LONGITUDINAL DATA ANALYSIS WITH STATISTICAL AND MACHINE LEARNING METHODS IN NEUROSCIENCE

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SERENAY ÇAKAR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
STATISTICS

AUGUST 2022

Approval of the thesis:

LONGITUDINAL DATA ANALYSIS WITH STATISTICAL AND MACHINE LEARNING METHODS IN NEUROSCIENCE

submitted by **SERENAY ÇAKAR** in partial fulfillment of the requirements for the degree of **Master of Science in Statistics Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Özlem İlk Dağ
Head of Department, **Statistics** _____

Assist. Prof. Dr. Fulya Gökalp Yavuz
Supervisor, **Statistics, METU** _____

Examining Committee Members:

Prof. Dr. Olçay Arslan
Statistics, Ankara University _____

Assist. Prof. Dr. Fulya Gökalp Yavuz
Statistics, METU _____

Prof. Dr. Ceylan Talu Yozgatlıgil
Statistics, METU _____

Date: 17.08.2022

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Serenay Çakar

Signature :

ABSTRACT

LONGITUDINAL DATA ANALYSIS WITH STATISTICAL AND MACHINE LEARNING METHODS IN NEUROSCIENCE

Çakar, Serenay
M.S., Department of Statistics
Supervisor: Assist. Prof. Dr. Fulya Gökalp Yavuz

August 2022, 103 pages

Exploration of brain activity under different conditions has been subject to many neuroscience studies. The recent developments in cognitive studies provide the opportunity to work on neural correlates of specific cognitive processes such as working memory, decision making, response inhibition, perception, and sensation. Brain response studies constitute multidimensional, multilevel or nested data sets formed by different parts of the brain of individuals. Hence, it is of significant importance to implement data analysis methods appropriate for the longitudinal structures. However, previous studies on brain response utilized methods that do not consider the dependency, multilevel and nested structure. In this thesis, we propose to apply different statistical and machine learning methods on cognitive data to fill the aforementioned deficiencies. We analyze open-access data, including optical density measures collected from 36 locations of the brain within 26 subjects through functional near-infrared spectroscopy (fNIRS). fNIRS signals are used to measure relative changes in oxyhemoglobin and deoxyhemoglobin concentrations. The nested structure of the data, which is having observations from different brain regions within subjects, is also considered. The content of this thesis provides a comprehensive implemen-

tation and comparison of several statistical and machine learning algorithms which are Linear Mixed Model (LMM) and its robustified version, Generalized LMM Tree (GLMM tree), Random Effects Expectation-Maximization Tree (RE-EM tree), Unbiased RE-EM tree, Longitudinal Classification and Regression Tree, and Gaussian Process Boosting. According to one of our findings, the GLMM tree with nested structure shows the best predictive performance as it provides the lowest model performance metrics. However, there is a trade-off between accuracy and speed since the speed of this algorithm is lower compared to other methods except robustified LMM.

Keywords: Cognitive Studies, fNIRS, N-back Data, Linear Mixed Model, Robust Modeling, Machine Learning Algorithms

ÖZ

SİNİRBİLİMDE İSTATİSTİKSEL VE MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE BOYLAMSAL VERİ ANALİZİ

Çakar, Serenay
Yüksek Lisans, İstatistik Bölümü
Tez Yöneticisi: Dr. Öğr. Üyesi. Fulya Gökalp Yavuz

Ağustos 2022 , 103 sayfa

Farklı koşullar altında beyin aktivitesinin araştırılması birçok sinirbilim çalışmasına konu olmuştur. Bilişsel araştırmalardaki son gelişmeler, işleyen bellek, karar verme, tepki ketleme, algılama ve duyum gibi belirli bilişsel süreçlerin sinirsel bağlantıları üzerinde çalışma fırsatı sağlamaktadır. Beyin tepki çalışmaları, bireylerin beyninin farklı bölümlerinin oluşturduğu çok boyutlu, çok düzeyli veya iç içe geçmiş veri kümelerini oluşturur. Bu nedenle, boylamsal yapılara uygun veri analiz yöntemlerinin uygulanması büyük önem taşımaktadır. Bununla birlikte, beyin tepkisi üzerine yapılan önceki çalışmalarda, verilerin bağımlılığını, çok düzeyli ve iç içe yapısını dikkate almayan yöntemler kullanılmıştır. Bu tezde, yukarıda bahsedilen eksiklikleri gidermek için bilişsel veriler üzerinde farklı istatistiksel ve makine öğrenmesi yöntemlerinin uygulanmasını öneriyoruz. Fonksiyonel yakın-kızılötesi spektroskopisi (fNIRS) aracılığıyla 26 deneğin 36 farklı beyin bölgesinden toplanan optik yoğunluk ölçümlerini içeren açık erişime sahip veri setini analiz ediyoruz. fNIRS sinyalleri, oksihemoglobin ve deoksihemoglobin konsantrasyonlarındaki nispi değişiklikleri ölçmek için kullanılır. Deneklerin farklı beyin bölgelerinden gözlemlerin yer aldığı iç içe veri

yapısı da dikkate alınmıştır. Bu tezin içeriği, Doğrusal Karma Model (LMM) ve bu modelin sağlamlaştırılmış versiyonu, GLMM ağacı, RE-EM ağacı, Yansız RE-EM ağacı, LongCART ve GPBoost gibi bilişsel veri yapısına uygun çeşitli algoritmaların kapsamlı bir uygulamasını ve karşılaştırmasını sağlar. Bulgularımızdan birine göre, iç içe yapıya sahip GLMM ağacı, en düşük model performans metriklerini sağladığı için en iyi tahmin performansını gösterir. Bununla birlikte, bu algoritmanın hızı, sağlamlaştırılmış LMM dışındaki diğer yöntemlere kıyasla daha düşük olduğundan, doğruluk ve hız arasında bir ödünleşim vardır.

Anahtar Kelimeler: Bilişsel Çalışmalar, fNIRS, N-geri Verisi, Doğrusal Karma Model, Sağlamlaştırılmış Modelleme, Makine Öğrenmesi Algoritmaları

To the memory of my best friend

ACKNOWLEDGMENTS

First and foremost, I must firstly thank my advisor, Assist. Prof. Fulya Gökalp Yavuz really does deserve the greatest of thanks since she has provided me with incredible support and invaluable advice in both the writing of this thesis and the work which preceded it. I feel very fortunate to have had the chance to work with her. My thesis process became an enjoyable and informative journey, even with all the difficulties, thanks to her smiling face and unconditional support.

I would like to express my gratitude to my examining committee members, Prof. Dr. Olçay Arslan and Prof. Dr. Ceylan Talu Yozgatlıgil, for their detailed reviews and valuable suggestions.

I would like to extend my heartfelt thanks to my colleagues Burcu Koca, İrem Tanrıverdi, Onur Çamlı, Ozancan Özdemir, Petek Aydemir, Rana Arslan and Sevilay Doğan.

Finally, I would like to express my deepest gratitude to my mom and grandma for being in my life and supporting me whenever I need.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

EEG	Electroencephalogram
PET	Positron Emission Tomography
NIR	Near-Infrared
NIRS	Near-Infrared Spectroscopy
fNIRS	Functional Near-Infrared Spectroscopy
fMRI	Functional Magnetic Resonance Imaging
HbO	Oxyhemoglobin
HbR	Deoxyhemoglobin
HbT	Total Hemoglobin
PFC	Prefrontal Cortical
MBLL	Modified Beer-Lambert Law
LMM	Linear Mixed Effects Model
GLMM tree	Generalized Linear Mixed-Effects Model Tree
RE-EM tree	Random Effects Expectation-Maximization Tree
CART	Classification and Regression Tree
LongCART	Longitudinal Classification and Regression Tree
GPBoost	Gaussian Process Boosting
EM	Expectation-Maximization
ERP	Event-Related Potential
PSMC	Primary Sensorimotor Cortex
SPM	Statistical Parametric Mapping
ANOVA	Analysis of Variance
WMS-R	Wechsler Memory Scale-Revise
WMS-R	Lower Limb Amputation

WTC	Wavelet Transform Coherence
dlPFC	Left/Right Dorsolateral Prefrontal Cortex
TPJ	Left/Right Temporoparietal Junction
ML	Machine Learning
MMSE	Mini-Mental Status Examination
kNN	K-Nearest-Neighbors
SVM	Support Vector Machine
RT	Reaction Time
HC	Healthy Control
RF	Random Forest
ET	Extra Tree
GB	Gradient Boosting
DPF	Differential Path Length Factor
OD	Optical Density
MLE	Maximum Likelihood Estimation
REML	Restricted Maximum Likelihood Estimation
LRT	Likelihood Ratio Test
MOB	Model-Based Recursive Partitioning
GLM	Generalized Linear Model
MSE	Mean Squared Error
RMSE	Root-Mean-Squared Error
MAE	Mean Absolute Error
DSR	Discrimination/Selection Response
WG	Word Generation
VIF	Variance Inflation Factor
AIC	Akaike Information Criterion
GD	Gradient Descent

WLS	Weighted Least Squares
NM	Nelder Mead
BFGS	Broyden–Fletcher–Goldfarb–Shanno
MERT	Mixed-Effect Regression Tree
MERF	Mixed-Effect Random Forest
REEMforest	Random Effect Expectation Maximization Forest

CHAPTER 1

INTRODUCTION

There are approximately 100 billion neurons in the human brain [1] and every neuron should pass through almost 10^{15} neurons to communicate with another neuron [2]. As a non-stop process, the brain functions compose an extensive data system. Considering the brain's complexity, the data from neuroscience studies require advanced analysis to extract features representing brain functionality. Recent developments on brain-computer interface techniques such as Electroencephalogram (EEG), Positron Emission Tomography (PET), Functional Near-Infrared Spectroscopy (fNIRS) and Functional Magnetic Resonance Imaging (fMRI) have opened up the possibility for observing and examining changes in subjects' brain activity. Many cognitive studies include the behavioral and cognitive data collected via these technologies, and focus on finding out connections between these measures for individuals or group of subjects. There exist many experimental studies which focus on explaining how the brain functions and what kind of factors are related to specific changes in the brain [3, 4, 5, 6, 7, 8, 9, 10, 11]. Many experimental data about the brain include several measurements taken over time from the same experimental units. So, longitudinal data structure is an integral part of many neuroscience studies. This type of data is commonly used in studies across a range of disciplines from the cognitive to the biomedical sciences. Hence, the number of algorithms to analyze such data continue to grow and they become more widespread compared to the past.

In the scope of this thesis study, we conduct our analyses using data from n-back task paradigm which is one of the most commonly used functional neuroimaging study of working memory [12, 13, 14, 15]. Also, we worked on fNIRS data that is one of the non-invasive and functional neuroimaging technique used to observe in-

dividuals' brain activation. With the help of fNIRS, optical density measures can be collected through two distinct wavelengths and it is possible to estimate alterations in the concentration of oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) by the use of Modified Beer-Lambert Law (MBLL) [16]. Although HbO and HbR concentrations are obtained for the subjects, the statistical analyses are conducted to assess the alterations in the HbO concentrations as this indicator is indicated to be the most reliable variable for examining the changes in regional cerebral blood flow [17]. We then combine the data for HbO measures of each subject and experiment related variables for analyses.

In this study, taking into account the dependency structure that is mainly ignored in the analyzes of neuroscience data [18, 19, 20, 21] and the model flexibility brought by using random parameters, we conduct our analyses through the approaches designed for longitudinal data. Unlike many studies including longitudinal data structure, our data have no missing cases. Also, multiple variables are included in the implemented methods simultaneously to reduce the modeling error arising from conducting univariate analysis several times in some statistical methods for cross-sectional data such as t-tests and Z-scores used in neuroscience studies dominantly [22, 23, 24, 25]. Initially, Linear Mixed Model (LMM) approach is used to develop a model for explaining the variation in HbO concentration changes. Additionally, we deal with the complexity of the neural data sets, including many levels in which some of their categorical responses are connected with the structure of brain imaging technologies. As each subjects may behave differently during the experiment, the random intercept term is added to the model so that alterations in HbO of each subject can differ from each other. On the other hand, we include the indices variable having 36 distinct categories nested within each subject with several covariance structures. Also, we implement a robust LMM approach into neuroscience data to reduce the effects of unusual observations in the data set.

In addition to LMM and its robustified version, we implement several ML algorithms for predicting mean HbO response on this data. The algorithms used in this study are Generalized Linear Mixed-Effects Model Tree (GLMM tree), Random Effects Expectation-Maximization Tree (RE-EM tree), unbiased RE-EM tree, Longitudinal Classification and Regression Tree (LongCART), and Gaussian Process Boosting

(GPBoost). As the last step, we provide comparisons of all the implemented algorithms. As one of our purpose is to compare their fitting performances, we evaluate them through three different statistical measures which are Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) once the algorithms are implemented on the data considering with and without nested structure. According to our findings, the model fitting performance of GLMM tree with nested structure performs best considering all these three metrics.

The rest of the study is organized as follows. The following chapter includes literature review on fNIRS and the studies that utilize statistical and machine learning methods to analyze data from cognitive science. Then, third chapter has three subsections. In the first subsection of this chapter, details on fNIRS modality are explained. The second subsection includes information on LMM and its robustified version while the last subsection describes hybrid methods for longitudinal data analysis utilized in this thesis study. The first part of the fourth chapter explains experimental data corresponding to n-back task, and implementation of different algorithms are included in the second part of this chapter. The results derived from the methods and comparison of algorithms are explained at the end of the same chapter. In the last chapter, the results of the methods are discussed and important findings are emphasized. Further, selection of the algorithm for such a data setting is considered with different circumstances. The study is finalized by explaining limitations of this study and our future plan for the analysis of the same data set. The table including accuracy measures of all subjects are presented in the Appendix.

CHAPTER 2

LITERATURE REVIEW ON COGNITIVE STUDIES

This literature review part is composed of two different sections due to the interdisciplinary characteristic of the thesis. The first part comprises the explanations for the historical perspective of fNIRS and the brief information on how fNIRS system acquire data from people's scalp. This part also covers literature for the some of the advantages and disadvantages of this technology. The second part dwells on the methods for data-driven analyses techniques as well as the issues and the limitations that they can suffer from. The methods mentioned in the second section are also explained in an order from simple to complex.

2.1 Historical Review on fNIRS

Neurons need nutrients such as glucose and oxygen to create cerebral activity so that they can generate energy and action potential during the execution of a particular task. Numerous experimental studies have been conducted to observe how people react to specific stimuli and what kind of potential sources of the activations can occur based on the characteristics of a certain task. Even an individual is on the rest, the brain does not stop, and the activation continues in the human brain. The neuroimaging modalities have been employed to assess the changes in brain activity. fNIRS is highlighted as one of the approaches that can reveal the hemoglobin reactions of participants using optical probes placed on the subjects' foreheads.

The possibility of measuring the changes of oxygenation in an adult's cortical area during hyperventilation with the help of Near-Infrared Spectroscopy (NIRS) was first demonstrated by Frans Jöbsis in 1977 [26]. After 1992, scientists started to explore

the human cerebral cortex by examining the alterations in oxygenation and hemodynamic functions of people's brain activity with the invention of fNIRS. Hence, the discovery of fNIRS added another dimension to the studies on the brain mapping research [27].

In light of the absorption spectra differences between HbO and HbR, the concentration changes of these two measures can be examined from diffusely scattered light sources [28, 26, 18]. Fantini and Sassaroli [29] stated that the estimated measurements for HbO and HbR could be obtained when MBLL is applied to the data obtained from the light within the near-infrared range of the electromagnetic spectrum, and can be absorbed by hemoglobin.

The history of the first single-site studies in which people's brain activities are examined through fNIRS is based on 1991. The publications of these studies were realized by four research groups belonging to distinct institutes such as Hokkaido University, Tokyo National Institute of Neuroscience, University of Munich, and the University of Pennsylvania. In 1995, the data collection system was improved as the scientists were able to use a multi-channel system with ten channels (indices) at the beginning instead of a single channel system in which there exists a poor sensitivity with a low temporal resolution. The 10-channel system is first introduced by Hitachi in 1994, and then multichannel systems with more than 10 channels have been developed after 1998 [27]. The development process of fNIRS modality from single channel to multi-channel system is illustrated in Figure 2.1 which was taken from [27].

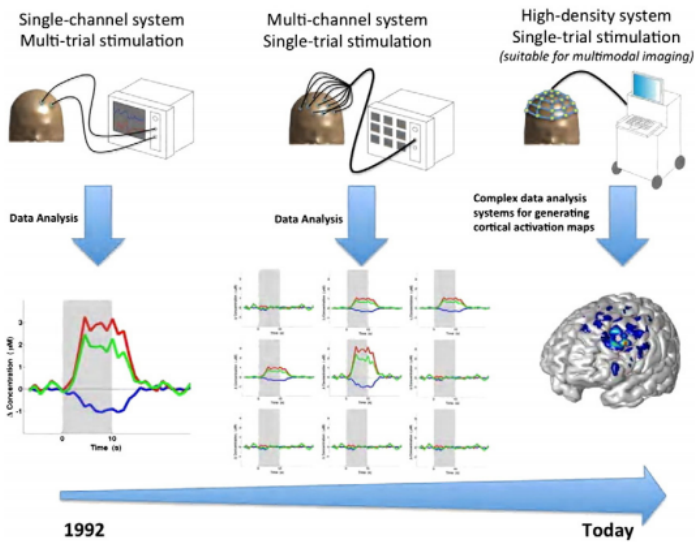


Figure 2.1: The development of fNIRS from 1992

As a complex data collection procedure, fNIRS has several advantages and limitations. [30, 31] can be examined to check out the advancements and pitfalls of fNIRS modality in a detailed version .

While analyzing the data, knowing the pros and cons would be necessary. First of all, fNIRS's spatial resolution is better than that of event-related potential (ERP) and EEG. At the same time, it has inferior spatial resolution when compared to fMRI [32]. Also, the measurements obtained from brain regions through fNIRS technology are shallower when compared to the results of fMRI and PET [33]. However, fNIRS has a better temporal resolution compared to fMRI. Specifically, fNIRS enable scientists to obtain brain signals with a temporal resolution of 0.01 second while this duration takes more time with fMRI. While it takes about 1-2 seconds to occur hemodynamic response to neural activity, this characteristic of fNIRS can provide direct measures of neural activity as the better resolution is essential to distinguish between signal contamination resulting from the physiological changes and motion artifacts. On the other hand, the spatial resolution of fNIRS is between 2-3 cm, and the signal obtained through the specific channel samples a 3D volume which is not known precisely. Hence, the signals can be recorded after fNIRS projects that volume onto a midchannel located on the scalp [34]. Additionally, it is not possible to obtain co-registration with anatomical records of the subject's brain through fNIRS.

The scientists should try to collect data from a separate structural MRI and align the location of the optodes with external skull landmarks or averages. Another alternative solution for that problem is that it is possible to use head size-appropriate subjects to provide co-registration using averaged skull landmarks [35]. Another advantage of this technology is that fNIRS is a noninvasive system. Thus, there is no danger in using this system repeatedly on individuals [32]. It is recorded as an indispensable feature for situations requiring repeated experiments.

2.2 Review on Cognitive Studies

This section provides illustrative examples for the statistical and machine learning analysis methods on different studies to explore brain activity in the literature.

In a hand grasping experiment including six healthy adults and six patients with brain ischemia, the concentration changes of HbO, HbR, and total hemoglobin (HbT) in the primary sensorimotor cortex (PSMC) contralateral are collected as a task-related measure [36]. In this study, they utilize both NIRS and BOLD- fMRI signals. The alterations in HbO, HbR, and HbT concentrations using NIRS in PSMC are measured. The evoked cerebral blood oxygenation (CBO) changes measured through NIRS are compared with the activation maps obtained through BOLD-fMRI. The Statistical Parametric Mapping (SPM) is used to get activation maps with a Z-score higher than 1.5. One of their findings indicates an increase in the PSMC on the lesion side while the patients performed the task. On the other hand, they also stated that this change is associated with increases in HbO and HbT concentrations for patients with brain ischemia, whereas a decrease is concluded for controls.

The t-test as a formal way of identifying any existing mean difference among two conditions is also widely used in neuroscience research. For instance, in an n-back study conducted by Herff et al. [37], the participants are expected to remember the last one, two, or three of rapidly flashing items. This study uses fNIRS to sample workload activity in the prefrontal cortex. T-tests with Bonferroni corrections to adjust type I error for each pairwise comparison are used to identify the percentage of missed targets differences among n-back conditions. They have concluded a signifi-

cant difference in difficulty levels of all the three n-back conditions.

Another most widely used statistical method among neuroscientists is the analysis of variance (ANOVA). In search of Google Scholar, which relates the terms "neuroscience" and "ANOVA" the number of hits is seen as 28,600, representing almost 36% of the search with keyword neuroscience in 2015 [38]. ANOVA can examine the difference in the dependent variable when there exist more than two factors or explanatory variables. However, as the ANOVA approach is a parametric method and uses the F-distribution, the reliability of the results of this procedure depends on a couple of assumptions on the data. Mainly, this procedure assumes that the observations included in data should be independent of one another, which can be potentially violated in neuroscience studies as the measures are collected from the same subjects through time and under different conditions. For instance, it is stated in [39] that among 314 papers examined from five most popular journals, 53% of these included this type of data where a single subject is reviewed. One study, including eight victims of the Tokyo Subway Sarin attack and the 26 people who are not, Watanabe et al. [40] use the fNIRS system with 24 channels. They conducted two-way ANOVA with two factors that are diagnosis and categories during the statistical analysis of the experimental design in which the subjects underwent a memory task using the Wechsler Memory Scale-Revised (WMS-R). Hence, the application of the ANOVA method on the data collected from an experimental design within the scope of neuroscience should be carefully examined due to the violation of independence assumption. The same attention must be taken into account for the t-test and Z-score.

As an extension of ANOVA, repeated measures ANOVA is an alternative method for analyzing the neural data. Although this statistical procedure can take into account both the between and within subjects' factors, the limitation of this method reveals from the fact that the independence can be accounted only when each observation in data is obtained in a different condition from the previously defined fixed conditions, not at distinct circumstances for the same factor [38]. When the observations are collected from the same subject, they tend to be more similar than those taken from the distinct cases. This situation can yield the nested designs in which the observations within each cluster cannot be thought of as being independent [39]. As the previously mentioned methods such as t-test and ANOVA assume the independent observations,

this potential failure for the neural data can threaten the validity of the results obtained from these procedures.

LMM can simultaneously consider the crossed structure of the data and its nested structure. Thereby, this modeling approach can control the type I error, which is not the case for the approaches covered up to now [41]. The difference of the LMM from the classical linear regression approach is that it includes both fixed and random effects. When the subjects and the different conditions are introduced as the random effects into the LMM, the model can allow us to generalize the results to the population [42].

In a study conducted to assess walking performance and prefrontal cortical activity of 39 people with lower limb amputation (LLA) and 33 persons without LLA, LMM is utilized. Prefrontal cortical (PFC) activity is measured using fNIRS in the experiment, and the walking performance is evaluated with a walk-test under three distinct conditions. The results of the LMM provided to detect how these measures are changing among different groups and according to other walking conditions [43].

In a prospective study including 20 healthy adults, the participants are expected to walk in two conditions. The study assesses whether complex walking situations are associated with increased PFC activity. The group hemodynamic response is calculated after neural data are collected from the subjects through two wireless near-infrared spectroscopy. They implement LMM, where 20 subjects are considered as a random effect. Their findings suggest that walking with unstable shoes engages with increasing PFC. Therefore, the walking activity, which is unstable and unpredictable, can increase PFC activity and they conclude that PFC can be associated with walking stability [44].

Another study investigates how PFC activation changes among depressed subjects and healthy controls (HC) when emotional and neutral facial expressions are provided. Manelis et al. [45] conduct two experiments to collect data on both behavioral and neural data from different participants. The first experiment aims to check the hypothesis that people with depression give more inaccurate responses than HC to recognize neutral facial expressions. In the behavioral study, there are 53 subjects, of which 33 depressed subjects suffer from either major depressive or bipolar disorders

and 20 HC. Response accuracy and reaction time (RT) measures are first calculated for each subject to analyze behavioral data in this experiment. To estimate group-level effects, LMM considering depressed and HC groups as independent variables and subjects as a random effect are conducted in `R` with `lmerTest` package using Satterthwaite's degrees of freedom approach to estimate denominator degrees of freedom required for F or t-statistic. On the other hand, in experiment 2, the researchers gathered neural data from 35 distinct individuals (19 diagnosed with either major depressive or bipolar disorders, 16 HC) through fNIRS to examine their PFC activation while they performed the same task. LMM is implemented on the neuroimaging data, and depressed and HC groups and three emotions such as happy, neutral and fearful are included in the set of explanatory variables. Also, they introduce subjects as random effect and interaction between two on PFC. The first experiment results reveal that depressed people diagnosed with either major depressive or bipolar disorders are slower in the task and respond less accurately in judging neutral facial expressions. On the other hand, they conclude that lower accuracy regarding neutral emotional expressions is associated with lower activation in the right PFC for depressed subjects, but this is not the case for HC. Also, depressed participants tend to provide lower right PFC activation in judging happy facial expressions than HC.

