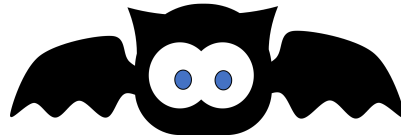# BATS: Development of a Biosignal Analysis Toolkit and Pipeline for Polytrauma Research

A Major Qualifying Project submitted to the faculty of

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the degree of Bachelor of Science

Submitted By:

Nathan Draudt

Claire Nicolas

Carlos A. Velasquez

Lauren Wasserman

Submitted to:

Professor Mark Claypool

Worcester Polytechnic Institute

Professor Adam Lammert

Worcester Polytechnic Institute

Submitted on: 04/26/2022

Abstract

Polytrauma is a serious health condition that arises following traumatic events. Research into new technologies for diagnosing polytrauma is essential to improving health outcomes for patients. Technology development in this domain would be aided by better research tools that allow for simultaneous analysis of traditional evaluations and novel biosignals. There are currently no analysis pipelines designed for polytrauma-specific research. Based on previous research and models, we developed a web application designed specifically for speech and neural data analysis relating to polytrauma research. It allows users to make a secure account, store and analyze study data and results. The user study showed a favorable response to our system and inspired ideas for future improvement to support various users.

Table of Contents

List of Figures

List of Tables

1: Introduction

Five percent of adults worldwide suffer from depression according to the World Health Organization (*Depression*, 2021). Treatment studies have shown that 20% of people with major depressive disorder (MDD) are retrospectively diagnosed with generalized anxiety disorder (GAD) (Zbozinek et al., 2012). MDD and GAD are comorbid, meaning they will commonly be diagnosed together. Although comorbid disorders are common, past experiences can contribute to the likelihood of people having various comorbidities. When comorbidities appear after traumatic events, it is called polytrauma. In this report, the polytrauma referenced will consist of post-traumatic stress disorder (PTSD), chronic pain, major depressive disorder (MDD), post-concussive syndrome (PCS), and general anxiety disorder (GAD) as comorbidities.

Research on polytrauma is relatively new and involves the analysis of biosignal data. Researchers are looking for new ways to diagnose polytraumas through speech and neural data alongside common diagnostic tools such as psychiatric evaluations. Common research in polytrauma uses psychiatric diagnostic tools such as the Diagnostic and Statistical Manual of Mental Disorders (DSM) to better understand what symptoms overlap for which comorbidities. New research, however, focuses on analyzing a person's biosignals such as neural or speech to model the presence of various comorbidities (Marmar et al., 2019; Sheerin et al., 2018). Currently, some software, toolboxes, and pipelines allow for the analysis of biosignals for all signal research. However, these are not specific to polytrauma research, do not allow for synchronous analysis of biodata, or do not have analysis tools specific to both neural and speech data.

There are no biosignal pipelines that focus on polytrauma related data such as non-invasive neural data, speech data, and psychiatric evaluations. There are analysis tools that handle either neural, speech, or log data, but these tools only allow users to analyze a single type of biosignal. There is currently no easy way to monitor polytrauma symptoms. Currently available systems, toolkits, and pipelines include Cerebral Blood Flow Biomedical Informatics Research Network (CBFBNIR). CBFBIRN is a highly specific and technical pipeline that provides data visualization and database tools. However, the system does not allow for the categorization and filtering of various data types at the same time nor is it web-based. MNE-NIRS has excellent functional near-infrared spectroscopy (fNIRS) preprocessing and analysis capabilities, but it cannot save data in the same location it performs the analysis. Finally, GILBERT, created by a previous MQP team, can save, categorize, and visualize analyzed various biodata. However, GILBERT is not web-based and does not have the proper analysis methods needed to complete proper polytrauma research.

