantly, the parameters of these Transformer layers remain frozen, distinguishing this approach from full fine-tuning, where all model parameters are subject to updates. By keeping these layers unaltered, it becomes possible to evaluate how well the features learned during pre-training generalize without further adjustments.

Only the final classification layer, which maps the extracted features to output classes, undergoes training. The training involves minimizing a task-specific loss function, such as binary cross-entropy for classification, via gradient descent. In this



(a) (b)

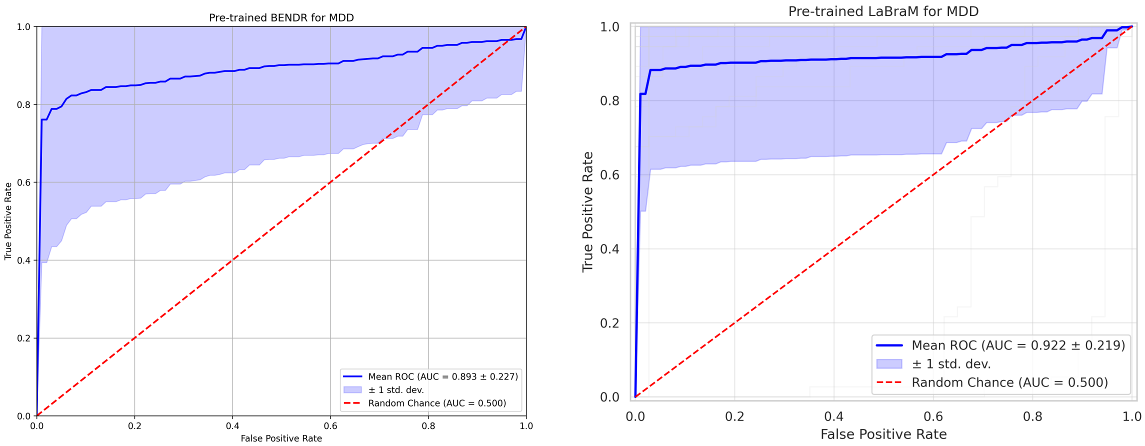
Figure 5.4: UMAP embeddings of EEG features extracted by the pre-trained model, illustrating neural representations associated with memory and stimulation conditions, with 95% confidence ellipses shown as dotted lines for each condition. (a) UMAP projection of EEG features derived from a 2-second window preceding stimulation, where blue data points and ellipse represent instances of remembered trials and red data points and ellipse indicate forgotten trials. (b) UMAP projection of EEG features from a 2-second window following stimulation, with blue data points and ellipse corresponding to sham or non-stimulated instances and red data points and ellipse representing stimulated instances.

case, only the weights of the classifier are updated. Notably, the weight decay is set to zero to avoid unnecessary reduction of the classifier's weights. This approach allows the classifier to learn the best weights for classification while effectively controlling model complexity.

The UMAP projection [248] of EEG features extracted by the pre-trained LaBraM for a 2-second segment of EEG signals prior to stimulation is illustrated in Figure 5.4a. In this figure, blue markers denote remembered instances, while red markers indicate forgotten instances. Figure 5.4a displays the corresponding EEG features extracted from a 2-second segment of signals following stimulation, with blue markers representing stimulated instances and red markers denoting sham or non-stimulated instances. Figure 5.4 demonstrates a more pronounced separation between the stimulation (red) and sham (blue) conditions compared to the memory conditions. The clusters exhibit clearer boundaries and reduced overlap between the confidence ellipses, indicating distinct neural patterns associated with stimulation and sham conditions.

5.5.8 Linear Probing vs. Latent Space Classification

As demonstrated in 5.4.6, the foundation model BENDR, pre-trained on the TUEG corpus [231], produced promising classification outcomes for the major depressive disorder (MDD) dataset [235]. Additionally, we pre-trained LaBraM on the TUEG corpus and fine-tuned it on the MDD dataset. Figure 5.5 presents the performance of the fine-tuned foundation models in classifying the MDD dataset into healthy and MDD subject groups, demonstrating the pre-trained model's adaptability to a smaller dataset not included in the initial pre-training phase.

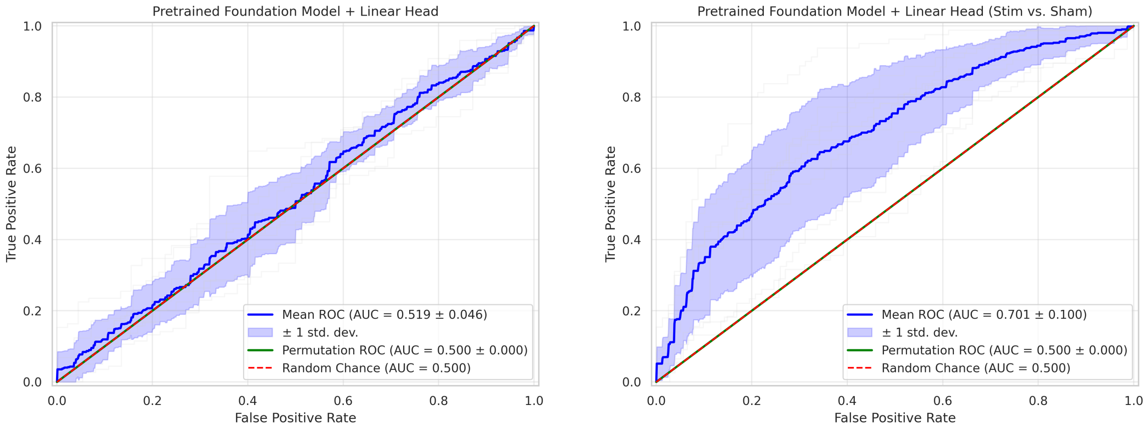


(a) (b)

Figure 5.5: ROC curves comparing the performance of pre-trained foundation models on MDD classification. (a) BENDR model achieving a mean AUC of 0.893 ± 0.227, and (b) LaBraM model achieving a mean AUC of 0.922 ± 0.219. The blue lines represent mean ROC curves, blue shaded areas indicate ±1 standard deviation, and red dashed lines show random chance performance (AUC = 0.500). Both models demonstrate strong discriminative ability between healthy and MDD subjects after fine-tuning on the MDD dataset.

To assess LaBraM's predictive capability for SEEG data of Chapter 4, we pre-trained it using the TUEG and DFR datasets and fine-tuned it on our SEEG dataset using the linear probing as suggested in [241]. Figure 5.6a illustrates that linear probing was ineffective in distinguishing between remembered and forgotten states. However, it demonstrated moderate predictive capability for classifying stimulation versus sham conditions, achieving a mean AUC of 0.701 (SD ± 0.101).

In addition to linear probing, we employed the latent space representations gener-



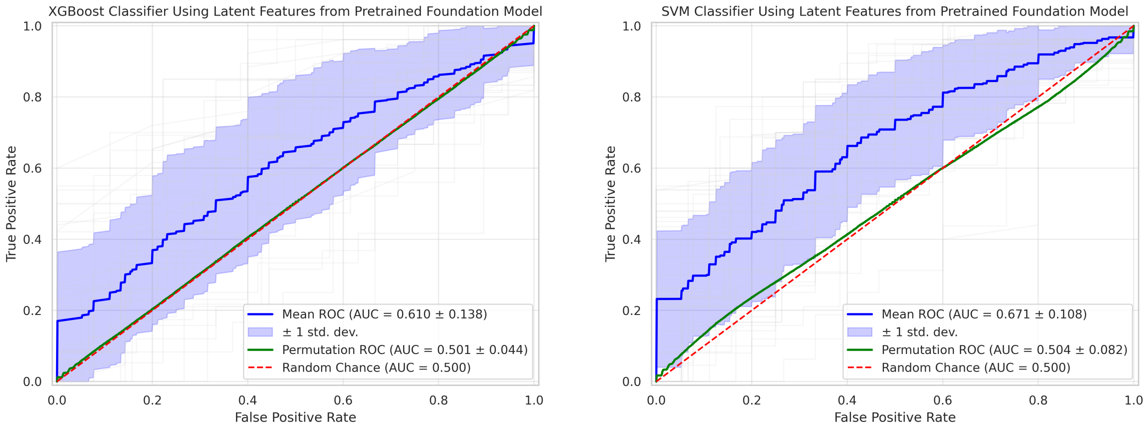
(a) (b)

Figure 5.6: Evaluation of the predictive performance of the pre-trained foundation model, fine-tuned by linear probing. (a) Performance of the pre-trained LaBraM with a linear head added to the representation layer, showing no significant predictive improvement (Mean ROC AUC = 0.519 ± 0.046) for *memory* classification task. (b) Performance of the same model for classification of *stimulation vs. sham*. Blue lines represent mean ROC curves with shaded regions indicating ±1 standard deviation, while red and green lines show random chance and permutation baselines, respectively.

ated by the pre-trained LaBraM model to train an XGBoost classifier, following the approach reported in Chapter 4. Given that XGBoost is more effective on tabular data [249], similar to the extracted features described in Chapter 4, we also experimented with an SVM classifier with a non-linear kernel, which in some cases can be better suited for high-dimensional representations [250]. Indeed, as shown in Figure 5.7, the SVM with a non-linear kernel achieved higher performance than XGBoost on the memory classification task. Notably, both classifiers outperformed linear probing in this context.

5.5.9 Enhancing Linear Probing

Revisiting Figure 5.4, it is apparent that distinct clusters emerge for each neural state, corresponding to the eight individual subjects in the dataset. Figure 5.8 illustrates the 2D UMAP projection of the latent space, where data points for each subject are uniquely color-coded. The clear separation in these UMAP projections indicates substantial inter-subject variability, which likely undermines the predictive



(a) (b)

Figure 5.7: Comparison of XGBoost and SVM classifiers on memory classification using latent features from the pre-trained foundation model. (a) ROC curve for the XGBoost classifier, achieving a mean AUC of 0.610 ± 0.138. (b) ROC curve for the SVM classifier with a non-linear kernel, yielding a higher mean AUC of 0.671 ± 0.108, outperforming XGBoost. Both classifiers performed above chance level (AUC = 0.500), but the SVM model demonstrated an advantage over XGBoost in distinguishing memory states.

performance of classifiers trained on this dataset. This variability poses a significant challenge, affecting both linear probing methods and classifiers trained directly on the latent space, by limiting their generalization capacity across subjects.

What follows next is the description of our proposed ensemble classifier that has domain (subject) adaptation capability, which fine-tunes subject-specific models in order to mitigate the challenge posed by inter-subject variability and limited sample size. The method is summarized in Algorithm 1.