In the study of exploring naturalistic interactions in parent-child dyads, Nguyen et al. [46] analyzed fNIRS hyperscanning data. They constructed a generalized linear mixed-effects model (GLMM) including Wavelet Transform Coherence (WTC), which is a measure for interpersonal neural synchrony, as a response variable, while fixed effects entered in the model are conditions with levels of cooperation and individual regions of interest with four levels which are left/right dorsolateral prefrontal cortex (dlPFC), left/right temporoparietal junction (TPJ), and pairing indicating true or random. Their proposed GLMM was composed of both main effect terms of these fixed effects and interaction among these variables. The contribution of an interaction effect among condition and pairing variables and region of interest as a fixed effect term to the model was significant. In terms of random effects, their model exhibited a random slope for condition variables as well as a random intercept for dyads. In their study of the Leiden 85-plus, Spagnoli et al. [47] utilized a mixed-effect logit model on Mini-Mental Status Examination (MMSE) index to assess the cognitive function-

ing of adults. One of their findings indicates that subjects with higher education levels have a lower probability of cognitive impairment, while cognitive functioning tends to decrease when subjects are getting older.

In addition to statistical methods, machine learning (ML) algorithms are widely used for classification and regression problems. However, we have recognized a few studies on implementing ML techniques using neural data in the literature mainly concentrating on classification problems. In this part, we review studies on both EEG and fNIRS since similar ML methods are implemented to analyze these two neuroscience modalities.

There are few studies in the literature, including the analysis of fNIRS data with ML algorithms ([48, 49, 50, 51]). For instance, Sitaram et al. [48] used Support Vector Machine (SVM) to differentiate left-hand imagery and right-hand imagery on the located NIRS signals. According to their finding, the average accuracy of the SVM classifier was 73% for all volunteers. Additionally, Girouard et al. [50] implemented k-nearest-neighbors (kNN) with $k = 3$ to classify two difficulty levels of the Pacman game and the resting state of participants on fNIRS data. The average classification accuracy of kNN for comparing these three conditions is 76.7%. In their experimental study on the estimation in a flight simulator using fNIRS, Gateau et al. [51] utilized an SVM-based classifier to differentiate between task difficulty levels which are low working memory load and high working memory load. According to their estimator tested on 19 pilots, they reached a classification accuracy of 80%.

Additional to the studies including fNIRS, we also review EEG data analysis cases in which ML algorithms are used. For instance, Plotnikov et al. [52] used SVM to analyze the boredom/flow conditions of subjects while they were playing Tetris games with different levels using EEG data. Considering the average accuracy for all participants, their 4-class SVM provides an accuracy of 57%. Papakostas et al. [53] investigated data on EEG signals to predict the outcome of a sequence learning task using SVM, Random Forest (RF), Extra Tree (ET), and Gradient Boosting (GB) algorithms. They obtain the highest average accuracy of 74% from the GB classifier.

Although the previously mentioned methods' predictive capabilities are sufficient to capture data information, their results are questionable due to the assumptions re-

quired to implement these algorithms. The methods used in the mentioned studies depend on the independency among measurement points. However, this assumption may not hold for the repeated measurements. Violation of dependency assumption may cause undetectable bias in predictions. To overcome the assumption limitations of these studies, we suggest methods that consider the dependency among repeated measures in neuroscience data analyses. In this study, we implement several methods, including statistical techniques, ML algorithms, and their hybrid forms, appropriate for analyzing longitudinal data structure. This is the first attempt in the literature that ML methods considering longitudinal data structure have been used to analyze data from cognitive studies to the best of our knowledge.

CHAPTER 3

METHODOLOGY

In the literature, many experimental cognitive studies aim to advance our understanding how nervous system functions and processes information using brain signals collected from different subjects during the specific task. To this end, many researchers adopt statistical and ML techniques to extract information from neural data.

This chapter mainly consists of four parts due to interdisciplinary aspect of this thesis. In the first part, we give an introduction to working principle of fNIRS modality and information on MBLL which is used to convert optical density measures into HbO and HbR concentration changes. In the second part, statistical and ML algorithms implemented to cognitive data are explained. Firstly, the concepts involved in LMM are explained in detail. In addition to theoretical framework of LMM, EM algorithm, which is one of the most commonly used method to find estimates of the model parameters in LMM, is explained. As the statistical modeling approaches require eliminating variables which do not have significant contribution to the model for explaining the variation in response, the hypothesis testing procedures for both fixed and random effect parameters in LMM are also explained, separately. Then, hybrid methods used in this study are explained. At the end of this section, the statistical measures used to compare different models are explained.

3.1 Functional Near-Infrared Spectroscopy (fNIRS)

fNIRS is a non-invasive light-based neuroimaging technique used to monitor changes in oxygenation status of tissue. fNIRS utilizes near-infrared (NIR) light to monitor functional neuroimaging. Functional neuroimaging indicates the use of neuroimaging

modalities to investigate brain functionality, often with a view to finding out the correlates of a certain brain region and cognitive processes. It is the use of wide variety of research areas such as neuropsychology and cognitive neuroscience. NIR light allows direct or indirect measurement of brain activity by evaluating changes in blood flow. From this information, it is possible to obtain HbO and HbR concentration changes with the help of this technology.

When brain areas become active, they need more oxygen and food. A specific protein called hemoglobin is responsible for carrying oxygen to the brain areas. To deliver oxygen to the brain areas quickly, blood vessels become wider and the blood flow increases. Hence, the amount of oxygenated hemoglobin arriving to the active brain region rises up. Further, when brain uses oxygen, the hemoglobin molecules change form from oxygenated to deoxygenated. Villringer and Chance [54] state that functional state of the brain can affect its optical properties. This concentration change can be measured by fNIRS as HbO molecules can absorb or swallows some of the light.

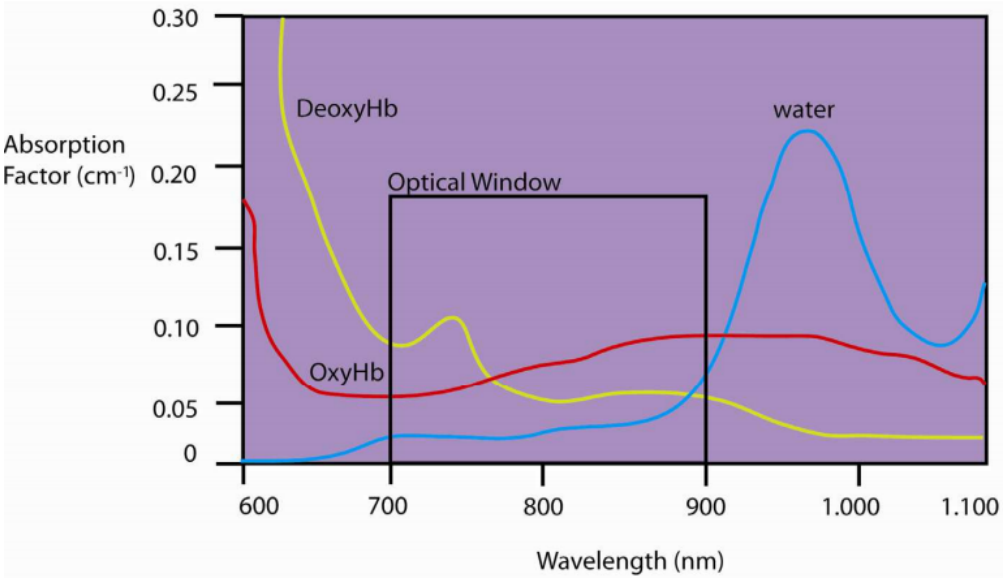


Figure 3.1: Absorption spectrum in optical window

The NIR light is the light with a wavelength between 700 and 900 nanometers (nm), and HbO and HbR molecules absorb light in an infrared spectrum. The absorption spectrum in NIR window can be examined in Figure 3.1 retrieved from [55]. If the used wavelengths were lower than 700 nm then the main absorber would be biolog-

ical tissue. If wavelengths higher than 900 nm were used, water would be the main absorber. Hence, devices are arranged according to this situation, and many of them emit light into the tissue with two wavelengths which are 760 nm and 850 nm as they give the best resolution that make it possible to distinguish the concentration changes in the oxygen-related chromophores, which are HbO and HbR, separately [56].

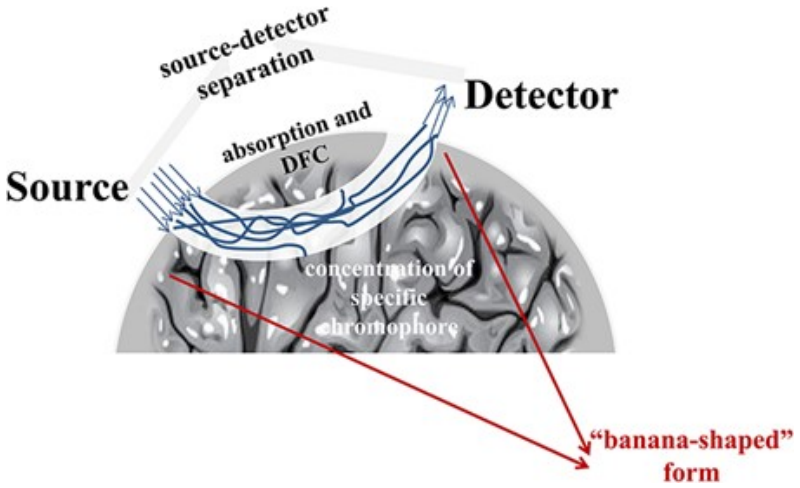


Figure 3.2: Banana shaped photon path of fNIRS

fNIRS modality has sources and detectors, and these two components are called optodes as shown in Figure 3.2 taken from [57]. The source can transmit light into the skin while detector receives the light back from the skin.

Two types of interactions are possible when photons enter the brain tissue: scattering and absorption [58]. Scattering can mainly cause deviation in the straight trajectory of the light beam while absorption indicates that photons to dissipate their energy into the medium. As can be seen from Figure 3.2, when photons are sent to the brain, they follow a banana shaped pathway along the path from a source to a detector if they are not absorbed. Photons are mainly absorbed by tissue due to the existence of chromophores such as HbO, HbR, and water. Scattering is assumed to be constant while absorption can be different depending on concentration changes of chromophores. The absorbance of these constituents is small within the optical window, and this allows photons to penetrate the tissue (see Figure 3.1). Considering the range of optical window, main absorbers are HbO and HbR as compared to other tissue chromophores such as water. Also, the differences in absorption conditions can be calculated from the variations in HbO and HbR concentrations in the brain due to the fact that these

two constituents have significantly different absorption spectra. This result indicates a relation between cerebral hemodynamic changes and brain functionality through the term "neurovascular coupling" [59]. Fortunately, functional brain imaging is possible from this relation and the ability of fNIRS to measure concentration changes in HbO and HbR.

In the following part, we explain MBLL which is used to obtain concentration changes in HbO and HbR from optical density measures obtained by fNIRS modality.

3.1.1 Modified Beer-Lambert Law (MBLL)

The fNIRS data including raw intensity measures are required to be converted into concentration changes of tissue chromophores, mainly HbO and HbR. MBLL is used to distract changes in HbO and HbR obtained from each optode of the fNIRS technology. In this section, the formula for Beer-Lambert Law (BLL) and its modified version included in the conversion of raw fNIRS measures are specified [60].

According to BLL, the following equation can be calculated.

$$\log_{10}\left(\frac{I_{inc}}{I_{det}}\right) = \varepsilon.C.l, \tag{3.1}$$

where I_{inc} shows the incident light intensity, I_{det} is for detected light intensity, ε indicates molar absorption coefficient and C is the concentration of substance in media while l represents the path length. However, as brain tissue does not transmit the light perfectly, this equation should be modified so that scattering can be taken into account. This can yield us to consider MBLL [61].

$$\mathrm{A} = \log_{10}\left(\frac{\mathrm{I_{inc}}}{\mathrm{I_{det}}}\right) = \varepsilon.C.DPF + \mathrm{G}, \tag{3.2}$$

where A is attenuation. The parameters such as G and differential path length factor (DPF) are introduced in MBLL formula. G is the geometry dependent factor and it represents the intensity lost resulted from scattering. The DPF is mainly the ratio of averaged path length of light to the distance between source and detector. As to

convert attenuation data into concentration changes, the distance travelled by photons has to be considered. To this end, DPF is used to realize this conversion.

The optical density (OD) for a given wavelength, denoted by (OD_λ), can be calculated from the following equation

$$OD_\lambda = \log\left(\frac{I_{\text{in}}}{I_{\text{out}}}\right) \approx \varepsilon_\lambda \cdot \text{c.d.DPF} + \text{G}. \tag{3.3}$$

The changes in optical density, represented by ΔOD, when having the same I_{in} in two distinct conditions can be written as

$$\Delta OD_\lambda = \log\left(\frac{I_{\text{rest}}}{I_{\text{test}}}\right) = \varepsilon_\lambda^{\text{HbR}} \cdot \Delta c^{\text{HbR}} \cdot \text{d.DPF} + \varepsilon_\lambda^{\text{HbO}} \cdot \Delta \text{c}^{\text{HbO}} \cdot \text{d.DPF}. \tag{3.4}$$

If the fNIRS measures are obtained in two distinct wavelengths denoted by λ_1 and λ_2, the formula becomes

$$\begin{bmatrix} OD_{\lambda 1} \\ OD_{\lambda 2} \end{bmatrix} = \begin{bmatrix} \varepsilon_{\lambda 1}^{\text{HbR}} \cdot \text{d.DPF} & \varepsilon_{\lambda 1}^{\text{HbO}} \cdot \text{d} \cdot \text{DPF} \\ \varepsilon_{\lambda 2}^{\text{HbR}} \cdot \text{d.DPF} & \varepsilon_{\lambda 2}^{\text{HbO}} \cdot \text{d.DPF} \end{bmatrix} \begin{bmatrix} \Delta \text{c}^{\text{HbR}} \\ \Delta \text{c}^{\text{HbO}} \end{bmatrix}. \tag{3.5}$$

The concentration changes for HbR and HbO denoted by $\Delta \text{c}^{\text{HbR}}$ and $\Delta \text{c}^{\text{HbO}}$, respectively, can be found from the Equation (3.6) if the 2x2 matrix is non-singular.

$$\begin{bmatrix} \Delta \text{c}^{\text{HbR}} \\ \Delta \text{c}^{\text{HbO0}} \end{bmatrix} = \begin{bmatrix} \varepsilon_{\lambda 1}^{\text{HbR}} \cdot \text{d.DPF} & \varepsilon_{\lambda 1}^{\text{HbO}} \cdot \text{d} \cdot \text{DPF} \\ \varepsilon_{\lambda 2}^{\text{HbR}} \cdot \text{d.DPF} & \varepsilon_{\lambda 2}^{\text{HbO}} \cdot \text{d.DPF} \end{bmatrix}^{-1} \begin{bmatrix} OD_{\lambda 1} \\ OD_{\lambda 2} \end{bmatrix}. \tag{3.6}$$

3.2 Linear Mixed Model (LMM)

LMM is an advanced regression model used for repeated measurements, including neuroscience experiments. Unlike classical linear regression models that include only fixed effects and the assumption of independent or uncorrelated errors, LMM can be constructed by incorporating random and fixed effects for correlated and dependent data. LMM does not have the disadvantages of ANOVA, which is often used in

neuroscience studies (see [40]). More clearly, nested/hierarchical structures can be included in LMM and the results for the subject or conditions can be generalized to the population with the help of random terms.

LMM relates the response variable to the random and fixed effects and the general matrix form of this model for each unit, i, can be expressed in Equation (3.7) [62].

$$\begin{aligned} \boldsymbol{Y}_i &= \underbrace{\boldsymbol{X}_i\boldsymbol{\beta}}_{\text{fixed}} + \underbrace{\mathbf{Z}_i\boldsymbol{u}_i + \boldsymbol{\varepsilon}_i}_{\text{random}}, \\ \boldsymbol{u}_i &\sim N(\mathbf{0}, \boldsymbol{D}), \\ \boldsymbol{\varepsilon}_i &\sim N\left(\mathbf{0}, R_i\right), \end{aligned} \tag{3.7}$$

where Y_i represents the $(n_i \times 1)$ vector of observed response values for the i^{th} individual. X_i having the dimension of $(n_i \times p)$ denotes the fixed effects design matrix including the observed values for p predictors on each subject while β represents a $(p \times 1)$ column matrix of unknown population parameters which are assumed to be the same for each subject in the data. That is why they are called fixed effects. On the other hand, Zi denotes an $(n_i \times q)$ matrix of the observed values of q covariates and u_i is a vector of q unknown random effects. The within-group error term, ε_i, consists of the vector of errors for each subject and it has dimension of $(n_i \times 1)$. This model assumes that the errors and the random effects are independent of each other for a given subject.

The errors are assumed to follow a normal distribution with 0 mean and variance of R_i, that is $\varepsilon_i \sim N(0, R_i)$. R_i is a positive definite symmetric matrix and different structures can be assumed for this covariance matrix. Unlike fixed effects, random effects (u_i) which can reflect the variability among individuals for each regressor in Z_i, are assumed to come from a statistical distribution. They follow a normal distribution with mean 0 and variance-covariance matrix of D. This symmetric and positive definite D matrix has the dimension of $(q \times q)$, where q represents the total number of random effect terms included in the model. Hence, it can be written as $u_i \sim N(0, D)$.

Depending on the assumptions that the random effects u_i and random errors ε_i are

independently normally distributed and they are linear in LMM model, we assume that each response vectors, Y_i, comes from a normal distribution with mean $X_i\beta$ and covariance matrix $V_i = Z_i D Z_i' + R_i$. The marginal covariances of the Y_i vector can be represented in the off-diagonal elements of V_i matrix having dimension of $(n_i \times n_i)$. Hence, the marginal distribution of Y_i can be expressed as $Y_i \sim N(X_i\beta, V_i)$.

The maximum likelihood estimation (MLE) and restricted maximum likelihood estimation (REML) methods are used to estimate the fixed-effect parameters, β, and the parameters in the variance-covariance matrices, shown with θ [63].

As the MLEs can provide the values for the unknown parameters, which can make the observed data most likely to occur under the distributional assumptions, the likelihood function for an observed data $Y_i = y_i$ should be derived. The marginal distribution of response Y_i which has multivariate normal distribution, $f\left(Y_i \mid \beta, \theta\right)$, is given by

$$\begin{aligned} f\left(Y_i \mid \beta, \theta\right) &= (2\pi)^{-n_i/2} \det\left(V_i\right)^{-1/2} \\ &\times \exp\left(-0.5\left(Y_i - X_i\beta\right)' V_i^{-1}\left(Y_i - X_i\beta\right)\right), \end{aligned} \tag{3.8}$$

where det is used to refer to the determinant [63]. On the other hand, the vector θ includes the variance-covariances included in both D and R_i matrices.

The likelihood function for observed data $Y_i = y_i$ for the $i-th$ subject can be stated as follows:

$$\begin{aligned} L_i(\beta, \theta) &= (2\pi)^{-n_i/2} \det(V_i)^{-1/2} \\ &\times \exp(-0.5(y_i - X_i\beta)' V_i^{-1}(y_i - X_i\beta)). \end{aligned} \tag{3.9}$$

When all m individuals are considered, the corresponding likelihood function which will be the product including the contribution of each m subjects $(i = 1, 2, \ldots.., m)$ can be derived as follows:

$$\begin{aligned} L(\beta, \theta) = \Pi_i L_i(\beta, \theta) = \Pi_i[&(2\pi)^{-n_i/2} \det\left(V_i\right)^{-1/2} \\ &\times \exp\left(-0.5\left(y_i - X_i\beta\right)' V_i^{-1}\left(y_i - X_i\beta\right)\right)]. \end{aligned} \tag{3.10}$$

When the logarithm is applied, the log-likelihood function, $l(\beta, \theta)$, can be defined as

$$l(\beta,\theta) = \ln L(\beta,\theta) = -\frac{1}{2}n\ln(2\pi) - \frac{1}{2}\sum_i \ln\left(\det\left(V_i\right)\right) - \frac{1}{2}\sum_i \left(y_i - X_i\beta\right)' V_i^{-1}\left(y_i - X_i\beta\right), \tag{3.11}$$

where n represents the total number of observations in the data and the function "ln" corresponds to the natural logarithm [63].

3.2.1 Expectation Maximization (EM) Algorithm in LMM

The MLEs of LMM include other unknown parameters. This consequent is required to use an iterative algorithm to implement LMM. We use the EM algorithm, introduced by [64], which is numerically stable and guarantees to have the parameters in the desired parameter spaces.

The EM algorithm consists of two steps which are Expectation and Maximization. In the E-step, the algorithm computes the conditional expectation of the log-likelihood of the complete data given the current parameter estimates. M-step is based on maximizing the conditional expectation derived in the E-step to obtain updated parameter estimates. The algorithm starts with the initial values for the unknown parameters and the E-step and M-step iterations continue until the convergence. As a result, when the algorithm stops, the obtained values for the unknown parameters are supposed to maximize the likelihood function [65].

At this point, the $\eta = (\beta, \theta)$ is considered the matrix of all unknown parameters in the LMM. The random effects are treated as unobserved data for implementing the EM algorithm. Let k represents the iteration number, $k = 0, 1, ...$ and the function calculated in the $k-th$ iteration of E-step of EM algorithm is given by

$$\begin{aligned} Q\left(\eta \mid \eta^{(k)}\right) &= E\left(\log L(\eta \mid y, u) \mid y, \eta^{(k)}\right) \\ &= E\left(\sum_{i=1}^{n}[\log f(y_i \mid u_i, \beta, R_i) + \log f(u_i \mid D) \mid y_i, \eta^{(k)}]\right), \end{aligned} \tag{3.12}$$

where the conditional expectation is derived from the conditional distribution $f\left(u_i \mid y_i, \eta^{(k)}\right)$.

The E-step provides the computations of the following sufficient statistics for the parameters included in the matrices of variance-covariance, D and R_i in LMM.

$$\sum_{i=1}^{n} E\left(\varepsilon_i'\varepsilon_i \mid y_i, \hat{\eta}^{(k)}\right) = \sum_{i=1}^{n} \left[\hat{\varepsilon}_i^{(k)\prime}\hat{\varepsilon}_i^{(k)} + \operatorname{tr}\left(\operatorname{Cov}\left(\varepsilon_i \mid y_i, \hat{\eta}^{(k)}\right)\right)\right], \tag{3.13}$$

$$\sum_{i=1}^{n} E\left(u_i u_i' \mid y_i, \hat{\eta}^{(k)}\right) = \sum_{i=1}^{n} \left[\hat{u}_i^{(k)}\hat{u}_i^{(k)\prime} + \operatorname{tr}\left(\operatorname{Cov}\left(u_i \mid y_i, \hat{\eta}^{(k)}\right)\right)\right], \tag{3.14}$$

where

$$\begin{aligned}
&\hat{\varepsilon}_i^{(k)} = y_i - X_i\hat{\beta}^{(k)} - Z_i\hat{u}_i^{(k)},\\
&\hat{u}_i^{(k)} = D\left(\hat{\eta}^{(k)}\right) Z_i' V_i^{-1}\left(\hat{\eta}^{(k)}\right)\left(y_i - X_i\hat{\beta}^{(k)}\right),\\
&V_i\left(\hat{\eta}^{(k)}\right) = Z_i D\left(\hat{\eta}^{(k)}\right) Z_i' + R_i\left(\hat{\eta}^{(k)}\right).
\end{aligned}$$

The final forms for the parameter estimators are derived as follows.

$$\begin{aligned}
&\hat{\beta}^{(k+1)} = \left[\sum_{i=1}^{n} X_i'\hat{V}_i^{-1}\left(\hat{\eta}^{(k)}\right) X_i\right]^{-1} \sum_{i=1}^{n} X_i'\hat{V}_i^{-1}\left(\hat{\eta}^{(k)}\right) y_i,\\
&\hat{\sigma}^{(k+1)2} = \frac{\sum_{i=1}^{n} E\left(\varepsilon_i'\varepsilon_i \mid y_i, \hat{\eta}^{(k)}\right)}{\sum_{i=1}^{n} n_i},\\
&\widehat{D}^{(k+1)} = \frac{\sum_{i=1}^{n} E\left(u_i u_i' \mid y_i \hat{\eta}^{(k)}\right)}{n}, \quad k = 0, 1, \ldots
\end{aligned}$$

The iteration of the procedures mentioned above continues until the convergence, and then the MLE of η can be obtained when the convergence is realized [65].

