3-26-2015

A Novel Analysis of Performance Classification and Workload Prediction Using Electroencephalography (EEG) Frequency Data

Donovan L. Ricks

A NOVEL ANALYSIS OF PERFORMANCE CLASSIFICATION AND WORKLOAD PREDICTION USING ELECTROENCEPHALOGRAPHY (EEG) FREQUENCY DATA

THESIS

Donovan L. Ricks, 1 Lt., USAF

AFIT-ENG-MS-15-M-012

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENG-MS-15-M-012

A NOVEL ANALYSIS OF PERFORMANCE CLASSIFICATION AND WORKLOAD PREDICITON USING ELECTROENCEPHALOGRAPHY (EEG) FREQUENCY DATA

THESIS

Presented to the Faculty

Department of Electrical and Computer Engineering

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Computer Engineering

Donovan L. Ricks, BS Computer Engineering

1 Lt., USAF

March 2015

AFIT-ENG-MS-15-M-012

A NOVEL ANALYSIS OF PERFORMANCE CLASSIFICATION AND WORKLOAD PREDICITON USING ELECTROENCEPHALOGRAPHY (EEG) FREQUENCY DATA

Donovan L. Ricks, BS Computer Engineering

1 Lt., USAF

Committee Membership:

Brett J. Borghetti, PhD
Chair

Brian G. Woolley, Maj, USAF, PhD
Member

Christina F. Rusnock, Maj, PhD
Member

AFIT-ENG-MS-15-M-012

Abstract

Many contexts across the Department of Defense (DOD) impose high levels of workload on operators involved in making decisions which can cause critical degradation of performance. These contexts, or circumstances that form an event [1], require varying levels of workload that the operator is faced with as he or she attempts to complete a task. The focus of the research presented in this thesis is to determine if those changes in workload can be predicted and to determine if individual task performance can be predicted using machine learning. Despite many efforts to predict workload and classify individuals with machine learning, there has been little exploration of the classification and predictive ability of Electroencephalography (EEG) frequency data at the individual EEG Frequency band level. In a 711th HPW/RCHP Human Universal Measurement and Assessment Network (HUMAN) Lab study, 14 subjects were asked to complete Surveillance and Tracking tasks with16 scenarios in each respectively. Their physiological data, including EEG frequency data, was recorded to capture the physiological changes their body went through over the course of the experiment. The research presented in this thesis focuses on EEG frequency data, and its' ability to predict task performance and changes in workload. This thesis contributes research to the medical and machine learning fields regarding the classification and workload prediction efficacy of EEG frequency data. Specifically, it presents a novel investigation of five EEG frequencies and their individual and combined abilities to predict task performance

and workload. It was discovered that using the Gamma EEG frequency and all EEG frequencies combined to predict task performance resulted in average classification accuracies of greater than 90%.

Acknowledgments

I would like to express my sincere appreciation to my faculty advisor, Dr. Brett Borghetti, for his guidance, patience and support throughout the course of this thesis effort. The insight and experience was certainly appreciated.

I would like to dedicate this thesis to my mother, who halted the pursuit of her career when I was a child, so that I could pursue mine. I would also like to dedicate this thesis to my father, who always set the bar high for my career. I love you both.

1Lt. Donovan L. Ricks

Table of Contents

Abstract iv

Acknowledgments vi

List of Figures x

List of Tables xi

I. Introduction 1

  Problem Statement 2

  Research Objectives/Hypothesis 2

  Methodology 4

  Assumptions/Limitations 5

  Implications 6

  Structure of the Document 7

II. Literature Review 8

  Chapter Overview 8

  Structure of the Literature Survey 8

  Section 1: Augmentation, Workload and Performance Decrement 9

  Section 2: Workload Prediction and Classification using EEG data 15

  Section 3: ANNs, their structure and Classification using Physiological data 20

    I. Classification 20

    II. Artificial Neural Networks 20

  Section 4: Summary 26

III. Methodology 28

  Introduction 29

  711 HPW/RHCP Human Lab Formal Study 1 Experiment Description 29

Performance Classification 33

Performance Classification on Dual Classified Subjects 37

Novel Task Prediction on Dual Classified Subjects 38

Workload Prediction 39

ANOVA and Canonical Correlation 40

One Way Analysis of Variance (ANOVA) 40

Two-Way Analysis of Variance (ANOVA) 41

Canonical Correlation Analysis 43

K-Fold Cross Validation 44

Methods for performance classification evaluation 45

IV. Analysis and Results 46

Performance classification with individual and combined EEG frequency bands 46

Performance Classification on Dual Classified Subjects 54

Novel Task Prediction on Dual Classified Subjects 57

Predicting Workload using Individual EEG Frequencies and All EEG frequencies Combined 59

V. Conclusions and Recommendations 67

Performance Classification with individual EEG frequencies: Is it possible to classify performance using EEG data exclusively? 67

Other Machine Learning Techniques Used in Combination 68

Feature Reduction of the EEG Frequencies 69

Workload Prediction using EEG frequency data: Is it possible to predict workload using EEG frequency data exclusively? 70

Feature Reduction *71*

Development of an EEG Baseline *71*

Summary 72

Appendix A. Root Mean Squared Error per EEG frequency using the ANN to predict VACP Workload Channel and Scenario (1-16) in the Surveillance and Tracking Tasks Respectively 74

Appendix B. One Way ANOVA *p* value results between Workload Prediction RMSE using ALL EEG Data Combined and Individual EEG Frequencies as Input to the ANN and the Uniform Naïve Predictor. One-Way ANOVA used the lowest RMSE from analysis using each EEG frequency to predict workload 79

Bibliography 80

List of Figures

Page

Figure 1. Example ANN 21

Figure 2. Comprehensive Timeline from HUMAN Lab Study 30

Figure 3. 5 Fold Cross Validation 44

Figure 4. Classification accuracy across 16 Scenarios Surveillance and 47

Figure 5. 1x7 and 3x7 Gamma EEG frequency. 52

Figure 6. Classification Accuracy of Dual Classified Subjects 55

Figure 7. Classification Accuracy of Novel 58

List of Tables

Page

Table 1. Multi-Layer Perceptrons and their use in classification 26

Table 2. Performance Data Label for Surveillance and Tracking 34

Table 3. Test Matrix for Effectiveness of ANN...36

Table 4. Results of pilot study varying the number of neurons. 36

Table 5. Dual Classified Subjects 37

Table 6. Factors and Levels for workload prediction analysis 40

Table 7. EEG Frequency Bands and Their Frequency Ranges..51

Table 8. Two-Way ANOVA results of Gamma Classification accuracies. 52

Table 9. EEG Frequency and number of Scenarios with 59

Table 10.(a-f) (a.) Possible Truth Values in Workload 60

Table 11. RMSE using Uniform Naïve Workload Predictor 62

Table 12. Average Canonical Correlation (Scenario 1 - 16) 65

Table 13. Average Canonical Correlation (Scenario 1 - 16) 66

A NOVEL ANALYSIS OF PERFORMANCE CLASSIFICATION AND WORKLOAD PREDICITON USING ELECTROENCEPHALOGRAPHY (EEG) FREQUENCY DATA

I. Introduction

Within the past decade, the cognitive demands we place on our military operators have increased significantly. Often the Remotely Piloted Aircraft (RPA) operator is asked to simultaneously track several targets, monitor an information news feed regarding the current task and relay information to forces on the ground or other aircraft. These demanding tasks require the RPA operator to maintain vigilance over extended periods of time. Vigilance, defined as the ability to maintain attention and alertness over prolonged periods of time while monitoring for rare stimuli among frequently occurring stimuli [2], is an important capability for human system operators to have and sustain. The point at which performance begins to degrade is different with each operator.

This thesis specifically explores whether the change in performance or workload can be detected using only EEG frequency data. In an experiment done by the 711th HPW HUMAN Lab participants were asked to complete 16 Surveillance and Tracking tasks while their physiological data was recorded. Score was simultaneously recorded over the duration of the tasks which showed that some participants excel at these tasks while others struggle to perform the same tasks. The goal of the 711th is to develop a method of providing adaptive aiding, similar to the capabilities of the operator trying to complete the task, using physiological triggers to initiate the aiding process. Finding a way to boost operator performance at the point of performance degradation would greatly reduce the operator error we see today. Analyzing the EEG frequencies and their ability to predict changes in workload and task performance may result in findings that indicate that it is possible to identify an individual based on their performance in a task and predict changes in workload before performance degradation occurs.

Problem Statement

RPA operators are required to track several targets simultaneously, report current location, and be aware of a constantly changing battle environment. Several techniques have been used that have shown to be accurate in their ability to predict changes in workload and performance. These techniques include recording electrical activity of the heart (electrocardiogram), brain waves (electroencephalogram), remote eye tracking, respiration data, and even saliva samples [3, 4, 5, 6]. Finding a way to use the RPA operator's physiological data to initiate performance augmentation would greatly reduce the amount of error seen due to high levels of workload or performance degradation.

Research Objectives/Hypothesis

The primary objective of this research is to evaluate the ability of EEG Frequency data to predict operator task performance and objective workload during surveillance and tracking tasks. The research presented in this thesis will concentrate heavily on the analysis of each EEG frequency's ability to predict workload and task performance using machine learning.

This thesis answers the following three questions:

- Can machine learning be used to predict workload and classify performance using only EEG data as input?
- Which EEG frequency best predicts workload?
- Which EEG frequency best predicts task performance?

Based on previous research of EEG data to predict operator state, this research addresses the following two hypotheses:

- $H_{1:}$ Each EEG Frequency individually has a different task performance (High Performers or Low Performers) prediction accuracy than the others.

- $H_{2:}$ The changes in workload are associated with changes in power in the individual EEG frequency bands and in the nodes within them

If we fail to reject $H_{1,}$ we have shown evidence that there exists an individual EEG frequency that provides a higher level of task prediction accuracy compared to the other EEG frequencies utilized in the HUMAN Lab experiment. That evidence would support the notion that accurate prediction of task performance using a single EEG frequency band is possible. Proving H_2 to be true means that the methods used to induce changes in workload have an equal effect on the EEG frequency data and those changes are detectable using machine learning. However, rejecting this hypothesis means that the methods used in this thesis to predict workload were not sufficient and that further work is needed before accurate workload prediction using only EEG frequency data is need.

Methodology

This research explores machine learning and its' ability to predict and classify data using only EEG frequency data. Existing data from the 711th HPW/RHCP's Human Universal Measurement and Assessment Network (HUMAN) Lab human performance experiment trials were used to train, validate and test the classifier used in this research effort. The research presented in this thesis explores the EEG frequency bands, individually and combined (Alpha, Beta, Gamma, Delta, Theta), as inputs to a classifier to analyze efficacy in predicting task performance and predicting workload.

There are two studies presented in this thesis, and in each study, the relationship between two variables is characterized. The first study explores the relationship between EEG power and task performance, expressed in two classes, "High Performer" or "Low Performer". Each subject's performance class was computed based on the average of their final scores across 16 trials in each task. The second study explores the relationship between EEG power and objective workload, or an objective numeric value of how difficult a task is at any time-step. Workload values were generated from IMPRINT [7] using an individual model of the task execution for each subject for each scenario.

Assumptions/Limitations

Several assumptions and limitations exist in this research effort. This research was conducted using existing data from a human performance experiment conducted previously by an external organization. The human experiment was limited in the number of subjects that could be recruited and tested. A consistent procedure was performed for all subjects during the experimental sessions. External factors that could affect a person's attention such as time of day, amount of sleep, or previous caffeine intake are not known or considered for this research. The efficacy of the selection of factors and levels used in the trials to induce workload variance was not analyzed prior to the experiment to determine which portion of the workload-performance profile it exercised for each subject.

Performance classes used in our research were computed using the external organization's scoring algorithm, and this scoring algorithm was not analyzed to verify correctness or applicability to performance assessment in any real-world mission scenario. There was no pre-defined standard for performance in the Surveillance or Tracking tasks before the data set was received. In order to construct performance labels for this research, these performance thresholds needed to be established in order to determine whether a subject's performance in a task was high or low. This threshold was established before classification analysis began. Final performance classes defined in this thesis were defined based on average scores over the 16 scenarios in both the Tracking and Surveillance tasks respectively. If the participant scored over 900 in the Tracking task, the individual was labeled as a High performer. There were no individuals whose 16-scenario average score was greater than 900 in the Surveillance task. For this reason, a separate threshold of 600 points was established to differentiate between high performers (average score > 600) and low performers. This separate threshold ensured a maximum level of difficulty for the classifier by allowing an even set of high performers and low performers in both tasks.

In the HUMAN Lab experiment, no clearly defined baseline was established for the participant for analysis of the EEG frequency data. An EEG baseline can be measured when the participant remains motionless, closes their eyes to remove external stimuli to the brain, and maybe listens to calming music to ease the individual before the start of the study [8]. EEG data contains noise ranging from muscle twitches, blinking and other functions of the body. Therefore, each subject's EEG data can be treated as an immediate response to their current environment. It is difficult to analyze the predictive ability of one physiological feature when the experiment was not specifically designed to do so. For these reasons, caution must be taken when generalizing the results to all reconnaissance tasks.

Implications

Identifying an EEG frequency band that best classifies performance or predicts workload will allow researchers to reduce the amount of features used when augmenting performance based on EEG data. Reducing the amount of physiological data needed to predict task performance and workload would result in improved algorithms that use EEG data as one of their inputs. Currently, operator state and performance are predicted using a combination of physiological sensors that can inhibit the performance of the operator during a given task. Reducing the number of sensors needed to predict performance would result in a less constraining environment for the operator while still allowing researchers to precisely predict operator state and performance. Identifying an EEG frequency band that best predicts task performance in activities such as surveillance or tracking would move researchers one step closer to this effort. Quantifying the utility of each EEG frequency's ability to accomplish performance classification and workload prediction would aid researchers in developing an algorithm that used physiological data to trigger augmentation.

Structure of the Document

A review of research relating to classification and prediction of workload is presented in Chapter 2. An exhaustive explanation of how all experiments and analysis were conducted is presented to the reader in Chapter 3. Chapter 4 draws conclusions based on the results achieved from following the Methodology presented in Chapter 3. A detailed summary of the results from the classification and workload prediction analysis is presented in Chapter 5. Future work for follow-on research is recommended in Chapter 5 as well.

II. Literature Review

Chapter Overview

The Air Force Research Laboratory (AFRL) is developing a real-time classifier and predictor of operator state to facilitate augmentation online [9, 10]. The model incorporates physiological inputs (Electroencephalography, Electrocardiography, eye-tracking activity and galvanic skin response), and subjective workload assessments of each condition measured using the NASA Task Load Index to increase prediction accuracy. AFRL uses the "Sense, Assess, Augment" taxonomy [9, 10] to include all possible inputs to make an operator state prediction, workload estimate, and augment performance of the subject as necessary. Similar methodologies have been utilized elsewhere in research to make predictions about operator state and perceived workload with varying levels of success. This literature survey seeks to examine past research as it applies to prediction and performance augmentation to highlight key discoveries regarding these efforts and findings using Electroencephalography (EEG) data as the key input feature. It also explores the use of Artificial Neural Networks as a classifier and their ability to use EEG data to predict workload and task performance.

Structure of the Literature Survey

Section 1 details efforts to predict workload and augment performance similar to 711th HPW/RHCP HUMAN Lab Experiment. A review of research efforts where only

EEG frequency data was used to predict workload and classify individuals is presented in Section 2. Section 3 presents different Neural Networks used to classify workload and the classification accuracy of those Neural Networks (See Table 1 for itemized description). A summary of research regarding the efforts related to this thesis and a proposed direction to proceed for the analysis of data is presented in Section 4.

Section 1: Augmentation, Workload and Performance Decrement

According to Hart, workload can be seen as a term that represents the cost of accomplishing mission requirements for the human operator [11]. Specifically, this informal definition simplifies down to the fatigue, stress, illness and accidents that an operator may incur while performing a given task. Workload is "human centered", and "emerges from the interaction between the requirements of a task, the circumstances under which it is performed, the skills, behaviors, and perceptions of the operator" [12]. One of the most widely accepted methods used to capture perceived workload was the National Aeronautics and Space Administration Task Load Index (NASA-TLX). This subjective performance survey is a multidimensional assessment tool that asked subjects to rate perceived workload after a given task was completed. It is widely used as the foundation for truth data regarding perceived operator workload, has been cited in over 4,400 studies [11] and has a large influence in the Human-Factors research domain. The NASA-TLX is broken up into six parts: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. NASA-TLX requires the subject to rate him/herself on a scale of each category from “Very Low” to “Very High” [12]. A benefit of the NASA-TLX is that it allows researchers to collect subjective workload values directly from the subjects participating in the study rather than estimating workload values based on evaluating subject behaviors - which may not be as accurate. A drawback of NASA-TLX is that after completing a certain amount of an activity, answers related to workload and difficulty of each task may not be accurate due to the subject’s ability to remember the intricacies of the tasks he or she endured.

Christensen et al used a multi-RPA operation task that was PC-based that simulated a mission involving the suppression of enemy air defenses [13]. Subjects monitored 8-16 RPA aircraft with specific flight plans and when the aircraft came within range of a target, the targets were to be engaged with a specific weapon (small, medium, or large) dependent on the type of aircraft being targeted. Physiological data was recorded over the course of the trial (Electrooculography (EOG), EEG (five channels), and Electrocardiography (ECG)).There were two types of augmentation triggers utilized in this study: 1. Physio-Activated and 2. Operator Activated. In the physio-activated augmentation, a classifier trained to detect high workload on 20 minutes of physiological data was used. The independent variables to the classifier were the physiological inputs recorded and the dependent variable was perceived workload. Operator-activated augmentation was triggered by the participant at their own discretion. Adaptive augmentation came in the form of automatic target prioritization and cued time-critical alerts requested by the operator or initialized automatically during periods of high workload. The study showed that over the course of 3 days, subjects who were assisted with physio-aided augmentation did better than those with operator-selected augmentation. Subjects using physio-aided augmentation saw an improvement from an 84% hit ratio (targets hit/possible targets) to 90%, while those using operator selected augmentation saw a decrease from an 87% hit ratio to 84%. The experiment conducted by Christensen et al shows the positive impact that physio-initiated augmentation could have on the performance of a subject. The operator themselves may trigger augmentation too late because they are too focused on the task at hand. This study is beneficial because it shows that human operators can potentially benefit from physio-initiated augmentation and that this augmentation can boost task performance.