The technologies mentioned above do not focus on learning more about polytrauma and its comorbidities but rather on general analysis. Researchers and healthcare diagnosticians in the polytrauma research domain would greatly benefit from a centralized pipeline that provides improved diagnostics, symptom monitoring, and treatment suggestions based on polytrauma-

specific research. Current solutions fulfill their goals; however, polytrauma research needs one system to complete all of the above.

To address the lack of tools in the polytrauma research domain, we created a data management and analysis pipeline named Biosignal Analysis Toolkit System (BATS). Our pipeline allows users to store, analyze, visualize, and retrieve data in a centralized web application. We tested our system with a user study. Seventeen users completed tasks related to accounting creation, data uploading, data analysis, analysis viewing, and data downloading. These user tests provided insightful knowledge that we used to define different metrics to evaluate BATS' effectiveness and user-friendliness as a web-based pipeline. The final BATS system used previous user test results and feedback to improve usability, showing statistical improvements based on our metrics. Our system is a strong foundation for a data management and analysis pipeline for biosignal-based polytrauma research.

Section 2 discusses relevant background information regarding identifying polytrauma, its comorbidities, how it affects various populations, the current research surrounding polytrauma, and currently available algorithms, toolboxes, and pipelines for biodata analysis. Section 3 discusses our design process and implementation, including feature selection, algorithm development, and system development. Section 4 describes BATS, its internal design, functions, and technologies. Section 5 outlines our methodology, including our needs definition, strategic development, system design, and user testing methods. Section 6 details the results of our user study. Section 7 reports the discussion of our user study results. Section 8 details future changes that can be made to improve BATS. Section 9 concludes our examination of BATS and its development.

2: Background

2.1: Polytrauma

   The development of polytrauma as a reaction to a traumatic event is not uncommon. It is essentially a name for presenting multiple conditions or comorbidities simultaneously. Historically, polytrauma has been used to refer to many different conditions, be it specific combinations of physical injuries to post-traumatic stress disorder (PTSD) itself (Keel & Trentz, 2005). In this paper, polytrauma refers to the specific comorbidities of PTSD, chronic pain, major depressive disorder (MDD), generalized anxiety disorder (GAD), and post-concussive syndrome (PCS). However, polytrauma is increasingly being considered a condition in itself and is being used to describe a group of people with multiple conditions.

   Comorbidity is when different conditions exist together in one person (Soo Hyun Rhee et al., 2004). As discussed in First 2005, an understanding of how often certain conditions are comorbid, the order in which these conditions appear, and the causation between them have been a subject of organized study. One of these areas of study is categorization, the effort to give labels to common condition clusters, such as polytrauma. This is opposed to the previous medical standard of evaluating every case individually (First, 2005). Developing a greater understanding of common comorbidity interactions could dramatically affect how the medical community understands and treats polytrauma and similar conditions as a whole and for each patient. In particular, it may be possible to determine and treat the source(s) of the issue to help patients with more of their symptoms.

2.1.1: Polytrauma's comorbidities

   Often the combination of symptoms can cause unique experiences and secondary symptoms. For example, a person with both anxiety and depression will make different decisions about being in public than those with only one of these conditions. These new behaviors are often not just an average of their conditions, but new behaviors caused by the symptoms playing off each other and restricting options they might otherwise have had to cope with their condition (Galyamina et al., 2017).

2.1.1.1: Introduction to PTSD

   PTSD is a psychiatric disorder that occurs in patients after witnessing or experiencing a traumatic event (Torres, n.d.). These events include active combat, sexual assault, natural disasters, and severe injuries. PTSD presents differently in individual patients but has common symptoms such as: "intrusion symptoms[,] avoidance of stimuli associated with the trauma[,] negative alterations of cognition and mood associated with the trauma[, and] alterations in arousal and reactivity associated with the trauma" (American Psychiatric Association, 2013). PTSD is one of the most prevalent diagnoses in researched polytraumas. On its own, 9.8%-23% of people exposed to trauma will meet the DSM-IV criteria for PTSD within a year, and 91% of those with PTSD develop comorbid conditions (Price & van Stolk-Cooke, 2015). In polytrauma research, the presence of PTSD as a diagnosis with PCT and its correlation with MDD and GAD

have been intensively researched (Finley et al., 2015). PTSD is an important diagnosis to consider when creating a polytrauma research tool due to its large presence in the field.