Once the data set is obtained, empirical Bayesian estimators specified in Equation 3.15 are used to determine the estimates of the random effects.

$$\hat{u}_i = E\left(\beta_i \mid y_i, \hat{\eta}\right) = \widehat{D}(\hat{\eta}) Z_i' V_i^{-1}(\hat{\eta})\left(y_i - X_i\hat{\beta}\right). \tag{3.15}$$

Using the estimates of the random effects obtained by the formula given in Equation

3.15, it is possible to make subject-specific inferences. For instance, the estimated response for the $i-th$ subject is shown in the following equation [65].

$$\hat{y}_i = X_i\hat{\beta} + Z_i\hat{u}_i. \tag{3.16}$$

After the definition of LMM and EM for LMM, the following sections provide details related to the hypothesis testing procedures for the elimination of insignificant random and fixed effect terms.

3.2.2 Hypothesis Testing Procedures in LMMs

In statistical modeling applications, hypothesis tests play an essential role in deciding which variables significantly contribute to the model. The hypotheses as statements about the model parameters in LMM can also be denoted by H_0 (null hypothesis) and H_1 (alternative hypothesis). In the formulation of test statistics, the term "reduced model" can be used and it can indicate the model where the statement under H_0 is assumed to be true. In other words, the reduced model, which contains only parameters not being tested, can be compared with the full model, including all the model parameters. In the following two sections, Likelihood Ratio Test (LRT) procedures for both fixed and random effects are mentioned.

3.2.2.1 Likelihood Ratio Tests

The likelihood functions of two distinct models can be compared within the context of LRTs so that one can assess the significance of single or multiple regression coefficients. As a test statistic, the following formula (Eq. 3.17), based on comparing the likelihood functions of the nested and reference models, can be defined in the applications of hypothesis tests for the fixed effect coefficients within the scope of LMMs.

In the following equation, the notations $L_{reduced}$ and L_{full} correspond to the likelihood of the reduced and full model, respectively. The full model is the model that includes

all variables of interest in data while the reduced model is constructed after removing the coefficients of variables whose contributions are in our interest.

$$-2\log\left(\frac{\mathrm{L}_{\text{reduced}}}{\mathrm{L}_{\text{full}}}\right) = -2\log\left(\mathrm{L}_{\text{reduced}}\right) - \left(-2\log\left(\mathrm{L}_{\text{full}}\right)\right) \sim \chi^2_{\text{df}}. \tag{3.17}$$

The critical value or p-value approaches can decide the stated hypotheses. In the critical value approach, the calculated test statistic can be compared with the corresponding quantile value from the χ^2 distribution having degrees of freedom that equals the difference in the number of fixed-effect parameters between the full and reduced model. The larger test statistic can provide evidence against the null hypothesis and one can continue with the full model in this case. On the other hand, the p-value can be calculated using the calculated test statistic and the obtained value can be compared with the predetermined significance level. The null hypothesis states that the reduced model fits the data as well as the full model and the full model can be rejected when the p-value is lower than the significance level.

In this part, we first briefly mention how to check the significance of the fixed-effect parameters, and then the procedures for examining the significance of random-effect parameters will be explained.

When we consider two models with the same random effects with distinct sets of fixed-effect parameters, LRT can be used to compare the likelihood functions of reduced and full models. The ML method instead of the REML algorithm should be used to apply LRT on only the fixed-effect parameters. The likelihood functions are calculated after the models with and without the parameters to be tested fitted based on the ML method. Then the corresponding test statistic can be obtained from the Equation 3.17. When the likelihood values of both the full and reduced models are close to each other, the test statistic will take a small value indicating that the reduced model can work as well as the full model. So, one can continue with the reduced model. On the other hand, the large test statistic value indicates evidence against the statement under the null hypothesis. The full model should be used as there can be stated that the contribution of the regressors under the null hypothesis to the model is significant.

LMM models during testing procedures of the fixed effects are required to use the ML method as a fitting algorithm. Unlike testing the significance of fixed effects, the model should be fitted with the REML method if one would like to test whether to include random effects into the model or not, given that the fixed-effects are the same [63]. Then, `anova` function can be used to compare log-likelihood values of both reduced and full models.

In the application of this procedure in the R programming language, the `gls` function fits marginal linear models without random effects using REML estimation. Then the model including the random effect term can be constructed with `lme` function [66]. After fitting these two models, the significance of a random effect term can be assessed through `anova` function.

3.3 Robust Estimation of Linear Mixed-Effects Models

In this method, the scoring equations of the log-likelihood derivatives are robustified. The residuals and spherical random effects with bounded functions are replaced with the scoring equations to obtain robust estimates. In M-estimation terminology, the bounded functions are called ψ-functions, the derivatives of ρ-function. The Huber function that takes quadratic form for the values around zero and becomes linear for the values outside $\pm k$ is a ρ-function in which k is called a tuning parameter. The choice of this tuning parameter can change the efficiency and robustness of the estimates. Specifically, the larger tuning parameter choices yield more efficient but less robust estimates, whereas smaller cases result in more robust but less efficient estimates. The most common choice is fixing the asymptotic efficiency at 95% of the ordinary estimates. Hence, $k = 1.345$ for the Huber function. By default, ρ-function used in the `rlmer` is the smoothed Huber ρ-function which is defined in Equation 3.18.

$$\psi(x, k, s) = \begin{cases} x & |x| \leq c \\ \text{sign}(x)\left(k - \frac{1}{(|x|-d)^s}\right) & \text{otherwise} \end{cases}, \tag{3.18}$$

where $c = k - s^{\frac{-s}{s+1}}$ and $d = c - s^{\frac{1}{s+1}}$. The value of $s = 10$ is recommended as the properties of both Huber function and the smoothed Huber function become identical asymptotically when this value is taken.

The weights can be obtained from the following formula for a given ψ function.

$$w_{\cdot}(v) = \begin{cases} \psi_{(v)}/x & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases}, \tag{3.19}$$

Although `rlmer` function does not provide fitting LMMs with specific correlation structures for random effects such as compound symmetry, the tuning parameter can be controlled to fit a robust model which includes correlation parameters. Specifically, while k is taken as 2.28 for the models with simple covariance components, this value becomes 5.11 for variance components with correlations for the smoothed Huber function to obtain 95% efficiency [67].

3.4 Hybrid Methods for Longitudinal Data Analysis

3.4.1 Generalized Linear Mixed-Effects Model Trees (GLMM trees)

GLMM trees were first proposed by Fokkema et al. [68] in 2018 so that one can detect interactions among the subgroups of treatments in clustered data. These models compose of two parts, namely global and local. The global part of the GLMM includes the random effect terms with all observations. On the other hand, the local part constitutes the fixed effect terms estimated locally, indicating that the algorithm partitions the observations in the data set based on the additional variables known as partitioning variables. Then, it estimates a fixed-effect model in each of the separated portions of the data [68]. The GLMM trees algorithm has the flexibility of detecting subgroups as this method works based on model-based recursive partitioning (MOB) introduced by Zeileis et al. [69] in 2008. MOB assumes that a single global model such as a generalized linear model (GLM) may not fully describe the data. Hence, it looks for any possible partitioning of the data with additional regressors to detect whether there is a better fit in each partition cell. To illustrate the working principle of

MOB, we will first handle the procedure for the model with only a fixed effect term. Then, we will explain the method for the clustered or longitudinal data structure.

3.4.2 Model-based recursive partitioning

The logic of MOB is that the data may not be described well by a single global model such as GLM in some cases, and the data can be partitioned concerning any other additional covariates. In such cases, it can be possible to find better fits in each partition for the data. For instance, one can fit a global GLM to examine the treatment effect on the response, but this effect will be the same for all observations as the fitted model can exhibit the same coefficient for the indicator of the treatment variable. On the other hand, the partition of data concerning other regressors can be needed to get separate models through different sets of observations in data if distinct groups exhibit different treatment effects.

We can write down a single global GLM which can model the expected response y_i given treatment covariate x_i through a link function as follows:

$$g\left(\mu_i\right) = x_i^\top \beta, \tag{3.20}$$

where x_i corresponds to a matrix of predictor values for $i-th$ data observation, β is the vector including the fixed effect regression coefficients. In this formulation, μ_i, which is the expected response given the predictor x_i, is linearly modeled by $x_i\beta$ through a link function of $g()$. The best-fitting may not be obtained with such a single model, especially when partitioning variables exist. This model cannot fit data well if there are potential partitioning variables as it considers the same effect/coefficient on the response for all data points. The MOB algorithm can take the other regressors into account while finding partitions of data, and it can fit better local models. The MOB algorithm utilizes parameter stability tests on a set of partitioning variables to achieve this.

3.4.2.1 Including random effects

As it is mentioned in the previous sections, using mixed-effects model would be more appropriate for the analysis of clustered or repeated data structure. The GLM extended from (3.20) for the analysis of such data sets can be written as follows:

$$g\left(\mu_i\right) = x_i^\top \beta + z_i^\top b. \tag{3.21}$$

If the model is composed of only a random intercept, z_i represents a unit vector with M components, of which the $m-th$ entry $(m = 1, ..., M)$ is 1 while the other entries take a value of 0. The m represents the cluster of the $i-th$ observation. Furthermore, b corresponds to a random vector of length M, of which each $m-th$ element represents the random intercept for the $m - th$ cluster.

In Equation (3.21), although the clustered structure of data can be accounted by the random part, the global fixed-effects part $x_i^\top \beta$ may not fit data well. Considering this limitation of the mixed-effects model, Fokkema et al. [68] proposed the GLMM tree method. In their proposed method, fixed-effect coefficients can also partition the data.

$$g\left(\mu_{ij}\right) = x_i^\top \beta_j + z_i^\top b. \tag{3.22}$$

As can be seen from the Equation (3.22), the fixed-effect coefficients β_j constitute the local part, and the values of these terms depend on the terminal node j while the random effects b corresponds to the global part of the algorithm.

The steps for estimating the Equation (3.22) using the GLMM tree algorithm are listed below.

Step 0: Set r and all values $\hat{b}_{(r)}$ to 0.

Step 1: Set $r = r + 1$. Fit a GLM tree by using $z_i^\top \hat{b}_{(r-1)}$.

Step 2: Estimate the mixed-effects model $g(\mu_{ij}) = x_i^\top \beta_j + z_i^\top b$ with terminal node $j(r)$ from the GLM tree in Step 1. Get posterior predictions $\hat{b}_{(r)}$ from the fitted model.

Step 3: Repeat Step 1 and Step 2 until convergence.

As the random effects are initially unknown, the algorithm starts by equating them to 0. In each iteration, the algorithm recursively fits the GLM tree in step 1 so that fixed and random effects can be re-estimated in step 2. There is no partitioning on the random effects, but these are estimated globally. On the other hand, the algorithm estimates fixed effects locally in each partition cell.

The GLMM trees can be applied using `lmertree` function under `glmertree` package in R [68]. The `glmertree` package utilizes package `partykit` [70] to obtain partitions and `lme4` package [71] to fit the mixed-effects model.

3.5 Random Effects Expectation-Maximization Tree (RE-EM Tree)

RE-EM tree was first released by Sela and Simonoff [72]. The most general LMM form can be specified to exhibit functional form for the relation between fixed effect terms and response. The underlying model with the distributional assumptions on the error term and random effects can be expressed as:

$$\begin{aligned} y_{it} &= Z_{it}\mathbf{b}_i + f\left(x_{it1}, \ldots, x_{itK}\right) + \varepsilon_{it}, \\ \varepsilon_{it} &\sim \mathrm{N}(0, R_i), \\ \mathbf{b}_i &\sim \mathrm{N}(0, D). \end{aligned} \tag{3.23}$$

In this formalization, while i stands for the longitudinal data objects $(i = 1, 2, \ldots, n)$, the indices of $t(t = 1, ..., T_i)$ is used to represent repeated measures which are especially taken through time. That is, objects in the data exhibit multiple observations in which each observation is represented by a vector of K variables $x_{it} = (x_{it1}, ..., x_{itK})^\top$. The function, f, is introduced to allow different forms for the relation among variables included as fixed effects and numerical response. Also, the known design matrix, Z_{it}, corresponds to the random effect terms and b_i is the unknown random effect coefficient to represent the subject-specific effects.

To fit the model in Equation (3.23), one can use the MLE technique with numerical algorithms such as EM. According to the structure and characteristics of the data set to be analyzed, these traditional methods may suffer from some problems. For instance,

the LMM assumes a parametric form for the function, f (where $f = X\beta$). However, this assumption might be too restrictive because f is generally unknown, and fitting a linear model on the data may not be the best choice. It is important to note that the linear model may not be flexible enough to approximate the real relation between response and explanatory variables. So, it may result in some bias in the estimate of f. Sela and Simonoff [72] proposed a method named Random Effects/EM tree, or RE–EM tree. Their method for longitudinal or clustered data is reminiscent of the EM algorithm, and it can estimate f by incorporating random effects, b_i. In this method, distinct observational units corresponding to the same subject may be included in different nodes. The advantage of the RE-EM tree is indicated as having a flexible structure on the fixed effects compared to the traditional LMM approach. We now move one step further by explaining the estimation methodology of the RE-EM tree method.

The steps for estimating the RE-EM tree are given as follows:

Step 0: Initialize by setting the estimated random effects, $\hat{b}_i$, to zero.

Step 1: Iterate the following two steps of the algorithm until the estimated random effect terms converge which is controlled based on alteration in the likelihood or restricted likelihood function to be less than determined tolerance value:

a. Fit a regression tree approximating the function, f, given that the target variable is $y_{it} - Z_{it}\hat{b}_i$ while attributes are expressed as $x_{it} = (x_{it1}, ..., x_{itK})$ for $(i = 1, 2, \ldots, n)$ and $t(t = 1, ..., T_i)$. Extract a set of indicator variables $I(\mathbf{x}_{it} \in g_p)$ where g_p ranges over all of the terminal nodes using the estimated regression tree.

b. Estimate the mixed effects model, $y_{it} = Z_{it}b_i + I(\mathbf{x}_{it} \in g_p)\mu_p + \varepsilon_{it}$ and get $\hat{b}_i$ from the fitted model.

Step 2: Replace the estimated response value at each terminal node of the fitted tree with the predicted mean $\hat{\mu}_p$ obtained from the mixed-effects model in Step 1b.

The function `REEMtree` under the package with the same name in `R` can be implemented to fit RE-EM tree. This function utilizes both `rpart()` from `rpart` package [73] to construct a regression tree, and `lme()` from `nlme` package [74] to

fit LMM part of the RE-EM tree.

3.6 Unbiased regression trees for longitudinal and clustered data

The RE–EM tree explained in the previous section can incorporate a mixed-effects regression model with tree-based algorithms. One of the advantages of this technique is that the RE-EM tree can assume a general structure for the relation between the fixed effects and response. This technique is more robust to parametric assumptions than the models that can handle random effects and regression trees with only fixed effects. The RE-EM tree's drawback is that it utilizes CART as a tree-based algorithm to split data into nodes. CART, which was firstly introduced by Breiman et al. [75], is one of the widely used algorithms to construct regression tree models for longitudinal and multivariate data. However, many researchers argue that the CART algorithm has limitations, such as overfitting and selection bias. Although pruning trees can avoid the overfitting problem, unsatisfactory results can be obtained due to the variable selection (splitting) bias of CART, in that CART tends to select the variables which have a large number of split points [76].

Fu and Simonoff [76] revised the RE-EM tree algorithm and suggested using the conditional inference tree (ctree) of Hothorn et al. [77] rather than using CART in Step 1a to construct an unbiased regression tree method for longitudinal and clustered data which can overcome bias problem.

The implementation of unbiased RE-EM tree can be realized through `REEMctree` function from `party` package in `R`.

3.7 Longitudinal Classification and Regression Tree (LongCART)

The LongCART algorithm, which is firt proposed by Kundu and Harezlak [78], utilizes baseline characteristics as partitioning variables. This algorithm can be utilized for the analysis of longitudinal data, including several distinct subgroups as the correct model parameter values of the mixed-effects can alter among these subgroups. For instance, Raudenbush [79] argued that although it may be applicable to assume

that all subjects in the study exhibit growth in a common function, the magnitude of this change might differ among these different observation units. Apart from the differences in the magnitudes, some studies may include the results changing their directions over time. To illustrate such a situation, Raudenbush [79] gave an example of an analysis of depression and argued that while some patients might experience a decrease in their depression level, some of them might have increasing depression levels. In such cases, a common parametric form assumption cannot capture the subgroup differences in the changes. As Kundu and Harezlak [78] argued, their proposed method of LongCART can overcome such limitations by taking a two-step approach into account. This method can control type I error while giving a decision on further splitting at each node, and it can be applied to situations when the data exhibit multiple measurements taken from the same subject at specific time points. To control the type I error rate; the LongCART algorithm decides whether there is a need for a further split through parameter instability tests at each node. If this decision is that the data can be divided further, the algorithm decides on the optimum number of splitting points.

The working mechanism of the LongCART algorithm can be summarized as follows:

Step 1. Conduct instability test on each partitioning variable separately at a predetermined significance level, α. The significance level is subject to adjustment to control type I error rate for pairwise comparisons.

Step 2. Stop the algorithm if there is no need for a further split at significance level α. Otherwise, select the variable having the smallest p-value for further split and continue to step 3.

Step 3. At each possible cut-off point of the selected partitioning variable, extract the improvement in the goodness of fit criterion such as deviance. Given that X^G is the partitioning variable chosen, the gain in the goodness of fit criterion can be calculated at the cut-off point $c_{(g)}$ through the following two steps:

a. Obtain two splits in which one group includes the observations from individuals with $X^G \leq c_{(g)}$ while the other consists of the observations from subjects with $X^G > c_{(g)}$.

b. Estimate longitudinal model on (i) all the subjects in the node, (ii) the subjects with $X^G \leq c_{(g)}$ and (iii) the subjects with $X^G > c_{(g)}$. Obtain the goodness of fit criteria of these three models. Then, denote them as GOF_{all}, GOF_I, and GOF_{II}, respectively.

c. Extract the improvement in goodness of fit criterion calculating $GOF_I + GOF_{II} - GOF_{all}$.

Step 4. Extract the cut-off value, which maximizes the goodness of fit measure improvement, and use this value for binary splitting.

Step 5. Repeat these four steps for each non-terminal node.

3.8 Gaussian Process Boosting

Boosting [80] is a general technique that can improve the predictive capability of an ML algorithm. Apart from increasing the predictive performance of the algorithm, boosting using trees as base learners can deal with multicollinearity, non-linearities, discontinuities, and high-order interactions. Also, it makes the method robust against potential outliers, and it can be used for data sets with missing values in explanatory variables without losing any information [81]. On the other hand, there are some limitations to boosting, such as the observations assumed to be independent among different data points. The algorithm cannot work accurately when the residuals exhibit a correlation. Besides the correlation structure in data, boosting has difficulty with the categorical variables on many levels.

Rasmussen [82] define Gaussian Process as "a collection of random variables, any finite number of which have (consistent) joint Gaussian distributions." These processes can allow flexible non-parametric models to provide superior predictive accuracy and make probabilistic predictions. As mentioned in the previous sections, mixed-effects models have widely used techniques for panel or longitudinal data. The observations may exhibit correlation due to the grouping structure of such data sets. These models can handle categorical variables with high cardinality.

The mean is assumed to be either zero or a linear function of given independent variables in the Gaussian process and mixed-effects regression models. The structured

residual variation can be modeled using a Gaussian process with zero-mean and a mixed-effects model. However, assuming both zero-mean and linearity is generally not realistic, and it may be required to relax these assumptions to get high predictive performance [81].

Considering the difficulties above in methods, Sigrist [81] proposed a technique that can combine tree-boosting with the Gaussian process and mixed-effects model. The `gpboost` package in `R` can be used to implement this method.

3.8.1 Combining Gaussian process and mixed-effects models with boosting

The mixed-effects model with a flexible functional form to represent the relation between fixed effects and response can be expressed as in Equation (3.24).

$$\begin{aligned} y &= F(X) + Zb + \varepsilon, \\ b &\sim N(0, \Sigma), \\ \varepsilon &\sim N\left(0, \sigma^2 I_n\right), \end{aligned} \tag{3.24}$$

where $y = (y_1, ..., y_n)^\top \in \mathbb{R}^n$ is response variable, $F(X) \in \mathbb{R}^n$ represent the fixed effects, $b \in \mathbb{R}^m$ are the random effects with covariance matrix $\Sigma \in \mathbb{R}^{m \times m}$ and $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)^\top \in \mathbb{R}^n$ is the error term. Also, n represents the number of observations, m represents the dimension of the random effects, and p represents the number of fixed effects. For the case of grouped random effects, Z matrix in Equation (3.24) is an incidence matrix $Z \in \{0, 1\}^{n \times m}$ and this matrix relates group-level random effects to data points.

The likelihood function, which is also known as the risk function, for the model given in (3.24) to be optimized, can be written as follows:

$$p(y \mid F, \theta) = \int p(y \mid b, F, \theta) p(b \mid \theta) db,$$

where

$$\begin{aligned} p(b \mid \theta) &= \exp\left(-\frac{1}{2} b^T \Sigma^{-1} b\right) |\Sigma|^{-1/2} (2\pi)^{-m/2}, \\ p(y \mid b, F, \theta) &= \exp\left(-\frac{1}{2\sigma^2} (y - F - Zb)^T (y - F - Zb)\right) \left(2\pi\sigma^2\right)^{-n/2}. \end{aligned}$$

In this formulation, $F = F(X)$ and $\theta \in \Theta \subset \mathbb{R}^q$ represents all variance and covariance parameters. According to the representation in [81], the function "$F(\cdot)$" will be used to denote the function F evaluated at X. Further, it will be assumed that θ_1, the first element of θ, equals to variance, σ^2.

The marginal distribution of the response variable is given by

$$y \sim N(F(X), \Psi), \quad \Psi = Z\Sigma Z^T + \sigma^2 I_n. \tag{3.25}$$

Hence, the negative log-likelihood of this model is

$$L(y, F, \theta) = \frac{1}{2}(y - F)^T \Psi^{-1}(y - F) + \frac{1}{2}\log\det(\Psi) + \frac{n}{2}\log(2\pi). \tag{3.26}$$

The following reparametrizations will be used for the ease of representations:

$$y \sim \mathcal{N}\left(F(X), \sigma^2 \Psi^\dagger\right), \tag{3.27}$$

where

$$\Sigma^\dagger = \Sigma/\sigma^2 \quad \text{and} \quad \Psi^\dagger = \Psi/\sigma^2. \tag{3.28}$$

The goal is to minimize the risk function for this given negative log-likelihood $L(y; F; \theta)$.

$$R(F(\cdot), \theta): \quad (F(\cdot), \theta) \mapsto L(y, F, \theta)|_{F=F(X)}, \tag{3.29}$$

As stated before, F(.) is a function at X, and its function space will be denoted by $\mathcal{H}$.

According to the algorithm proposed by Sigrist [81], the following joint minimizer can be found combining the Gaussian process with LMM:

$$(\hat{F}(\cdot), \hat{\theta}) = \underset{(F(\cdot),\theta)\in(\mathcal{H},\Theta)}{\operatorname{argmin}} R(F(\cdot), \theta). \tag{3.30}$$

3.8.2 Boosting when θ is fixed

In this section, the algorithm will be explained under the assumption of a fixed variance-covariance matrix, θ. When θ is given, boosting algorithm performs in an iterative manner updating the estimate $F_{m-1}(\cdot)$ by adding $f_m(\cdot)$ to its current value.

$$F_m(\cdot) = F_{m-1}(\cdot) + f_m(\cdot), \quad f_m \in \mathcal{S}, \quad m = 1, \ldots, M, \tag{3.31}$$

In gradient boosting, $f_m(\cdot)$ is found by the least-squares approximation

$$f_m(\cdot) = \underset{f(\cdot)\in\mathcal{S}}{\operatorname{argmin}} \left\| \Psi^{-1}\left(y - F_{m-1}\right) - f \right\|^2, \tag{3.32}$$

where $f = (f(X_1), ..., f(X_M))^T$. For a fixed θ, Newton boosting finds the minimizer of a second-order functional Taylor approximation

$$f_m(\cdot) = \underset{f(\cdot)\in\mathcal{S}}{\operatorname{argmin}} \left(y - F_{m-1} - f\right)^T \Psi^{\dagger^{-1}} \left(y - F_{m-1} - f\right). \tag{3.33}$$

There is also a hybrid gradient-Newton boosting method which combines these two methods. In this case, the base learners are assumed in the following form:

$$f(\cdot) = h(\cdot;\alpha)^T\gamma, \quad h(\cdot;\alpha), \gamma \in \mathbb{R}^K, \quad \alpha \in \mathbb{R}^Q, \tag{3.34}$$

where α and γ represent parameters of the base learners and $h(\cdot;\alpha) : \mathbb{R}^p \to \mathbb{R}^K$. This denotes the regression tree case in which α is for the splitting variables and split locations while γ includes the values for terminal nodes. Further, $h(\cdot;\alpha)$ represents a function that maps the predictors to terminal nodes of the tree. In such a case, this boosting method initially learns α_m using gradient boosting defined in (3.32), and then it learns γ_m using a Newton boosting step defined in (3.33). An explicit form of the generalized least squares solution is given by

$$\gamma_m = \left(h_{\alpha_m}^T \Psi^{\dagger^{-1}} h_{\alpha_m}\right)^{-1} h_{\alpha_m}^T \Psi^{\dagger^{-1}} \left(y - F_{m-1}\right), \tag{3.35}$$

where $h_{\alpha_m} \in \mathbb{R}^{n\times K}$ is the matrix with elements $(h_{\alpha_m})_{ik} = (h(X_i;\alpha_m))_k$, $i = 1, ..., n$, $k = 1, ..., K$.