Base Classifier

The core of the classifier is a streamlined, single-hidden-layer neural network architecture comprising 256 nodes, designed specifically for binary classification tasks. To enhance stability across the diverse subject-specific data distributions, batch normalization (BatchNorm1d) is applied. Additionally, an adjustable dropout layer is used to prevent overfitting, especially given the limited dataset size. The final output layer produces a single logit for binary classification.

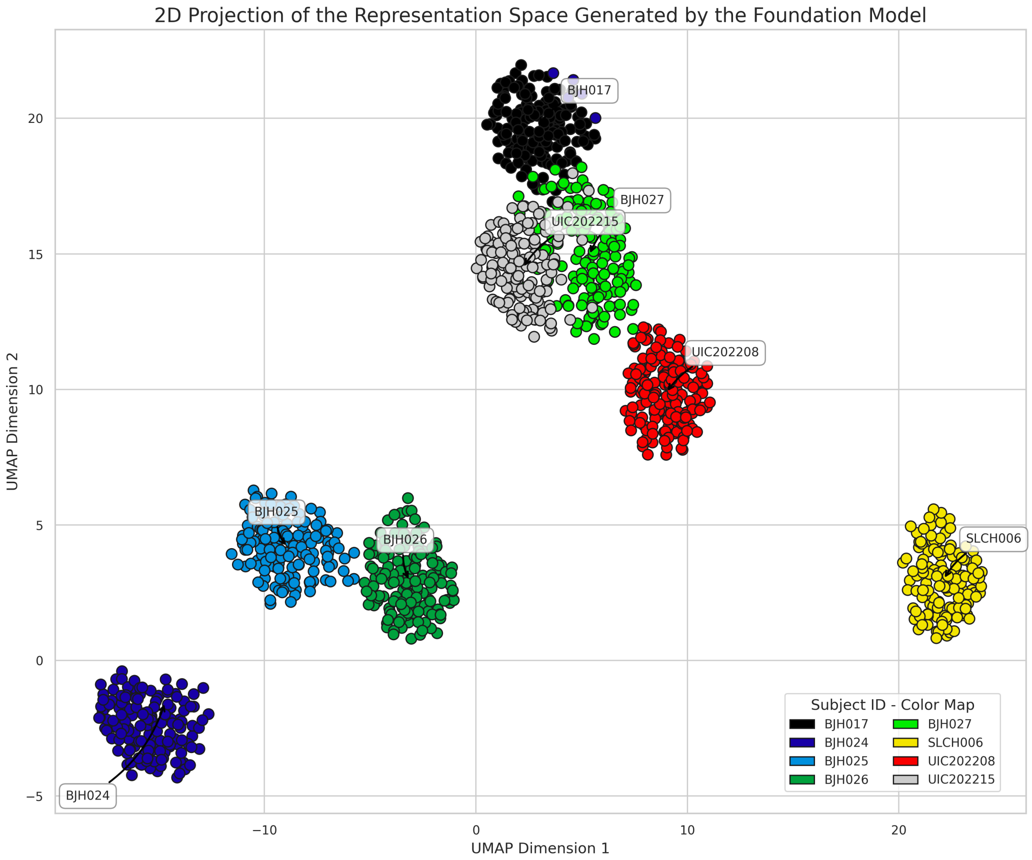


Figure 5.8: 2D UMAP Projection of the Representation Space Generated by the pre-trained foundation model. Each point represents data from a specific subject, with color coding used to distinguish between the eight individual subjects. The distinct clustering suggests significant inter-subject variability.

An ensemble of independent classifiers is utilized to minimize variance and enhance predictive stability. To foster model diversity, each classifier in the ensemble is trained on a different bootstrap sample. The ensemble prediction is then generated by averaging the outputs from all the individual models, leading to a more robust and stable final prediction.

Procedure

The algorithm employs a Leave-One-Subject-Out cross-validation strategy, which allows systematic evaluation and adaptation of models to each subject in turn. The procedure consists of four key stages:

1. **Leave-One-Subject-Out Cross-Validation:**

Step 1.1 Select one subject as the validation subject, leaving the rest as the training set.

Step 1.2 Preprocess the data for each subject:

  Step 1.2a Apply subject-specific normalization.

  Step 1.2b Select optimal features based on performance metrics.

2. **Training Ensemble Models:**

Step 2.1 Generate multiple classifier instances to create the ensemble.

Step 2.2 Train each classifier independently:

  Step 2.2a Train each on a bootstrap sample of the training data.

  Step 2.2b Apply early stopping during validation to prevent overfitting.

3. **Domain Adaptation:**

Step 3.1 Fine-tune each classifier in the ensemble using the data from the validation subject to adapt the models to individual-specific data.

Step 3.2 Adjust learning parameters, such as reducing the learning rate, to prevent drastic modifications to the learned features while adapting to new data.

4. **Evaluation:**

Step 4.1 Evaluate and record metrics to determine the best-performing model.

Feature Selection

Managing the high-dimensional feature space ($d = 200$) and relatively sparse sample size ($n \approx 1000$) requires an effective feature selection strategy to avoid overfitting. We use mutual information scores to measure the correlation between each feature and the target labels. Subsets of features ranging from 20 to 200 are evaluated iteratively, and the optimal subset is chosen based on balanced accuracy and the Area Under the Curve (AUC) metric. This approach ensures the selection of features with the highest discriminative power while reducing dimensionality.

Domain Adaptation Logic

To enhance generalization across different subjects in memory classification tasks, we combine an ensemble learning framework with a two-phase fine-tuning process.

- **Initial Training Phase:** The ensemble of base models is trained on data that includes all training subjects. Each model is trained on bootstrap samples to encourage diversity within the ensemble, helping to reduce variance and achieve generalization.

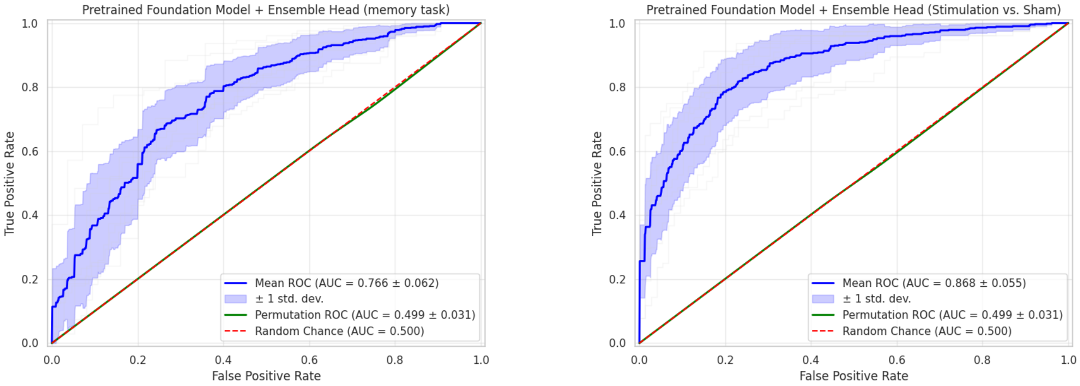
- **Fine-Tuning Phase:** After initial training, the models undergo fine-tuning on subject-specific data. In this phase, the learning rate is reduced to 10% of its original value, allowing the models to adapt to the nuances of the individual without overfitting. This controlled adaptation retains the general features learned during initial training while tuning for subject-specific characteristics.

We evaluated the performance of our ensemble classifier on the memory task and the Stimulation vs. Sham tasks. The ROC curves for these tasks, shown in Figure 5.9, demonstrate the classifier's effectiveness in distinguishing between classes in both cases.

In Figure 5.9a, the ROC curve for the memory task achieves a mean AUC of 0.766 ± 0.062, indicating a fair discrimination ability that surpasses linear probing and classification of representation space. Similarly, Figure 5.9b presents the ROC curve for the Stimulation vs. Sham task, yielding an even higher mean AUC of 0.868 ± 0.055.

This result highlights the classifier's strong discriminative capability in identifying stimulation effects, as reflected in the narrower confidence interval for this task.

These ROC results demonstrate the ensemble classifier's robustness and adaptability across diverse tasks, validating the utility of our domain adaptation approach. By fine-tuning each model to subject-specific data, we were able to effectively mitigate inter-subject variability, resulting in enhanced predictive performance across both tasks.



(a)

(b)

Figure 5.9: Receiver Operating Characteristic (ROC) curves for the proposed ensemble classifier with domain adaptation capability, applied to two binary classification tasks. Panel (a) shows the ROC curve for the memory task, achieving a mean AUC of 0.766 ± 0.062, indicating reliable model performance. Panel (b) displays the ROC curve for the Stimulation vs. Sham task, with a higher mean AUC of 0.868 ± 0.055, demonstrating greater discriminative ability. Shaded areas represent ±1 standard deviation.

Algorithm 1

Algorithm: Ensemble Classifiers with Domain Adaptation

```
Input: Encoded EEG, Labels, Subject_IDs, Parameters
Output: metrics, best_model

Procedure:
1.For each fold in LeaveOneGroupOut:
  1.1 Select validation subject
  1.2 Perform preprocessing:
    a.Apply subject-specific normalization
    b.Select optimal features

2.Train Ensemble Models:
  2.1 Create multiple classifier instances
  2.2 For each model:
    a. Train on bootstrap samples
    b. Validate and apply early stopping

3.Domain Adaptation:
  3.1 Fine-tune ensemble on validation subject data
  3.2 Adjust learning parameters

4.Evaluate Performance:
  4.1 Record best metrics
```

Chapter 6

Conclusion and Future Work

Electrophysiological recordings, in particular, modalities like EEG and LFP, have considerably furthered our understanding of neural mechanisms of sensorimotor and cognitive processes. These modalities yield physiomarkers or biomarkers important to understand neural mechanisms of diseases, the development of therapeutic interventions, and the study of treatment mechanisms. Despite progress in AI methodologies, current approaches are battling the problem of limited labeled data, which is a common issue in neuroimaging because of practical and ethical considerations. This dissertation explored the limitation of working with limited sample sizes, especially while analyzing these signals and their relationship with electrical neuromodulation. The key objective of this dissertation was to devise tailored automated data representation frameworks that assist in biomarker identification on small datasets.

The dissertation spans several core contributions. In Chapter 2, we presented a novel data acquisition schema to determine the effect of electrically induced Vagus Nerve Stimulation (VNS) on electrogastrography (EGG) signals, a domain that has been less attended to in comparison with other electrophysiological recordings like EEG. Characterizing these effects contributed to a better understanding of the GI regulation mechanisms and designing informed closed-loop VNS systems. The nature of data acquisition was not trivial, involving complex electrode implantation. For this and other technical and ethical reasons, a frequent problem with these types of studies is the extremely small and noisy dataset. We utilized a large set of engineered features and accounted for the high predictor-to-sample ratio using a voting algorithm that combined several feature selection methods. Our pipeline proved effective in improving the classification performance and interpretability to programmatically distinguish between baseline and VNS states. Feature selection significantly improved the interpretability of our models by allowing us to draw clear distinctions between feature importance. This method provided insights into VNS effects, revealing frequency-specific differences in the impact of 10 Hz and 30 Hz VNS stimulation.