2.1.1.2: Introduction to Chronic Pain

Chronic pain is a persistent experience of pain that occurs with chronic illness or past the usual recovery period (*Chronic Pain*, n.d.). Chronic pain can be due to other medical diagnoses such as cancer, nerve pain, fibromyalgia, back pain, headaches, and arthritis (*Acute Pain vs. Chronic Pain*, n.d.). In some cases, the pain can become unbearable to the point where the patient experiences changes in personality and a decrease in their quality of life (*Chronic Pain*, n.d.; Naylor et al., 2019). Changes in personality due to chronic pain can often lead to other comorbidities. Of patients dealing with chronic pain three months after a motor vehicle collision, 33.8% had PTSD, and 52.1% had PTSD and PCS (Peixoto et al., 2018). Similarly, 30%-54% of clinic-based patients suffering from chronic pain also suffered from MDD as comorbidity (Banks & Kerns, 1996). In patients, the important relationship between the presence of chronic pain and other comorbidities makes this condition an important aspect of polytrauma.

2.1.1.3: Introduction to MDD

Major depressive disorder (MDD), also known as clinical depression, is a mood disorder that causes sadness, loss of interest, impaired cognition, and vegetative symptoms (Otte et al., 2016). It can affect feelings and thought processes and lead to changes in behavior. Symptoms include extended depressed moods, loss of interest, significant weight loss, insomnia or hypersomnia, psychomotor agitation or retardation, loss of energy, feelings of worthlessness, diminished ability to concentrate, and suicidal ideations (American Psychiatric Association, 2013). As discussed above, depression can be a comorbidity of PTSD and chronic pain. MDD is also commonly associated with GAD having comorbidity rates of 58.4%-71.7%. This is most likely due to the similarities in neurobiological mechanisms associated with anxiety and depression (Zhou et al., 2017). MDD and other types of depressive disorders are closely linked to other common comorbidities found in polytraumas. This makes MDD an important comorbidity to consider when researching polytraumas.

2.1.1.4: Introduction to GAD

General anxiety disorder (GAD) is an anxiety disorder that is characterized by worry, fear, and constantly being overwhelmed (Munir & Takov, 2021). GAD presents symptoms such as excessive anxiety, difficulty controlling worry, "[being] on edge[,] easily fatigued[,] difficulty concentrating[,] irritability[,] muscle tension[, and] sleep disturbance" (American Psychiatric Association, 2013). GAD and MDD overlap in diagnosis, sharing four symptoms per the Diagnostic and Statistical Manual of Mental Disorders (DSM) – Fourth Edition – Text Revision (American Psychiatric Association, 2000; Zbozinek et al., 2012). GAD is also one of the most prominent comorbidities in polytrauma (Noyes, 2001). A study from the University of Vermont showed that 25.4% of patients with PTSD presented symptoms of GAD (Price et al., 2019). GAD is a common disorder and is often found with MDD. Due to this, GAD is both a commonly researched disorder on its own and with other comorbid disorders found in polytrauma.

2.1.1.5: Introduction to PCS

Post-concussive syndrome (PCS) is the occurrence of concussion symptoms after the normal healing time from a concussion (Mayo Clinic Staff, 2020). Symptoms of PCS include headache, fatigue, vision changes, confusion, insomnia, and learning difficulties (Permenter et al., 2022). PCS can also cause long-lasting effects on a person's cognitive skills, memory, and executive function (Permenter et al., 2022). PCS is a part of the Polytraumatic Clinical Triad with chronic pain and PTSD (Lew et al., 2009). PCS has also been linked to anxiety and depression. In particular, one study showed that those with PCS had at least a 52% risk of developing anxiety and a 42% risk of developing depression. PCS is a common disorder for those who suffer physical trauma and mild traumatic brain injuries (Mayo Clinic Staff, 2020).

