
Masters Theses

Student Theses and Dissertations

Summer 2020

Quantifying effects of sleep deprivation on cognitive performance

Quang Nghia Le

Follow this and additional works at: https://scholarsmine.mst.edu/masters_theses



Part of the [Statistics and Probability Commons](#)

Department:

Recommended Citation

Le, Quang Nghia, "Quantifying effects of sleep deprivation on cognitive performance" (2020). *Masters Theses*. 8004.

https://scholarsmine.mst.edu/masters_theses/8004

This thesis is brought to you by Scholars' Mine, a service of the Missouri S&T Library and Learning Resources. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact scholarsmine@mst.edu.

QUANTIFYING EFFECTS OF SLEEP DEPRIVATION ON COGNITIVE
PERFORMANCE

by

QUANG NGHIA LE

A THESIS

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

in

APPLIED MATHEMATICS WITH STATISTICS EMPHASIS

2020

Approved by

Dr. VA Samaranayake, Advisor

Dr. Matthew S. Thimgan

Dr. Xuerong Wen

Copyright 2020
QUANG NGHIA LE
All Rights Reserved

ABSTRACT

The most commonly used metric for evaluating alertness and vigilance is the Psychomotor Vigilance Test (PVT), previous studies have indicated that alertness and vigilance can be affected by the lack of sleep as a function of sleep loss. This study explores methods to predict median psychomotor vigilance reaction times. The data used in this study comes from a series of tests and surveys conducted on volunteer students. The data set contains many potential predictors of PVT and one aspect of the study was to identify variables that are useful in prediction. The performances of various prediction methods that allow for feature selection were evaluated. Prediction errors were estimated by using ten-fold validation method and root mean squared error was employed to compare the methods.

Results show that the linear model with LASSO feature selection provide the best predictions of psychomotor vigilance test median reaction time in this context. Moreover, we were able to identify subsets of predictors that lead to reduced prediction error and are useful for extracting biological insights. The linear mixed model and canonical correlation analysis provided information on what factors affect vigilance attention, and what cognitive functions are affected by sleep quality.

ACKNOWLEDGMENTS

First and foremost, I would like to thank Dr. Samaranayake. I would not be here in this program without him nor would this work be possible without his knowledgeable academic insight and guidance.

Dr. Samaranayake, I am so grateful for you. It has been a privilege to be your student and to work under your supervision. Thank you for your patience, kindness, understanding and support. This project has been very difficult and it would be much harder without your guidance. One of the reasons I have made it this far is because of your belief in me. I genuinely would not have made it without you.

Dr. Thimgan, your knowledge about biology has been crucial and the most important factor to help me complete this project. Without the data collected by you, this project and thesis would not have existed at all. During the time when I worked with you, I realized a very obvious but often neglected fact, statistics is just a tool, you have to know what you are working with.

I would like to give my gratitude to other members of this project, who had handled data collection, without them this project would be not possible.

To Joshua Maxwell, your invaluable advice came when I needed it the most. Although you didn't teach any class, I have learned so much from you, both academically and personally. To Austin, without your advice of using LaTeX to do my thesis, I would have had a very difficult time trying to figure out how to type and manage complex equations in Microsoft Word.

And lastly, to my loving family. You supported me both financially and emotionally without any second thought. Many thanks from the bottom of my very full heart. I truly am so grateful for you all.

TABLE OF CONTENTS

	Page
ABSTRACT	iii
ACKNOWLEDGMENTS	iv
LIST OF ILLUSTRATIONS	vii
LIST OF TABLES	viii
NOMENCLATURE	ix
 SECTION	
1. INTRODUCTION	1
2. DESCRIPTION OF THE DATA SET	3
2.1. PSYCHO-MOTOR VIGILANCE TASK	3
2.2. EPWORTH SLEEPINESS SCALE	4
2.3. DIGIT SYMBOL SUBSTITUTION TEST	5
2.4. BALLOON ANALOGUE RISK TASK	5
2.5. N-BACK TASK	6
2.6. EMOTION RECOGNITION TASK	6
3. STATISTICAL ANALYSES	7
3.1. MODELLING METHODS	7
3.1.1. Linear Model	7
3.1.2. Linear Mixed Model	9
3.1.3. Canonical Correlation Analysis	11

3.1.4. Support Vector Regression	12
3.1.5. Gradient Boosting Tree.....	13
3.2. FEATURE SELECTION METHODS	15
3.3. ERROR ESTIMATION	16
3.4. COMPARISON OF PREDICTION MODELS	17
4. RESULTS	18
4.1. PREDICTION USING PARAMETRIC MODELS	18
4.2. PREDICTION USING NON-PARAMETRIC MODELS	19
4.3. LINEAR MIXED MODEL	20
4.4. CANONICAL CORRELATION ANALYSIS	22
5. CONCLUSIONS	25
APPENDICES	
A. MODEL FIT DIAGNOSTICS AND PREDICTION INTERVAL	27
B. FEATURE SELECTION AND MULTI-COLLINEARITY DIAGNOSTIC	31
REFERENCES	34
VITA	36

LIST OF ILLUSTRATIONS

Figure	Page
4.1. Visualization of the first two canonical dimensions.	24

LIST OF TABLES

Table	Page
4.1. Prediction accuracy of the models.	18
4.2. Summary result for linear regression models.	19
4.3. Hyper-parameter for gradient boosted tree model.	20
4.4. Hyper-parameter for support vector machine model.	20
4.5. Confidence intervals for fixed effects.	21
4.6. Confidence intervals for random effect.	21
4.7. The X set of variables.	23
4.8. The Y set of variables.	23
4.9. Results of canonical correlation analysis.	23
4.10. Standardized canonical variables for the X set.	24
4.11. Standardized canonical variables for the Y set.	24

NOMENCLATURE

Symbol	Description
Acronym	
AIC	Akaike information criterion
BART	Balloon Analogue Risk Task
BIC	Bayesian information criterion
DSST	Digit Symbol Substitution Test
ERT	Emotion Recognition Task
ESS	The Epworth Sleepiness Scale
LASSO	Least absolute shrinkage and selection operator
MCAR	Missing completely at random
MNAR	Missing not at random
OLS	Ordinary least squares
PVT	Psycho-motor Vigilance Task
RMSE	Root mean squared error
SVM	Support vector machine
SVR	Support vector regression
VIF	Variance Inflation Factor

1. INTRODUCTION

In modern society, with a lot of stress and responsibilities, loss of sleep is a common problem among many individuals. Most adults admit that they have faced sleep loss at least once in their life. Sleep deprivation is when a person has less sleep than they need to be able to feel fully awake and alert. The amount of sleep a person needs to have is, of course, subjective and it varies significantly between people. Some people seem to suffer less to the effects of lack of sleep, while others, like children and young adults, are more vulnerable. Although occasional sleeplessness generally has mild effects, it is well-established that long term sleep loss leads to subjective feelings of increased fatigue and decreased motivation and alertness (Browne *et al.*, 1999). Chronic or ongoing insomnia can lead to emotional changes, degradation of mental health, weight gain, and poor job or academic performance.

Long term sleep loss sharply decreases human alertness and ability to maintain attention. Lack of sleep, not surprisingly, leads to an increase in the risk of vehicle and machinery-related accidents (Pérez-Chada *et al.*, 2005), as well as increases in on-the-job errors in many occupations, from train operators to pilots and medical professionals who are required to work long shifts with little opportunity for rest. When the body doesn't receive sufficient amount of quality sleep, it does not release the hormones necessary to regulate growth and appetite. Instead, it forms an overabundance of stress chemicals like cortisol (Bassett *et al.*, 2015).