Chapter 3 focused on leveraging ear-EEG for seizure detection using custom-designed hardware, the Aware hearable device. Ear-EEG offered a non-invasive and cost-effective alternative for long-term seizure monitoring. The Aware hearable captured EEG signals suitable for clinical seizure detection and distinguishing between wakefulness and sleep. The study demonstrated the feasibility of capturing accurate signals for real-world ambulatory use. However, certain limitations of scalp EEG, such as failing to detect deeply localized seizures, also applied to the Aware system. Nonetheless, by utilizing classical machine learning algorithms, including tree-based models, the study achieved excellent results for the seizure detection task. The device offered a promising tool for long-term seizure monitoring and potentially expanded EEG use cases, such as sleep state detection, through wearable, user-friendly technology. Future work will include improving artifact removal techniques and incorporating additional sensors, like accelerometers, to capture motor activity.

Chapter 4 explored non-linear feature integration to enhance SEEG classification in memory tasks. Traditional band-power features, commonly used for EEG analysis, failed to offer adequate discriminative power. Phase-Amplitude Coupling (PAC) emerged as a prominent feature, suggesting that the stimulation of the amygdala during encoding enhanced memory consolidation by increasing theta-gamma coupling in the hippocampus. This observation aligned with existing literature on the mechanisms underlying working memory, highlighting the importance of coupling between brain rhythms during memory processes. This chapter underlined the significance of using advanced features over canonical features, emphasizing the need for non-linear analyses to unravel complex brain dynamics.

In Chapter 5, we showed the promise of EEG foundation models as general feature extractors for unseen EEG datasets by pre-training two foundation models and fine-tuning them on major depressive disorder classification. While fine-tuned models exhibited excellent discriminative power in this case, they failed in the SEEG memory classification task. BENDR was not applicable to the SEEG dataset of Chapter 4 and its associated memory classification task due to its architectural limitation in ingesting SEEG data with a variable number of channels, and fine-tuning LaBraM via linear probing failed as the classification performance was at the chance level. The reason for this failure was the extremely small size of the SEEG dataset (less than 30 minutes per class). We went further by proposing to classify the latent space of SEEG data generated by LaBraM with traditional machine learning algorithms (XGBoost and SVM). In addition, we were able to significantly enhance the classification performance via our introduced domain adaptation algorithm.

These are the aspects that should be explored further in EEG foundation models: 1) their zero- or few-shot adaptability to very small-size data (A.2), in contrast with most of the literature testing EEG foundation models on datasets orders of magnitude larger than our SEEG dataset; and 2) interpretability. While the features extracted in the form of latent space can provide useful insights, as in our case, they revealed a significant inter-subject variability, but more work needs to be done to make these features interpretable for the users compared to the manually engineered ones.

Future work will focus on advancing foundation model adaptability for broader neuroimaging applications. These improvements have the potential to substantially impact the clinical use of neuroimaging biomarkers, paving the way for more effective, and personalized closed-loop neuromodulation.

Appendix A

Appendix

A.1 Chapter 2

The Petrosian Fractal Dimension (PFD) is a method designed to measure the complexity or irregularity of a signal and is computationally efficient compared to other traditional fractal dimension estimation techniques. The subsequent equation illustrates the computation of the Petrosian Fractal Dimension:

$\mathrm{PFD} = \frac{\log_{10} N}{\log_{10} N + \log_{10}\left(\frac{N}{N+0.4N\delta}\right)}$

where N is the window length, and $N\delta$ is the number of sign changes in the signal derivative.

Figure A.1 demonstrates the histogram plot of bootstrapped RMS values of baseline, VNS at 10 Hz, and VNS at 30 Hz. RMS values of baseline were statistically significantly different from VNS at 10 Hz and VNS 30 Hz (p-values = 0.000 and t-statistic = 118.862 and -69.637, respectively). However, the difference between RMS values of baseline and VNS at 10 Hz was greater than of baseline and VNS at 30 Hz. (52.82 μV vs -33.42 μV)

Figure A.2 demonstrates the accuracy, F1-, and F2-score of the trained Random Forest for the first scenario (baseline vs. VNS at 10 Hz).

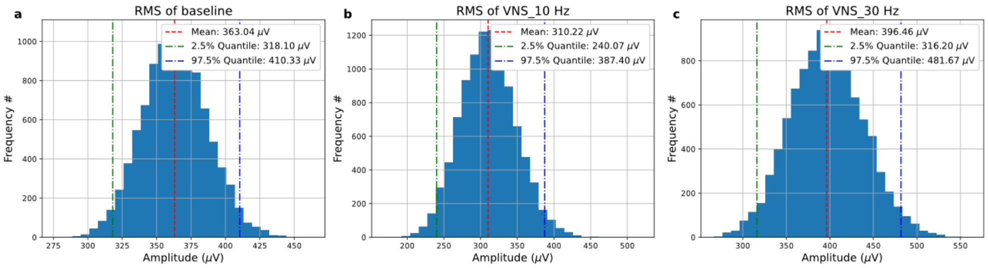


Figure A.1: Comparison of the RMS values. a) baseline, b) VNS at 10 Hz, c) VNS at 30 Hz.

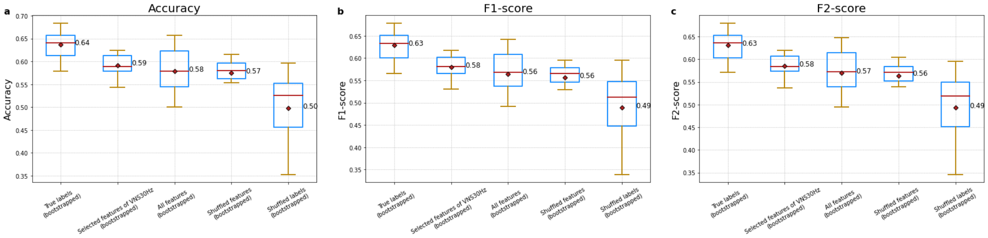


Figure A.2: Performance of the trained Random Forest for the first scenario. a) Accuracy, b) F1-score, c) F2-score.

Figure A.3 demonstrates the histogram plot of bootstrapped dominant frequency (DF) values of baseline, VNS at 10 Hz, and VNS at 30 Hz. DF values of baseline were statistically significantly different from VNS at 10 Hz and VNS 30 Hz (p-values = 0.000 and t-statistic = -106.132 and -198.124, respectively). However, the difference between DF values of baseline and VNS at 30 Hz was greater than of baseline and VNS at 10 Hz. (-1.05 cpm vs -0.59 cpm)

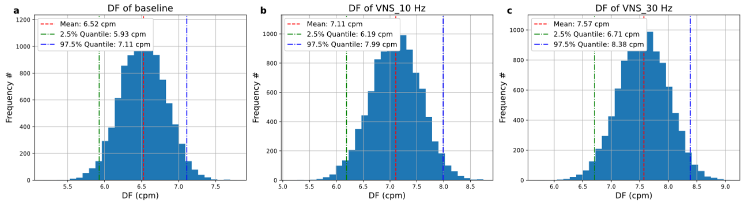


Figure A.3: Comparison of the DF values. a) baseline, b) VNS at 10 Hz, c) VNS at 30 Hz.

Figure A.4 demonstrates the accuracy, f1-, and fr2-score of the trained Random Forest for the second scenario (baseline vs. VNS at 30 Hz).

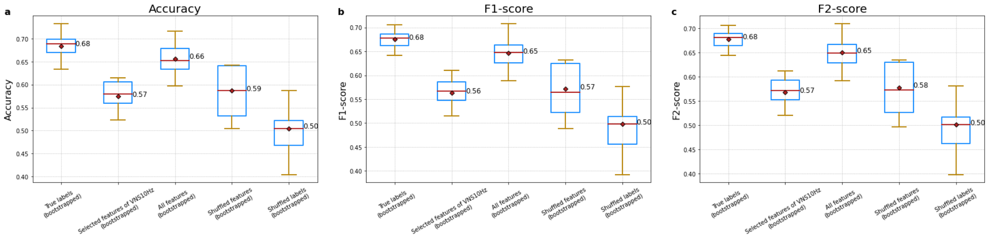


Figure A.4: Performance of the trained Random Forest for the second scenario. a) Accuracy, b) F1-score, c) F2-score.

A.2 Chapter 5

A.2.1 Inference Without Fine-Tuning

In addition to the linear probing and latent space classification, we investigated the concept of "Inference Without Fine-Tuning" which refers to the ability of machine learning models to perform tasks on new data without additional training or parameter adjustments such as in zero- shot learning [251]. One way to perform inference without fine-tuning is to construct class prototypes [252]. We begin by gathering representative samples from our SEEG dataset. Specifically, for each class in the memory classification task ("Forgotten" as Class 0 and "Remembered" as Class 1), a set of representative samples is extracted from this dataset. These samples are subsequently encoded into latent space feature vectors using the pre-trained encoder model (LaBraM).

Let the set of encoded vectors for Class 0 be denoted by $\{\mathbf{v}_i^{(0)}\}$ and for Class 1 by $\{\mathbf{v}_i^{(1)}\}$. The class prototype for each category is computed by calculating the mean vector of the corresponding encoded vectors. To ensure consistency in similarity measurements, each prototype is normalized to a unit vector:

$$\mathbf{p}^{(0)} = \frac{1}{N_0}\sum_{i=1}^{N_0}\mathbf{v}_i^{(0)}, \quad \mathbf{p}^{(1)} = \frac{1}{N_1}\sum_{i=1}^{N_1}\mathbf{v}_i^{(1)}$$

$$\mathbf{p}^{(0)} \leftarrow \frac{\mathbf{p}^{(0)}}{\|\mathbf{p}^{(0)}\|}, \quad \mathbf{p}^{(1)} \leftarrow \frac{\mathbf{p}^{(1)}}{\|\mathbf{p}^{(1)}\|}$$

where N_0 and N_1 are the number of samples in Class 0 and Class 1, respectively.

