university of groningen / faculty of science and engineering / artificial intelligence

Exploring EEG Signal Analysis for Perceived Number Classification and Transferability Across Individuals

Bachelor's Thesis

Malte L. Krambeer, s4251970, m.l.krambeer@student.rug.nl
Supervisor: Dr. Lisa-Marie Vortmann

Abstract: This study investigates the feasibility of classifying EEG signals recorded during number perception and the transferability of these classification models across individuals. We explore distinct brain activity patterns in subjects viewing numeric stimuli, using EEG recordings from a consumer-grade, four-channel device. Our evaluation focuses on the distinguishability of signals both within and across subjects to identify individual and subject-independent patterns. While classification results do not support number-specific EEG recordings, statistical differences are observed in individual cases. Subjects vary in the informativeness of their data, with more number-specific information found across individuals in tempo-parietal regions and 300-400 ms after stimulus presentation. The present methodology appears limited in its capability to consistently detect number-specific information within or across individuals. Future research is encouraged to improve on our findings by utilizing sophisticated recording equipment, informative recording locations and temporal windows in EEG waveform, as well as extensive data collection. We will discuss implications of the present findings with respect to current theoretical knowledge, previous EEG classification endeavors, and practical applications in brain-computer interface development. - Study resources, additional results, and the appendix to this paper can be found at: `https://osf.io/z3ctp/`

1 Introduction

The fascinating concept of Mind Reading has long been a matter of fictional narratives and served as intriguing but fantastical fuel in many works of popular culture (Luckhurst, 2002). With advancements in modern technology, the once imaginary idea now seems to gradually become a realistic possibility (Coles, 1989; Nicolas-Alonso & Gomez-Gil, 2012; Norman, Polyn, Detre, & Haxby, 2006; Rathkopf, Heinrichs, & Heinrichs, 2023). Promising approaches conduct analyses of brain activity measures, which are thought to relate to specific neural patterns. Through the identification of these patterns, researchers aim to trace distinct cognitive processes, essentially reading the contents of the mind holding them (Craik, He, & Contreras-Vidal, 2019; Kumar, Saini, Roy, Sahu, & Dogra, 2018; Mishra, Sharma, & Bhavsar, 2021; Saeidi et al., 2021; Yi et al., 2022). The present research aims to explore these possibilities by examining electroencephalographic (EEG) measures of neural activation to classify numeric stimuli perceived by participants. By leveraging state-of-the art signal processing and machine learning techniques, we investigate the neural correlates of numeric perception, as well as the potential for realistic applications of their classification.

Research into the technological realizations of Mind Reading is exceedingly relevant, as it combines the findings of cognitive neuroscience, engineering, and machine learning in one scientific endeavor. Firstly, it provides insights about an individual's neural representation of numbers by identifying and characterizing the neural patterns associated with perceiving and processing numerical information. The ability to classify EEG signals suggests the presence of distinct neural representations for different numbers, offering insights into how the brain encodes numerical information. Secondly, by evaluating the transferability of these methods we gain a better understanding of the universality and variability of numerical representations in the brain. This knowledge can then inform the design of practical applications that extend and assist cognitive functioning, for example in individuals with learning disabilities. Overall, this strongly interdisciplinary approach advances both fundamental and practical knowledge and validates the ties between all areas of research involved.

Throughout our research, we systematically distinguish four steps that we see as critical elements of Mind Reading as enabled by technological means: First, we conceptualize thoughts as specific patterns of neural activity. Second, neural activity is made quantifiable using EEG, a non-invasive method that measures electrical activity on the scalp corresponding to neural activation (Chaddad, Wu, Kateb, & Bouridane, 2023; Khosla, Khandnor, & Chand, 2020). Third, after a collection and processing of recordings, signals occurring during the perception of numeric stimuli will be classified to interpret the type of stimuli perceived. Fourth and finally, the process of analysis and classification will be evaluated to identify systematic patterns of neural activity within and across individuals and stimuli. The existence of such patterns may inform the design of practical applications and validate or extend fundamental knowledge in cognitive neuroscience research. The methods applied in this research as well as later sections below will largely follow this four-step process.

1.1 Related Work

Current neuroscience evidence strongly supports the notion of localized brain regions activated during number perception. Research by Dehaene, Piazza, Pinel, and Cohen (2003) identified three parietal circuits responsible for number processing by examining fMRI activations during numerical tasks. Specifically, the horizontal segment of the intraparietal sulcus was identified as a core region in which activation scales with the processing of numbers and quantity, while more lateral areas activate during verbal and spatial number tasks. More recent research by Marlair, Crollen, and Lochy (2022) supports these findings with evidence from frequency-tagged EEG, which similarly identified the bilateral intraparietal sulci as a source of activity during number-related tasks. Additional findings by Arsalidou and Taylor (2011) further point to the involvement of parietal regions in number representation, while additional prefrontal regions are activated during number manipulation tasks. Nieder (2016) further identified specific 'number neurons' in the intraparietal sulcus and lateral prefrontal cortex in non-human primates, which encode numerosity abstractly across modalities. Emphasizing strong similarities in numerical representation between human and non-human primates, their research points to an evolutionary pre-adaptation for semantic number representations in the parietofrontal network. In summary, the current body of research indicates a potential of using EEG to access specific brain regions and record measurements of activation that distinctly relate to numeric perception.

The possibility of interpreting neural representations from EEG recordings has been demonstrated by a multitude of studies, which achieved substantial accuracy in their predictions (see Saeidi et al. (2021) for a review). Mahapatra and Bhuyan (2023) successfully classified imagined numbers of a single subject using a recurrent neural network (RNN) architecture, reporting around 95% accuracy with low-density EEG recordings. Similarly, Kumar et al. (2018) applied a random forest method to classify imagined speech, achieving 67-80% accuracy for a classification of digits, characters, and objects. Mishra et al. (2021) further validated the potential of EEG-based classification by using a convolutional neural network (CNN) to decode visual perception of digits, obtaining approximately 70% accuracy. Many other reports of successful EEG signal classification for various stimuli exist, using various methods ranging from statistical regression models to deep neural networks (Alazrai, Abuhijleh, Ali, & Daoud, 2022; Bird, Faria, Manso, Ekárt, & Buckingham, 2019; Kalita, 2023; Saeidi et al., 2021; Yi et al., 2022). The abundance of research provides strong support for the notion that EEG signals can reliably indicate specific neural representations and underlying cognitive processes.

Building on the success of classification methods working with measurements of neural activity, brain-computer interface (BCI) design has emerged as a fascinating field of study. BCIs, which enable direct communication between the brain and external devices without the need of peripheral nerves or muscles (Nicolas-Alonso & Gomez-Gil, 2012), offer potential for a variety of applications. Such applications include neuroprosthetics, assistive devices for individuals with physical disabilities, and cognitive enhancement tools (Baniqued et al., 2021; Bird et al., 2019; Nicolas-Alonso & Gomez-Gil, 2012). However, using non-invasive EEG for its high temporal resolution and affordability, BCIs suffer from low signal-to-noise ratios in high dimensional data and require computationally expensive preprocessing and classification methods. Still, successful BCI applications demonstrate the efficacy of EEG-based methods (see Baniqued et al. (2021) for a review), highlighting their potential to improve assistive technologies and cognitive tasks.

1.2 The Present Study

The presented research lays the foundation for the analyses done in this project, which aims to validate and extend the existing fundamental knowledge and investigate practical implications for BCI development. It has been shown that EEG signals can be classified to interpret such processes with substantial accuracy. To further validate these findings, the present study aims to use a lightweight, consumer-grade EEG recording device for data collection, which would greatly facilitate BCI development due to its affordability and ease of use. A replication of the classification accuracy encountered in existing works would have significant implications for BCI design, since current applications come at a substantial cost, making them unaffordable for most people. Additionally, this study seeks to extend fundamental knowledge by evaluating the similarities and differences in recorded patterns of neural activity across different participants. By investigating the generalization of classification methods and the consistency of numeric representation patterns, the present research may provide insights into the robustness of EEG-based classifications and their potential applicability across individuals. In the pursuit of these goals we thus hope to contribute to a growing body of knowledge on BCI development while validating the fundamental insights of cognitive neuroscience on which this development is rooted.

The remainder of this paper will follow the four steps of technological Mind Reading outlined above: First, we will experimentally induce the thought of numbers in subjects. Second, the subject's neural activity will be recorded using an EEG device. Third, we will develop two classification models that learn to interpret the EEG signals to predict the original stimuli that subjects perceived during recording. Fourth and finally, we evaluate the systems for their effectiveness, while statistically analyzing the EEG signals for patterns indicating significant similarities or differences of number representation between individuals. In doing so, we aim to answer two fundamental questions:

1. Is it possible to achieve above-chance number classification performance through the analysis of EEG signals using a Random Forest or Neural Network approach?
2. To what extent does a similarity in neural number representation exist between users, as evaluated through a comparison of EEG signals and classification models?

