Single-Trial ERSP Classification of EEG Data Across Multiple Axes Using Deep and Machine Learning: Spectrograms vs. Frequency- and Subject-Level Analysis

by

Arian Yavari

A thesis submitted to the
Department of Computer Science
in conformity with the requirements for
the degree of Master of Science

Bishop's University
Canada
May 2025

Abstract

This study presents a novel approach for classifying visual stimuli from single-trial Event-Related Spectral Perturbation (ERSP) responses derived from electroencephalographic (EEG) recordings. To improve the reliability and interpretability of visual Brain-Computer Interfaces (BCIs), we propose a classification approach that systematically evaluates neural responses across five key dimensions: subject variability, frequency bands, trial-level differences, stimulus categories, and post-stimulus time windows.

To address the multi-dimensional nature of ERSP data, both deep learning and classical machine learning approaches were applied to EEG recordings from thirty-one neurologically and psychiatrically healthy participants. Deep learning models—including convolutional neural networks (CNNs) and recurrent architectures such as bidirectional long short-term memory (LSTM) and gated recurrent units (GRUs)—were trained on spectrogram representations of ERSPs. A CNN architecture incorporating dropout and batch normalization achieved validation accuracies of up to 85.57% in binary classification tasks involving high-contrast stimuli and participants with strong gamma-band responses. However, performance declined in multiclass tasks involving all stimuli and participants, with validation accuracy plateauing around 30%, primarily due to substantial inter-subject variability and limited training data.

In parallel, classical machine learning approaches were applied in a subject- and frequency-specific manner. Random Forest classifiers delivered strong and interpretable results, particularly in the gamma frequency range (20–40 Hz), early post-stimulus intervals, and among participants with pronounced gamma activity. Ensemble-based models consistently outperformed support vector machines, logistic regression, and gradient boosting, underscoring the effectiveness of individualized modeling in data-constrained settings.

This study is among the first to systematically investigate ERSP classification across multiple predictive dimensions using both deep spectrogram-based models and subject- and frequency-aware classical approaches. The findings emphasize the complementary strengths of these methods: deep learning performs well with structured, high-resolution inputs when sufficient data are available, while classical approaches offer robustness and interpretability in subject-specific contexts. Future

research should explore hybrid strategies, subject-informed training methods, and the application of self-supervised or transfer learning approaches fine-tuned for task-relevant scenarios to compare and investigate the possibility of improving generalizability across participants and experimental conditions.

Acknowledgments

I extend my deepest gratitude to my supervisor, Dr. Russell Butler, for his invaluable support, guidance, and mentorship throughout the course of my studies. His expertise and encouragement have been instrumental in shaping the outcomes of this research.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Electroencephalography (EEG) is a widely used, noninvasive neuroimaging technique that captures the brain's electrical activity through electrodes placed on the scalp. These signals reflect the synchronized activity of large neuronal populations and are particularly useful in visual neuroscience, where EEG helps decode brain responses to structured visual stimuli such as contrast, luminance, spatial frequency, and object orientation. These visual properties modulate cortical oscillations across distinct frequency bands—especially within the gamma band (30–100 Hz)—and at specific latencies, providing high-resolution temporal insight into perceptual and cognitive processes [56, 115, 140].

Unlike functional MRI, PET, or MEG, EEG offers millisecond-level temporal resolution, is portable, and remains more affordable, making it suitable for real-time applications such as Brain-Computer Interfaces (BCIs) [38, 41]. Despite these strengths, EEG data are often noisy, high-dimensional, and subject-specific, making manual analysis labor-intensive and prone to bias [33]. These challenges necessitate machine learning approaches for automated and scalable interpretation.

Recent advances in machine learning (ML) and deep learning (DL) have revolutionized EEG decoding by capturing complex spatial, temporal, and spectral patterns. While unsupervised and self-supervised methods—such as autoencoders and contrastive learning—are popular for representation learning in low-resource contexts, their efficacy in EEG-based visual classification remains limited. These methods tend to underperform in tasks requiring precise stimulus-response alignment and suffer from reduced interpretability, limiting their clinical and cognitive neuroscience applicability [126, 137].

Supervised learning, on the other hand, excels in scenarios where labeled data are available, offering superior accuracy and model interpretability. By directly learning mappings between stimuli and evoked brain activity, supervised models better capture the neural signatures of visual features such as contrast and spatial layout. Our study adopts a supervised learning approach to decode gamma-band visual responses, demonstrating improved classification and neurophysiological

relevance over unsupervised alternatives [72, 140].

Furthermore, BCIs that rely on EEG for decoding user intent have shown significantly improved performance when powered by supervised ML models. These systems leverage labeled neural data to optimize decoding of visual evoked potentials (VEPs), facilitating adaptive neurotechnologies tailored to individual cognitive and perceptual profiles [5, 11, 45, 83, 125].

EEG-based BCIs have garnered significant attention in healthcare and assistive technology due to their noninvasive nature and real-time capabilities. These systems have been successfully implemented in a range of clinical applications, from neurorehabilitation for stroke patients to communication aids for individuals with severe motor impairments, such as those with amyotrophic lateral sclerosis (ALS) or locked-in syndrome. The ability to capture neural signals and translate them into control signals for assistive devices—such as robotic arms or communication boards—holds immense potential for improving the quality of life and independence of individuals with disabilities. As machine learning and deep learning techniques continue to evolve, the accuracy and adaptability of EEG-based BCIs are expected to improve, allowing for more personalized and efficient interventions.

Recent studies have expanded the scope of EEG-based BCIs in healthcare. For instance, a comprehensive review highlighted their applications across eight critical areas: rehabilitation, daily communication, epilepsy, cerebral resuscitation, sleep, neurodegenerative diseases, anesthesiology, and emotion recognition [136]. This broad applicability underscores the versatility of EEG-based BCIs in addressing diverse medical needs.

Innovations in BCI technology have also led to the development of hybrid systems that integrate EEG with other modalities. A notable example is the hybrid brain-machine interface that combines steady-state visually evoked potential (SSVEP)-based EEG with facial electromyography (EMG). This integration allows for reduced physical demand and improved user experience in assistive applications such as virtual navigation tasks [119].

Furthermore, advancements in wearable BCI devices are making EEG-based systems more accessible and practical for everyday use. Recent progress in wearable EEG-based BCI devices for medical applications has focused on improving signal acquisition methods and enhancing user comfort. These developments aim to make EEG-based BCIs more suitable for continuous monitoring and long-term use in various medical settings [133].

1.1 Problem Statement

While the majority of Brain-Computer Interface (BCI) research has focused on motor imagery tasks, decoding EEG signals elicited by visual stimuli remains a comparatively underexplored domain—despite its increasing relevance for developing non-motor-based BCIs, particularly for individuals with severe motor impairments

[58, 87]. This gap is especially significant given the potential of visually driven BCIs as low-effort, cognitively intuitive alternatives for user interaction.

Recent studies have demonstrated the feasibility of decoding visual content from EEG signals using both classical and deep learning models [14, 130]. However, several challenges persist, including high inter-subject variability, susceptibility to physiological artifacts, and limited generalizability across diverse stimulus types [36, 71]. Moreover, most existing research focuses on isolated feature domains—spectral, spatial, or temporal—without fully exploring their interactions or the impact of subject-specific variability [8, 40].

Visual EEG decoding continues to lag behind motor-based BCIs in both development and application. This discrepancy stems in part from the inherent complexity of visual processing, which involves multi-dimensional spatial and temporal dynamics that are more difficult to capture and model than motor-related signals [66, 105]. Additionally, motor imagery tasks benefit from clearer and more consistent neural signatures, whereas visual tasks are confounded by considerable noise and variability, making it challenging to extract reliable patterns from EEG data [50, 54].

To address these limitations, this study systematically investigates how the brain responds to structured visual stimuli—ranging from contrast-based patterns to randomized designs—and evaluates the extent to which these responses can be reliably classified using both classical machine learning (ML) and deep learning (DL) techniques. Specifically, we aim to identify the most discriminative combinations of features across four analytical axes: spectral (e.g., frequency bands), temporal (e.g., task-locked windows), repetition-based (e.g., trial averaging), and individual-level (e.g., subject-specific variability).

This multi-dimensional analytical framework leverages the complementary strengths of classical ML and spectrogram-based DL models. Unlike conventional approaches that assume uniform neural responses across individuals or employ static frequency band definitions, our method dynamically adapts to both individual neural profiles and stimulus-specific characteristics. This adaptability enhances model interpretability, generalizability, and classification performance.

By identifying the most predictive and robust feature combinations, our study establishes a foundational basis for the development of adaptive and personalized BCIs. The implications are twofold: advancing theoretical understanding of how visual stimuli modulate neural dynamics, and informing the design of practical neurotechnological solutions that utilize EEG as a scalable, non-invasive, and real-time tool for assistive and rehabilitative applications.

Although EEG-based BCIs have made substantial progress in motor control—particularly in applications such as prosthetics and rehabilitation—the decoding of visual perception from EEG signals remains comparatively underdeveloped. This shortfall is primarily due to the challenges of capturing the neural signatures associated with visual attributes like contrast, luminance, and spatial orientation

amid noisy and variable EEG data. Moreover, the spatial and temporal dynamics of visual processing present greater modeling complexity than those of motor activity. Nevertheless, the potential for EEG-based visual decoding in cognitive neuroscience and assistive technology is substantial, especially when paired with advanced ML and DL approaches. These models can deepen our understanding of visual processing in the brain and support the development of new applications for individuals with visual impairments or cognitive disabilities.

1.2 Thesis Objectives

This thesis aims to advance EEG-based Brain-Computer Interface (BCI) classification by systematically analyzing visual stimulus responses in single-trial Event-Related Spectral Perturbation (ERSP) data across five key dimensions: Subjects, Frequencies, Stimuli, Trials, and Timepoints. Furthermore, it compares the performance of deep learning (DL) models with traditional machine learning (ML) approaches. The specific objectives are as follows:

i. To evaluate and optimize baseline deep learning models for spectrogram-based EEG classification, with a focus on hyperparameter tuning, overfitting prevention, and the influence of temporal feature selection on model performance.

ii. To compare the effectiveness of classical machine learning techniques across frequency bands and individual subjects, examining how different combinations of frequency ranges, participant groups, stimulus types, and time segments affect classification accuracy.

iii. To systematically assess the contribution of key features—including subjects, frequency bands, time windows, stimulus properties, and trial repetitions—to overall classification performance, with the goal of identifying the most predictive and generalizable feature sets.

Chapter 2

Literature Review

2.1 Impact of Visual Stimuli on Neural Dynamics and Predictability

Visual stimuli modulate oscillatory EEG dynamics across multiple frequency bands, including gamma (30–100 Hz), alpha (8–13 Hz), theta (4–8 Hz), and delta (1–4 Hz). These frequency-specific responses are closely linked to key visual attributes—such as contrast, spatial frequency, structure (e.g., coherent vs. randomized patterns), motion, and spatial positioning—which shape perceptual encoding and influence the performance of EEG classification pipelines.

High-contrast stimuli reliably enhance gamma-band activity in occipital and parietal cortices, indicating increased neural synchronization and sensory gain [1, 48, 64, 80, 114]. Alpha-band modulation is more complex: while some studies associate increased alpha power with cortical disengagement and inhibitory processes [57, 68], others link it to attentional allocation and perceptual binding. Conversely, randomized or unpredictable stimuli often reduce phase-locked responses while increasing activity in lower frequency bands such as theta and delta—oscillations associated with novelty detection, working memory, and cognitive load [29, 35].

The structural regularity of visual input also plays a critical role. Coherent, repetitive patterns evoke stronger gamma and theta oscillations than incoherent stimuli, suggesting that neural predictability and stimulus regularity enhance cortical entrainment. Similarly, motion stimuli and those with low spatial frequency preferentially elicit gamma-band responses due to the activation of motion-sensitive visual areas, thereby improving signal discriminability for classification tasks [20, 46].

Spatial positioning further modulates EEG dynamics. Stimuli presented in the lower visual field consistently evoke stronger alpha and gamma responses in occipital-parietal regions [35, 68, 114]. Spatial lateralization effects have also been observed: stimuli appearing in distinct visual hemifields or quadrants engage different cortical regions. For example, linguistic and symbolic stimuli predominantly

activate left-lateralized networks, whereas spatial or emotionally salient stimuli elicit stronger right-hemispheric responses [139]. These findings underscore the importance of integrating spatial topography into EEG decoding frameworks.

Collectively, these observations highlight the critical role of stimulus-dependent, frequency-specific, and spatially localized oscillatory dynamics in shaping visual EEG responses. Accounting for these multidimensional neural signatures can substantially enhance the sensitivity, specificity, and generalizability of EEG-based classification models.

2.2 Single-Trial ERSPs and Spectrogram-Based Time-Frequency Analysis

Single-trial analysis of visual evoked EEG responses has garnered increasing interest due to its relevance for brain-computer interface (BCI) applications and individualized neural decoding. Among the most informative approaches for characterizing neural dynamics in this context are event-related spectral perturbations (ERSPs) and spectrogram-based time-frequency representations (TFRs), which capture stimulus-induced changes in oscillatory activity across time and frequency domains.

Traditionally, ERSPs are computed by averaging spectral responses across trials, effectively revealing phase-locked activity patterns consistent across repetitions. While this enhances signal-to-noise ratios, it also attenuates transient, non-phase-locked dynamics and masks inter-trial variability—features often critical for understanding cognitive variability and enabling single-trial classification [52, 74, 142].

In contrast, spectrogram-based TFRs preserve trial-specific temporal and spectral details, enabling the analysis of both evoked (phase-locked) and induced (non-phase-locked) components of the EEG signal [3, 59, 110, 138]. These representations are particularly advantageous for decoding visual responses at the single-trial level, where individual neural variability encodes task-relevant information [63, 117].

Recent advances in adaptive time-frequency analysis—such as wavelet transforms, chirplet decompositions, and empirical mode decomposition (EMD)—have further enhanced the sensitivity of single-trial analyses. These methods effectively capture transient and localized spectral features, especially in non-stationary frequency bands associated with visual attention and perception, including alpha and gamma [24, 34, 62, 134, 143].

Beyond temporal and spectral domains, multi-channel spectrogram analysis facilitates the exploration of spatial dynamics in visual EEG data. This includes identifying lateralized cortical activations and region-specific frequency responses, thus supporting more interpretable and spatially informed decoding models [13, 106, 108].

The integration of machine learning techniques with spectrogram-based and

ERSP analyses has further improved decoding performance by leveraging the high-dimensional, time-frequency structure of EEG signals. These approaches contribute to more robust and generalizable classification across subjects and stimulus types [75, 90].

2.3 Single-Trial ERSPs and Spectrogram-Based EEG Classification Using Machine Learning and Deep Learning

Classifying single-trial EEG responses to visual stimuli remains a major challenge in neuroscience and brain-computer interface (BCI) research, primarily due to high trial-to-trial variability and low signal-to-noise ratios. To address these challenges, researchers increasingly employ time-frequency representations—such as event-related spectral perturbations (ERSPs) and spectrograms—alongside machine learning (ML) and deep learning (DL) methods to extract discriminative features from EEG data.

ERSPs capture transient oscillatory activity across frequency bands following visual stimuli, preserving both temporal and spectral dynamics at the single-trial level. Feature extraction commonly involves time-frequency decomposition techniques, including wavelet transforms and short-time Fourier transforms (STFT), yielding high-dimensional ERSP maps. These are often vectorized or reduced using dimensionality reduction methods such as principal component analysis (PCA) or t-distributed stochastic neighbor embedding (t-SNE), which help retain informative variance while mitigating redundancy [73, 121].

Classical ML algorithms—including support vector machines (SVM), random forests (RF), and ensemble classifiers—have been widely applied to these ERSP-derived features [61, 123, 132]. These models support subject-specific classification and enable the identification of frequency bands most predictive of visual perception. Random forests, in particular, offer interpretability through feature importance scores, facilitating per-frequency and per-subject performance analysis.

Deep learning models, particularly convolutional neural networks (CNNs), are well-suited for learning hierarchical spatial and temporal features directly from spectrograms or ERSP heatmaps. These models obviate the need for manual feature engineering and have demonstrated strong performance in decoding visual stimuli from raw or minimally processed EEG inputs [49, 53, 96]. When augmented with attention mechanisms or transformer architectures, CNNs can localize salient time-frequency regions and improve interpretability and generalization [127, 128].

Hybrid deep learning architectures—such as EEG-ConvTransformers and attention-augmented CNNs—further enhance accuracy by combining local feature extraction with global context modeling. These models are particularly effective in capturing stimulus-specific temporal dynamics and are adaptable across subjects via fine-tuning or domain adaptation strategies [16, 65, 101].