To classify a new sample, we proceed by encoding it using the same pre-trained encoder. Specifically, given a new EEG sample s, we obtain its corresponding encoded feature vector, denoted by $\mathbf{v}$. The encoded vector is then normalized to a unit vector to match the format of the class prototypes.

$$\mathbf{v} = \text{Encoder}(s)$$

$$\mathbf{v} \leftarrow \frac{\mathbf{v}}{\|\mathbf{v}\|}$$

The classification decision is made based on cosine similarity. Since all vectors are normalized, the cosine similarity can be directly computed as the dot product between $\mathbf{v}$ and each class prototype. The class label is assigned to the sample based on the highest computed similarity value.

$$\text{Similarity to Class 0} = \cos\theta_0 = \mathbf{v} \cdot \mathbf{p}^{(0)}$$

$$\text{Similarity to Class 1} = \cos\theta_1 = \mathbf{v} \cdot \mathbf{p}^{(1)}$$

$$\text{Predicted Class} = \begin{cases} 0, & \text{if } \cos\theta_0 > \cos\theta_1 \\ 1, & \text{otherwise} \end{cases}$$

Bibliography

[1] A. Zandvakili, N. S. Philip, S. R. Jones, A. R. Tyrka, B. D. Greenberg, and L. L. Carpenter, "Use of machine learning in predicting clinical response to transcranial magnetic stimulation in comorbid posttraumatic stress disorder and major depression: A resting state electroencephalography study," *Journal of Affective Disorders*, 2019.

[2] P. Durongbhan, Y. Zhao, L.-Y. Chen, P. Zis, M. D. Marco, Z. C. Unwin, A. Venneri, X. He, S. Li, Y. Zhao, D. Blackburn, and P. G. Sarrigiannis, "A dementia classification framework using frequency and time-frequency features based on eeg signals," *Ieee Transactions on Neural Systems and Rehabilitation Engineering*, 2019.

[3] S. Jóźwiak, A. Becker, C. Cepeda, J. Engel, V. Gnatkovsky, G. Huberfeld, M. Kaya, K. Kobow, M. Simonato, and J. A. Loeb, "Wonoep appraisal: Development of epilepsy biomarkers—what we can learn from our patients?," *Epilepsia*, 2017.

[4] L. R. Trambaiolli, A. C. Lorena, F. J. Fraga, P. A. Medeiros Kanda, R. Anghinah, and R. Nitrini, "Improving alzheimer's disease diagnosis with machine learning techniques," *Clinical Eeg and Neuroscience*, 2011.

[5] G. Chen, Z. Zhang, M. Wang, G. Yang, B. Jin, and T. Aung, "Update on the

neuroimaging and electroencephalographic biomarkers of epileptogenesis: A literature review," *Frontiers in Neurology*, 2021.

[6] L. Lei, Z. Liu, M. Gao, Z. Liu, N. Sun, C. Yang, A. Zhang, Y. Wang, and K. Zhang, "Theta oscillations: A rhythm difference comparison between major depressive disorder and anxiety disorder," *Frontiers in Psychiatry*, 2022.

[7] E. C. Irelli, G. Leodori, A. Morano, and C. D. Bonaventura, "Eeg markers of treatment resistance in idiopathic generalized epilepsy: From standard eeg findings to advanced signal analysis," *Biomedicines*, 2022.

[8] C. Moral-Rubio, P. Balugo, A. Fraile-Pereda, V. Pytel, L. Fernández-Romero, C. Delgado-Alonso, A. Delgado-Álvarez, J. Matías-Guiu, and J. L. Ayala, "Application of machine learning to electroencephalography for the diagnosis of primary progressive aphasia: A pilot study," *Brain Sciences*, 2021.

[9] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (eeg) classification tasks: A review," *Journal of Neural Engineering*, 2019.

[10] C. Joucla, D. Gabriel, J.-P. Ortega, and E. Haffen, "Three simple steps to improve the interpretability of eeg-svm studies," *Journal of Neurophysiology*, 2022.

[11] H. Chahal, H. Toner, and I. Rahkovsky, "Small data's big ai potential," *Center for Security and Emerging Technology. https://cset. georgetown. edu/publication/small-datas-big-ai-potential*, 2021.

[12] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[13] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.

[14] S. Wager, S. Wang, and P. S. Liang, "Dropout training as adaptive regularization," *Advances in neural information processing systems*, vol. 26, 2013.

[15] L. Brigato and L. Iocchi, "A close look at deep learning with small data," in *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 2490–2497, IEEE, 2021.

[16] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "Randaugment: Practical automated data augmentation with a reduced search space," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 702–703, 2020.

[17] M. Sonogashira, M. Shonai, and M. Iiyama, "High-resolution bathymetry by deep-learning-based image superresolution," *Plos one*, vol. 15, no. 7, p. e0235487, 2020.

[18] S. Zini, M. Buzzelli, B. Twardowski, and J. van de Weijer, "Planckian jitter: enhancing the color quality of self-supervised visual representations," *arXiv preprint arXiv:2202.07993*, 2022.

[19] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," *arXiv preprint arXiv:1708.04552*, 2017.

[20] G. B. P. da Costa, W. A. Contato, T. S. Nazare, J. E. Neto, and M. Ponti, "An empirical study on the effects of different types of noise in image classification tasks," *arXiv preprint arXiv:1609.02781*, 2016.

[21] E. Castro, J. S. Cardoso, and J. C. Pereira, “Elastic deformations for data augmentation in breast cancer mass detection,” in *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 230–234, IEEE, 2018.

[22] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, “mixup: Beyond empirical risk minimization,” *arXiv preprint arXiv:1710.09412*, 2017.

[23] Y.-J. Lin and I.-F. Chung, “Medical data augmentation using generative adversarial networks: X-ray image generation for transfer learning of hip fracture detection,” in *2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, pp. 1–5, IEEE, 2019.

[24] C. He, J. Liu, Y. Zhu, and W. Du, “Data augmentation for deep neural networks model in eeg classification task: a review,” *Frontiers in Human Neuroscience*, vol. 15, p. 765525, 2021.

[25] S. Motamed, P. Rogalla, and F. Khalvati, “Data augmentation using generative adversarial networks (gans) for gan-based detection of pneumonia and covid-19 in chest x-ray images,” *Informatics in Medicine Unlocked*, vol. 27, p. 100779, 2021.

[26] T. Iqball and M. A. Wani, “Generative adversarial networks for data augmentation in x-ray medical imaging,” in *Generative Adversarial Learning: Architectures and Applications*, pp. 341–355, Springer, 2022.

[27] M. Pesteie, P. Abolmaesumi, and R. N. Rohling, “Adaptive augmentation of medical data using independently conditional variational auto-encoders,” *IEEE transactions on medical imaging*, vol. 38, no. 12, pp. 2807–2820, 2019.

[28] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[29] D. P. Kingma, M. Welling, *et al.*, “An introduction to variational autoencoders,” *Foundations and Trends® in Machine Learning*, vol. 12, no. 4, pp. 307–392, 2019.

[30] J. J. Bird, M. Pritchard, A. Fratini, A. Ekart, and D. R. Faria, “Synthetic Biological Signals Machine-Generated by GPT-2 Improve the Classification of EEG and EMG through Data Augmentation,” *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3498–3504, 2021.

[31] R. Niu, Y. Wang, H. Xi, Y. Hao, and M. Zhang, “Epileptic Seizure Prediction by Synthesizing EEG Signals through GPT,” *ACM International Conference Proceeding Series*, pp. 419–423, 2021.

[32] Y. Chen, X.-H. Yang, Z. Wei, A. A. Heidari, N. Zheng, Z. Li, H. Chen, H. Hu, Q. Zhou, and Q. Guan, “Generative adversarial networks in medical image augmentation: a review,” *Computers in Biology and Medicine*, p. 105382, 2022.

[33] Nature Publishing Group, “Biophysical models - latest research and news,” 2023.

[34] A. Mang, S. Bakas, S. Subramanian, C. Davatzikos, and G. Biros, “Integrated biophysical modeling and image analysis: application to neuro-oncology,” *Annual review of biomedical engineering*, vol. 22, pp. 309–341, 2020.

[35] M. X. Cohen, “Where does eeg come from and what does it mean?,” *Trends in neurosciences*, vol. 40, no. 4, pp. 208–218, 2017.

[36] C. Cakan, N. Jajcay, and K. Obermayer, “neurolib: A simulation framework for whole-brain neural mass modeling,” *Cognitive Computation*, pp. 1–21, 2021.

[37] W. Hao, M. El-Khamy, J. Lee, J. Zhang, K. J. Liang, C. Chen, and L. C. Duke, “Towards fair federated learning with zero-shot data augmentation,” in

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3310–3319, 2021.

[38] K. Yue, R. Jin, R. Pilgrim, C.-W. Wong, D. Baron, and H. Dai, “Neural tangent kernel empowered federated learning,” in *International Conference on Machine Learning*, pp. 25783–25803, PMLR, 2022.

[39] E. Lashgari, D. Liang, and U. Maoz, “Data augmentation for deep-learning-based electroencephalography,” *Journal of Neuroscience Methods*, vol. 346, no. July, p. 108885, 2020.

[40] A. G. Habashi, A. M. Azab, S. Eldawlatly, and G. M. Aly, “Generative adversarial networks in EEG analysis: an overview,” *Journal of NeuroEngineering and Rehabilitation*, vol. 20, no. 1, pp. 1–24, 2023.

[41] F. P. Carrle, Y. Hollenbenders, and A. Reichenbach, “Generation of synthetic EEG data for training algorithms supporting the diagnosis of major depressive disorder,” *Frontiers in Neuroscience*, vol. 17, no. October, pp. 1–17, 2023.

[42] K. G. Hartmann, R. T. Schirrmeister, and T. Ball, “Eeg-gan: Generative adversarial networks for electroencephalograhic (eeg) brain signals,” *arXiv preprint arXiv:1806.01875*, 2018.

[43] N. Donges, “What is transfer learning? exploring the popular deep learning approach,” *Built in*, 2019.

[44] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, “A comprehensive survey on transfer learning. arxiv e-prints,” *arXiv preprint arXiv:1911.02685*, 2019.

[45] N. Gupta, “A pre-trained vs fine-tuning methodology in transfer learning,” in

Journal of Physics: Conference Series, vol. 1947, p. 012028, IOP Publishing, 2021.

[46] M. Iman, K. Rasheed, and H. R. Arabnia, "A review of deep transfer learning and recent advancements," *arXiv preprint arXiv:2201.09679*, 2022.

[47] J. H. Harkness, J. J. Theis, W. M. O'Keefe, G. W. Wade, and K. J. Lawton, "Leveraging ai transfer learning for rapid and accurate identification and quantification of cellular biomarkers in microscopy images," *The FASEB Journal*, vol. 36, 2022.

[48] Z. Wan, R. Yang, M. Huang, N. Zeng, and X. Liu, "A review on transfer learning in eeg signal analysis," *Neurocomputing*, vol. 421, pp. 1–14, 2021.