The update in (3.31) can be damped by a factor ν to get higher predictive performance:

$$F_m(\cdot) = F_{m-1}(\cdot) + \nu f_m(\cdot), \quad \nu > 0, \tag{3.36}$$

where ν is called the learning rate.

3.9 Model Performance Metrics

In the literature, there are several measurements to evaluate the goodness of models with continuous response. They can measure the model performance by compar-

ing the predicted values of the dependent variable with actual response values. In this study, we examine the Mean Squared Error (MSE), Root-Mean-Squared Error (RMSE), and Mean Absolute Error (MAE) results of different methods to compare their fitting performances. In principle, although these mentioned performance measures can evaluate the model performance using the train data used to develop the algorithm, it is also possible to get these metrics based on the newly seen test data set. We demonstrate the only train data set approach results after explaining the model-performance measures in the next section.

MSE is based on the averaged squared residuals, which are the differences between the fitted and actual response values. In other words, MSE can be defined as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{3.37}$$

where $\hat{y}_i$ is the fitted response value while y_i is the observed response for the $i-th$ observation, and n represents the number of data points.

Although MSE is a widely used metric for evaluating the model performance, this measure can be affected by the outliers, which are observations showing different fashion than most of the remaining data points, as it gives them the equal weight of all differences. In other words, MSE is sensitive to the data points with large residuals.

RMSE is the squared version of MSE as given in the following equation.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{3.38}$$

While MSE shows the averaged squared error, RMSE demonstrates the average deviation between the predicted dependent variable and observed response in the model. RMSE is used more often since this measurement has the same unit as the response variable, unlike MSE, which is measured in the squared unit of the dependent variable.

MAE measures the averaged magnitude of residuals. The formula of this measure is given as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{3.39}$$

Unlike MSE, which is calculated by squaring the errors, MAE takes the absolute difference between the fitted and actual response values. This is the principal reason why MAE is known to be more robust against outliers compared to MSE.

CHAPTER 4

DATA ANALYSIS

In this section, we first give an introduction to data. Then, pre-processing steps implemented on raw data in MATLAB are explained. After explaining these steps, the exploratory data analysis results are provided to get insight about the individual variables and their relations with each other. At the end of the following section, we investigate the fitting performances of the different methods through n-back data. The implementations of the mentioned algorithms are realized in `R` programming language. After constructing the models, MSE, RMSE, and MAE values are estimated to evaluate their performances. Additional to these metrics, the speed of the implemented methods is assessed through elapsed time obtained by `system.time` function in `R`.

4.1 Data on n-back task

An open-access EEG-fNIRS data include brain activity measurements of 26 healthy participants who performed three distinct cognitive tasks in order of n-back, discrimination/selection response (DSR) task, and word generation (WG) tasks [83]. The demographic characteristics of data indicated that there were nine males and 17 females in the experimental study while their ages were ranged from 17 to 33 years. Also, the mean age of all participants was reported as 26.1, with a standard deviation of 3.5 years. There were no reported participants who have been diagnosed with any neurological, psychiatric, or other brain-related diseases which can potentially affect the results.

The collection of fNIRS data set during each task was realized with 16 sources and

16 detectors system that was placed at frontal, motor, parietal, and occipital areas of the subjects. According to the international 10-5 system, the fNIRS channels were created around AFpz, AFp3, AFp4, AFp7, AFp8, AF1, AF2, AF5h, AF6h, AF7, AF8, AFFz, AFF3h, AFF4h, AFF5h, AFF6, FCC3, FCC4, C3h, C4h, C5h, C6h, CCP3, CCP4, CPP3, CPP4, P3h, P4h, P5h, P6h, PPOz, PPO3, PPO4, PO1, PO2, and POOz. In addition to the fNIRS data set for the n-back task, including neural activity obtained through NIRScout technology at a sampling rate of 10.4 Hz, behavioral measures of 26 different experimenters who participated in the n-back task are provided.

The subjects are presented with a series of numbers as visual stimuli. In the experiment, participants are asked to respond when the correct number matches the previously seen in 2 or 3 trials is displayed on the screen. As behavioral data, the response times are recorded on each trial even if the corresponding participant did not give the correct answer for the stimuli. Hence, it is also possible to check the accuracy measures of the subjects as the data set include correct and incorrect responses.

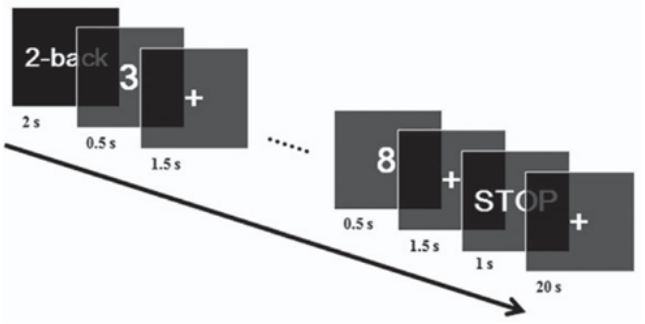


Figure 4.1: Modality of single n-back task

In this study, only data for the n-back task is analyzed. In the n-back dataset, three sessions correspond to each subject's behavioral measures. Participants are provided with nine equally distributed 0-, 2-, and 3-back tasks in each session in different orders. As can be seen in Figure 4.1 which was taken from [83], each n-back session begins with a 2 s instruction showing whether the task is 0-, 2- or 3-back task. After a short beep (250 ms), the task period starts and the numbers are seen on the screen. The timing order of each of these nine series is constructed with 20 s instruction, 40 s task period, and 20 s resting state. A short beep with the same duration is also provided to indicate the end of the task period additional to the 'STOP' word, which

displays the duration of 1 s on the screen. After this sound, participants switch to a rest period. Also, a fixation cross appears on the screen when the participants are in the resting state. In the n-back task period, the random numbers ranging from 0 to 9 were displayed on the screen for 0.5 s and then fixation cross appeared with a duration of 1.5 s. In each task period, there were 20 trials in which the targets occurred with a probability of 30% while non-targets have 70% of a chance to be seen. During the 0-back, participants are asked to press either button numbered 7 with their right index finger or button 8 with their right middle finger so that the analyst can be ensured that the subjects are engaged with the experiment. In other 2- and 3-back conditions, the number 7 button was considered a target button, while non-target cases were represented with a button of 8. Unlike the 0-back task period, experimenters should remember the previous numbers during 2- and 3- back tasks. They were asked to press the target button if the seen number is matched with the 2 or 3 preceding digits, respectively. If the displayed number is not matched with the previously seen one, they should press the non-target button. The participants are provided with a resting period when each session ends to relax, and their brain state can return to a baseline.

This experiment is designed and conducted by Shin et al. [83]. The open raw data can be found at `http://doc.ml.tu-berlin.de/simultaneous_EEG_NIRS`. After gathering the raw data from the source, MBLL is implemented to convert optical density measures to HbO and HbR concentrations. Biologically, an increased blood flow occurs to the active brain region [84]. This results in a decreasing HbR and increasing HbO concentrations [85] which are indirect indicators of cortical activity. Only HbO concentrations are analyzed since Hoshi [17] stated that they are the most reliable measures to work on alterations in regional cerebral blood flow. Then Butterworth low-pass filter is implemented on the data, including concentration changes with the help of `butter()` function in MATLAB. The observations corresponding to each block (i.e., 0-, 2-, and 3-back) were separated with the command of `proc_segmentation()`.

The rest of the analyses such as LMM and ML algorithms are implemented in `R` programming language [86]. The variable names with their descriptions are listed in Table 4.1.

Table 4.1: Variable Descriptions

Variable Name	Variable Type	Description
Values	Numeric	Indicating HbO values
Accuracy	Numeric	The ratio of correct responses over all responses
MeanRT	Numeric	Mean response time on each n-back condition
Age	Numeric	The age of each subject
n-back	Factor	Indicating 0-, 2- and 3-back condition levels
Subject	Factor	ID for each of the 26 subjects
Gender	Factor	Indicating whether the subject is Female or Male
Session	Factor	Indicating the levels for each three sessions
Indices	Factor	Indicating the locations of fNIRS channels

The descriptive statistics for quantitative variables in terms of both demographical and experiment-related variables in this data set are summarized in Table 4.2.

Table 4.2: Summary statistics of the variables

	Values	Accuracy	MeanRT	Age
Min.	-4.39×10^{-3}	51.85	0.31	17.00
1st Qu.	-4.37×10^{-4}	80.62	0.64	24.00
Median	5.31×10^{-5}	89.06	0.81	26.00
Mean	4.84×10^{-5}	88.18	0.82	26.00
3rd Qu.	5.14×10^{-4}	96.29	0.98	28.00
Max.	8.68×10^{-3}	100.00	1.28	33.00

In addition to the descriptive statistics for quantitative variables, MeanRT and Accuracy measures of each subject for three different sessions of the experiment are obtained. The distribution of MeanRT according to n-back conditions for distinct sessions of the experiment can be seen in Figure 4.2 for a randomly selected subject. Also, error bars are added on each bar to indicate the variability in the obtained measures. According to bar plot in Figure 4.2, highest MeanRT results are occurred at 3-back condition in all sessions. Specifically, the highest MeanRT is obtained as 0.96 ± 0.24 and this result belongs to the 3-back condition in the first session. Also, 2- and 3-back conditions have higher MeanRT compared to 0-back condition in which

participants do not have to remember the previously seen numbers.

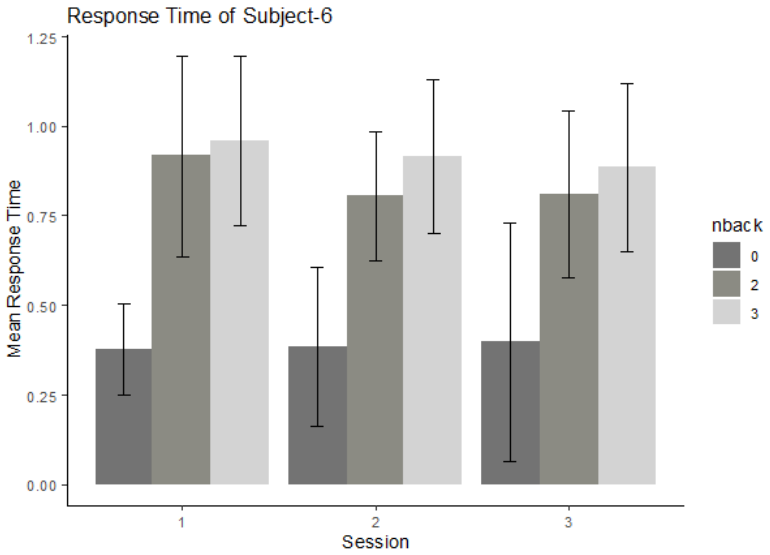


Figure 4.2: Distribution of MeanRT for Subject-6 (randomly selected)

The rest of the MeanRT plots of all subjects can be examined at `https://github.com/statserenay/Subjects_MeanRT_Plots`. As 0-back condition is considered as a control state for detecting whether the subject is dealing with the task, Accuracy results corresponding to only 2- and 3- back conditions are illustrated in Table A.1 in Appendix. According to this table, the accuracy is higher in 2-back condition compared to 3-back for most of the subjects.

Further, the distribution of the response variable indicating mean HbO measures is obtained for each subject in the data. The resulted plot is shown in Figure 4.3. According to this plot, there are some differences in the distribution of response variable for distinct subjects. To illustrate, while the distribution has a symmetric shape for subject-7, it exhibits a right-skewed shape for subject-3.

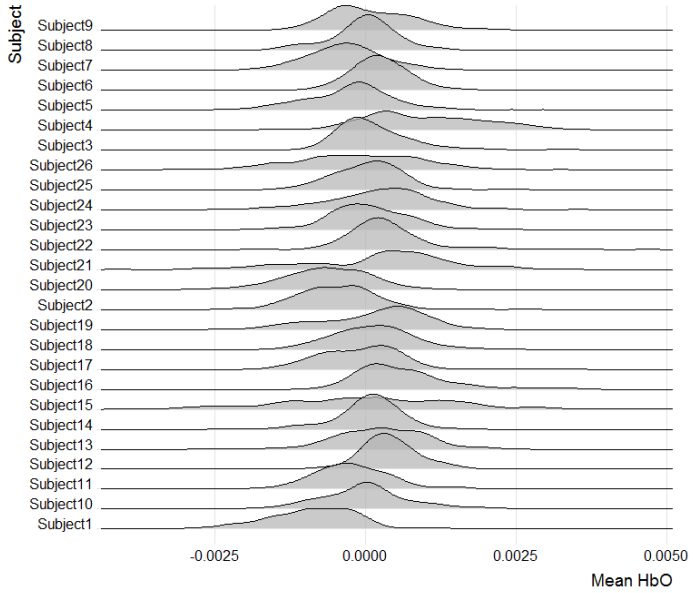


Figure 4.3: Distribution of Mean HbO Values for each Subject

In Figure 4.4, a 3-dimensional scatter plot is shown to examine how the response variable is associated with MeanRT and Accuracy together. This graph indicates that the mean HbO values differ for each subject and fluctuations between observations.

Up to now, exploratory data analysis results on variables for randomly selected subject or all subjects in the data are covered. In addition to these findings, the relations between quantitative independent variables and response are examined for each different indices. The relation between response with Accuracy and MeanRT variables taken from different locations (i.e., the location of optode) within all the subjects are obtained, and can be found at `https://github.com/statserenay/SubjectsRTandAccuracyvsY`. We only illustrated the resulted plots for a randomly selected subject in Figure 4.5 and Figure 4.6. The different characteristics from these plots are observed. This first information kept from the graphs pointed out that the nested structures of the locations within the individuals should be added to the model, and the intra-observational dependency structures should be included in the analysis. In addition, the apparent differences observed in the graphs according to the n-back labels are a sign that this variable may also have significant effects on the analysis results.

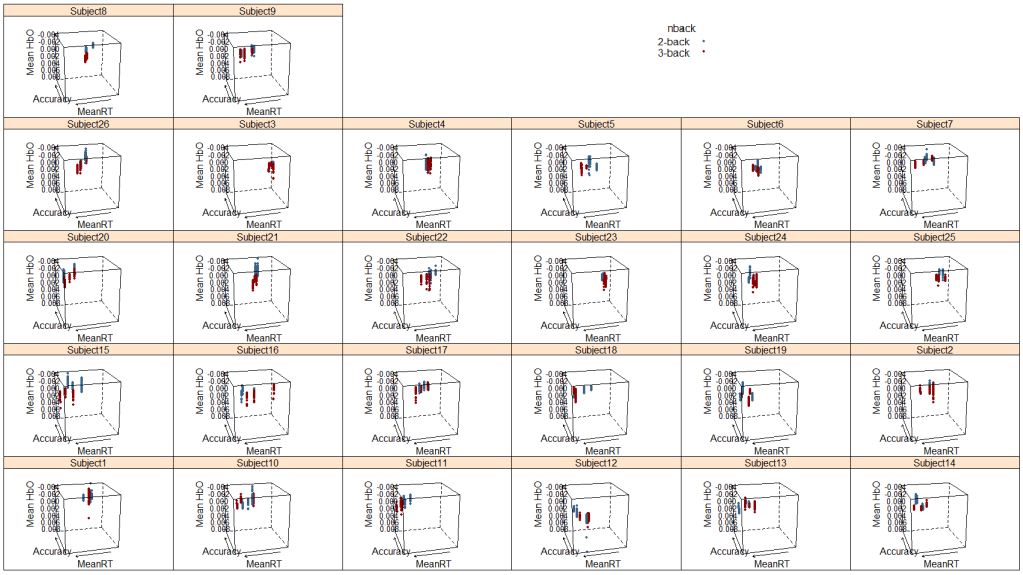


Figure 4.4: Relation between MeanRT, Accuracy and Mean HbO Values

4.2 Implementation of the Methods

As mentioned before, the n-back task includes three distinct types of activity. In 0-back condition, the experimenters must only press a specific button without considering previously seen numbers. This condition is considered to detect whether the subject is dealing with a task or not. In this data set, all experimenters seem to engage with the task as their accuracy measures are 100% whenever they are in the 0-back condition. Hence, we conduct our analyses without considering the observations corresponding to this condition. The data used in the statistical analysis steps are obtained after taking mean HbO values coming from the same n-back conditions in each session. In this data, there are 5616 observations in which each subject has an equal number of data points.

The two distinct models in which one only includes random effect with subject variable while the other takes the nested structure of subject within indices are considered for the analyses. The implementation of statistical modeling approaches and hybrid methods are mentioned in the following sections. After implementing the algorithms on data, we examine three goodness of fit measures which are MSE, RMSE, and MAE. Then we summarized their results of to compare the fitting performances of

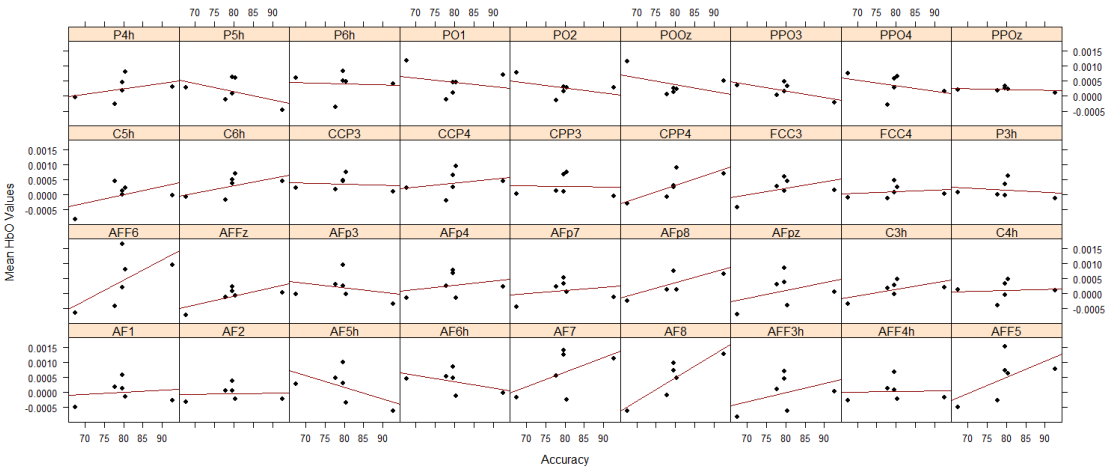


Figure 4.5: Relation between Accuracy and Mean HbO Values for Indices within Subject-6 (randomly selected)

different methods.

4.2.1 Linear Mixed Modeling HbO

In the modeling part, we follow a step-by-step approach which is illustrated in Figure 4.7. All steps and model selections are explained in detail throughout this section.

In Figure 4.8, differences in HbO values between subjects are obvious and so adding a random effect associated with the intercept for each subject is planned to capture the between-subject variation in modeling. Also, measurements are collected from several channels which are nested within each subject. As a result, a full model including all fixed effects is constructed, considering the subject variable as a random effect with the nested structure of indices corresponding to the location of fNIRS channels (Step 1). LRT approaches are applied to identify significant fixed and random effects in the model. Then, variance inflation factor (VIF) values are calculated to check the high correlation between independent variables, so-called multicollinearity. As a result of these calculations, it is worked with independent variable groups with a VIF value below 10 (Step 2). The model without the nested structure is tested and decided to continue with the nested structure (Step 3). After deciding on the significant random and fixed main effects, interaction terms are also added into the models

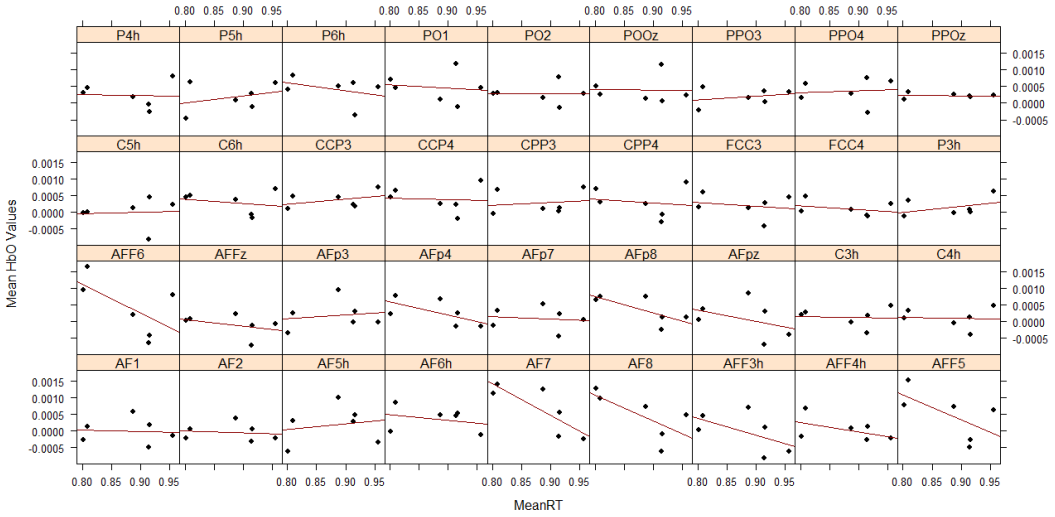


Figure 4.6: Relation between MeanRT and Mean HbO Values for Indices within Subject-6 (randomly selected)

and the procedure is repeated to identify significant terms (Step 4). Furthermore, four different covariance structures, which are Autoregressive process of order 1 (AR(1)), Continuous AR(1), a general positive definite matrix and a compound-symmetric matrix for the random effects are employed and these models are compared using Akaike Information Criterion (AIC). The final decision is given based on the model having the lowest AIC (Step 5). The model procedure is finalized with diagnostics.