There is no definitive research that asserts one physiological feature as a better predictor than any other physiological feature. There is wide variation amongst researchers that shows operator workload can be predicted using a multitude of features together, or using different physiological features exclusively as predictors. Fong et al was able to predict mental workload accurately using only eye metric data (pupil diameter, divergence, fixation, movement) [4]. The experiment used the Automated Operation Span (OSPAN) task that has also previously been used to measure working memory capacity [4]. The OSPAN task has been seen as highly effective because it is believed that, "As working memory processing requirements increase, mental workload increases" [4]. For the OSPAN task, subjects are presented with basic arithmetic questions of different set sizes and given a limited amount of time to provide a correct answer. In the context of the OSPAN task, a set is a grouping of arithmetic problems given to the subject. After all the questions in the set are answered, the subject is presented a letter; following the display of all the questions and letters, subjects must recall all the letters in the correct order they were displayed. The experiment presented in Fong's research was a three-class classification problem, where he tried to predict High, Medium and Low workload (dependent variables).

The independent variables in this study were pupil diameter, divergence, fixation, movement. Fong et al were able to achieve workload classification rates as high as 85% using just pupil metrics and the OSPAN testing method. Fong et al not only showed that workload prediction from ocular physiological data was possible, but he also showed the efficacy of Artificial Neural Networks (ANNs). In his study the ANNs had a higher classification rate than the Classification tree he used to complete the same analysis. This bolsters the argument that ANNs are best suited at handling a high level of inputs.

Similarly, Song et al focused primarily on the P300 measurement to correlate changes in mental workload to physiological changes [14]. P300 is an EEG measurement recorded at the scalp and consists of the electrophysiological response to a stimulus evoked in the process of decision making. Its' activity is directly related to a person's reaction to a stimulus and not the physical attributes of the stimulus itself. It is said that P300 is closely related to the information processing capacity of the operator and can be applied to the classification and evaluation of operator's mental workload [14]. Mental workload was varied by changing the refresh rates of the target information that was supposed to be responded to by the participant. High mental workload was caused by high refresh frequencies. Conversely, Low mental workload was induced by lowering the information refresh rate. In Song's study, the independent variable was the mental workload seen by the participant, and the dependent variable was the P300 component from the EEG. Song et al were able to show that "the main effect of mental workload on the peak amplitude of P300 was significant" ($P = 0.031$, $P<0.05$). His study showed that peak amplitude of P300 under the low mental workload was higher than that of high mental workload. Song et al was able to show that when set up properly, an experiment with high workload can induce changes in participant EEG. Although P300 does not represent the EEG frequency data in its unprocessed form, it still shows the information processing capacity of the participant [14, 15]. Research by Song et al shows that it is possible to set up a study where workload has a direct effect on the physiological data. Results from Song et al show that an experiment can be created where the workload induces changes in the physiological data such that the changes in the physiological data correlate with the changes in workload.

Shaw et al examined the subjective and physiological workload seen by participants during a 3-D audio vigilance task. The study explored the benefits of using Multi-Modal Communication (MMC) as a means of delivering instruction and communication to Airborne Warning and Control System (AWACS) operators as opposed to the standard monaural method. In this experiment, the MMC method delivers audio to the operator via 6 different channels in both a 3-D spatial audio condition and the same amount of audio chatter with a monaural radio. Mental workload was measured via cerebral blood flow velocity (CBFV) and compared with a subjective measure of workload, the NASA-TLX. Participants were asked to detect hostile phrases read to them with both the 3-D spatial audio and monaural audio. Results showed that there was a significant vigilance decrement over time, but that overall detection probability was higher in the 3D Spatial Audio than in the Monaural Radio condition [16]. Research conducted by Shaw et al suggests the NASA-TLX may not be the best means of measuring operator mental workload because responses to the TLX after experiment could suffer from memory lapses and operator bias. Shaw et al also show that performance decrement is likely in tasks requiring vigilance from the operator, but that the means of information transmission can serve as a form of augmentation.

Tiwari et al also saw a decline in performance due to an increase in vigilance required to complete a task requiring subjects to detect critical and non-critical objects on a screen [17]. The high task condition had 300 events, 60 of which were critical targets. The low task condition was comprised of 150 events, 30 of which were critical targets. Performance was measured by the correct identification of critical targets. Tiwari et al were able to show that within 15 minutes of beginning a task, vigilance decrement can occur. His research also reports that when task demand conditions are high, the decrement can occur as quickly as the first 5 minutes of a task.

Saxby et al explores the theory that there are two types of fatigue: active and passive, and that introducing automation to alleviate workload may actually have a negative result. Active fatigue can be defined as an operator state where the operator is physically or mentally exhausted from maintaining a high level of vigilance in a task. Passive fatigue can be defined as an operator state where the operator has such a low level of consciousness that he/she is highly inattentive. Participants were required to keep a vehicle within the lanes on a simulated highway for 10, 30 and 50 minute durations. In the active fatigue simulation, "wind gusts" were used to make it harder to keep the vehicle inside the driving lanes. In the passive fatigue simulation, speed and steering were under full automation. On average, there was a decline in task engagement within the first 10 minutes of the experiment [18]. This correlates positively with past research stating that vigilance decrement can occur within the first 15 minutes of a task [17, 19, 20]. In a similar research study by Helton et al to assess the change in vigilance over time, participants saw an 8% drop in detection rate within the first 12 minutes of the task at hand [2]. It is clear that in research where workload is held constant and sustained vigilance is required, a decrement in performance and participant engagement will drop within some period of time. However, excessive augmentation can actually hinder the performance of a subject because it does not require the individual to maintain a high level of alertness. This passive fatigue caused by excessive augmentation is undesired in DOD where the war-time environment can change rapidly. Similarly, active fatigue from a lack of augmentation can cause performance decrement as well and is undesirable in the DOD.

Section 2: Workload Prediction and Classification using EEG data

Researchers have even used EEG based systems to predict cognitive workload and operator state with varying degrees of success, but have not been able to focus their efforts on the predictive capabilities of the individual EEG frequencies themselves. Only notional conclusions have been drawn regarding various combinations of EEG frequencies and singular EEG bands. Declaring an EEG frequency dominant in workload prediction would allow researchers to focus on including one or a combination of EEG frequency bands as features to their workload and performance prediction systems.

Borghini et al looked to study the variation of power in the EEG frequency bands as a subject started a new task. The goal of their study was to find the differences from the beginning of the training to the session in which the performance level is good enough for considering him/her able to complete the task without any problems [3, 21]. While novices of the study were engaged in flight simulation tasks, brain activity was recorded with the hope of seeing a notable change in the EEG data. EEG frequency data from 61 channels were recorded, band pass filtered (low-pass filter cut-off frequency: 40 (Hz), high-pass filter cut-off frequency: 1 (Hz)) and then ran through Independent Component Analysis to remove any artifacts from the data. Borghini et al showed that the brain activity in the Theta band over the left, central and right frontal areas decreased with respect to the session in which they got completely into the tasks (T3) [3]. Using cortical maps that depict brain activity visually, Borghini et al was also able to note the trend of the supposed learning process using only the Theta EEG frequency band [3, 21, 22]. Borghini noted that brain activity in the Theta band increases as subjects learn a new task and test strategies in pursuit of success within the study. Once a strategy is developed and implemented, power in the Theta band decreases. This was a notable finding by Borghini et al, but does not quantify the predictive ability of the Theta Band itself. EEG data was also focused on in a different study conducted by Borghini et al, where subjects were asked to drive in a simulated environment at a constant speed for an extended period of time. Results showed a burst (in the Alpha EEG Frequency data) occurred during the monotonous driving task as signal of drowsiness and reduced vigilance [22]. After the occurrence of the variation in the EEG signal subjects drove off from the correct trajectory lane with a high statistical occurrence when compared to the drive errors performed during standard driving conditions ($p < 0.05$) [22].

Ebrahimi et al and Lin et al were able to correctly identify different sleep and cognitive state respectively using EEG signals alone [23, 24]. Ebrahimi sought to identify between four subdivisions of the non-rapid eye movement (NREM) sleep state. NREM Stage 1 is a transitional stage “between wakefulness and sleep” [24]. NREM Stage 2 is the baseline of sleep respective to each subject. NREM Stage 3 is defined as the period of sleep where 20% to 50% of EEG signals with frequencies less than 2 Hz within the Delta waves and amplitudes more than 75 microvolts occur. Similar to Stage 3, NREM Stage 4 is the period of sleep where “Delta waves cover 50% or more of the record”. Sleep data from PhysioBank Database were used for his research. EEG signals recorded from seven Caucasian males and females (21-35 years old) without any medication for 24 hours sampled at 100 Hz were selected and EEG artifacts were manually removed to work with clean EEG data. Ebrahimi et al used a three-layer, feed-forward, ANN trained with standard back propagation to classify the different sleep stages. The output layer of the ANN had 4 neurons that signified the 4 different sleep stages respectively. Ebrahimi was able to achieve 93% sleep stage classification accuracy across all four sleep stages, obtaining no less than 84% classification accuracy for any one particular sleep stage. Accuracy was measured by correctly identified EEG sample data. This research is helpful because the research done by Ebrahimi et al with sleep stage identification is similar to the research conducted in this thesis with performance class identification. His research shows that it is possible to train an ANN to identify differences in EEG frequency data based on data labels placed on the EEG frequency data samples.

Similarly, Lin et al used a virtual-reality highway-driving environment to monitor and observe differences in EEG frequency data [23]. The goal of the research conducted by Lin et al was to develop an alert model system based off the EEG power in the Alpha and Theta EEG frequency bands. Lin et al completed a moving-averaged spectral analysis of the EEG data using a 500-point Hanning window without overlap. The result of the moving-averaged spectral analysis on the EEG was then compared to level of alertness of the participant.   The subject’s alertness level was defined as the deviation between the center of the vehicle and the center of the cruising lane. His research suggests that EEG power in the Theta and Alpha bands increase monotonically in tasks that require sustained attention. Lin et al was able to show that the Malahanobis Distance between the EEG power and a derived alert model strongly correlated with subject drowsiness in a linear fashion. Lin et al was able to induce change in the Alpha and Theta EEG frequency bands from the difficulty of the driving task. However, the research in these two studies still does not identify the weak and strong features in the EEG frequency dimensional space.

Belyavin et al used Independent Component Analysis(ICA) to remove artifacts within the EEG frequency data before finding which frequency best indicated verbal and spatial workload [5]. The verbal workload was induced using a visual task where subjects had to report the numbers presented to them directly after they disappeared from a screen. Workload was varied from low to high by increasing or decreasing the frequency with which the pictures appeared on the screen. The spatial task was a two dimensional simulated flying task using a joystick. Workload was induced using an increasing or decreasing amount of forcing functions affecting the frequency of subject interventions. Belyavin et al developed a conceptual model of cognitive workload named Prediction of Operator Performance (POP). One of the assumptions of this model is that only a small number of cognitive activities can be undertaken in parallel, even if there are multiple motor actions that can be done simultaneously. Research conducted by Nicholls et al on dual task experiments indicates that two important activities that can be done without significant interference are verbal and spatial tasks [25]. Belyavin et al used a kurtosis based Independent Component Analysis procedure to remove anomalous signals in the EEG frequency data. The EEG data was gathered from an array of 14 electrodes at a frequency of 1024 kHz was analyzed between 2 and 4 seconds long. Therefore, each block, or vector of EEG data was between 2048 and 4096 samples long. The first stage of the artifact removal process for a single block of EEG data was to calculate the contracted kurtosis tensor using the outputs of all the EEG. The kurtosis method of Independent Component Analysis (ICA) removed the spikes in the data at each time step to hinder any negative effects noise had on the EEG frequency recordings. After each block had been subjected to artifact removal, the spectrum was calculated into nine frequency bands (Delta, Theta, Alpha 1, Alpha 2, Beta 1, Beta 2, Gamma Low, Gamma Mid, Gamma High) using cross-spectrum analysis. The nine EEG recordings were then aligned with the total time of each task (150 seconds) to determine the effect of the verbal and spatial workload on the EEG data. The occurrences of different frequency components in the nine EEG frequencies were summed and tabulated at the conclusion of the alignment process. It was revealed that the Gamma 3 EEG frequency (70-100) Hz was the best indicator of verbal workload, while the Gamma 2 EEG frequency(53-70) Hz was the best indicator of spatial workload based on the number of frequency component occurrences in the frequency bands themselves(160 and 155 occurrences respectively). This was followed by Gamma 1 (30-47 Hz), Beta 2 (20-30 HZ), Beta 1 (14.1 – 20 Hz), and Alpha 2(10.2 – 14.1 Hz) which were third through sixth in their responsiveness to verbal and spatial workload. This study is extremely helpful because it provides a detailed review of EEG frequency bands themselves and reports the effect verbal and spatial workload has on each EEG frequency band. The study increases the number of EEG frequency band representations from the seven used in the 711th HPW/ RHCP HUMAN Lab study to nine. This study also provides some insight into what the expected workload prediction ability of the EEG frequencies themselves may be by reporting the number of frequency component occurrences due to the workload from the verbal and spatial tasks. However, Belyavin et al tackles the problem of finding the utility of the EEG frequency bands to predict workload from a signal analysis standpoint and not machine learning. Therefore, it cannot be assumed that the same EEG frequencies that had high frequency component occurrences in Belyavin's research will do well predicting workload in another study using machine learning.

Section 3: ANNs, their structure and Classification using Physiological data

I. Classification

Some researchers use Neural Networks to facilitate the process of identifying patterns and handling the complex computations needed to identify these patterns and groupings that exist in the data they handle. The act of determining those groups and finding those patterns with Neural Networks is called 'Classification'. Classification is used in this research to identify performance levels of participating subjects.

II. Artificial Neural Networks

An Artificial Neural Network (ANN) is a fully connected directed graph of artificial neurons that uses a mathematical model for information processing and data classification. The ANN closely emulates the neuron activity in the brain and its method to classify data it processes. An ANN can be used to classify data with complex relationships and find patterns in data [26]. The ANNs used to classify performance and predict workload presented in this research will be a feed forward neural network trained using scaled conjugate gradient back propagation. The Scaled Conjugate Gradient (SCG) calculates the approximation of the error term and uses a scalar α_n to regulate the indefiniteness of the Hessian term. Gradient descent takes the error seen at each epoch and reports it back to the neurons in the hidden layers to facilitate convergence. The output value for the ANN will be numeric when classifying workload and performance. The transfer functions for the hidden and output layer in the ANN used for task performance prediction are tan-sigmoid and log-sigmoid respectively. The transfer functions for the hidden and output layer in the ANN used for workload prediction are tan-sigmoid and pure linear respectively.

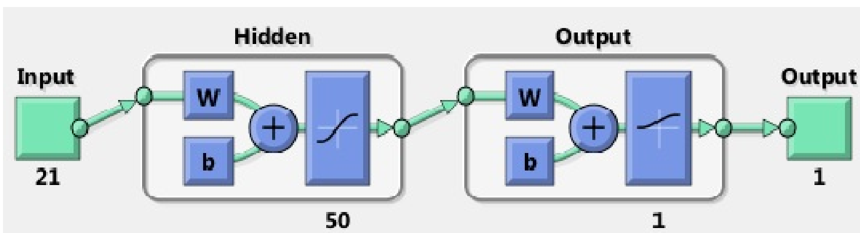

Figure 1. Example ANN

There are several means of classification including K-Means Clustering, Classification Trees, and Artificial Neural Networks (ANN). As it relates to workload prediction and classification, Classification Trees and ANNs have been most popular in research related to mapping several inputs to one or two outputs. Classification trees are used as a predictive model that map observations about an input to conclusions about the inputs' target value. An ANN is a computational model that projects sets of input data onto a set of appropriate outputs. The ANN consists of multiple layers in a directed graph, each layer fully connected between the input and output of the ANN.

Fong et al did a comparison of classification accuracy between Logistic Regression, Artificial Neural Networks, and Classification Trees to predict mental workload (High, Medium, Low) [4] using ocular physiological data as input (Pupil Divergence, Fixation and Movement). The ANN and Classification Tree were the two highest performing structures achieving classification ratios of 86.8% and 82.9% respectively. Fong's research notes that the ANN has "Good Predictive performance", "Handles complex relationships well", and has a "High tolerance to noisy data" [4]. Conversely, his research suggests that while Classification trees are "Good for variable selection", they are also sensitive to small changes in data. Considering the high variability in an EEG signal and the amount of noise within any EEG frequency recorded, evidence suggests it would be useful to use an ANN to attempt workload classification of EEG data.

Ebrahimi et al desired to classify sleep state based on EEG signals alone using a "three-layer feed forward perceptron" [24] with 12 inputs, 8 neurons in the hidden layer and 4 output neurons signifying the 4 sleep stages. Input to the ANN consisted of the 12 features used to represent the EGG data from the Beta, Alpha, Beta, and Theta EEG frequency bands. The ANN used in his research was able to achieve sleep stage classification accuracies no less than 84.2% and as high as 94.9%. His research suggests that with a higher amount of neurons in the hidden layer, "accuracy increases and standard deviation decreases" [24].

Correa et al tested 25 ANNs' in his research to differentiate alertness and drowsiness stages [27]. The EEG data was pre-labeled with respect to its stage before analysis according to Rechtschaffen and Kales method [28]. EEG records of ten subjects were selected from the MIT-BIH Polysomnographic Database whose ages were between 32-56 years old. The single available EEG signal was acquired between C3 and O1 positions in the 10-20 international node placement system with a sample frequency of 250

Hz. During testing, size of the hidden layer varied from 5 to 30 neurons. All EEG records were preprocessing with a 2nd order, bidirectional, Butterworth, band - pass filter with cut-off frequencies of 0.5 and 60 Hz. The ANN had one output neuron whose categories were “0” for the alertness stage and “1” for the drowsiness stage. The best ANN architecture (12-20-1, Input-Hidden layer-output layer) was able to achieve “86.5% of alertness stage detection and 81.7% of drowsiness stage detection” [24].

Wilson et al used an ANN to detect High Performers and Low performers in activities with Easy and Difficult task levels amongst 10 subjects in an RPA simulation task. Performance of the individuals was measured by the mean level of the scores within the tasks of the study. Those who fell above the mean were labeled “High Performers”, and those who fell below the mean were labeled “Low Performers”. ‘Easy’ was defined a low level of distractors in the RPA simulation task when tracking the target and vice versa for the ‘Hard’ task level. Wilson was able to achieve 89.7% classification accuracy of the easy condition and 80.1% classification accuracy of the difficult condition. Input to the ANN consisted of EEG data and Electrocardiogram (ECG) data. The EEG data were recorded from scalp sites F7, Fz, Pz, T5, and O2 of the 10/20 electrode system using an Electrocap. The EEG frequencies recorded were: Delta 2.0 to 4.0 Hz, Theta 5.0 to 8.0 Hz, Alpha 9.0 to 13.0 Hz, Beta 14.0 to 32.0 Hz, and Gamma 33.0 to 43.0 Hz. Wilson’s research showed that augmentation (slowing target velocity and displaying vehicle health task messages) based on an ANN classifier resulted in a “50% improvement in performance” [29]. Research conducted by Wilson et al proposes a meaningful way to label physiological data and also provides support to the notion that using an ANN to train and classify physiological data will result in high classification accuracy. In Wilson’s study, he was able to achieve classification accuracy levels above 90% labeling his data based on performance (High Performer, Low Performer).