[49] C.-Y. Yang, P.-C. Chen, and W.-C. Huang, "Cross-domain transfer of eeg to eeg or ecg learning for cnn classification models," *Sensors*, vol. 23, no. 5, p. 2458, 2023.

[50] C. Doersch and A. Zisserman, "Multi-task self-supervised visual learning," in *Proceedings of the IEEE international conference on computer vision*, pp. 2051–2060, 2017.

[51] X. Liu, F. Zhang, Z. Hou, Z. Wang, L. Mian, J. Zhang, and J. Tang, "Self-supervised learning: Generative or contrastive. corr abs/2006.08218 (2020)," *arXiv preprint arXiv:2006.08218*, 2020.

[52] Y. Ou, S. Sun, H. Gan, R. Zhou, and Z. Yang, "An improved self-supervised learning for eeg classification," *Mathematical Biosciences and Engineering*, vol. 19, no. 7, pp. 6907–6922, 2022.

[53] C. Yang, D. Xiao, M. B. Westover, and J. Sun, "Self-supervised eeg representation learning for automatic sleep staging," *arXiv preprint arXiv:2110.15278*, 2021.

[54] A. Gramfort, H. Banville, O. Chehab, A. Hyvärinen, and D. Engemann, “Learning with self-supervision on eeg data,” in *2021 9th International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1–2, IEEE, 2021.

[55] X. Li and V. Metsis, “Spp-eegnet: An input-agnostic self-supervised eeg representation model for inter-dataset transfer learning,” in *Proceedings of the 18th International Conference on Computing and Information Technology (IC2IT 2022)*, pp. 173–182, Springer, 2022.

[56] H. Banville, I. Albuquerque, A. Hyvärinen, G. Moffat, D.-A. Engemann, and A. Gramfort, “Self-supervised representation learning from electroencephalography signals,” in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, IEEE, 2019.

[57] S. Shurrab and R. Duwairi, “Self-supervised learning methods and applications in medical imaging analysis: A survey,” *PeerJ Computer Science*, vol. 8, p. e1045, 2022.

[58] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, “Context encoders: Feature learning by inpainting,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2536–2544, 2016.

[59] V. Rani, S. T. Nabi, M. Kumar, A. Mittal, and K. Kumar, “Self-supervised learning: A succinct review,” *Archives of Computational Methods in Engineering*, vol. 30, no. 4, pp. 2761–2775, 2023.

[60] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, “Emerging properties in self-supervised vision transformers,” in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9650–9660, 2021.

[61] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar, *et al.*, "Bootstrap your own latent-a new approach to self-supervised learning," *Advances in neural information processing systems*, vol. 33, pp. 21271–21284, 2020.

[62] X. Chen and K. He, "Exploring simple siamese representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 15750–15758, 2021.

[63] I. Goodfellow, "Deep learning," 2016.

[64] D. P. Kingma, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[65] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning*, pp. 1096–1103, 2008.

[66] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[67] M. Zeydabadinezhad, C. C. Horn, and B. Mahmoudi, "Quantifying the effects of vagus nerve stimulation on gastric myoelectric activity in ferrets using an interpretable machine learning approach," *Plos one*, vol. 18, no. 12, p. e0295297, 2023.

[68] K.-H. Lu, J. Cao, S. Oleson, M. P. Ward, R. Phillips, T. L. Powley, and Z. Liu, "Vagus nerve stimulation promotes gastric emptying by increasing pyloric opening measured with magnetic resonance imaging," *Neurogastroenterology & Motility*, vol. 30, no. 10, p. e13380, 2018.

[69] G. O'Grady, A. A. Gharibans, P. Du, and J. D. Huizinga, "The gastric conduction system in health and disease: a translational review," *American Journal of Physiology-Gastrointestinal and Liver Physiology*, vol. 321, no. 5, pp. G527–G542, 2021.

[70] W. C. Alvarez, "The electrogastrogram and what it shows," *Jama*, vol. 78, pp. 1116–1119, 1922.

[71] M. D. Poscente and M. P. Mintchev, "Enhanced electrogastrography: A realistic way to salvage a promise that was never kept?," *World journal of gastroenterology*, vol. 23, no. 25, p. 4517, 2017.

[72] A. A. Gharibans, B. L. Smarr, D. C. Kunkel, L. J. Kriegsfeld, H. M. Mousa, and T. P. Coleman, "Artifact Rejection Methodology Enables Continuous, Noninvasive Measurement of Gastric Myoelectric Activity in Ambulatory Subjects," *Scientific Reports*, vol. 8, no. 1, pp. 1–5, 2018.

[73] S. A. Shikora, "Implantable gastric stimulation for the treatment of severe obesity," *Obesity surgery*, vol. 14, no. 4, pp. 545–548, 2004.

[74] C. N. Price, D. T. Westwick, and M. P. Mintchev, "Analysis of canine model of gastric electrical uncoupling using recurrence quantification analysis," *Digestive Diseases and Sciences*, vol. 50, no. 5, pp. 885–892, 2005.

[75] N. B. Popović, N. Miljković, and M. B. Popović, "Simple gastric motility assessment method with a single-channel electrogastrogram," *Biomedizinische Technik*, vol. 64, no. 2, pp. 177–185, 2019.

[76] T. Gruden, N. B. Popović, K. Stojmenova, G. Jakus, N. Miljković, S. Tomažič, and J. Sodnik, "Electrogastrography in autonomous vehicles—an objective method for assessment of motion sickness in simulated driving environments," *Sensors*, vol. 21, no. 2, p. 550, 2021.

[77] N. Miljković, N. B. Popović, M. Prodanov, and J. Sodnik, “Assessment of sickness in virtual environments,” in *Proceedings of the 9th International Conference on Information Society and Technology (ICIST 2019)*, 2019.

[78] D. Grundy and T. Scratcherd, “Effect of stimulation of the vagus nerve in bursts on gastric acid secretion and motility in the anaesthetized ferret.,” *The Journal of Physiology*, vol. 333, no. 1, pp. 451–461, 1982.

[79] J. A. Shulgach, D. W. Beam, A. C. Nanivadekar, D. M. Miller, S. Fulton, M. Sciullo, J. Ogren, L. Wong, B. L. McLaughlin, B. J. Yates, C. C. Horn, and L. E. Fisher, “Selective stimulation of the ferret abdominal vagus nerve with multi-contact nerve cuff electrodes,” *Scientific Reports*, vol. 11, no. 1, pp. 1–30, 2021.

[80] A. C. Nanivadekar, D. M. Miller, S. Fulton, L. Wong, J. Ogren, G. Chitnis, B. McLaughlin, S. Zhai, L. E. Fisher, B. J. Yates, C. C. Horn, A. C. N. Id, D. M. Miller, S. Fulton, L. Wong, J. Ogren, G. Chitnis, B. McLaughlin, S. Zhai, L. E. Fisher, B. J. Y. Id, and C. C. H. Id, “Machine learning prediction of emesis and gastrointestinal state in ferrets,” *PLoS ONE*, vol. 14, no. 10, pp. 1–20, 2019.

[81] D. Komorowski, “EGG DWPack: System for Multi-Channel Electrogastrographic Signals Recording and Analysis,” *Journal of Medical Systems*, vol. 42, no. 11, pp. 26–28, 2018.

[82] S. J. Roberts, W. Penny, and I. Rezek, “Temporal and spatial complexity measures for electroencephalogram based brain-computer interfacing,” *Medical \& biological engineering \& computing*, vol. 37, no. 1, pp. 93–98, 1999.

[83] S. Pei, F. Nie, R. Wang, and X. Li, “Efficient clustering based on a unified view of k-means and ratio-cut,” *Advances in Neural Information Processing Systems*, vol. 2020-Decem, no. c, pp. 1–12, 2020.

[84] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalography and clinical neurophysiology*, vol. 29, no. 3, pp. 306–310, 1970.

[85] H. S. Shin, C. Lee, and M. Lee, "Ideal filtering approach on DCT domain for biomedical signals: Index blocked DCT filtering method (IB-DCTFM)," *Journal of Medical Systems*, vol. 34, no. 4, pp. 741–753, 2010.

[86] E. o. f. printedfiber reinforcedcomposites, annealing in enhancing the interfacialbonding, and mechanical properties of 3D, "ce d M us pt," *Nanotechnology*, pp. 0–22, 2019.

[87] N. P. Du Sert, K. M. Chu, M. K. Wai, J. A. Rudd, and P. L. Andrews, "Reduced normogastric electrical activity associated with emesis: A telemetric study in ferrets," *World Journal of Gastroenterology*, vol. 15, no. 48, pp. 6034–6043, 2009.

[88] G. Jakus, J. Sodnik, and N. Miljković, "Electrogastrogram-derived features for automated sickness detection in driving simulator," *Sensors*, vol. 22, no. 22, p. 8616, 2022.

[89] Y. Saeys, I. Inza, and P. Larranaga, "A review of feature selection techniques in bioinformatics," *bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.

[90] J. J. Stewart, M. J. Wood, C. D. Wood, and M. E. Mims, "Effects of motion sickness and antimotion sickness drugs on gastric function," *The Journal of Clinical Pharmacology*, vol. 34, no. 6, pp. 635–643, 1994.

[91] M. F. Amri, A. R. Yuliani, A. I. Simbolon, R. Ristiana, and D. E. Kusumandari, "Toward early abnormalities detection on digestive system: Multi-features electrogastrogram (egg) signal classification based on machine learning," in *2021 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET)*, pp. 185–190, IEEE, 2021.

[92] A. Ali, I. Cosic, D. Kumar, B. Polus, and C. Da_Costa, “Power changes of eeg signals associated with muscle fatigue: The root mean square analysis of eeg bands,” in *Proceedings of the 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference*, IEEE, 2004.

[93] M. Ribeiro, S. Singh, and C. Guestrin, ““why should i trust you?”,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2016.

[94] G. Chandrashekar and F. Sahin, “A survey on feature selection methods,” *Computers & Electrical Engineering*, 2014.

[95] V. Teckentrup, S. Neubert, J. C. Santiago, M. Hallschmid, M. Walter, and N. B. Kroemer, “Non-invasive stimulation of vagal afferents reduces gastric frequency,” *Brain Stimulation*, vol. 13, no. 2, pp. 470–473, 2019.

[96] B. Cheung and P. Vaitkus, “Perspectives of electrogastrography and motion sickness,” *Brain Research Bulletin*, 1998.

[97] V. Berger and Y. Zhou, “Kolmogorov–smirnov test: Overview,” in *Wiley StatsRef: Statistics Reference Online*, Wiley, 2014.