Quantifying effects of sleep deprivation, both short term and long term, becomes an important task as it is the first step to understanding sleep deprivation and developing effective interventions. This project aims to study the effects of sleep deprivation and other cognitive behavior on human attention. Over decades, the metric that has been used primarily in the study of vigilant attention with the presence of sleep deprivation is the psychomotor vigilance test (PVT). This simple but extremely informative measure has

been utilized as a numeric marker of the deficit in attention in subjects (Lee *et al.*, 2010), (Thomann *et al.*, 2010). Its high sensitivity to sleep deprivation combined with its reliability and well studied characteristics (Basner and Rubinstein, 2012) are reasons why we use it to measure vigilant attention of subjects in this project. The first goal of this project is to find out if we can predict vigilant attention of a subject with reasonable accuracy. The second goal is to figure out what factors affect vigilant attention of a subject, and the last goal is to study the effect of sleep deprivation on other cognitive functions.

2. DESCRIPTION OF THE DATA SET

In this section, we briefly go through descriptions of the data set and predictors utilized in the predictive models. The data set consists of many variables, and each observation is a result of several cognitive tests of an individual at a given time. Each individual is tested on several occasions over time yielding many observations per test subject.

The experiment was conducted by Dr. Thimgan and his collaborators. The test subjects are Missouri University of Science and Technology students, who volunteered for this study. Two sessions were conducted per week for 15 weeks straight (1 semester), with one in the morning and one in the afternoon. Hence potentially each participant would have thirty observations. During sessions each participants had to go through multiple tests to measure their cognitive performance. In the following sections we will discuss cognitive performance tests that were used in the experiment.

2.1. PSYCHO-MOTOR VIGILANCE TASK

The Psycho-motor Vigilance Task (PVT) (Drummond *et al.*, 2005) is a test designed to evaluate our ability of sustained-attention. It measures the speed we respond to a specific stimulus. The tested subjects have to press a button as soon as the light appears, the light turns on randomly every few seconds for 5-10 minutes. The goal of this task is not to evaluate the subject's reaction time but to see how many times the button is not pressed when the light is on so we can measure sustained attention; this gives a numerical metric of how sleepy the subject is by counting the number of lapses in attention of the tested subject.

2.2. EPWORTH SLEEPINESS SCALE

The Epworth Sleepiness Scale (or ESS) (Johns, 1991) asks the subject to rate on a 4-point scale their chances of falling asleep while doing eight widely different activities.

The activities included are:

- Sitting and reading
- Watching TV
- Sitting inactive in a public place, such as a meeting or theatre
- Riding as a passenger in a car for an hour without a break
- Lying down to rest in the afternoon when circumstances permit
- Sitting and talking to someone
- Sitting quietly after a lunch without alcohol
- Sitting in a car stopped for a few minutes in traffic.

These activities vary in their somnificity, which describes how different postures and activities impact your readiness to fall asleep. Each of the activities listed has an assigned score from 0 to 3 that indicates how likely a person is to fall asleep during the activity:

- 0 = Would never doze
- 1 = Slight chance of dozing
- 2 = Moderate chance of dozing
- 3 = High chance of dozing.

The total score can range from 0 to 24, which is the sum of all individual ESS scores for each activity. A higher score is associated with increased sleepiness.

2.3. DIGIT SYMBOL SUBSTITUTION TEST

Digit Symbol Substitution Test (DSST) (Jaeger, 2018) is an assessment used to survey psychological functioning. It was at first a piece of the Wechsler Adult Intelligence Test (WAIS), a notable test that quantifies a person's IQ (Lafont *et al.*, 2010).

Digit Symbol Substitution Test includes a key comprising of the numbers 1-9, each combined with an exceptional, simple to-draw image, for example, a "V", "+" or ">". Beneath the key are a progression of the numbers 1-9 in irregular order and rehashed a few times. The test taker is then permitted 90 or 120 seconds (contingent upon the test form) to fill in the comparing image for each number. This task requires the person to outwardly examine the appropriate response key given at any moment of the test and afterward compose the right image by each number.

Digit Symbol Substitution Test assesses preparing speed, working memory, visuospatial handling, and attention. Specifically, this test, by all accounts, is sensitive to deviations in individuals' perception from the normal. In contrast, other tests may be not able to separate between people with ordinary comprehension and those with a gentle psychological disability.

2.4. BALLOON ANALOGUE RISK TASK

The Balloon Analogue Risk Task (BART) is a task designed to assess risk-taking behavior. In this task, the subject is shown a balloon which the subject can pump. Each pump comes with two options:

First option: Pump the balloon up, each pump causing the balloon to inflate and money to be added, until a threshold at which the balloon explodes. If the balloon explodes, the subject will lose all the money.

Second option: Stop pumping and cash-out, the subject can collect the money earned for that trail.

Thus, each pump comes with a risk, but also a potential reward. Participants do not know the threshold in which the balloon explodes. The absence of this information allows for testing both participants' initial responses to the task and changes in responding as they gain experience with how the game is conducted. Adjusted average number of pumps on unexploded balloons is the score we record, with higher scores mean greater risk-taking tendency.

2.5. N-BACK TASK

The test shows the subject a sequence of stimuli and then asks the subject to identify whether the current stimulus matches the one from N steps earlier in the sequence. The load factor N can be changed to adjust the difficulty of the test. In this test, there is only one item that appears in different positions during each turn.

The n-back test catches the dynamic piece of working memory (Borkowska *et al.*, 2009). At the point when N rises to at least 2, it isn't sufficient to remember the portrayal of the latest item. The working memory support likewise should be refreshed consistently to monitor with what the present item must be contrasted against. To complete the assignment, the subject needs to both keep up and control data in working memory.

2.6. EMOTION RECOGNITION TASK

Emotion Recognition Task (Montagne *et al.*, 2007) measures the capacity to recognize six fundamental facial feelings along a continuum of expression magnitude.

Computer-transformed pictures obtained from the exaggerated facial features of real people, each indicating a particular feeling, are shown on the screen, one at a time. Each face is shown for 200ms and quickly concealed afterward to prevent residual processing of the image. The subject must choose which feeling the face showed from 6 choices (bitterness, satisfaction, dread, outrage, appall, or shock).

3. STATISTICAL ANALYSES

Biological data require special considerations when conducting statistical analyses because of the high variability that is present in such data. In the study that generated the data analyzed in this investigation, subjects answered many survey questions, resulting in a large number of variables. In addition, the time, resource and other limitations limited the number of observations made during the study. If we build a prediction model without performing feature selection or dimension reduction, our models will suffer from the "curse of dimensionality" and our out of sample prediction would not be accurate.

In this project, we had to deal with two types of missing data. Subjects of the study were not compulsorily required to participate in every session, hence resulted in missing data. We assume that this type of missing data is missing completely at random (MCAR). We handled this kind of missing data by dropping the entire row that has missing data, or if one variable has so many missing values, we will drop that variable. The second type of missing data we had to deal with is missing not at random, also as known as non-ignorable missing data (or MNAR). MNAR is usually a hard problem to solve, and how we dealt with it is discussed in detail in this section.

3.1. MODELLING METHODS

In this section, we discuss statistical models that we investigated as potential prediction models. Both parametric and non-parametric models were employed to achieve this goal.