Amarasinghe et al proposed a novel methodology to recognize thought patterns using Self Organizing Maps (SOM) for unsupervised clustering of raw EEG data and a feed forward ANN for classification [6]. The EEG frequency data was converted to the time domain using Discrete Fourier Transformation, which enabled segmentation of the EEG data with respect to the five frequencies that exist in brain signals (Alpha, Beta, Gamma, Delta, and Theta). The study was used on 5 participants to identify two different thought patterns; “move forward” and “rest”. These thought processes represented the brain signals used to control a virtual 3D GUI controlled by the participant. Amarasinghe et al proposed a methodology where the SOM clustered the processed EEG frequency data, and then passed it along to the ANN for classification. Average classification accuracies for the SOM and ANN after 5 participants were 96.6% and 88.4% respectively. This research is unique due to the novel labeling method used to identify the EEG frequency data samples. Instead of manually labeling each data sample, Amarasinghe et al trusted the accuracy of the SOM to label the data correctly. This may have contributed to lower classification accuracy for the ANN due to improper labeling of the EEG data.

From a medical standpoint, EEG frequency data is widely explored to help doctors understand the brain activity and what it reveals about the human state. Almahasneh et al proposed Singular Value Decomposition (SVD) for EEG feature extraction and then Support Vector Machines (SVM) for accurate detection of the participant’s cognitive state [30]. The use of SVD in Almahasneh’s research focuses only on EEG data related to the changes of driver cognitive distraction which he found made it very efficient to investigate the driver cognitive distraction. For each participant (42), after collecting 128-channel EEG data from each session, the EEG data was first preprocessed using low-pass and high-pass filters with cut-off frequencies of 0.5 Hz to 50 Hz to remove the line noise and high frequency noise. Then, the data from each subject for each session has been filtered by a Chebyshev band-pass filter of order 6 in order to extract the EEG frequency bands. Almahasneh hoped to develop a system that was able to detect changes in cognitive state when a driver was "Driving with Distraction" and "Driving" using EEG data exclusively. Support Vector Machine (SVM) classifiers from Waikato Environment for Knowledge Analysis (WEKA) were used for classifying the data into distracted and non-distracted classes. Using SVM and SVD, Almahasneh et al were able to achieve an average classification accuracy of 96.78% of cognitive state.

Table 1. Multi-Layer Perceptrons and their use in classification

Researcher	MLP Structure (Input Layer-Hidden Layer-Output Layer)	Best classification Accuracy	Algorithm used	Classification Labels
Fong	5-5-3	86.8%	Feed-forward-back propagation	Low Work level, Medium Work Level, High Work Level
Correa	12-20-1	86.5%	Levenberg-Marquardt back propagation	Awake, Drowsy
Ebrahimi	12-8-4	94.9%	Feed forward – back propagation	Stage 1-4 of NREM sleep
Wilson	37-37-2	89.7%	Not reported	Easy, difficult tasks

Section 4: Summary

There is no definitive evidence regarding the workload predictive ability and performance classification ability of EEG frequency data at the individual frequency band level using machine learning. The most closely aligned research was conducted by Belyavin et al and explored the efficacy of each individual EEG frequency band to predict workload, but took an Electrical Engineering perspective to analyze them. The Alpha, Beta, and Delta and Theta EEG frequency bands are the most referenced in research, but the reported behavior of the frequency bands may have been exclusively driven by the activities of the task given to the subject. The ANN was shown to be the best suited structure to find relationships between uncorrelated data when predicting workload and performance classification. Finding the best EEG frequency to predict workload and classify performance would decrease computational time for augmentation algorithms and reduce augmentation algorithm complexity. Identifying an EEG frequency band best suited for classification and prediction would provide a better means of augmenting based off of EEG data when needed and bolster existing augmentation algorithms used today.

III. Methodology

This research analyzes the efficacy of using Electroencephalography (EEG) data for two purposes: 1) to predict workload and 2) to classify performance of human subjects. The main goal of this research is to find the best subset of EEG waveforms which predict workload and task performance. We assume that when using an Artificial Neural Network (ANN) the waveforms that are most predictive of cognitive workload will be those which have the best classification and regression results.

This section also explains the methods used to determine whether changes in objective workload are associated with changes in power in the individual EEG frequency bands and in the nodes within them (F7, Fz, F8, T8, T7, Pz, O2, T8). To better determine the plausibility of this hypothesis, a Canonical Correlation analysis between each node in each EEG frequency and workload value was completed. Completion of the steps explained in this methodology will allow us to determine which subset of EEG wavelengths best classifies performance and predicts workload. An ANN will be used to evaluate the ability of individual and combined EEG data to predict task performance and predict objective workload values. MATLAB 2014a was used to evaluate the ANN in both the classification experiments and the workload prediction experiments.

Introduction

This section describes the methods used for the analysis presented in this thesis. Section I presents a description of the experiment used to generate the HUMAN Lab data set. Section II, III, and IV describe the methods used for performance classification on 10 subjects, dual classified subjects and using novel scenario data. Section IV describes the methods used for workload prediction. Section V describes ANOVA, Canonical Correlation, and K-Fold Cross Validation respectively.

711 HPW/RHCP Human Lab Formal Study 1 Experiment Description

AFRL's Human Universal Measurement and Assessment Network (HUMAN) laboratory's first formal study used a virtual remotely piloted aircraft (RPA) program called Vigilant Spirit [9]. Over the course of six days, participants experienced two training sessions and four data collection sessions, each with four trials. Every trial had a primary task which consisted of a surveillance phase followed by a tracking phase, with secondary communications task which consisted of answering cognitive questions throughout the entire trial. Each trial followed a scripted time-line (Figure 2, Comprehensive Timeline from Human Lab Study 1), lasting a total of seventeen minutes.

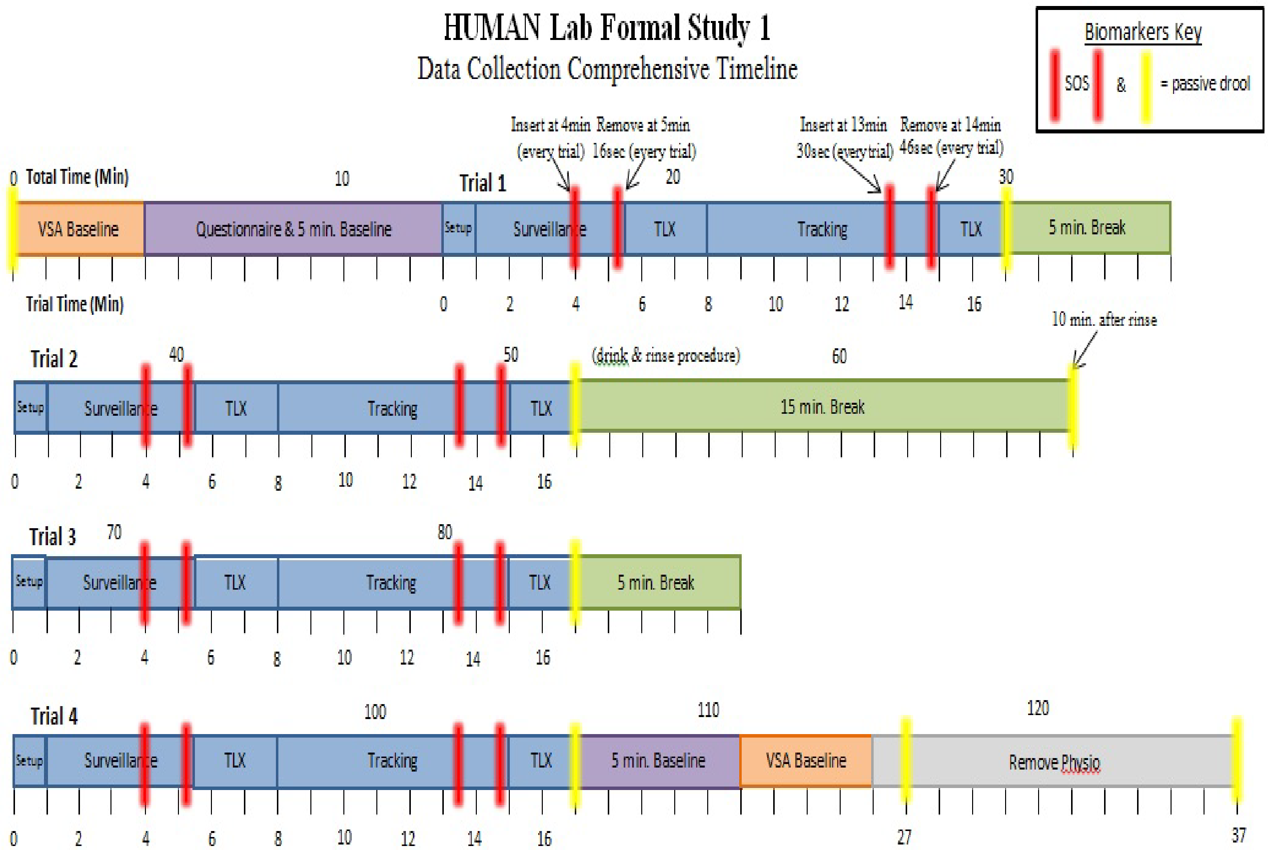

Figure 2. Comprehensive Timeline from HUMAN Lab Study

A trial begins with one minute allowed for taking control and setting up the RPA. This is followed by the four and a half minute surveillance phase, the goal of which is to monitor the market and attempt to locate the four high value targets, one at a time (HVTs). Each HVT carries a rifle (AK47), but irrelevant personnel (non-HVT distractors) are also present. They may carry a handgun, shovel, or nothing. The subject executes the surveillance mission by continuously operating the RPA camera, with the goal of following the HVT. Subjects search the market using the RPA camera by clicking where they want the RPA camera to center, while zooming the camera's field of view with the mouse scroll wheel. This technique enables the subject to determine whether a person was one of the HVTs or a distractor. Once the subject finds an HVT, (indicated by pressing F key on keyboard), the HVT is tracked until the target walks under one of twenty tents in the market, at which point the participant begins looking for the next HVT. Independent variables in the surveillance phase include the number of distractors (high or low), and sensor fuzz (either absent or present). When a target is found, and points are accrued as long as the target is visible in the simulated field of view of the user-controlled UAV camera (4.0 points per second tracked).

At the completion of the surveillance phase, participants have three minutes to complete the NASA-TLX subjective workload questionnaire. The seven minute tracking phase then begins. Thirty seconds into the tracking phase, the first HVT walks out from underneath a tent and walks to a different tent where he mounts a motorcycle. The participant attempts to track the HVT as it leaves the market on the motorcycle and rides to a new location. In half of the trials, a second HVT leaves in a similar manner, thirty seconds after the first, and must also be tracked. If an HVT is lost, participants are instructed to zoom out and search the surrounding area in order to reacquire the HVT. In half of the trials the HVTs travel along city roads and in the other half they travel along country roads. Independent variables in the tracking phase include number of HVTs (one or two) and route (city or country). When the tracking phase ends, participants are asked to fill out a final NASA-TLX questionnaire and given two minutes to do so. In the Tracking task, points are accrued in a similar manner, except points are accrued differently depending on the optical zoom the participant is able to track the target (High zoom = 1.429 points per second, Low Zoom = .715 points per second).

The secondary task occurs concurrently with the primary task. The secondary task uses the Multi-Modal Communications (MMC) tool to present participants with four questions at one minute intervals during both the surveillance and tracking phases, for a total of eight questions per trial. These operationally relevant questions require mental computations to calculate time and altitude values based on differing levels of distance and speed. Participants respond verbally with a push-to-talk space-bar while simultaneously continuing their primary task. During the entire 17 minute script, time-series EEG signal data (as well as other physiological data) is recorded and stored to measure the brain's electronic response to different tasks.

Participants were automatically scored on a scale from 0 to 1000 based on the subject's ability to complete the given RPA Surveillance and Tracking tasks. The rate at which a subject earned points increased or decreased in relation to their performance in a given task. For example, tracking a High Value Target (HVT) at magnification level two compared to tracking at lower zoom levels earned the participant 2.857 points per second and 1.429 points per second respectively.

The human experiment generated a dataset that we use. The goal of our research is to determine if EEG data measured from the brain activity can be used as an input to the ANN to classify high performing subjects and low performing subjects based on data labels derived from performance scores. Similarly, we want to see if the same EEG data can be used to predict the workload values derived from the IMPRINT program.

Performance Classification

For the classification problem we attempt to classify our subjects based on task performance similar to Wilson et al, which used the two classes “High Performer” and “Low Performer”. The subjects’ final score over the course of 16 scenarios were averaged and 2 categories (High Performer or Low Performer) were created that were based on the subjective need to augment performance or not. 10 subjects were selected for classification for both the Tracking and Surveillance task (5 High Performers, 5 Low Performers). For the Tracking Task, subjects who had an average score over 16 trials greater than 900 points were labeled as “High Performers” and subjects who scored less than 900 were labeled as “Low Performers” (See Table 2. Performance Data Label). For the Surveillance Task, subjects whose average score over 16 scenarios was greater than 600 were labeled as “High Performers” and subjects who scored less than 600 points were labeled as “Low Performers”. No subject had an average score greater than 900 for the Surveillance task, so a different performance threshold had to be used for the Surveillance tasks. These thresholds allowed for maximum classification difficulty for the ANN because it will be presented with an equal amount of High and Low performers.

Classification accuracy of the Artificial Neural Network (ANN) was recorded to determine the best method available to classify subjects. To measure the classification accuracy, Alpha, Beta, Gamma, Delta and Theta EEG frequency data were used as control variables and the corresponding response variable was the output from the ANN. Classification accuracy can be defined using the equation, $Hit\ Ratio = w/t$, where w is the class identified correctly and t is the total amount of data samples.

Table 2. Performance Data Label for Surveillance and Tracking Classification analysis. Data is labeled with either a' 0' or '1'. (S = Subject)

Task	**Performance Threshold (Avg.)**	**Subjects in Class**	**Label**
Surveillance **Tracking**	a. > 600 b. > 900	a. S2, S5, S7, S9, S10 b. S2, S5, S8, S9, S10	"High Performer" [0]
Surveillance **Tracking**	a. < 600 b. < 900	a. S4, S6, S8, S14,S12 b. S4, S6, S7, S11, S13	"Low Performer" [1]

Due to the wide frequency range of the Gamma EEG frequency band (30-100 Hz) the Gamma frequency was broken up into 3 parts; Gamma 1, Gamma 2 and Gamma 3. Each sub-frequency of the Gamma Frequency band was collected from seven scalp nodes (F7, Fz, F8, T8, T7, Pz, O2, T8), resulting in 21 feature values per time-step. To ensure the classification results of the original Gamma frequency representation are not skewed due to the detailed representation (Gamma 1, Gamma 2, Gamma 3), a separate trial was run using a Gamma frequency kept as one frequency band with only one set of seven features (F7,Fz,F8,T8,T7,Pz,O2,T8) per time-step.

A series of tests varying the number of neurons in the hidden layer (See Table 3. Test Matrix for Effectiveness of ANN) were completed to find the best configuration for the classification problem. Analysis of the ANN classification performance, using 10, 25, and 50 neurons in the hidden layer, revealed that using 50 neurons in the hidden layer was suitable for the classification and workload problem presented in this thesis. Error was extremely close to zero for all structures and using 50 neurons as opposed to 25 neurons in the hidden layer resulted in a classification accuracy decline of only 1%. A T-test was done to test the null hypothesis that the error using 50 neurons as opposed to 25 neurons in the hidden layer was equal when predicting task performance. The results of the T-test reveal we would fail to reject the null hypothesis and that the two structures are statistically similar with respect to task prediction error ($p = 0.9862$). Failing to reject this null hypothesis shows that the 1% decline in classification accuracy using 50 neurons in the hidden layer is negligible due to the statistical similarity of the error seen using either 25 or 50 neurons in the hidden layer.

In an extremely similar study done by Wilson et al classifying High and Low Performers in an RPA task, neurons in the hidden layer were not decreased beyond that of the number of neurons in the input layer [29]. Ebrahimi et al found that, "By varying the number of neurons in the hidden layer, it was observed that with increasing the number of neurons the mean of accuracy increases and standard deviation decreases" [24]. Research linked to ANN structure also validates the use of increased size of the hidden layer with respect to the input layer. "The network acquires a global perspective despite its local connectivity, due to the extra set of synaptic connections and the extra dimension of neural interactions" [31]. Increasing the number of neurons in the hidden layer give the ANN increased flexibility because there are more parameters the ANN can optimize [32, 31]. Our biggest test case will have an input size of 49 features using all EEG data combined as input. A well-performing ANN will have most error as close to zero as possible. Table 4 shows the error of the ANN at each fold in the cross validation process when predicting task performance.

Table 3. Test Matrix for Effectiveness of ANN

Parameter	Values
EEG Data	Gamma
Neurons in Hidden Layer	a) 10 b) 25 c) 50

Table 4. Results of pilot study varying the number of neurons in the hidden layer of the Artificial Neural Network. Mean Squared Error(MSE) is reported per validation stage along with correctly classified percentage of samples. Average from every K fold is reported as 'Average MSE'

	1st K-Fold MSE	2nd K-Fold MSE	3rd K-Fold MSE	4th K-Fold MSE	5th K-Fold MSE	Average MSE	Percentage Correctly Classified
50 Neurons	0.0181	0.013	0.0148	0.0152	0.0174	0.01570	96.10%
25 Neurons	0.0154	0.0155	0.0181	0.0152	0.0144	0.01572	97.20%
10 Neurons	0.0146	0.018	0.0234	0.0195	0.0161	.01832	96.90%

Performance Classification on Dual Classified Subjects

In the 711th HPW/RHCP HUMAN Lab experiments there were subjects that could be considered "dual classified". Dual classified means that their performance classification in the Surveillance task was that of a High Performer, while their performance in the Tracking task was that of a Low Performer, or vice versa. To validate that performance truly has an effect on EEG data, we will perform the same classification test mentioned previously in this thesis, but specifically on the same person, using EEG frequency data from a scenario where the participant struggled at one task, and flourished in the other (See Table 5. Dual Classified Subjects). Performance classification thresholds remained the same in these tests as those used in the first performance classification analysis that were completed (See Table 2. Performance Data Label) for Surveillance and Tracking performance classification analysis. The participants' physiological data was labeled according to their performance in the same manner according to Table 2. Performance Data Label for Surveillance and Tracking Classification analysis and the same ANN structure was used to predict task performance using EEG data.