Specifically, the variables nback, Session, Accuracy, MeanRT, Age and Gender are included as fixed effects in a full model in which indices are nested within subjects. The argument `random = 1 | Subject/indices` is specified into lme function to fit a model with nesting structure of the random effects. In this specification, the `random = 1` indicates that the random effects are associated with the intercept term. The first term specified in the right side of the vertical bar is the random factor at the highest level (i.e., Subject), while the variable given after a forward slash (i.e., indices) corresponds to a factor whose levels are nested within the levels of the first one. The specification of the model formula (Eq. 4.1) is given as follows:

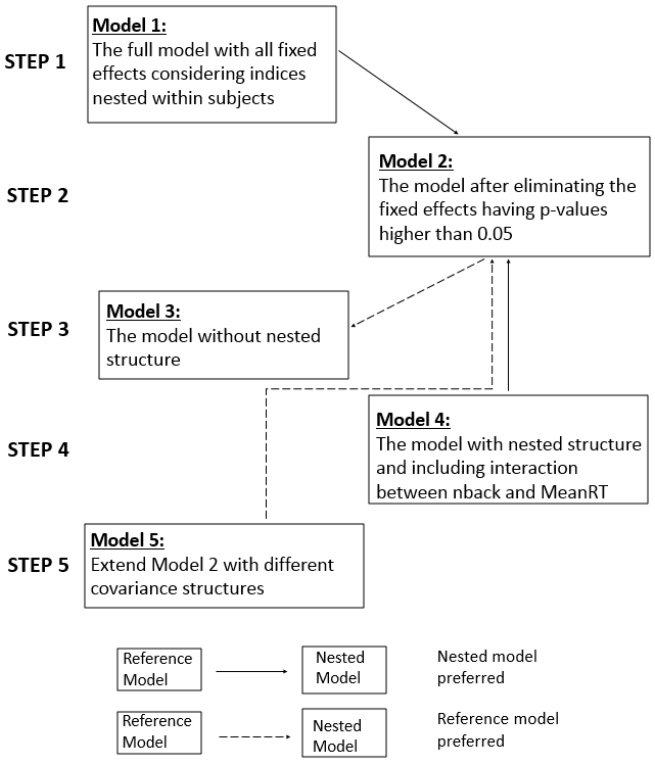


Figure 4.7: Modeling scheme

$$\begin{aligned} \text{values}_{ijk} = \beta_0 + \beta_1 \times \text{nback3}_{ijk} + \beta_2 \times \text{MeanRT}_{ijk} + \\ \beta_3 \times \text{Age}_{ijk} + \beta_4 \times \text{Gendermale}_{ijk} + \beta_5 \times \text{Session2}_{ijk} + \\ \beta_6 \times \text{Session3}_{ijk} + \beta_7 \times \text{Accuracy}_{ijk} \\ + u_{0k} + u_{0j|k} + \varepsilon_{ijk}, \end{aligned} \tag{4.1}$$

where values_{ijk} represents the value of the dependent variable for observation i from channel j nested within subject k; β_0 represents the fixed intercept and β_1 through β_7 represent the fixed effects of the covariates which are nback3, MeanRT, Age, Gendermale, Session2, Session3 and Accuracy; u_{0k} is the random effect associated with the intercept for subject k; $u_{0j|k}$ is the random effect associated with the intercept for channel j within-subject k; and ε_{ijk} represents the error term. The summary of the full model is illustrated in Table 4.3. According to individual p-values, one can conclude that there is a significant difference in mean HbO values among 2- and 3- back conditions when all the other variables are fixed at 5 % significance level. Further, MeanRT contributes to the model significantly once we include all the demographical and experimental-related variables. When the significance of the dummy variables

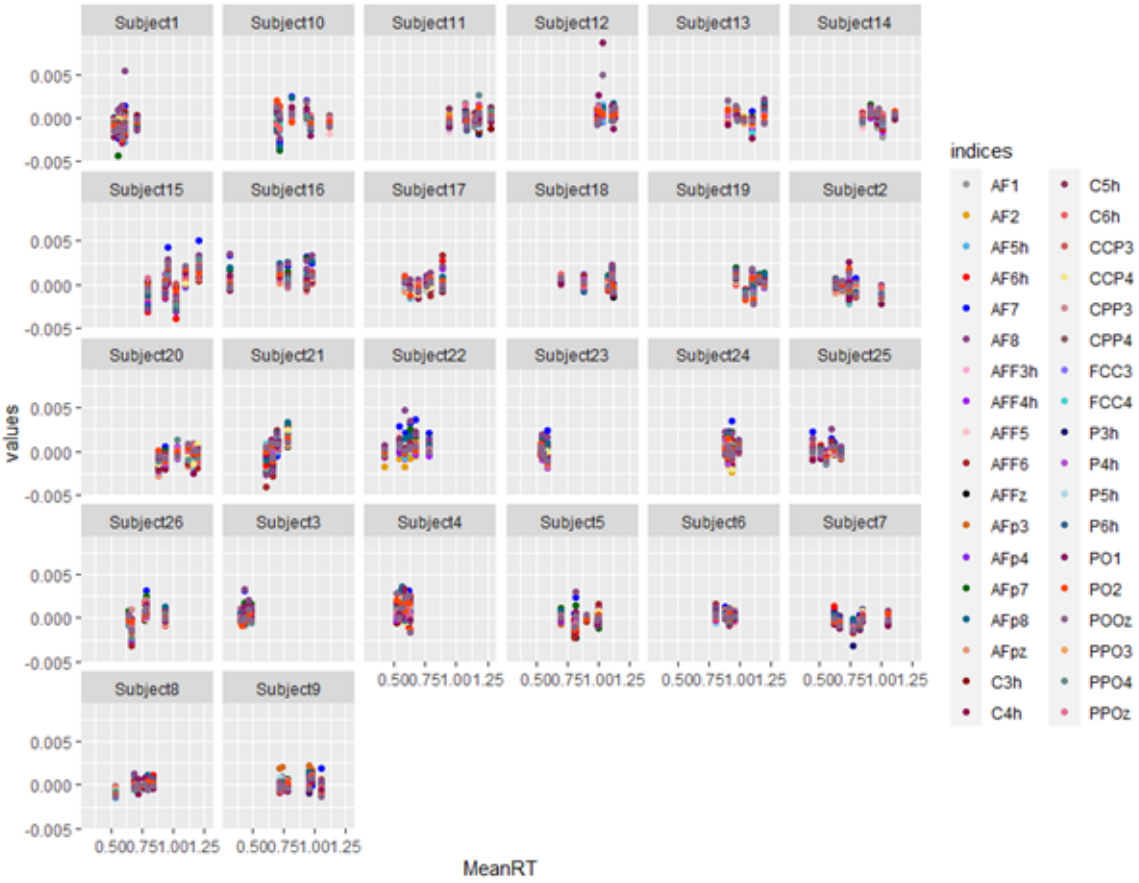


Figure 4.8: Plot of the MeanRT and HbO values for each subject according to different indices

for the levels of Session variable is checked, we can conclude there is a statistically significant difference between the first session and the second one, while the third session does not have significant difference in mean HbO compared to the first session.

The reduced model given in Equation 4.2 is constructed after removing the fixed effects having p-values higher than the predetermined significance level, 0.05. In this stage of the analysis, Session, Accuracy, Age and Gender are removed from the full model and there is concluded that the reduced model performs as well as the full model according to LRT ($\chi^2_5 = 9.45, \text{p}-\text{value} = 0.093$).

$$\begin{aligned} \text{values}_{ijk} = \beta_0 + \beta_1 \times \text{nback3}_{ijk} + \beta_2 \times \text{MeanRT}_{ijk} + \\ + u_{0k} + u_{0j|k} + \varepsilon_{ijk}, \end{aligned} \tag{4.2}$$

Table 4.3: The summary table for the full model

Term	Value	Standard Error	t value	p value
(Intercept)	-1.05×10^{-3}	6.58×10^{-4}	-1.59	0.11
nback3	8.60×10^{-5}	2.77×10^{-5}	3.11	$0.002^{\star}$
Session2	6.46×10^{-5}	2.70×10^{-5}	2.40	$0.02^{\star}$
Session3	5.77×10^{-6}	2.92×10^{-5}	0.20	0.84
Accuracy	-1.81×10^{-6}	2.33×10^{-6}	-0.78	0.44
MeanRT	7.70×10^{-4}	1.18×10^{-4}	6.52	$0.00^{\star}$
Age	2.02×10^{-5}	2.29×10^{-5}	0.88	0.39
Gendermale	8.64×10^{-5}	1.71×10^{-4}	0.51	0.62

$\ast p < 0.05$

After deciding on the significant fixed effects for this model, the model with only a random intercept for each subject (Eq. 4.3) is constructed.

$$\begin{aligned} \text{values}_{ij} = \beta_0 + \beta_1 \times \text{nback3}_{ij} + \beta_2 \times \text{MeanRT}_{ij} + \\ + u_{0j} + \varepsilon_{ij}, \end{aligned} \tag{4.3}$$

Then, LRT is applied for comparing the model with and without nested structure in random effects. The result indicated that the model including indices nested within the subject levels performs better compared to the model without nested structure ($\chi^2_1 = 29.12, \text{p}-\text{value} < 0.0001$). Then, the significance of Subject term is concluded after comparing the models with only fixed effect with the one including only random intercept for each subject ($\chi^2_1 = 1035.5, \text{p}-\text{value} < 0.0001$). The interaction between n-back and MeanRT is added to the model in Equation 4.2 and the new model given in Equation 4.4 is constructed. However, this term was not significant according to the result of LRT. Hence, the model including n-back and MeanRT in which indices are nested within subjects is reached.

$$\begin{aligned} \text{values}_{ijk} = \beta_0 + \beta_1 \times \text{nback3}_{ijk} + \beta_2 \times \text{MeanRT}_{ijk} + \\ + \beta_2 \times \text{nback3}_{ijk} \times \text{MeanRT}_{ijk} \\ + u_{0k} + u_{0j|k} + \varepsilon_{ijk}, \end{aligned} \tag{4.4}$$

The models obtained are fitted under the assumption of no within-group correlations. Considering that the response values obtained from the same fNIRS channels, which are nested within-subjects, may exhibit dependency among distinct subjects, the models are constructed under the assumption that HbO values from the same channels of different subjects have a specific correlation structure. For instance, HbO concentrations taking similar values from Subject15 and Subject22 on the channel coded as AF7 can be detected from Figure 4.8. So, it seems better to look for a model which takes the within-group correlation into account. These models are compared according to their AIC values and the one with the lowest AIC is preferred over the others. To this end, first, the correlation argument is added in the `lme` function corresponding to the within-group correlation structure. As is stated before, as a default, the `lme` function fits a model including no within-group correlations. We fit the models changing this correlation structure and select the final model comparing their AIC values. The results indicating that the argument used in the `lme` function to change the correlation structure and obtained AIC values are summarized in Table 4.4. The AIC of the model assuming no within-group correlation is -64177.27, while the model assuming a general positive definite matrix is -64714.36. The final decision is the model having a general positive definite matrix.

Table 4.4: AIC values with different correlation structures

Correlation Structure	Argument	AIC
No within-group correlation	Default	-64177.27
Autoregressive process of order 1 (AR(1))	corAR1()	-64251.95
Continuous AR(1)	corCAR1()	-64251.95
A general positive definite matrix	**corSymm()**	$\mathbf{-64714.36}$
A compound-symmetric matrix	corCompSymm()	-64175.27

To sum up, the results of the selected model (Eq. 4.4) fitted with a correlation structure of a general positive definite matrix are summarized in Table 4.5.

From Table 4.5, the intercept is interpreted as the mean of oxyhemoglobin concentration when all the predictors have a zero value. In other words, the estimated intercept term, -5.7×10^{-4} is the mean HbO value for a 2-back condition in which MeanRT

Table 4.5: The summary table for the selected model

Term	Value	Standard Error	t value	p value
`(Intercept)`	-5.7×10^{-4}	1.05×10^{-4}	-5.37	$0.00^{\star}$
`nback3`	1.95×10^{-4}	5.59×10^{-5}	7.51	$0.00^{\star}$
`MeanRT`	6.08×10^{-4}	8.76×10^{-5}	6.94	$0.00^{\star}$

$\ast p < 0.05$

is zero. However, as there is no case having a MeanRT as zero, interpreting the estimated intercept term can cause an extrapolation problem for our case. N-back conditions are statistically significant in the model (p-value=0.00). The estimated coefficient of the n-back condition in this model indicates that the mean HbO concentration is 1.95×10^{-4} unit higher for 3-back condition compared to 2-back condition for a fixed MeanRT. In addition, the mean HbO values tend to increase by 6.08×10^{-4} corresponding to each one-unit increase in MeanRT for a fixed n-back condition.

For a more detailed investigation, the residual plots belonging to this model are illustrated in Figure 4.9 and fitted versus actual response values plots for each subject are shown in Figure 4.10. We observe the residuals around zero with several outlier points in Figure 4.9. The plot of the estimated intercept terms for the random effect of Subject is located in Figure 4.11.

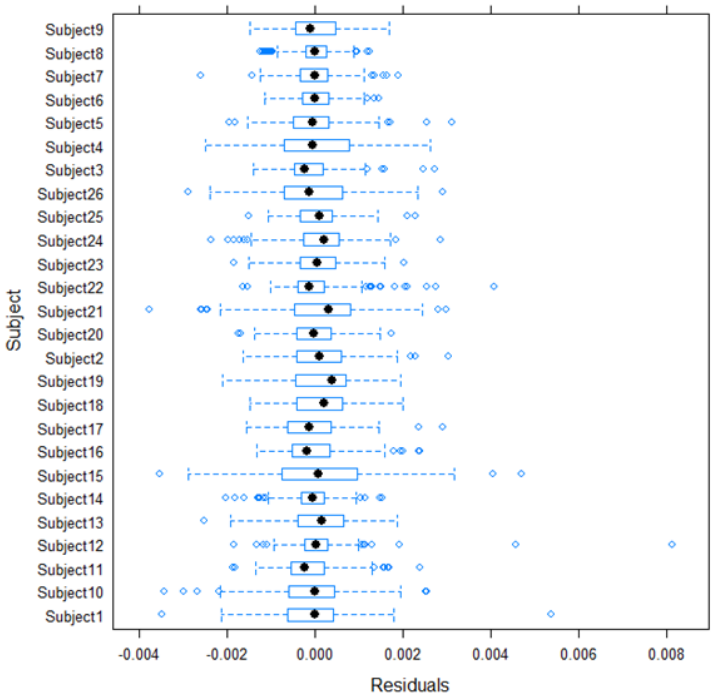


Figure 4.9: Box-plot of residuals for each subject

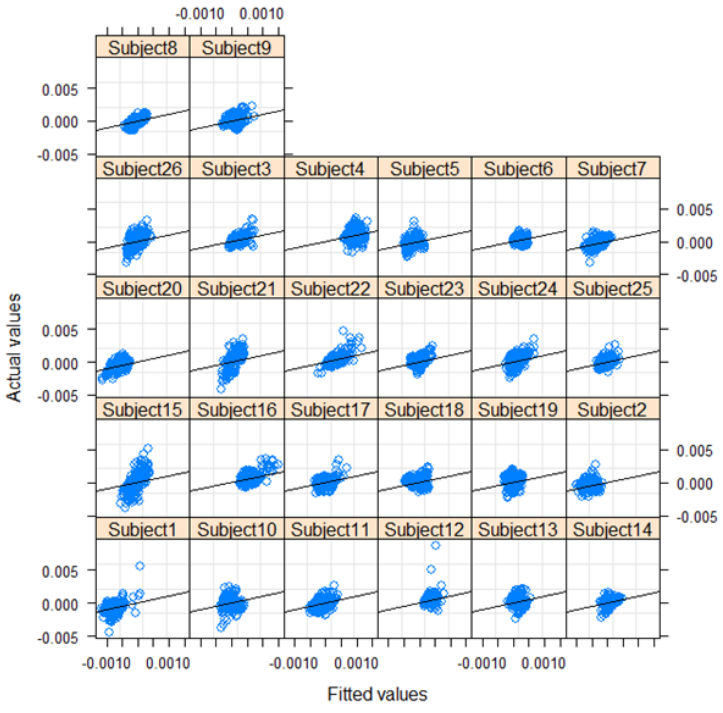


Figure 4.10: Plot of the fitted versus actual values for LMM

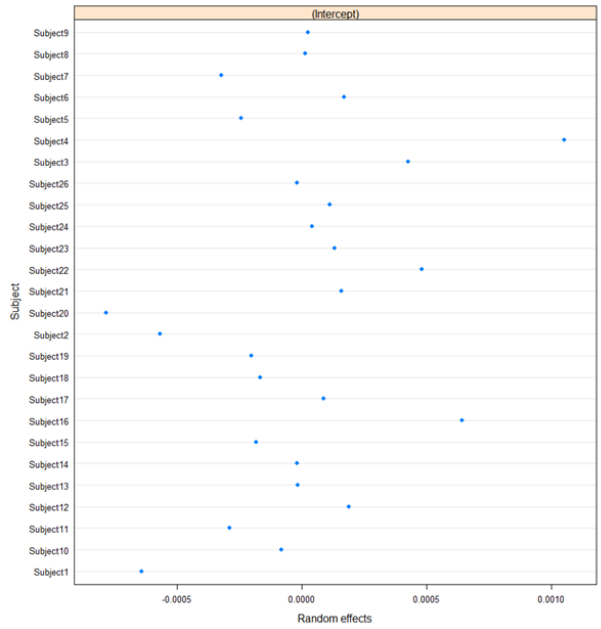


Figure 4.11: Plot of the estimated random intercept terms for subject variable

Also, the estimated intercept terms corresponding to the random effect of the indices nested within subjects for Subject-6 (randomly selected) is illustrated in Figure 4.12. The plots of random effect estimates for all subjects are available at `https://github.com/statserenay/RandomEffectsNested`.

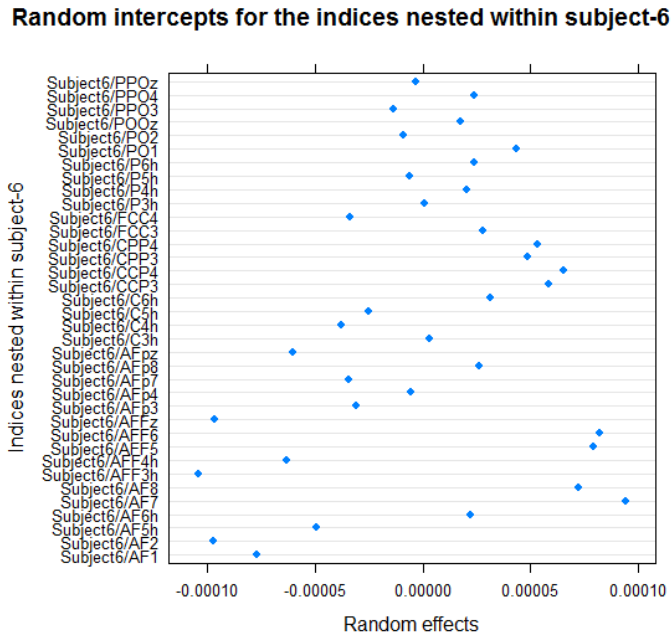


Figure 4.12: Plot of the estimated random intercepts for the indices nested within subject-6

4.2.2 Robustified Linear Mixed Modeling with HbO

In many applications of statistical modeling on real data sets, it is almost inevitable to observe some cases which behave differently compared to the rest of the data points. This situation is also the case in many neural data set as there may exist data points for irregular experimental units.

According to the residual plots given in Figure 4.10, some of the observations may be outliers which can potentially affect the underlying assumptions for the LMM approach, such as normality. Hence, the robust LMMs with and without nested structure are also constructed using `rlmer` function from `robustlmm` package in R [67]. The robustified models include only nback and MeanRT which were significant according to LMM to assess the effect of unusual observations. Further, considering that we have chosen LMM with a specific correlation structure, the robust model is constructed taking $k = 5.11$ as well as with default covariance structure (i.e., $k = 2.28$) of `rlmer` function in R. As log-likelihood and information criteria measures such as AIC are not supported by this function, we use sigma values to compare the models

with these two cases. The model constructed with default covariance parameters are selected as it has lower sigma value (6.91×10^{-4}) compared to the model when k is taken as 5.11 (7.26×10^{-4}).

The resulting residual plots obtained from the robust LMM with nested structure are shown in Figure 4.13 and Figure 4.14.

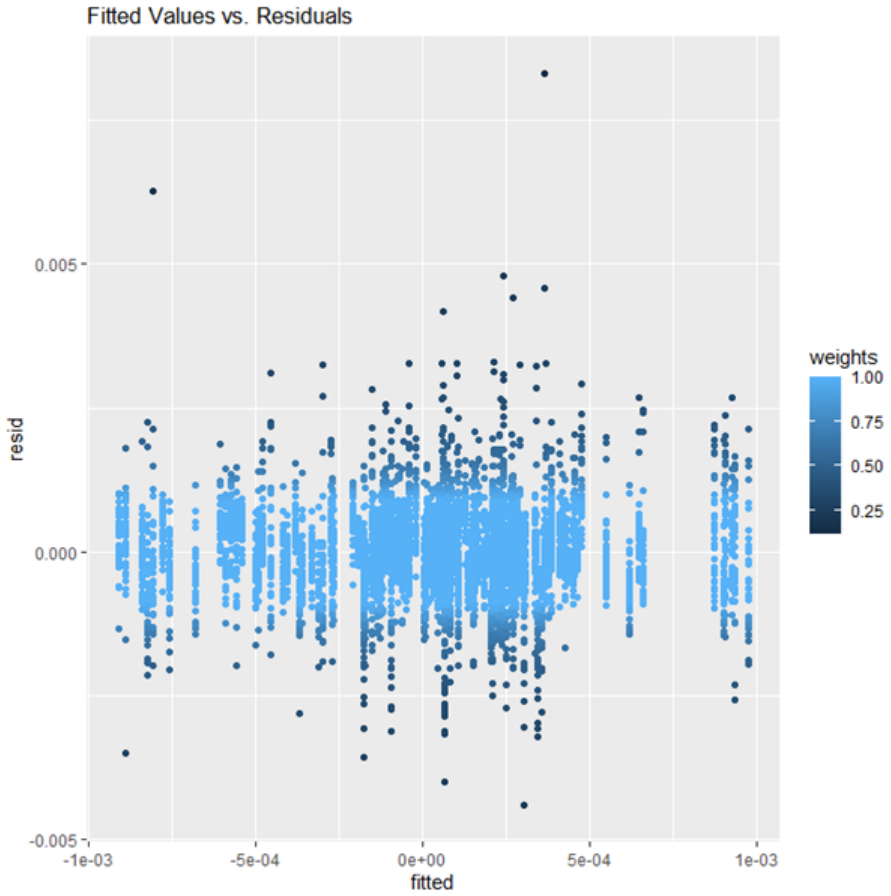


Figure 4.13: Plot of the fitted values vs. residuals from robust LMM with weights

A darker color indicates the observations with a low robustness weight w. From Figure 4.13, some of our observations seem to get low robustness weights compared to the other observations. The fitted values versus the actual response values plot in which the observations are colored according to their weights resulting from the robust regression approach are also obtained for each subject and shown in Figure 4.15. According to this plot, for instance, while some of the individuals mostly have weights closer to 1, the weights for the observations of some of those are arranged to decrease their effect on the estimated regression coefficients. We can again

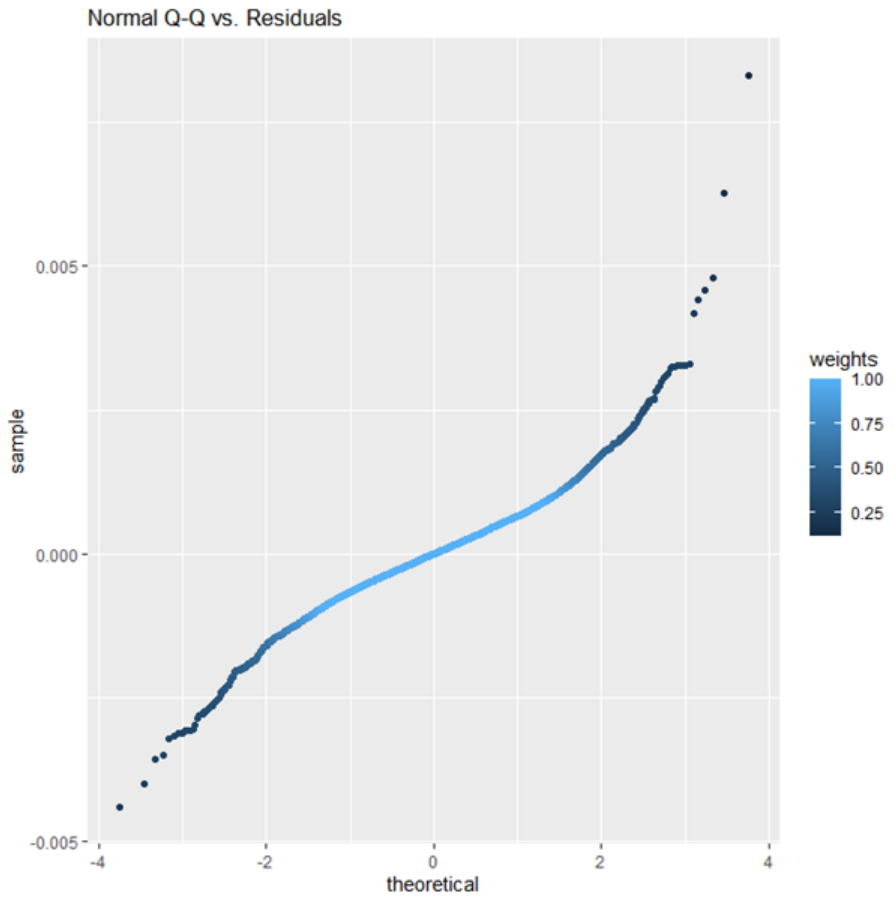


Figure 4.14: Normal Q-Q vs. residuals plot of robust LMM with weights

conclude that considering a robust model seems reasonable. The estimated fixed effects obtained from robust regression are almost the same as the ones obtained from the LMM, which gives the weight of 1 to all data points. So, similar interpretations concluded before according to the results of the fitted LMM can be made from the estimated fixed effect coefficients of the robust model. Although similar estimations were obtained from both models for the fixed effects, the robust model may be preferred as it would reduce the effect of outlier observations. In contrast, when the performance metric results of LMM with its robustified version from Table (4.7) are compared with each other, we can conclude a better fitting performance when LMM is implemented to the data as it provides lower results. On the other hand, while we conclude differences in two model versions including with and without nested structure when LMM fit is obtained, there is no difference in these two cases for robustified model. When the output of robustified LMM is examined, the estimated variance of

the intercept term corresponding to indices:Subject is almost 0. Hence, this model concludes no significant contribution made by nested structure of data. This situation can also be observed from Figure 4.13, as the plot does not demonstrate a random pattern.