3.1.1. Linear Model. The first model we considered is the linear regression model to describe the data. Given a sample of N observations $i = 1, \dots, N$. Every observation i follows the equation of the linear regression model (Rosenfeld, 2016):

$$y_i = x_i' \beta + \epsilon_i. \quad (3.1)$$

where $i = 1, \dots, N$ and β is a $(K + 1)$ -dimensional column vector of parameters, x_i' is a $(K+1)$ -dimensional row vector, and ϵ_i is the error or residual. Equation (3,1) can also be expressed in matrix notation:

$$Y = X\beta + \epsilon_i, \quad (3.2)$$

or as

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1K} \\ 1 & x_{21} & \dots & x_{2K} \\ & x_{31} & \dots & x_{3K} \\ \vdots & & & \\ 1 & x_{N1} & \dots & x_{NK} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_N \end{bmatrix}, \quad (3.3)$$

given observed values $y_1, y_2, y_3, \dots, y_N$ and features x_{ij} for $i = 1, \dots, N$ and $j = 1, \dots, K$. Ordinary least squares (OLS) minimizes the squared distances between the observed y and the predicted y :

$$S(\beta) = \sum_{i=1}^N (y_i - x_i' \beta)^2 = (Y - X\beta)'(Y - X\beta), \quad (3.4)$$

and the OLS estimator of β would be:

$$\hat{\beta} = (X'X)^{-1} X'Y. \quad (3.5)$$

Now we discuss the assumptions made when carrying out OLS linear regression model. The OLS linear regression model consists of a set of assumptions. Those assumptions are (Diego, 2011):

1. Full rank of the X matrix: Full rank means that there is no exact linear relationship between independent variables.

2. Exogeneity of the independent variables: This means each of the disturbance terms (which are also called residuals) is assumed to have zero expected values conditioned on the X values.:

$$E[\epsilon|X] = 0. \quad (3.6)$$

3. Homoscedasticity: This refers to the variance of the residuals, That is the residuals should have finite and constant variance:

$$Var[\epsilon|X] = \sigma^2. \quad (3.7)$$

4. Normally distributed residuals: This assumption, combine with assumption 3 and 4 can be expressed as the following:

$$\epsilon|X \sim N[0, \sigma^2]. \quad (3.8)$$

We fit the OLS regression model using function `lm()` in R. We used studentised error and adjusted R^2 to check the fit of the model. If the model is a good fit, the residuals should behave like a random sample from a normal distribution with mean zero and variance σ^2 without showing systematic patterns in the residual plots.

3.1.2. Linear Mixed Model. When we deal with repeated measures from several units (like the data we are dealing with), the fixed effects linear model is generally not enough to model the relationship between predictors and response variables, This is when the linear mixed model becomes useful.

Generally, in the linear model, the distributional assumption about the error term is too restrictive. The mixed model (SAS documents, 1999) extend this by allowing both correlation and heterogeneous variances of the covariance matrix of the error term, while still assuming normality. The mixed model is written as:

$$y = X\beta + Z\gamma + \epsilon. \quad (3.9)$$

We can clearly see that it is nearly the same as that of linear model. However, there is one new term, that is random-effects parameter γ and the matrix Z . The matrix Z can contain either continuous or dummy variables. We call this "mixed model" because it contains both fixed-effects β and random-effects γ . The assumption of γ and ϵ is that:

$$E \begin{bmatrix} \gamma \\ \epsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad (3.10)$$

$$Var \begin{bmatrix} \gamma \\ \epsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}. \quad (3.11)$$

The variance of y is $V = ZGZ' + R$. Now we discuss the estimation of G and R . In a mixed model, we can no longer use the least-squares method. Instead, we use Generalized least squares (GLS), minimizing the following function:

$$(y - X\beta)'V^{-1}(y - X\beta). \quad (3.12)$$

We need G and R to model V and we need V to minimize or loss function to find G and R . One approach is to insert a reasonable estimate for V into the loss function, from which we find a reasonable estimate of G and R . One way to estimate G and R when we have the estimate of V is using maximum likelihood, the corresponding log-likelihood function is as follows:

$$l(G, R) = \frac{-1}{2} \log|V| - \frac{1}{2} r'V^{-1}r - \frac{n}{2} \log(2\pi), \quad (3.13)$$

where $r = y - X(X'V^{-1}X)^{-1}X'V^{-1}y$ and p is the rank of X . When we have the estimate of G and R , we can estimate the $\hat{\beta}$ and $\hat{\gamma}$ as

$$\hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y, \quad (3.14)$$

$$\hat{\gamma} = \hat{G}Z'\hat{V}^{-1}(y - X\hat{\beta}). \quad (3.15)$$

3.1.3. Canonical Correlation Analysis. One goal of the project is to study the effect of sleep deprivation on the cognitive performance of the subject. Statistically speaking, this is the problem of studying the correlation between two sets of variables, with the first set containing information about subjects sleeping characteristics and the second set containing information about subjects cognitive performance. One way to address the problem is to use canonical correlation analysis; which is a multivariate statistical method to explore the relationship between two sets of variables. Complex multivariate structures of the data are better studied when we can compress it into low-dimensional projections. Canonical correlation analysis aims to find dimensions that maximize the correlation between the two data sets. This method allows us to describe the relationships with a lower number of dimensions while preserving the main facets of the relationships. In a way, the goal of canonical correlation is very similar to principal component analysis. We discuss the theory of canonical correlation analysis as follows (Fang, 2017). First, we define two sets of variable X and Y :

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, \quad (3.16)$$

and

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_q \end{bmatrix}. \quad (3.17)$$

Without loss of generality we assume that $p \leq q$. Similar to principal components analysis, we look at linear combinations of the variables. We define a set of linear combinations labeled U and V . U corresponds to the linear combinations from the first set of

variables, X , and V corresponds to the second set of variables, Y . Each member of U is paired with a member of V . For example, U_1 below is a linear combination of the p X variables and V_1 is the corresponding linear combination of the q Y variables. Similarly, U_2 is a linear combination of the p X variables, and V_2 is the corresponding linear combination of the q Y variables. We can repeat this process up to the set U_p and V_p .

Define (U_i, V_i) as the i -th canonical variate pair. (U_1, V_1) is the first canonical variate pair, (U_2, V_2) is the second canonical variate pair, and so on. Note that these pairs are selected such that the linear combination that define U_i is independent of the linear combinations that defined $U_k, k = 1, 2, 3, \dots, i - 1$, and a similar independent structure holds for the V_i . With $p \leq q$ there are p canonical variate pairs. The goal is to find linear combinations that maximize the correlations between the members of each canonical variate pair. The correlation of the i^{th} canonical variate pair is:

$$\rho_i = \frac{cov(U_i, V_i)}{\sqrt{Var(U_i)Var(V_i)}}. \quad (3.18)$$

3.1.4. Support Vector Regression. For the non-parametric models, we considered Support Vector Regression (SVR). Some regression problems cannot be correctly described using a linear model. In such a case, SVR with nonlinear kernel (Evgeniou and Pontil, 2011) is a good model to consider. In this project we use `svm()` function in R with the `e1071` package. The traditional belief is that SVR algorithm will try to find an optimal hyper-plane to optimize a given criterion, which in the simplest case is the linear hyper-plane. In this case the hypothesis space is a collection of all hyper-planes of the form:

$$f(x) = \mathbf{w} \cdot \mathbf{x} + b. \quad (3.19)$$

In the above formulation, SVR algorithm tries to find a hyper-plane in a space that the input data \mathbf{x} is in. This hyper-plane exists in a feature space which is created by a kernel K . This space can be expressed as the set:

$$S = \{f : \|f\|_k^2 < \infty\}. \quad (3.20)$$

where K is the kernel that defines Reproducing Kernel Hilbert Space. For the linear case, K is the dot product such that $K(x_1, x_2) = x_1 \cdot x_2$, the functions considered are (3.10). In non-linear problems (Mathworks, 2013), we replace linear kernel with a non-linear kernel K such that $K(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle$ where $\phi(x)$ is the transformation that maps features \mathbf{x} to a high-dimensional space where a linear representation of \mathbf{x} is possible. The loss function of support vector regression is defined as:

$$|y - f(\mathbf{x})|_\epsilon = \begin{cases} |y - f(\mathbf{x})| - \epsilon & \text{if } |y - f(\mathbf{x})| > \epsilon \\ 0 & \text{otherwise} \end{cases}, \quad (3.21)$$

where ϵ is the tolerable error. Having hyper-plane and loss function defined, support vector regression works by minimizing trade off between error and complexity of hypothesis space. This minimization problem can be written as:

$$\min \|f\|_k^2 + C \sum_{i=1}^l |y_i - f(\mathbf{x})|_\epsilon, \quad (3.22)$$

where f belongs to the set S .