Table 5. Dual Classified Subjects

Subject	Surveillance Classification	Tracking Classification
Subject 7	High Performer	Low Performer
Subject 8	Low Performer	High Performer
Subject 12	Low Performer	High Performer
Subject 14	Low Performer	High Performer

Novel Task Prediction on Dual Classified Subjects

To test the predictive ability of the EEG frequencies themselves, we will train the ANN on data from 15 scenarios, and evaluate its ability to predict the aforementioned task performance (High Performer, Low Performer) using only EEG data from the remaining scenario. The EEG frequency data will be labeled as a high performance scenario or low performance scenario according to the individual's average score over the 16 scenarios in the Surveillance and Tracking respectively. This labeling technique is different from the group comparison technique done in the initial task prediction analysis. Using the individual's mean provides a higher level of accuracy when labeling the EEG data because the threshold is set according to the individual's performance over 16 scenarios in a specific task and not the performance of the group. These tests will be completed on the dual classified subjects using each scenario (1-16) as a hold out set to test the ANN with after it has been trained on all other scenarios. Each scenario is used as a holdout set to test the ANN's ability to predict task performance using novel scenario data. Each Individual EEG individual frequency band, as well as All EEG frequency bands combined, will be used to test the efficacy of the ANN to predict task performance from novel scenario data. The performance of the ANN to predict novel task performance will be measured using the $Hit\ Ratio = w/t$ equation. The classification accuracy of the individual and combined EEG frequencies will be compared to an Uninformed Naïve Classifier. The Uninformed Naïve Classifier predicts task performance based on the proportion of High Performance Scenarios to Low Performance Scenarios per person. It then classifies each EEG data sample based on the higher of the two (High, Low Performer). Classification rate is calculated based on the number of EEG data samples correctly predicted as High Performance or Low Performance. The average classification accuracy of the ANN to predict task performance from each novel scenario per dual classified subject and EEG frequency will be reported. A One-Way ANOVA will be conducted on the average classification rate per dual classified subject, including the classification accuracy of the uniformed Naïve Classifier. This will test the null hypothesis that the Naïve Classifier is equal to the individual and combined EEG frequencies to predict task performance on novel scenario data.

Workload Prediction

Workload truth data was generated using a program called IMPRINT that uses a function to create objective values reflecting the workload the operator endured during the task. IMPRINT is a "dynamic, stochastic discrete event network modeling tool designed to help assess the interaction of warfighter and system performance throughout the system lifecycle—from concept design to field testing and system upgrades" [7]. Objective workload values were estimated for the Auditory, Cognitive, Fine Motor, Speech, Visual, and Overall workload. These values served as truth data for training and testing of the ANN. The goal of the workload prediction analysis is to see if accuracy is greater when using a single EEG frequency band, or when using all EEG Frequencies combined. Table *6*, "Factors and Levels for workload prediction analysis", shows the factors and levels that will be used for the workload prediction tests. During this phase of analysis, all EEG frequencies will be used as inputs into the ANN individually and combined (Alpha, Beta, Gamma, Delta, and Theta) to see how well they are able to predict the IMPRINT VACP Workload values. The RMSE is a commonly used general purpose error metric for numerical predictions. The average RMSE after a Five Fold Cross Validation process will be reported for the entirety of this thesis.

$$\text{Root Mean Squared Error (RMSE)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y' - Y)^2} \qquad (1)$$

where *Y'* = Predicted and *Y* = actual, will be recorded and compared to the RSME of a Naïve Predictor to gauge the prediction accuracy of the EEG frequency. In this thesis, the Naïve Predictor of workload randomly chooses a workload value at each time-step based on the distribution of the workload values seen in the workload truth data set.

Table 6. Factors and Levels for workload prediction analysis

Factor	**Levels**
EEG Data	Alpha, Beta, Gamma, Delta, Theta, All EEG Frequencies combined
Task	Surveillance Scenarios 1-16 and Tracking Scenarios 1-16
VACP Workload	Auditory, Cognitive, Fine Motor, Overall, Speech, Visual

ANOVA and Canonical Correlation

One Way Analysis of Variance (ANOVA)

Analysis of Variance (ANOVA) is a statistical technique used to analyze the means between two or more groups of data. A One-Way ANOVA can also be used as a form of hypothesis testing, where the *p* value calculated from the analysis of variance will help reject or fail to reject a given null hypothesis. If the *p* value (*p* = probability or likelihood of occurring) is less than the given significance level then it casts doubt on the plausibility of the null hypothesis and suggests that at least one sample mean from the group being tested is significantly different from the others. The default significance level for MATLAB's ANOVA is $p = 0.05$. If the *p*-value is greater than the significance level, we fail to reject the null hypothesis and assume the group means are equal. A One-Way ANOVA between the average classification accuracies per scenario was done to test the following alternative hypothesis: "Each EEG Frequency individually has a different task performance prediction (High Performers or Low Performers) accuracy than the others". This average was taken after completing a Five Fold Cross Validation process on the data set using the ANN. A probability reported from the ANOVA test greater than 0.05 would cause us to fail to reject the null hypothesis that: "The EEG frequencies are equal in their abilities to predict task performance".

Two-Way Analysis of Variance (ANOVA)

A Two-Way ANOVA will be completed on the average classification percentages of both Gamma EEG frequency representations. This test will be completed to determine how statistically similar the two Gamma EEG frequency representations are in their ability predict task performance. The two-way ANOVA compares the mean differences between groups that have been split on two independent variables also known as factors [33]. Factor, defined as “one of the elements contributing to a particular result or situation” [1], can be seen as the two Gamma EEG representations (3x7 features, 1x7 features) used as inputs to the ANN to predict task performance (dependent variable) and the scenario (1-16) of which task performance is predicted. The primary purpose of a two-way ANOVA is to understand if there is any interaction between the two independent variables on the dependent variable [33]. There are three null hypotheses that the Two-Way ANOVA will evaluate:

1. The population classification accuracy means of the two Gamma EEG frequency representations are equal for the Surveillance and Tracking tasks respectively.

2. The two Gamma EEG frequency representations are able to predict task performance equally across the 16 scenarios in both the Surveillance and Tracking Tasks respectively.

3. There is no interaction between the two Gamma EEG frequency representations and their ability to predict task performance per scenario in the Surveillance and Tracking Tasks respectively.

If the p values from the ANOVA for hypotheses one and two are insignificant ($p > 0.05$), then there is no difference between the two representations, and one is essentially just as good as the other at predicting task performance per scenario (1-16). If there is no interaction between the groups, the factors can be considered as being statistically similar regardless of the levels of detail in the Gamma EEG representation (3x7 versus 1x7 Gamma EEG frequency representation). If interaction is present between the two groups, the effects of the level of detail (Gamma 1, Gamma 2, and Gamma 3) are not the same. As it relates to this thesis, if there is no interaction between the two groups of classification accuracies per scenario, there is no effect that reducing the amount of Gamma features has on its' ability to predict task performance. Similar to the One-Way ANOVA, if the *p* value is greater than the significance level, $p > 0.05$, then we fail to reject the null hypothesis. Conversely, if $p < 0.05$, we reject the null hypothesis in favor of the alternative hypothesis that the two Gamma EEG frequency representations are significantly different in their ability to predict task performance.

Canonical Correlation Analysis

Canonical Correlation Analysis is used to measure the relationship between two sets of variables. When doing regression analysis, linear correlation analysis tools like Canonical Correlation report how well multiple variables align with another. For example say we have two sets of variables $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2}, \ldots \mathbf{X_n})$ and $\mathbf{Y} = (\mathbf{Y_1}, \mathbf{Y_2}, \ldots \mathbf{Y_n})$, a linear combination of these two groups would be named **U** and **V**. U is a linear combination of the **X** variables ($\mathbf{U} = a_{11}\mathbf{X_1} + a_{12}\mathbf{X_2} + a_{1n}\mathbf{X_n}$) and **V** is a linear combination of the **Y** variables ($V = b_{11}\mathbf{X_1} + b_{12}\mathbf{X_2} + b_{1n}\mathbf{X_n}$). The Canonical Correlation of the *i*-th pair ($\mathbf{U}_i, \mathbf{V}_i$) is

$$p_i = \frac{cov(Ui, Vi)}{\sqrt{Var(Ui)var(Vi)}} \qquad (2)$$

A Canonical Correlation Analysis will report the correlation between the output power of each node (F7, F8, Fz, T3, Pz, O2, T4) in each EEG frequency and the workload seen in each VACP workload channel.

K-Fold Cross Validation

A Five-Fold Cross Validation process will be used to validate the results achieved by the ANN in both the workload prediction and performance classification analysis. Specifically in K-Fold Cross Validation is used in performance classification using 10 subjects in both the Surveillance and Tracking tasks, and on the dual classified subjects. Cross validation is a way of testing the accuracy of an ANN before using it in the real world. The cross validation process validates the accuracy of the ANN because the ANN is being tested on data that it hasn't been exposed to in its training epochs. Figure 3 shows a conceptual view of how the data is structured for performance classification analysis with data from the Tracking tasks (Scenario 1-16).

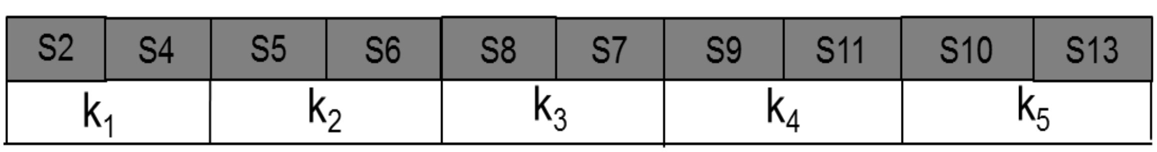

Figure 3. 5 Fold Cross Validation Diagram for Tracking task performance classification analysis. Figure 3 is a Graphical representation of how ANN will separate the data to validate the classification accuracy of the ANN in each fold of the 5 Fold Cross Validation Process. (S2 = Subject 2, k_1 = 1^{st} fold)

In each cross validation stage, one fold worth of data (k_1, k_2, k_3, k_4 or k_5) is held out from the training set. All other folds that weren't held out are used to train the ANN for a certain amount of epochs as designated by the algorithms creator. When all training epochs have finished, the k-fold that was held out is used to test the accuracy of the ANN. The average classification rate using each k-fold is reported and averaged over the *k* folds. This method will be used to report the accuracy of the ANN in both the performance classification and workload prediction analysis.

Methods for performance classification evaluation

Histograms will be used to visually depict the classification accuracy of each Frequency band for every task and scenario. The X axis will indicate the EEG frequency used as input to the ANN and the Y-axis will indicate the percentage of samples correctly identified or predicted. An ANOVA on the mean classification ratios per Task and EEG data frequency will be done to reject or fail to reject the null hypotheses presented in the Introduction of this thesis.

IV. Analysis and Results

Performance classification with individual and combined EEG frequency bands

The hit-ratio computed from correctly classified data samples was recorded for each scenario and each EEG frequency and tabulated over all scenarios and tasks. The average classification accuracy after a 5 Fold Cross Validation process using each respective EEG frequency to predict task performance is shown in Figure 4. The results show (Figure 4) that over the 16 scenarios in both the Tracking and Surveillance tasks, performance classification was better when using Gamma frequency-band EEG features than when using any other individual EEG frequency band. The Delta EEG frequency data was the worst input to the ANN classifying less data samples consistently over the course of the 16 scenarios in both the Surveillance and Tracking tasks. In literature, it is said that the Delta frequency is generally active in subjective cognitive states when the subject is in a deep, dreamless sleep, or unconscious. This EEG frequency (Delta) has most been associated with non-REM sleep, periods where there is a lack of movement, and low-levels of arousal. The classifier used in the research presented in this thesis may have struggled to classify individuals using this particular EEG frequency (Delta) because the task required some level of constant alertness [27, 34].

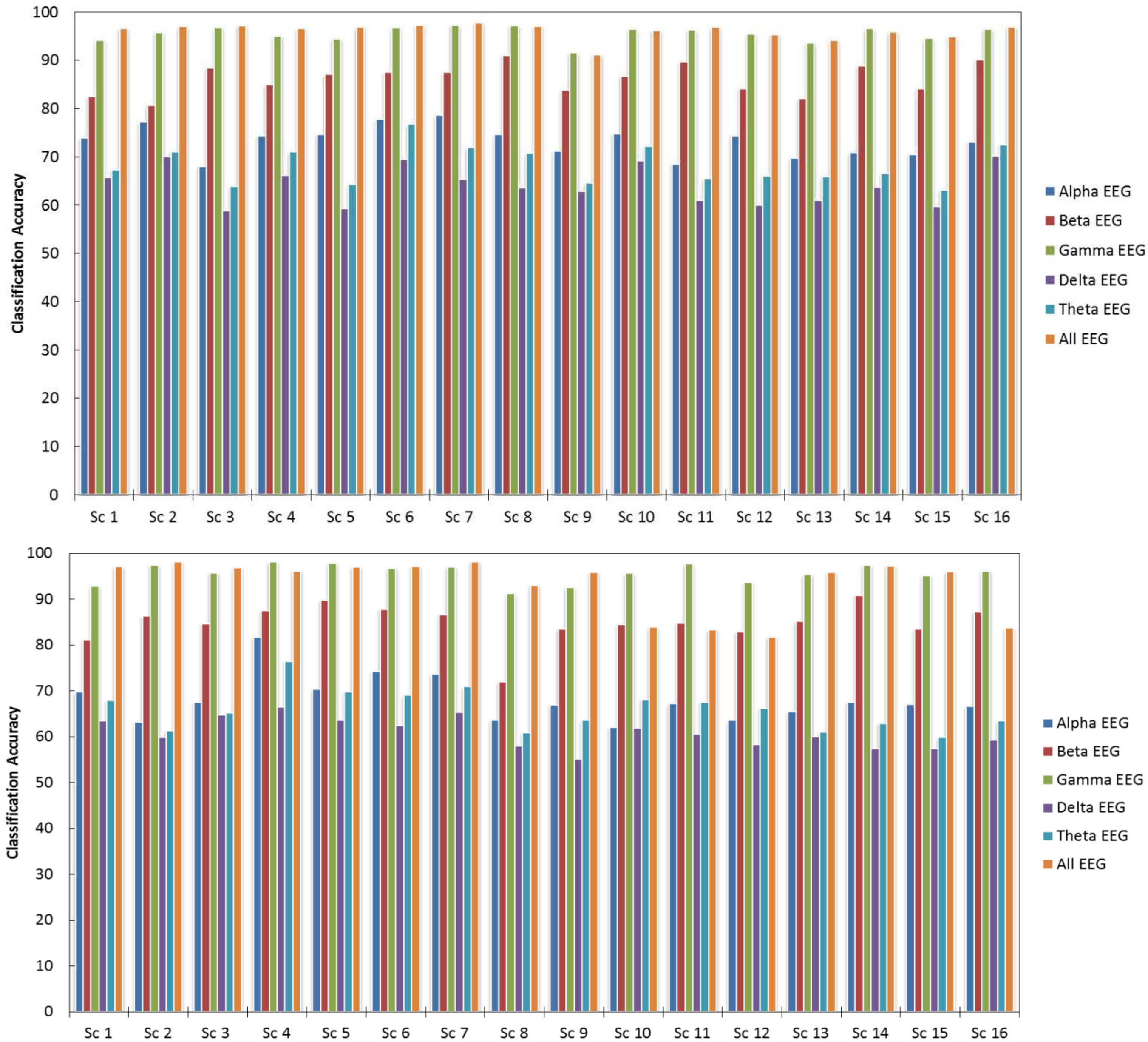

Figure 4. (top Tracking task, bottom Surveillance task) Shows the classification accuracy of the EEG frequencies and their ability to predict task performance individually across the 16 scenarios (Sc) in both the Tracking and Surveillance tasks.

Interestingly, the Beta EEG frequency was close in its ability to predict task performance with an average hit ratio of 84.565% over the course of 16 scenarios in both the Tracking and Surveillance tasks. Gamma EEG data had an average classification hit ratio of 94.495% over the 16 trials in both the Tracking and Surveillance tasks. This is compared to a uniform naïve predictor that achieved average classification accuracy of 50.887% and 50.789% over the 16 scenarios in both the Surveillance and Tracking tasks respectively. The uniform naïve predictor chose a random classification per time-step based on the likelihood of the performance classification (high or low performer). Gamma's ability to predict task performance data labels is significantly better than the Alpha, Delta and Theta EEG Frequencies.

The high classification accuracies of the Gamma and Beta EEG frequencies could be because they more closely align with brain activity that would be utilized during the Surveillance and Tracking tasks in the HUMAN Lab experiments. The Beta EEG Frequency has been associated with general activation of mind and body functions. In the medical domain, focused study of EEG frequency data breaks the Beta EEG frequency into three parts; Low Beta (12-15 Hz), Midrange Beta (15-18Hz), and High Beta (above 18 Hz) [34]. Low Beta can be detected anywhere on the cortex and is considered the "Sensorimotor Rhythm" or "SMR". Sensorimotor is defined as "of, relating to, or functioning in both sensory and motor aspects of bodily activity [1]. Midrange Beta is associated with subjective cognitive states such as "thinking, aware of self & surroundings". High Beta has been associated with feeling states such as alertness and agitation [34]. Similarly, the Gamma frequency has been associated with cognitive states such as thinking and integrated thought, and thought to indicate periods of high level information processing [34].

It seems intuitive that these two frequency bands would be more indicative of task performance as they closely align with the implied cognitive requirements of the Surveillance and Tracking task. Conversely, the Delta, Theta, and Alpha EEG frequencies have been associated with cognitive states of non-Rapid Eye Movement (REM) sleep, drowsiness, and meditation respectively [27, 34]. These actions do not correlate with the level of consciousness and focus required to be successful in a Surveillance or Tracking task. A classifier constructed to detect differences in EEG data based on levels of performance may struggle to do so using EEG frequency bands such as these (Alpha, Delta, and Theta). But, this may explain why the same classifier built to identify differences in EEG based on performance in the two tasks did so well using the Beta and Gamma EEG frequency data as inputs.

A One-Way ANOVA was used to test the hypothesis, "Each individual EEG Frequency is equal to one another in their ability to predict task performance (High Performers or Low Performers)". The test was run on a matrix containing the average classification hit ratios of all the EEG frequencies (Alpha, Beta, Gamma, Delta Theta) revealing that all EEG frequencies individually do not have the same ability to predict task performance (Surveillance- $p = 1.7730e\text{-}34$, Tracking – $p=8.1339e\text{-}40$). This supports the notion that, there are significant differences in the individual frequencies ability to predict task performance. We would reject the null hypothesis stated above in favor of the alternative hypothesis that there is at least one EEG frequency that is significantly different from the others at predicting task performance.