The first question intends to validate the proof of principle that has previously been provided by other works utilizing classification algorithms to interpret EEG data (Alazrai et al., 2022; Kumar et al., 2018; Mahapatra & Bhuyan, 2023; Mishra et al., 2021), and we expect to obtain similar findings by adapting their data processing and classification methods for our purpose. The second question is more open-ended, allowing for an exploration of the classification results and patterns in the EEG data itself. Based on neuroscience research on shared neural circuitry for number representation (Arsalidou & Taylor, 2011; Dehaene et al., 2003; Marlair et al., 2022; Nieder, 2016; Plodowski, Swainson, Jackson, Rorden, & Jackson, 2003), we expect to find significant similarities in EEG interpretability between individuals. Ultimately, we will evaluate our results under acknowledgement of the limitations, draw conclusions for practical applications, and discuss implications for future research.

2 Data Collection and Processing

This section outlines the first two steps of the research process, namely the creation of numeric perceptions in participants, and the capturing of corresponding neural activation using an EEG recording device. We thus talk about the experimental design, recording, data processing and cleaning methods, as well as the tools and software used to accomplish these tasks.

2.1 Methods

2.1.1 Experimental Setup

Defining the numbers one to six as the distinct experimental conditions, we chose to present numeric stimuli in the form of dice faces. This choice of stimulus was intended to present a number as conceptually as possible, preventing potential noise in neural activation that might result from a reading-process (when presenting digits or number words) or semantic interpretation (when presenting quantities of other objects). Previous research indicates differential processing for different numerical stimuli (Plodowski et al., 2003; Wei, Chen, Yang, Zhang, & Zhou, 2014), which highlights the need to induce a conceptual thought of numbers in participants. Additionally, the limitation to the numbers from one to six allowed for a more simplistic research design, in which less data was required to investigate classification performance and patterns in EEG signals.

The experimental design required participants to visually perceive sequences of stimulus- or response trials. Trials were presented in random order, with at least one stimulus trial in between response trials. Each stimulus trial consisted of two phases: (1) the baseline period, in which a blank white square is shown and neural activity is meant to return to a baseline, and (2) the stimulus period, in which a complete dice face is presented. Each phase lasted exactly two seconds. Response trials similarly start with a baseline period, but then prompt participants to indicate the most recently perceived stimulus using a number key. Response trials differed from stimulus trials in that they only continued once a response was given and provided feedback on the correctness of that response. Response trials were included to assess and focus the attention of participants during the task, and allowed for the removal of participant data when too many incorrect responses were made. A visualization of a recording session can be seen in Figure 2.1.

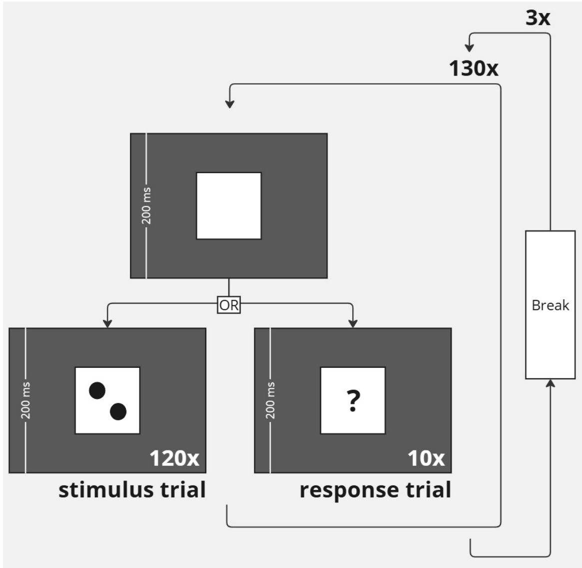

Figure 2.1: Flowchart of a typical recording session with three experimental blocks containing 130 trials each. Six subjects were asked to complete two such sessions, while the rest completed one.

2.1.2 Recording

EEG data were recorded using the Interaxon Muse 2 device, with a sampling rate of 256 Hz, over four channels: TP9, AF7, AF8, and TP10, which were placed according to the 10-20 international system (Klem, Lüders, Jasper, & Elger, 1999) and can be seen in Figure 2.2. The device was used to record neural activation within each experimental block, using the BlueMuse software at version 2.4 (Kowaleski, 2024). Recordings were synchronized with the stream of stimuli using the LabStreamingLayer (LSL), through the LabRecorder software (*Labstreaminglayer/App-LabRecorder*, 2024).

2.1.3 Processing and Cleaning

We performed various signal processing and cleaning steps to improve the signal-to-noise ratio of the recordings. For this purpose, the MNE python package (Gramfort et al., 2013; Larson et al., 2024) was used within a Python 3.4.12 environment (Guido Van Rossum & Drake, 2009). Firstly, we decided to exclude recordings from experimental blocks in which more than two out of the ten response trials were answered incorrectly, assuming a lack of attention to the task. After concatenating individual recordings from each experimental block, we applied a bandpass filter to extract frequencies from the range of 0.1 to 40 Hz, removing the DC component and higher frequencies. We thereby removed noise such as 50 Hz interference

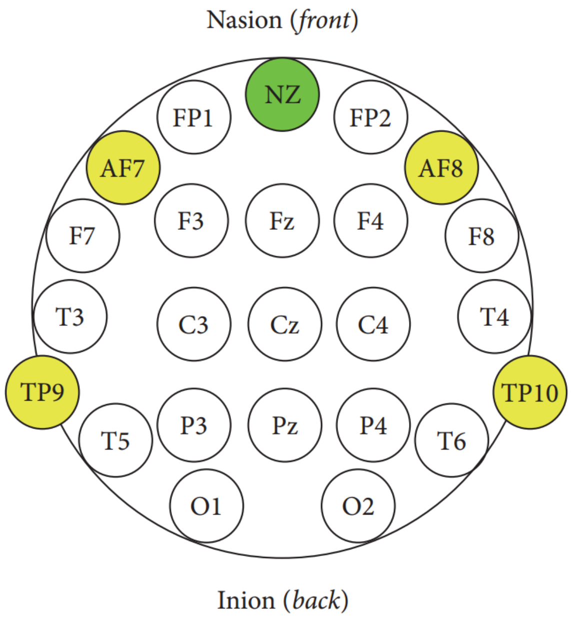

Figure 2.2: Electrode locations of the Muse2 headset. NZ indicates the location of a reference electrode.

resulting from power grid voltages. Ultimately, one raw EEG signal per participant was saved for further cleaning.

The additional processing steps consisted of epoching, baseline correction, outlier removal, detrending, and a discrete wavelet transformation. During epoching, we confined each signal to the interval of 0.2s before, and 0.5s after the presentation of the stimulus. In doing so, we allowed for an observation of signal changes in immediate response to the stimulus, and were able to apply a baseline correction using the mean amplitude of the 0.2s interval before stimulus application. Since initial recordings with the Muse 2 showed extreme fluctuations in amplitude during body movements or adjustment of the headband, we defined signals with amplitude values of two standard deviations from the mean as outliers. Along with epochs containing missing measurements, such outliers were removed from the data. The remaining signal was then detrended to correct for a potential increase in average amplitude, which may occur from external influence or an accumulated sensor error over time. Finally, we performed a discrete wavelet transformation (DWT) on each epoch to further remove artifacts and noise. The DWT decomposes an original signal into a set of wavelet coefficients at varying scales and positions. This is achieved by passing the signal through a sequence of high- and low-pass filters, separating the signal into different frequency components. Resulting coefficients provide a time-frequency representation of the signal, allowing an identification and removal of noise and artifacts while preserving the underlying neural activity (Chaddad et al., 2023; Dautov & Özerdem, 2018). We adopted the choice of a level three Daubechies-4 wavelet for the DWT, based on similar work on Muse recordings by Mahapatra and Bhuyan (2023). The use of this wavelet enabled us to retain or discard specific frequency components, enhancing the signal quality for further analysis without removing its distinctive features.

2.2 Results

EEG data was successfully recorded from all 18 participants, albeit with a considerable loss of data due to connectivity problems and the subsequent cleaning process. We recruited nine male and nine female participants for our study, who reported to have normal or lens-corrected eyesight and no learning disabilities. Their average age at the time of recording was 21.44 years ($SD = 2.39$), and two of the 18 participants reported left-handedness. Six participants (three male and three female) completed six experimental trials, resulting in an average of 706.3 ($SD = 65.83$) recorded signals out of 720 maximally possible. No recordings were discarded due to inattention during response trials. The data cleaning procedure resulted in an additional removal of signals which contained missing measurements or exceeded the amplitude threshold of two standard deviations. After removing an average of 107 signals for each of the six participants, the remaining number of recordings were reduced to 598.7 ($SD = 79.05$) across subjects. For the remaining twelve participants who completed three experimental trials, the average number of recordings was reduced from 284 ($SD = 157,45$) at the completion of the experiment to 216 ($SD = 150.5$) after cleaning. Ultimately, we are left with an average of 1031.3 ($SD = 13.16$) recordings per condition overall, out of the 1440 that were maximally possible, assuming perfect connectivity and signal-to-noise ratio. While the removal of noisy data can be expected in EEG research, the loss of recordings due to connectivity issues pose a serious limitation on the use of the Muse 2 headset for our research and its application for potential BCI implementations. Despite its ease of use and precautions against signal interruptions taken during the experiment, it appears that a reliable connection cannot be established in all cases. Nonetheless, a substantial amount of usable data has been collected, and further analyses and classification attempts can be made.