3.1.5. Gradient Boosting Tree. In reality, linear regression models do not perform well in some situations where highly non-linear or complex relationships exist between the response variable and the independent variables, such as with some biological data. The data also has to meet certain assumptions which are not always satisfied. This leads to inaccuracy in prediction and unreliable model interpretation. In that case, we can use support vector regression model with a nonlinear kernel but this sacrifices interpretability of the model. Gradient boosting tree (GBT) is a good alternative that balances between prediction power and interpret-ability of the model while making little assumption about the data and can handle non-linearity pretty well.

To understand gradient boosting tree, we will first describe the regression trees that forms the basis for the boosted trees (Hastie *et al.*, 2016). Suppose that our data has p inputs $x_1, x_2, x_3, \dots, x_p$, one response y , and N observations. Thus, the data can be represented

as (x_i, y_i) for $i = 1, \dots, N$ and $x_i = x_{i1}, x_{i2}, \dots, x_{ip}$. The regression tree model has to automatically find splitting variables and splitting points based on a criterion. Suppose that we have M regions $R_1, R_2, R_3, \dots, R_m$ and c_m is the response in each region and we define:

$$f(x) = \sum_{m=1}^M c_m I(x \in R_m). \quad (3.23)$$

Our criterion is minimizing the sum of square $\sum (y_i - f(x_i))^2$. Then in this case the best \hat{c}_m is the average of y_i in region R_m . That is:

$$\hat{c}_m = \text{ave}(y_i | x_i \in R_m), \quad (3.24)$$

define j as splitting variable, s as splitting point and R_1, R_2 pairs of half-plane such that:

$$R_1(j, s) = \{X | X_j \leq s\} \text{ and } R_2(j, s) = \{X | X_j \geq s\}, \quad (3.25)$$

we seek to find j and s that solve:

$$\min_{j, s} [\min_{c_1} \sum_{x_i \in R_1(j, s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j, s)} (y_i - c_2)^2], \quad (3.26)$$

for every j and s , the minimization problem can be solved by:

$$\hat{c}_1 = \text{ave}(y_i | x_i \in R_1(j, s)) \text{ and } \hat{c}_2 = \text{ave}(y_i | x_i \in R_2(j, s)). \quad (3.27)$$

After we find the best split, we subdivide the data into the two regions and repeat the whole process on each of the two regions, this process is repeated on all of the resulting regions. Then we can fit the gradient boosting tree using algorithm (Hastie *et al.*, 2016) below:

1. Initialize $f_0(x) = \text{argmin}_{\gamma} (\sum_{i=1}^N L(y_i, \gamma))$.
2. item For $m = 1$ to M :

(a) For $i = 1, 2, \dots, N$ compute:

$$r_{im} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f=f_{m-1}}. \quad (3.28)$$

(b) Fit a regression tree to the targets r_{im} with terminal regions $R_{jm}, j= 1, 2, \dots, J_m$.

(c) For $R_{jm}, j= 1, 2, \dots, J_m$ compute:

$$\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma). \quad (3.29)$$

(d) Update $f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$.

3. Output $\hat{f}(x) = f_M(x)$.

3.2. FEATURE SELECTION METHODS

Biological data sets are very rich in information from multiple attributes. This makes such data high dimensional. It is not unusual to see data sets with hundreds of variables (features).

Feature Selection is a crucial component in projects with high dimensional data. When we fit a model with high dimensional data, its performance is usually inferior because the training time increases exponentially with the number of features, and models tend to over-fit when they consist of a large number of variables. Feature Selection methods help with this problem by reducing the dimensions without losing much information. It also helps to make sense of the features and their importance. In this project, we used LASSO (Least Absolute Shrinkage and Selection Operator) regression to remove unimportant variables. Following is a discussion of the theory behind Lasso regression.

Lasso regression, in theory, is very similar to linear regression, with the only different being its loss function. In linear regression, the loss function is given by:

$$\sum_{i=1}^N (y_i - \hat{y}_i)^2 = \sum_{i=1}^N (y_i - \sum_{j=1}^K \beta_j x_{ij})^2. \quad (3.30)$$

The Lasso loss function is given as (Tibshirani, 1996):

$$\sum_{i=1}^N (y_i - \hat{y}_i)^2 = \sum_{i=1}^N (y_i - \sum_{j=1}^K \beta_j x_{ij})^2 + \lambda \sum_{j=0}^P |\beta_j|, \quad (3.31)$$

$$\text{subject to } \sum_{j=1}^K |\beta_j| \leq t. \quad (3.32)$$

This loss function takes the magnitude of coefficients into account. This type of regularization (termed L1 regularization) can lead to zero coefficients; that is, the model ignores some variables when it evaluates the output. These characteristics not only help in reducing over-fitting, but it can also help in feature selection. The regularization parameter λ can be controlled. In this project, we used 10 fold cross-validation and selected λ that minimized mean squared error of the test data sets.

3.3. ERROR ESTIMATION

In an ideal scenario where we have an abundant number of observations, we would split the data set into training sets and testing sets randomly to do the error estimation. In this project, the number of observations is small so we can't afford to lose a considerable amount of data for the testing set. One way to overcome the limitation is to do K-fold validation. The procedure for K-fold validation is that (Shao, 1991):

1. Randomly shuffle the data set
2. Divide the data set into K disjoint subsets of equal sizes
3. Pick one subset as the testing set, remaining K-1 subsets as the training set and compute prediction error of the model using the testing set
4. Repeat Step 3 until we have used every subsets of the data as a testing set
5. Estimated prediction error is the average of K estimations of error

With this method, every data point is in a test set exactly one time and gets to be in a training set K-1 times. The variance of the error estimate is reduced as we increase K. In this project, we used root mean squared error (*RMSE*) as one criteria to evaluate regression

models prediction accuracy. *RMSE* is the squared root of the average of squared differences between predictions \hat{y} and actual observations y . We define:

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_i - \hat{y}_i)^2}. \quad (3.33)$$

3.4. COMPARISON OF PREDICTION MODELS

Akaike information criterion (*AIC*) is based on in-sample fit to estimate the likelihood of a model to predict/estimate the future values (Emad *et al.*, 2015) . A good model is the one that has minimum *AIC* among all the other models. Bayesian information criterion (*BIC*) is another criteria for model selection that measures the trade-off between model fit and complexity of the model. A lower *AIC* or *BIC* value indicates a better fit. The following equations are used to estimate the *AIC* and *BIC* of a model:

$$AIC = -2 \ln L + 2k, \quad (3.34)$$

$$BIC = -2 \ln L + 2k \ln N, \quad (3.35)$$

where L is the value of the likelihood function at the estimated parameters, N is the number of recorded measurements, and k is the number of estimated parameters. In this project we used *AIC*, *BIC* and *RMSE* as model selection criteria.

4. RESULTS

4.1. PREDICTION USING PARAMETRIC MODELS

The method we used to evaluate prediction accuracy is ten-fold validation. Metrics used to compare these models are *RMSE*, *AIC*, *BIC* and adjusted R^2 . In this section, the linear model without feature selection is abbreviated to *lm_model*, the linear model with feature selection is abbreviated to *lm_fs_model*. The summary results of the models are outlined in Table 4.1 and Table 4.2.

We can see that the linear model with feature selection is better in terms of both *AIC* and *BIC* with a reduction in adjusted R^2 , but this reduction of adjusted R^2 is not practically significant. We present the performance of these two models in Table 4.2 which illustrates prediction accuracy. *lm_model* produced results with decent accuracy but there is evidence for over-fitting (testing *RMSE* is 24% larger than training *RMSE* error). *lm_fs_model* out-performed *lm_model* by producing results with smaller *RMSE* while improved the over-fitting issue.