A test was also run to see how well all EEG frequencies combined as input to the ANN (Alpha, Beta, Gamma, Delta, Theta) could predict task performance. Using all EEG frequencies to predict task performance resulted in classification accuracy greater than 90% with only 4 instances less than 80% (Scenario 10, 11, 12 and 15; Surveillance Task). A One-Way ANOVA on the classification percentages after a 5 Fold Cross Validation on the data was done to test the null hypothesis, "Using All EEG frequencies combined as input to the ANN will result in equal classification accuracy in each scenario in the Surveillance and Tracking tasks" (Surveillance p = .9754, Tracking p = .7642). This means that using All EEG frequencies combined to predict task performance is consistent over all scenarios in the Tracking and Surveillance tasks.

Experiment facilitators with the 711th HPW/RHCP HUMAN LAB reported the Gamma frequency (See Table 7) in three, seven-feature scalp-node observations (F7, Fz, F8, Pz, T7, T8, O2) as opposed to one, seven-feature observation like the rest of the EEG frequencies. Representing the Gamma frequency in three parts (Gamma 1, Gamma 2, and Gamma 3) allowed the researchers with 711th HPW/RHCP to represent the Gamma frequency with a greater level of detail. The raw Gamma EEG frequency band was re-filtered to represent 1x7 sub-band features (F7, Fz, F8, Pz,T7, T8, O2) in the same way that the 711th HPW/RHCP Human Lab experiment facilitators filtered the other EEG frequencies (Alpha, Beta, Delta, Theta).

Table 7. EEG Frequency Bands and Their Frequency Ranges

EEG Frequency	Frequency Range (Hz)
Alpha	8-15
Beta	16-30
Gamma	30-100
Delta	0.1-4
Theta	4-6

The same classification experiment was run using the 1x7 feature Gamma EEG frequency data and compared to the 3x7 feature Gamma EEG frequency data to see which representation was most advantageous for classification. The results show only a slight decline in the 1x7 Gamma EEG frequency's ability to predict task performance. Specifically, there were 2 instances (Scenario 9, Surveillance and Tracking, see Figure 5) where the 1x7 feature Gamma EEG frequency data classified below 90%. In comparison, the 3x7 feature Gamma EEG frequency data had no scenarios where it classified with less than 90% accuracy. A Two-Way ANOVA was used to test the hypothesis, "Both Gamma EEG frequency representations are equal in their ability to classify based on performance" (See VII. ANOVA and Canonical Correlation for all 3 hypotheses). This analysis was run on a matrix containing classification percentages using the 3x7 Gamma EEG frequency representation and the filtered 1x7 Gamma EEG frequency representation from the Surveillance and Tracking tasks respectively.

Table 8. Two-Way ANOVA results of Gamma Classification accuracies.

Task	**Two-Way ANOVA results**
Surveillance	1. Gamma representations equal: p = 0.4120 2 Scenario classification equal: p = .2564 3. Interaction: p = .9981
Tracking	1. Gamma representations equal: p = 0.1395 2 Scenario classification equal: p = 0.2793 3. Interaction: p = .8350

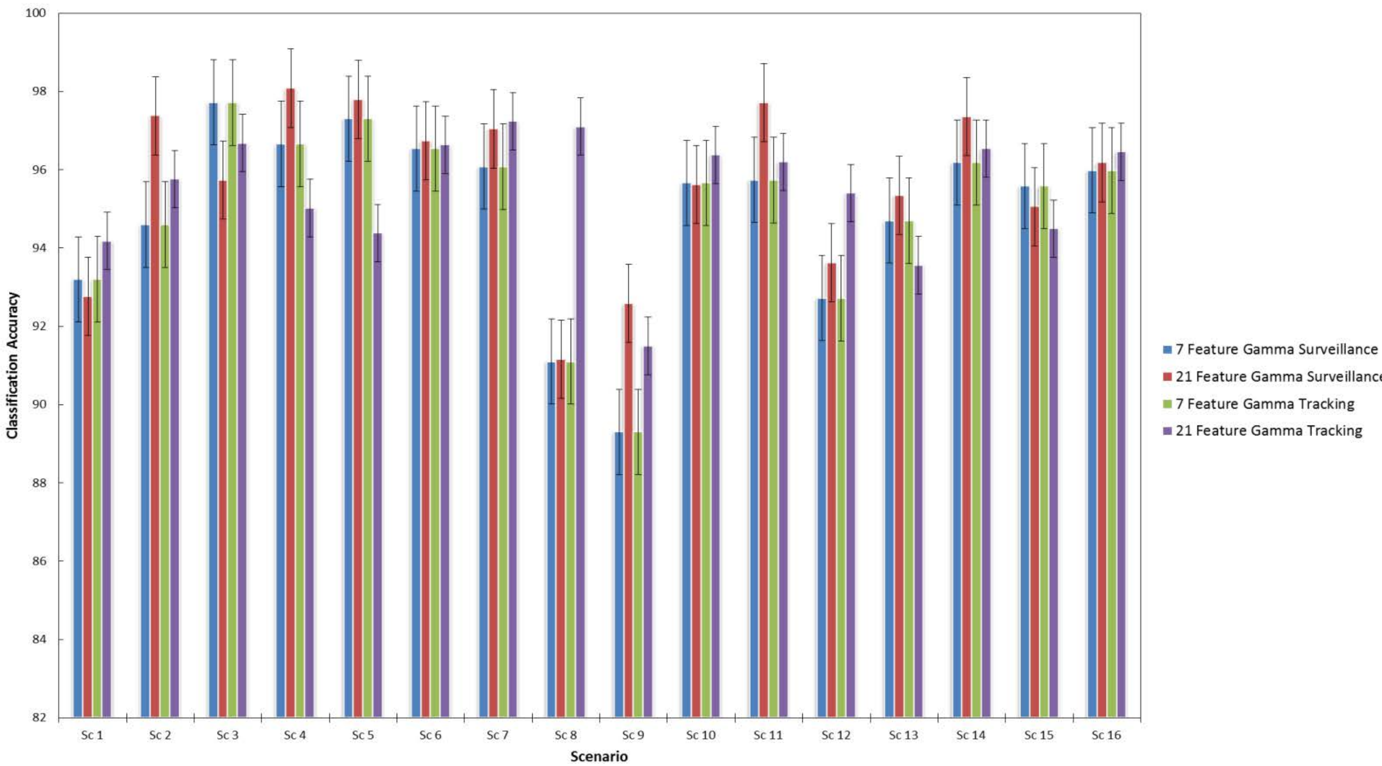

***Figure 5. 1x7 and 3x7 Gamma EEG frequency classification accuracy across 16 Scenarios in both Surveillance and Tracking Tasks*. (Classification accuracy on 10 subjects using 95% confidence intervals)**

Results show that we would fail to reject the null hypothesis, "Both Gamma EEG frequency representations are equal in their ability to classify based on performance". In the both the Tracking and Surveillance tasks, the probability for the classification accuracy per scenario and Gamma EEG representations being equal is greater than the significance level, $p > .05$. Reducing the amount of features in the Gamma EEG frequency representation has no statistically significant effect on its ability predict task performance (see Table 8. Interaction: Surveillance = 0.998, Tracking = 0.8350). The lowest the classification accuracy dropped to in the 1x7 feature Gamma EEG frequency representation was 87% in the Tracking task. Filtering the Gamma EEG frequency data to represent one set of seven features did hinder its ability to classify above 90% across both tasks and all scenarios, but was not statistically significant enough to cause us to reject our null hypothesis. When interaction is absent, as it is in the Two-Way ANOVA results between the two Gamma EEG frequency representations, the effects of the representations can be seen as being statistically similar.

Classification based on performance using EEG frequency data sheds more light on the use of EEG frequency data, and how it can be used in combination with machine learning. The results presented in this thesis regarding classification based on performance show that it is possible to use machine learning to classify based on thresholds defined by performance using only EEG data. They also show that it is possible to predict task performance using one EEG frequency alone and all EEG frequencies combined with a high level of classification accuracy.

Using the 3x7-Feature Gamma EEG frequency alone to classify individuals in the Surveillance and Tracking task resulted in greater than 90% classification accuracy in both tasks. This alludes to the fact that the Gamma EEG frequency (21 features) could be used exclusively to predict task performance in a system designed to detect a low performing individual, instead of all EEG frequencies combined (49 features).

Performance Classification on Dual Classified Subjects

Performance classification on dual classified subjects revealed that there was a change in the EEG frequency in situations where the individual struggled or excelled in a task. Specifically, the Gamma EEG frequency data was more consistent in classifying individuals based on performance with greater than 90% classification accuracy. The Beta EEG frequency was second best at identifying this change in performance within the individual, but was only able to do so in one instance with Subject 12, Scenario 1. This rare instance could be because of the great difference in performance with Subject 12. In Scenario 1 of the Tracking task the subject’s final score was a 943, where as in the Surveillance task their final score was 388.2. This range in score between tasks, but in the same scenario, consisted of 506.73 points and was the greatest variability seen amongst the dual classified subjects.

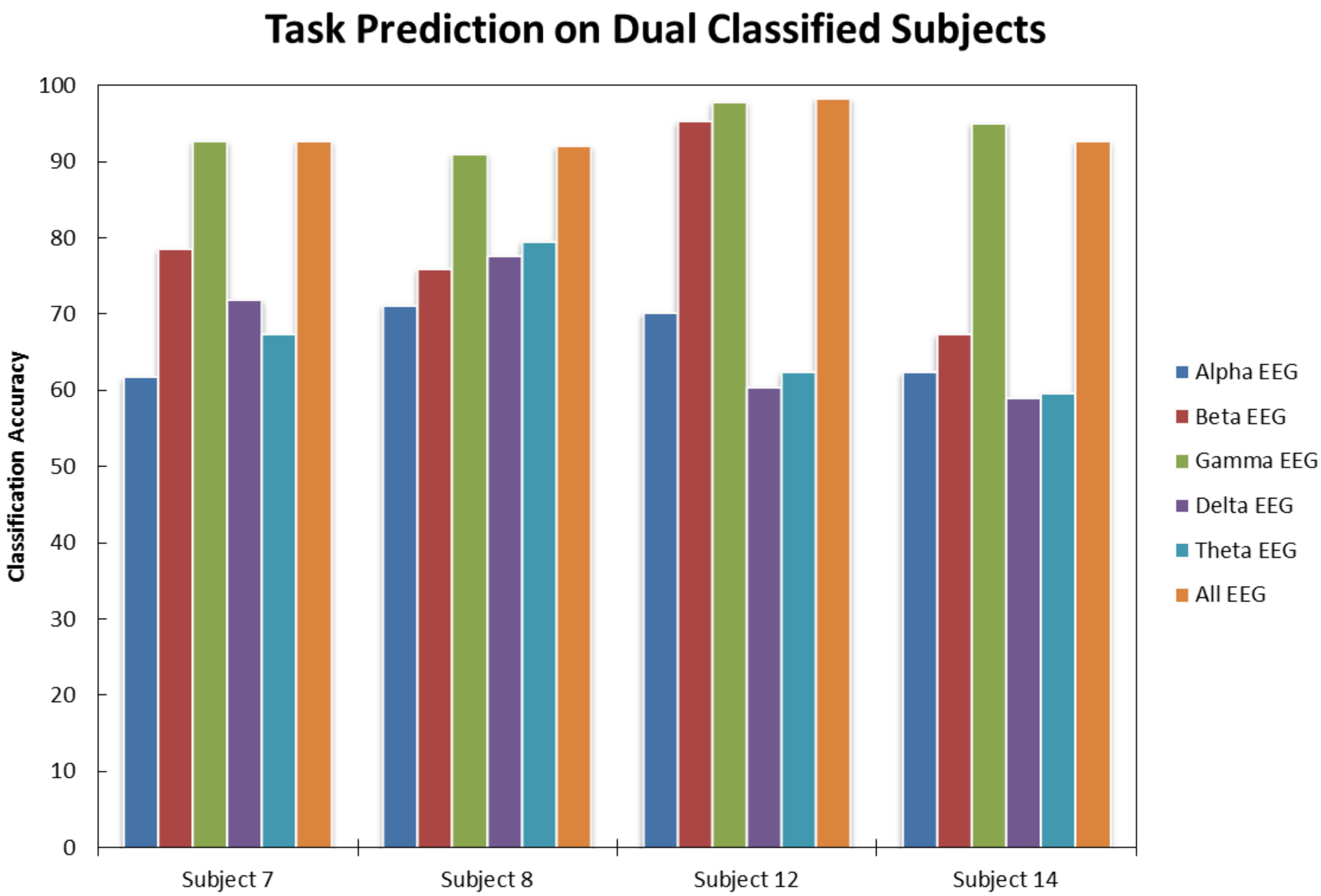

Figure 6. Task prediction of Dual Classified Subjects using EEG frequency data

Results presented of the performance classification analysis on dual classified subjects confirm results regarding power activity in the Gamma EEG frequency band. A One-Way ANOVA on the average classification accuracies of the individual EEG frequencies after the 5 Fold Cross Validation process per dual classified subject revealed no statistical similarities (p = 7.311e-04). This means that there is a difference in the EEG frequencies and their ability to predict task performance of dual classified subjects. It is widely believed in the neuroscience and psychology fields that oscillations in the Gamma EEG frequency range, specifically (30-70 Hz), are associated with basic aspects of brain functioning such as conscious perception, feature and temporal binding, attention, memory, and information processing integrated with related motor response (sensorimotor processing) [35, 36, 37, 38]. In an experiment where participants were required to perform tracking, wrist extension, and finger sequencing tasks, Aoki et al were able to show a peak in the (30-40 Hz) Gamma frequency range during the tracking task in all subjects. Results from his study indicate that gamma oscillations corresponding to sensorimotor tasks became synchronized across multiple node sites [37]. Experiments conducted by Yordavana et al and Struber et al were able to show that the spontaneous gamma activity when identifying cube reversals was greater at the frontal node sites and decreased in the anterior and posterior regions [35, 36]. Specifically, Gamma EEG frequency power was significantly greater at the left than at the right frontal sites.

These findings in prior research begin to explain the high classification accuracy seen in the performance classification trials and classification of the dual classified subjects. Gamma EEG has been shown to be highly responsive to activities requiring consciousness and arousal because it attenuates over the course of long term stimulation and disappears during deep sleep and anesthesia [39]. Figure 6 shows that an ANN trained on EEG frequency data from two tasks can delineate between high performance and low performance. Research from Yordavana et al and Struber et al suggest that the frontal regions of the brain are most sensitive to these changes in arousal and situational awareness that may be required in a tracking or surveillance task. Interestingly, three of the seven node features used to record the EEG frequency data in the HUMAN Lab study were that of the frontal region (F7, F8, Fz). The other four came from the Parietal, Occipital, and Temporal regions of the brain (T4, T3, Pz, O2), and made up less of the feature space than the Frontal regions. The situational awareness required from the Surveillance and Tracking tasks in the HUMAN Lab study, combined with the high ratio of frontal lobe region features and its sensitivity to sensorimotor processing could explain the high classification accuracy of the Gamma EEG frequency over the course of the performance classification analysis in dual classified subjects.

Novel Task Prediction on Dual Classified Subjects

Task prediction using novel scenario data resulted in poor classification accuracy when using both the individual EEG frequencies and combined EEG frequencies as input to the ANN. We would fail to reject the null hypothesis that the Uninformed Naïve Classifier and the EEG frequencies are equal in their ability to predict task performance (Surveillance: p = 0.915, Tracking: p = 0.724). The ANOVA results reveal that the ANN struggles to make accurate predictions per EEG data sample regarding task performance in novel scenarios after being trained on EEG frequency data and task performance results from other scenarios. This means that EEG frequency data is highly unique to the scenario the individual is participating in and is that changes in the EEG frequency bands is not generalizable to other scenarios.

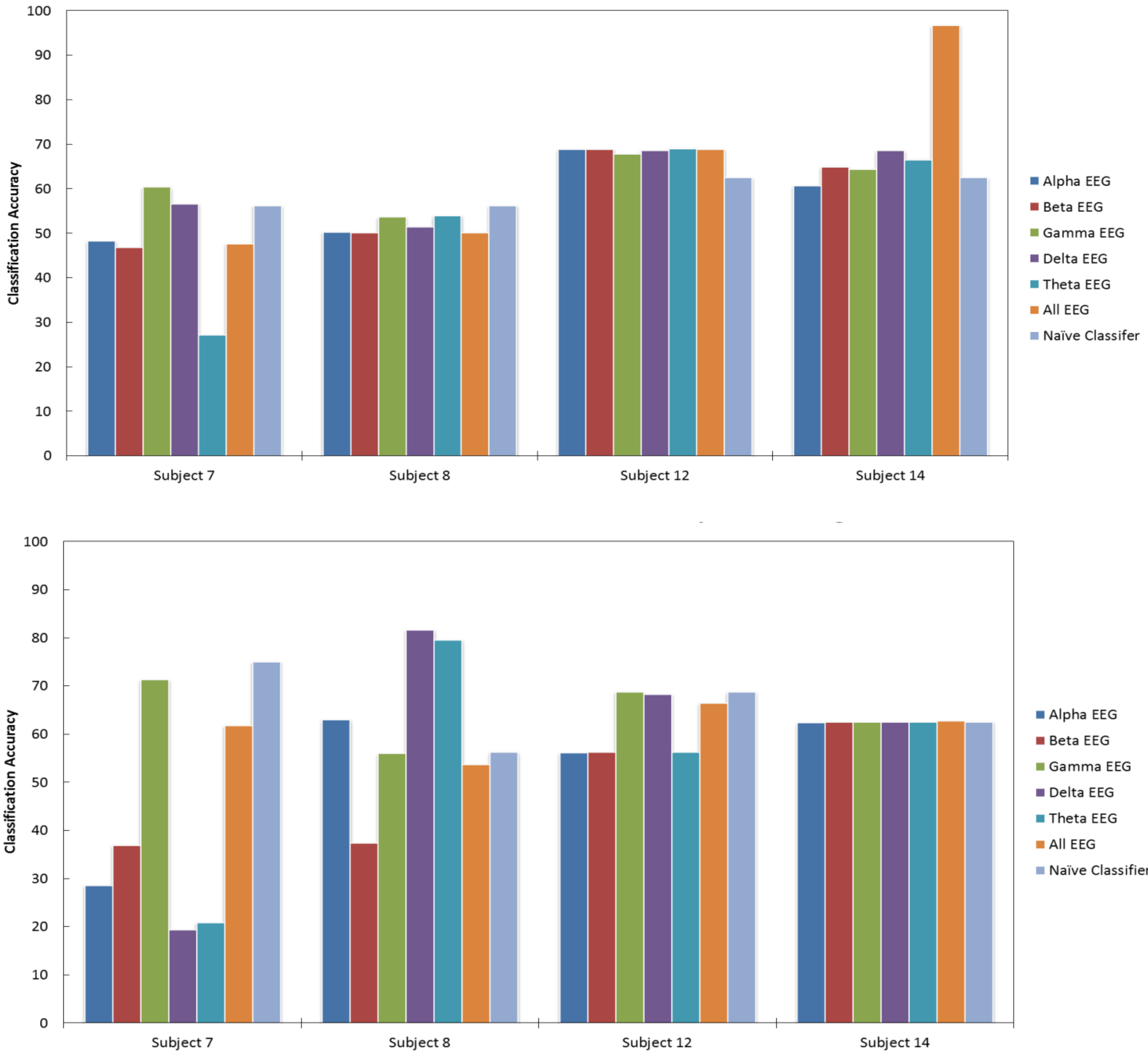

Figure 7. Classification Accuracy of Novel Task Prediction on Dual Classified Subjects (top Surveillance, bottom Tracking)

Predicting Workload using Individual EEG Frequencies and All EEG frequencies Combined

Workload Prediction was done using an ANN and accuracy of prediction was measured using the Root Mean Squared Error (RMSE). Table 10 shows the Root Mean Squared Error of both the Naïve Workload Predictor and the ANN when used to predict VACP Workload. Each column in Table 10 represents the individual VACP Workload Channel (Auditory, Cognitive, Fine Motor, Overall, Speech and Visual), while each row represents each Scenario (1-16) in the given Surveillance and Tracking tasks. Each row and column used in Table 10 shows the EEG frequency with the lowest RMSE used as input to the ANN when used to predict VACP Workload per scenario. Results show that the Delta EEG frequency had the most scenarios with the lowest RMSE over the 16 trials in both the Surveillance and Tracking tasks, while Beta had the least (See Table 9).