2.1.2: Populations Affected by Polytrauma

Most psychological conditions have a wide range of victims with a variety of backgrounds and triggers. The most common causes of polytrauma are physical or long-term traumas, typically vehicle collisions, childhood trauma, active military duty, and gender basis.

Car crashes often meet all of the qualifications for inflicting polytrauma. While focused on one incident, the emotional toll can be as significant as the long-term trauma commonly present in polytrauma (Peixoto et al., 2018). In general, the injury sustained must be reasonably serious, or the crash must be severely traumatizing for this response to be provoked. Symptoms of polytrauma at the crash may not be immediately noticeable. However, it is still possible to develop polytrauma, and in particular, late-onset PTSD is highly probable in motor vehicle accidents (Bryant & Harvey, 2002).

Childhood trauma has also been noted as a possible cause of polytrauma. Often the least physical of the conditions discussed in this section. Agorastos et al. (2014) showed that many people diagnosed with PTSD as adults have been able to link the cause to abuse in their childhood. The observation related personal polytrauma to the development of other negative health effects, such as an increased tendency towards other conditions, making polytrauma likely in this population (Agorastos et al., 2014).

Soldiers who have seen active combat are disproportionately affected by polytrauma. Indeed, soldiers are the vast majority of polytrauma diagnoses (Laughter et al., 2021). Cases in soldiers are much more likely to be diagnosed than those in civilians because of the knowledge of the risk of developing PTSD for soldiers. However, PTSD and polytrauma are not unique to soldiers; many situations can produce similar traumas that lead to polytrauma. Experiences of repeated trauma such as domestic abuse, warzone experience, homelessness, incarceration and traumatic instances like car crashes and natural disasters all meet the criteria of conditions that cause extensive trauma and physical strain over several months (Peixoto et al., 2018; Sareen, 2014). These situations are only starting to be understood as producing associated conditions instead of minimally related trauma responses. The combination of PTSD and chronic pain are often the precursors of comorbidities with MDD and GAD. This combination of sustained mental and physical trauma generally provokes polytrauma instead of a more standard trauma response.

There is a disparity of prevalence of polytrauma between genders. PTSD, despite common belief, is more prevalent in women than men in America (Sareen, 2014). This is largely due to the disparity in the amount and severity of sexual assault between the genders, leading to about three times more PTSD in women. However, current polytrauma research is focused on members of the military and veterans because they often have comorbid conditions associated with polytrauma. There is a significant gender imbalance in the military, especially among veterans that have seen active combat. This is partly due to women only recently being legally allowed to deploy in active combat. Hence, most dedicated polytrauma data exists from male samples, despite the possibility of many polytrauma cases in women who have not served.

2.1.3: Effects on Affected Populations

Polytrauma affects various groups of people, but they are all similarly influenced. Polytrauma can present itself with a variety of symptoms from the comorbidities. This intersection of symptoms can allow more dangerous behaviors to arise if left untreated and is ultimately more difficult to treat.

Trauma is more difficult to manage when more than one disorder is present. In general, people with polytrauma are considered "polytraumatized" when they reach the DSM's, or other diagnostic tools' criteria for more than one disorder (Peixoto et al., 2018). Diagnostic tools, such as the DSM, diagnose using symptoms present in a patient. For example, a person is diagnosed with MDD and PTSD if they have greater than or equal to 5 listed symptoms and one repeated described symptom present during the same two-week period (American Psychiatric Association, 2013). A polytraumatic patient is then someone who must vocalize or show a wide variety of symptoms, each related to one or several comorbidities. This could make life more difficult and potentially decrease the quality of life of a person if they must combat a wider variety of symptoms than someone who only has one disorder. Polytraumatized people can be affected greatly by their symptoms and issues that can stem from the initial trauma.