While supervised learning remains the dominant paradigm in single-trial EEG classification, recent advancements in self-supervised learning (SSL) offer promising alternatives for scenarios with limited labeled data. SSL approaches, such as contrastive learning and autoencoders, aim to learn robust EEG representations without requiring explicit stimulus labels [4, 122]. Nonetheless, supervised models continue to outperform SSL in visual classification tasks due to their superior capacity for learning task-specific discriminative features [131, 141, 146].

Interpretability remains a critical issue in deploying ML and DL models to EEG data. Techniques such as class activation mapping (CAM), saliency maps, and layer-wise relevance propagation (LRP) are increasingly employed to visualize which time-frequency regions contribute most to classification decisions [51, 90]. These tools enhance transparency and support the translation of model predictions into neuroscientific insights.

Chapter 3

Materials and Methods

This chapter provides a comprehensive overview of the EEG recordings used for analysis, including the design and presentation of visual stimuli, participant demographics, trial structure, and data acquisition protocols. It describes the preprocessing steps applied to the raw EEG data, outlines the overall dataset structure, and details the classification models employed. Both classical machine learning and deep learning models are presented, along with the rationale for their selection, training configurations, and hyperparameter settings. The software frameworks and analytical tools used throughout the pipeline are also specified.

To facilitate interpretation of the data, representative plots of averaged event-related spectral perturbations (ERSPs) across subjects and stimulus types are provided. Single-trial classification of ERSP responses presents a complex, high-dimensional challenge that requires careful consideration of multiple factors—including inter-subject physiological variability, frequency band interactions, stimulus characteristics, trial-to-trial fluctuations, and the temporal alignment of EEG signals with stimulus onset. These complexities necessitate the use of diverse modeling approaches, including both classical machine learning and deep learning methods, and the incorporation of multiple input representations. Such strategies aim to capture both shared and subject-specific neural dynamics over time in response to various visual stimuli.

3.1 Participants, Stimuli, EEG Acquisition, and Preprocessing

This study utilized EEG recordings collected in accordance with the ethical guidelines of the Centre Hospitalier Universitaire de Sherbrooke's Internal Review Board. Participants had normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. Visual stimuli were generated using the Psychophysics Toolbox [9] and presented on a uniform gray background with luminance matched

to the mean stimulus luminance. A baseline grating drifted horizontally within a 7° circular aperture at a spatial frequency of 3 cycles per degree and a temporal frequency of 6 cycles per second. Michelson contrast was varied at three levels—100%, 33%, and 5%—by adjusting brightness ranges while maintaining consistent mean luminance. Spatial randomization was introduced by randomly exchanging $0.15° \times 0.15°$ patches until contrast levels of 10% and 60% were achieved, without altering overall contrast [111, 144].

Stimuli were displayed on a CRT monitor with a resolution of 800×600 pixels and a refresh rate of 85 Hz. Participants were seated comfortably in a dimly lit room. The session comprised nine blocks, each lasting 8.5 minutes, with 1-minute rest breaks between blocks to maintain participant alertness and data quality. Each participant completed 810 trials. Six distinct stimulus types were presented 135 times in a pseudo-randomized order. Each trial began with a fixation cross that turned red, followed by a jittered delay between 0.5 and 1 second, and a 2-second grating presentation. After the stimulus presentation, the fixation cross returned to black, signaling a 2-second rest period. Additional 8-second rest intervals were introduced every two minutes to allow for blinking and posture adjustments, minimizing muscle-related artifacts.

EEG data were recorded using a 64-channel Brain Products system, with electrode placement following the international 10-20 system. The ground electrode was placed between FPz and Fz, while the reference electrode was placed between CPz and Cz. Data were sampled at 500 Hz and subsequently downsampled to 256 Hz for computational efficiency. Channels with poor signal quality were identified using z-score thresholds and interpolated. Continuous EEG recordings from the 90-minute session were bandpass filtered between 1 and 120 Hz. Preprocessing was conducted using EEGLAB [23] and custom MATLAB scripts. Independent component analysis (ICA) was applied to remove artifacts, with the guidance of a neuronal response function extracted from posterior electrodes filtered in the 10–25 Hz and 40–110 Hz bands. Components exhibiting the highest Pearson correlation with this response function were retained, and visual inspection confirmed their occipital origin [43, 85].

For spectral analysis, EEG trials were epoched from -0.85 seconds to 2.85 seconds relative to stimulus onset. Event-related spectral perturbations (ERSPs) were computed for the interval between 500 milliseconds and 2000 milliseconds following stimulus presentation. Further preprocessing steps, including baseline correction and spectrogram generation, were implemented using custom Python scripts.

The visual stimuli were designed to assess the effects of luminance contrast and spatial-temporal randomization on EEG responses. Grayscale images were normalized for global luminance to ensure consistency across conditions. Six experimental conditions were used:

- **Contrast Levels**: High (100%), Medium (33%), and Low (5%)—designed to modulate the salience of the visual input and the corresponding neural

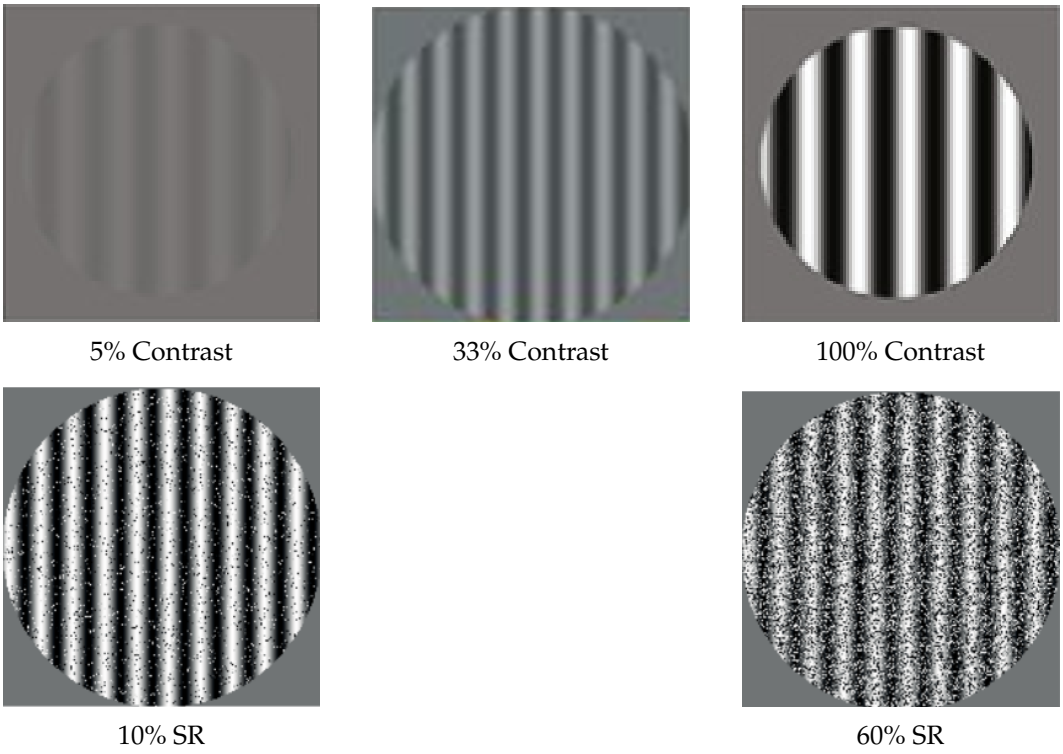


Figure 3.1: Visual stimuli: Contrast levels (top row) and Spatial Randomization (SR) levels (bottom row).

response strength.

- **Randomization Levels**: None (0%), Low (10%), and High (60%)—used to control the degree of spatial structure disruption in the images, assessing sensitivity to spatial predictability.

Stimuli used during recordings are presented in Figure 3.1.

3.2 Dataset Structure

This section provides detailed information about the dataset structure used as input for both machine learning (ML) and deep learning (DL) models. The ML models were trained on hand-crafted, feature-engineered datasets, while the DL models utilized spectrogram representations derived from single-trial ERSPs.

The initial single-trial ERSP dataset, gathered from previously recorded EEG data, had the shape `(subjects, stimuli, trials, frequencies, time points)` with the following parameters:

- **Stimulus conditions**: Six types, varying by contrast (100%, 33%, 5%) and spatial randomization (0%, 10%, 60%).
- **Trials**: 135 trials per stimulus type, totaling 810 trials per participant.
- **Frequencies**: 60 frequency bands ranging from 1 to 120 Hz in 2 Hz intervals.
- **Time points**: 350 time points per trial.

The input data for the DL models consisted of spectrograms represented as arrays of shape `(60, 350)`, where the first axis corresponds to frequency values and the second axis represents time points. Spectrograms were computed separately for each trial, stimulus type, and subject. In total, 31 participants with 6 stimulus types and 135 trials per type yielded $31 \times 6 \times 135$ spectrograms using this method.

For the machine learning analysis, the dataset had to be reshaped further, as most machine learning models require two-dimensional input. In the case of subject-level analysis (targeting individual-level responses), the remaining factors—stimuli and trials—formed the first dimension of the input, while frequencies and time points were flattened into the second dimension. For frequency-level analysis (capturing shared responses), the first dimension included subjects, stimuli, and trials, and the second dimension included only the time points. The exception to this structure was the per-subject, per-frequency analysis, where the first dimension included stimuli and trials, and the second dimension consisted of time points only.

Across both approaches, participants exhibiting strong gamma-band responses across stimulus types and trials were identified as 1, 4, 8, 9, 10, 12, 13, 14, 18, 20, 22, 23, 24, 25, 27, and 31. These individuals, identified based on the ERSP plots shown in Figure 3.3, were of particular interest, especially in response to 100% and 5% contrast stimuli, which represented the strongest and weakest levels of visual stimulation across the gamma and alpha frequency bands.

3.3 Deep Learning Models

Different deep learning model architectures were tested on spectrograms of single-trial ERSP data, including fully connected feedforward networks, CNNs, LSTMs, GRUs, and SimpleRNN structures. These models were selected to account for model simplicity (fully connected networks), the ability to extract low-level features from images (CNNs), and the capacity of recurrent neural networks to capture and memorize temporal dependencies. Additionally, hidden layer activations (ReLU and ELU) were tested, along with dropout, L1/L2 regularization, and MaxNorm kernel constraints to prevent overfitting. A batch size of 32 was used, and early stopping was employed to further mitigate the risk of overfitting. The optimizers employed were Adam and RMSprop, with learning rate schedules designed to fine-tune the training process. The models were implemented using TensorFlow version 2.10, employing both subclassing and the functional API.

The data were split into training (70%) and validation (30%) sets. All models were trained for 30 epochs to ensure adequate model convergence and to assess performance across various stages of training. The choice of 30 epochs was made to balance training duration and computational efficiency, as early stopping was applied to halt training when performance on the validation set stopped improving. This strategy mitigates overfitting and ensures that the model doesn't overtrain on the data.

Transformers were not included in this study due to the limited size of our dataset. Despite their success in various domains, transformer models typically require large amounts of data to effectively capture complex patterns and avoid overfitting. In contrast, our dataset was insufficient to support the training of such models without the risk of poor generalization. Recent studies have highlighted similar challenges when applying transformer-based models to EEG data, particularly in contexts with small datasets. For example, a study by [116] discusses the limitations of transformers in EEG analysis, emphasizing the need for large-scale datasets to achieve optimal performance in classification tasks. Therefore, to ensure robust and generalizable results, we opted for alternative deep learning architectures better suited to the constraints of our data.

The selection of models and hyperparameters was guided by recent research in EEG classification tasks and general deep learning best practices:

- **Fully Connected Feedforward Networks**: These models serve as a baseline and are useful for initial comparisons, providing insights into the basic capacity of deep learning for EEG analysis.
- **Convolutional Neural Networks (CNNs)**: CNNs are effective for spatial feature extraction, which is crucial when working with spectrograms derived from EEG data. CNNs capture frequency-domain patterns that are important for decoding visual stimuli [103].
- **Long Short-Term Memory (LSTM) Networks**: LSTMs are ideal for sequential data and are particularly beneficial for capturing long-term dependencies, which are crucial for EEG signals that exhibit temporal dynamics [124].
- **Gated Recurrent Units (GRUs)**: GRUs are simpler than LSTMs but still perform well on sequential data, providing a more computationally efficient alternative for temporal feature learning [15].
- **SimpleRNNs**: Although less effective in capturing long-range dependencies, SimpleRNNs were tested as a baseline model to assess the advantage of more advanced recurrent architectures.

The choice of hyperparameters is as follows:

- **Activation Functions (ReLU and ELU)**: ReLU is widely used due to its simplicity and effectiveness in avoiding vanishing gradients. ELU is tested for its ability to allow negative activations, which can help improve learning dynamics [18].
- **Dropout**: Dropout is a regularization technique that helps prevent overfitting by randomly setting a fraction of input units to zero during training [109].
- **L1/L2 Regularization**: These regularization techniques help reduce overfitting by adding penalty terms to the loss function [82].
- **MaxNorm Kernel Constraint**: This constraint helps stabilize training by limiting the weight values, which can lead to better generalization [6].
- **Batch Size (32)**: A batch size of 32 is commonly used, providing a balance between efficient training and stable optimization [6].
- **Early Stopping**: Early stopping prevents overfitting by halting training when performance on a validation set stops improving [88].
- **Optimizers (Adam and RMSprop)**: Adam is an adaptive optimizer that works well in various contexts, including EEG classification, while RMSprop is effective for non-stationary data like EEG signals [42, 55].
- **Learning Rate Schedulers**: Learning rate schedules adjust the learning rate during training, helping the model converge more efficiently and avoiding overshooting minima [70].
- **Epochs (30)**: The models were trained for 30 epochs to ensure sufficient convergence and to monitor performance stability, with early stopping employed to prevent overfitting and ensure generalization.

Deep Learning model architectures used for analysis are as follows:

- **3-Layer Dense (Binary Classification, Multiclass Classification)**: 128-32 units, ReLU activation, last dense layer with value of 1 or 6 depending on the classification type.
- **5-Layer Dense (Binary Classification, Multiclass Classification)**: 512-256-128-32 units, last dense layer with value of 1 or 6 depending on the classification type, ReLU activation. A regularized form using L1/L2 regularization, MaxNorm constraints, and dropout layers of (0.4, 0.3, 0.2).
- **6-Layer Dense (Binary Classification, Multiclass Classification)**: 1024-512-256-128-32 units, last dense layer with value of 1 or 6 depending on the classification type, ReLU activation.

- **3-Layer CNN (Multiclass Classification)**: 32-64 filters, max pooling, last dense layer with value of 1 or 6 depending on the classification type.
- **4-Layer CNN (Binary Classification, Multiclass Classification)**: 32-64-128 filters, last dense layer with value of 1 or 6 depending on the classification type, batch normalization, ReLU activation, dropout (0.1-0.15-0.35).
- **8-Layer SimpleRNN (Binary Classification, Multiclass Classification)**: 64-128-256 units of SimpleRNN layers, followed by 128-64-32-16 units of dense layers, tanh activation, last dense layer with value of 1, dropout values of 0.2 and 0.3.
- **3-Layer LSTM (Binary Classification, Multiclass Classification)**: 64-32 units, tanh activation, last dense layer with value of 1.
- **4-Layer LSTM (Binary Classification)**: 128-64-32 units, tanh activation, last dense layer with value of 1.
- **3-Layer GRU (Binary Classification, Multiclass Classification)**: 64-32 units, tanh activation, last dense layer with value of 1.

The classification accuracies of the deep learning models and detailed hyperparameters are explained in the results section for the deep learning approach.

3.4 Machine Learning Models

Several machine learning models were initially tested to address the multi-class classification task across multiple dimensions of the dataset. The subset used for this evaluation comprised data from half of the subjects, all stimulus types, the first half of the trials, all frequency bands, and the first quarter of the time points. Both One-vs-Rest and One-vs-One classification schemes were employed to accommodate the nature of the models [2, 22]. Among all tested models, the Random Forest classifier achieved the highest validation accuracy on this data subset and was subsequently chosen for further investigation.

The full suite of evaluated models included: Support Vector Machines (SVM) [19], Random Forests (RF) [10], Logistic Regression (LR) [37], Gradient Boosting [26], HistGradient Boosting [98], Gaussian Process Classifier [91], Linear SVC [12], Stochastic Gradient Descent (SGD) Classifier [93], Perceptron [94], Passive-Aggressive Classifier [21], Ridge Classifier with Cross-Validation [44], Quadratic Discriminant Analysis (QDA) [78], Nearest Centroid [100], Multi-layer Perceptron (MLP) [7], Label Spreading [145], Gaussian Naive Bayes [25], Extra Trees Classifier [32], Decision Tree Classifier [89], and Bernoulli Naive Bayes [77]. The hyperparameters for each model were set to the default values as specified in scikit-learn version 1.6.1 [86].