The result is clear, for linear models, the model with feature selection out-performed model without feature selection in terms of prediction accuracy and goodness of fit. But with the analysis of residuals, the model with feature selection violated one of the assumptions

Table 4.1. Prediction accuracy of the models.

Model	Training RMSE	Testing RMSE
<i>lm_model</i>	28.61	35.41
<i>lm_fs_model</i>	30.61	34.29
<i>tr_model</i>	2.67	32.11
<i>svm_model</i>	30.635	34.71

Table 4.2. Summary result for linear regression models.

Model	Adjusted R^2	AIC	BIC
lm_model	0.5745	5812.92	6241.67
lm_fs_model	0.5713	5781.52	6096.52

of the linear model (homoscedasticity of residuals) as seen in plots in Appendix A. The Q-Q plot, residuals vs. observations plot and probability density of the residuals, as well as prediction interval of the linear model with feature selection, can be found in appendix A.

4.2. PREDICTION USING NON-PARAMETRIC MODELS

For the non-parametric models, we used gradient boosted tree (`tr_model`) and support vector machine (`svm_model`). The method we used to validate the result is ten-fold validation; the metric we used to measure the performance of the models is root mean squared error.

Fitting a gradient boosted tree model is a challenging task; we have to input many hyper-parameters to start the fitting process. To deal with this problem, we used Bayesian optimization to tune the hyper-parameters of `tr_model`. A brief introduction to Bayesian optimization is included in Appendix C and the set of hyper-parameter found by the algorithm is presented in Table 4.3. Prediction accuracy is presented in Table 4.1.

From Table 4.1, we can see that `tr_model` gives a slightly better result in terms of accuracy comparing to linear models. We should notice that the training error of the model is smaller than testing error many times over. Usually, we can expect that testing error is larger than training error, but when this difference is significant, it indicates that our model over-fitted the training data.

If we take a look at Table 4.1, the support vector machine model gives a similar result to that of the linear model with feature selection. With Table 4.4, we see hyper-parameters tuned using Bayesian optimization indicate that the real relationship between predictors and

Table 4.3. Hyper-parameter for gradient boosted tree model.

Distribution	Number of Trees	Interaction Depth	Minobsinnode	Shrinkage
Laplace	3000	8	7	0.01

Table 4.4. Hyper-parameter for support vector machine model.

Kernel	Gamma	Degree	Cost	Tolerance	Epsilon
polynomial	1	1	2	0.01	0.01

response variable is linear. Therefore, the linear model with feature selection is the best model (out of every considered models) to model the data without sacrificing the prediction accuracy.

4.3. LINEAR MIXED MODEL

With data consisting of repeated measures within individuals, the linear model generally can't adequately accommodate dependent structure within an individual; this is why we used a mixed model to make inference on the data. With the data set, the ID of subjects is considered as a random effect and the rest as fixed effects.

With nearly 200 variables at the beginning, it is difficult to make any meaningful inference because of noise from other unrelated variables and multi-collinearity. In this project, we used L1 regularization (LASSO) to select important features. After unrelated variables were discarded, we did multi-collinearity diagnostic to detect variables with high variance inflation. We present variables selected by LASSO and multi-collinearity diagnostic in Appendix B of this thesis. As seen in that table, non of the selected variables have high VIF values, this indicating the absence of multi-collinearity in the model. Confidence interval for fixed effects are presented in Table 4.5, Note that only significant variables are displayed (p-value < 0.05), confidence interval for random effect is presented in Table 4.6.

Table 4.5. Confidence intervals for fixed effects.

Variable	Estimate	Lower	Upper	p-value
Intercept	342.940	269.241	416.642	0.000
TST	-0.013	-0.0257	-0.001	0.043
KSS2	5.475	3.878	7.073	0.021
AM.medRT	-0.016	-0.025	-0.006	0.013
DSST.coicidentFalseStarts	11.075	1.420	20.730	0.012
ERT.eff	-0.043	-0.087	-0.003	0.043
NBACK.accuracy	-0.039	-0.057	-0.022	0.036

Table 4.6. Confidence intervals for random effect.

Random effect: ID	Lower	Estimated	Upper
Intercept	18.63	27.46	40.46
Residual	37.90	34.71	26.07

Results in Table 4.5 shows that one minute increase in total sleep time (TST) leads to -0.013 ms decrease in PVT median reaction time with the 95% confidence interval given by (-0.0257, -0.001). We can see that the effect of TST is significant but is very weak, one hour (60 minutes) increase in total sleep time leads to only 0.78 ms decrease in reaction time with the 95% confidence interval given by (-1.542, -0.060). Strong effects we can identify are DSST.coicidentFalseStart, with one unit increase in DSST.coicidentFalseStart leading to 11.075 ms increase in median reaction time with the 95% confidence interval of (1.42, 20.730). Another strong effect is KSS2 with one unit increase in KSS2 leads to 5.475 ms increase in median reaction time with 95% confidence interval is (3.878, 7.073).

For random effect, we can interpret Table 4.6 as follows: it is estimated that 44.17% (or $\frac{7.46}{27.46+35.71}$) of left-over variance can be explained by random effect (difference between people), with lower bound of 32.95% and upper bound of 60.81%. The difference between people is significant and it can explain much of the variance that is not explained by the fixed effects.

4.4. CANONICAL CORRELATION ANALYSIS

In this subsection, we used canonical correlation analysis to explore the correlation between two sets of variables. We present the description of the X set of variables in Table 4.7 and the description of the Y set in Table 4.8.

In Table 4.9, we present the result of canonical correlation analysis. We were able to describe the relationship between two sets of variables in 3 dimensions. With the 0.05 significant level, we conclude that all three dimensions are significant. Standardized canonical coefficients for both x and Y set can be found in Table 4.10 and Table 4.11. Figure 4.1 is provided so we can easily understand the relationship between X and Y , with red vectors represent X and blue vectors represent Y .

With the first canonical variate pair, for X set, the first canonical variable V_1 has a strong correlation with Sex (0.79). For the Y set, the first canonical variable U_1 has a strong correlation with AM.accuracy (0.89). The interpretation is that for the first canonical dimension, Sex (encoded as 1 for female and 0 for male) is the strongest predictor for the accuracy of abstract matching, with females tending to have better performance in abstract matching than their male counterparts.

With the second canonical variate pair, for X set, ESS.Total has a strong positive correlation (1.04), while Sex has a negative correlation (-0.75) with the second canonical variable V_2 . For Y set, variables with significant correlation to the canonical variable U_2 can be accounted for are BART.balloonsPopped (0.54) and BART.accuracy (0.71). We can conclude that for the second set of canonical correlations, ESS.Total is the strongest predictor for BART.balloonsPopped and BART.accuracy. The more sleepy a person feels, the more they tend to take risks. With the negative correlation of Sex, we can say that male subjects less likely to take risks than their female counterparts.

With the final canonical variate pair, for X set, AM.PM (encoded as 1 for AM and 0 for PM) has a strong negative correlation (-1.00) with the 3rd canonical variable V_3 . For Y set, we can account for BART.balloonPopped (-0.59) and DSST accuracy (-0.57). We

conclude that subjects are more accurate in these tasks (DSST and BART) in the morning than that of the afternoon.

Table 4.7. The X set of variables.

Name	Description
ESS.Total	Measures the sleepiness of a subject
Sex	Indicates gender of a subject
AM.PM	Indicates the time of day when a subject took the test

Table 4.8. The Y set of variables.