3 Classification

We made use of two approaches to examine the classifiability of the recorded EEG data. Firstly, we trained Random Forest models on frequency-band features extracted from the power density spectrum representation of each participant's EEG signal. Importantly, the Random Forest serves as an interpretable architecture, which can be compared between individuals to explore differences in classification and underlying EEG, highlighting its suitability for this study. Secondly, a recurrent neural network was trained on the EEG amplitude sequences to investigate the importance of temporal patterns for classification success. The following sections outline both architectures and their specific implementation in detail and present the results of their application to the collected data. In doing so, we aim to cover steps (3) and (4) of the research process: the classification of EEG recordings, and an evaluation of individual differences.

3.1 Methods

3.1.1 Random Forest

Random Forest (RF) classifiers are an ensemble learning method that combines multiple decision trees to enhance classification accuracy and robustness (Breiman, 2001). Each tree is trained on a different data subset using bootstrap aggregating, reducing overfitting and improving generalization. This approach is particularly effective for classifying data with feature vectors,as it can handle large datasets with many non-parametric features (Genuer, Poggi, & Tuleau-Malot, 2010). RF classifiers also evaluate feature importance by quantifying each feature's contribution to predictions (Breiman, 2001; Scornet, Biau, & Vert, 2015; Speiser, Miller, Tooze, & Ip, 2019). This allows for an identification and prioritization of the most significant features in their data, facilitating model comparison and selection based on the relative importance of different features for classification tasks. RF models are thus suited for classifying perceived numbers from EEG data and comparing individual models by assessing features that drive classification success.

RF models have successfully been applied to the task of EEG classification (Dweiri, Jadallah, Shannaq, & Alasasleh, 2022; Kumar et al., 2018; Zhang, Chen, & Li, 2018), demonstrating their suitability for the present study. Specifically the work of Kumar et al. (2018) highlights the ability of RF classifiers to distinguish between the imagined speech of digits, characters, and objects, based on features extracted from EEG recordings. Taking inspiration from their research, we extracted 140 features from the power spectral density of each signal. For each frequency band and electrode, seven metrics were computed which may reflect signal differences relating to specific number perceptions. Specifically, computed the (1) mean, (2) minimum, (3) maximum, (4) standard deviation, (5) root mean square, (6) simple sum, and (7) sum of squares for the power spectral density values of Delta ($0.1-4Hz$), Theta ($4-7Hz$), Alpha ($8-12Hz$), Beta ($12-25Hz$), and Gamma ($25-40Hz$) frequency bands (Nayak & Anilkumar, 2023). In doing so, we defined each signal as a feature vector of length $5\cdot 7\cdot 4 = 140$ ($n_{frequency\_bands}\cdot n_{metrics}\cdot n_{channels}$). For the training of various RF models, we utilized the entirety of recorded signals as well as subject-specific recordings for individual models, each of which were split into training (80%) and evaluation sets (20%). A comprehensive approach was intended to utilize the robustness and feature selection capabilities of RF classifiers to detect unique patterns in EEG data which link to the perceived number stimuli. We implemented the Random Forest classifier architecture using the sklearn library (Pedregosa et al., 2011) in Python. The final model specifications were determined after conducting a hyperparameter grid search across various parameters on an initial set of data, including data split criterion, the number of estimators, maximum depth, and bootstrap options. The final classifier used the Gini impurity criterion to measure the quality of splits at each node in the decision trees. The ensemble consisted of 500 individual decision trees of unrestricted depth, while the minimum number of samples required to split was set to two. In order to determine the best split at each node, the model considered a random subset of features equal to the square root of the total number of features. Such randomness further helped in reducing overfitting and improving generalization. Overall, the Random Forest classifier was intended to effectively handle numerous features, providing accurate and reliable predictions. For the classification of the EEG signals, as well as an evaluation of the underlying data in terms of informative features chosen by the model, we fitted an RF classifier to both the full set of data, and to subject-specific data sets. For these specific models, we utilized only the data from subjects who attended six experimental blocks, to ensure a larger amount of training data. The classifier was trained on 80% of data, with the remaining 20% serving for performance evaluation. The data split was randomized and stratified on the target variable, accounting for potential imbalances. Fitted models were evaluated on their final prediction performance in terms of accuracy and f1 score. A confusion matrix was generated, indicating the specific predictions on the validation set. In addition, we assessed the feature importances of models that achieved above-chance classification accuracy, to explore potential similarities and differences in the way such models make predictions.

3.1.2 Recurrent Neural Network

Recurrent Neural Networks (RNNs) are a class of neural networks designed to handle sequential data by maintaining a memory of previous inputs in the sequence. Unlike feedforward neural networks, RNNs have loops in their architecture, allowing information to persist and influence future inputs, which is essential for processing time-series data containing temporal dependencies (Sherstinsky, 2020). This method is thus especially useful for the classification of data characterized by sequential patterns, such as amplitude sequences in EEG recordings (Al-Saegh, Dawwd, & Abdul-Jabbar, 2021; Lipton, Berkowitz, & Elkan, 2015; Saeidi et al., 2021). Furthermore, advanced variants like Long Short-Term Memory (LSTM) networks address the limitations of traditional RNNs by handling long-range dependencies and mitigating issues like vanishing gradients (Hochreiter & Schmidhuber, 1997). The ability to model temporal relationships makes RNNs suitable for classifying EEG signals, where the sequence of amplitude values over time can provide insights into the perceived stimuli. For our purpose, RNN models are thus well-suited to demonstrate the feasibility of classifying perceived numbers based on amplitude sequences from EEG data.

The successful application of RNNs for the task of EEG classification has been demonstrated previously (Ma, Wang, Hu, Zhang, & Hua, 2021; Saeidi et al., 2021, see Craik et al. (2019) for a review). Specifically, Mahapatra and Bhuyan (2023) report the use of a recurrent neural network for the classification of Muse-recorded EEG with high accuracy. We took inspiration from their multilayer bidirectional LSTM architecture, which was reported to classify EEG recordings of the imagined speech of single-digit numbers with an accuracy of 96.18% (Mahapatra & Bhuyan, 2023). Following their example, our model consisted of three layers of bidirectional LSTM cells, with 440, 220, and 110 units in layers 1, 2, and 3, respectively. The input to the model was a sequence of 180 timesteps, with each timestep represented by a feature vector of 4 elements containing the z-standardized measurements per electrode at this time. The output layer was a dense layer with six units, using softmax activation to provide a confidence estimate for each classifiable number. The network was trained to minimize categorical cross entropy loss using the Adam optimizer with a learning rate of 0.001. The model and training schedule were implemented using the tensorflow package in python (TensorFlow Developers, 2024). Models were trained over ten epochs using a batch size of 32, using the same data splitting procedure as for the RF models.

Successfully classifying EEG signals to match the stimuli presented during recording will support the notion of distinctive neural activity relating to number perception. In addition, such a finding would complement the results of the Random Forest-based approach outlined above, since the informative value of a time or frequency-based representation of the EEG signals may differ. Different performances may thus inform future applications of EEG-based number classification, both in terms of data representation and choice of classifier architecture.

3.2 Results

3.2.1 Random Forest

Subject-Independent Effects We fitted one Random Forest classifier to the entirety of recorded data, testing the assumption that EEG signals across individuals contain distinct patterns relating to perceived numbers. However, an evaluation of the model's performance does not support this assumption. The final RF classifier predicts perceived numbers with an accuracy of 16.16% on the evaluation data, slightly less than the random guessing accuracy of 16.67%. An additional 20-fold cross validation run on subsets of the stratified training data produced an average accuracy of 16.65% ($SD = 0.35$), indicating that the model indeed performs at baseline accuracy. Investigating its feature importances, the quantification of the contribution each feature makes to the prediction accuracy of the model, we observed average importances of 0.007 ($SD = 0.001$) across the set of 140 features. An overview of all feature importances can be found in Appendix 1*. The low average, in conjunction with low variability between importances points to a lack of predictive value across all features. The failure of the model to learn from its training data is further exemplified by the pattern of predictions shown in Figure 3.1 and the F1-scores per condition listed in Table 3.1. The model appeared to predominantly predict the three most common conditions, here the numbers 1, 3, and 6, of which a few more training examples existed. Overall, we can conclude that no shared patterns exist in frequency spectrum information of the EEG signals, from which an RF classification algorithm could predict perceived numbers.

Subject-Dependent Effects Since the creation of an overarching RF model trained on all recorded EEG did not succeed, we investigated the possibility of subject-specific activation patterns by fitting RF models to individual data from the six participants attending six experimental blocks. The results parallel those of the overarching model. A low average classification accuracy of 19.59% ($SD = 4.04$) indicates a lack of distinct relations between signal features and perceived numbers. Low F1-scores, as well as confusion matrices further support the notion that the individual models performed at baseline, essentially guessing the target numbers. Additionally, overviews of the feature importances show a lack of significantly informative features, with no importance values exceeding 0.012 in any subject-specific classifier. The F1-scores on prediction performance on single numbers for each subject can be found in Table 3.1, while detailed reports, feature importances, and confusion matrices can be found in Appendix 1. While we can observe individual results exceeding baseline prediction performance, for instance in subject 1, these findings are likely chance effects. In summary, we conclude that the training of subject-specific RF classifiers using frequency features extracted from EEG signals cannot identify distinct patterns relating to perceived numbers.