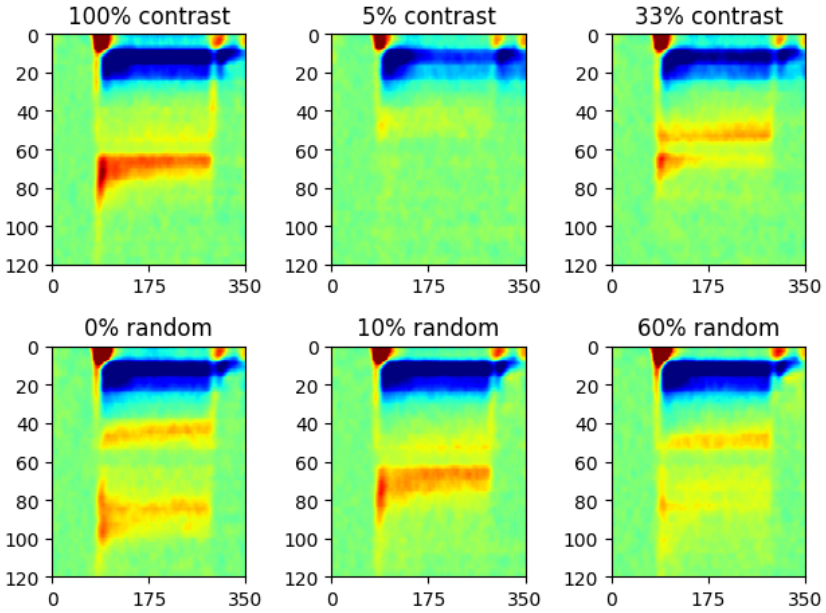


Figure 3.2: Averaged ERSP responses to different stimuli type across all subjects

To incorporate temporal dynamics into the classification pipeline, several strategies were applied at both the per-subject and per-frequency levels: (1) using all time points, (2) averaging across all time points, (3) averaging the first and second halves of the time series, and (4) segmenting time points into distinct intervals. These approaches enabled a comprehensive analysis of the temporal evolution of neural activity, providing detailed insights into how visual stimuli modulate brain responses over time [79]. For each case, the classifier was trained on 70% of the data and validated on the remaining 30%.

The classification accuracies for the Random Forest model under different conditions for per-subject and per-frequency splits are discussed in the results section for the machine learning approach.

To visualize the consistent changes in power in frequency bands during recording, the averaged baseline-corrected ERSP is displayed for each stimulus type, which are as shown in Figure 3.2. Starting from top left to bottom right, the stimuli are 100% contrast, 5% contrast, 33% contrast, 0% random, 10% random, and 60% random, respectively. Figure 3.3 shows averaged ERSP activity across all stimuli for each subject.

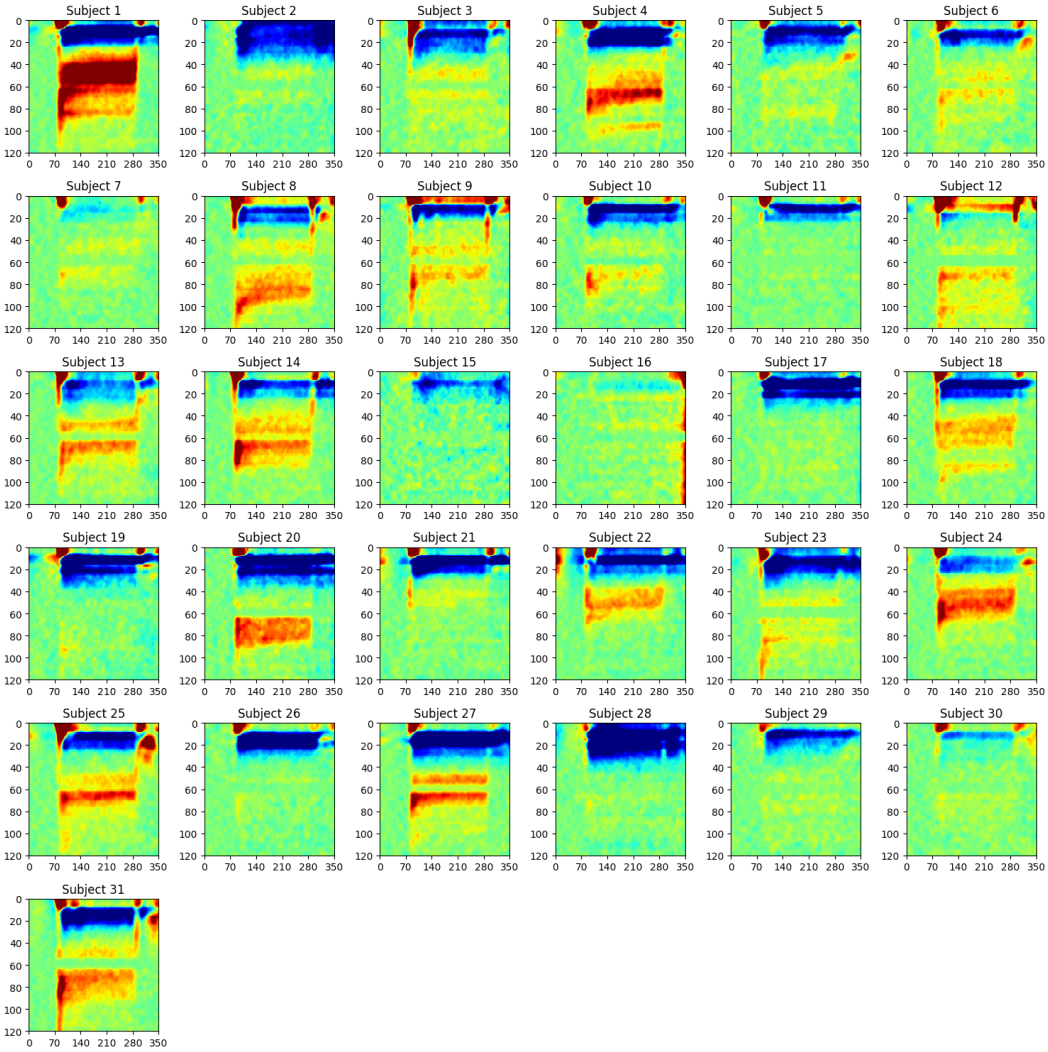


Figure 3.3: Averaged ERSP responses to all stimuli per subject

Chapter 4

Results

This chapter presents the analysis results of deep learning in section 4.1 and classical machine learning in section 4.2 approaches. Deep learning models applied to spectrograms were divided into two main categories: (1) selected subjects with strong gamma responses exposed to 100% and 5% contrast stimuli, referred to as the binary classification task in the following context; and (2) all subjects across all stimulus types, referred to as the multiclass classification task. For the deep learning approach, a subset of five figures is presented in section 4.1, each corresponding to the deep learning model that achieved the highest validation accuracy among those listed in Table 4.1. Each figure contains two plots: one showing training and validation accuracy, and the other showing training and validation loss. Figure 4.1 shows the 5-Layer Dense model for the binary classification task; Figure 4.2 presents the 5-Layer Dense model for the multiclass classification task; Figure 4.3 displays the 4-Layer CNN model for the binary classification task; Figure 4.4 illustrates the 4-Layer regularized CNN model for the binary classification task; and Figure 4.5 shows the 4-Layer LSTM model for the binary classification task.

The machine learning approach investigates classification performance on a per frequency and per subject basis, assessing the impact of feature combinations within this multi dimensional problem space. Figures 4.6 to 4.14 illustrate the validation accuracy of the random forest model for various feature combinations at both the per subject and per frequency levels. These results are discussed in greater detail in the Machine Learning Approach section 4.2.

4.1 Deep Learning Approach

Each single trial event related spectral perturbation (ERSP) yields a spectrogram of size (60,350), where each row represents a specific frequency band and each column corresponds to a time point in the recording. These spectrograms serve as input features for the deep learning classification models described in the Deep Learning Models section of the Materials and Methods chapter. As noted earlier,

Architecture	Validation Accuracy
3-Layer Dense NN	82.33% (Binary Classification), 31.59% (Multiclass Classification)
5-Layer Dense NN	83.56% (Binary Classification, Regularized), 83.10% (Binary Classification), 31.53% (Multiclass Classification)
6-Layer Dense NN	82.56% (Binary Classification), 30.94% (Multiclass Classification)
3-Layer CNN	28.91% (Multiclass Classification)
4-Layer CNN	81.56% (Binary Classification), 85.57% (Binary Classifcation, Regularized), 29.27% (Multiclass Classification)
3-Layer LSTM	75.85% (Binary Classification), 79.55% (Binary Classification, Bidirectional)
4-Layer LSTM	81.64% (Binary Classification)
3-Layer GRU	79.09% (Binary Classification, Bidirectional)
8-Layer SimpleRNN	48.84% (Binary Classification, Regularized)

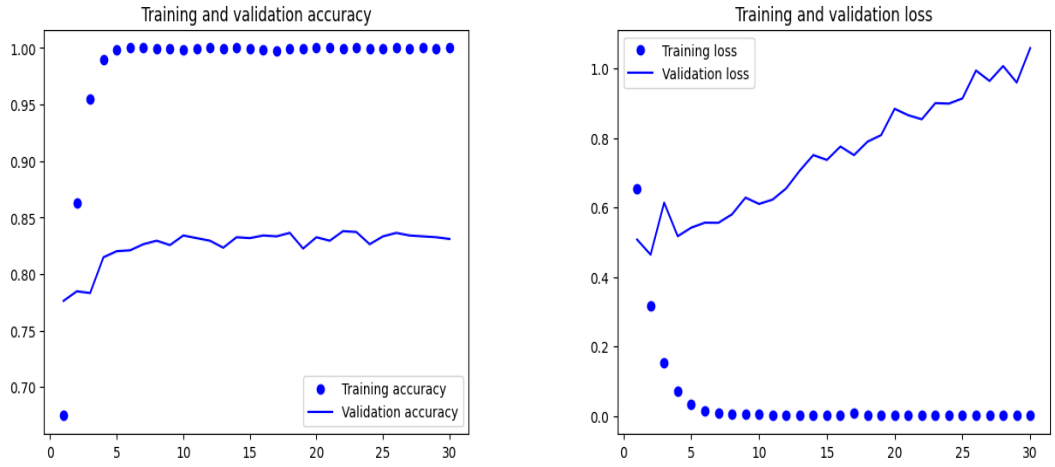
Table 4.1: Summary of grouped architectures and their validation accuracies.

two distinct datasets were constructed. The first dataset comprises trials from all subjects across six stimulus conditions: 100% contrast, 5% contrast, 33% contrast, 0% random (plaid), 10% random, and 60% random, and is treated as a multiclass classification problem to distinguish among these conditions using the full range of trials, frequency bands, and time points. The second dataset includes only participants with strong gamma band responses and contains trials from the 100% and 5% contrast conditions, and is treated as a binary classification task. For both tasks, the primary evaluation metric was validation accuracy, supplemented by training accuracy and loss metrics.

Among the models tested, a simple feedforward network the 5-Layer regularized Dense model achieved satisfactory validation accuracy with low computational cost. This model was selected for hyperparameter tuning using Keras Tuner [84], with the tuning search space defined as described in the Materials and Methods section. However, the highest validation accuracy of 85.57% was achieved by a 4-Layer regularized CNN model.

The remainder of this section presents a summary of the models evaluated and their corresponding validation accuracies.

The 5-Layer Dense model comprises fully connected layers of sizes 512, 256, 128, 32, and 1. It uses ReLU activations between layers and a sigmoid activation in the final layer for binary classification of strong gamma band responders to 5% and 100% contrast stimuli. As shown in Figure 4.1, the model exhibits clear signs



Accuracy metric across epochs Loss metric across epochs

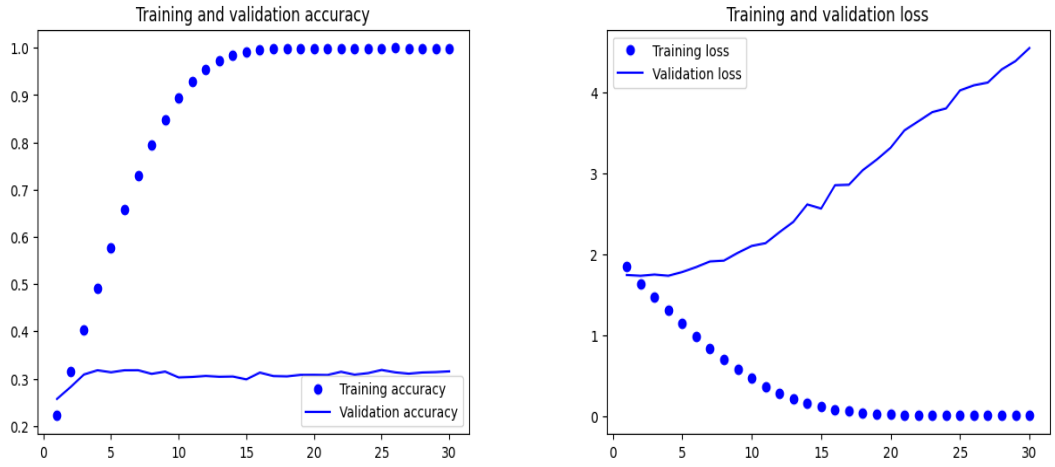
Loss: 8.6518e−04, Accuracy: 1, Validation Loss: 1.0583, Validation Accuracy: 0.8310

Figure 4.1: Performance of the 5-Layer Dense Model on Selected Subjects with strong gamma response

of overfitting: the training loss rapidly drops to zero with perfect (100%) accuracy, while the validation loss steadily increases and validation accuracy plateaus. This behavior indicates that the model memorizes the training data but fails to generalize to unseen examples.

The 5-Layer Dense model comprises fully connected layers with sizes 512, 256, 128, 32, and 6, designed for multiclass classification across all subjects and stimulus types. ReLU activations are applied between layers, and a softmax activation in the final layer predicts one of the six stimulus classes. Figure 4.2 illustrates that the model quickly achieves near perfect training accuracy; however, the validation loss steadily increases, and validation accuracy plateaus at approximately 31%. This behavior indicates significant overfitting: the model memorizes the training data but fails to generalize to unseen examples.

The 4-Layer CNN model begins with a Conv2D layer consisting of 32 filters with a kernel size of 3 and ReLU activation, followed by a MaxPooling2D layer to reduce spatial dimensions. This is succeeded by two additional Conv2D layers with 64 and 128 filters, respectively, each employing ReLU activation to extract progressively more complex features. The resulting feature maps are then flattened, and the network concludes with a dense output layer containing a single neuron with sigmoid activation for binary classification tasks. This compact architecture balances effective feature extraction with computational efficiency for selected subjects with strong gamma response to 5% and 100% contrast stimuli. As shown in Figure 4.3, the model exhibits steadily increasing training accuracy, reaching 99.5%



Accuracy metric across epochs Loss metric across epochs

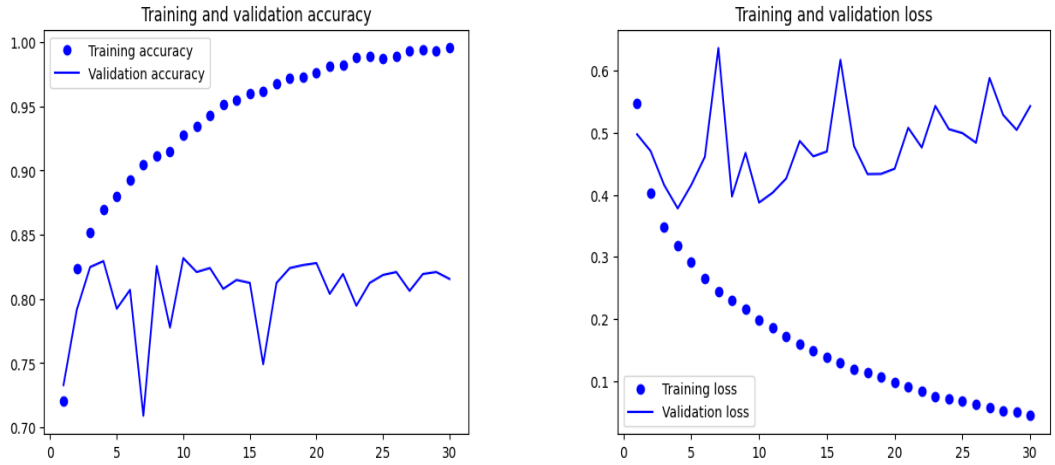
Loss: 4.7e−2, Accuracy: 0.9991, Validation Loss: 4.5485, Validation Accuracy: 0.3153

Figure 4.2: Performance of the 5-Layer Dense Model on All Subjects Across All Stimuli

by epoch 30; however, validation accuracy plateaus around 81–83%, and validation loss increases, indicating overfitting and limited generalization capability.