When left untreated, polytrauma can cause larger issues in those affected. These issues include substance abuse, suicidal behavior, ideations, and an overall decrease in quality of life (Griffin, 1990). Substance abuse, the excessive use of drugs or alcohol in a way that is detrimental to the self and their interactions with society, is a common way the polytraumatized population may try to cope with their symptoms and the causing trauma (Griffin, 1990). Of veterans of the Iraq and Afghanistan wars, those with PTSD were twice as likely to report alcohol abuse than those without any polytrauma comorbidities (Smith & Cottler, 2018). This leans into the self-medication hypothesis, stating that substance consumption is used to numb the pain and symptoms of anxiety, depression, and PTSD (Joseph Volpicelli et al., 1999). Substance abuse, and polytrauma, can lead to suicidal behaviors (Esang & Ahmed, 2018; Finley et al., 2015). It was found that people affected with MMD or GAD were more likely to commit suicide than the general population, and these likelihoods only increase when the person has a history of substance abuse (Esang & Ahmed, 2018). Ultimately, substance abuse and suicidal ideations caused by polytrauma can decrease a person's quality of life quite starkly. A better understanding

of polytrauma would produce more reliable diagnoses, allowing people to get the proper care they need.

Polytrauma is often novel and can have a larger impact on people's lives than those presenting a single mental disorder. Polytrauma is harder to cope with since there is a larger scope of presented symptoms due to the variety of comorbidities. In addition, polytrauma has a higher correlation to substance abuse and suicidal ideations than a singular disorder. These come together to show the importance of being able to understand and treat people affected by polytrauma.

2.1.4: Current Polytrauma Research

2.1.4.1: Papers of Particular Interest

There are many studies on each of the conditions that make up polytrauma independently (depression, anxiety, PTSD, and chronic pain). A strong understanding of the current study of polytrauma as a category is important to be able to predict the needs of researchers and the following general papers allow for a thorough understanding of the subject.

"Prevalence of Polytrauma Clinical Triad Among Active-Duty Service Members" is one of the first long-term studies explicitly made to study polytrauma and its effect on the military (Laughter 2021). This study contains general statistics about polytrauma in the military, which is the current pool of polytrauma victims. It is well constructed with many resources that communicate to the reader the statistics of polytrauma in the military and the current understanding of its effect on their lives.

"Prevalence of chronic pain, posttraumatic stress disorder, and persistent post-concussive symptoms in OIF/OEF veterans: Polytrauma clinical triad" is similar in some ways to Laughter et. al. but is focused on veterans (Lew et al., 2009). This is a good source on veteran experiences of PTSD, albeit with a smaller sample size than Laughter. It has similar information on the commonness of these conditions, which comorbidities are the most common, and how these problems tend to happen.

Blakey et al., 2018 looks at depression in relation to the classical polytrauma trio of PTSD, chronic pain, and traumatic brain injury (TBI). This paper explores how those with polytrauma can develop depression and sometimes its common comorbidity anxiety, and how that can lead to suicidal ideation and extreme states of depression.

2.1.4.2: Data Types

Polytrauma research focuses on determining the correlation between comorbidities by viewing the physiological symptoms and/or neural causes. But they all do so differently. One of the main ways the research can be differentiated is through the data types they use to determine these causes. The majority of polytrauma uses log and neural data in their analyses.
Log data consists of any data that is written down and is not a biosignal. This could be answers to a survey, notes from an interaction with another person, or self-written responses. Most polytraumas are diagnosed with a psychiatrist and diagnostic tools such as the DSM-5. Polytrauma log data takes the form of notes on symptom presence per the DSM 5 and is often

taken to view the overlap between comorbidities. Cayden Peixoto et al., viewed the overlap between the presence of chronic pain and PCT after a motor vehicle crash by using a variety of questionnaires based on current diagnostic tools such as the DSM-5 (Peixoto et al., 2018). Similarly, research from the University of Vermont viewed the correlation between PTSD, MDD, and GAD by using the DSM-5's criteria in a Human Intelligence Task survey (Price & van Stolk-Cooke, 2015). Log data is the most common form of data and the easiest to collect. This, and the current presence of symptom-based diagnostic criteria, has made log data useful in polytrauma research.