*See `https://osf.io/z3ctp/` for all appendices

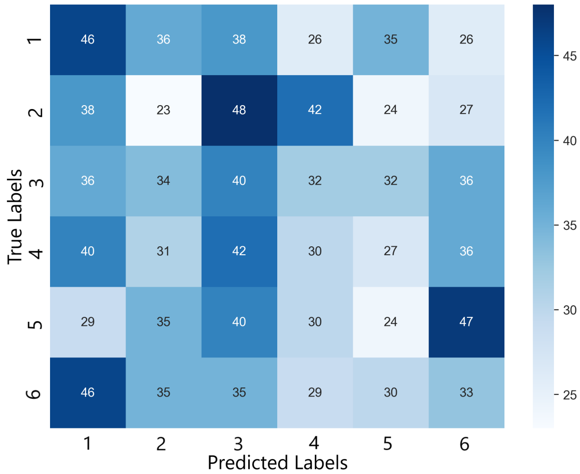

Figure 3.1: The confusion matrix of the Random Forest classifier trained on all available data.

3.2.2 Recurrent Neural Network

Subject-Independent Effects To investigate the subject-independent predictability of perceived numbers based on the temporal sequence of recorded EEG amplitudes, we trained a recurrent neural network on the entirety of recorded data. As with the RF models described above, its results do not support the notion of shared activity patterns that can be linked to perceived numbers. The

Digit	All Data	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6
1 (F1-score)	0.188	0.383	0.111	0.279	0.148	0.118	0.083
2 (F1-score)	0.116	0.216	0.154	0.182	0.108	0.2	0.204
3 (F1-score)	0.174	0.158	0.218	0.171	0.303	0.148	0.255
4 (F1-score)	0.153	0.279	0.217	0.111	0.235	0.103	0.158
5 (F1-score)	0.145	0.158	0.276	0.195	0.207	0.146	0.085
6 (F1-score)	0.182	0.343	0	0.35	0.312	0.14	0.216
Accuracy	0.162	0.261	0.169	0.219	0.219	0.142	0.165

Table 3.1: Classification results of the RF model on subject-independent and subject-specific EEG.

classifier performed with a near-chance accuracy of 19%, while individual F1-scores fluctuated around a mean of 0.181 ($SD = 0.055$), indicating low classification success for all conditions. Accuracy and F1-scores can be found in Table 3.2. Examining the specific predictions made by the model in a confusion matrix (Figure 3.2) further suggests the random guessing behaviour of the model, where no relation between target and predicted classes can be identified. In addition, the history of training and validation loss show a very early divergence of losses before the second epoch, after which validation loss increases and training loss decreases as the model starts to overfit. The loss history and a detailed classification report can be found in Appendix 2. In summary, these results support our previous findings with the RF classifier, which point to the absence of distinct patterns of activation in recorded data. It appears that, in addition to frequency features, no subject-independent patterns of temporal waveform can be identified.

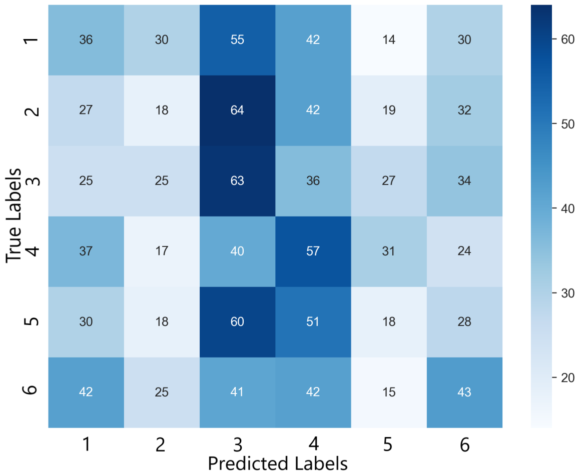

Figure 3.2: The confusion matrix of the RNN classifier trained on all available data.

Subject-Dependent Effects To investigate subject-specific patterns between conditions in EEG amplitude sequence, we fitted six additional models to data of the participants attending six experimental blocks. We thus extended the previous classification attempt using the RF approach with the RNN method previously applied to the complete data. Again, results do not support the notion of distinct patterns in EEG differing between experimental conditions. With an average classification accuracy of 0.185 ($SD = 0.027$) none of the models exhibited clear above-chance performance, which is further supported by low F1-scores and random confusion matrices. Furthermore, the loss trajectories of all subject-specific models recreated the development of the subject-independent model, where each model almost immediately begins to overfit. Both accuracy and F1-score metrics can be found in Table 3.2, while confusion matrices, detailed classification reports, and loss histories can be found in Appendix 2. Again, we conclude that, within subjects, an RNN could not identify temporal patterns in EEG amplitude that can reliably predict the number perceived.

4 Statistical Analysis

Since a classification of recorded signals proved unsuccessful, we performed statistical analysis on EEG waveforms to examine differences between experimental conditions for significance. The lack of classification success suggested that the recorded data does not contain distinct patterns relating to perceived numbers, both in temporal sequences and

Digit	All Data	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6
1 (F1-score)	0.178	0.222	0.108	0.125	0.214	0.164	0.227
2 (F1-score)	0.107	0.067	0.162	0.065	0.069	0.128	0.233
3 (F1-score)	0.236	0.14	0.163	0.333	0.148	0.163	0.178
4 (F1-score)	0.239	0.346	0.226	0.186	0.292	0.154	0.138
5 (F1-score)	0.109	0.195	0.207	0.133	0.188	0.163	0.261
6 (F1-score)	0.216	0.056	0.129	0.393	0	0.108	0.167
Accuracy	0.19	0.185	0.169	0.228	0.167	0.149	0.211

Table 3.2: Classification results of the RNN model on subject-independent and subject-specific EEG.

frequency spectrum information. However, it is possible that the classification methods were unsuitable for this specific task or data. To investigate this possibility, we can directly compare recorded signals using functional data analysis (FDA). In the sections below, we define the concept of FDA along with its application for our purpose and describe the results obtained.

4.1 Methods

Functional data analysis is a statistical tool suited to handle the increasing abundance of sequential data generated by advanced longitudinal measurement capabilities. Consequently, FDA usually considers many repeated measurements per subject, often represented as curves or continuous processes, and focuses on covariance structure and smoothing techniques (Martínez-Camblor & Corral, 2011). The main purpose of FDA is to provide methodologies for statistically describing and modeling these sets of functions or curves (Górecki & Smaga, 2015; Ramsay, 2005; Ramsay, Hooker, & Graves, 2009). In our analysis, FDA treated the EEG data as random functions, characterized by their mean and variance functions. This approach allowed us to examine and quantify differences between groups of data for statistical significance. These groups correspond to signals recorded during different experimental conditions, namely the perception of numbers one to six. Specifically, we applied functional ANOVA and post-hoc t tests to examine the potential differences between such groups. Previous studies have already demonstrated the efficacy of such methodology in EEG classification and analyses (Thivierge, 2007; Tian, 2010; Yi et al., 2022), leveraging the expression of EEG as a function over time. As a result, we see FDA as a useful tool for comparing groups of EEG signals, facilitating a functional analysis of differences between conditions.

4.1.1 Functional Analysis of Variance

Functional Analysis of Variance (fANOVA) is a statistical method used to analyze significant differences between groups of EEG waveforms. In this analysis, we investigated six such groups corresponding to the six experimental conditions for each channel to determine significant differences. Importantly, each channel was analyzed separately to capture localized patterns that may hold relevant information. The EEG signals consisted of 180 measurements taken over an interval of 0.7 seconds (0.2 seconds before and 0.5 seconds after the stimulus). The data were represented as six groups of independent random functions $X_{ij}(t)$, where $i \in [1, 6]$ and $j = 1, \ldots, n_i$, defined over the interval $T = [1, 180]$. We assumed these functions to be stochastic processes with mean functions $\mu_i(t), t \in T$ and covariance functions $\gamma(s, t), s, t \in T$ (Górecki & Smaga, 2015). The functional ANOVA tested the null hypothesis $H0 : \mu_1(t) = \mu_2(t) = \ldots = \mu_6(t), t \in T.$, implying that all groups of functions share the same mean function over time. The alternative hypothesis stated that the true mean functions are not equal across all groups (Górecki & Smaga, 2015). Tests of the null hypothesis are based on the pointwise F-test statistic (Ramsay, 2005), allowing us to examine whether the signals recorded under different conditions were significantly different both across and within individuals. This approach should provide more detailed insight into the EEG data than classification models alone, highlighting significant functional differences corresponding to varying conditions. We conducted a Functional Analysis of Variance (fANOVA) both across and within subjects, using data from all subjects. Tests will be conducted using the fdANOVA package (Górecki & Smaga, 2015) in R 4.4.1 (Team, 2024). Specifically, we used the permutation F test developed by Górecki and Smaga (2015), which handles functional data in the form of basis function representations. Basis function representations produce an approximation of a given function through a combination of simpler, predefined functions. In this case, the basis function representation was created using splines, which are piecewise polynomials connected smoothly at so-called knots. For the across-subjects analysis, we assumed that all signals originate from the same source and performed one fANOVA for each channel. Per channel, we will thus obtain an F statistic and a significance value, with significance values adjusted for multiple comparisons using Bonferroni correction (Bonferroni, 1936). Significant F statistics indicate the existence of differences between conditions across individuals. For the within-subjects analysis, we applied the same procedure to each subject's data. Significant F statistics indicate the existence of significant differences between conditions within a single subject. Based on the RF and RNN classification results, we expected to see no or few significant differences.