The 4-Layer regularized CNN model begins with a Conv2D layer containing 32 filters, a kernel size of 3, and same padding to preserve spatial dimensions, followed by a BatchNormalization layer to stabilize and accelerate training. A ReLU activation function introduces nonlinearity, and a Dropout layer with a rate of 0.1 helps mitigate overfitting. A MaxPooling2D layer with a pool size of 2 then reduces the spatial resolution of the feature maps. This pattern is repeated with a Conv2D layer comprising 64 filters, followed by BatchNormalization, ReLU activation, a Dropout rate of 0.15, and another MaxPooling2D layer. A third convolutional block with 128 filters follows the same sequence, incorporating a higher dropout rate of 0.35 for stronger regularization. The resulting feature maps are flattened, and a Dropout layer with a rate of 0.5 is applied before the final output layer. The model concludes with a Dense layer comprising a single neuron and sigmoid activation, making it suitable for binary classification of selected subjects under 5% and 100% contrast visual stimuli.

This architecture emphasizes training stability and generalization by incorporating progressive dropout and normalization across layers, offering a balanced tradeoff between depth, expressiveness, and regularization. The model exhibited consistent improvement, reaching 92.2% training accuracy by epoch 30, while validation accuracy peaked at 85.6% with a corresponding loss of 0.3554, indicating effective learning and minimal overfitting. The use of progressive dropout and



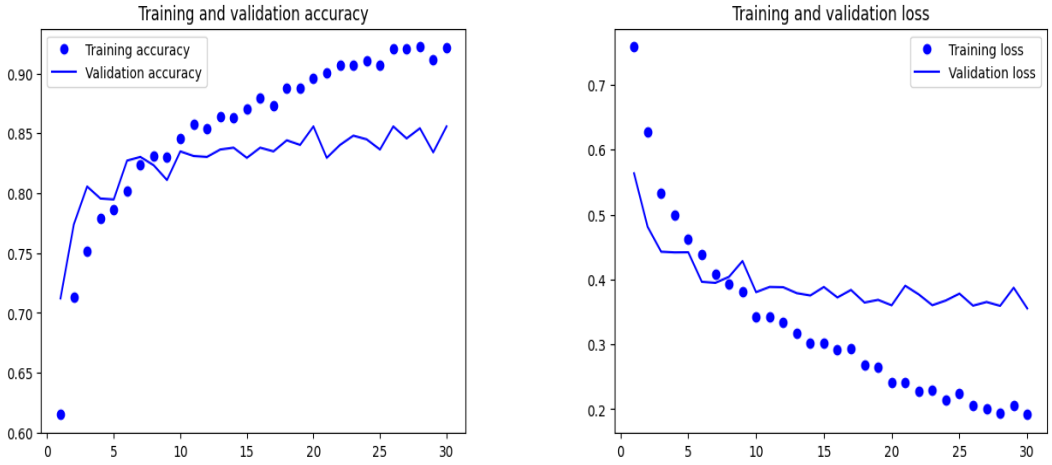
Accuracy metric across epochs Loss metric across epochs

Loss: 4.36e−2, Accuracy: 0.9954, Validation Loss: 0.5431, Validation Accuracy: 0.8156

Figure 4.3: Performance of the 4-Layer CNN Model on Selected Subject with Strong Gamma Response

normalization techniques helped the model generalize better. Model performance is illustrated in Figure 4.4.

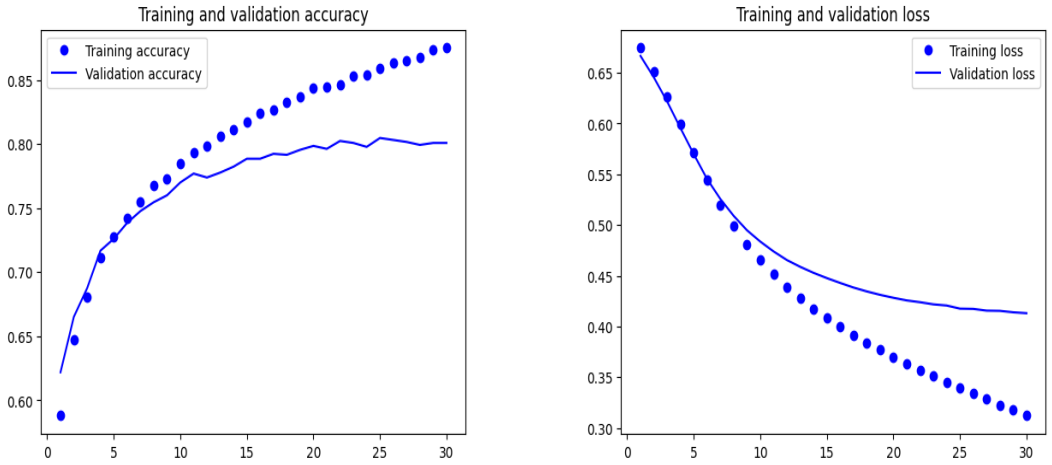
The 4-Layer LSTM model employs a sequential architecture composed of three stacked LSTM layers to capture complex temporal dependencies in EEG signals. It begins with an LSTM layer of 128 units configured with an input shape of (350,60) and `return_sequences=True` to preserve the full temporal structure across time steps. A second LSTM layer with 64 units and a third with 32 units follow, both maintaining the sequence format. The output is then flattened into a single vector and passed to a `Dense` layer with one neuron and sigmoid activation, making it suitable for binary classification of visual stimuli across selected subjects at 5% and 100% contrast levels. Unlike earlier designs that relied on bidirectional or handcrafted recurrent structures, this model emphasizes progressive hierarchical abstraction of temporal features through depth and sequence retention. During training over 30 epochs, the model showed steady learning progression: training accuracy increased from 48.8% to 86.1%, and validation accuracy rose from 53.5% to 81.6%. Meanwhile, validation loss decreased from 0.6908 to 0.4177, suggesting effective generalization without clear signs of overfitting. This performance is illustrated in Figure 4.5.



Accuracy metric across epochs Loss metric across epochs

Loss: 1.916e−1, Accuracy: 0.9216, Validation Loss: 0.3554, Validation Accuracy: 0.8557

Figure 4.4: Performance of the regularized 4-Layer CNN Model on Selected Subjects



Accuracy metric across epochs Loss metric across epochs

Loss: 3.261e−1, Accuracy: 0.8611, Validation Loss: 0.4177, Validation Accuracy: 0.8164

Figure 4.5: Performance of the 4-Layer LSTM Model on Selected Subjects with Strong Gamma Response

4.2 Machine Learning Approach

The classification of baseline corrected single trial ERSP data was formulated as a multidimensional multitask problem, requiring analysis along two axes: frequency level (to assess shared patterns across subjects) and subject level (to capture individual neural responses). A broad range of classical machine learning models was first applied to the multiclass classification task. This section presents results for various feature combinations in both frequency level and subject level classification. These analyses employed the Random Forest model alongside other models tested on an initial representative subset of the dataset, as described in the Materials and Methods section.

The following sections detail the outcomes of the frequency level and subject level analyses.

4.2.1 Subject-level Analysis

For this analysis, classification was performed at the subject level, training a separate model for each individual's single trial ERSP data. This within subject validation ensured that accuracy metrics reflected each model's ability to generalize to unseen data from the same individual. During training, the classifier incorporated variation across stimulus types, trials, frequency bands, and time points, enabling it to learn subject specific patterns while accounting for multidimensional variability within the same individual's data. The conditions tested were as follows:

i. All stimuli, all trials, all frequencies, all time points.

ii. Stimuli: '100% contrast', '5% contrast', '33% contrast', all trials, all frequencies, all time points.

iii. Stimuli: '0% random', '10% random', '60% random', all trials, all frequencies, all time points.

iv. Per frequency, stimuli: '100% contrast', '5% contrast', '33% contrast', all trials, all time points.

v. Per frequency, stimuli: '0% random', '10% random', '60% random', all trials, all frequencies, all time points.

vi. Per frequency, stimuli: '100% contrast', '5% contrast', '33% contrast', all trials, time points segmented into (0–0.33, 0.33–0.66, 0.66–1.0).

vii. Per frequency, stimuli: '0% random', '10% random', '60% random', all trials, time points segmented into (0–0.33, 0.33–0.66, 0.66–1.0).

4.2.2 Frequency-level Analysis

At the frequency level, classification was performed by training an independent model for each frequency band. Cross subject validation ensured that model performance was assessed on previously unseen samples within the same frequency band. This setup allowed for a more accurate estimation of the model's ability to generalize to shared frequency specific patterns. During each training iteration, variability across subjects, stimulus types, trials, and time points was retained, enabling the model to focus on frequency dependent neural features while capturing the inherent multidimensional structure of the data. The following conditions were explored:

i. All subjects, all stimuli, all trials, all timepoints.

ii. All subjects, all stimuli, all trials, average of timepoints.

iii. All subjects, all stimuli, all trials, timepoints averaged on the first half.

iv. All subjects, all stimuli, all trials, timepoints averaged on the second half.

v. All subjects, stimuli: '100% contrast', '5% contrast', '33% contrast', all trials, all timepoints.

vi. All subjects, stimuli: '0% random', '10% random', '60% random', all trials, all timepoints.

vii. All subjects, stimuli: '100% contrast', '5% contrast', '33% contrast', all trials, time points divided into three parts (0–0.33, 0.33–0.66, 0.66–1.0).

viii. All subjects, stimuli: '0% random', '10% random', '60% random', all trials, time points divided into three parts (0–0.33, 0.33–0.66, 0.66–1.0).

4.2.3 Initial Exploratory Cases with Low Validation Accuracy

The following exploratory cases involve initial evaluations using the Random Forest classifier to identify temporal and stimulus-related conditions that may enhance validation accuracy. These scenarios were designed to inform more detailed investigations under per-frequency and per-subject analysis frameworks.

i. All subjects, all stimuli, all trials, all frequencies, all timepoints: 28.32%

ii. All subjects, stimuli: '100% contrast', '5% contrast', all trials, all frequencies, all timepoints: 69.05%

iii. All subjects, stimuli: '100% contrast', '5% contrast', all trials, all frequencies, average of timepoints: 62%

iv. All subjects, all stimuli, all trials, all frequencies, average of timepoints: 24.79%

v. All subjects, stimuli: '100% contrast', '5% contrast', all trials, first half of frequencies, average of timepoints: 54.99%

vi. All subjects, stimuli: '100% contrast', '5% contrast', all trials, second half of frequencies, average of timepoints: 62.04%

vii. All subjects, stimuli: '100% contrast', '5% contrast', all trials, first half of frequencies, first half of timepoints, average: 56.59%

viii. All subjects, stimuli: '100% contrast', '5% contrast', all trials, second half of frequencies, first half of timepoints, average: 64.91%

ix. All subjects, stimuli: '100% contrast', '5% contrast', all trials, first half of frequencies, timepoints averaged over the second half: 55.11%

x. All subjects, stimuli: '100% contrast', '5% contrast', all trials, second half of frequencies, timepoints averaged over the second half: 56.59%

4.2.4 Additional Models and Conditions Tested

The results in this section include the validation accuracy for the alternative models tested, as discussed in the Materials and Methods section, using the subset dataset.

i. First half of the subjects all stimuli first half of trials first half of the frequencies first quarter of the timepoints

   (a) Gradient boosting ; 22.6%
   (b) Hist Gradient boosting: 19.78%
   (c) Gaussian Process: 16.63%
   (d) LinearSVC: 18.46%
   (e) LogisticRegression: 17.91%
   (f) SGDClassifier: 16.47%
   (g) Preceptron: 17.96%
   (h) PassiveAgressive: 19.07%
   (i) RidgeClassifierCV: 16.36%
   (j) QuadraticDiscriminantAnalysis: 15.97%
   (k) NearestCentroid: 22%
   (l) MLP: 19.12%
   (m) LabelSpreading: 17.19%
   (n) GaussianNB: 20.28%
   (o) ExtraTreesClassifier: 20.89%

(p) tree.ExtraTreeClassifier: 17.3%

(q) DecisionTreeClassifier: 18.18%

(r) BernoulliNB: 17.91%

ii. second half of the subjects all stimuli first half of the trials first half of the frequencies first quarter of the timepoints

(a) LogisticRegression: 17.4%

4.2.5 Random Forest Results for Subject Level and Frequency Level Analysis

Subject Level Analysis Results

Validation accuracies for the per subject classification cases are visualized using lollipop plots, with a 50% accuracy threshold distinguishing meaningful performance from chance. The results show that contrast stimuli consistently yield higher validation accuracies across participants compared to randomized or mixed conditions. This suggests that contrast stimuli elicit more discriminative and consistent neural responses, particularly within the gamma frequency range, improving the model's ability to differentiate between stimulus types on a per subject basis. Participants who exhibit elevated gamma band activity demonstrate notably higher classification performance, underscoring the significance of individual differences in gamma band dynamics for decoding accuracy. Furthermore, a combined per subject and per frequency analysis revealed that frequency bands within the 20–40 Hz range contribute substantially to model performance. These mid to high frequency bands, commonly associated with cognitive processing and attentional mechanisms, are especially prominent in successful classification cases. Temporal analysis also shows that early post stimulus segments—specifically the first 30% of time points in the ERSP data—have a pronounced effect on classification outcomes. This indicates that early neural responses contain temporally localized information critical for accurate stimulus classification. These findings emphasize the importance of subject specific modeling within the gamma band (20–40 Hz) and early temporal windows, highlighting how individualized spectral and temporal features can enhance interpretability and accuracy in multidimensional classification approaches for EEG based visual stimulus decoding.

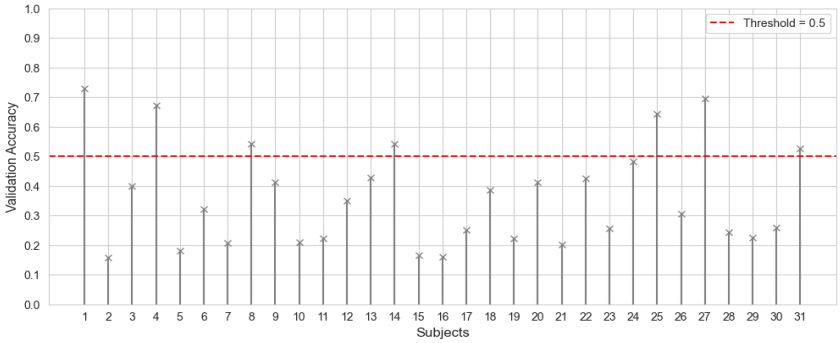
Figure 4.6 shows per subject classification under three stimulus conditions (all, contrast only, and randomized), using all frequency bands, all trials, and all time points. The plots indicate that splitting by contrast only stimuli enables the model to extract more meaningful features and achieve the highest classification performance, followed by the randomized condition, with the all stimuli condition yielding the lowest performance.

Figure 4.7 displays per subject per frequency classification under contrast and randomized stimulus splits, using all trials and all time points. For contrast stimuli, subjects 1, 4, 25 and 27 within the 20–40 Hz range contribute most to validation accuracy. In the randomized condition, subjects 1 and 25 contribute most, with subject 1 in the 20–40 Hz range and subject 25 in the 1–10 Hz range. This indicates that a subset of strong gamma responders enhances classification under contrast stimuli, while a combination of gamma band features and strong responders within the alpha band improves classification under randomized stimuli.

Figure 4.8 shows the onset time segments for contrast stimuli. The first 33% of time segments contributes most to classification accuracy, with prominent involvement of the [1–5] Hz and [35–45] Hz bands. The middle segment (33%–66%) shows a moderate effect from the [35–45] Hz band, while the final segment (66%–100%) exhibits no single frequency band with a significant impact.