Name	Description
AM.accuracy	Accuracy of abstract matching
BART.balloonsPopped	Number of balloon popped in Balloon analogue risk task
BART.accuracy	Accuracy of Balloon analogue risk task
DSST.accuracy	Accuracy of Digit symbol substitution test
ERT.accuracy	Accuracy of Emotion recognition test
NBACK.accuracy	Accuracy of N-back test
PVT.medRT	Measures reaction time of Psychomotor vigilance test

Table 4.9. Results of canonical correlation analysis.

Dimension	Approx	df1	df2	p-value
1 to 3	8.12	21	1694.71	0.00
2 to 3	1.26	12	1182.00	0.23
3 to 3	1.32	5	592.00	0.25

Table 4.10. Standardized canonical variables for the X set.

Variable	V1	V2	V3
AM.PM	0.06	-0.09	-1.00
sex	0.79	-0.75	0.07
ESS.Total	0.36	1.04	0.04

Table 4.11. Standardized canonical variables for the Y set.

Variable	U1	U2	U3
AM.accuracy	0.89	0.10	0.14
BART.balloonsPopped	0.35	0.54	-0.59
BART.accuracy	-0.15	0.71	0.46
DSST.accuracy	-0.03	-0.23	-0.57
ERT.accuracy	-0.03	-0.04	0.37
NBACK.accuracy	-0.20	0.14	-0.08
PVT.medRT	-0.19	0.44	-0.49

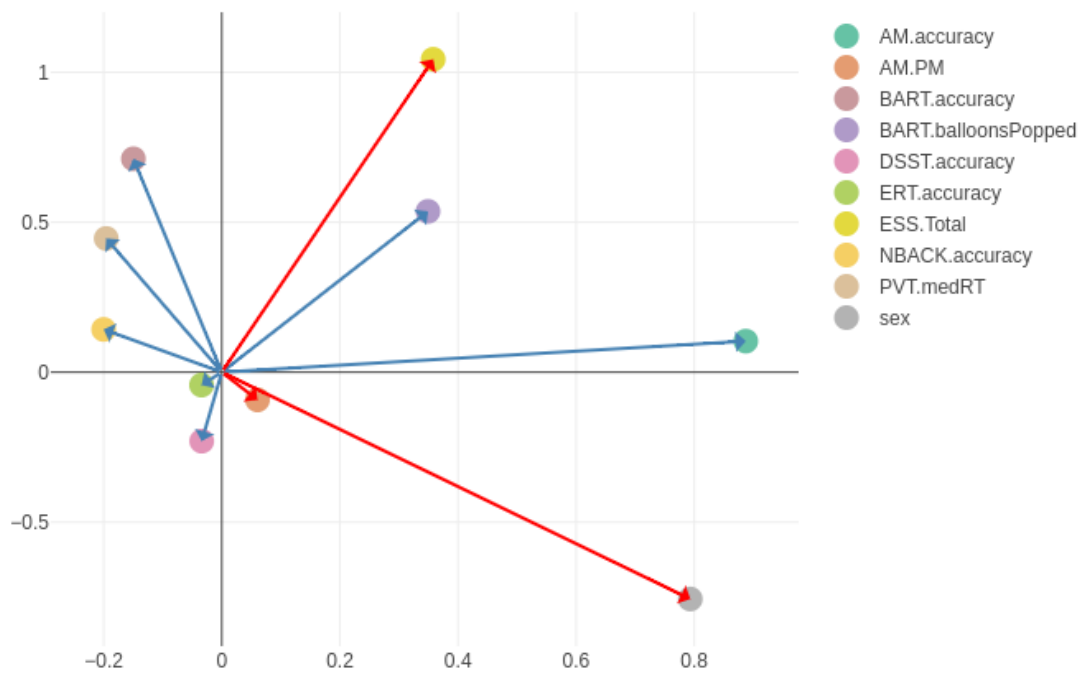


Figure 4.1. Visualization of the first two canonical dimensions.

5. CONCLUSIONS

This study compared several models for predicting Psychomotor Vigilance Test (PVT) median reaction times using variables that measure sleep deficits, gender and other variables. We conclude that the linear model with feature selection is the best among our considered models. Other non-parametric models may give slightly better RMSE but we would lose our ability to construct a meaningful prediction model. Overall, model fit diagnostics point out that our linear model with feature selection is a good fit, with RMSE of 34.29. Approximately 95% of observed values lie inside our prediction interval, verifying the validity of the interval and the model that yielded this interval. Therefore it is safe to say that we can predict PVT median reaction time with reasonable accuracy using the selected model.

With the mixed model, we were able to quantify the effects of sleeping time and other cognitive tests on the PVT median reaction time. Some of the effects are very weak, and unfortunately, the effect of ESS, which measures the sleepiness of an individual, is insignificant. The effect of total sleep time is also very weak, the lack of a large range of total sleep time may explain this. The subjects were not intentionally sleep deprived before taking cognitive tests, thus making the variations seen in total sleep time due only to individual behaviors that may not affect PVT in a significant way, and the same may be the reason why the effect of ESS variables were not significant. Furthermore, for some people, sleep deprivation for one day is not enough for effect to have a significant impact on PVT.

For future work, we suggest a more controlled experiment, meaning we should randomly sleep-deprive subjects before doing cognitive tests. That way, we would be able to have more insights about the effects of sleep deprivation on vigilant attention. Because subjects committed to the project from the start to the end of a semester (4 months) and with the same cognitive tests were used repeatedly, a learning effect is inevitably introduced. To

get rid of the learning effect, we suggest that one subject should only be sleep deprived and tested for cognitive performance no more than 5 sessions in 10 months, and the time when these sessions are arranged should be randomly generated.

APPENDIX A.

MODEL FIT DIAGNOSTICS AND PREDICTION INTERVAL

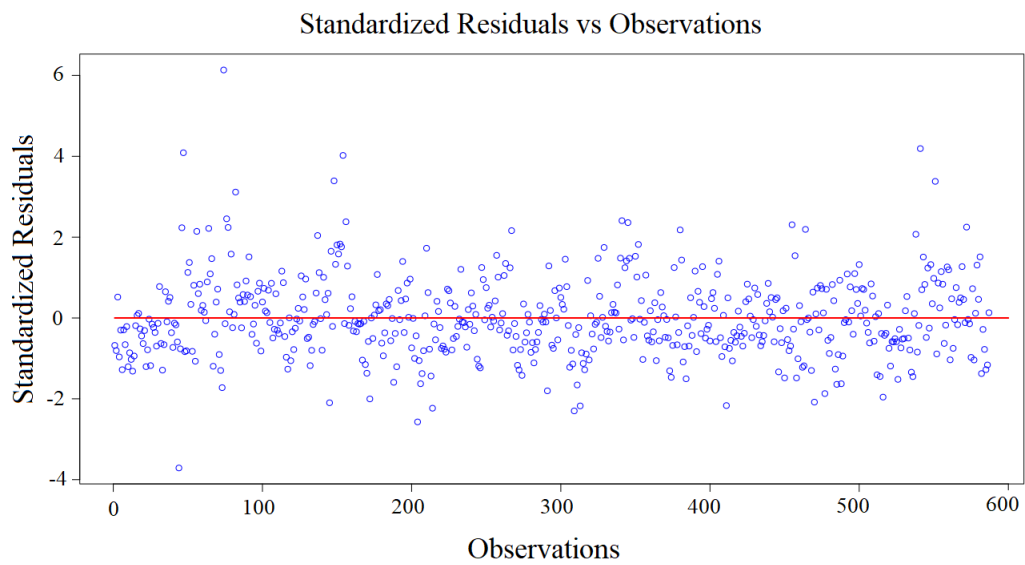


Figure A.1. Standardized Residuals vs Observations for the linear model with feature selection.