[98] D. Komorowski and B. Mika, “A new approach for denoising multichannel electrogastrographic signals,” *Biomedical Signal Processing and Control*, 2018.

[99] K. Koch and R. Stern, *Handbook of electrogastrography.* New York: Oxford University Press, 2003.

[100] P. Perez-Alcazar and A. Santos, “Relationship between sampling rate and quantization noise,” in *2002 14th International Conference on Digital Signal Processing Proceedings DSP 2002 (Cat No02TH8628)*, IEEE, 2002.

[101] R. Lyons, *Understanding Digital Signal Processing.* Boston: Addison-Wesley Longman Publishing, 2010.

[102] R. Esteller, G. Vachtsevanos, J. Echauz, and B. Lilt, “A comparison of fractal dimension algorithms using synthetic and experimental data,” in *ISCAS’99 Proceedings of the 1999 IEEE International Symposium on Circuits and Systems VLSI (Cat No99CH36349)*, IEEE, 1999.

[103] H. Shin, C. Lee, and M. Lee, “Ideal filtering approach on dct domain for biomedical signals: Index blocked dct filtering method (ib-dctfm),” *Journal of Medical Systems*, 2009.

[104] Y. Zhang, P. Li, X. Zhu, S. Su, Q. Guo, P. Xu, *et al.*, “Extracting time-frequency feature of single-channel vastus medialis emg signals for knee exercise pattern recognition,” *PLOS ONE*, 2017.

[105] A. Al-Timemy, R. Khushaba, G. Bugmann, and J. Escudero, “Improving the performance against force variation of emg controlled multifunctional upper-limb prostheses for transradial amputees,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2016.

[106] S. Sharanya and S. Arjunan, “Fractal dimension techniques for analysis of cardiac autonomic neuropathy (can),” *Biomedical Engineering: Applications, Basis and Communications*, 2023.

[107] J. He, L. Yang, D. Liu, and Z. Song, “Automatic recognition of high-density epileptic eeg using support vector machine and gradient-boosting decision tree,” *Brain Sciences*, 2022.

[108] A. Perez-Sanchez, M. Valtierra-Rodriguez, C. Perez-Ramierz, J. De-Santiago-Perez, and J. Amezquita-Sanchez, “Epileptic seizure prediction using wavelet

transform, fractal dimension, support vector machine, and eeg signals," *Fractals*, 2022.

[109] P. Alagumariappan, K. Krishnamurthy, S. Kandiah, E. Cyril, and R. V, "Diagnosis of type 2 diabetes using electrogastrograms: Extraction and genetic algorithm–based selection of informative features," *JMIR Biomedical Engineering*, 2020.

[110] B. Bein, "Entropy," *Best Practice & Research Clinical Anaesthesiology*, 2006.

[111] Z. Liang, X. Duan, and X. Li, "Entropy measures in neural signals," in *Signal Processing in Neuroscience*, Springer, 2016.

[112] E. Peri, L. Xu, C. Ciccarelli, N. Vandenbussche, H. Xu, X. Long, *et al.*, "Singular value decomposition for removal of cardiac interference from trunk electromyogram," *Sensors*, 2021.

[113] D. Komorowski, "Egg dwpack: System for multi-channel electrogastrographic signals recording and analysis," *Journal of Medical Systems*, 2018.

[114] N. Sert, K. Chu, M. Wai, J. Rudd, and P. Andrews, "Reduced normogastric electrical activity associated with emesis: A telemetric study in ferrets," *World Journal of Gastroenterology*, 2009.

[115] A. Chhabra, R. Subramaniam, A. Srivastava, H. Prabhakar, M. Kalaivani, and S. Paranjape, "Spectral entropy monitoring for adults and children undergoing general anaesthesia," *Cochrane Database of Systematic Reviews*, 2016.

[116] R. Esteller, J. Echauz, T. Tcheng, B. Litt, and B. Pless, "Line length: an efficient feature for seizure onset detection," in *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, pp. 1707–1710, IEEE, 2001.

[117] C. Bandt and B. Pompe, “Permutation entropy: A natural complexity measure for time series,” *Physical Review Letters*, 2002.

[118] F. Chang, “Electrogastrography: Basic knowledge, recording, processing and its clinical applications,” *Journal of Gastroenterology and Hepatology*, 2005.

[119] A. Petrosian, “Kolmogorov complexity of finite sequences and recognition of different preictal eeg patterns,” in *Proceedings eighth IEEE symposium on computer-based medical systems*, pp. 212–217, IEEE, 1995.

[120] P. Namadurai, V. Padmanabhan, and R. Swaminathan, “Multifractal analysis of uterine electromyography signals for the assessment of progression of pregnancy in term conditions,” *IEEE Journal of Biomedical and Health Informatics*, 2019.

[121] D. Donoho, “High-dimensional data analysis: The curses and blessings of dimensionality,” in *AMS math challenges lecture*, p. 32, American Mathematical Society, 2000.

[122] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[123] L. Breiman, *Machine Learning.* Springer, 2001.

[124] P. Cohen and D. Jensen, “Overfitting explained,” in *Sixth International Workshop on Artificial Intelligence and Statistics*, pp. 115–122, PMLR, 1997.

[125] C. Hsu, C. Chang, and C. Lin, “A practical guide to support vector classification,” tech. rep., National Taiwan University, 2003.

[126] M. J. der Laan, E. C. Polley, and A. E. Hubbard, “Super learner,” *Statistical applications in genetics and molecular biology*, vol. 6, no. 1, 2007.

[127] I. Rish, “An empirical study of the naive bayes classifier,” in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, pp. 41–46, AAAI Press, 2001.

[128] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ACM, 2019.

[129] C. Ling, J. Huang, and H. Zhang, "Auc: a statistically consistent and more discriminating measure than accuracy," *IJCAI*, vol. 3, pp. 519–524, 2003.

[130] M. Ganz and E. Konukoglu, "Permutation tests for classification: Revisited," in *2017 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, IEEE, 2017.

[131] J. Huang and C. Ling, "Using auc and accuracy in evaluating learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, 2005.

[132] S. Okkesim, A. Al Kafee, M. Yildirim, Y. Kayar, A. Danalioglu, M. Aydin, *et al.*, "Line length feature of the electrogastrogram for delayed gastric emptying diagnosis," in *2016 IEEE International Conference on Signal and Image Processing (ICSIP)*, IEEE, 2016.

[133] M. Zeydabadinezhad, J. Jowers, D. Buhl, B. Cabaniss, and B. Mahmoudi, "A personalized earbud for non-invasive long-term eeg monitoring," *Journal of Neural Engineering*, 2024.

[134] R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, *et al.*, "Ilae official report: a practical clinical definition of epilepsy," *Epilepsia*, vol. 55, pp. 475–482, 2014.

[135] U. Seneviratne, D. Rajendran, M. Brusco, and T. Phan, "How good are we at diagnosing seizures based on semiology?," *Epilepsia*, vol. 53, pp. e63–e66, 2012.

[136] L. Kiloh, A. McComas, and J. Osselton, *Clinical electroencephalography*. Butterworth-Heinemann, 2013.

[137] S. Benbadis and W. Tatum, "Overinterpretation of eegs and misdiagnosis of epilepsy," *Journal of Clinical Neurophysiology*, vol. 20, pp. 42–44, 2003.

[138] S. Smith, "Eeg in the diagnosis, classification, and management of patients with epilepsy," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 76, pp. ii2–ii7, 2005.

[139] D. Dash, L. Hernandez-Ronquillo, F. Moien-Afshari, and J. Tellez-Zenteno, "Ambulatory eeg: a cost-effective alternative to inpatient video-eeg in adult patients," *Epileptic Disorders*, vol. 14, pp. 290–297, 2012.

[140] H. Faulkner, H. Arima, and A. Mohamed, "The utility of prolonged outpatient ambulatory eeg," *Seizure*, vol. 21, pp. 491–495, 2012.

[141] G. Cascino, "Clinical indications and diagnostic yield of video-electroencephalographic monitoring in patients with seizures and spells," *Mayo Clinic Proceedings*, pp. 1111–1120, 2002.

[142] J. Liou, E. Smith, L. Bateman, S. Bruce, G. McKhann, R. Goodman, *et al.*, "A model for focal seizure onset, propagation, evolution, and progression," *Elife*, vol. 9, p. e50927, 2020.

[143] D. Burkholder, J. Britton, V. Rajasekaran, R. Fabris, P. Cherian, K. Kelly-Williams, E. So, K. Nickels, L. Wong-Kisiel, T. Lagerlund, and G. Cascino, "Routine vs extended outpatient eeg for the detection of interictal epileptiform discharges," *Neurology*, vol. 86, no. 16, pp. 1524–1530, 2016.

[144] A. Pasupuleti, J. Amling, T. Chang, J. Scafidi, and T. Tsuchida, "Skin integrity

during prolonged eeg recording in hospitalized neonatal and pediatric patients (p3. 247)," in *AAN Enterprises*, 2016.

[145] V. Goverdovsky, W. Von Rosenberg, T. Nakamura, D. Looney, D. J. Sharp, C. Papavassiliou, M. J. Morrell, and D. P. Mandic, "Hearables: Multimodal physiological in-ear sensing," *Scientific Reports*, vol. 7, no. 1, pp. 1–10, 2017.

[146] I. C. Zibrandtsen, P. Kidmose, C. B. Christensen, and T. W. Kjaer, "Ear-EEG detects ictal and interictal abnormalities in focal and generalized epilepsy–A comparison with scalp EEG monitoring," *Clinical Neurophysiology*, vol. 128, no. 12, pp. 2454–2461, 2017.

[147] K. B. Mikkelsen, Y. R. Tabar, S. L. Kappel, C. B. Christensen, H. O. Toft, M. C. Hemmsen, M. L. Rank, M. Otto, and P. Kidmose, "Accurate whole-night sleep monitoring with dry-contact ear-EEG," *Scientific reports*, vol. 9, no. 1, p. 16824, 2019.

[148] J. Nielsen, I. Zibrandtsen, P. Masulli, T. Sørensen, T. Andersen, and T. Kjær, "Towards a wearable multi-modal seizure detection system in epilepsy: a pilot study," *Clinical Neurophysiology*, vol. 136, pp. 40–48, 2022.

[149] D. Zambrana-Vinaroz, J. Vicente-Samper, and J. Sabater-Navarro, "Validation of continuous monitoring system for epileptic users in outpatient settings," *Sensors*, vol. 22, no. 8, p. 2900, 2022.