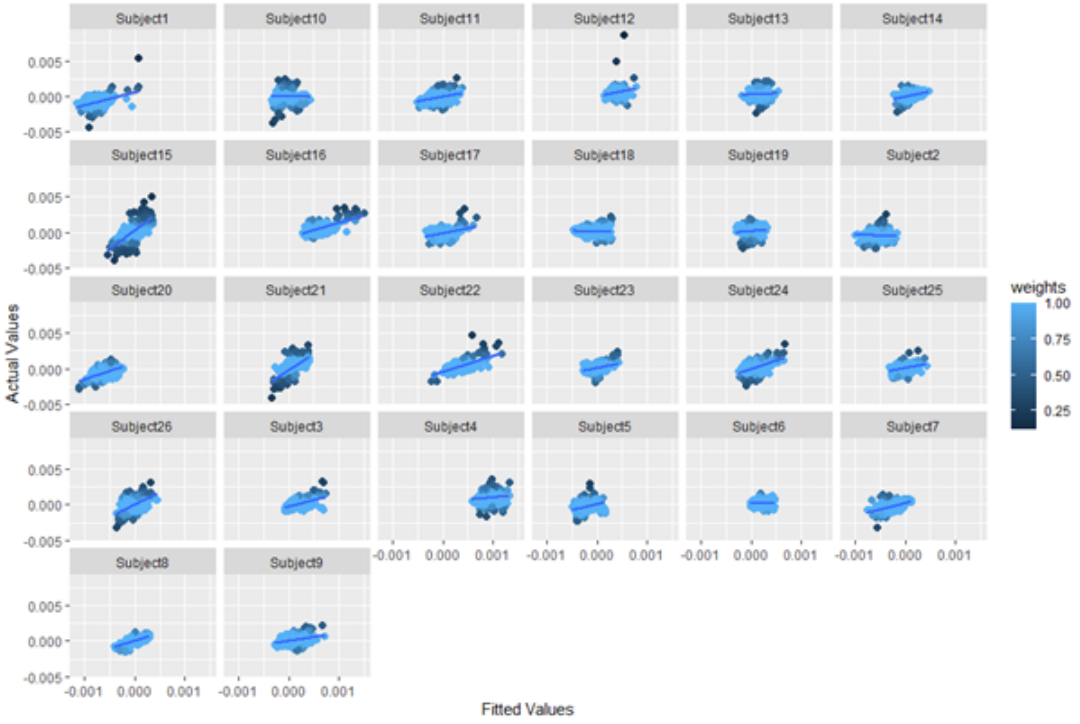


Figure 4.15: Plot of the fitted vs. actual values for each subject with weights

4.3 Hybrid Methods

Hybrid methods represent the models merging LMM with ML algorithms to improve prediction accuracy. First, we have implemented GLMM tree method, both using subjects as only random effects and indices nested within subjects. The resulted regression trees are shown in Figure 4.16 and Figure 4.17, respectively. We first mention the interpretation of the general tree structure and then use this terminology to comment on the plotted GLMM tree for the model without and with the nested structure. In a tree data structure, the variables are represented as the nodes of a tree, and these variables have a significant association with the response, mean HbO values. In any tree structure, the variable in the origin of the splitting can be called a root node, while the nodes at the end of this data structure are known as terminal nodes. Also, each line connecting any two nodes is called an edge. At first, we use this terminology to interpret the part of the GLMM tree shown in Figure 4.16.

This data representation shows that Accuracy, which corresponds to the ratio of correct answers to overall responses given by experimenters, is a root node as this variable initializes the process. In other words, the most crucial variable to split the data is Accuracy which is not significant according to LMM results. First, the algorithm splits the tree with Accuracy at 98.147, forming two heterogeneous groups. Considering the first condition, we can conclude that if Accuracy is less than or higher than a value shown in the edge, which is 98.147, the data can be partitioned accordingly. Since the experimental design requires remembering the numbers in a short time, it does not require a tremendous cognitive effort. So accuracy is significantly high.

The second split assigns observations in Session 1 and 2 to the left branch, and then that group is further subdivided by MeanRT. Observations in Session 3 are assigned to the right branch. That can be explained by the fact that there is no significant difference in mean HbO values among the first and second levels of the session variable, while Session 3 differs from these two levels.

Following a similar strategy with the other variables: Age, MeanRT, Gender, and n-back, the tree is finalized by the terminal nodes, which can be seen from the last line of this tree in Figure 4.16. Overall, the tree stratifies or segments the observations

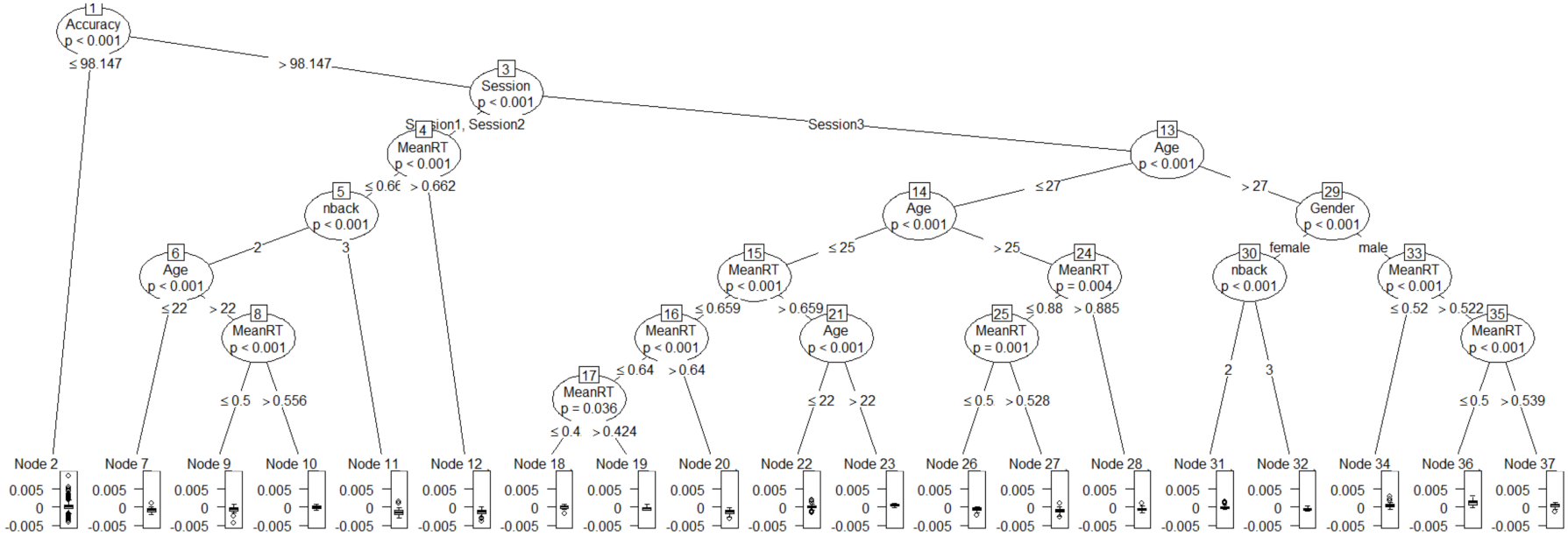


Figure 4.16: GLMM tree without nested structure

into 19 regions of predictor space. From this part of the tree, the distribution of HbO values is grouped into each of these terminal nodes. For instance, we can conclude a right-skewed shape for the distribution of HbO values observed at Node 2, while the observations in Node 7 exhibit a more symmetric distribution. Similar conclusions can be drawn from the remaining partitions of data observed in other nodes. We now compare GLMM trees with nested and without nested data structures. Similar to the case seen without nested data structure in Figure 4.16, Accuracy is the most crucial variable in determining HbO values, and the Session variable follows it in Figure 4.17, represents with nested structure results. In both plots, Session 1 and 2 are in one branch while Session 3 is in another branch. This can be resulted from decreases in cognitive efforts of participants in Session 3 compared to the first two sessions. Unlike the pattern without the nested data structure, the fourth split occurs with the Gender variable rather than MeanRT. Following a similar strategy to interpret the GLMM tree, it is possible to compare the splitting variables and edge values for the rest of the GLMM tree given in Figure 4.17 with the corresponding parts of Figure 4.16.

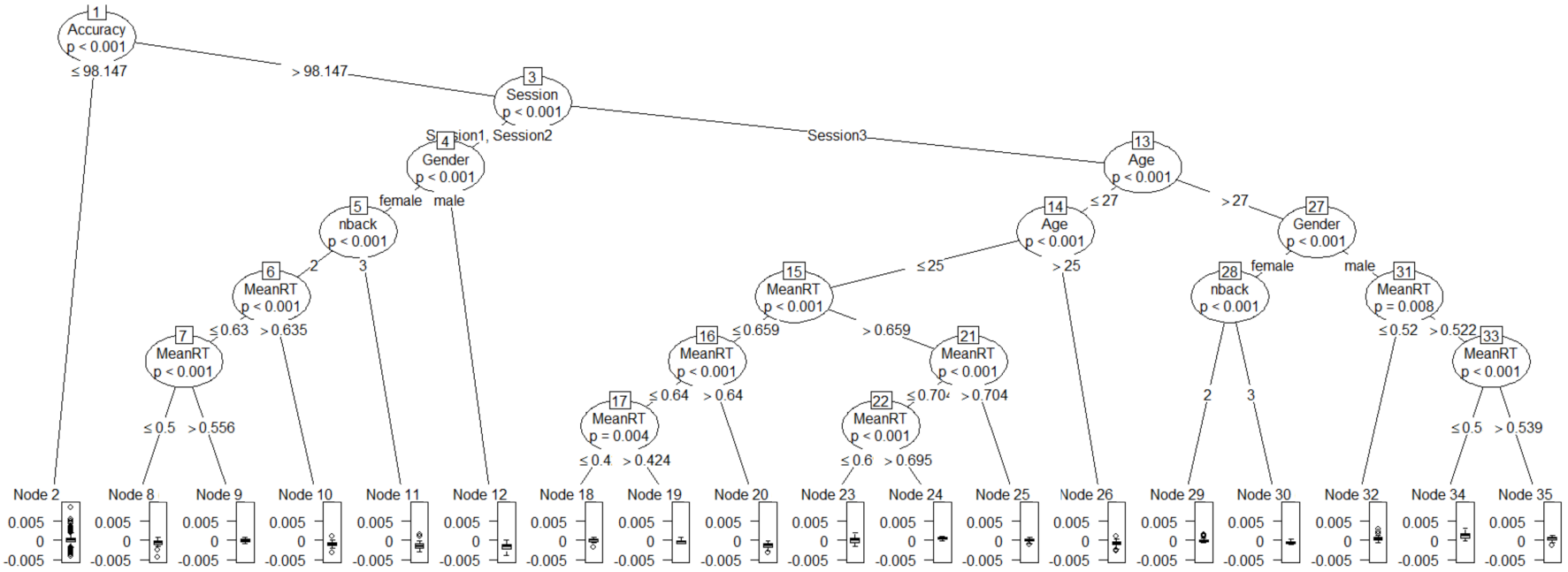


Figure 4.17: GLMM tree with nested structure

As a second hybrid method, RE-EM tree is constructed with the help of `REEMtree` function from `REEMtree` package in `R`. After estimating RE-EM tree, unbiased RE-EM tree is constructed through `REEMctree` function from `party` package in `R` [77]. The codes for REEMctree function are available at `https://pages.stern.nyu.edu/~jsimonof/unbiasedREEM/REEMctree.r`. The functions to construct the RE-EM tree and its unbiased version provide to implement the methods incorporating different correlation structures. The results for the RE-EM tree are summarized in Table B.1 and Table B.2 in Appendix. When AIC values are compared for the trees with different correlation structures, AR(1) is selected for the model without a nested structure, while compound symmetry is preferred for the nested model. After deciding on the models with the lowest AIC values, performance measures are summarized in Table B.5 to compare the models with the default correlation structure in the `REEMtree` function in R. The same procedure is repeated for the Unbiased RE-EM tree, but log-likelihood values are compared since AIC is not available in the R function. The results are displayed in Table B.3 and Table B.4. According to these tables, log-likelihood is highest when compound symmetry is assumed for the model without nested structure, while AR(1) provides better results for the model with nested structure. Hence, the performance metrics are obtained for the Unbiased RE-EM tree with these correlation structures, and the results are presented in Table B.6. As seen in Tables B.5 and B.6, there are no big differences among the performance measures of the methods with different correlation structures for the RE-EM tree and its unbiased version compared to the cases using the default correlation structure. Hence, the results for hybrid methods are provided with a default correlation structure in the next section.

In the next step, we have implemented the LongCART algorithm. The function in `R` requires the partitioning variables, which are not changing over time. Only age and gender variables satisfy this condition. After specifying all possible combinations of these two variables, the algorithm converged when only gender was used as a partitioning variable. However, the resulting model indicated no split, and performance measures of LongCART are not calculated.

We have finalized the analyses with GPBoost algorithm using the `gpboost` package in `R` [87]. The GPBoost algorithm is constructed with different optimizer algorithms,

Table 4.6: GPBoost with different optimizers

	GD	WLS	NM	BFGS
Without nested structure				
RMSE	7.91×10^{-4}	7.90×10^{-4}	7.98×10^{-4}	7.98×10^{-4}
MSE	6.24×10^{-7}	6.24×10^{-7}	6.37×10^{-7}	6.37×10^{-7}
MAE	5.80×10^{-4}	5.80×10^{-4}	5.80×10^{-4}	5.80×10^{-4}
Number of iterations until convergence	150	4	2	3
System time	0.06	0.00	0.01	0.02
Negative log-likelihood	-32079.7	-32084.6	-32017.1	-32026.8
With nested structure				
RMSE	7.72×10^{-4}	7.71×10^{-4}	7.80×10^{-4}	7.80×10^{-4}
MSE	5.96×10^{-7}	5.95×10^{-7}	6.08×10^{-7}	6.08×10^{-7}
MAE	5.68×10^{-4}	5.68×10^{-4}	5.70×10^{-4}	5.70×10^{-4}
Number of iterations until convergence	127	6	2	2
System time	0.14	0.10	0.07	0.01
Negative log-likelihood	-32161.6	-32166.4	-32060.6	-32060.6

which are Gradient Descent (GD), Weighted Least Squares (WLS), Nelder Mead (NM), and Broyden–Fletcher–Goldfarb–Shanno (BFGS). The performance metrics calculated from the GPBoost algorithm with different optimizer functions are illustrated in Table 4.6. According to these results, although each technique provides close results, the one with the lowest RMSE is WLS, the default optimizer for the Gaussian data.

4.4 Results of all methods

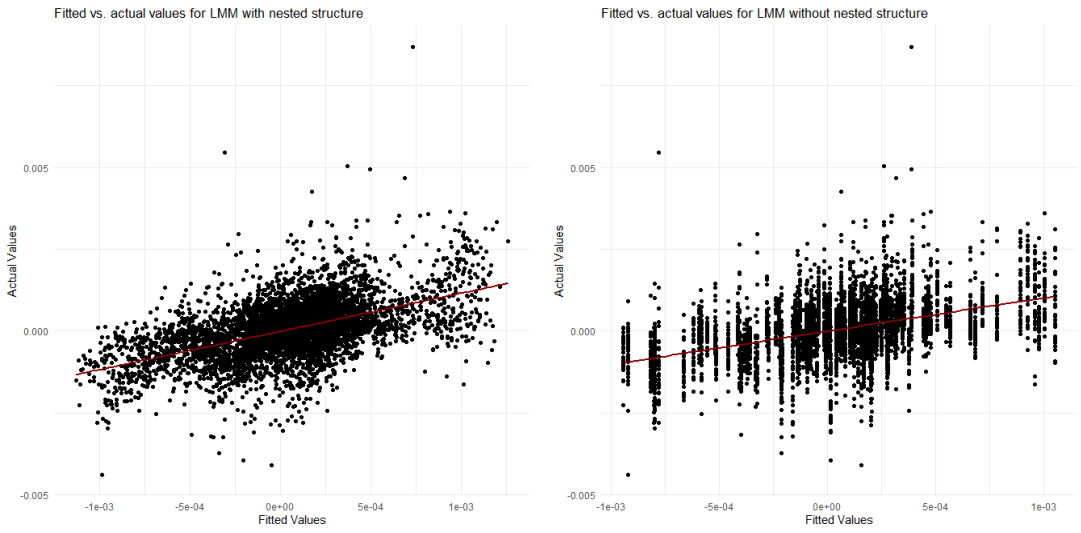
After obtaining model fits and implementing hybrid algorithms on the data, it is intended to compare their prediction performances. However, as LongCART algorithm shows no partition of data, the results for the performance measures of the methods except LongCART are depicted in Table (4.7).

Table 4.7: Results for all methods

Method	RMSE	MSE	MAE	System Time
LMM				
without nested structure	7.9×10^{-4}	6.2×10^{-7}	5.8×10^{-4}	0.11
LMM				
with nested structure	7.6×10^{-4}	5.7×10^{-7}	5.5×10^{-4}	0.95
Robustified LMM				
without nested structure	7.9×10^{-4}	6.3×10^{-7}	5.8×10^{-4}	284.98
Robustified LMM				
with nested structure	7.9×10^{-4}	6.3×10^{-7}	5.8×10^{-4}	591.08
GLMM tree				
without nested structure	6.1×10^{-4}	3.8×10^{-7}	4.3×10^{-4}	230.67
GLMM tree				
with nested structure	$\mathbf{5.1 \times 10^{-4}}$	$\mathbf{2.6 \times 10^{-7}}$	$\mathbf{3.5 \times 10^{-4}}$	105.63
RE-EM tree				
without nested structure	6.3×10^{-4}	4.0×10^{-7}	4.5×10^{-4}	5.18
RE-EM tree				
with nested structure	5.4×10^{-4}	2.9×10^{-7}	3.7×10^{-4}	8.32
Unbiased RE-EM tree				
without nested structure	6.5×10^{-4}	4.2×10^{-7}	4.6×10^{-4}	4.11
Unbiased RE-EM tree				
with nested structure	6.0×10^{-4}	3.6×10^{-7}	4.4×10^{-4}	16.89
GPBoost				
without nested structure	7.9×10^{-4}	6.2×10^{-7}	5.8×10^{-4}	**0.00**
GPBoost				
with nested structure	7.7×10^{-4}	5.9×10^{-7}	5.7×10^{-4}	0.10

According to Table (4.7), GLMM tree with nested structure outperforms other methods demonstrated in terms of all three model performance metrics as it leads to the lowest RMSE, MSE, and MAE values. On the other hand, the time required for convergence of the GLMM tree algorithm is highest compared to the other methods.

It is also better to examine how the fitted and actual values are related. To this end, the fitted versus actual values plots are constructed for each of these methods according to with and without nested data structure to check how these values are close. First, we obtain the plots for LMMs with and without nested structure, and they are illustrated in Figure 4.18. By looking at Figure 4.18, the performance of LMM can be examined. We conclude that most of the data points are not scattered homogeneously around the fitted line especially in the case of model without nested data structure. On the other hand, the performance of LMM is better when we include indices nested within the levels of subject seen on the left in Figure 4.18.



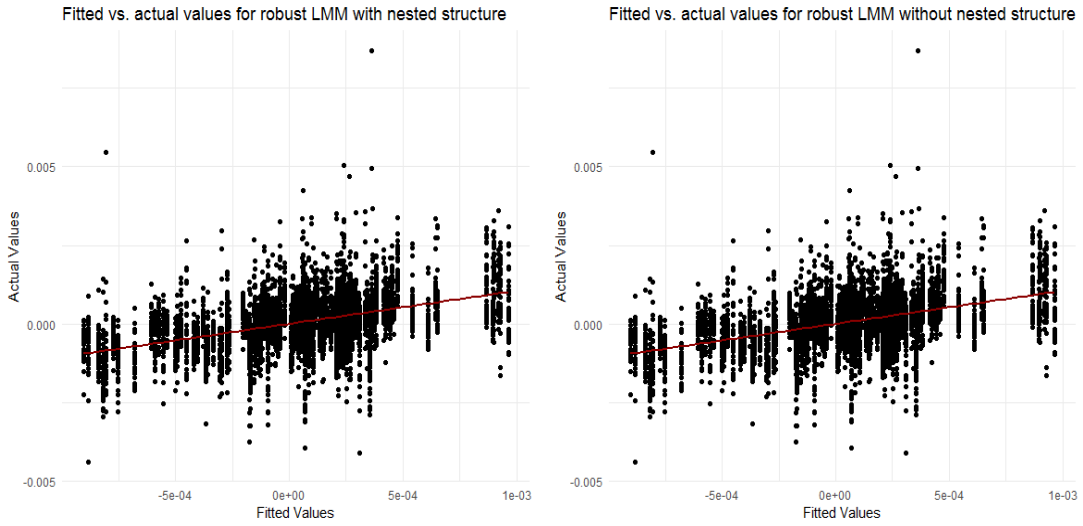
(a) LMM with nested data structure

(b) LMM without nested data structure

Figure 4.18: Fitted vs. Actual Values for LMM

As a different way of modeling relation between quantitative response and covariates, we used robust LMM as there exist observations which deviate from the rest of the data. The resulted fitted versus actual values plots are included in Figure 4.19 for with and without nested cases. It is observed that there is no apparent difference between these two plots. It was also concluded in the previous section, we obtained the same

model performance metrics. Hence, there is no improvement in this robustified model including nested structure unlike the other implemented methods.



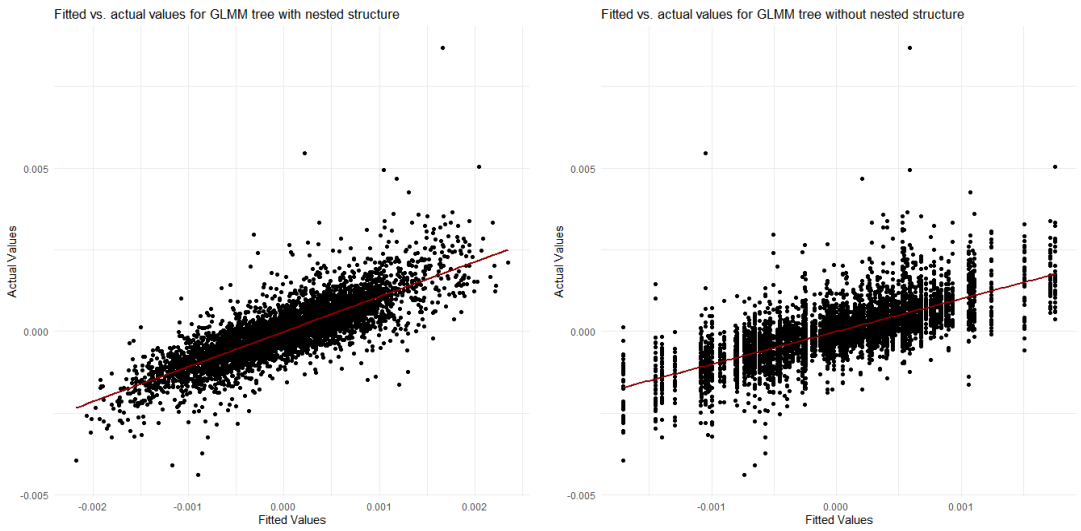
(a) Robust LMM with nested data structure (b) Robust LMM without nested data structure

Figure 4.19: Fitted vs. Actual Values for Robust LMM

The plots for the GLMM tree, which has the best predictive performance, with and without nested structure, are illustrated in Figure 4.20. The scatter plot on the left at Figure 4.20 is the output of the model with the indices nested within-subjects, while the right one is the output of the model without nested structure. The predicted values are clustered around specific values in the latter. As expected from an accurate model, the actual and predicted values are spread around a linear line on the left. We conclude that the performance of the GLMM tree is better when we include indices nested within the levels of the subject.

The plot of the RE-EM tree with nested structure (part (a) in Figure 4.21) is also promising in terms of prediction accuracy. The model performance criterion for this model is also close to the chosen model (GLMM tree with nested structure) with a much lower computation time.

The resulted fitted values versus actual response plots of Unbiased RE-EM tree for two cases are given in Figure 4.22. The similar conclusions made for the previously mentioned algorithms can be concluded for the comparison of fitting performances of

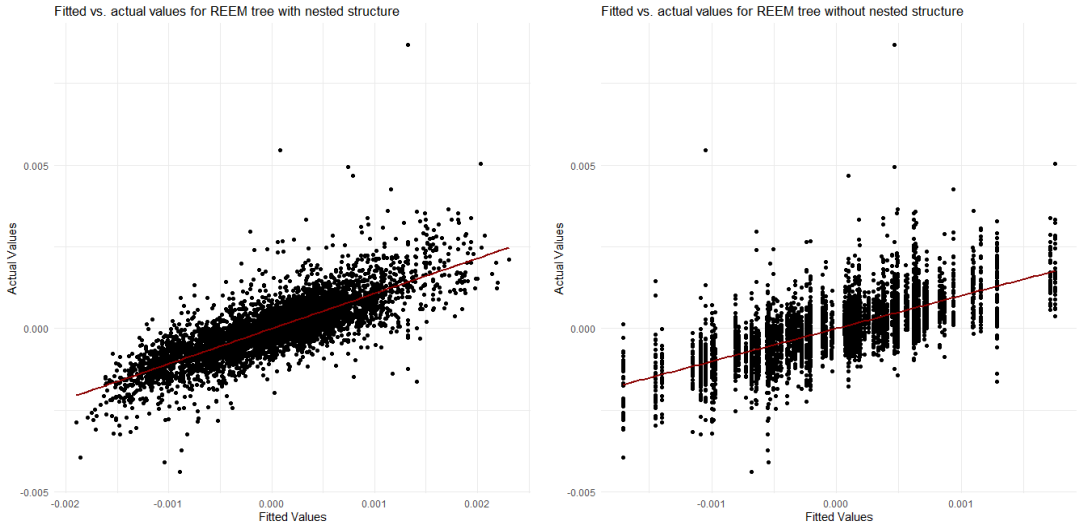


(a) GLMM tree with nested data structure

(b) GLMM tree without nested data structure

Figure 4.20: Fitted vs. Actual Values for GLMM tree

models including nested and without nested structures. Also, as there is a connection between RE-EM tree and Unbiased RE-EM tree, it is better to compare plots for these two methods. The plots for the models without nested structure obtained from RE-EM tree and its unbiased version (part (b) in Figure 4.21 and Figure 4.22, respectively) exhibits almost the same pattern through different fitted values. When the plots for the models including nested structure are compared, we can make conclusion in a similar way.