Studentized Breusch-Pagan test

H_0 : Residuals are homoscedastic

H_A : Residuals are non-homoscedastic

Data: Residuals from lm_fs_model

BP = 106.81, df = 70, p-value = 0.003044

Conclusion: Reject Null hypothesis, the residuals are non-homoscedastic

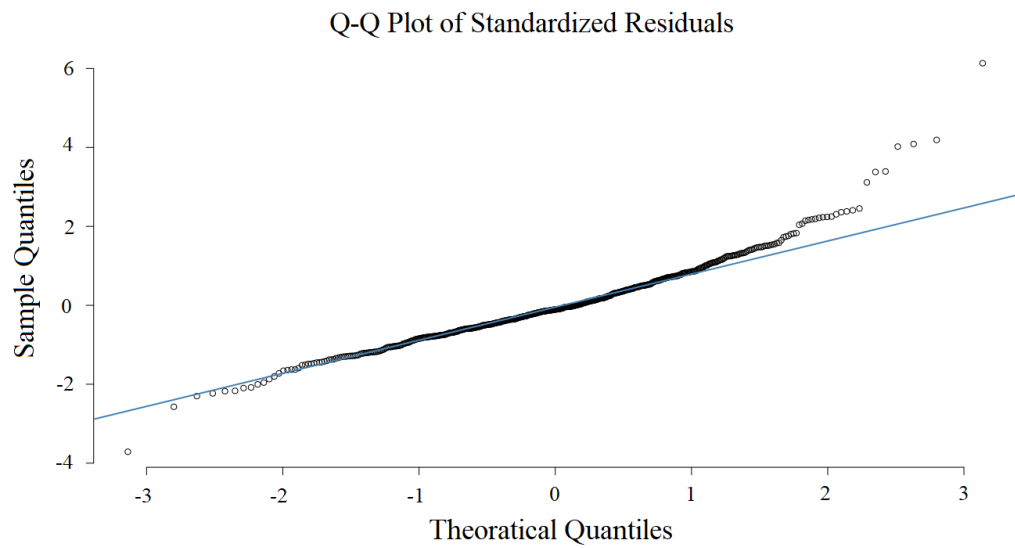


Figure A.2. Q-Q plot of the standardized residuals.

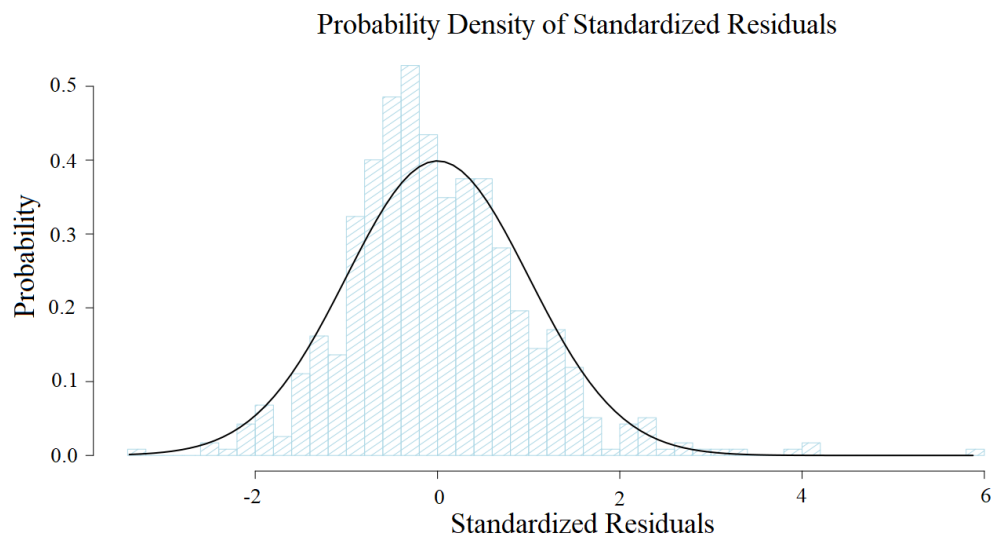


Figure A.3. Probability density of the standardized residuals.

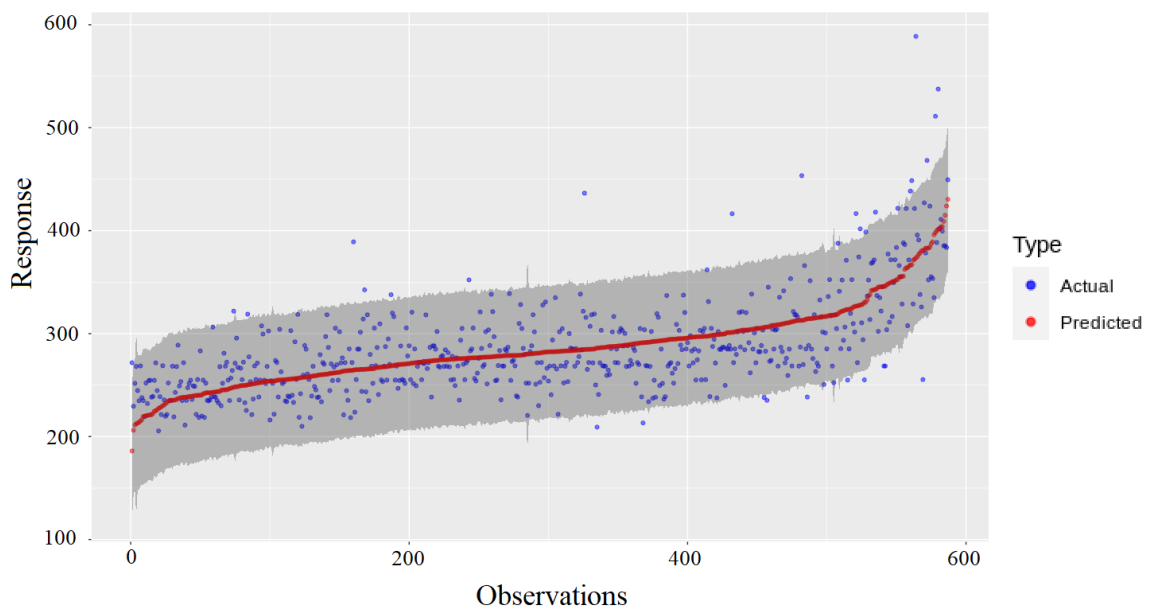


Figure A.4. Prediction interval for the linear model with feature selection (Shaded area is the prediction interval).

APPENDIX B.

FEATURE SELECTION AND MULTI-COLLINEARITY DIAGNOSTIC

Table B.1. Feature selected and multi-collinearity diagnostic.

Feature selected by LASSO	VIF	Removal
TST	1.135936	
sick.lvl	1.456087	
KSS2	1.254057	
AM.meanRT	23.390448	Removed
AM.medRT	23.600809	
AM.eff	1.386447	
BART.medRTpump	17.115560	Removed
BART.meanRTpump	19.688752	
BART.SDRTpump	2.969898	
BART.Sdpumps	1.748687	
BART.meanPumpOptimalDifference	1.762794	
DSST.throughput	1.621286	
DSST.coicidentFalseStarts	1.110574	
DSST.incorrectResponses	1.302286	
ERT.correct.fear	Inf	
ERT.correct.sad	Inf	Removed
ERT.incorrect.fear	Inf	Removed
ERT.incorrect.sad	Inf	Removed
ERT.meanRT.correct	9.706738	Removed
ERT.meanRT.correct.fear	2.879216	
ERT.meanRT.correct.anger	2.001093	
ERT.meanRT.correct.happy	2.992456	
ERT.meanRT.correct.sad	1.720077	
ERT.meanRT.incorrect.happy	1.182714	
ERT.eff	2.895108	
ERT.falsePos.sad	2.367009	
ERT.falsePos.neutral	2.602774	
NBACK.correct	7.189816	Removed
NBACK.accuracy	6.062331	
NBACK.meanRT	1.509278	
ESS.cat.A	1.971490	
ESS.cat.B	1.845447	
NBACK.medRT.dummy	Inf	
NBACK.medRT.correct.dummy	Inf	Removed
NBACK.medRT.correctMatch.dummy	Inf	Removed
NBACK.meanRT.dummy	Inf	Removed
NBACK.SDRT.dummy	Inf	Removed

Table B.2. Final set of features and their variance inflation factor (VIF).