Table 9. EEG Frequency and number of Scenarios with lowest RMSE (32 Scenarios Tracking and Surveillance x 6 VACP Workload Channels)

EEG Frequency	Percentage of Scenarios with Lowest RMSE when used to predict workload
Alpha	23%
Beta	7%
Gamma	9%
Delta	41%
Theta	19%

Table 10.(a-f) (a.) Possible Truth Values in Workload Channel (b.) Surveillance Naïve Predictor (c.) Surveillance Best Predictor of VACP Workload indicated by color (d.) Tracking Naïve Predictor (e.)Best Predictor of VACP Workload indicated by color (f.) legend indicating color indicative of corresponding EEG frequency. Error of ANN is presented as RMSE.

Possible Truth Values in Workload Channel	
Auditory	0 , 6
Cognitive	0, 4.6 , 7, 11.6
Fine Motor	2.6 , 4.8
Speech	0 , 2
Visual	4.4 , 6
Overall	6, 7,11.6 , 13.2 , 15.8 , 17.4 , 17.6 , 18.6 , 19.2 , 20.2

Surveillance Naïve Predictor

	Auditory	Cognitive	Fine Motor	Speech	Visual	Overall
1	3.47	4.03	1.27	1.16	0.93	4.22
2	3.45	4.50	1.49	1.16	3.00	5.51
3	3.49	4.04	1.27	1.15	0.93	4.28
4	3.47	4.61	1.19	1.15	3.15	6.71
5	3.45	4.72	1.20	1.14	3.01	6.58
6	3.46	4.11	1.27	1.15	0.91	4.22
7	3.45	4.54	1.18	1.14	2.93	6.33
8	3.46	4.49	1.20	1.16	3.00	6.40
9	3.44	4.49	1.18	1.14	3.10	6.22
10	3.45	4.68	1.19	1.16	3.03	6.50
11	3.45	4.73	1.18	1.17	3.10	6.71
12	3.46	4.05	1.28	1.14	0.92	4.15
13	3.44	4.66	1.19	1.16	3.05	6.35
14	3.43	4.03	1.26	1.17	0.91	4.04
15	3.44	4.02	1.26	1.16	0.93	4.12
16	3.43	3.98	1.27	1.16	0.93	4.10

Surveillance Best Predictor of VACP workload

	Auditory	Cognitive	Fine Motor	Speech	Visual	Overall	Best EEG Freq.
1	1.79	2.97	0.46	0.41	0.74	2.98	Delta
2	1.80	2.67	0.48	0.39	0.83	3.75	Alpha
3	1.91	3.05	0.45	0.41	0.73	4.16	Delta
4	1.91	2.81	0.42	0.40	0.74	5.40	Delta
5	1.98	1.55	0.45	0.43	0.84	3.61	Alpha
6	1.83	2.84	0.44	0.39	0.78	2.87	Delta
7	1.66	3.27	0.46	0.44	0.92	3.88	Alpha
8	1.93	3.10	0.46	0.42	0.85	5.10	Theta
9	1.88	2.68	0.43	0.42	0.89	4.16	Gamma
10	1.52	3.21	0.44	0.42	0.86	3.86	Delta
11	1.87	3.05	0.42	0.43	0.91	4.24	Delta
12	1.74	2.71	0.46	0.40	0.79	3.84	Theta
13	1.74	3.04	0.45	0.43	0.89	4.00	Delta
14	1.89	1.98	0.45	0.40	0.76	4.25	Theta
15	1.90	2.50	0.46	0.42	0.82	3.03	Delta
16	1.93	2.13	0.47	0.43	0.78	3.40	Delta

Legend	Alpha	Beta	Gamma	Delta	Theta

Tracking Naïve Predictor

	Auditory	Cognitive	Fine Motor	Speech	Visual	Overall
1	3.14	4.85	1.99	1.06	3.37	8.79
2	3.16	4.56	1.77	1.04	2.84	8.11
3	3.18	4.92	2.08	1.05	3.37	9.11
4	3.09	5.17	2.08	1.06	3.38	9.13
5	3.20	5.01	2.08	1.05	3.32	9.09
6	3.13	4.95	1.97	1.04	3.20	8.82
7	3.20	4.97	2.08	1.06	3.40	8.36
8	3.18	4.66	1.90	1.05	2.99	8.35
9	3.22	4.88	2.00	1.05	3.12	8.63
10	3.14	5.12	2.18	1.06	3.55	9.12
11	3.18	4.63	1.84	1.04	2.98	8.10
12	3.14	4.65	1.90	1.04	3.06	8.28
13	3.09	4.82	1.94	1.04	3.16	8.57
14	3.21	4.80	2.02	1.03	3.21	8.29
15	3.15	4.67	1.88	1.05	2.97	8.24
16	3.16	4.92	2.13	1.05	3.42	8.86

Best Predictor of VACP workload

	Auditory	Cognitive	Fine Motor	Speech	Visual	Overall	Best EEG Freq.
1	1.67	2.08	1.35	0.40	1.81	7.61	Delta
2	1.63	3.06	0.90	0.39	2.20	8.31	Delta
3	1.72	3.09	1.44	0.39	1.52	6.44	Delta
4	1.65	2.51	1.55	0.40	2.67	7.73	Delta
5	1.51	2.34	1.16	0.34	2.02	8.11	Alpha
6	1.37	3.19	1.51	0.34	2.38	8.48	Theta
7	1.42	2.59	1.38	0.40	2.34	7.73	Delta
8	1.74	2.96	1.47	0.39	2.58	6.75	Theta
9	1.56	3.33	1.26	0.40	2.30	8.47	Theta
10	1.58	3.57	1.28	0.38	2.36	5.15	Delta
11	1.73	3.36	0.95	0.36	2.04	11.19	Alpha
12	1.33	3.17	1.34	0.42	2.21	11.35	Delta
13	1.69	2.45	1.26	0.38	1.60	8.54	Theta
14	1.35	4.06	1.36	0.39	2.10	9.56	Alpha
15	1.60	3.38	1.31	0.41	2.77	9.02	Delta
16	1.11	2.82	1.42	0.39	1.67	8.88	Delta

Contrary to the performance classification analysis trials, the Beta and Gamma were the first and second worst EEG frequencies to use as an input to the ANN to predict VACP workload (See *Table 10*. EEG Frequency and number of Scenarios with lowest RMSE). These rankings are justified by the small percentage of scenarios where Beta and Gamma had the lowest RMSE to predict the respective VACP workload channel (Table 9,Table 10). Table 10 shows the frequencies with wider ranges (Hz) are actually worse at predicting VACP Workload values and that it is much harder for the ANN to distinguish some relationship between the EEG input and desired VACP workload value as the EEG frequency range increases with size. A strong indicator of this notion is the poor performance of the ANN predicting VACP workload when ALL EEG frequencies are used as inputs (See Table 11).

Table 11. RMSE using Uniform Naïve Workload Predictor (Left) compared to the ANN using ALL EEG Frequencies Combined (Right). Red indicates High RMSE in comparison to the Naïve Predictor or a scenario where the Naïve Predictor actually did better than the ANN.

Surveillance Naïve Predictor

	Auditory	Cognitive	Fine Motor	Speech	Visual	Overall
1	3.47	4.03	1.27	1.16	0.93	4.22
2	3.45	4.50	1.49	1.16	3.00	5.51
3	3.49	4.04	1.27	1.15	0.93	4.28
4	3.47	4.61	1.19	1.15	3.15	6.71
5	3.45	4.72	1.20	1.14	3.01	6.58
6	3.46	4.11	1.27	1.15	0.91	4.22
7	3.45	4.54	1.18	1.14	2.93	6.33
8	3.46	4.49	1.20	1.16	3.00	6.40
9	3.44	4.49	1.18	1.14	3.10	6.22
10	3.45	4.68	1.19	1.16	3.03	6.50
11	3.45	4.73	1.18	1.17	3.10	6.71
12	3.46	4.05	1.28	1.14	0.92	4.15
13	3.44	4.66	1.19	1.16	3.05	6.35
14	3.43	4.03	1.26	1.17	0.91	4.04
15	3.44	4.02	1.26	1.16	0.93	4.12
16	3.43	3.98	1.27	1.16	0.93	4.10

Surveillance All EEG Combined

	Auditory	Cognitive	Fine Motor	Speech	Visual	Overall
1	2.23	4.43	0.71	0.60	1.27	10.37
2	2.85	4.70	0.68	0.59	1.26	5.06
3	2.14	4.31	1.32	0.65	0.91	4.39
4	1.94	5.37	0.55	0.48	1.72	8.55
5	1.86	6.24	0.60	0.45	1.92	7.42
6	1.99	4.90	1.02	0.48	1.02	7.51
7	1.83	6.46	0.72	0.50	1.37	7.90
8	2.08	5.15	0.75	0.44	1.40	9.36
9	2.72	6.77	0.98	0.86	1.21	9.70
10	2.86	5.70	0.51	0.53	1.70	6.85
11	2.78	8.37	0.46	0.58	1.15	15.56
12	2.05	4.52	0.59	0.50	0.88	5.09
13	2.66	5.98	0.73	0.57	1.53	8.14
14	2.02	3.99	0.67	0.63	1.80	10.79
15	1.89	3.65	0.68	0.63	0.80	5.01
16	2.93	5.14	0.64	0.55	1.47	10.36

Tracking Naïve Predictor

	Auditory	Cognitive	Fine Motor	Speech	Visual	Overall
1	3.14	4.85	1.99	1.06	3.37	8.79
2	3.16	4.56	1.77	1.04	2.84	8.11
3	3.18	4.92	2.08	1.05	3.37	9.11
4	3.09	5.17	2.08	1.06	3.38	9.13
5	3.20	5.01	2.08	1.05	3.32	9.09
6	3.13	4.95	1.97	1.04	3.20	8.82
7	3.20	4.97	2.08	1.06	3.40	8.36
8	3.18	4.66	1.90	1.05	2.99	8.35
9	3.22	4.88	2.00	1.05	3.12	8.63
10	3.14	5.12	2.18	1.06	3.55	9.12
11	3.18	4.63	1.84	1.04	2.98	8.10
12	3.14	4.65	1.90	1.04	3.06	8.28
13	3.09	4.82	1.94	1.04	3.16	8.57
14	3.21	4.80	2.02	1.03	3.21	8.29
15	3.15	4.67	1.88	1.05	2.97	8.24
16	3.16	4.92	2.13	1.05	3.42	8.86

Tracking All EEG Combined

	Auditory	Cognitive	Fine Motor	Speech	Visual	Overall
1	2.37	5.47	1.51	0.53	4.64	21.88
2	2.07	9.34	1.63	0.46	3.52	11.93
3	2.61	8.95	1.35	0.85	2.69	10.14
4	2.24	6.70	3.17	0.50	4.05	15.52
5	2.16	10.03	1.60	0.50	2.94	17.11
6	3.01	5.75	1.62	0.45	4.08	15.45
7	2.29	8.80	1.33	0.62	3.37	13.74
8	1.85	4.64	1.33	0.48	4.29	8.93
9	2.13	6.11	1.97	0.83	4.42	10.20
10	2.42	6.62	1.71	0.52	2.65	18.78
11	2.40	4.42	1.67	0.46	3.13	10.45
12	2.72	10.04	3.32	0.57	4.15	19.58
13	1.90	5.60	1.44	0.44	2.20	16.65
14	1.99	8.74	1.96	0.55	2.84	12.90
15	1.61	5.74	2.00	0.43	3.48	9.77
16	2.24	6.74	1.83	0.45	4.21	11.51

The highest RMSE reported after using all EEG frequencies as inputs was 21.88 in Scenario 1 Predicting Overall VACP workload. The RMSE was greater than 5 in all but one attempt at predicting Overall workload in the Surveillance Task (Scenario 3, See Table 11). Error predicting overall workload could mean the ANN is predicting workload when there is none, over predicting the amount of workload seen, or is grossly wrong predicting workload based on the EEG frequency input. Error when predicting Auditory or Speech workload channels workload is highly undesired. The ANN would actually be predicting that the subject is listening or speaking when he or she really isn't. This may result in triggering augmentation when it really isn't needed in cases where workload is over-predicted. Scenarios highlighted in red in Table 11 indicate high RMSE with respect to the Naïve Workload Predictor or scenarios that were higher than those of the Naïve Workload Predictor. Using all EEG frequencies combined (49 features) as input to the ANN seemed to hinder its' ability to predict VACP workload. These results allude to the fact that feature reduction would be beneficial to an ANN trying to predict VACP Workload using EEG frequency data.

A One-Way ANOVA between the RMSE from the ANN predicting workload and the Naïve Predictor was conducted. The One-Way ANOVA comparing the individual EEG frequencies to the Naïve Predictor used the lowest RMSE from the 5 EEG frequencies when predicting the particular workload channel per scenario. The results from the One-Way ANOVA show a statistical difference in the classification ability of the Naïve classifier and the ANN except when using All EEG combined to predict Fine Motor workload and Visual Workload (*p* = .3270, *p* = .0948 respectively(See Appendix B).

ANOVA on RMSE for full One-Way ANOVA results). This means that there is a statistical difference when using the EEG Frequencies as input to the ANN to predict workload and the Naïve Predictor. But, when predicting Fine Motor and Visual workload with the EEG frequency data, the Naïve Predictor is statistically similar to the ANN in doing so. Table 10 & 11 show that overall, the ANN is better at predicting workload than the Naïve Predictor. However, the RMSE seen when using EEG data to predict workload show this method does not facilitate accurate workload prediction (See Table 10. Possible Truth Values in Workload Channel).

Canonical Correlation Analysis between the EEG frequencies and the VACP Workload values revealed little to no correlation between the two (See Figure 12 and 13). There was almost no negative or positive correlation between the EEG nodes and the VACP workload values. Based on the poor workload prediction results (Table 10 and Table 11) and the lack of correlation between the EEG data and the VACP workload values (Table 12 and Table 13), we can conclude that the objective workload seen by the participant had no direct effect on the power in the EEG frequency bands. These results indicate that using EEG data in the form presented in this thesis do not facilitate accurate workload prediction. Therefore, we can conclude that the changes in objective workload do not cause associated changes in the individual EEG frequency bands based on the physiological data retrieved from the HUMAN Lab Surveillance and Tracking tasks.

Table 12. Average Canonical Correlation (Scenario 1 - 16) between each node in Alpha, Beta, Gamma (1-3, Delta, and Theta EEG frequency respectively and VACP workload channels in the Tracking Tasks.

Alpha Frequency	F7	Fz	F8	Pz	T7	T8	O2
Avg Auditory	0.02	-0.01	-0.02	0.01	-0.01	-0.04	0.08
Avg Cognitive	0.02	-0.02	-0.02	0.00	0.00	-0.04	0.09
Avg Fine Motor	-0.03	-0.01	0.13	0.07	-0.06	-0.09	-0.04
Avg Speech	0.06	0.00	-0.03	-0.02	0.00	-0.02	0.07
Avg Visual	-0.01	0.03	-0.03	-0.07	0.06	0.04	0.12
Avg Overall	0.02	-0.02	-0.02	0.00	0.00	-0.04	0.09

Beta Frequency	F7	Fz	F8	Pz	T7	T8	O2
Avg Auditory	0.03	0.01	0.03	-0.09	-0.05	0.02	0.10
Avg Cognitive	0.01	0.01	0.05	-0.07	-0.07	-0.01	0.08
Avg Fine Motor	-0.07	0.01	0.19	-0.17	-0.06	-0.19	-0.02
Avg Speech	0.03	0.02	0.03	-0.09	-0.03	0.04	0.08
Avg Visual	-0.02	0.03	-0.04	0.24	0.09	0.16	-0.07
Avg Overall	0.01	0.01	0.06	-0.08	-0.06	0.00	0.08

Gamma 1 Frequency	F7	Fz	F8	Pz	T7	T8	O2
Avg Auditory	0.00	-0.01	-0.02	-0.03	-0.02	-0.02	0.04
Avg Cognitive	0.00	-0.01	-0.01	-0.03	-0.02	-0.02	0.03
Avg Fine Motor	-0.04	0.01	0.07	0.00	0.01	-0.04	-0.02
Avg Speech	-0.01	0.02	-0.01	-0.02	-0.02	0.00	0.05
Avg Visual	-0.03	0.01	0.03	0.06	0.01	0.04	-0.03
Avg Overall	0.00	-0.01	-0.01	-0.03	-0.02	-0.02	0.03
Gamma 2 Frequency	**F7**	**Fz**	**F8**	**Pz**	**T7**	**T8**	**O2**
Avg Auditory	0.02	0.00	0.02	-0.07	-0.06	0.05	0.02
Avg Cognitive	0.01	-0.01	0.01	-0.06	-0.06	0.06	0.02
Avg Fine Motor	-0.05	-0.01	0.01	-0.09	0.04	-0.03	0.17
Avg Speech	0.01	0.02	0.02	-0.06	-0.08	0.05	0.02
Avg Visual	0.09	-0.02	-0.12	0.06	0.03	-0.03	-0.10
Avg Overall	0.01	-0.01	0.01	-0.06	-0.06	0.06	0.02
Gamma 3 Frequency	**F7**	**Fz**	**F8**	**Pz**	**T7**	**T8**	**O2**
Avg Auditory	-0.02	0.00	0.07	0.02	-0.03	0.04	0.00
Avg Cognitive	-0.02	0.00	0.07	0.03	-0.03	0.04	0.01
Avg Fine Motor	0.02	0.12	0.01	0.02	-0.26	0.03	-0.10
Avg Speech	0.01	-0.05	0.06	0.00	0.02	0.03	-0.01
Avg Visual	-0.03	-0.08	0.00	0.01	0.28	0.00	0.04
Avg Overall	-0.02	0.00	0.07	0.02	-0.01	0.04	0.01

Delta Frequency	F7	Fz	F8	Pz	T7	T8	O2
Avg Auditory	0.01	-0.01	-0.01	0.01	0.01	-0.02	0.02
Avg Cognitive	0.01	-0.01	-0.01	0.01	0.01	-0.02	0.02
Avg Fine Motor	0.04	0.04	-0.05	0.01	-0.05	-0.04	-0.09
Avg Speech	0.03	0.00	-0.02	0.02	0.00	-0.02	0.01
Avg Visual	-0.03	-0.09	0.02	0.07	-0.10	0.05	0.21
Avg Overall	0.00	0.00	0.00	0.02	-0.01	0.00	-0.01

Theta Frequency	F7	Fz	F8	Pz	T7	T8	O2
Avg Auditory	0.01	-0.01	-0.01	0.02	-0.03	0.00	0.03
Avg Cognitive	0.01	-0.01	-0.01	0.02	-0.03	0.00	0.03
Avg Fine Motor	-0.03	-0.05	0.07	0.12	0.01	-0.04	-0.10
Avg Speech	0.03	0.01	-0.03	0.02	-0.03	0.01	0.01
Avg Visual	0.00	0.09	-0.13	-0.03	0.00	-0.01	0.06
Avg Overall	0.01	-0.01	-0.01	0.03	-0.03	0.00	0.03

Table 13. Average Canonical Correlation (Scenario 1 - 16) between each node in Alpha, Beta, Gamma (1-3, Delta, and Theta EEG frequency respectively and VACP workload channels in the Surveillance Tasks.