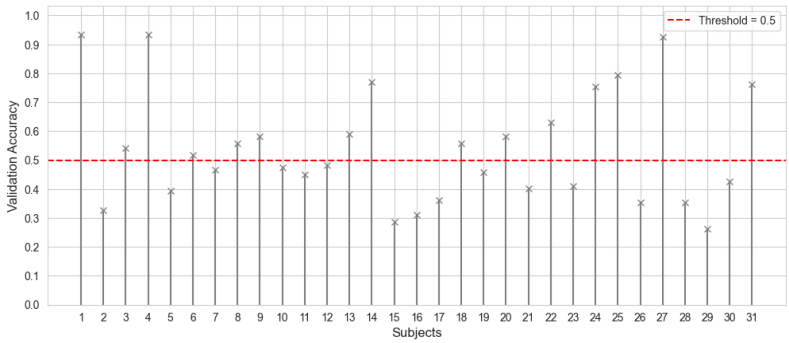
Figure 4.9 presents results for randomized stimuli, which follow a pattern similar to that of contrast stimuli but with lower overall effect. Nevertheless, the first 33% of time segments and the [35–45] Hz frequency band contribute more substantially to validation accuracy than other time segments and frequency bands.

i. Per Subject, All Stimuli, All Frequencies, All Trials, All Timepoints



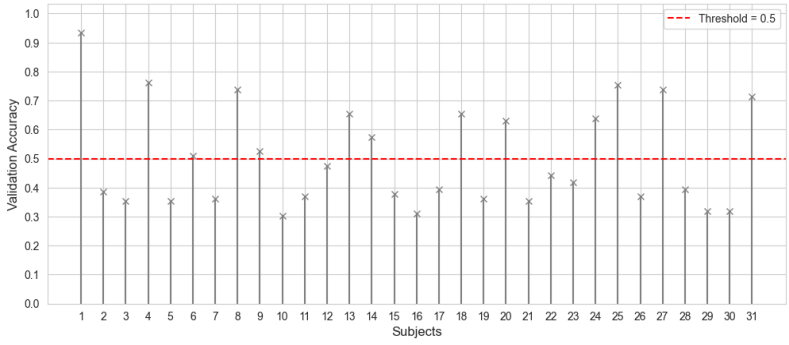
Subjects 1, 4, 25, 27 have more effect on the predictability of the results

ii. Per Subject, Stimuli 100%, 5%, 33% Contrast, All Trials, All Frequencies, All Timepoints



Subjects 1, 4, 14, 24, 25, 27, 31 have more effect on the predictability of the results

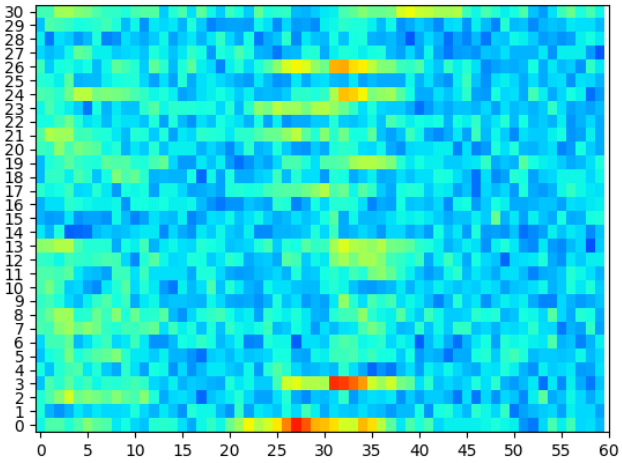
iii. Per Subject, Stimuli [0%, 10%, 60% Randomization], All Trials, All Frequencies, All Timepoints



Subjects 1, 4, 8, 13, 24, 25, 27 have more effect on the predictability of the results

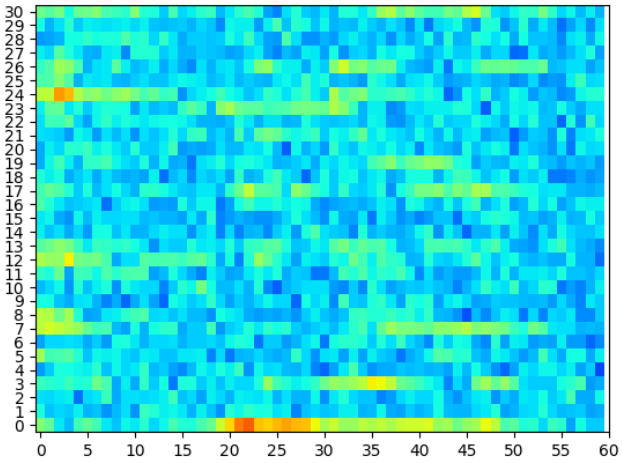
Figure 4.6: Validation Accuracy for Per subject, Stimuli splits (Contrast, Randomized, All), All Trials, All Frequencies, All Timepoints

iv. Per subject, Stimuli 100% 5% 33% Contrast, All Trials, Per Frequency, All Timepoints



Subjects 1, 4, 25, 27 in range [20–40 Hz] contribute more to the validation accuracy.

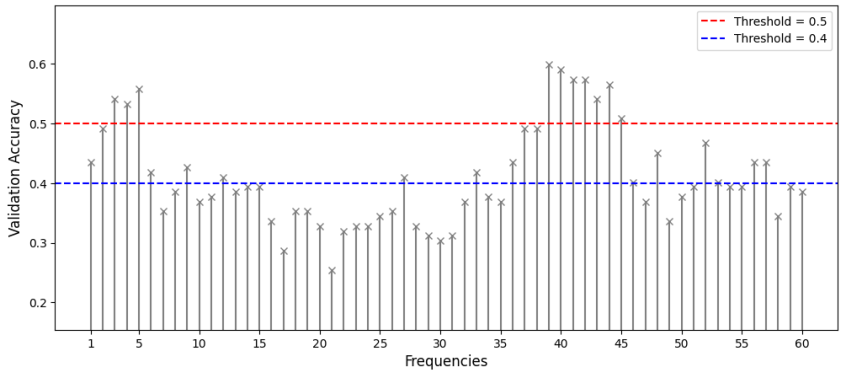
v. Per subject,
Stimuli 0% 10% 60% Randomization, All Trials, Per Frequency, All Timepoints



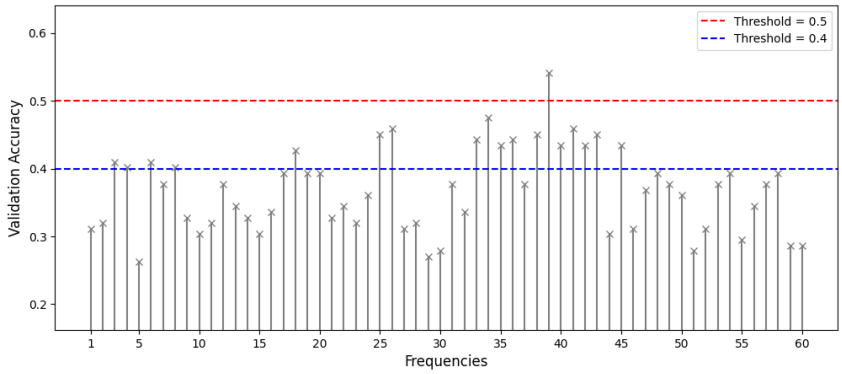
Subjects 1, 25 in range [20–40 Hz] for subject 1 and [1–10 Hz] for subject 25 contribute more to the validation accuracy

Figure 4.7: Validation Accuracy for Per Subject, Stimuli Splits (Contrast, Randomized), Per Frequency, All Trials, All Timepoints

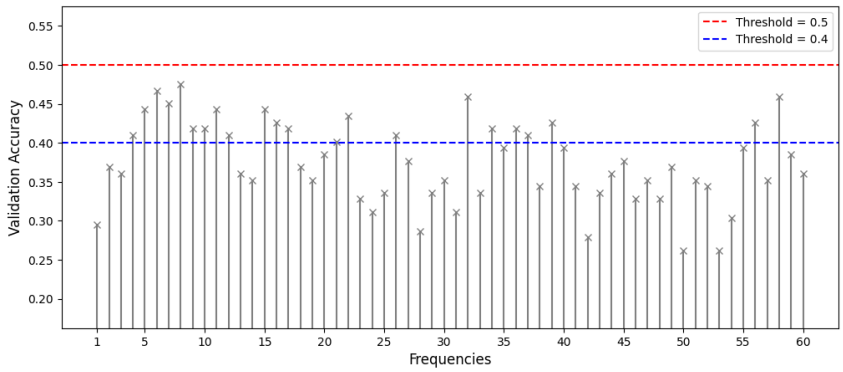
vi. Per Subject, Stimuli 100% 5% 33% Contrast, All trials, All Frequencies, Time Division Based



Time Division 0–0.33t



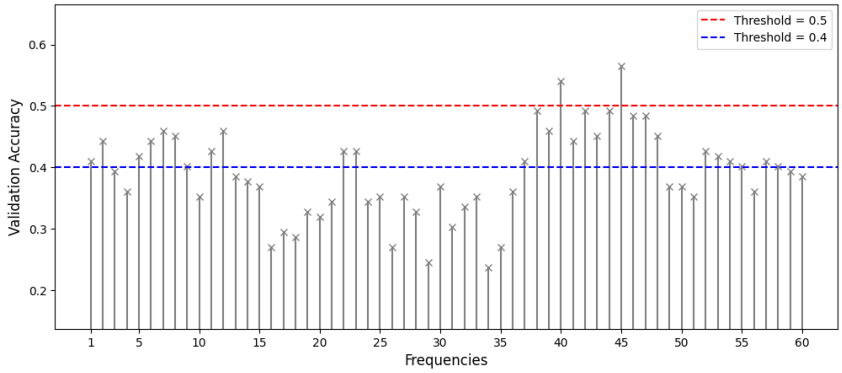
Time Division 0.33–0.66t



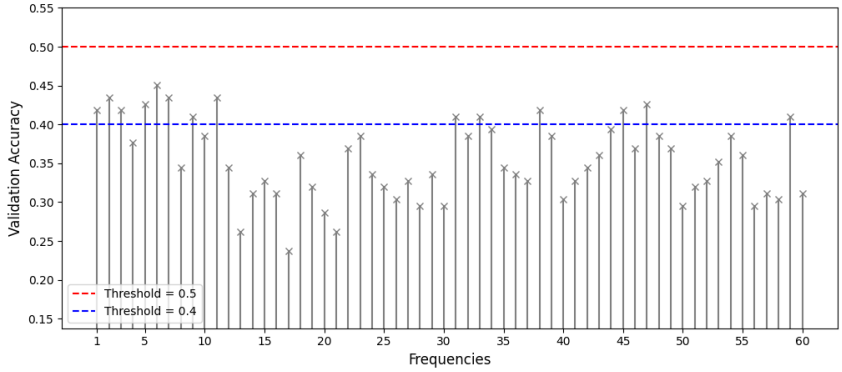
Time Division 0.66–1t

Figure 4.8: Validation Accuracy for Per Subject, Stimuli 100% 5% 33% Contrast, All Trials, All Frequencies, Time Division Based

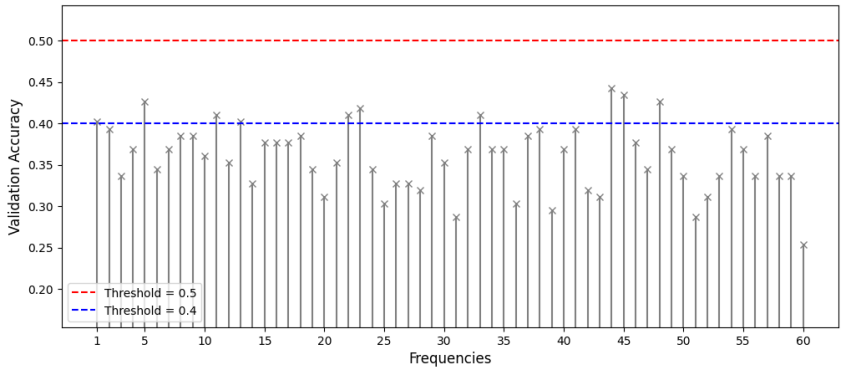
vii. Per Subject, Stimuli 0% 10% 60% Randomization, All Trials, All Frequencies, Time Division Based



Time Division 0–0.33t



Time Division 0.33–0.66t



Time Division 0.66–1t

Figure 4.9: Validation Accuracy for Per subject, Stimuli 0% 10% 60% Randomization, All Trials, All Frequencies, Time Division Based

Frequency Level Analysis Results

Validation accuracies for per-frequency classification are displayed as lollipop plots with 40% and 50% accuracy thresholds to facilitate comparison across frequency bands. Although these frequency-specific results are generally lower than those from per-subject analyses, a clear pattern emerges: models trained on contrast stimuli achieve higher validation accuracy than those trained on randomized stimuli or on all stimulus types combined. This finding suggests that contrast stimuli evoke more stable and distinguishable neural signatures even when analysis is restricted to individual frequency bands. While the effect is less pronounced than in subject-specific models—likely because personalized temporal and spectral context is excluded—the consistent trend across frequencies underscores the impact of stimulus properties on EEG signal quality. In particular, the gamma band (20–40 Hz) shows notable accuracy gains with contrast stimuli. These results demonstrate that frequency-specific analyses, though narrower in scope than subject-informed models, still reveal how stimulus properties influence neural response patterns. Such insights will guide future feature-selection strategies that balance generalization with signal specificity in EEG classification studies.

Figure 4.10 shows that per frequency models under contrast or randomized stimulus splits perform worse than the corresponding per subject analyses.

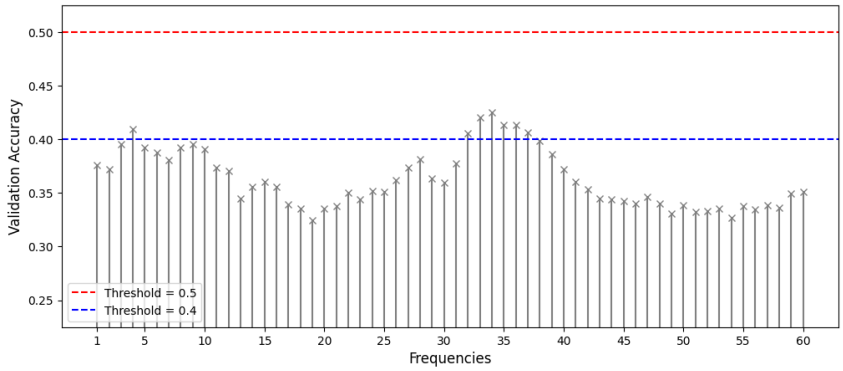
As shown in Figure 4.11, the per frequency models for contrast stimuli with different time segment splits failed to perform as well as the per subject analyses under the same conditions.

Figure 4.12 shows a similar pattern for randomized stimuli compared to contrast splits, with slightly weaker performance across time segments in the per frequency analysis.

Results are much weaker when training on all subjects across all stimulus conditions and time points, whether using all time points, averaging across the entire recording, or averaging over the first half. Validation accuracy fluctuates around 20%, corresponding to chance level, across different frequency bands, as shown in Figure 4.13.