4.1.2 Functional Post-Hoc t Testing

In order to examine the results of the fANOVA more closely, and to find specific pairs of significantly different conditions, a series of functional post-hoc t tests were conducted. This approach paralleled the fANOVA described above in the assumptions it makes about the groups of functional data and tested the null-hypothesis of equal mean functions $H0 : \mu_1(t) = \mu_2(t)$ (Ramsay, 2005; Ramsay et al., 2009). The tests were conducted using the fda package (Ramsay, 2003) in R 4.4.1 (Team, 2024), and made use of its pointwise functional permutation t test. Again, tests were applied both across and within subjects, to investigate shared and individual effects. Specifically, we extracted every channel or subject-channel pair that has been found to contain significant differences between conditions during the across-subjects or within-subjects fANOVA, respectively. Signals from each condition were then transformed into basis function representations. A 1000-permutation t test was used to obtain a sequence of test statistics over the signal interval, as well as significance levels for the pointwise sequence and the overall test. One test was applied to each of the 15 pairs of conditions, and significance values were adjusted accordingly using Bonferroni correction (Bonferroni, 1936). Notably, the sequence of t statistics over the signals allowed for a localization of significant effects, where the global significance threshold is exceeded by a pointwise test result. As a consequence, the tests offer detailed insights into the temporal differences in EEG waveform between conditions. Assuming perfect predictability of perceived numbers based on EEG signals, we would have expected to see significant differences between all combinations of conditions (see Figure 4.2.1 for an illustration).

4.2 Results

4.2.1 Functional Analysis of Variance

Subject-Independent Effects Conducting a fANOVA for each channel on all collected signals independent of subjects, we found significant differences in both tempo-parietal (TP) channels. As shown in Table 4.1, we observed a greater variance between conditional mean functions than within conditions when considering the amplitude measurements of the TP9 channel ($F = 1.74, p < .001$). The same could be observed for recordings of the TP10 channel ($F = 2.47, p < .001$). Both variance ratios were interpreted as significant at $\alpha = 0.05$. For the anterior-frontal (AF) channels, we observed p-values above this threshold after correcting for multiple comparisons. As a result, we conclude that significant differences between conditions exist between groups of EEG as recorded by TP electrodes, but not in recordings of AF electrodes.

Subject-Dependent Effects An fANOVA conducted on groups of signals recorded for each subject suggests the existence of significant differences in some subject-channel pairs. Specifically, five out of the 72 pairs of subjects and channels indicate the existence of significant differences between experimental conditions. A complete list of F-statistics and significance values can be found in Appendix

Channel	F	p	p (corrected)
TP9	1.735	$< .001$	$< .001$
AF7	0.921	.752	3.008
AF8	1.202	.047	.188
TP10	2.472	$< .001$	$< .001$

Table 4.1: fANOVA results of subject-independent data.

3. Interestingly, three of the subject-channel pairs involved the TP10 channel with $F = 2.24$ ($p < .001$), $F = 1.38$ ($p = .008$), and $F = 1.31$ ($p = .044$), with one involving the TP9 channel ($F = 2.04, p < .001$). Recordings of the AF8 channel have been found to differ significantly in one subject (F=1.215, p=0.032). Finally, no differences have been found in recordings of the AF7 channel in any of the subjects. These findings parallel those of the subject-independent analysis, where conditions varied most notably in the TP10 channel recordings, followed by the TP9 recordings, with the AF8 recordings missing the significance threshold after correction. It can be concluded that differences between conditions can only be found in very few subjects and channels. The majority of the data did not allow a rejection of the null-hypothesis of equal mean functions between groups of EEG amplitude functions.

4.2.2 Functional Post-Hoc t Testing

Subject-Independent Effects Application of functional post-hoc t tests on groups of signals recorded from the TP channels indicate very low distinguishability between pairs of conditions. Considering the TP9 and TP10 channels, which were flagged as sources of significant differences in the previous fANOVA, we compared 15 pairs of conditions in each channel's recordings across individuals. Per channel, 4 pairs were found to differ significantly after correcting for multiple comparisons, which can be observed in the detailed results in Appendix 3. The results of these tests are illustrated in Figure 4.2.2, where significantly different pairs of conditions are represented as nodes connected by a line. Thicker edges between nodes indicate significance in both TP9 and TP10, which is the case only for pairs 3-4 and 3-6. For the pairs of conditions that do differ, the functional t test also provided a temporal localization of the difference, as can be seen in Figure 4.1. For the result of comparing numbers one and five in the TP9 recordings, it can be seen that a significant difference in waveform occurs between 300 and 400 ms after stimulus presentation. Other significant pairs of conditions, which can be seen in Appendix 3, exhibit a similar result and suggest a generally informative interval between 300 ms and 500 ms. Given additional significant findings, such a temporal evaluation can be helpful in identifying the time during which perceived numbers can be distinguished. However, the lack of more distinguishable pairs further points to an absence of informative value in our EEG data, with the individual findings lacking the consistent distinguishability that would be needed to conclude number-specific differences.

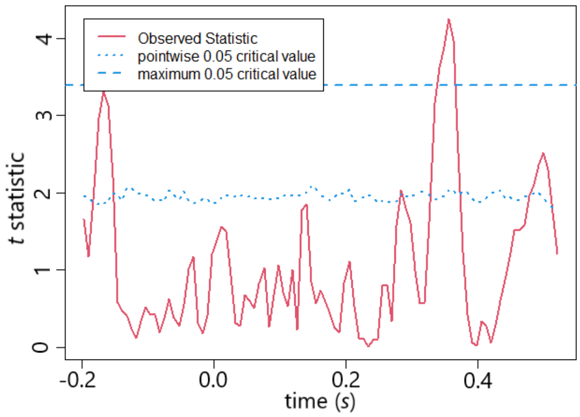

Figure 4.1: An example of pointwise t statistics for the functional t test of the subject-independent difference between numbers one and five. Both stimuli were presented at time 0.

Subject-Dependent Effects A post-hoc, within-subject analysis of differences between pairs of conditions provided similar results as the subject-independent analysis, but points to the existence of subject-specific differences in data quality. A functional t test was applied to test for the difference between 15 pairs of conditions for each of the four subject-channel pairs identified in the previous fANOVA. For the TP9 channel, 4 out of 15 comparisons were found to be significant, while 5 out of 45 comparisons of the TP10 channel were found to be significant. A detailed report of test results can be found in Appendix 3. Averaged results are illustrated in Figure 4.2.3, where thicker edges between nodes indicate consistently significant differences between conditions. Interestingly, all the significant results were obtained from the data of a single subject, with other subject's results proving insignificant after correction. This indicates a subject-specific difference in the informative value of data, while suggesting that previous fANOVA results may have rested predominantly on this subject's recordings. The subject's individual results are illustrated in comparison with cross-subject findings in Figure 4.2.4. An illustrative t test result shown in Figure 4.3 demonstrates a functional difference between conditions starting at 300 ms after stimulus presentation. In conjunction with other t test results of significant pairs, this finding supports the previous conclusion of a temporal localization of condition-specific information at 300-400 ms. The total of nine out of 30 significant test results do not suggest consistent distinguishability of this subject's recordings, however. The temporal exclusivity of a functional difference between signals, paired with the sparse occurrence across the four channels and 15 pairs of conditions, leaves very little information for classification models to work with. This sparsity of information also explains the previous inability of a subject-specific classifier to distinguish between perceived numbers. However, our identification of informative channels and time-periods may inform future designs of classification models based on EEG sequences.

In summary, the application of functional analysis methods offered more detailed insights into recorded data than previous classification methods did, identifying temporally and physically localized differences in EEG waveform between conditions. Despite the clear informative value of the TP channels in the detection of significant differences in both subject-independent and subject-specific analysis, the general signal-to-noise ratio is low. This is evidenced by an overall lack of significant differences between conditions in the data of most subjects, as well as the exclusion of the anterior-frontal channels as sources of relevant information. Nonetheless, the identification of subject-specific variations in data quality point to the possibility of a successful distinction between perceived numbers based on EEG waveform. Through a collection of additional data, and a greater focus on informative channels and temporally localized patterns, the creation of a successful classifier appears feasible in theory.

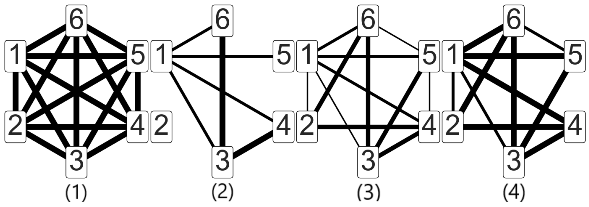

Figure 4.2: An illustration of functional t test results, showing significant differences in EEG waveform between pairs of conditions. Figure (1) demonstrates perfect distinguishability between conditions; (2) shows the results of the across-subject analysis, averaging the results across both TP channels investigated; (3) shows the results of the within-subject analysis, averaging the results across all subject-channel pairs; (4) shows the results of subject 1 alone, averaging across both TP channels. Thicker lines correspond to a larger proportion of significant differences across subjects and/or channels.