Feature	VIF
TST	1.127967
sick.lvl	1.339635
KSS2	1.230240
AM.medRT	1.611656
AM.eff	1.371101
BART.meanRTpump	2.091357
BART.SDRTpump	1.379046
BART.Sdpumps	1.673135
BART.meanPumpOptimalDifference	1.665977
DSST.throughput	1.605764
DSST.coincidentFalseStarts	1.093301
DSST.incorrectResponses	1.281517
ERT.correct.fear	1.864330
ERT.meanRT.correct.fear	1.434122
ERT.meanRT.correct.anger	1.387543
ERT.meanRT.correct.happy	1.547624
ERT.meanRT.correct.sad	1.204952
ERT.meanRT.incorrect.happy	1.136398
ERT.eff	1.956233
ERT.falsePos.sad	2.001514
ERT.falsePos.neutral	2.207772
NBACK.accuracy	1.420021
NBACK.meanRT	1.465857
ESS.cat.A	1.727275
ESS.cat.B	1.628909
NBACK.medRT.dummy	1.190946

REFERENCES

- Basner, M. and Rubinstein, J., 'Fitness for duty: A 3 minute version of the psychomotor vigilance test predicts fatigue related declines in luggage screening performance,' <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3190077/>, 2012, accessed: 2020-04-11.
- Bassett, S., Lupis, S., Gianferante, D., Rohleder, N., and Wolf, J., 'Sleep quality but not sleep quantity effects on cortisol responses to acute psychosocial stress,' <https://www.tandfonline.com/doi/abs/10.3109/10253890.2015.1087503>, 2015, accessed: 2020-04-13.
- Borkowska, A., Drożdż, W., Jurkowski, P., and Rybakowski, J. K., 'The wisconsin card sorting test and the n-back test in mild cognitive impairment and elderly depression,' <https://www.tandfonline.com/doi/abs/10.1080/15622970701557985/>, 2009, accessed: 2020-05-21.
- Browne, B., Van, S. T., Onsager, D., Simpson, D., Salaymeh, B., and Condon, R., 'Influence of sleep deprivation on learning among surgical house staff and medical students,' <https://europepmc.org/article/med/8178259>, 1999, accessed: 2020-04-13.
- Diego, E., 'Linear Regression Models, OLS, Assumptions and Properties,' <https://faculty.utrgv.edu/diego.escobari/teaching/Econ8370/Handouts/Chapter02.pdf>, 2011, accessed: 2020-03-11.
- Drummond, S., Amanda, B.-G., F., D. D., Ayalon, L., and Mednick, M. M. J., Sara C, 'The neural basis of the psychomotor vigilance task,' <https://academic.oup.com/sleep/article/28/9/1059/2708157>, 2005, accessed: 2020-05-21.
- Emad, M., Naugler, C., and Far1, B., 'Emerging business intelligence framework for a clinical laboratory through big data analytics,' <https://www.sciencedirect.com/topics/medicine-and-dentistry/akaike-information-criterion>, 2015, accessed: 2020-03-13.
- Evgeniou, T. and Pontil, M., 'Workshop on support vector machines: theory and applications,' https://www.researchgate.net/publication/221621494_Support_Vector_Machines_Theory_and_Applications, 2011, accessed: 2020-03-11.
- Fang, E., 'Canonical correlation analysis,' <https://online.stat.psu.edu/stat505/book/export/html/682>, 2017, accessed: 2020-04-22.
- Hastie, T., Tibshirani, R., and Friedman, J., *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, Springer, 2016.

- Jaeger, J., 'Digit symbol substitution test, the case for sensitivity over specificity in neuropsychological testing,' <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6291255/>, 2018, accessed: 2020-05-21.
- Johns, M. W., 'A new method for measuring daytime sleepiness: The epworth sleepiness scale,' <https://academic.oup.com/sleep/article/14/6/540/2742871>, 1991, accessed: 2020-05-21.
- Lafont, S., Marin-Lamellet, C., Paire-Ficout, L., Thomas-Anterion, C., Laurent, B., and Fabrigoule, C., 'The wechsler digit symbol substitution test as the best indicator of the risk of impaired driving in alzheimer disease and normal aging,' <https://www.karger.com/Article/Abstract/264631>, 2010, accessed: 2020-05-21.
- Lee, I.-S., Bardwell, W., Ancoli-Israel, S., and Dimsdale, J., 'Number of lapses during the psychomotor vigilance task as an objective measure of fatigue,' <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2854704/>, 2010, accessed: 2020-05-21.
- Mathworks, 'Workshop on support vector machines: theory and applications,' <https://www.mathworks.com/help/stats/understanding-support-vector-machine-regression.html>, 2013, accessed: 2020-03-11.
- Montagne, B., Kessels, R. P. C., De Haan, E. H. F., and Perrett, D. I., 'The emotion recognition task: A paradigm to measure the perception of facial emotional expressions at different intensities,' <https://journals.sagepub.com/doi/abs/10.2466/pms.104.2.589-598>, 2007, accessed: 2020-05-21.
- Pérez-Chada, D., Videla, A., O'Flaherty, M., Palermo, P., Meoni, J., Sarchi, M., Khoury, M., and Durán-Cantolla, J., 'Sleep habits and accident risk among truck drivers: A cross-sectional study in argentina,' <https://academic.oup.com/sleep/article/28/9/1103/2708214>, 2005, accessed: 2020-05-21.
- Rosenfeld, M., 'Ols in matrix form,' https://web.stanford.edu/~mrosenfe/soc_meth_proj3/matrix_OLS_NYU_notes.pdf, 2016, accessed: 2020-03-11.
- SAS documents, 'Mixed models theory,' <https://v8doc.sas.com/sashtml/stat/chap41/sect23.htm>, 1999, accessed: 2020-03-13.
- Shao, J., 'Linear model selection by cross-validation,' https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1993.10476299#.Xpent_I0luQ, 1991, accessed: 2020-03-12.
- Thomann, J., Baumann, C., Landolt, H.-P., and Werth, E. a., 'Psychomotor vigilance task demonstrates impaired vigilance in disorders with excessive daytime sleepiness,' <https://jcsn.aasm.org/doi/10.5664/jcsn.4042>, 2010, accessed: 2020-05-21.
- Tibshirani, R., 'Regression shrinkage and selection via the lasso,' <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1996.tb02080.x>, 1996, accessed: 2020-03-11.

VITA

Quang Le studied in Danang University of Technology, Danang, Vietnam where he achieved Bachelor of Science in Electrical Engineering in January 2018. He then studied in Missouri S&T, Rolla, Missouri, USA where he received a Master of Science in Applied Mathematics with Statistics Emphasis in August 2020.

Quang worked with Dr. Nguyen as an undergraduate research assistant from May 2017 - January 2018 in Danang University of Technology, Department of Electrical and Electronic Engineering. He worked with Dr. Thimgan and Dr. Samaranayake as a research assistant from September 2019 to May 2020 in Missouri S&T, Department of Mathematics and Statistics.

Quang worked as a teaching assistant from Jan 2017 - Jan 2018 under the supervision of Anh Nguyen in Danang University of Technology, Department of Electrical and Electronic Engineering. The course he assisted with teaching was Introduction to C programming. He worked as a graduate teaching assistant from Sep 2019 - May 2020 under the supervision of Kimberly Kinder in Missouri S&T Department of Mathematics and Statistics. The course he assisted with teaching was MATH1215 Calculus for Engineers II.