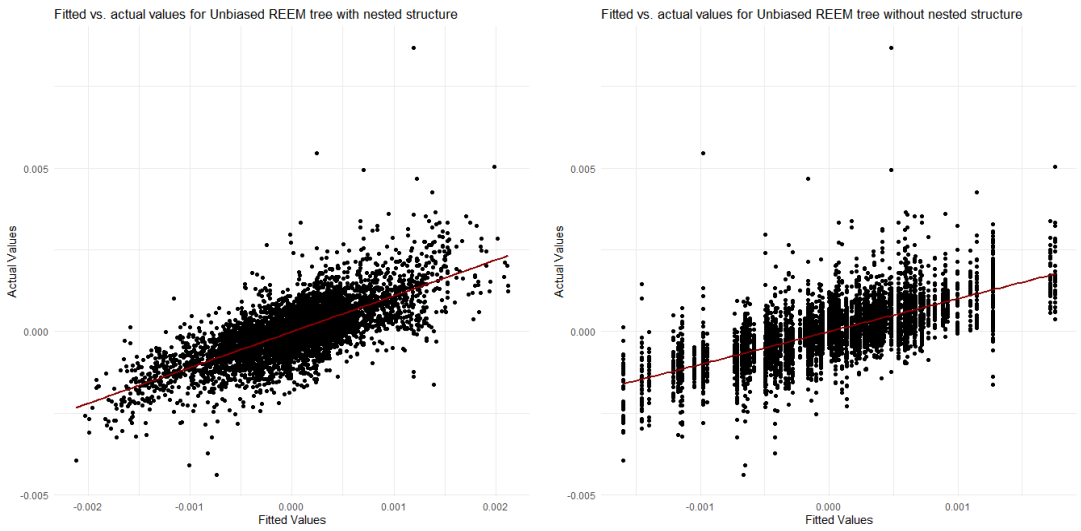


(a) RE-EM tree with nested data structure

(b) RE-EM tree without nested data structure

Figure 4.21: Fitted vs. Actual Values for RE-EM tree

As a next algorithm, the performance of GPBoost is assessed. We provide the results for the GPBoost in Figure 4.23. When part (a) of Figure 4.23 is examined for the fitting performance of the algorithm constructed with nested structure, one can conclude that the spread of the points around the fitted line are wider similar to the case obtained from LMM and shown in part (a) of Figure 4.18.

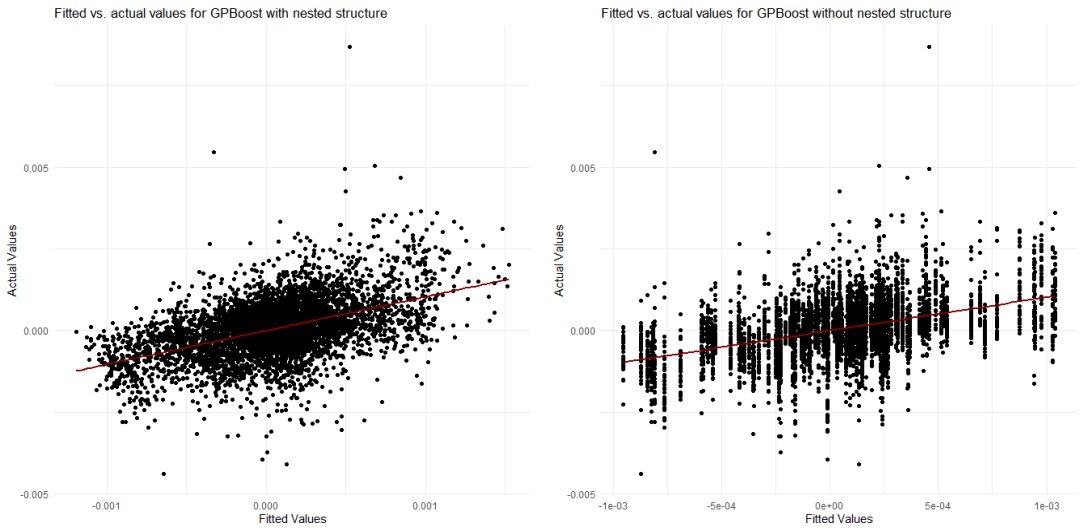


(a) Unbiased RE-EM tree with nested data structure

(b) Unbiased RE-EM tree without nested data structure

Figure 4.22: Fitted vs. Actual Values for Unbiased RE-EM tree

In addition to the plots indicating fitted versus actual values through all subjects, we partition the plots according to distinct subjects in data for the case without nested data structure (Figure 4.24).



(a) GPBoost with nested data structure

(b) GPBoost without nested data structure

Figure 4.23: Fitted vs. Actual Values for GPBoost

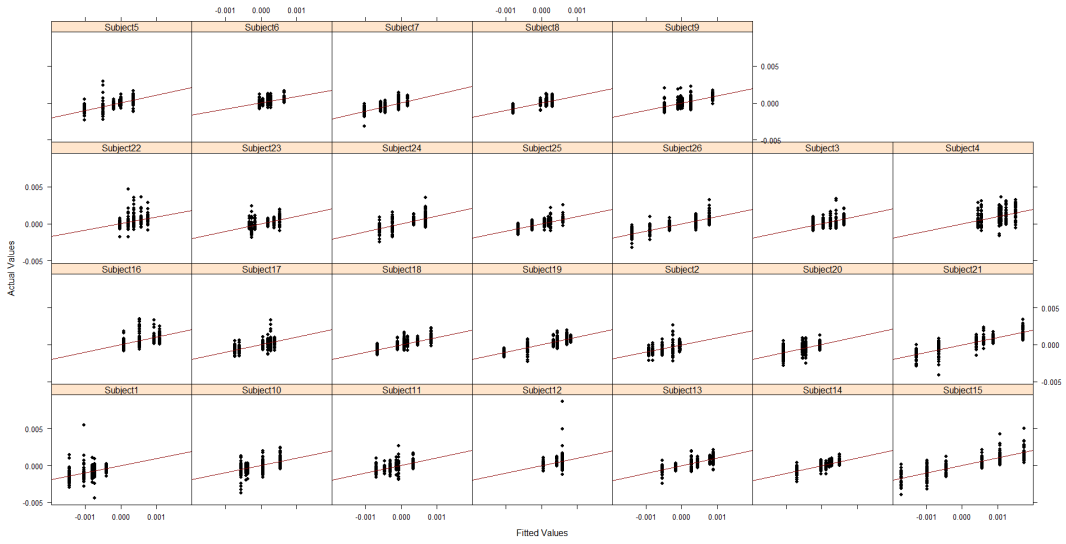


Figure 4.24: Fitted versus Actual Values Plot of GLMM tree for each Subject

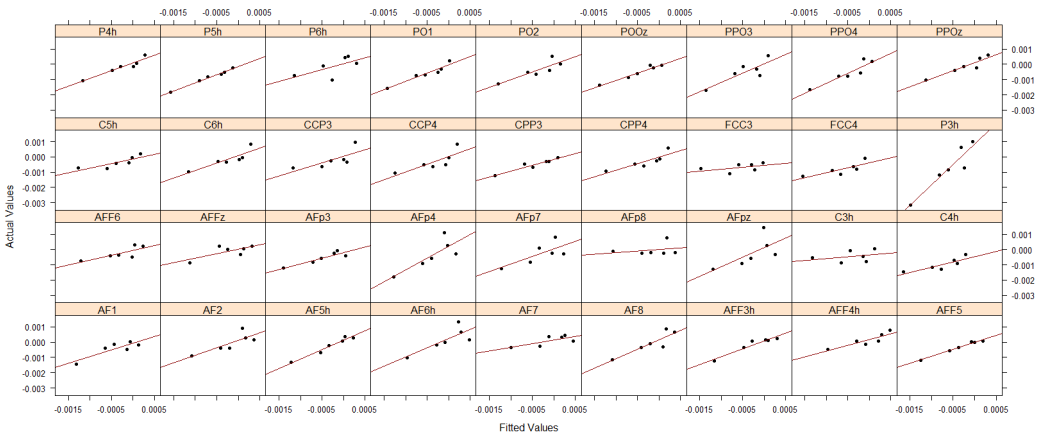


Figure 4.25: Fitted versus Actual Values Plot of GLMM tree for Indices Nested within Subject-7 (randomly selected)

Considering the nested data structure, we randomly select one subject and obtain the plot of fitted versus actual response values for each indices (i.e., the location of optode) (Figure 4.25). We added only the GLMM tree plot to illustrate the model performance according to each subject in Figure 4.24. Also, the performance of this method for Subject-7 concerning different indices is shown in Figure 4.25 in this part. When Figure 4.24 is examined to have an idea about the predictive performance of the GLMM tree, it is concluded that the observations for each subject mostly deviate from the line indicating the predictions. Hence, the method does not perform well in terms of providing close predicted values to actual response cases. On the other hand, according to Figure 4.25, most of the points scatter closer to the line for the randomly selected subject. In comparison to the model without a nested structure, it is concluded that the model performs better in making predictions closest to the observed response values when we introduce a term for indices nested within the levels of the subject. The rest of the lattice plots for the first five subjects can be found at `https://github.com/statserenay/FittedvsActualSubject1-5`.

4.5 Comparison of different algorithms

In this section, we explain the differences among the implemented methods. First, we focus on pure prediction algorithms, ML's, and LMM's pros and cons. The aim of the studies comparing the performances of different ML algorithms is under which conditions a specific algorithm significantly outperforms on a given problem instead of deciding whether an algorithm is best compared to others [88]. In this study, our aim is not to determine which algorithm is superior to others, as their results can be examined under different circumstances. Instead, we focus on the differences between the LMM and pure prediction algorithms and examine their performances based on the data from cognitive research. As in many studies comparing different ML algorithms' performance metrics and theoretical settings, we also explain the potential advantages and disadvantages of the applied methods for this data set. First, we present the main differences between the LMM and ML algorithms, then compare the algorithms built from a similar theoretical framework.

Before going into detail, it would be useful to make a little reminder about the data. The cognitive data set was not used as a tool in algorithm comparisons in this study; on the contrary, this data set was preferred intentionally because it was not analyzed with appropriate methods in the literature before. These data structures, which can reveal many unknowns in neuroscience, were analyzed very primitively, and even the dependency structure between repeated observations, for example, was ignored.

Efron [89] explained the differences between the traditional regression models with the pure prediction algorithms in terms of the assumptions that are needed, the scientific philosophy behind these methods, and their objectives. As Efron [89] expressed, the first difference can result from the assumptions which should be satisfied for the implementation of the methods. ML algorithms are non-parametric ways of data analysis; they are flexible against the assumptions required to implement LMM, a parametric approach. For instance, the adequate LMM should exhibit normally distributed residuals while there is no such requirement for applying any ML algorithms.

The second difference occurs because model parsimony is carried out by the researcher for traditional regression models as the variable selection can be conducted

through the significance testing procedures. On the other hand, the pure prediction algorithms decide on the predictors without considering the scientific fact, resulting in uncontrolled variable selection. To illustrate, we obtain the parsimonious LMM after implementing LRTs, whereas we include all the variables in the data while conducting ML algorithms.

Third, generalization of the results is possible for only LMM once the sample represents the population. However, the form of the pure prediction algorithms can change according to the data set in the particular study as these methods learn from the training data set.

We mainly mentioned the differences between LMM and ML algorithms from advantages and disadvantages perspectives. We now focus on the similarities and dissimilarities of these methods based on their theoretical background. As we have mentioned, LMM is most widely used to analyze data with a clustered structure for which the random part of LMM can account. However, the fixed effects part may not fit data well in some cases. Hence, the fixed effect terms may also be needed to take into account to partition data. GLMM tree depends on such a strategy as it utilizes MOB. That is to say, while LMM fits a parametric model on data, the MOB algorithm is used to construct a GLMM tree, first fitting a parametric model and then partitioning the data concerning splitting variables that have p-values lower than the significance level α with the help of parameter stability tests.

Because the assumption of a parametric form for the relation among the fixed effect terms and response may not be the best choice, Sela and Simonoff [72] suggest a method called RE-EM tree which can handle other types of functions for expressing such relations. In addition to being able to handle variables changing over time, observations from the same unit can be partitioned into different terminal nodes in the RE-EM tree approach [90]. However, unlike the GLMM tree approach, the RE-EM tree utilizes a mixed-effects model to handle random effects while modeling fixed effects using a CART approach. Then, Fu and Simonoff [76] proposed an unbiased version of the RE-EM tree after changing the CART algorithm into the RE-EM tree. The motivation of their method is to overcome the problems such as overfitting and variable selection bias of the CART algorithm using the conditional inference tree

instead of CART for the partition of the fixed effects. According to our findings, both the RE-EM tree and unbiased RE-EM tree provide close results for the cases of data with and without the nested structure. The reason behind such a conclusion might be that the categorical variables in our data, n-back, gender, and session, have a similar number of levels. So, adapting conditional inference instead of CART in variable selection does not make a difference in model performance metrics for the analysis of our data set.

As another ML approach that can be used to analyze longitudinal data, LongCART is considered. However, we observe that the function in `R` for the implementation of this algorithm requires baseline regressors as partitioning variables that are not changing over time. When the algorithm is constructed accordingly, we cannot reach a solution for a complete set of variables. That can be resulted from the restriction on the selection of partitioning variables as only age and gender, which may not provide any further split, are appropriate to construct this algorithm.

As one of the latest approaches for analyzing clustered or longitudinal data, GPBoost combining Gaussian process and mixed-effects regression model with boosting is also considered. Similar to the RE-EM tree approach, any functional form to represent the relation between fixed effect terms and response can be assumed in this approach. Further, unlike LMM, this method does not require any assumptions regarding the model adequacy.

In terms of goodness of fit measures to evaluate the predictive capability of algorithms, the GLMM tree, including indices nested within subject variable, outperforms all the other mentioned methods since it has the lowest RMSE, MSE, and MAE. The interpretation of the resulted tree is also not that difficult. As a result, considering the simplicity of interpretation and model performance metrics, we recommend using the GLMM tree to analyze such a cognitive data setting. On the other hand, we recommend using LMM if the aim is to make inferences on the model parameters since the GLMM tree approach does not provide inferences on the model parameters while LMM has such a property.

CHAPTER 5

CONCLUSION AND DISCUSSION

The data used for this study has a block design that exposes repeated blocks of different experimental conditions (such as n-back) besides some demographics. So, defining the change in the values, or other words, defining the active times of the brain in different conditions is crucial to locate when/where the brain is more active in neuroscience studies. Analyses carried out in this thesis mainly aimed to explore the changes in mean HbO values through different models by using data obtained via fNIRS brain imaging technology.

If one aims to check how different subjects exhibit differences in the oxygenated hemoglobin values, conducting classical regression analysis with a 25 dummy variables approach may be the first option. However, this model setting generates many parameters to be estimated. Hence, the linear regression model does not seem to consider the differences among distinct subjects efficiently. Also, another limitation of the classical linear regression approach is that it can allow one to examine the difference between each subject and Subject1 as this category is considered as a reference level decided by `lm()` function in `R`. Therefore, one of the assumptions of this kind of model is based on the independence of the observations in data. However, it is not appropriate for our cases because the data sets include measures from the same subjects. As a result, the classical linear regression model approach is not suitable for examining the changes in HbO values with the given set of explanatory variables in data. Also, one of the most used analysis types for data obtained from neuroscience studies is ANOVA which also not suitable because of the dependency structure between repeated measures of subjects.

LMM approach is taken into account, considering the challenges above. In fNIRS

analysis, the LMM estimates HbO values as a linear combination of several components such as Gender, Age, n-back, Sessions, Accuracy and MeanRT while taking the subject and channel-specific differences into account with specifying channels nested within subject as a random effect. The model selection procedure is achieved through selecting possible significant fixed effects after constructing full model and eliminating insignificant variables from the model through LRT. And all these fixed-effect components and their interactions are tested whether any of them is linked to the change in response, but the selected model was based on only the main effects. We find that 2- and 3- back conditions and MeanRT play statistically significant roles in modeling HbO values after adjusting the model with random channel effect nested within subjects. This is the first attempt in the literature, taking into account many variables in the same model, including dependency structure, randomness and nested structures. So, it naturally reduces the type-I error, which may be arisen when multiple models are used instead of a comprehensive model such as LMM.

The correlation among random effects is also explored by fitting the models under different correlation structures, and comparing them according to their AIC values. As the model with the lowest AIC indicates a better fit, we have found that the model including a positive definite matrix structure performs better. According to the LRT results, the significant main effects are n-back and MeanRT for a model considering random effects as indices nested within subjects. It is also found that the model including nested structure performs better than the model including only random effects of subject. Furthermore, the estimated coefficient of the n-back condition in this model indicates that the mean HbO concentration is 1.95×10^{-4} unit higher for 3-back condition compared to 2-back condition for a fixed MeanRT. We have also found that the mean HbO values tend to increase by 6.08×10^{-4} for each one-unit increase in MeanRT given that the n-back condition is fixed.

After deciding on the model formula, model diagnostics are checked from the residual plots. As some data points were deviating from the rest of the observations, a robust linear mixed model approach, based on down weighting the observations having larger residuals, is also conducted with the smoothed Huber function. The robust regression model is preferable as the assumptions affected by unusual observations are controlled. The resulting model gave the same estimated fixed effect coefficients

as those obtained from an ordinary linear mixed model approach which assumes equal effect of each observation on the model fit.

In addition to statistical modeling approaches, hybrid methods merging mixed models and ML appropriate to analyze longitudinal data are also constructed. The parametric models are based on some assumptions, which are mostly difficult to be verified. However, ML algorithms can be implemented to derive insight from data without a priori assumptions as opposed to these parametric models. After we implement the methods covered in the study, we compare their performances in terms of both performance metrics and advantage-disadvantage perspective. Our findings on performance metrics of different algorithms conclude that the one with the lowest resulted RMSE, MSE, and MAE is the GLMM tree, including random terms as indices nested within the Subject variable. It was also observed that different models suggest different significant variables. For instance, LMM suggested that nback and MeanRT are significant while Accuracy is the most significant variable according to RE-EM tree. On the other hand, if the researcher aims to make inferences on the model parameters, we recommend using LMM over the other mentioned methods. Mainly, we offer to use LMM and robust LMM to investigate the independent variables and demographics on HbO values. In this model, while determining the factors associated with the HbO values, the model is also controlled for the indices within-subjects and eliminated the homogeneity of variance assumption. Additionally, robust LMM provides more flexibility to reduce the effects of outlier observations. However, when the model performance metrics are compared for the robustified model with and without nested structure, there are no differences in our findings. Also, fitted versus actual values plots exhibit the same pattern for these two cases. Hence, we think that robust LMM does not take nested structure appropriately.

Further, there are more ML algorithms for longitudinal data such as Mixed-Effect Regression Tree (MERT) [91], Mixed-Effect Random Forest (MERF) [92], Random Effect Expectation Maximization Forest (REEMforest) [93], Historical Tree Ensembles for Longitudinal Data [94] in the literature. However, one limitation of our study is that there is no time variable in data unlike many experimental cognitive data sets possessing longitudinal structure. Hence, we could not use them in R programming language as the construction of these algorithms required time variable.

In the future studies, we plan to focus on the following aspects to improve this thesis study. At first sight, we plan to implement clustering algorithms for longitudinal data structure on our data set. For example, we constructed the models and different algorithms by including nested structure of all indices. In the future, we try to group these different locations after implementing clustering algorithms. Then, the model fits will be examined with these grouped indices. Moreover, we included all the data for the methods as one of our aim is to examine whether robustified mixed model performs better than the ordinary LMM. In some of the similar studies, analyses include a step for eliminating the data for channels or subjects with excessive motion artifacts by using signal to noise ratio approach based on pre-specified criteria. Our another further study will be based on comparing the results of the algorithms implemented on data obtained after removing such cases. Lastly, the performance measures of the robustified model with and without nested structure are the same as the model including random intercept for indices nested within subjects is 0. It contradicts with the other model results as we have found improvements in these metrics when the nested structure is incorporated into the model. Thus, we also plan to extend this study by incorporating different robustness structures such as heavy-tailed distributions into the model for fNIRS data structures.

REFERENCES

[1] C. S. von Bartheld, "Myths and truths about the cellular composition of the human brain: A review of influential concepts," *Journal of Chemical Neuroanatomy*, vol. 93, pp. 2–15, 2018.

[2] J. M. Murre and D. P. Sturdy, "The connectivity of the brain: multi-level quantitative analysis," *Biological Cybernetics*, vol. 73, no. 6, pp. 529–545, 1995.

[3] A. J. Golby, R. A. Poldrack, J. B. Brewer, D. Spencer, J. E. Desmond, A. P. Aron, and J. D. Gabrieli, "Material-specific lateralization in the medial temporal lobe and prefrontal cortex during memory encoding," *Brain*, vol. 124, no. 9, pp. 1841–1854, 2001.

[4] K. Hugdahl, B. R. Rund, A. Lund, A. Asbjørnsen, J. Egeland, L. Ersland, N. I. Landrø, A. Roness, K. I. Stordal, K. Sundet, *et al.*, "Brain activation measured with fmri during a mental arithmetic task in schizophrenia and major depression," *American Journal of Psychiatry*, vol. 161, no. 2, pp. 286–293, 2004.

[5] M. Hampson, N. R. Driesen, P. Skudlarski, J. C. Gore, and R. T. Constable, "Brain connectivity related to working memory performance," *Journal of Neuroscience*, vol. 26, no. 51, pp. 13338–13343, 2006.

[6] J. Hirano, A. Takamiya, B. Yamagata, S. Hotta, Y. Miyasaka, S. Pu, A. Iwanami, H. Uchida, and M. Mimura, "Frontal and temporal cortical functional recovery after electroconvulsive therapy for depression: A longitudinal functional near-infrared spectroscopy study," *Journal of Psychiatric Research*, vol. 91, pp. 26–35, 2017.

[7] W. M. Kelley, F. M. Miezin, K. B. McDermott, R. L. Buckner, M. E. Raichle, N. J. Cohen, J. M. Ollinger, E. Akbudak, T. E. Conturo, A. Z. Snyder, *et al.*, "Hemispheric specialization in human dorsal frontal cortex and medial temporal lobe for verbal and nonverbal memory encoding," *Neuron*, vol. 20, no. 5, pp. 927–936, 1998.

[8] K. Tsunoda, S. Sekimoto, and K. Itoh, "Near-infrared-spectroscopic study on processing of sounds in the brain; a comparison between native and non-native speakers of japanese," *Acta Oto-Laryngologica*, vol. 136, no. 6, pp. 568–574, 2016.

[9] M. Okamoto, M. Matsunami, H. Dan, T. Kohata, K. Kohyama, and I. Dan, "Prefrontal activity during taste encoding: an fnirs study," *Neuroimage*, vol. 31, no. 2, pp. 796–806, 2006.

[10] L. Yang, R. Ma, H. M. Zhang, W. Guan, and S. Jiang, "Driving behavior recognition using eeg data from a simulated car-following experiment," *Accident Analysis & Prevention*, vol. 116, pp. 30–40, 2018.

[11] X. Hou, Y. Liu, O. Sourina, Y. R. E. Tan, L. Wang, and W. Mueller-Wittig, "Eeg based stress monitoring," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 3110–3115, IEEE, 2015.

[12] S. M. Jaeggi, M. Buschkuehl, W. J. Perrig, and B. Meier, "The concurrent validity of the n-back task as a working memory measure," *Memory*, vol. 18, no. 4, pp. 394–412, 2010.

[13] J. M. Jansma, N. F. Ramsey, R. Coppola, and R. S. Kahn, "Specific versus nonspecific brain activity in a parametric n-back task," *Neuroimage*, vol. 12, no. 6, pp. 688–697, 2000.

[14] P.-O. Harvey, P. Fossati, J.-B. Pochon, R. Levy, G. LeBastard, S. Lehéricy, J.-F. Allilaire, and B. Dubois, "Cognitive control and brain resources in major depression: an fmri study using the n-back task," *Neuroimage*, vol. 26, no. 3, pp. 860–869, 2005.

[15] K. T. Ciesielski, P. G. Lesnik, R. L. Savoy, E. P. Grant, and S. P. Ahlfors, "Developmental neural networks in children performing a categorical n-back task," *Neuroimage*, vol. 33, no. 3, pp. 980–990, 2006.