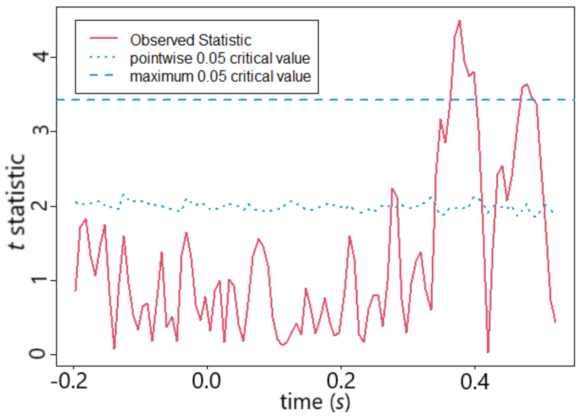

Figure 4.3: An example of pointwise t statistics for the functional t test of the difference between numbers one and four in recordings of subject 1. Both stimuli were presented at time 0.

5 Discussion

The present study aimed to validate and extend the current research on the analysis of number perception through an assessment of neural activity measurements. Seeking to contribute to a small-scale, technological realization of Mind Reading, we investigated the possibility of classifying visually perceived numbers. To that extent, neural activity was recorded using a 4-channel EEG device and classified using two machine learning architectures. Additionally, we explored the extent to which subject-independent classification is possible by evaluating and comparing classifiers trained on individual or overall data, and by statistically analyzing EEG data recorded during the perception of different numbers. Our findings aim to validate existing knowledge of the shared functional localization of number processing in the brain and to extend previous studies which succeeded in the task of EEG data classification. Furthermore, the overall endeavor of interpreting recordings of neural activity and the use of a consumer-grade EEG device informs the design of practical applications such as brain-computer interfaces.

The application of a Random Forest (RF) classifier, an ensemble method fitted to features extracted from frequency spectrum representations of signals, was unsuccessful in classifying subject-independent and subject-specific data. None of the models achieved notable prediction performance above baseline, either in terms of overall accuracy or condition-specific F1-scores. Additionally, an analysis of feature importances indicating the informative value of specific features for prediction success suggests an overall lack of information across all metrics and frequency bands assessed. As a result, we cannot conclude a proof of principle for the overall possibility of interpreting EEG signals by using a Random Forest model on our data. We thus fail to replicate the success of previous studies such as Kumar et al. (2018). This discrepancy can be attributed to several factors, for instance the lack of suitable data and the limited sensitivity of the EEG recording equipment. Notably, their use of a 14-channel headset with higher sampling rate and larger amounts of data from more participants may have provided a necessary advantage, especially since all their data was used for the training of a single overall model (Kumar et al., 2018). As suggested by findings in cognitive neuroscience (Dehaene et al., 2003; Marlair et al., 2022; Nieder, 2016), as well as our own statistical analysis, the inclusion of parietal electrodes may be critical for the detection of number-specific information. The availability of such information marks an additional difference between our work and that of Kumar et al. (2018). Independent of such limitations, however, a frequency spectrum-based approach may prove unsuitable for the classification of perceived numbers. This is suggested by our statistical results, which show a temporal localization of difference effects between conditions. An approach evaluating a time-invariant frequency spectrum representation of the entire signal may disregard such effects, which would further explain the failure of our RF models to learn. It can be concluded that the use of RF models working with frequency domain features is to be discouraged when aiming to interpret perceived numbers. Future research may want to emphasize the use of time-dependent features or a different classification architecture altogether, while making sure to collect sufficient data, ideally including parietal sources.

The application of a recurrent neural network (RNN) model, interpreting the temporal sequence of EEG amplitudes, was also unsuccessful in classifying subject-independent and subject-specific data. Prediction performances in terms of accuracy and F1-scores, as well as loss trajectories both suggest a random-guessing behaviour due to a lack of information in the data. These results align with those of the RF model, indicating that neither frequency features nor temporal waveforms of EEG signals contain patterns predictive of perceived numbers. Our results contrast previous work such as that of Mahapatra and Bhuyan (2023), who succeeded in the use of a sequence-based classifier for the task of EEG classification. Again, several factors may have contributed to our inability to replicate their findings. Firstly, their work uses a very large dataset of a single subject, collected over 2 years (Vivancos & Cuesta, 2022), which may have provided more consistent learning data. More importantly however, their success may have been overstated, since a direct replication of their study using the same methods and data, as well as the present study using largely analogous methods, did not yield comparable results. Unfortunately, the authors did not provide their specific implementation or a statement on this discrepancy as of yet. In conjunction with the results of the RF classification above, we conclude that a significant limitation exists in the type and amount of data collected, while the efficacy of our specific RNN implementation is questioned. Furthermore, the classification results do not support the existence of distinct, classifiable patterns in the temporal domains of EEG signals. While the effectiveness of RNN-based classifiers should remain under investigation, statistical results do support their suitability for perceived number classification as a result of their capability to detect temporal features in amplitude sequences. Future research is thus encouraged to continue exploring their usability, while aiming to collect larger amounts of higher-dimensional data.

The application of functional data analysis methods, specifically the functional analysis of variance (fANOVA) and functional t-tests, allowed for some more nuanced insights into EEG waveforms and the existence of shared features. The fANOVA identified significant differences between experimental conditions in the tempo-parietal (TP) channels, but not anterior-frontal (AF) channels. This finding is in line with cognitive neuroscience research pinpointing the parietal lobes as most active during number processing (Dehaene et al., 2003; Marlair et al., 2022; Nieder, 2016). While results based on the four electrode locations investigated here are limited, a greater focus on more parietal recording locations may prove useful for future applications of perceived number classification. Further post-hoc tests examining differences between pairs of conditions identified very few sources in which significant differences could be observed. In the across-subject analysis, only four pairs of conditions could be distinguished from each other in either TP channel, indicating a lack of number-specific waveform patterns that would be needed for predictability. Similarly, a subject-specific analysis identified a single subject whose data contained significantly different waveforms between conditions. This points to a subject-specific difference in the informative value of EEG data, but also suggests the potential for better distinguishability between conditions when more higher-quality recordings can be obtained. Nonetheless, it emphasizes the sparsity of information contained in the collected data and explains the lack of success of our previous classification attempts. Interestingly, the functional t-tests allowed for a temporal analysis of amplitude sequences, and consistently identified the interval of 300-400 ms as the source of significant differences. While the absolute number of differences itself is low, this finding suggests a pattern which can inform subsequent analyses and prediction attempts. If future research manages to identify this time interval as a consistent source of information, applications of EEG classification may perform efficiently with lower amounts of data. This will be especially useful if the inclusion of additional electrodes proves necessary, since higher-dimensional data can still be reduced to maintain fast performance. Consequently, optimizing data collection and focusing on this informative time window could pave the way for more successful EEG-based classification models.

Our first question of research - '*Is it possible to achieve above-chance number classification performance through the analysis of EEG signals using a Random Forest or Neural Network approach?*' - must be concluded with a 'No' in light of the present results, methodology, and data. Importantly, this null-result is limited to the context of the present study, and is likely not indicative of a general impossibility of its overall aim. As previously discussed, recent research has provided strong support for the general feasibility of EEG-based classification of perceived numbers. The inability of our results to further validate this idea is likely due to the study's limitations already broached above, but will be discussed in more detail below. With respect to the second question of research - '*To what extent does a similarity in neural number representation exist between users, as evaluated through a comparison of EEG signals and classification models?*' - we can draw subject-independent and subject-specific conclusions: Overall, we observe a lack of patterns, shared or otherwise, in recorded EEG data. Across individuals, there appears to be a sparsity of significant differences which can be detected between pairs of perceived numbers. Nonetheless, it appears that electrodes in the tempo-parietal regions are more sensitive to such differences when they do occur, independent of subject identity. Furthermore, our data suggest a subject-specific variability in the informative value of EEG recordings. This finding may result from individual differences in neural activation in response to perceived numbers, but also from differences in behaviour or conditions during the experiment. A differentiation of such sources of variability in signal-to-noise ratio is critical for the further development of subject-independent approaches, and should be pursued in future research. Another result which is likely to generalize across individuals is the temporal localization of differences between EEG waveforms. Despite overall significant differences being sparse, a somewhat shared temporal localization is still conceivable. Since the identification of visually presented numbers is likely realized in a biologically-determined sequence of early, low-level perceptual processes and later semantic interpretation, there should be a common window during which number-specific differences are most salient. Similar to the P300 ERP component, which appears during the perception of unexpected stimuli (Valakos et al., 2020; van Dinteren, Arns, Jongsma, & Kessels, 2014), a number-specific difference in EEG may be located at the intersection of visual perception and semantic processing. In conclusion, despite the limitations in the present study, the potential for EEG-based number classification remains promising, encouraging future efforts to concentrate on enhancing data quality, understanding individual differences, and identifying shared temporal patterns.