Alpha Frequency	F7	Fz	F8	Pz	T7	T8	O2
Avg Auditory	0.09	0.09	0.05	-0.24	-0.02	-0.02	-0.06
Avg Cognitive	-0.13	-0.01	0.03	-0.02	-0.01	-0.02	0.13
Avg Fine Motor	-0.11	-0.16	0.13	0.12	-0.08	-0.01	-0.01
Avg Speech	0.11	0.14	0.01	-0.27	0.01	-0.02	-0.05
Avg Visual	0.06	0.05	-0.05	-0.12	0.00	0.05	0.09
Avg Overall	0.04	0.05	-0.03	-0.12	0.02	0.04	0.09

Beta Frequency	F7	Fz	F8	Pz	T7	T8	O2
Avg Auditory	0.02	0.10	-0.20	-0.24	0.21	0.01	-0.02
Avg Cognitive	-0.04	-0.05	0.01	0.11	0.06	0.00	0.05
Avg Fine Motor	-0.08	0.11	-0.08	-0.08	-0.21	-0.08	0.02
Avg Speech	0.05	0.06	-0.21	-0.21	0.32	0.04	0.00
Avg Visual	0.03	-0.11	0.07	0.00	0.14	0.13	0.03
Avg Overall	0.00	-0.04	0.00	0.04	0.19	0.17	-0.02

Gamma 1 Frequency	F7	Fz	F8	Pz	T7	T8	O2
Avg Auditory	0.04	-0.03	-0.03	0.05	-0.03	0.07	0.02
Avg Cognitive	-0.01	-0.09	0.05	0.03	-0.05	0.02	0.05
Avg Fine Motor	-0.05	0.05	-0.02	0.01	-0.07	0.05	0.01
Avg Speech	0.06	-0.09	0.00	0.06	-0.02	0.06	0.03
Avg Visual	0.02	0.02	0.02	0.03	-0.02	0.00	-0.02
Avg Overall	0.03	0.03	0.01	0.05	0.00	0.00	-0.04
Gamma 2 Frequency	**F7**	**Fz**	**F8**	**Pz**	**T7**	**T8**	**O2**
Avg Auditory	0.00	0.06	-0.04	-0.02	-0.02	0.04	0.02
Avg Cognitive	0.01	-0.15	0.02	0.02	-0.03	-0.03	0.01
Avg Fine Motor	-0.07	0.12	-0.02	0.00	0.08	-0.01	0.01
Avg Speech	0.03	-0.04	-0.05	-0.02	-0.06	0.04	0.02
Avg Visual	0.03	-0.08	0.08	0.02	0.00	-0.01	0.03
Avg Overall	0.03	-0.08	0.05	0.05	0.01	0.00	0.00
Gamma 3 Frequency	**F7**	**Fz**	**F8**	**Pz**	**T7**	**T8**	**O2**
Avg Auditory	-0.03	-0.04	0.12	-0.13	-0.08	-0.04	0.00
Avg Cognitive	0.03	0.18	-0.02	-0.02	0.20	-0.06	-0.10
Avg Fine Motor	0.05	-0.07	0.05	-0.05	-0.33	0.06	0.06
Avg Speech	-0.03	0.07	0.10	-0.13	0.12	-0.09	-0.06
Avg Visual	0.00	-0.05	-0.06	-0.02	0.25	-0.06	-0.04
Avg Overall	-0.03	-0.02	-0.04	-0.06	0.23	-0.06	-0.02

Delta Frequency	F7	Fz	F8	Pz	T7	T8	O2
Avg Auditory	-0.03	0.05	0.13	-0.06	0.01	-0.02	0.00
Avg Cognitive	-0.04	-0.02	0.11	0.07	-0.08	-0.02	0.09
Avg Fine Motor	-0.05	0.03	0.05	-0.08	-0.02	0.03	-0.13
Avg Speech	-0.01	0.03	0.11	-0.01	0.01	-0.04	0.05
Avg Visual	-0.01	0.02	0.05	0.02	-0.12	0.06	0.10
Avg Overall	-0.02	0.02	0.07	0.03	-0.12	0.08	0.10

Theta Frequency	F7	Fz	F8	Pz	T7	T8	O2
Avg Auditory	0.09	0.11	-0.09	-0.14	0.07	-0.11	0.05
Avg Cognitive	0.05	0.09	-0.16	-0.05	-0.04	-0.07	0.04
Avg Fine Motor	-0.09	-0.02	0.16	-0.05	-0.06	0.02	-0.10
Avg Speech	0.12	0.13	-0.15	-0.12	0.08	-0.11	0.07
Avg Visual	0.06	0.10	-0.18	0.00	0.02	-0.08	0.07
Avg Overall	0.06	0.11	-0.16	-0.01	0.02	-0.07	0.06

V. Conclusions and Recommendations

The results presented in this thesis show that there is promise in using EEG frequency data for performance classification. This thesis presented an in depth look into each EEG frequency band used in the 711th HPW/ RHCP HUMAN LAB experiment to give future researchers more insight towards what each EEG frequency is capable of with respect to classification of operator performance and operator workload prediction. There is still much work to be done before EEG Data can be relied on heavily as an indicator of performance or workload.

Performance Classification with individual EEG frequencies: Is it possible to classify performance using EEG data exclusively?

The results of performance classification show that High performers and Low performers can be detected using only EEG data and machine learning. From the results presented in this thesis, we can conclude that detecting these two different classes (High performer, Low performer) is possible using either the Gamma EEG data or the Beta EEG data. Similarly, these two classes can be detected using all EEG frequencies combined. Based on the results reported in this thesis, it may be possible to rely on only Gamma EEG data to predict task performance as opposed to all EEG frequencies combined. Reducing the number of features used when classifying with EEG data from 49 (All EEG frequencies) to 21 (Gamma 3x7) would improve the ANN used to classify performance and decrease computational load. This would make implementation in the field much easier and make classification online a more feasible effort. Instead of designing an algorithm to utilize some combination of EEG frequency bands, researchers could use just one. Task prediction using EEG frequency data from dual classified subjects indicate that there is a difference in EEG data when an individual struggles in a task as opposed to when that individual excels in another. The results show that the Gamma EEG frequency was the best and most consistent in its ability to predict task performance in the dual classified subjects.

Results from Task Prediction using novel EEG frequency scenario data show that the methods used in this thesis will not facilitate accurate prediction of High or Low Performers using raw EEG data the ANN has not been trained on. An ANOVA showed the classification accuracy of the ANN on novel scenario data was statistically similar to the classification accuracy of the Naïve classifier. The results from the classification analysis suggest some areas for future work to validate or improve the results.

Other Machine Learning Techniques Used in Combination

It would be beneficial to compare these results against the use of another Machine Learning technique like Self Organizing Maps (SOM) or Radial Basis Function Neural Network (RBFNN) to see if better results can be produced using EEG data to classify individuals. The scoring algorithm used in the HUMAN LAB study was not tested for accuracy before its inception. It is possible that the task performance labeling technique used was not the best way to identify the EEG data samples for classification, resulting in lower classification ratios for the Alpha, Delta, and Theta EEG frequencies. Patterns may exist in these frequency bands that cannot be determined by any individual. Instead, these patterns may be better identified using a SOM or RBFNN to identify and label the data initially, and then attempt classification using each individual EEG frequency band. Amarasinghe et al proposed a novel methodology to recognize thought patterns using Self Organizing Maps (SOM) for unsupervised clustering of raw EEG data and a feed forward ANN for classification [6]. This same method may be helpful in distinguishing different ways to label to the data to improve the low classification results of the Alpha, Delta and Theta EEG frequencies.

Feature Reduction of the EEG Frequencies

Experiments conducted by Yordavana et al and Struber et al were able to show that the spontaneous gamma activity was greater at the frontal node sites and decreased in the anterior and posterior regions [35, 36]. Specifically, Gamma EEG frequency power was significantly greater at the left than at the right frontal node sites. It may be beneficial to only include Frontal lobe node sights to truly test their responsiveness to sensorimotor information processing and their ability to predict task performance. A classification study could be done similar to the one presented in this thesis, but using EEG frequency data with only Frontal lobe features. Once the features from the Frontal lobes have been isolated, noisy data should then be removed using Independent Component Analysis (ICA) similar to Belyalvin et al [5]. This process would decrease the amount of remaining muscle and eye movement noise from the data that hinder the ANN’s ability to predict task performance. If the classification results beat that of the classification results presented in this thesis, it would further benefit algorithms designed to trigger augmentation based on physiological features.

Workload Prediction using EEG frequency data: Is it possible to predict workload using EEG frequency data exclusively?

The results presented in this thesis during the workload prediction analysis suggest there is still much research to be done in this area. Currently there is very little research that has been done to explore the abilities of each individual EEG frequency band to predict objective operator workload values. The work presented in this thesis can act as a starting point for future research in this area. The prediction accuracy of the ANN was recorded using Root Mean Squared Error and compared to a Naïve Predictor. After analysis on 192 combinations of scenarios, and VACP channels, there was no evidence that workload can be accurately predicted using raw EEG data with the techniques presented in the Methodology of this thesis. Specifically, predicting Overall VACP workload based on EEG frequency data proved difficult for the ANN. In most scenarios, the ANN was extremely close to the high error results of the Naïve Predictor. In 5 scenarios (Scenarios 1-16, Table 11) the Naïve predictor actually beat the predictive accuracy of the ANN. Also, there was no correlation with the EEG frequency data and the VACP workload values. Using all of the EEG frequency data combined to predict the VACP workload data actually produced more error than individual EEG frequency bands. Does this mean that there is such a thing as too many features in the input data with regards to predicting VACP workload values with only EEG data? The high error and poor correlation results are clear indicators that there is much work to be done before prediction with EEG frequency data is used in the field. It may be beneficial to conduct tests in the future with minor changes to improve workload prediction

Feature Reduction

In the future, it may be beneficial to do some feature reduction on the EEG frequency data before attempting to use it as input to the ANN for workload prediction. Reducing the features used to represent the EEG frequency bands to 2 or 3 may improve the workload prediction RMSE results. Employing further filtering techniques on the EEG data to reduce the feature size of each EEG frequency data may also be beneficial to the ANN to increase prediction accuracy. It was clear that when features and granularity were increased to predict workload, the ANN performed worse with higher RMSE.

Development of an EEG Baseline

To get a more precise idea of how well EEG data can predict or classify, the experiment itself has to be set up to do so from onset. An “EEG baseline” must be established so that changes in the EEG data due to increases workload are more distinguishable when using machine learning. As stated earlier in this conclusion, there is noise in the EEG frequency data generated from muscle movements, eye blinks, and other functions of the body. An EEG baseline could be defined as a period of time before the task where the individual closes their eyes, sits motionless, and is given noise muffling headphones to reduce recorded EEG noise from these bodily functions [8]. This baseline would make it easier for the ANN to distinguish between periods with no workload and where workload was induced. This increased ability to distinguish between changes in workload may reduce prediction error. During the task it would be beneficial to remove all persons from the room and turn off all lights in the room to reduce distraction from the task. The participant should only be able to see the apparatus being used to conduct the experiment. Developing a baseline where this noise has less of an impact on the noise captured by the EEG nodes would be beneficial to a study that looked to deeply analyze EEG frequency data and its' ability to predict workload.

Summary

There is great promise in researching the classification and predictive abilities of EEG frequency data. EEG data is a fairly untapped resource in the Machine Learning community, but with further research, EEG frequency data could become a strong physiological feature used in a system designed to augment human performance using physiological data. Further investigation of the EEG frequencies is needed before this step can be taken. Evidence presented in this thesis suggest the Gamma EEG frequency is the best EEG frequency to use to classify individuals as High or Low performers in tasks requiring alertness such as Surveillance and Tracking.

Further research is needed when it comes to predicting workload based on EEG data exclusively. It would be extremely helpful to explore feature reduction techniques to reduce the amount of data the ANN used as an input to predict workload values. The workload prediction results show that too much granularity in the EEG frequency data is disadvantageous to the ANN and hinders its ability to predict. They also show the importance of properly setting up an experiment to analyze desired features. In the future, it would be beneficial to set up a baseline for any physiological feature to be analyzed post experiment. This would make changes in the physiological data more apparent, specifically the EEG frequency data.

Appendix A. Root Mean Squared Error per EEG frequency using the ANN to predict VACP Workload Channel and Scenario (1-16) in the Surveillance and Tracking Tasks Respectively

	Alpha Surveillance					
	Auditory	Cognitive	Fine Motor	Overall	Speech	Visual
1	1.79	4.14	0.46	4.89	0.41	1.05
2	1.80	2.67	0.50	7.30	0.39	0.91
3	1.91	3.92	0.48	5.11	0.42	0.83
4	1.91	4.57	0.45	7.00	0.41	1.15
5	1.98	1.55	0.48	6.59	0.43	0.84
6	2.00	4.14	0.47	5.55	0.39	0.78
7	1.66	3.42	0.46	6.55	0.45	0.92
8	2.19	3.24	0.49	6.29	0.42	1.63
9	1.88	2.68	0.45	5.43	0.43	1.00
10	2.09	5.84	0.48	9.47	0.43	1.84
11	1.99	4.45	0.45	8.35	0.44	0.98
12	1.74	3.52	0.46	5.23	0.42	0.94
13	1.74	4.16	0.48	6.83	0.44	1.00
14	2.10	1.98	0.46	5.83	0.43	0.91
15	2.00	3.20	0.46	3.03	0.43	0.87
16	2.05	3.23	0.49	3.40	0.43	0.83

	Beta Surveillance					
	Auditory	Cognitive	Fine Motor	Overall	Speech	Visual
1	1.99	4.42	0.57	8.64	0.45	1.27
2	2.13	3.74	0.54	6.36	0.44	1.10
3	2.04	3.42	0.49	5.51	0.44	0.82
4	2.18	3.53	0.47	9.80	0.40	0.74
5	2.04	2.77	0.47	5.44	0.43	0.96
6	2.23	4.35	0.48	7.84	0.45	0.91
7	2.01	5.15	0.67	10.34	0.45	1.12
8	2.09	3.10	0.51	10.50	0.46	1.22
9	2.10	7.00	0.49	5.70	0.48	0.91
10	1.52	3.51	0.48	5.77	0.52	0.86
11	2.11	7.46	0.42	6.69	0.44	2.52
12	2.04	3.70	0.48	9.61	0.40	1.09
13	2.04	4.73	0.53	10.10	0.46	1.11
14	2.15	4.43	0.52	5.17	0.43	0.98
15	1.93	3.41	0.54	3.37	0.45	0.82
16	1.94	2.13	0.51	7.31	0.44	0.85

	Gamma Surveillance					
	Auditory	Cognitive	Fine Motor	Overall	Speech	Visual
1	2.18	5.20	0.84	10.00	0.62	0.95
2	1.93	3.73	0.57	5.02	0.43	1.56
3	2.09	4.57	0.57	7.44	0.47	0.90
4	1.37	6.01	0.63	9.74	0.51	1.13
5	2.11	4.42	0.51	7.12	0.45	1.27
6	2.00	3.99	0.44	8.47	0.41	1.08
7	2.21	4.92	0.67	5.91	0.67	1.39
8	2.06	3.83	0.74	11.18	0.45	2.00
9	2.10	3.79	0.43	5.20	0.42	1.35
10	1.85	5.86	0.47	7.39	0.54	1.02
11	2.47	6.79	0.64	8.04	0.45	1.31
12	2.13	5.03	0.51	7.18	0.42	1.16
13	2.13	3.87	0.54	6.45	0.46	1.25
14	2.42	2.88	0.53	5.08	0.46	1.35
15	1.90	3.24	0.57	3.86	0.49	0.86
16	2.11	3.60	0.58	6.34	0.47	0.87

	Delta Surveillance					
	Auditory	Cognitive	Fine Motor	Overall	Speech	Visual
1	1.91	2.97	0.43	3.30	0.41	0.74
2	2.05	3.43	0.48	3.99	0.43	0.83
3	1.98	3.05	0.45	4.23	0.41	0.73
4	1.94	2.81	0.42	5.40	0.40	0.82
5	1.99	2.69	0.46	3.61	0.43	0.88
6	1.83	2.84	0.44	2.87	0.41	0.79
7	1.87	3.27	0.46	3.97	0.45	0.93
8	2.03	3.16	0.46	5.37	0.45	0.85
9	1.88	2.86	0.45	4.16	0.43	0.89
10	2.01	3.21	0.44	3.86	0.43	0.91
11	1.87	3.05	0.44	4.24	0.43	0.91
12	1.97	2.71	0.46	3.84	0.42	0.81
13	1.97	3.04	0.45	4.00	0.43	0.89
14	1.89	3.22	0.48	4.25	0.44	0.76
15	1.95	2.50	0.46	3.38	0.42	0.80
16	2.26	3.19	0.47	3.91	0.43	0.78