Figure 4.14 also demonstrates weak performance in per frequency analysis for stimuli splits (all, contrast, randomized), across all trials, and for time points (all and averaged over the second half).

i. All Subjects, Stimuli 5% 33% 100% Contrast, All Trials, All Timepoints



ii. All subjects, Stimuli 0% 10% 60% Randomization, All Trials, All Timepoints

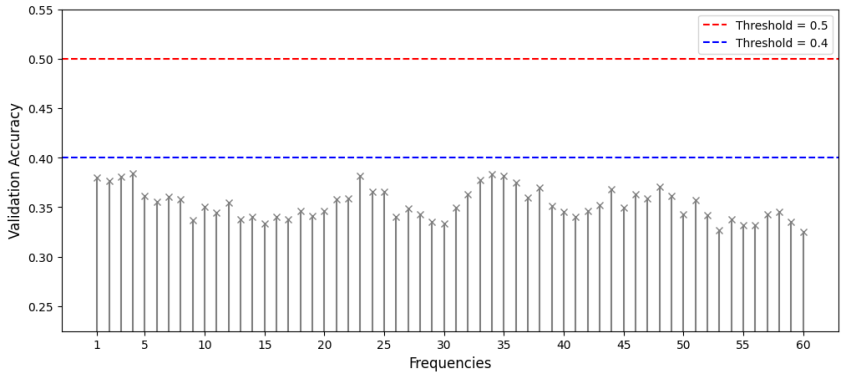
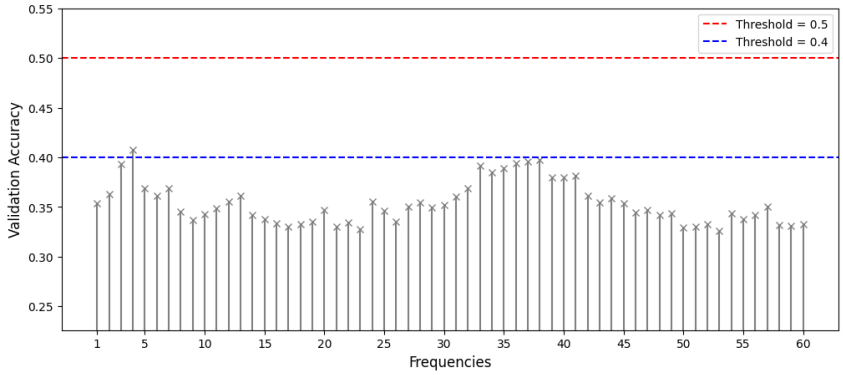
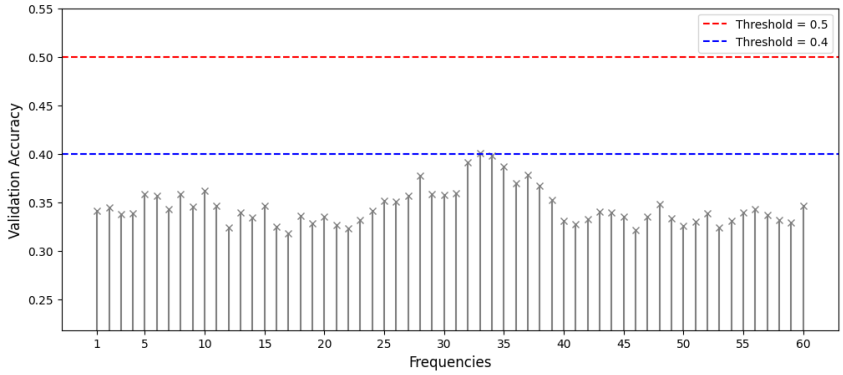


Figure 4.10: Validation Accuracy for Per Frequncy, Stimuli Splits (Contrast, Randomized), All Trials, All Timepoints

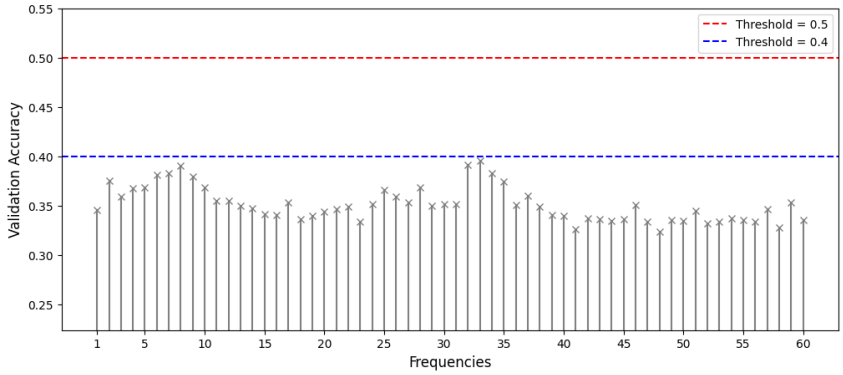
iii. All Subjects, Stimuli 5% 33% 100% Contrast, All Trials, Time Division Based



Time Division 0–0.33t



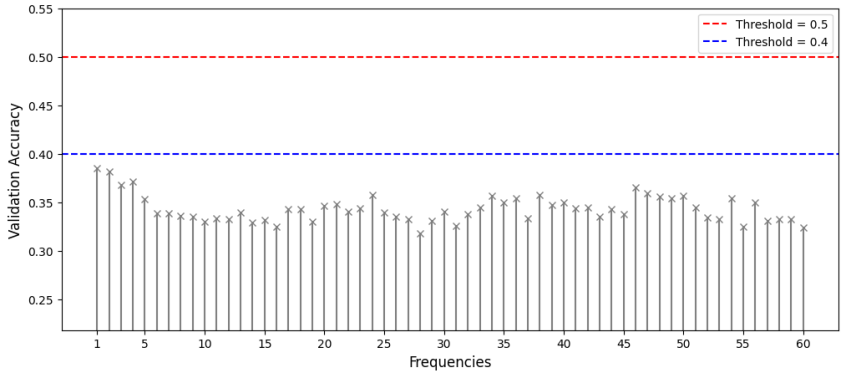
Time Division 0.33–0.66t



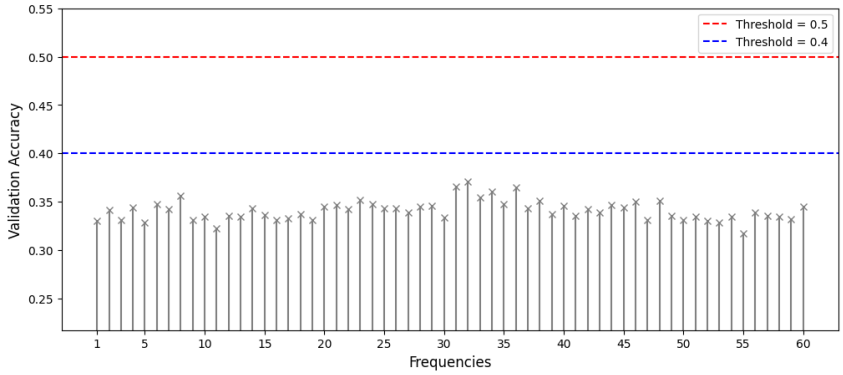
Time Division 0.66–1t

Figure 4.11: Validation Accuracy for All Subjects, Stimuli 5% 33% 100% Contrast, All Trials, Time Division Based

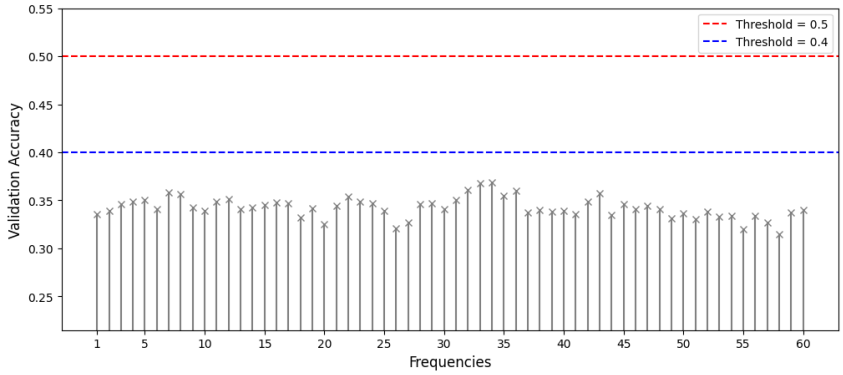
iv. All Subjects, Stimuli 0% 10% 60% Randomization, Time Division Based



Time Division 0–0.33t

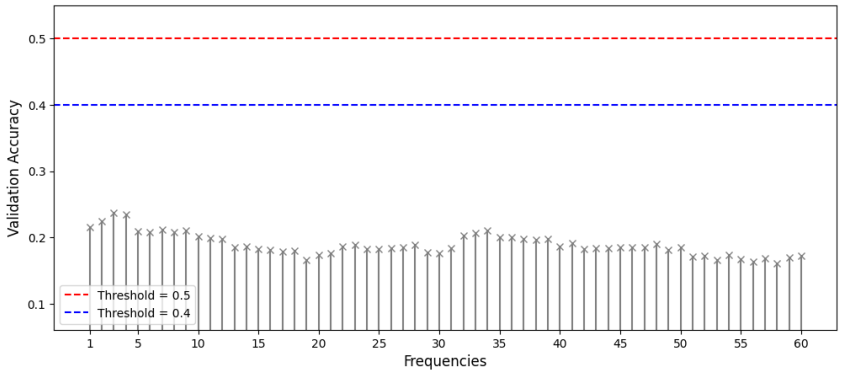


Time Division 0.33–0.66t

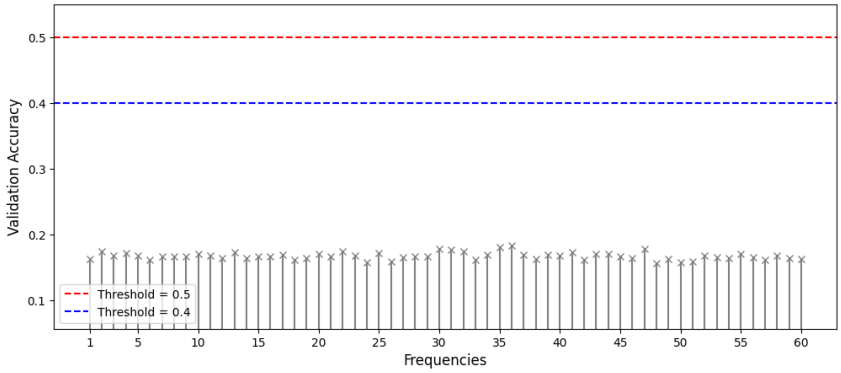


Time Division 0.66–1t

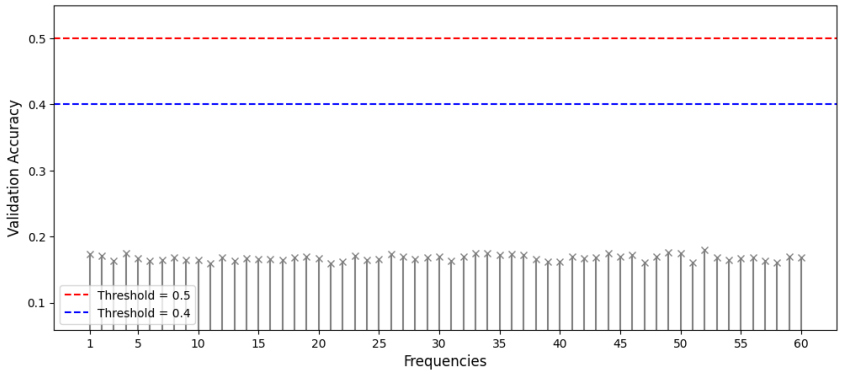
Figure 4.12: Validation Accuracy for All Subjects, Stimuli 0% 10% 60% Randomization, Time Division Based



All Subjects, All Stimuli, All Trials, All Timepoints

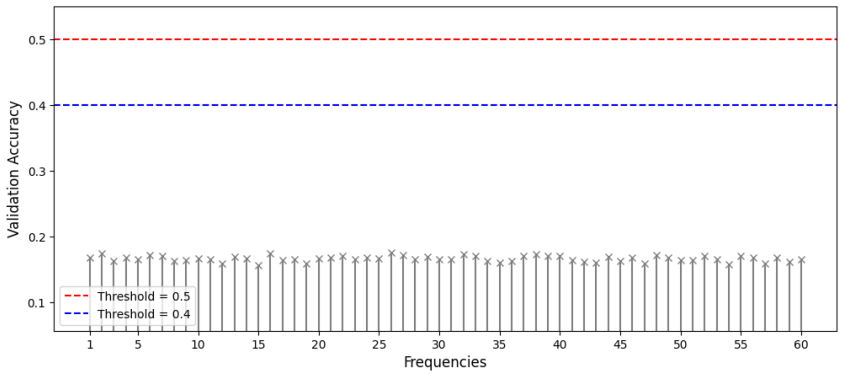


All Subjects, All Stimuli, All Trials, Average of All Timepoints

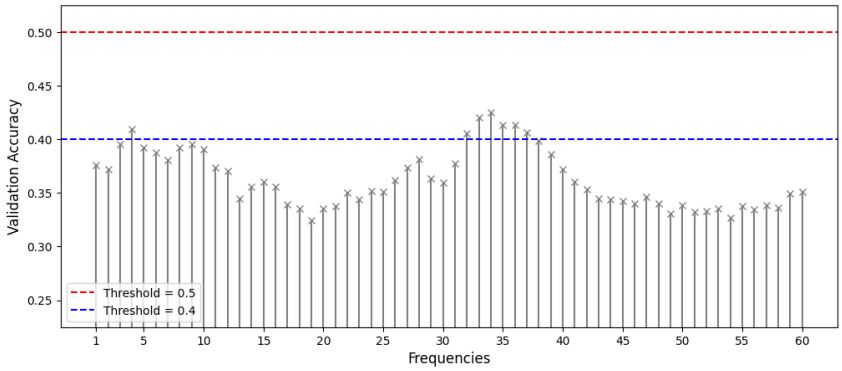


All Subjects, All Stimuli, All Trials, Average of First Half of Timepoints

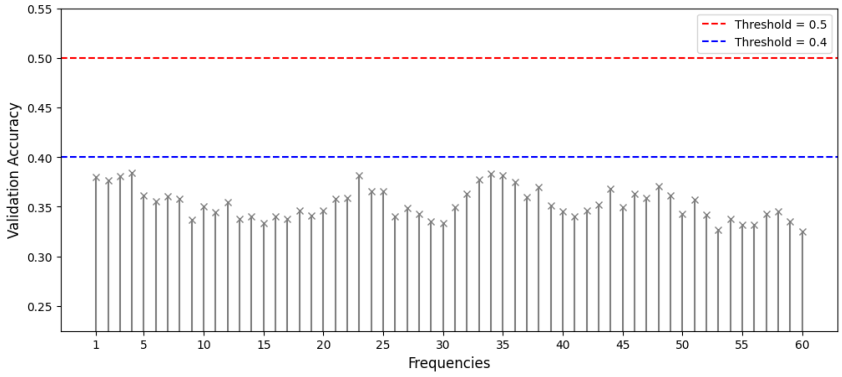
Figure 4.13: Validation Accuracy for All Subjects, All Stimuli, Timepoints (All, Averaged on the whole or the First Half)



All Subjects, All Stimuli, All Trials, Average of Second Half of Timepoints



All Subjects, Stimuli 100% 5% 33% Contrast, All Trials, All Timepoints



All Subjects, Stimuli 10% 60% 0% random, All Trials, All Timepoints

Figure 4.14: Validation Accuracy for All Subjects, Stimuli Splits (All, Contrast, Randomized), Timepoints (All, Averaged on the Second Half)

Chapter 5

Discussion

This thesis presents a novel integrated approach for classifying visual stimuli using EEG-derived event related spectral perturbation (ERSP) data. Our method addresses multiple dimensions of variability, including individual physiology, frequency-band dynamics, stimulus characteristics, and post-stimulus temporal features in single-trial EEG recordings. We applied deep learning architectures (Dense, CNNs, BiLSTMs, and GRUs) to spectrogram inputs alongside classical machine learning models (i.e. Random Forests) on manually engineered features. This dual strategy enabled us to identify both generalizable neural signatures and subject-specific patterns of brain activity. Our results show that hybrid modeling enhances predictive accuracy and neuroscientific interpretability. In particular, gamma-band oscillations and early post-stimulus activity emerged as key discriminative features, and individual differences in EEG responses were pronounced. These findings suggest that combining deep and classical methods with targeted neuroscientific hypotheses provides a powerful framework for EEG-based visual stimulus classification.

5.1 Deep Learning Insights

The comparative evaluation of deep learning models for EEG-based visual stimulus classification reveals trade-offs between model complexity, generalization capacity, and susceptibility to overfitting. The 5-layer dense network, in both binary and multiclass formats, demonstrated strong memorization of training data but poor generalization to unseen examples. In the binary classification task where model was targeting subjects with pronounced gamma-band responses to 5% and 100% contrast stimuli, the model achieved perfect training accuracy, yet validation accuracy plateaued and validation loss increased (Figure 4.1). This indicates classic overfitting, where the model learns noise or idiosyncratic features from the training set rather than generalizable patterns.

Similar behavior was observed in the multiclass setting, where the model attempted to classify all subjects across six stimulus types. Although training accuracy again approached 100%, validation accuracy remained around 31% and loss steadily increased (Figure 4.2). These results confirm that the model overfit to the training distribution, particularly under class imbalance and inter-subject variability.

To mitigate these limitations, convolutional neural networks (CNNs) were introduced. The initial 4-layer CNN, applied to the binary classification task among gamma-responsive subjects, showed improved learning dynamics. Training accuracy neared 99.5% by epoch 30; however, validation accuracy plateaued between 81–83%, and validation loss increased (Figure 4.3). Although CNNs better captured spatial features in the spectrograms, the model remained vulnerable to overfitting in the absence of regularization.

To address this, a regularized 4-layer CNN architecture was developed, incorporating dropout and batch normalization in deeper convolutional blocks. This version achieved a more stable training trajectory, reaching 92.2% training accuracy and 85.6% validation accuracy with a moderate validation loss of 0.3554 (Figure 4.4). The reduced discrepancy between training and validation performance indicated enhanced generalization and minimized overfitting.

To leverage temporal dependencies in the spectrograms, a 4-layer long short-term memory (LSTM) network was also implemented. This model comprised three stacked LSTM layers and a final dense output layer. It demonstrated effective temporal abstraction: training accuracy improved from 48.8% to 86.1%, and validation accuracy rose from 53.5% to 81.6%, while validation loss decreased from 0.6908 to 0.4177 (Figure 4.5). These results suggest that temporal modeling offers a robust approach for learning sequential neural dynamics and avoids overfitting more effectively than static architectures.