[150] K. Vandecasteele, T. De Cooman, C. Chatzichristos, E. Cleeren, L. Swinnen, J. Macea Ortiz, S. Van Huffel, M. Dümpelmann, A. Schulze-Bonhage, M. De Vos, and W. Van Paesschen, "The power of ecg in multimodal patient-specific seizure monitoring: added value to an eeg-based detector using limited channels," *Epilepsia*, vol. 62, no. 10, pp. 2333–2343, 2021.

[151] K. Vandecasteele, T. De Cooman, J. Dan, E. Cleeren, S. Van Huffel, B. Hunyadi, and W. Van Paesschen, "Visual seizure annotation and automated seizure detection using behind-the-ear electroencephalographic channels," *Epilepsia*, vol. 61, no. 4, pp. 766–775, 2020.

[152] S. You, B. Cho, Y. Shon, D. Seo, and I. Kim, "Semi-supervised automatic seizure detection using personalized anomaly detecting variational autoencoder with behind-the-ear eeg," *Computer Methods and Programs in Biomedicine*, vol. 213, p. 106542, 2022.

[153] C. Musaeus, G. Waldemar, B. Andersen, P. Høgh, P. Kidmose, M. Hemmsen, M. Rank, T. Kjær, and K. Frederiksen, "Long-term eeg monitoring in patients with alzheimer's disease using ear-eeg: A feasibility study," *Journal of Alzheimer's Disease*, vol. 90, no. 4, pp. 1713–1723, 2022.

[154] B. B. Ma and V. R. Rao, "Responsive neurostimulation: candidates and considerations," *Epilepsy & Behavior*, vol. 88, pp. 388–395, 2018.

[155] M. Koelle, "What Makes Wearable Technologies Socially Acceptable?," *XRDS: Crossroads, The ACM Magazine for Students*, vol. 29, no. 2, pp. 30–35, 2022.

[156] A. J. Casson, "Wearable EEG and beyond," *Biomedical engineering letters*, vol. 9, no. 1, pp. 53–71, 2019.

[157] J. Tang, R. El Atrache, S. Yu, U. Asif, M. Jackson, S. Roy, *et al.*, "Seizure detection using wearable sensors and machine learning: Setting a benchmark," *Epilepsia*, vol. 62, pp. 1807–1819, 2021.

[158] R. Picard and E. Boyer, "Smartwatch biomarkers and the path to clinical use," *Med*, vol. 2, pp. 797–799, 2021.

[159] T. Tobollik, S. Shalawadi, C. Getschmann, and F. Echtler, "Exploring epileptic seizure detection with commercial smartwatches," in *2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, pp. 636–641, 2021.

[160] J. Macea, M. Bhagubai, V. Broux, M. De Vos, and W. Van Paesschen, "In-hospital and home-based long-term monitoring of focal epilepsy with a wearable electroencephalographic device: Diagnostic yield and user experience," *Epilepsia*, 2023.

[161] F. Onorati, G. Regalia, C. Caborni, M. Migliorini, D. Bender, M.-Z. Poh, *et al.*, "Multicenter clinical assessment of improved wearable multimodal convulsive seizure detectors," *Epilepsia*, vol. 58, pp. 1870–1879, 2017.

[162] J. Téllez-Zenteno and L. Hernández-Ronquillo, "A review of the epidemiology of temporal lobe epilepsy," *Epilepsy Research and Treatment*, vol. 2012, 2012.

[163] S. L. Kappel, S. Makeig, and P. Kidmose, "Ear-EEG Forward Models: Improved Head-Models for Ear-EEG," *Frontiers in Neuroscience*, vol. 13, no. September, 2019.

[164] A. Meiser, F. Tadel, S. Debener, and M. Bleichner, "The sensitivity of ear-eeg: Evaluating the source-sensor relationship using forward modeling," *Brain Topography*, vol. 33, pp. 665–676, 2020.

[165] S. Beniczky, P. Karoly, E. Nurse, P. Ryvlin, and M. Cook, "Machine learning and wearable devices of the future," *Epilepsia*, vol. 62, p. S116–S124, 2021.

[166] S. You, B. Cho, S. Yook, J. Kim, Y. Shon, D. Seo, and I. Kim, "Unsupervised automatic seizure detection for focal-onset seizures recorded with behind-the-ear eeg using an anomaly-detecting generative adversarial network," *Computer Methods and Programs in Biomedicine*, vol. 193, p. 105472, 2020.

[167] S. Gedam and S. Paul, “A review on mental stress detection using wearable sensors and machine learning techniques,” *IEEE Access*, vol. 9, pp. 84045–84066, 2021.

[168] L. Abdel-Hamid, “An efficient machine learning-based emotional valence recognition approach towards wearable eeg,” *Sensors*, vol. 23, p. 1255, 2023.

[169] K. Rasheed, J. Qadir, T. OB́rien, L. Kuhlmann, and A. Razi, “A generative model to synthesize eeg data for epileptic seizure prediction,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 2322–2332, 2021.

[170] S.-H. Oh, Y.-R. Lee, and H.-N. Kim, “A novel EEG feature extraction method using Hjorth parameter,” *International Journal of Electronics and Electrical Engineering*, vol. 2, no. 2, pp. 106–110, 2014.

[171] S. Singh and H. Kaur, “An intelligent method for epilepsy seizure detection based on hybrid nonlinear eeg data features using adaptive signal decomposition methods,” *Circuits, Systems, and Signal Processing*, vol. 42, no. 5, pp. 2782–2803, 2023.

[172] T. Haddad, N. Ben-Hamida, L. Talbi, A. Lakhssassi, and S. Aouini, “Temporal epilepsy seizures monitoring and prediction using cross-correlation and chaos theory,” *Healthcare Technology Letters*, vol. 1, pp. 45–50, March 2014.

[173] V. Sip, J. Scholly, M. Guye, F. Bartolomei, and V. Jirsa, “Evidence for spreading seizure as a cause of theta-alpha activity electrographic pattern in stereo-eeg seizure recordings,” *PLOS Computational Biology*, vol. 17, p. e1008731, February 2021.

[174] F. Zubler, A. Rubino, G. Lo Russo, K. Schindler, and L. Nobili, “Correlating interictal spikes with sigma and delta dynamics during non-rapid-eye-movement-sleep,” *Frontiers in Neurology*, vol. 8, p. 288, June 2017.

[175] K. Edakawa, T. Yanagisawa, H. Kishima, R. Fukuma, S. Oshino, H. Khoo, M. Kobayashi, M. Tanaka, and T. Yoshimine, "Detection of epileptic seizures using phase–amplitude coupling in intracranial electroencephalography," *Scientific Reports*, vol. 6, p. 25422, May 2016.

[176] Z. Wang and P. Mengoni, "Seizure classification with selected frequency bands and eeg montages: a natural language processing approach," *Brain Informatics*, vol. 9, p. 11, December 2022.

[177] L. Abou-Abbas, I. Jemal, K. Henni, A. Mitiche, and N. Mezghani, "Focal and generalized seizures distinction by rebalancing class data and random forest classification," in *Bioengineering and Biomedical Signal and Image Processing*, pp. 63–70, Springer, 2021.

[178] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International journal of data mining & knowledge management process*, vol. 5, no. 2, p. 1, 2015.

[179] S. Sinha, M. Brady, C. Scott, and M. Walker, "Do seizures in patients with refractory epilepsy vary between wakefulness and sleep?," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 77, no. 9, pp. 1076–1078, 2006.

[180] M. Hills, "Seizure detection using FFT, temporal and spectral correlation coefficients, eigenvalues and Random Forest," *Github, San Francisco, CA, USA, Tech. Rep*, pp. 1–10, 2014.

[181] H. Abdi and L. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, pp. 433–459, 2010.

[182] E. Combrisson and K. Jerbi, "Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment

of decoding accuracy," *Journal of Neuroscience Methods*, vol. 250, pp. 126–136, 2015.

[183] G. Ramantani, L. Maillard, and L. Koessler, "Correlation of invasive eeg and scalp eeg," *Seizure*, vol. 41, pp. 196–200, 2016.

[184] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *Information Fusion*, vol. 81, pp. 84–90, 2022.

[185] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?," in *Thirty-sixth Conference on Neural Information Processing Systems (NeurIPS 2022) Datasets and Benchmarks Track*, NeurIPS, 2022.

[186] Y. Qi, "Random forest for bioinformatics," in *Ensemble Machine Learning: Methods and Applications*, pp. 307–323, Springer, 2012.

[187] G. Sintotskiy and H. Hinrichs, "In-ear-EEG–a portable platform for home monitoring," *Journal of Medical Engineering & Technology*, vol. 44, no. 1, pp. 26–37, 2020.

[188] M. T. Kucewicz, G. A. Worrell, and N. Axmacher, "Direct electrical brain stimulation of human memory: lessons learnt and future perspectives," *Brain*, vol. 146, no. 6, pp. 2214–2226, 2023.

[189] A. J. McDonald and D. D. Mott, "Functional neuroanatomy of amygdalohippocampal interconnections and their role in learning and memory," *Journal of neuroscience research*, vol. 95, no. 3, pp. 797–820, 2017.

[190] C. S. Inman, M. K. Hollearn, L. Augustin, J. M. Campbell, K. L. Olson, and K. L. Wahlstrom, "Discovering how the amygdala shapes human behavior: From

lesion studies to neuromodulation," *Neuron*, vol. 111, no. 24, pp. 3906–3910, 2023.

[191] C. S. Inman, J. R. Manns, K. R. Bijanki, D. I. Bass, S. Hamann, D. L. Drane, R. E. Fasano, C. K. Kovach, R. E. Gross, and J. T. Willie, "Direct electrical stimulation of the amygdala enhances declarative memory in humans," *Proceedings of the National Academy of Sciences*, vol. 115, no. 1, pp. 98–103, 2018.

[192] C. S. Inman, K. R. Bijanki, D. I. Bass, R. E. Gross, S. Hamann, and J. T. Willie, "Human amygdala stimulation effects on emotion physiology and emotional experience," *Neuropsychologia*, vol. 145, p. 106722, 2020.

[193] M. S. E. Sendi, C. S. Inman, K. R. Bijanki, L. Blanpain, J. K. Park, S. Hamann, R. E. Gross, J. T. Willie, and B. Mahmoudi, "Identifying the neurophysiological effects of memory-enhancing amygdala stimulation using interpretable machine learning," *Brain Stimulation*, vol. 14, no. 6, pp. 1511–1519, 2021.

[194] S. Pahuja, K. Veer, *et al.*, "Recent approaches on classification and feature extraction of eeg signal: A review," *Robotica*, vol. 40, no. 1, pp. 77–101, 2022.