	Theta Surveillance					
	Auditory	Cognitive	Fine Motor	Overall	Speech	Visual
1	1.97	3.64	0.52	2.98	0.42	0.82
2	2.05	3.41	0.52	3.75	0.42	0.89
3	2.04	4.04	0.46	4.16	0.42	0.81
4	1.92	3.48	0.44	5.42	0.41	0.85
5	1.90	4.14	0.45	5.19	0.43	0.94
6	1.95	4.08	0.46	3.42	0.41	0.82
7	1.93	4.08	0.47	3.88	0.44	0.91
8	1.93	3.22	0.49	5.10	0.43	0.89
9	2.21	3.04	0.47	6.55	0.43	0.98
10	2.03	3.30	0.46	4.70	0.42	1.01
11	1.91	4.48	0.47	5.62	0.46	1.10
12	1.89	3.26	0.46	4.71	0.41	0.79
13	1.89	3.18	0.48	5.26	0.43	0.98
14	2.02	3.74	0.45	5.50	0.40	0.91
15	1.96	3.19	0.47	3.33	0.42	0.82
16	1.93	4.77	0.50	3.75	0.43	0.92

	Alpha Tracking					
	Auditory	Cognitive	Fine Motor	Overall	Speech	Visual
1	1.67	8.65	2.10	15.38	0.43	4.68
2	1.63	5.48	1.16	15.81	0.39	2.44
3	2.03	3.71	1.48	13.73	0.42	3.27
4	1.90	4.62	1.55	13.29	0.40	2.67
5	1.51	5.25	1.16	14.60	0.38	2.02
6	1.91	6.23	2.14	14.90	0.34	4.49
7	1.95	4.64	1.74	10.68	0.40	2.81
8	1.77	4.76	1.47	9.98	0.40	2.58
9	2.13	6.23	1.35	19.49	0.40	2.49
10	1.72	5.07	1.28	11.19	0.41	2.76
11	1.73	3.36	0.95	17.48	0.37	2.04
12	1.33	5.60	1.98	15.61	0.42	4.92
13	1.76	4.54	1.26	8.54	0.38	2.63
14	1.35	4.96	1.50	13.69	0.39	2.85
15	1.73	5.36	1.74	10.57	0.41	3.86
16	1.44	5.47	1.92	12.00	0.41	2.78

	Beta Tracking					
	Auditory	Cognitive	Fine Motor	Overall	Speech	Visual
1	1.78	6.31	1.43	14.39	0.42	4.23
2	1.97	8.24	1.17	14.79	0.47	3.70
3	2.27	5.48	1.93	12.87	1.03	4.44
4	1.96	6.02	3.62	13.09	0.56	4.15
5	2.09	6.99	1.67	14.76	0.34	3.64
6	1.50	5.72	2.02	14.30	0.36	3.60
7	1.94	3.99	1.87	15.55	0.42	3.13
8	1.95	8.21	1.73	26.73	0.41	4.16
9	1.56	7.31	2.08	10.57	0.46	2.82
10	2.02	7.55	1.86	17.88	0.42	3.12
11	1.85	5.91	1.95	22.34	0.36	3.28
12	1.95	8.49	1.41	24.09	0.42	3.05
13	1.80	5.50	1.64	13.52	0.38	2.61
14	2.11	6.58	2.16	21.57	0.41	3.44
15	1.60	5.37	2.10	9.88	0.43	3.98
16	1.76	6.97	2.14	18.47	0.40	3.78

	Gamma Tracking					
	Auditory	Cognitive	Fine Motor	Overall	Speech	Visual
1	1.93	4.19	2.55	10.03	0.45	3.70
2	1.70	7.33	1.86	13.38	0.41	4.73
3	2.14	6.72	1.45	11.60	0.56	3.12
4	1.79	9.32	2.84	16.48	1.00	3.81
5	2.22	7.22	1.52	26.54	0.50	3.24
6	2.05	6.26	1.99	11.94	0.43	3.36
7	1.73	8.02	1.38	13.45	0.49	2.34
8	1.91	5.20	1.53	11.04	0.49	3.56
9	2.23	4.96	1.71	8.55	0.53	3.02
10	1.58	6.40	1.61	7.52	0.50	2.36
11	2.00	6.06	1.93	11.26	0.39	3.36
12	1.96	6.52	3.27	20.29	0.48	6.54
13	1.76	6.24	1.54	16.02	0.45	2.91
14	2.24	6.32	1.61	10.91	0.51	2.10
15	1.73	5.31	1.93	9.02	0.42	3.41
16	1.11	6.21	1.79	13.44	0.42	3.55

Delta Tracking						
	Auditory	Cognitive	Fine Motor	Overall	Speech	Visual
1	1.78	2.08	1.57	7.61	0.40	2.51
2	1.88	3.06	1.79	8.31	0.39	2.20
3	1.72	3.09	1.44	6.44	0.39	2.42
4	1.65	2.51	1.61	7.73	0.40	2.68
5	2.07	2.34	1.66	8.11	0.38	2.64
6	1.81	3.19	1.59	8.98	0.36	2.38
7	2.09	2.59	1.52	7.73	0.40	2.62
8	1.77	3.76	1.80	6.75	0.39	2.88
9	1.93	3.33	1.71	8.47	0.40	2.75
10	1.86	3.57	1.57	5.15	0.38	2.53
11	2.09	3.78	1.93	11.44	0.39	3.10
12	1.61	3.17	1.46	11.35	0.44	2.21
13	1.86	2.45	1.73	9.38	0.40	2.39
14	1.76	4.06	1.69	9.56	0.41	2.84
15	1.71	3.38	1.74	10.71	0.41	2.77
16	1.56	2.82	1.42	8.88	0.39	1.67

Theta Tracking						
	Auditory	Cognitive	Fine Motor	Overall	Speech	Visual
1	1.68	3.99	1.35	11.75	0.42	1.81
2	1.71	6.48	0.90	14.55	0.40	2.79
3	1.88	6.07	1.44	13.79	0.39	1.52
4	1.78	5.24	1.76	13.15	0.41	3.63
5	1.85	5.03	1.21	18.74	0.37	2.82
6	1.37	4.82	1.51	8.48	0.36	3.22
7	1.42	4.49	1.79	10.26	0.41	2.82
8	1.74	2.96	1.71	14.93	0.39	2.88
9	1.57	5.12	1.26	11.28	0.42	2.30
10	2.03	5.38	1.66	10.91	0.38	2.78
11	1.89	3.88	1.49	11.19	0.40	3.09
12	1.81	6.52	1.34	17.76	0.47	2.83
13	1.69	4.69	1.50	9.46	0.39	1.60
14	1.66	5.91	1.36	9.62	0.43	3.53
15	1.69	5.06	1.31	9.99	0.41	3.22
16	1.73	3.98	1.71	10.85	0.40	2.92

Appendix B. One Way ANOVA *p* value results between Workload Prediction RMSE using ALL EEG Data Combined and Individual EEG Frequencies as Input to the ANN and the Uniform Naïve Predictor. One-Way ANOVA used the lowest RMSE from analysis using each EEG frequency to predict workload

One-Way ANOVA between All EEG Combined and Naïve Predictor		
	Surveillance	**Tracking**
Auditory	4.037e-12	1.708e-11
Cognitive	0.003	1.708e-11
Fine Motor	5.079e-10	0.327
Overall	8.375e-04	6.437e-06
Speech	1.604e-20	4.357e-16
Visual	0.011	0.095

One-Way ANOVA between Individual EEG Frequencies and with Naïve Predictor		
	Surveillance	**Tracking**
Auditory	1.446e-31	1.446e-31
Cognitive	1.317e-12	1.317e-12
Fine Motor	1.006e-27	1.006e-27
Overall	0.001	0.001
Speech	1.071-45	1.071e-45
Visual	4.6211e-05	4.621e-05

Bibliography

[1] Merriam-Webster, The Merriam-Webster Dictionary, Massachusetts: Merriam-Webster Mass Market, 2010.

[2] W. S. Helton, D. Hollander, W. J, D. Tripp, K. Parsons, G. Matthews, D. W, P. R. and H. P., "The Abbreviated Vigilance and Cerebral Hemodynamics," *Journal of Clinical and Experimental Neuropsycology*, vol. 29, no. 5, pp. 545-552, 2007.

[3] G. Borghini, L. Astolfi, F. Toppi, D. Cincotti, P. Mattia, G. Cherubino, A. Vecchiato, I. Maglione, A. Babiloni, I. Graziana and P. Arico, "Frontal EEG Theta Changes Asses the Training Improvements of Novices in Flight Simulation Tasks.," in *35th Annual International Conference of the IEEE EMBS*, Osaka, 2013.

[4] A. Fong, C. Sibley, A. Cole, C. Baldwin and J. Coyne, "A Comparisoon of Artificial Neural Networks, Logistic Regressions, and Classification Trees for Modeling Mental Workload in Real-Time," in *54th Human Factors and Ergonomics Society Annual Meeting 2010*, San Francisco, 2010.

[5] A. Belyavin, C. Ryder and B. Dickson, "A statistics based approach to the use of EEG signals to measure workload," QinetiQ Ltd, Farmborough, UK, 2002.

[6] K. Amarasinghe, D. Wijayasekara and M. Manic, "EEG Based Brain Activity Monitoring using Artifial Neural Networks," University of Idaho, Idaho Falls, 2014.

[7] US Army Research Lab, "Improved Performance Research Integration Tool," [Online]. Available: http://www.arl.army.mil/www/default.cfm?page=445. [Accessed 2010].

[8] S. Fishel, E. Muth and A. Hoover, "Establishing approriate physiological baseline procedures for real-time physiological measurement," *Journal of Cognitive Engineering and Decision Making*, vol. 1.3, pp. 286-308, 2007.

[9] K. Durkee, A. Geyer, S. Pappada, A. Ortiz and S. Galster, "Real-Time Workload Assessment as a Foundation for Human Performance Augmentation," in *15th International Conference on Human-Computer Interaction, HCI International*, 2013.

[10] S. M. Galster and E. Johnson, "Sense-assess-augment: A taxonomy for human effectiveness," Air Force Research Laboratory, Wright Patterson Air Force Base, OH 45433, 2013.

[11] S. G. Hart, "NASA-Task Load Index (NASA-TLX); 20 Years Later," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2006.

[12] S. Hart, "Development of Nasa-TLX: Results of Empirical and Theoretical research," NASA Ames Research Center, Moffet Field California, 1988.

[13] J. Christensen and J. R. Estepp, "Coadaptive Aiding and Automation Enhance Operator Performance," *Human Factors,* pp. 965-975, 2013.

[14] J. Song, Z. Damin, S. Guoling and W. Xiaoru, "Pilot Mental Workload Measurement and Evaluation under Dual Task," in *4th International Conference on Biomedical Engineering and Informatics,* 2011.

[15] L. Prinzell III, F. Freeman, M. Scerbo, P. Mikulka and A. Pope, "Effects of a Physiological System for Adaptive Automation on Performance, Workload, and the Event-Related Potential P300 Component," *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 2003.

[16] T. H. Shaw, K. Satterfield, R. Ramirez and V. Finomore, "A Comparison of Subjective and Physiological Workload assessment techniques during a 3-dimensional audio vigilance task," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2012.

[17] T. Tiwari, A. Singh and I. Singh, "Task Demand and Workload: Effects on Vigilance Performance and Stress," *Journal of the Indian Academy of Applied Physcology,* vol. 35, no. 2, pp. 265-275, 2009.

[18] D. J. Saxby, G. Matthews, E. Hitchcock and J. Warm, "Development of Active and Passive Fatigue Manipulations using a Driving Simulatior," in *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting*, 2007.

[19] J. Jeroski, "Physiological Investigation of Localized Temperature Effects on Vigilance Performance," Air Force Institute of Technology, WPAFB, 2014.

[20] A. Azadeh, M. Rouzbahman, M. Saberi, F. Valianpour and A. Keramati, "Improved prediction of mental workload versus HSE and ergonomoics factors by an adaptive intelligent algorithm," *Safety Science,* vol. 58, pp. 59-75, 2013.

[21] G. Borghini, L. Astolfia, G. Vecchiatoa, D. Mattiaa and F. Babilonia, "Measuring Neurophysiological Signals in Aircraft Pilots and Car Drivers for the Assessment of Mental Workload," *Neuroscience and Biobehaviroal Reviews,* vol. 44, pp. 58-75, 2012.

[22] G. Borghini, G. Vecchiato, A. Colosimo, D. Wei, A. G. Maglione, W. Kong, F. Bablioni and L. Astolfi, "Assessment of Mental Fatigue During Car Driving by Using High Resolution EEG Activity and Neurophysiologic Indices," in *Concerence of the IEEE EMBS*, San Diego, 2012.

[23] C.-T. Lin, R. P. Nikhil, C.-Y. Chuang, J. Tzyy-Ping, K. Li-Wei and L. Sheng-Fu, "An EEG-based Subject and Session-independent Drowsiness Detection," in *2008 International Joint Conference on Neural Networks*, Hong Kong, 2008.

[24] F. Ebrahimi, M. Mikaeili, E. Estrada and H. Nazeran, "Automatic Sleep Stage Classification Based on EEG Signals by Using Neural Networks and Wavelet Packet Coefficients," in *30th Annual International IEEE EMBS Conference Vancouver*, Vancouver, 2008.

[25] P. Nichols, W. Farmer, I. Peachey and J. Belyavin, "Dual task interference: Using Hierarchical Cluster Analysis to identify underlying cognitive mechanisms," in *Proceedings of the 47th Human Factors and Ergonomics Meeting*, Denver, Colorado, 2003.

[26] S. Haykin, Neural Networks and Learning Machines, McMaster University: Prentice Hall, 2008.

[27] A. G. Correa and E. L. Leber, "An Automatic Detector of Drowsiness Based on Spectral Analysis and Wavelet Decopomposition of EEG Records," in *Conference of the IEEE EMBS Buenos Aires*, Buenos Aires, 2010.

[28] A. Rechtschaffen and A. Kales, A Manual of Standardized Terminology, Techniques and scoring system for Sleep Stages of Human Subjects, Los Angeles: Brain Information service/Brain Research Institute, 1968.

[29] G. F. Wilson and C. A. Russell, "Performance Enhancement in an Uninhabited Air Vehicle Task Using Psychophysiological Determined Adaptive Aiding," in *Human Factors: The Journal of the Human Factor and Ergonomics Society,* 2008.

[30] H. S. Almahasneh, N. Kamel, A. Malik, N. Wlater and W. Chooi, "EEG Based Driver Cognitive Distraction Assessment," Universiti Teknologi, Petrak, Malaysia, 2014.

[31] T. Sejnowski and P. S. Churchland, The Computational Brain, Cambridge: MIT Press.

[32] M. H. Beale, M. T. Hagan and H. B. Demuth, Neural Network Toolbox User's Guide 2014b, Natick: The MathWorks Inc, 2014.

[33] A. Lund PhD and L. Mark, "Two-Way ANOVA in SPSS Statistics," 1 January 2013. [Online]. Available: https://statistics.laerd.com/spss-tutorials/two-way-anova-using-spss-statistics.php. [Accessed 22 January 2015].

[34] Thomas, F. Collura Ph.D, "The Measurement Interpretation and Use of EEG Frequency Bands," 2005. [Online]. Available: http://www.brainmaster.com/generalinfo/eegbands/eegbands.html. [Accessed 22 January 2013].

[35] J. Yordanova, V. Kolev, H. Heinrich, W. Woerner, T. Banaschewski and A. Rothenberger, "Developmental event-related Gamma Oscillations: effects of auditory attention," *European Journal of Neuroscience,* vol. 16, pp. 2214-2224, 2002.

[36] D. Struber and C. Basar-Eroglu, "EEG Gamma-band response during the perception of Necker cube reversals," *Vision Cognition,* vol. 8, pp. 609-621, 2001.

[37] F. Aokia, E. Fetza, L. Shupea, E. Lettich and G. A. Ojemannc, "Increased Gamma-range activity in human sensorimotor cortex during performance of visuomotor tasks," *Clinical Neurophysiology,* vol. 110, pp. 524-537, 1999.

[38] J. O. Willoughby, S. P. Fitzgibbon, K. J. Pope, L. Mackenzie, M. Davey, R. A. Wilcox and C. R. Clark, "Mental Tasks Induce Gamma EEG with Reduced Responsiveness in Primary Generalized Epilepsies," *Epilepsia,* vol. 44, pp. 1406-1412, 2003.

[39] P. May, H. Tiitinen, J. Sinkonen and R. Naatanen, "Long-term stimulation attenuates the transient 40 Hz response," *Neuroreport*, vol. 5, pp. 1918-1920, 1994.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 074-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
26-03-2015	Master's Thesis	October 2013 – March 2015

4. TITLE AND SUBTITLE
A Novel Analysis of Performance Classification and Workload Prediction Using Electroencephalography (EEG) Frequency Data

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Ricks, Donovan L., 1Lt, USAF

5d. PROJECT NUMBER
15G129

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
Wright-Patterson AFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENG-MS-15-M-012

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
711th Human Performance Wing
Attn: Scott Galster
2510 Fifth Street, Bldg. 840
Wright-Patterson Air Force Base, Ohio 45433
(937)-798-3632 scott.galster@us.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)
711th HPW/AFRL/RHCPA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Statement A. Approved for public release; Distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Across the DOD each task an operator is presented with has some level of difficulty associated with it. This level of difficulty over the course of the task is also known as workload, where the operator is faced with varying levels of workload as he or she attempts to complete the task. The focus of the research presented in this thesis is to determine if those changes in workload can be predicted and to determine if individuals can be classified based on performance in order to prevent an increase in workload that would cause a decline in performance in a given task. Despite many efforts to predict workload and classify individuals with machine learning, the classification and predictive ability of Electroencephalography (EEG) frequency data has not been explored at the individual EEG Frequency band level. In a 711th HPW/RCHP Human Universal Measurement and Assessment Network (HUMAN) Lab study, 14 Subjects were asked to complete two tasks over 16 scenarios, while their physiological data, including EEG frequency data, was recorded to capture the physiological changes their body went through over the course of the experiment. The research presented in this thesis focuses on EEG frequency data, and its' ability to predict task performance and changes in workload. Several machine learning techniques are explored in this thesis before a final technique was chosen. This thesis contributes research to the medical and machine learning fields regarding the classification and workload prediction efficacy of EEG frequency data. Specifically, it presents a novel investigation of five EEG frequencies and their individual abilities to predict task performance and workload. It was discovered that using the Gamma EEG frequency and all EEG frequencies combined to predict task performance resulted in average classification accuracies of greater than 90%.

15. SUBJECT TERMS
Human Performance, Machine Learning, Artificial Neural Network, Workload

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Borghetti, Brett, Ph. D., Research Advisor
					19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	97	(937) 255-3636 x 4612

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18