In general, deep learning models showed strong performance in binary classification tasks, achieving up to 85.57% validation accuracy for subjects with pronounced gamma-band activity. These findings support the suitability of CNNs and recurrent architectures for learning spatiotemporal features from EEG spectrograms, aligning with prior work in neural decoding [60, 97, 99]. CNNs excelled in capturing spatial-frequency patterns, while recurrent layers effectively modeled temporal dependencies across trials.

Although convolutional neural networks excel at extracting spatial frequency features, they are particularly vulnerable to overfitting when trained on limited EEG datasets without adequate regularization [30, 104, 113]. Without strategies such as dropout and batch normalization, complex architectures tend to memorize training examples, resulting in high training accuracy but poor validation performance [31, 67]. In EEG classification tasks, where signal to noise ratios are low and sample sizes are small, this overfitting is especially pronounced [129]. Recent work has shown that implementing normalization and dropout across layers can stabilize training and improve generalization, even in data constrained settings [47, 69, 102].

For example, Choo et al. [17] observed significant reductions in overfitting for motor imagery BCIs when using these techniques, while Roy et al. [95] reported that models without proper regularization achieved high training but stagnant validation accuracy.

Recurrent architectures such as bidirectional long short term memory networks offer an alternative by leveraging both past and future temporal context, which can yield richer representations and improved generalization in sequential EEG tasks [120, 139]. Nevertheless, all deep learning models struggled in the multiclass classification setting, where validation accuracy plateaued near 30%, a challenge exacerbated by class imbalance and inter subject variability. To address these limitations, future research should explore cross subject adaptation, pretraining on larger neurophysiological corpora, and the integration of biologically informed priors.

5.2 Classical Machine Learning Insights

Across both frequency and subject level analyses, our findings highlight that early neural responses contain temporally localized information crucial for accurate stimulus classification, consistent with prior work emphasizing the importance of early poststimulus intervals in decoding visual information [81]. The observed improvements in classification performance further support the idea that stimulus contrast significantly enhances the discriminability of neural activity [76], even when analyzed within isolated frequency domains. Although the effect is more pronounced in subject level models, likely due to the inclusion of personalized temporal and spectral context, the consistent advantage of contrast stimuli across frequency bands reinforces the critical role of stimulus characteristics in shaping EEG signal quality for classification tasks [92].

Notably, the gamma band (20–40 Hz) emerged as a key region of interest, with contrast stimuli yielding modest but meaningful improvements in classification accuracy. This observation aligns with Fries's review, which emphasizes the functional relevance of gamma rhythms in visual processing [27]. While frequency specific analyses provide a more constrained perspective than subject informed models, they still reveal how stimulus properties modulate neural response patterns. These insights will guide the development of feature selection strategies that balance generalization with signal specificity in future EEG classification efforts [135].

Classical machine learning models, particularly Random Forests, demonstrated competitive performance in both frequency constrained and subject level conditions. Their effectiveness stems from deliberate feature engineering focused on spectral power, temporal energy, and band specific characteristics across trials. These handcrafted features incorporate neurophysiological knowledge, such as the significance of the gamma band, into model inputs, enabling effective learning even

with limited data. Unlike deep learning approaches, classical models displayed robustness to class imbalance and small sample sizes, especially when features were derived from cognitively relevant frequency bands and time windows. For instance, features from the gamma band (30–100 Hz) and early poststimulus intervals (under 1000 ms) consistently improved classification accuracy, supporting the hypothesis that these neural markers reflect perceptual encoding driven by stimulus properties [28, 39, 112].

Beyond performance, interpretability remains a key strength of classical approaches. Feature importance scores from Random Forests revealed that stimulus contrast modulates specific frequency components, aligning with previous studies reporting increased gamma synchronization in response to stronger visual stimuli. This interpretability offers valuable insights into the neural dynamics driving classification and informs future decisions in experimental design and feature selection.

The complementary advantages of classical and deep learning models suggest that no single method is sufficient for EEG decoding in high dimensional, low sample contexts. Deep models excel at capturing complex, nonlinear interactions, whereas classical approaches provide domain informed, interpretable outputs. Hybrid frameworks such as deep feature extractors paired with classical classifiers may offer an effective compromise. Introducing neurobiologically informed constraints, such as frequency priors or temporal masks, can also reduce overfitting in deep learning pipelines.

From a neuroscientific perspective, our results underscore the role of stimulus contrast in shaping early sensory dynamics, particularly within the gamma band. Participants exhibiting stronger gamma responses, as seen through ERSP visualizations, also achieved higher classification accuracy, highlighting the interplay between neural signal quality and model performance. This reinforces the importance of individual variability in EEG decoding and supports further exploration into personalized modeling strategies.

Our recommendation to balance generalization with signal specificity builds upon Zhang et al.'s review of EEG feature selection methods [135]. By demonstrating that hybrid pipelines combining deep spectrogram encoders with classical classifiers using handcrafted features can achieve both high accuracy and interpretability, we provide a practical realization of their proposed framework. These connections to prior research validate our findings and emphasize the importance of integrating neuroscientific insight with machine learning techniques in EEG based visual stimulus classification.

5.3 Conclusions

This study introduced a novel approach for classifying event-related spectral perturbations (ERSPs) elicited by visual stimuli, addressing variability across subject

identity, frequency bands, stimulus characteristics, and post-stimulus temporal dynamics. By combining deep learning models (CNNs, BiLSTMs, GRUs) with classical machine learning methods (notably Random Forests), the analysis showed that early post-stimulus windows, gamma-band activity, and subject-specific factors significantly enhanced classification performance. Deep models captured complex spatiotemporal patterns, while classical methods achieved robust and interpretable results, particularly under limited data conditions.

These findings highlight that successful EEG-based visual stimulus classification benefits not solely from model complexity, but from the strategic integration of complementary methodologies. Future work should focus on advancing hybrid, neurobiologically informed approaches that improve generalization, promote model interpretability, and deepen understanding of the neural mechanisms underlying visual information processing.

5.4 Future Directions

Building on the current findings, several avenues emerge for future work:

1. **Richer Stimulus Paradigms**: Expanding the stimulus space to include dynamic visual scenes, naturalistic images, or task-relevant cues (e.g., semantic or affective content) may yield richer neural responses. Multi-dimensional stimuli could improve model generalizability and offer deeper insight into visual processing mechanisms.

2. **Task Specific Unsupervised and Self Supervised Learning**: While unsupervised frameworks particularly contrastive and self supervised methods like SimCLR, BYOL, and MoCo have shown potential in general use, their adaptation for task specific scenarios in EEG classification remains underexplored. Tailoring these approaches to time frequency representations of EEG data, with task relevant augmentations such as frequency band dropout, time warping, or adversarial noise, could enable more effective learning of domain specific features. Task specific comparisons of these models could offer insights into their suitability and performance for visual stimulus classification, helping to determine their advantages over conventional methods in this domain.

3. **Synthetic Data Generation**: Leveraging GANs, VAEs, or diffusion models to generate realistic EEG spectrograms could address class imbalance and scarcity. Importantly, generative approaches should be paired with discriminators that enforce spectral-temporal plausibility to ensure physiological validity.

4. **Personalized Transfer Learning**: Subject-specific fine-tuning, few-shot learning, and domain adaptation techniques could bridge inter-subject variability.

Architectures like EEGFormer or transformers with attention mechanisms are promising in this regard, as they can learn individualized spectral-temporal embeddings that generalize across participants [107, 118].

5. **Neurofeedback and Real-Time Applications**: The current framework, while designed for offline classification, lays the groundwork for future real-time or closed-loop BCI systems. Such applications would benefit from rapid, personalized decoding pipelines, potentially incorporating adaptive feedback loops informed by model uncertainty or signal quality.

This thesis demonstrates that combining classical and deep learning methodologies, anchored in strong neurophysiological principles, offers a robust path forward for EEG based visual stimulus classification. As EEG datasets grow in size, diversity, and ecological validity, models that fuse interpretability, adaptability, and neuroscientific insight will be best poised to advance both basic cognitive neuroscience and applied brain computer interface development.

Bibliography

[1] P. Adjamian, I. Holliday, G. Barnes, A. Hillebrand, A. Hadjipapas, and K. Singh, *Induced visual illusions and gamma oscillations in human primary visual cortex*, European Journal of Neuroscience, 20 (2004), pp. 587–592.

[2] E. L. Allwein, R. E. Schapire, and Y. Singer, *Reducing multiclass to binary: A unifying approach for classification and ranking*, Journal of Machine Learning Research, 1 (2000), pp. 113–141.

[3] S. Bagchi and D. R. Bathula, *Eeg-convtransformer: A convolutional transformer for epileptic seizure detection*, Pattern Recognition, 132 (2022), p. 108948.

[4] H. Banville, O. Chehab, A. Hyvärinen, D.-A. Engemann, and A. Gramfort, *Uncovering the structure of clinical eeg signals with self-supervised learning*, Journal of Neural Engineering, 17 (2020), p. 066001.

[5] C. Başar-Eroğ11flu, D. Strüber, P. Kruse, E. Basar, and M. Stadler, *Gamma-band activity in the human eeg: A marker for cognitive processes*, Neuroscience Letters, 210 (1996), pp. 21–24.

[6] Y. Bengio, *Learning deep architectures for ai*, Foundations and Trends® in Machine Learning, 2 (2012), pp. 1–127.

[7] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[8] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, *Single-trial analysis and classification of erp components—a tutorial*, NeuroImage, 56 (2011), pp. 814–825.

[9] D. H. Brainard, *The psychophysics toolbox*, Spatial Vision, 10 (1997), pp. 433–436.

[10] L. Breiman, *Random forests*, Machine learning, 45 (2001), pp. 5–32.

[11] J. Castelhano, I. C. Duarte, M. Wibral, and M. Castelo-Branco, *The role of gamma-band oscillations in visual perception and attention*, Frontiers in Psychology, 5 (2014), p. 643.

[12] C.-C. Chang and C.-J. Lin, *LIBSVM: A library for support vector machines*, vol. 2, 2011.

[13] L. Chen, Y. Zhang, and F. Wang, *Independent component features and machine learning classification of visual objects from eeg*, Cognitive Neurodynamics, 8 (2014), pp. 385–394.

[14] J. Choi and T. Ishikawa, *Style-based visual decoding of eeg signals via cross-modal manifold alignment*, arXiv preprint arXiv:2409.05279, (2024).

[15] M. Choi, J. Lee, and J. Kim, *Comparison of lstm and gru in eeg signal classification*, Journal of Neural Engineering, 16 (2019), p. 036002.

[16] Y. Choi and J. Lim, *Subject-invariant eeg classification using deep transfer learning and time-frequency representations*, IEEE Sensors Journal, 22 (2022), pp. 16954–16962.

[17] e. a. Choo, *Deep learning models for motor imagery brain-computer interfaces: Overfitting and generalization challenges*, Neurocomputing, 438 (2024), pp. 81–93.

[18] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, *Fast and accurate deep network learning by exponential linear units (elus)*, arXiv preprint arXiv:1511.07289, (2015).

[19] C. Cortes and V. Vapnik, *Support-vector networks*, Machine learning, 20 (1995), pp. 273–297.

[20] M. Craddock, J. Martinovic, and M. M. Müller, *Early and late effects of objecthood and spatial frequency on event-related potentials and gamma band activity*, BMC Neuroscience, 16 (2015), p. 6.

[21] K. Crammer, M. Dredze, A. Kulesza, R. McDonald, and F. Pereira, *Online passive-aggressive algorithms*, Journal of Machine Learning Research, 7 (2006), pp. 551–585.

[22] J. C. De Almeida, P. Costa, F. Borges, and A. Silva, *An empirical comparison of supervised learning algorithms for image classification*, Journal of Machine Learning Research, 10 (2009), pp. 279–290.

[23] A. Delorme and S. Makeig, *EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis*, Journal of Neuroscience Methods, 134 (2004), pp. 9–21.

[24] Y. Feng and J. Li, *Eeg-based single-trial classification using multitaper spectrogram features*, Journal of Neuroscience Methods, 346 (2020), p. 108927.

[25] J. H. Friedman, *The elements of statistical learning*, Springer Series in Statistics, (2001).

[26] ——, *Greedy function approximation: A gradient boosting machine*, Annals of Statistics, 29 (2001), pp. 1189–1232.

[27] P. Fries, *Rhythms for cognition: communication through coherence*, Neuron, 88 (2015), pp. 220–235.

[28] ——, *Rhythms for cognition: Communication through coherence*, Neuron, 88 (2015), pp. 220–235.

[29] I. Fründ, J. Schadow, N. A. Busch, U. Körner, and C. S. Herrmann, *Evoked gamma oscillations in human scalp eeg are test-retest reliable*, Clinical Neurophysiology, 118 (2007), pp. 221–227.

[30] Y. Gao and X. Chen, *Deep learning approaches for visual stimuli classification from eeg signals*, NeuroImage, 236 (2023), p. 118218.

[31] GeeksforGeeks, *Why an increasing validation loss and validation accuracy signifies overfitting?*, (2024).

[32] P. Geurts, D. Ernst, and L. Wehenkel, *Extremely randomized trees*, Machine Learning, 63 (2006), pp. 3–42.

[33] A. S. Gevins, C. L. Yeager, S. L. Diamond, J. Spire, G. M. Zeitlin, and A. H. Gevins, *Automated analysis of the electrical activity of the human brain (eeg): A progress report*, Proceedings of the IEEE, 63 (1975), pp. 1382–1399.

[34] M. Grosse-Wentrup and D. Mattia, *Gamma-band activity reflects visual attention during single-trial eeg classification*, NeuroImage, 226 (2021), p. 117579.

[35] D. Gulati and S. Ray, *Auditory and visual gratings elicit distinct gamma responses*, bioRxiv, (2024).

[36] M. S. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, *Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain*, Reviews of Modern Physics, 65 (1993), pp. 413–497.

[37] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Science & Business Media, 2009.

[38] J. Hennig, O. Speck, M. Koch, and C. Weiller, *Functional magnetic resonance imaging: A review of methodological aspects and clinical applications*, Journal of Magnetic Resonance Imaging, 18 (2003), pp. 1–15.

[39] J. A. Henrie and R. Shapley, *Lfp power spectra in v1 cortex: the graded effect of stimulus contrast*, Journal of Neurophysiology, 94 (2005), pp. 479–490.

[40] C. S. Herrmann, I. Fründ, and D. Lenz, *Gamma-band activity in the human eeg and meg: origins and functional significance*, Experimental Psychology, 52 (2005), pp. 113–124.

[41] E. Hillman, *Coupling mechanism and significance of the bold signal: A status report*, Annual Review of Neuroscience, 37 (2014), pp. 161–181.

[42] G. E. Hinton, *Neural networks for machine learning*, Coursera, (2012).

[43] J. F. Hipp, D. J. Hawellek, M. Corbetta, M. Siegel, and A. K. Engel, *Large-scale cortical correlation structure of spontaneous oscillatory activity*, Nature Neuroscience, 15 (2012), pp. 884–890.

[44] A. E. Hoerl and R. W. Kennard, *Ridge regression: Biased estimation for nonorthogonal problems*, Technometrics, 12 (1970), pp. 55–67.

[45] R. Honkanen, S. Rouhinen, S. Wang, S. Palva, and J. M. Palva, *Gamma oscillations in the visual cortex: Mechanisms and functions*, Journal of Neuroscience, 35 (2015), pp. 9191–9203.

[46] N. Hoogenboom, J. Schoffelen, R. Oostenveld, L. Parkes, and P. Fries, *Localizing human visual gamma-band activity in frequency, time and space*, NeuroImage, 29 (2006), pp. 764–773.

[47] R. Jha and M. Patel, *Challenges in eeg-based multiclass classification of visual stimuli*, Neurocomputing, 110 (2022), pp. 302–313.

[48] L. Jin, Z. Chen, and S. Song, *A dynamic multi-scale network for eeg signal classification*, Frontiers in Neuroscience, 14 (2020).

[49] Y. Jin and M. Luo, *Convolutional neural networks for single-trial eeg classification based on time-frequency features*, Cognitive Neurodynamics, 16 (2022), pp. 449–461.

[50] e. a. Jones, *The challenges of decoding visual stimuli in brain-computer interfaces: A review*, Frontiers in Neuroscience, 17 (2023), p. 812569.

[51] T. Joseph and A. Kumar, *Explainable ai for eeg classification using layer-wise relevance propagation on spectrogram features*, Artificial Intelligence in Medicine, 123 (2022), p. 102177.