While we perceive the theoretical and methodological foundation of the present study as robust, several factors must be considered as potential causes for an inability to reliably classify recorded EEG. Firstly, we may have underestimated the amount of data required to effectively train classification models. Unexpected loss of data during recording, in addition to further removal during data cleaning likely resulted in insufficient information for the classifiers to learn from. To address this, further research should aim to collect extensive data sets, potentially focusing on fewer subjects to ensure depth over breadth. Secondly, the quality of collected data may have been compromised by several factors, such as interference during collection, and differences in participants' approaches to the experimental task. For instance, some might internally visualize or vocalize numbers upon presentation, or use their hands to represent numbers physically for easier recall during response trials. Furthermore, recording on different days may have contributed to variation between subject-specific signals. However, we aimed to investigate realistic applications of our approach for instance in BCI implementations, which should be robust to such variability in recordings. It must therefore be concluded that the presently used equipment and processing methods do not suffice to adequately capture the necessary information. Future studies should prioritize high-quality recordings to establish a proof of principle before investigating the robustness of working classification methods to variable participant behaviour. Third, and finally, we see a major limitation in the use of a four-channel recording device for our purpose. Specifically the Muse 2 model used may have been inadequate, given that channel locations do not cover the parietal areas identified to activate during number processing. Additionally, a low sampling rate and connection issues during recording further limit the device's utility. Future studies may want to employ more comprehensive recording systems with additional parietal electrodes and higher sampling rate to capture detailed and relevant neural activity.

Our findings suggest the current infeasibility of performing perceived number classification using practical applications such as brain-computer interfaces (BCIs). Lightweight, consumer-grade headsets like the Muse2, while cost-effective and easy to use, are likely insufficient for capturing the necessary information for the task. Additionally, our results indicate that subject-independent information is unreliable for creating out-of-the-box BCI tools that perform universally without additional fine-tuning. Instead, substantial amounts of data are needed to train reliable and accurate classifiers, especially when creating subject-specific applications. However, considering these suggestions, future research may progress towards practical applications of EEG classification. Specifically when realizing perceived number identification, we determined that a focus on parietal electrodes and temporal windows in waveform may provide additional benefits. While current methods may face limitations, the insights gained suggest key areas for improvement and future exploration. With targeted efforts, the potential of reliable EEG-based number classification - and a small realization of actual Mind Reading - remains within reach.

References

Alazrai, R., Abuhijleh, M., Ali, M. Z., & Daoud, M. I. (2022, October). A deep learning approach for decoding visually imagined digits and letters using time–frequency–spatial representation of EEG signals. *Expert Systems with Applications*, *203*, 117417. Retrieved 2024-03-19, from https://linkinghub.elsevier.com/retrieve/pii/S0957417422007576 doi: 10.1016/j.eswa.2022.117417

Al-Saegh, A., Dawwd, S. A., & Abdul-Jabbar, J. M. (2021, January). Deep learning for motor imagery EEG-based classification: A review. *Biomedical Signal Processing and Control*, *63*, 102172. Retrieved 2024-07-01, from https://www.sciencedirect.com/science/article/pii/S1746809420303116 doi: 10.1016/j.bspc.2020.102172

Arsalidou, M., & Taylor, M. J. (2011, February). Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage*, *54*(3), 2382–2393. Retrieved 2024-02-25, from https://linkinghub.elsevier.com/retrieve/pii/S1053811910013017 doi: 10.1016/j.neuroimage.2010.10.009

Baniqued, P. D. E., Stanyer, E. C., Awais, M., Alazmani, A., Jackson, A. E., Mon-Williams, M. A., ... Holt, R. J. (2021, January). Brain–computer interface robotics for hand rehabilitation after stroke: a systematic review. *Journal of NeuroEngineering and Rehabilitation*, *18*(1), 15. Retrieved 2024-06-23, from https://jneuroengrehab.biomedcentral.com/articles/10.1186/s12984-021-00820-8 doi: 10.1186/s12984-021-00820-8

Bird, J. J., Faria, D. R., Manso, L. J., Ekárt, A., & Buckingham, C. D. (2019, March). A Deep Evolutionary Approach to Bioinspired Classifier Optimisation for Brain-Machine Interaction. *Complexity*, *2019*, 1–14. Retrieved 2024-06-21, from https://www.hindawi.com/journals/complexity/2019/4316548/ doi: 10.1155/2019/4316548

Bonferroni, C. E. (1936). *Teoria statistica delle classi e calcolo delle probabilità*. Seeber.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. Retrieved 2024-03-05, from http://link.springer.com/10.1023/A:1010933404324 doi: 10.1023/A:1010933404324

Chaddad, A., Wu, Y., Kateb, R., & Bouridane, A. (2023, July). Electroencephalography Signal Processing: A Comprehensive Review and Analysis of Methods and Techniques. *Sensors*, *23*(14), 6434. Retrieved 2024-06-25, from https://www.mdpi.com/1424-8220/23/14/6434 doi: 10.3390/s23146434

Coles, M. G. H. (1989). Modern Mind-Brain Reading: Psychophysiology, Physiology, and Cognition. *Psychophysiology*, *26*(3), 251–269. Retrieved 2024-06-25, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8986.1989.tb01916.x doi: 10.1111/j.1469-8986.1989.tb01916.x

Craik, A., He, Y., & Contreras-Vidal, J. L. (2019, April). Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of Neural Engineering*, *16*(3), 031001. Retrieved 2024-07-01, from https://dx.doi.org/10.1088/1741-2552/ab0ab5 (Publisher: IOP Publishing) doi: 10.1088/1741-2552/ab0ab5

Dautov, Ç. P., & Özerdem, M. S. (2018, May). Wavelet transform and signal denoising using Wavelet method. In *2018 26th Signal Processing and Communications Applications Conference (SIU)* (pp. 1–4). Retrieved 2024-06-29, from https://ieeexplore.ieee.org/document/8404418 doi: 10.1109/SIU.2018.8404418

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003, May). THREE PARIETAL CIRCUITS FOR NUMBER PROCESSING. *Cognitive Neuropsychology*, *20*(3-6), 487–506. Retrieved 2024-02-20, from http://www.tandfonline.com/doi/abs/10.1080/02643290244000239 doi: 10.1080/02643290244000239

Dweiri, Y., Jadallah, S., Shannaq, Y., & Alasasleh, A. (2022, April). Sleep Stage Classification Using Random Forest Method. In *Proceedings of the 12th International Conference on Biomedical Engineering and Technology* (pp. 84–88). Tokyo Japan: ACM. Retrieved 2024-04-17, from https://dl.acm.org/doi/

10.1145/3535694.3535709 doi: 10.1145/3535694.3535709

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010, October). Variable selection using random forests. *Pattern Recognition Letters*, *31*(14), 2225–2236. Retrieved 2024-03-18, from https://linkinghub.elsevier.com/retrieve/pii/S0167865510000954 doi: 10.1016/j.patrec.2010.03.014

Górecki, T., & Smaga, Ł. (2015, December). A comparison of tests for the one-way ANOVA problem for functional data. *Computational Statistics*, *30*(4), 987–1010. Retrieved 2024-06-10, from http://link.springer.com/10.1007/s00180-015-0555-0 doi: 10.1007/s00180-015-0555-0

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... Hämäläinen, M. (2013, December). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*. Retrieved 2024-06-29, from https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2013.00267/full (Publisher: Frontiers) doi: 10.3389/fnins.2013.00267

Guido Van Rossum, & Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace. Retrieved 2024-06-29, from https://docs.python.org/3/reference/index.html

Hochreiter, S., & Schmidhuber, J. (1997, December). Long Short-term Memory. *Neural computation*, *9*, 1735–80. doi: 10.1162/neco.1997.9.8.1735

Kalita, D. (2023, July). *Decoding Thoughts with Deep Learning: EEG-Based Digit Detection using CNNs.* Retrieved 2024-02-24, from https://dxganta.medium.com/decoding-thoughts-with-deep-learning-eeg-based-digit-detection-using-cnns-cdf7eee20722

Khosla, A., Khandnor, P., & Chand, T. (2020, April). A comparative analysis of signal processing and classification methods for different applications based on EEG signals. *Biocybernetics and Biomedical Engineering*, *40*(2), 649–690. Retrieved 2024-06-25, from https://www.sciencedirect.com/science/article/pii/S0208521620300231 doi: 10.1016/j.bbe.2020.02.002

Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalography and Clinical Neurophysiology. Supplement*, *52*, 3–6.

Kowaleski, J. (2024, May). *kowalej/BlueMuse.* Retrieved 2024-06-29, from https://github.com/kowalej/BlueMuse (original-date: 2017-08-19T05:24:36Z)

Kumar, P., Saini, R., Roy, P. P., Sahu, P. K., & Dogra, D. P. (2018, February). Envisioned speech recognition using EEG sensors. *Personal and Ubiquitous Computing*, *22*(1), 185–199. Retrieved 2024-02-24, from http://link.springer.com/10.1007/s00779-017-1083-4 doi: 10.1007/s00779-017-1083-4

Labstreaminglayer/app-labrecorder. (2024, June). LabStreamingLayer submodules. Retrieved 2024-06-29, from https://github.com/labstreaminglayer/App-LabRecorder (original-date: 2018-02-28T12:17:09Z)

Larson, E., Gramfort, A., Engemann, D. A., Leppakangas, J., Brodbeck, C., Jas, M., ... luzpaz (2024, June). *MNE-Python.* Zenodo. Retrieved 2024-06-29, from https://zenodo.org/doi/10.5281/zenodo.11662646 doi: 10.5281/ZENODO.11662646

Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015, October). *A Critical Review of Recurrent Neural Networks for Sequence Learning.* arXiv. Retrieved 2024-07-01, from http://arxiv.org/abs/1506.00019 (arXiv:1506.00019 [cs]) doi: 10.48550/arXiv.1506.00019

Luckhurst, R. (2002). *The Invention of Telepathy, 1870-1901.* Oxford University Press.