[195] T. S. Davis, R. M. Caston, B. Philip, C. M. Charlebois, D. N. Anderson, K. E. Weaver, E. H. Smith, and J. D. Rolston, "Legui: a fast and accurate graphical user interface for automated detection and anatomical localization of intracranial electrodes," *Frontiers in Neuroscience*, vol. 15, p. 769872, 2021.

[196] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

[197] H. Wen and Z. Liu, "Separating fractal and oscillatory components in the power spectrum of neurophysiological signal," *Brain topography*, vol. 29, pp. 13–26, 2016.

[198] Z. J. Lau, T. Pham, S. A. Chen, and D. Makowski, "Brain entropy, fractal dimensions and predictability: A review of complexity measures for eeg in healthy and neuropsychiatric populations," *European Journal of Neuroscience*, vol. 56, no. 7, pp. 5047–5069, 2022.

[199] R. T. Canolty and R. T. Knight, "The functional role of cross-frequency coupling," *Trends in cognitive sciences*, vol. 14, no. 11, pp. 506–515, 2010.

[200] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, 2005.

[201] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

[202] F. K. Ewald, L. Bothmann, M. N. Wright, B. Bischl, G. Casalicchio, and G. König, "A guide to feature importance methods for scientific inference," in *World Conference on Explainable Artificial Intelligence*, pp. 440–464, Springer, 2024.

[203] T. Harmony, G. Hinojosa, E. Marosi, J. Becker, M. Rodriguez, A. Reyes, and C. Rocha, "Correlation between eeg spectral parameters and an educational evaluation," *International Journal of Neuroscience*, vol. 54, no. 1-2, pp. 147–155, 1990.

[204] O. Dressler, G. Schneider, G. Stockmanns, and E. Kochs, "Awareness and the eeg power spectrum: analysis of frequencies," *British journal of anaesthesia*, vol. 93, no. 6, pp. 806–809, 2004.

[205] W. S. Neo, D. Foti, B. Keehn, and B. Kelleher, "Resting-state eeg power

differences in autism spectrum disorder: a systematic review and meta-analysis," *Translational Psychiatry*, vol. 13, no. 1, p. 389, 2023.

[206] M. Ursino and G. Pirazzini, "Theta–gamma coupling as a ubiquitous brain mechanism: implications for memory, attention, dreaming, imagination, and consciousness," *Current Opinion in Behavioral Sciences*, vol. 59, p. 101433, 2024.

[207] M. Abubaker, W. Al Qasem, K. Pilátová, P. Ježdík, and E. Kvašňák, "Theta-gamma-coupling as predictor of working memory performance in young and elderly healthy people," *Molecular Brain*, vol. 17, no. 1, pp. 1–20, 2024.

[208] M. S. Barr, T. K. Rajji, R. Zomorrodi, N. Radhu, T. P. George, D. M. Blumberger, and Z. J. Daskalakis, "Impaired theta-gamma coupling during working memory performance in schizophrenia," *Schizophrenia research*, vol. 189, pp. 104–110, 2017.

[209] M. Tamura, T. J. Spellman, A. M. Rosen, J. A. Gogos, and J. A. Gordon, "Hippocampal-prefrontal theta-gamma coupling during performance of a spatial working memory task," *Nature communications*, vol. 8, no. 1, p. 2182, 2017.

[210] M. Saeidi, W. Karwowski, F. V. Farahani, K. Fiok, R. Taiar, P. A. Hancock, and A. Al-Juaid, "Neural decoding of eeg signals with machine learning: a systematic review," *Brain Sciences*, vol. 11, no. 11, p. 1525, 2021.

[211] W.-B. Jiang, Y. Wang, B.-L. Lu, and D. Li, "Neurolm: A universal multi-task foundation model for bridging the gap between language and eeg signals," *arXiv preprint arXiv:2409.00101*, 2024.

[212] W. Weng, Y. Gu, S. Guo, Y. Ma, Z. Yang, Y. Liu, and Y. Chen, "Self-supervised learning for electroencephalogram: A systematic survey," *arXiv preprint arXiv:2401.05446*, 2024.

[213] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, *et al.*, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, 2021.

[214] D. Wu, S. Li, J. Yang, and M. Sawan, "neuro2vec: Masked fourier spectrum prediction for neurophysiological representation learning," *arXiv preprint arXiv:2204.12440*, 2022.

[215] W. Cui, W. Jeong, P. Thölke, T. Medani, K. Jerbi, A. A. Joshi, and R. M. Leahy, "Neuro-gpt: Towards a foundation model for eeg," in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5, IEEE, 2024.

[216] E. Shi, K. Zhao, Q. Yuan, J. Wang, H. Hu, S. Yu, and S. Zhang, "Fome: A foundation model for eeg using adaptive temporal-lateral attention scaling," *arXiv preprint arXiv:2409.12454*, 2024.

[217] R. Sharma and H. K. Meena, "Emerging trends in eeg signal processing: A systematic review," *SN Computer Science*, vol. 5, no. 4, pp. 1–14, 2024.

[218] M. Pang, H. Wang, J. Huang, C.-M. Vong, Z. Zeng, and C. Chen, "Multi-scale masked autoencoders for cross-session emotion recognition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.

[219] W. Khan, S. Leem, K. B. See, J. K. Wong, S. Zhang, and R. Fang, "A comprehensive survey of foundation models in medicine," *arXiv preprint arXiv:2406.10729*, 2024.

[220] A. Thomschewski, N. Gerner, P. B. Langthaler, E. Trinka, A. C. Bathke, J. Fell, and Y. Höller, "Automatic vs. manual detection of high frequency oscillations in intracranial recordings from the human temporal lobe," *Frontiers in Neurology*, vol. 11, 2020.

[221] D. Kostas, S. Aroca-Ouellette, and F. Rudzicz, “BENDR: Using Transformers and a Contrastive Self-Supervised Learning Task to Learn From Massive Amounts of EEG Data,” *Frontiers in Human Neuroscience*, vol. 15, no. June, pp. 1–15, 2021.

[222] X. Liu, F. Zhang, Z. Hou, M. Li, Z. Wang, J. Zhang, and J. Tang, “Self-supervised learning: Generative or contrastive,” *Ieee Transactions on Knowledge and Data Engineering*, 2021.

[223] C. Rommel, J. Paillard, T. Moreau, and A. Gramfort, “Data augmentation for learning predictive models on EEG: a systematic comparison,” *Journal of Neural Engineering*, vol. 19, no. 6, 2022.

[224] S. Aroca-Ouellette and F. Rudzicz, “On losses for modern language models,” *arXiv preprint arXiv:2010.01694*, 2020.

[225] D. Jiang, W. Li, R. Zhang, M. Cao, N. Luo, Y. Han, W. Zou, K. Han, and X. Li, “A further study of unsupervised pretraining for transformer based speech recognition,” in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6538–6542, IEEE, 2021.

[226] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*. 2016.

[227] Y. Rho and X. Shao, “Inference for time series regression models with weakly dependent and heteroscedastic errors,” *Journal of Business and Economic Statistics*, 2015.

[228] A. v. d. Oord, Y. Li, and O. Vinyals, “Representation learning with contrastive predictive coding,” *arXiv preprint arXiv:1807.03748*, 2018.

[229] A. Baevski, H. Zhou, A. Mohamed, and M. Auli, “wav2vec 2.0: A framework for self-supervised learning of speech representations,” *Advances in Neural Information Processing Systems*, vol. 2020-Decem, no. Figure 1, pp. 1–12, 2020.

[230] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in *International conference on machine learning*, pp. 1597–1607, PMLR, 2020.

[231] I. Obeid and J. Picone, “The temple university hospital eeg data corpus,” *Frontiers in neuroscience*, vol. 10, p. 196, 2016.

[232] V. Jurcak, D. Tsuzuki, and I. Dan, “10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems,” *Neuroimage*, vol. 34, no. 4, pp. 1600–1611, 2007.

[233] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.

[234] K. Sohn, “Improved deep metric learning with multi-class n-pair loss objective,” *Advances in neural information processing systems*, vol. 29, 2016.

[235] W. Mumtaz, “Mdd patients and healthy controls eeg data (new),” *Figshare, Dataset*, 2016.

[236] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, “Bci competition 2008–graz data set a,” *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, vol. 16, pp. 1–6, 2008.

[237] B. Kemp, “Sleep-EDF Database Expanded.” `https://doi.org/10.13026/C2X676`, 2013.

[238] J.-S. Bang, M.-H. Lee, S. Fazli, C. Guan, and S.-W. Lee, "Spatio-spectral feature representation for motor imagery classification using convolutional neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 7, pp. 3038–3049, 2021.

[239] Y. Song, Q. Zheng, B. Liu, and X. Gao, "Eeg conformer: Convolutional transformer for eeg decoding and visualization," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 710–719, 2022.

[240] C. Zhang, Y.-K. Kim, and A. Eskandarian, "Eeg-inception: an accurate and robust end-to-end neural network for eeg-based motor imagery classification," *Journal of Neural Engineering*, vol. 18, no. 4, p. 046014, 2021.

[241] W. Jiang, L. Zhao, and B. liang Lu, "Large brain model for learning generic representations with tremendous EEG data in BCI," in *The Twelfth International Conference on Learning Representations*, 2024.

[242] H. Bao, L. Dong, S. Piao, and F. Wei, "Beit: Bert pre-training of image transformers," *arXiv preprint arXiv:2106.08254*, 2021.

[243] H. G. Herrema and M. J. Kahana, "Delayed free recall of word lists," 2024.

[244] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[245] Y. Wu and K. He, "Group normalization," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, 2018.

[246] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)," *arXiv preprint arXiv:1606.08415*, 2016.

[247] A. Van Den Oord, O. Vinyals, *et al.*, "Neural discrete representation learning," *Advances in neural information processing systems*, vol. 30, 2017.

[248] L. McInnes, J. Healy, and J. Melville, “Umap: Uniform manifold approximation and projection for dimension reduction,” *arXiv preprint arXiv:1802.03426*, 2018.

[249] R. Shwartz-Ziv and A. Armon, “Tabular data: Deep learning is not all you need,” *Information Fusion*, vol. 81, pp. 84–90, 2022.

[250] N. Cristianini and B. Scholkopf, “Support vector machines and kernel methods: the new generation of learning machines,” *Ai Magazine*, vol. 23, no. 3, pp. 31–31, 2002.

[251] M. Wortsman, G. Ilharco, J. W. Kim, M. Li, S. Kornblith, R. Roelofs, R. G. Lopes, H. Hajishirzi, A. Farhadi, H. Namkoong, *et al.*, “Robust fine-tuning of zero-shot models,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7959–7971, 2022.

[252] W. Wang, V. W. Zheng, H. Yu, and C. Miao, “A survey of zero-shot learning: Settings, methods, and applications,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–37, 2019.