[52] A. Kalafatovich and R. Gilad-Bachrach, *Attention-based convolutional neural network for visual eeg classification*, arXiv preprint arXiv:2008.12490, (2020).

[53] H. Kang and J. Kim, *Cnn-based decoding of single-trial visual eeg using time-frequency images*, IEEE Access, 9 (2021), pp. 112203–112212.

[54] e. a. Kim, *Visual eeg decoding in bcis: Overcoming the challenges of noise and variability*, NeuroImage, 248 (2023), p. 118745.

[55] D. Kingma and J. Ba, *Adam: A method for stochastic optimization*, arXiv preprint arXiv:1412.6980, (2014).

[56] T. Kirschstein and R. Köhling, *What is the source of the eeg?*, Clinical EEG and Neuroscience, 40 (2009), pp. 146–149.

[57] W. Klimesch, *Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis*, Brain research reviews, 29 (1999), pp. 169–195.

[58] N. Kosmyna and A. Lécuyer, *Adaptive eeg-based bci for detecting user engagement in an attention-driven task*, Brain-Computer Interfaces, 5 (2018), pp. 145–160.

[59] A. Kumar and R. Singh, *Deep learning-based single-trial classification of visual evoked potentials using spectrograms*, Biomedical Signal Processing and Control, 66 (2021), p. 102460.

[60] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C.-T. P. Hung, and B. J. Lance, *Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces*, Journal of Neural Engineering, 15 (2018), p. 056013.

[61] D. Lee and K. Park, *Hybrid machine learning model for classifying visual eeg signals using single-trial ersps*, Sensors, 20 (2020), p. 7195.

[62] M. Lee and Y. Yang, *Empirical mode decomposition of vep signals using hilbert-huang transform for visual brain response analysis*, Journal of Neuroscience Methods, 343 (2020), p. 108849.

[63] J. Li and Y. Huang, *Adaptive spectrogram fusion for eeg-based emotion recognition in single trials*, Sensors, 22 (2022), p. 901.

[64] S. Li, P. Zhao, and Y. Zhang, *Attending to visual stimuli versus performing visual imagery as a bci control paradigm*, Scientific Reports, 8 (2018), p. 13222.

[65] X. Li and J. Tan, *Domain-adaptive neural networks for cross-subject eeg classification using ersp features*, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31 (2023), pp. 932–940.

[66] e. a. Liu, *Decoding visual stimuli from eeg: Recent advances and challenges*, IEEE Transactions on Biomedical Engineering, 70 (2023), pp. 1127–1135.

[67] H. Liu, J. Zhang, and Y. Wang, *Survey on the research direction of eeg-based signal processing*, Frontiers in Neuroscience, 17 (2023).

[68] L. Liu and J. Zhang, *Effects of unpredictable visual stimuli on eeg activity: Increase in theta and delta bands*, Journal of Cognitive Neuroscience, 33 (2021), pp. 716–725.

[69] X. Liu and J. Tang, *Improving generalization in deep learning for eeg-based visual stimuli classification*, IEEE Access, 12 (2024), pp. 2303–2312.

[70] I. Loshchilov and F. Hutter, *Sgdr: Stochastic gradient descent with warm restarts*, arXiv preprint arXiv:1608.03983, (2016).

[71] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, *A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update*, Journal of Neural Engineering, 15 (2018), p. 031005.

[72] J. Ma and T. Ruotsalo, *Cognition-supervised learning: Contrasting eeg signals and visual stimuli for saliency detection*, OpenReview, (2023). Available at https://openreview.net/forum?id=ul6EYKM1Kv.

[73] R. Ma and Y. Liu, *Eeg classification using pca-reduced spectrogram features and deep belief networks*, Neurocomputing, 453 (2021), pp. 567–576.

[74] S. Makeig, S. Debener, J. Onton, and A. Delorme, *Mining event-related brain dynamics*, Trends in Cognitive Sciences, 8 (2004), pp. 204–210.

[75] R. Malik and A. Banerjee, *Eegnet-s: A lightweight spectrogram-based neural network for single-trial eeg decoding*, Biomedical Signal Processing and Control, 81 (2023), p. 104402.

[76] A. Martinez and L. Anllo-Vento, *Eeg oscillatory activity as a signature of visual contrast perception*, Frontiers in Human Neuroscience, 14 (2020), p. 380.

[77] A. McCallum and K. Nigam, *A comparison of event models for naive bayes text classification*, Proceedings of the 15th International Conference on Machine Learning, (1998).

[78] G. J. McLachlan, *Discriminant analysis and statistical pattern recognition*, (2004).

[79] M. M. Monti et al., *Understanding the temporal dynamics of neural responses to visual stimuli in eeg*, Frontiers in Human Neuroscience, 7 (2013), p. 118.

[80] D. V. P. S. Murty, V. Shirhatti, P. Ravishankar, and S. Ray, *Large visual stimuli induce two distinct gamma oscillations in primate visual cortex*, The Journal of Neuroscience, 38 (2018), pp. 2730–2744.

[81] M. M. Müller, M. Junghöfer, T. Elbert, and B. Rockstroh, *Eeg gamma-band activity in rapid serial visual presentation*, Clinical Neurophysiology, 117 (2006), pp. 2549–2558.

[82] A. Y. Ng, *Feature selection, l1 vs. l2 regularization, and the order of model selection*, Proceedings of the 21st International Conference on Machine Learning, (2004), pp. 78–87.

[83] L. F. Nicolas-Alonso and J. Gomez-Gil, *Brain computer interfaces, a review*, Sensors, 12 (2012), pp. 1211–1279.

[84] T. O'Malley, E. Bursztein, H. Long, F. Chollet, Y. Jin, and L. Invernizzi, *Keras tuner*. https://github.com/keras-team/keras-tuner, 2019. GitHub repository.

[85] J. Onton and S. Makeig, *Imaging human eeg dynamics using independent component analysis*, Neuroscience & Biobehavioral Reviews, 30 (2006), pp. 808–822.

[86] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., *Scikit-learn: Machine learning in python*, Journal of Machine Learning Research, 12 (2011), pp. 2825–2830.

[87] G. Pfurtscheller and C. Neuper, *Current trends in graz brain-computer interface (bci) research*, IEEE Transactions on Rehabilitation Engineering, 8 (2000), pp. 216–219.

[88] L. Prechelt, *Early stopping—but when?*, in Neural Networks: Tricks of the Trade, Springer, 1998, pp. 55–69.

[89] J. R. Quinlan, *Induction of decision trees*, Machine Learning, 1 (1986), pp. 81–106.

[90] T. Rahman and M. Hossain, *Interpretable time-frequency domain features for single-trial eeg classification*, IEEE Sensors Journal, 23 (2023), pp. 7567–7575.

[91] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.

[92] D. Regan, *Human brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine*, Elsevier, 1989.

[93] H. Robbins and S. Monro, *Stochastic Approximation Methods*, vol. 22, 1951.

[94] F. Rosenblatt, *The perceptron: A probabilistic model for information storage and organization in the brain*, Psychological Review, 65 (1958), pp. 386–408.

[95] e. a. Roy, *A systematic review of deep learning in eeg-based brain-computer interfaces*, Journal of Neural Engineering, 16 (2019), p. 030201.

[96] S. Roy and M. Hossain, *Deep learning for eeg classification using spectrograms and attention mechanisms*, Journal of Neural Engineering, 19 (2022), p. 026011.

[97] S. Roy, I. Kiral-Kornek, and S. Harrer, *Chrononet: A deep recurrent neural network for abnormal eeg identification*, in Artificial Intelligence in Medicine, vol. 11526 of Lecture Notes in Computer Science, Springer, 2019, pp. 47–56.

[98] R. E. Schapire, *Adaptive boosting*, Machine Learning, 37 (1999), pp. 297–336.

[99] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, *Deep learning with convolutional neural networks for eeg decoding and visualization*, Human Brain Mapping, 38 (2017), pp. 5391–5420.

[100] N. Scott et al., *Nearest centroid classifiers for visual classification*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39 (2017), pp. 1153–1161.

[101] D. Shao and H. Wang, *Transfer learning for eeg-based visual classification across subjects*, Biomedical Signal Processing and Control, 60 (2020), p. 101966.

[102] W. Shi and Y. Zhou, *Effectiveness of dropout and batch normalization in eeg classification tasks*, Journal of Neuroscience Methods, 285 (2024), pp. 61–72.

[103] J. Shih, S. Wang, and B. He, *Deep learning for eeg-based brain-computer interfaces: A survey*, IEEE Transactions on Biomedical Engineering, 67 (2020), pp. 1–13.

[104] A. Smith and R. Johnson, *Deep learning for eeg signal classification: Overfitting and regularization techniques*, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29 (2021), pp. 576–585.

[105] e. a. Smith, *Challenges in decoding visual stimuli from eeg: A comparison with motor bcis*, Journal of Neural Engineering, 20 (2023), p. 066001.

[106] L. Smith and M. Kutas, *Time-frequency analysis of interhemispheric transfer in visual processing*, Biological Cybernetics, 91 (2004), pp. 1–10.

[107] T. Song, W. Zheng, C. Lu, Y. Zong, and X. Zhang, *A comprehensive review of few-shot learning techniques for eeg signal classification*, Frontiers in Neuroscience, 16 (2022), p. 1219133.

[108] Y. Song and J. Wang, *Spatial-spectral neural features for single-trial eeg classification in visual tasks*, IEEE Access, 9 (2021), pp. 47395–47405.

[109] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, *Dropout: A simple way to prevent neural networks from overfitting*, Journal of Machine Learning Research, 15 (2014), pp. 1929–1958.

[110] I. Sturm, S. Bach, W. Samek, and K.-R. Müller, *Interpretable deep neural networks for single-trial eeg classification*, Journal of Neuroscience Methods, 274 (2016), pp. 141–145.

[111] H.-C. Sun, H. Ban, M. Di Luca, and A. E. Welchman, *fMRI evidence for areas that process surface gloss in the human visual cortex*, Vision Research, 109 (2015), pp. 149–157.

[112] C. Tallon-Baudry and O. Bertrand, *Oscillatory gamma activity in humans and its role in object representation*, Trends in Cognitive Sciences, 3 (1999), pp. 151–162.

[113] H. Van and Y. Lee, *Convolutional neural networks for eeg classification: A review*, Frontiers in Neuroscience, 17 (2023), pp. 1241–1258.

[114] H. Van Praag, *The effects of random stimuli on brain activity in theta and delta bands*, Neuroscience Letters, 742 (2019), pp. 48–54.

[115] R. VanRullen and S. J. Thorpe, *The time course of visual perception: From the retina to the decision*, Visual Cognition, 8 (2001), pp. 437–455.

[116] M. Vaziri, B. He, and Z. Wang, *Limitations of transformer models in eeg classification: Challenges and alternatives*, Journal of Neuroscience Methods, 365 (2023), p. 109406.

[117] H. Wang and Z. Lin, *Multiview spectrogram learning for robust eeg classification in visual tasks*, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31 (2023), pp. 56–65.

[118] J. Wang, W. Ye, J. He, L. Zhang, G. Huang, Z. Yu, and Z. Liang, *Integrating biological and machine intelligence: Attention mechanisms in brain-computer interfaces*, arXiv preprint arXiv:2502.19281, (2025).

[119] J. Wang, Z. Zhang, and H. Chen, *Hybrid brain-machine interfaces: Integration of eeg with facial emg for assistive applications*, Journal of Neuroengineering and Rehabilitation, 22 (2025), pp. 111–120.

[120] L. Wang and Y. Zhang, *Temporal pattern recognition with bidirectional lstms for eeg classification*, IEEE Transactions on Biomedical Engineering, 70 (2023), pp. 1681–1693.

[121] Y. Wang and X. Zhao, *Dimensionality reduction of time-frequency eeg features using t-sne for visual stimulus classification*, Cognitive Neurodynamics, 14 (2020), pp. 587–598.

[122] 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).

[123] Z. Xu and M. Chen, *Single-trial visual stimulus classification using wavelet-based spectro-temporal features*, Biomedical Signal Processing and Control, 68 (2021), p. 102705.

[124] Y. Yadid, H. Lior, and S. Moses, *Deep learning in eeg signal analysis: A survey*, Neural Computation, 33 (2021), pp. 2222–2239.

[125] Y. Yang, Z. Zhao, Q. Wang, Y. Yang, and J. Chen, *Mapping eeg signals to visual stimuli: A deep learning approach to match vs. mismatch classification*, arXiv preprint, (2023).

[126] Y. Yang, Z. Zhao, Q. Wang, Y. Yang, and J. Chen, *Self-supervised learning based emotion recognition using eeg*, Frontiers in Human Neuroscience, 18 (2024), p. 1334721.

[127] Z. Yang and L. Zhou, *Attention-enhanced cnn for decoding visual stimulus from single-trial eeg spectrograms*, Pattern Recognition Letters, 167 (2023), pp. 75–82.

[128] F. Yu and C. Chen, *Multi-scale convolutional neural networks for single-trial visual eeg classification*, IEEE Transactions on Cognitive and Developmental Systems, 13 (2021), pp. 915–924.

[129] e. a. Zhang, *Deep learning for brain-computer interface: Current trends and challenges*, IEEE Transactions on Biomedical Engineering, 69 (2022), pp. 3170–3185.

[130] H. Zhang, B. Zhao, X. Yang, K. Shi, and M. Tan, *Cognitioncapturer: Brain-visual representation learning for eeg-based visual stimuli reconstruction*, arXiv preprint arXiv:2412.10489, (2024).

[131] M. Zhang and Y. Cui, *Self supervised learning based emotion recognition using physiological signals*, Frontiers in Human Neuroscience, 18 (2024), p. 1334721.

[132] T. Zhang and Y. Li, *Multi-class single-trial eeg classification using time-frequency features and ensemble learning*, Neurocomputing, 448 (2021), pp. 212–221.

[133] W. Zhang, F. Yang, and Y. Li, *Wearable eeg-based brain-computer interfaces for medical applications: Advances and challenges*, Sensors, 23 (2023), p. 1442.

[134] X. Zhang and J. Li, *Matching pursuit with chirplet atoms for visual evoked potential characterization*, Biomedical Signal Processing and Control, 63 (2021), p. 102206.

[135] Y. Zhang, Z. Guo, and S. Wang, *Eeg-based feature selection methods in brain-computer interfaces: A review*, Frontiers in Human Neuroscience, 15 (2021), p. 779709.

[136] Y. Zhang, X. Li, and L. Wang, *A comprehensive review of eeg-based brain-computer interfaces in healthcare: Current applications and future trends*, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 33 (2025), pp. 720–731.

[137] Y. Zhang, Q. Wu, L. Wang, and X. Li, *Self-supervised learning for electroencephalogram: A systematic review*, arXiv preprint, (2023).

[138] Y. Zhang, W.-L. Zheng, and B.-L. Lu, *Spectral-temporal feature extraction for eeg-based emotion recognition using deep learning*, Neurocomputing, 388 (2020), pp. 173–180.

[139] X. Zhao, L. Ma, and Z. Wang, *Motor imagery eeg signal classification with a multivariate time-frequency analysis*, Biomedical Engineering Online, 22 (2023), p. 25.

[140] X. Zheng, W. Chen, Y. You, Y. Jiang, M. Li, and T. Zhang, *Ensemble deep learning for automated visual classification using eeg signals*, Pattern Recognition, 102 (2020), p. 107147.

[141] X. Zheng, W. Chen, Y. You, Y. Jiang, M. Li, and T. Zhang, *Ensemble deep learning for automated visual classification using eeg signals*, Pattern Recognition, 102 (2020), p. 107433.

[142] Y. Zheng, Y. Zhang, W. Zhang, and M. Xu, *An erp-guided lstm model for visual stimulus classification from eeg*, arXiv preprint arXiv:2006.15357, (2020).

[143] L. Zhou, Y. Xu, and C. Wang, *A robust sparse time-frequency representation method for vep signal analysis*, Computers in Biology and Medicine, 133 (2021), p. 104561.

[144] Z. Zhou, M. R. Bernard, and A. B. Bonds, *Deconstruction of spatial integrity in visual stimulus detected by modulation of synchronized activity in cat visual cortex*, Journal of Neuroscience, 28 (2008), pp. 3759–3768.

[145] X. Zhu, Z. Ghahramani, and J. D. Lafferty, *Semi-supervised learning with label propagation*, Proceedings of the ICML, (2003).

[146] M. Zuo, X. Chen, and L. Sui, *Evaluation of machine learning algorithms for classification of visual stimulation-induced eeg signals in 2d and 3d vr videos*, Brain Sciences, 15 (2025), p. 75.