Ma, Q., Wang, M., Hu, L., Zhang, L., & Hua, Z. (2021, March). A Novel Recurrent Neural Network to Classify EEG Signals for Customers' Decision-Making Behavior Prediction in Brand Extension Scenario. *Frontiers in Human Neuroscience*, *15*. Retrieved 2024-07-01, from https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2021.610890/full (Publisher: Frontiers) doi: 10.3389/fnhum.2021.610890

Mahapatra, N. C., & Bhuyan, P. (2023, April).

EEG-based classification of imagined digits using a recurrent neural network. *Journal of Neural Engineering*, *20*(2), 026040. Retrieved 2024-02-20, from https://iopscience.iop.org/article/10.1088/1741-2552/acc976 doi: 10.1088/1741-2552/acc976

Marlair, C., Crollen, V., & Lochy, A. (2022, August). A shared numerical magnitude representation evidenced by the distance effect in frequency-tagging EEG. *Scientific Reports*, *12*(1), 14559. Retrieved 2024-02-20, from https://www.nature.com/articles/s41598-022-18811-7 doi: 10.1038/s41598-022-18811-7

Martínez-Camblor, P., & Corral, N. (2011, December). Repeated measures analysis for functional data. *Computational Statistics & Data Analysis*, *55*(12), 3244–3256. Retrieved 2024-07-08, from https://linkinghub.elsevier.com/retrieve/pii/S0167947311002106 doi: 10.1016/j.csda.2011.06.007

Mishra, R., Sharma, K., & Bhavsar, A. (2021, August). Visual Brain Decoding for Short Duration EEG Signals. In *2021 29th European Signal Processing Conference (EUSIPCO)* (pp. 1226–1230). Dublin, Ireland: IEEE. Retrieved 2024-02-20, from https://ieeexplore.ieee.org/document/9616192/ doi: 10.23919/EUSIPCO54536.2021.9616192

Nayak, C. S., & Anilkumar, A. C. (2023). EEG Normal Waveforms. In *StatPearls*. Treasure Island (FL): StatPearls Publishing. Retrieved 2024-07-03, from http://www.ncbi.nlm.nih.gov/books/NBK539805/

Nicolas-Alonso, L. F., & Gomez-Gil, J. (2012, January). Brain Computer Interfaces, a Review. *Sensors*, *12*(2), 1211–1279. Retrieved 2024-03-19, from http://www.mdpi.com/1424-8220/12/2/1211 doi: 10.3390/s120201211

Nieder, A. (2016, June). The neuronal code for number. *Nature Reviews Neuroscience*, *17*(6), 366–382. Retrieved 2024-03-18, from https://www.nature.com/articles/nrn.2016.40 doi: 10.1038/nrn.2016.40

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006, September). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. Retrieved 2024-06-25, from https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(06)00184-7 (Publisher: Elsevier) doi: 10.1016/j.tics.2006.07.005

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011, November). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, *12*(null), 2825–2830.

Plodowski, A., Swainson, R., Jackson, G. M., Rorden, C., & Jackson, S. R. (2003, December). Mental Representation of Number in Different Numerical Forms. *Current Biology*, *13*(23), 2045–2050. Retrieved 2024-06-28, from https://linkinghub.elsevier.com/retrieve/pii/S0960982203008650 doi: 10.1016/j.cub.2003.11.023

Ramsay, J. (2003, December). *fda: Functional Data Analysis.* Retrieved 2024-07-07, from https://CRAN.R-project.org/package=fda (Institution: Comprehensive R Archive Network Pages: 6.1.8) doi: 10.32614/CRAN.package.fda

Ramsay, J. (2005, April). Functional Data Analysis. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (1st ed.). Wiley. Retrieved 2024-07-07, from https://onlinelibrary.wiley.com/doi/10.1002/0470013192.bsa239 doi: 10.1002/0470013192.bsa239

Ramsay, J., Hooker, G., & Graves, S. (2009). *Functional Data Analysis with R and MATLAB.* New York, NY: Springer New York. Retrieved 2024-06-13, from https://link.springer.com/10.1007/978-0-387-98185-7 doi: 10.1007/978-0-387-98185-7

Rathkopf, C., Heinrichs, J. H., & Heinrichs, B. (2023, February). Can we read minds by imaging brains? *Philosophical Psychology*, *36*(2), 221–246. Retrieved 2024-06-25, from https://doi.org/10.1080/09515089.2022.2041590 doi: 10.1080/09515089.2022.2041590

Saeidi, M., Karwowski, W., Farahani, F. V., Fiok, K., Taiar, R., Hancock, P. A., & Al-Juaid, A. (2021, November). Neural Decoding of EEG Signals with Machine Learning: A

Systematic Review. *Brain Sciences*, *11*(11), 1525. Retrieved 2024-05-06, from https://www.mdpi.com/2076-3425/11/11/1525 doi: 10.3390/brainsci11111525

Scornet, E., Biau, G., & Vert, J.-P. (2015, August). Consistency of random forests. *The Annals of Statistics*, *43*(4). Retrieved 2024-03-05, from https://projecteuclid.org/journals/annals-of-statistics/volume-43/issue-4/Consistency-of-random-forests/10.1214/15-AOS1321.full doi: 10.1214/15-AOS1321

Sherstinsky, A. (2020, March). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *Physica D: Nonlinear Phenomena*, *404*, 132306. Retrieved 2024-07-01, from http://arxiv.org/abs/1808.03314 (arXiv:1808.03314 [cs, stat]) doi: 10.1016/j.physd.2019.132306

Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019, November). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, *134*, 93–101. Retrieved 2024-03-05, from https://linkinghub.elsevier.com/retrieve/pii/S0957417419303574 doi: 10.1016/j.eswa.2019.05.028

Team, R. C. (2024). *R: A Language and Environment for Statistcial Computing.* R Foundation for Statistical Computing. Retrieved from www.R-project.org

TensorFlow Developers. (2024, June). *TensorFlow.* Zenodo. Retrieved 2024-07-01, from https://zenodo.org/doi/10.5281/zenodo.4724125 doi: 10.5281/ZENODO.4724125

Thivierge, J.-P. (2007, October). Functional data analysis of cognitive events in EEG. In *2007 IEEE International Conference on Systems, Man and Cybernetics* (pp. 2473–2478). Montreal, QC, Canada: IEEE. Retrieved 2024-07-07, from http://ieeexplore.ieee.org/document/4413811/ doi: 10.1109/ICSMC.2007.4413811

Tian, T. S. (2010). Functional Data Analysis in brain imaging studies. *Frontiers in Psychology*, *1*. Retrieved 2024-07-07, from http://journal.frontiersin.org/article/10.3389/fpsyg.2010.00035/abstract doi: 10.3389/fpsyg.2010.00035

Valakos, D., d'Avossa, G., Mylonas, D., Butler, J., Klein, C., & Smyrnis, N. (2020, June). P300 response modulation reflects breaches of non-probabilistic expectations. *Scientific Reports*, *10*(1), 10254. Retrieved 2024-07-17, from https://www.nature.com/articles/s41598-020-67275-0 (Publisher: Nature Publishing Group) doi: 10.1038/s41598-020-67275-0

van Dinteren, R., Arns, M., Jongsma, M. L. A., & Kessels, R. P. C. (2014, February). P300 Development across the Lifespan: A Systematic Review and Meta-Analysis. *PLoS ONE*, *9*(2), e87347. Retrieved 2024-07-17, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3923761/ doi: 10.1371/journal.pone.0087347

Vivancos, D., & Cuesta, F. (2022). MindBigData 2022 A Large Dataset of Brain Signals. Retrieved 2024-02-20, from https://arxiv.org/abs/2212.14746 (Publisher: arXiv Version Number: 1) doi: 10.48550/ARXIV.2212.14746

Wei, W., Chen, C., Yang, T., Zhang, H., & Zhou, X. (2014). Dissociated neural correlates of quantity processing of quantifiers, numbers, and numerosities. *Human Brain Mapping*, *35*(2), 444–454. Retrieved 2024-06-28, from https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.22190 doi: 10.1002/hbm.22190

Yi, Y., Billor, N., Liang, M., Cao, X., Ekstrom, A., & Zheng, J. (2022, July). Classification of EEG signals: An interpretable approach using functional data analysis. *Journal of Neuroscience Methods*, *376*, 109609. Retrieved 2024-06-13, from https://linkinghub.elsevier.com/retrieve/pii/S0165027022001364 doi: 10.1016/j.jneumeth.2022.109609

Zhang, T., Chen, W., & Li, M. (2018, January). Generalized Stockwell transform and SVD-based epileptic seizure detection in EEG using random forest. *Biocybernetics and Biomedical Engineering*, *38*(3), 519–534. Retrieved 2024-06-30, from https://www.sciencedirect.com/science/article/pii/S0208521617302577 doi: 10.1016/j.bbe.2018.03.007