[16] T. Kono, K. Matsuo, K. Tsunashima, K. Kasai, R. Takizawa, M. A. Rogers, H. Yamasue, T. Yano, Y. Taketani, and N. Kato, "Multiple-time replicability of near-infrared spectroscopy recording during prefrontal activation task in healthy men," *Neuroscience Research*, vol. 57, no. 4, pp. 504–512, 2007.

[17] Y. Hoshi, "Functional near-infrared optical imaging: Utility and limitations in human brain mapping," *Psychophysiology*, vol. 40, no. 4, pp. 511–520, 2003.

[18] A. Kleinschmidt, H. Obrig, M. Requardt, K.-D. Merboldt, U. Dirnagl, A. Villringer, and J. Frahm, "Simultaneous recording of cerebral blood oxygenation changes during human brain activation by magnetic resonance imaging and near-infrared spectroscopy," *Journal of Cerebral Blood Flow & Metabolism*, vol. 16, no. 5, pp. 817–826, 1996.

[19] M. L. Schroeter, M. M. Bücheler, K. Müller, K. Uludağ, H. Obrig, G. Lohmann, M. Tittgemeyer, A. Villringer, and D. Y. von Cramon, "Towards a standard analysis for functional near-infrared imaging," *NeuroImage*, vol. 21, no. 1, pp. 283–290, 2004.

[20] A. F. Abdelnour and T. Huppert, "Real-time imaging of human brain function by near-infrared spectroscopy using an adaptive general linear model," *Neuroimage*, vol. 46, no. 1, pp. 133–143, 2009.

[21] L. Chen, Q. Li, H. Song, R. Gao, J. Yang, W. Dong, and W. Dang, "Classification of schizophrenia using general linear model and support vector machine via fnirs," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 4, pp. 1151–1160, 2020.

[22] Y. Hoshi, B. H. Tsou, V. A. Billock, M. Tanosaki, Y. Iguchi, M. Shimada, T. Shinba, Y. Yamada, and I. Oda, "Spatiotemporal characteristics of hemodynamic changes in the human lateral prefrontal cortex during working memory tasks," *Neuroimage*, vol. 20, no. 3, pp. 1493–1504, 2003.

[23] K. Isobe, T. Kusaka, K. Nagano, K. Okubo, S. Yasuda, M. Kondo, S. Itoh, and S. Onishi, "Functional imaging of the brain in sedated newborn infants using near infrared topography during passive knee movement," *Neuroscience Letters*, vol. 299, no. 3, pp. 221–224, 2001.

[24] M. Okamoto, H. Dan, K. Shimizu, K. Takeo, T. Amita, I. Oda, I. Konishi, K. Sakamoto, S. Isobe, T. Suzuki, *et al.*, "Multimodal assessment of cortical activation during apple peeling by nirs and fmri," *Neuroimage*, vol. 21, no. 4, pp. 1275–1288, 2004.

[25] S. Tsujimoto, T. Yamamoto, H. Kawaguchi, H. Koizumi, and T. Sawaguchi, "Prefrontal cortical activation associated with working memory in adults and preschool children: an event-related optical topography study," *Cerebral Cortex*, vol. 14, no. 7, pp. 703–712, 2004.

[26] F. F. Jobsis, "Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters," *Science*, vol. 198, no. 4323, pp. 1264–1267, 1977.

[27] M. Ferrari and V. Quaresima, "A brief review on the history of human functional near-infrared spectroscopy (fnirs) development and fields of application," *Neuroimage*, vol. 63, no. 2, pp. 921–935, 2012.

[28] M. Ferrari, L. Mottola, and V. Quaresima, "Principles, techniques, and limitations of near infrared spectroscopy," *Canadian Journal of Applied Physiology*, vol. 29, no. 4, pp. 463–487, 2004.

[29] S. Fantini and A. Sassaroli, "Frequency-domain techniques for cerebral and functional near-infrared spectroscopy," *Frontiers in Neuroscience*, vol. 14, p. 300, 2020.

[30] F. Scholkmann, S. Kleiser, A. J. Metz, R. Zimmermann, J. M. Pavia, U. Wolf, and M. Wolf, "A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology," *Neuroimage*, vol. 85, pp. 6–27, 2014.

[31] A. Torricelli, D. Contini, A. Pifferi, M. Caffini, R. Re, L. Zucchelli, and L. Spinelli, "Time domain functional nirs imaging for human brain mapping," *Neuroimage*, vol. 85, pp. 28–50, 2014.

[32] T. Wilcox and M. Biondi, "fnirs in the developmental sciences," *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 6, no. 3, pp. 263–283, 2015.

[33] P.-H. Chou, C.-J. Huang, and C.-W. Sun, "The potential role of functional near-infrared spectroscopy as clinical biomarkers in schizophrenia," *Current Pharmaceutical Design*, vol. 26, no. 2, pp. 201–217, 2020.

[34] R. N. Aslin, M. Shukla, and L. L. Emberson, "Hemodynamic correlates of cognition in human infants," *Annual Review of Psychology*, vol. 66, pp. 349–379, 2015.

[35] S. Lloyd-Fox, A. Blasi, C. Elwell, T. Charman, D. Murphy, and M. Johnson, "Reduced neural sensitivity to social stimuli in infants at risk for autism," *Proceedings of the Royal Society B: Biological Sciences*, vol. 280, no. 1758, p. 20123026, 2013.

[36] Y. Murata, K. Sakatani, Y. Katayama, and C. Fukaya, "Increase in focal concentration of deoxyhaemoglobin during neuronal activity in cerebral ischaemic patients," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 73, no. 2, pp. 182–184, 2002.

[37] C. Herff, D. Heger, O. Fortmann, J. Hennrich, F. Putze, and T. Schultz, "Mental workload during n-back task—quantified in the prefrontal cortex using fnirs," *Frontiers in Human Neuroscience*, vol. 7, p. 935, 2014.

[38] M. P. Boisgontier and B. Cheval, "The anova to mixed model transition," *Neuroscience & Biobehavioral Reviews*, vol. 68, pp. 1004–1005, 2016.

[39] E. Aarts, M. Verhage, J. V. Veenvliet, C. V. Dolan, and S. Van Der Sluis, "A solution to dependency: using multilevel analysis to accommodate nested data," *Nature Neuroscience*, vol. 17, no. 4, pp. 491–496, 2014.

[40] A. Watanabe, K. Matsuo, N. Kato, and T. Kato, "Cerebrovascular response to cognitive tasks and hyperventilation measured by multi-channel near-infrared spectroscopy," *The Journal of Neuropsychiatry and Clinical Neurosciences*, vol. 15, no. 4, pp. 442–449, 2003.

[41] R. H. Baayen, D. J. Davidson, and D. M. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," *Journal of Memory and Language*, vol. 59, no. 4, pp. 390–412, 2008.

[42] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of Memory and Language*, vol. 68, no. 3, pp. 255–278, 2013.

[43] J. Schack, a. H. Pripp, P. Mirtaheri, H. Steen, E. Güler, and T. Gjøvaag, "Increased prefrontal cortical activation during challenging walking conditions in persons with lower limb amputation–an fnirs observational study," *Physiotherapy Theory and Practice*, pp. 1–11, 2020.

[44] Y. Koren, Y. Parmet, and S. Bar-Haim, "Treading on the unknown increases prefrontal activity: A pilot fnirs study," *Gait & Posture*, vol. 69, pp. 96–100, 2019.

[45] A. Manelis, T. J. Huppert, E. Rodgers, H. A. Swartz, and M. L. Phillips, "The role of the right prefrontal cortex in recognition of facial emotional expressions in depressed individuals: fnirs study," *Journal of Affective Disorders*, vol. 258, pp. 151–158, 2019.

[46] T. Nguyen, S. Hoehl, and P. Vrtička, "A guide to parent-child fnirs hyperscanning data processing and analysis," *Sensors*, vol. 21, no. 12, p. 4075, 2021.

[47] A. Spagnoli, J. Houwing-Duistermaat, and M. Alfó, "Mixed-effect models for longitudinal responses with different types of dropout: an application to the leiden 85-plus study," *Journal of Applied Statistics*, vol. 42, no. 9, pp. 1896–1910, 2015.

[48] R. Sitaram, H. Zhang, C. Guan, M. Thulasidas, Y. Hoshi, A. Ishikawa, K. Shimizu, and N. Birbaumer, "Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain–computer interface," *NeuroImage*, vol. 34, no. 4, pp. 1416–1427, 2007.

[49] L. M. Hirshfield, E. T. Solovey, A. Girouard, J. Kebinger, R. J. Jacob, A. Sassaroli, and S. Fantini, "Brain measurement for usability testing and adaptive interfaces: an example of uncovering syntactic workload with functional near infrared spectroscopy," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2185–2194, 2009.

[50] A. Girouard, E. T. Solovey, L. M. Hirshfield, K. Chauncey, A. Sassaroli, S. Fantini, and R. J. Jacob, "Distinguishing difficulty levels with non-invasive brain activity measurements," in *IFIP Conference on Human-Computer Interaction*, pp. 440–452, Springer, 2009.

[51] T. Gateau, G. Durantin, F. Lancelot, S. Scannella, and F. Dehais, "Real-time state estimation in a flight simulator using fnirs," *PloS One*, vol. 10, no. 3, p. e0121279, 2015.

[52] A. Plotnikov, N. Stakheika, A. De Gloria, C. Schatten, F. Bellotti, R. Berta, C. Fiorini, and F. Ansovini, "Exploiting real-time eeg analysis for assessing flow in games," in *2012 IEEE 12th International Conference on Advanced Learning Technologies*, pp. 688–689, IEEE, 2012.

[53] M. Papakostas, K. Tsiakas, T. Giannakopoulos, and F. Makedon, "Towards predicting task performance from eeg signals," in *2017 IEEE International Conference on Big Data (Big Data)*, pp. 4423–4425, IEEE, 2017.

[54] A. Villringer and B. Chance, "Non-invasive optical spectroscopy and imaging of human brain function," *Trends in Neurosciences*, vol. 20, no. 10, pp. 435–442, 1997.

[55] J. León-Carrión and U. León-Domínguez, "Functional near-infrared spectroscopy (fnirs): principles and neuroscientific applications," *Neuroimaging Methods*, pp. 48–74, 2012.

[56] D. Cheong, F. Zhang, K. Kim, A. Reid, C. Hanan, L. Ding, and H. Yuan, "Task-related systemic artifacts in functional near-infrared spectroscopy," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 948–951, IEEE, 2020.

[57] I. M. Kopton and P. Kenning, "Near-infrared spectroscopy (nirs) as a new tool for neuroeconomic research," *Frontiers in Human Neuroscience*, vol. 8, p. 549, 2014.

[58] M. Izzetoglu, S. C. Bunce, K. Izzetoglu, B. Onaral, and K. Pourrezaei, "Functional brain imaging using near-infrared technology," *IEEE Engineering in Medicine and Biology Magazine*, vol. 26, no. 4, pp. 38–46, 2007.

[59] H. Obrig and A. Villringer, "Beyond the visible—imaging the human brain with light," *Journal of Cerebral Blood Flow & Metabolism*, vol. 23, no. 1, pp. 1–18, 2003.

[60] M. Kohl, C. Nolte, H. R. Heekeren, S. Horst, U. Scholz, H. Obrig, and A. Villringer, "Determination of the wavelength dependence of the differential path-length factor from near-infrared pulse signals," *Physics in Medicine & Biology*, vol. 43, no. 6, p. 1771, 1998.

[61] P. Kyriacou, K. Budidha, and T. Y. Abay, "Optical techniques for blood and tissue oxygenation," *Encyclopedia of Biomedical Engineering*, vol. 3, pp. 461–472, 2019.

[62] N. M. Laird and J. H. Ware, "Random-effects models for longitudinal data," *Biometrics*, pp. 963–974, 1982.

[63] B. T. West, K. B. Welch, and A. T. Galecki, *Linear mixed models: a practical guide using statistical software*. Chapman and Hall/CRC, 2006.

[64] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[65] L. Wu, *Mixed effects models for complex data*. CRC press, 2009.

[66] J. C. Pinheiro and D. M. Bates, "Linear mixed-effects models: basic concepts and examples," *Mixed-effects Models in S and S-Plus*, pp. 3–56, 2000.

[67] M. Koller, "robustlmm: an r package for robust estimation of linear mixed-effects models," *Journal of Statistical Software*, vol. 75, no. 1, pp. 1–24, 2016.

[68] M. Fokkema, N. Smits, A. Zeileis, T. Hothorn, and H. Kelderman, "Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees," *Behavior Research Methods*, vol. 50, no. 5, pp. 2016–2034, 2018.

[69] A. Zeileis, T. Hothorn, and K. Hornik, "Model-based recursive partitioning," *Journal of Computational and Graphical Statistics*, vol. 17, no. 2, pp. 492–514, 2008.

[70] T. Hothorn and A. Zeileis, "partykit: A modular toolkit for recursive partytioning in r," *The Journal of Machine Learning Research*, vol. 16, no. 1, pp. 3905–3909, 2015.

[71] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *arXiv preprint arXiv:1406.5823*, 2014.

[72] R. J. Sela and J. S. Simonoff, "Re-em trees: a data mining approach for longitudinal and clustered data," *Machine Learning*, vol. 86, no. 2, pp. 169–207, 2012.

[73] T. Therneau and B. Atkinson, "rpart: Recursive partitioning. r port by brian ripley," *R Package Version 3.1*, vol. 41, 2008.

[74] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, S. Heisterkamp, B. Van Willigen, and R. Maintainer, "Package 'nlme'," *Linear and nonlinear mixed effects models, version*, vol. 3, no. 1, 2017.

[75] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*. Routledge, 2017.

[76] W. Fu and J. S. Simonoff, "Unbiased regression trees for longitudinal and clustered data," *Computational Statistics & Data Analysis*, vol. 88, pp. 53–74, 2015.

[77] T. Hothorn, K. Hornik, and A. Zeileis, "Unbiased recursive partitioning: A conditional inference framework," *Journal of Computational and Graphical statistics*, vol. 15, no. 3, pp. 651–674, 2006.

[78] M. G. Kundu and J. Harezlak, "Regression trees for longitudinal data," *arXiv preprint arXiv:1309.7733*, 2013.

[79] S. W. Raudenbush, "Comparing personal trajectories and drawing causal inferences from longitudinal data," *Annual Review of Psychology*, vol. 52, no. 1, pp. 501–525, 2001.

[80] R. E. Schapire and Y. Freund, "Boosting: Foundations and algorithms," *Kybernetes*, 2013.

[81] F. Sigrist, "Gaussian process boosting," *arXiv preprint arXiv:2004.02653*, 2020.

[82] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*, pp. 63–71, Springer, 2003.

[83] J. Shin, A. Von Lühmann, D.-W. Kim, J. Mehnert, H.-J. Hwang, and K.-R. Müller, "Simultaneous acquisition of eeg and nirs during cognitive tasks for an open access dataset," *Scientific Data*, vol. 5, no. 1, pp. 1–16, 2018.

[84] M. D'Esposito, E. Zarahn, and G. K. Aguirre, "Event-related functional mri: implications for cognitive psychology.," *Psychological Bulletin*, vol. 125, no. 1, p. 155, 1999.

[85] E. M. Peck, D. Afergan, B. F. Yuksel, F. Lalooses, and R. J. Jacob, "Using fnirs to measure mental workload in the real world," in *Advances in Physiological Computing*, pp. 117–139, Springer, 2014.

[86] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021.

[87] F. Sigrist, *gpboost: Combining Tree-Boosting with Gaussian Process and Mixed Effects Models*, 2022. R package version 0.7.8.

[88] S. Borah, S. K. Mishra, B. K. Mishra, V. E. Balas, and Z. Polkowski, *Advances in Data Science and Management: Proceedings of ICDSM 2021*. Springer, 2022.

[89] B. Efron, "Prediction, estimation, and attribution," *International Statistical Review*, vol. 88, pp. S28–S59, 2020.

[90] Y. Xu, A. Zafirov, R. M. Alvarez, D. Kojis, M. Tan, and C. M. Ramirez, "Freetree: a tree-based approach for high dimensional longitudinal data with correlated features," *arXiv preprint arXiv:2006.09693*, 2020.

[91] A. Hajjem, F. Bellavance, and D. Larocque, "Mixed effects regression trees for clustered data," *Statistics & Probability Letters*, vol. 81, no. 4, pp. 451–459, 2011.

[92] A. Hajjem, F. Bellavance, and D. Larocque, "Mixed-effects random forest for clustered data," *Journal of Statistical Computation and Simulation*, vol. 84, no. 6, pp. 1313–1328, 2014.

[93] L. Capitaine, R. Genuer, and R. Thiébaut, "Random forests for high-dimensional longitudinal data," *Statistical Methods in Medical Research*, vol. 30, no. 1, pp. 166–184, 2021.

[94] J. Sexton, *htree: Historical Tree Ensembles for Longitudinal Data*, 2018. R package version 2.0.0.

Appendix A

ACCURACY MEASURES

Table A.1: Accuracy measures for each subject

Subject ID	Session	N-back	The number of missed responses	The number of wrong responses	Accuracy (%)
1	S1	2	0	1	98.15
		3	0	1	98.04
	S2	2	0	0	100
		3	0	0	100
	S3	2	0	0	100
		3	0	0	100
2	S1	2	0	3	94.44
		3	0	11	78.43
	S2	2	0	0	100
		3	0	8	84.32
	S3	2	0	5	90.74
		3	0	10	80.39
3	S1	2	0	9	83.33
		3	0	10	80.39
	S2	2	0	9	83.33
		3	0	12	76.47
	S3	2	0	11	79.63
		3	0	10	80.39

Continued on next page

Table A.1 – *Continued from previous page*

Subject ID	Session	n-Back	The number of missed responses	The number of wrong responses	Accuracy (%)
4	S1	2	0	1	98.15
		3	0	3	94.12
	S2	2	0	1	98.04
		3	0	2	96.08
	S3	2	0	0	100
		3	0	0	100
5	S1	2	0	11	79.63
		3	0	10	80.39
	S2	2	0	7	86.71
		3	0	12	76.47
	S3	2	1	9	84.86
		3	0	9	82.35
6	S1	2	0	12	77.78
		3	0	10	80.39
	S2	2	0	4	92.59
		3	0	17	67.54
	S3	2	0	11	79.63
		3	1	10	79.66
7	S1	2	1	2	96.19
		3	5	5	89.13
	S2	2	0	0	100
		3	0	1	98.04
	S3	2	2	0	96.30
		3	0	0	100
8	S1	2	0	6	88.89
		3	3	15	64.71
	S2	2	0	5	90.74
		3	1	13	72.55

Continued on next page

Table A.1 – *Continued from previous page*

Subject ID	Session	n-Back	The number of missed responses	The number of wrong responses	Accuracy (%)
	S3	2	1	3	92.59
		3	0	13	74.51
9	S1	2	1	4	90.74
		3	1	10	78.43
	S2	2	0	2	96.30
		3	0	9	84.31
	S3	2	0	4	92.59
		3	0	7	86.27
10	S1	2	2	2	92.59
		3	1	11	76.17
	S2	2	0	1	98.15
		3	3	2	90.20
	S3	2	0	1	98.15
		3	0	0	100
11	S1	2	0	2	96.30
		3	0	9	82.35
	S2	2	1	1	96.30
		3	0	6	88.24
	S3	2	0	0	100
		3	1	6	86.03
12	S1	2	0	10	81.48
		3	0	27	52.94
	S2	2	0	26	51.85
		3	1	27	52.94
	S3	2	0	16	70.37
		3	1	30	58.82
13	S1	2	1	7	85.19
		3	2	10	76.47

Continued on next page

Table A.1 – *Continued from previous page*

Subject ID	Session	n-Back	The number of missed responses	The number of wrong responses	Accuracy (%)
	S2	2	0	4	92.48
		3	0	8	84.31
	S3	2	0	4	92.59
		3	0	4	92.16
14	S1	2	0	7	87.04
		3	0	8	84.31
	S2	2	0	3	94.44
		3	0	10	80.39
	S3	2	0	0	100
		3	1	9	80.39
15	S1	2	1	0	98.15
		3	3	3	88.24
	S2	2	0	1	98.15
		3	3	3	88.24
	S3	2	0	5	90.63
		3	0	4	92.16
16	S1	2	0	5	90.74
		3	0	11	78.43
	S2	2	1	2	94.44
		3	0	4	92.16
	S3	2	0	3	94.44
		3	0	10	80.39
17	S1	2	1	3	92.59
		3	0	6	88.24
	S2	2	1	3	92.59
		3	0	3	94.12
	S3	2	0	1	98.15
		3	1	3	92.16

Continued on next page

Table A.1 – *Continued from previous page*

Subject ID	Session	n-Back	The number of missed responses	The number of wrong responses	Accuracy (%)
18	S1	2	0	6	88.89
		3	1	8	82.35
	S2	2	1	4	90.74
		3	6	9	70.59
	S3	2	0	0	100
		3	1	7	84.31
19	S1	2	2	10	77.78
		3	1	13	72.55
	S2	2	2	11	75.93
		3	0	8	84.31
	S3	2	0	6	88.89
		3	1	17	64.71
20	S1	2	0	8	85.19
		3	0	14	72.55
	S2	2	0	2	96.30
		3	0	5	90.20
	S3	2	0	0	100
		3	1	4	90.20
21	S1	2	0	1	98.15
		3	0	6	88.24
	S2	2	0	1	98.04
		3	0	8	84.31
	S3	2	0	0	100
		3	0	4	92.16
22	S1	2	0	1	98.15
		3	0	6	88.24
	S2	2	0	3	94.44
		3	0	5	90.20

Continued on next page

Table A.1 – *Continued from previous page*

Subject ID	Session	n-Back	The number of missed responses	The number of wrong responses	Accuracy (%)
	S3	2	0	0	100
		3	0	7	86.27
23	S1	2	0	6	88.89
		3	0	13	74.51
	S2	2	0	7	87.04
		3	0	12	76.47
	S3	2	0	7	87.04
		3	0	11	78.43
24	S1	2	0	6	88.89
		3	1	12	74.51
	S2	2	0	2	96.30
		3	0	6	88.24
	S3	2	0	4	92.59
		3	2	8	80.39
25	S1	2	0	6	88.89
		3	0	9	82.35
	S2	2	0	4	92.59
		3	0	6	88.24
	S3	2	0	1	98.15
		3	0	10	80.39
26	S1	2	0	2	96.30
		3	1	11	76.47
	S2	2	0	0	100
		3	0	2	96.08
	S3	2	0	0	100
		3	0	0	100

Appendix B

CORRELATION STRUCTURES

Table B.1: AIC values with different correlation structures for RE-EM tree without nested structure

Correlation Structure	Argument	AIC
No within-group correlation	Default	-65555.57
Autoregressive process of order 1 (AR(1))	**corAR1()**	$\mathbf{-65875.02}$
Continuous AR(1)	corCAR1()	-65861.01
A compound-symmetric matrix	corCompSymm()	-65592.47

Table B.2: AIC values with different correlation structures for RE-EM tree with nested structure

Correlation Structure	Argument	AIC
No within-group correlation	Default	-65956.89
AR(1)	corAR1()	-66017.18
Continuous AR(1)	corCAR1()	-66002.16
A compound-symmetric matrix	**corCompSymm()**	$\mathbf{-66065.96}$

Table B.3: Log-likelihood values with different correlation structures for Unbiased RE-EM tree without nested structure

Correlation Structure	Argument	Log-Likelihood
No within-group correlation	**Default**	**32771.34**
AR(1)	corAR1()	32669.06
Continuous AR(1)	corCAR1()	32669.06
A compound-symmetric matrix	**corCompSymm()**	**32771.34**

Table B.4: Log-likelihood values with different correlation structures for Unbiased RE-EM tree with nested structure

Correlation Structure	Argument	Log-Likelihood
No within-group correlation	Default	32663.85
AR(1)	**corAR1()**	**32932.03**
Continuous AR(1)	**corCAR1()**	**32932.03**
A compound-symmetric matrix	corCompSymm()	32663.85

Table B.5: Performance metrics for RE-EM tree with default and selected correlation structures

RE-EM tree	RMSE	MSE	MAE
Without nested structure			
Default	6.3×10^{-4}	4.0×10^{-7}	4.5×10^{-4}
AR(1)	6.4×10^{-4}	4.1×10^{-7}	4.5×10^{-4}
With nested structure			
Default	5.4×10^{-4}	2.9×10^{-7}	3.7×10^{-4}
Compound Symmetry	5.4×10^{-4}	2.9×10^{-7}	3.8×10^{-4}

Table B.6: Performance metrics for Unbiased RE-EM tree with default and selected correlation structures

Unbiased RE-EM tree	RMSE	MSE	MAE
Without nested structure			
Default	6.5×10^{-4}	4.2×10^{-7}	4.6×10^{-4}
Compound Symmetry	6.5×10^{-4}	4.2×10^{-7}	4.6×10^{-4}
With nested structure			
Default	6.0×10^{-4}	3.6×10^{-7}	4.4×10^{-4}
AR(1)	5.7×10^{-4}	3.3×10^{-7}	4.1×10^{-4}