Neural data includes any sort of biosignals taken from the brain either invasively or non-invasively. In polytrauma research, it is common to take neural data in the form of electroencephalograms (EEG) or functional near-infrared spectroscopy (fNIRS), two common methods of non-invasive neural data collection (Sheerin et al., 2018). Neural data offers direct windows into the neurological causes of polytrauma and their comorbidities' symptoms. Sheerin et al., used EEG data to prove the neurobiological basis of four PTSD symptom factors in veterans of the Iraq and Afghanistan wars (Sheerin et al., 2018). Similarly, current research has determined the use of fNIRS to create biomarkers to better understand the severity of fibromyalgia, a condition whose primary symptom is chronic pain (Daniela Gabiatti Donadel et al., 2021). Non-invasive neural data is an excellent way to determine the causation of various symptoms and their severity directly. As such, neural data has become an important part of polytrauma research.

Speech data is an emerging technology in the research field of polytrauma and its comorbidities. Despite not being formally used to find causes of polytrauma, speech data can help solve some of the longstanding issues with polytrauma diagnosis. The first of these issues is that the usual self-reports or clinical assessment measures can often over and under-report symptoms. Speech data is both commonly available and can be used to predict 89% of PTSD cases (Marmar et al., 2019). For MDD, speech patterns were deemed statistically robust, such that they can be used to recommend patients for evaluation (Mundt et al., 2012). These examples of speech data represent how promising speech is as a new form of psychological data and why researchers will need a flexible way to work with this type of biodata. It also has the unique advantage of being backward compatible in a way many data types are not. This lends to a decently sized backlog of data on conditions as interview recordings and the like are standard in many places.

In the context of polytrauma, log data is standard and useful for formal medical diagnosis. In contrast, neural and speech data are emerging as the basis for biomarkers that can promote disease progression diagnosis and tracking, and they have proved their importance to polytrauma research. Using log, neural, and speech data together could improve the current understanding of polytrauma's physiological and neurological causations.

There is still much research to be done on the topic to better understand the disorders' physiological and neural casualties. This research is necessary to create better standards of care for the polytraumatized population. Current research has shown promising discoveries in the

neural and auditive datasets regarding polytrauma as a whole and their comorbidities individually. Creating a tool that would simplify the analysis of polytraumatic data sets could increase and improve research in this area.

2.2: Biomedical Data Pipelines

A biomedical data pipeline (a.k.a biodata pipeline) is defined in this paper as a system that stores and processes biological signals for scientific analysis. They contain various analysis tools such as toolboxes and algorithms. They also provide the ability to easily manipulate and share data. Understanding what makes a pipeline useful for researchers is key to building robust pipelines. Biomedical data pipelines are used for a variety of purposes including research, clinical trials, diagnostics, and medical devices. However, biodata pipelines are most commonly created for and by academic researchers leading to many hyper-specific biodata pipelines.

Researchers from the Journal of Pediatric Health Care produced a biodata pipeline that takes in diagnostic and medical data from a patient, including family history and vitals over time (Honigfeld et al., 2017). It is similar to the typical medical databases that doctors use. The pipeline was made due to the increased need for diagnoses of anxiety and depression in adolescents. Researchers found that having a tool that makes diagnoses easier leads to diagnosis screenings happening more often. This study also shows that multiple levels of analysis can be beneficial. The algorithm used in this study first marked and counted symptomatic behaviors and trends in the patient's existing medical data to determine if the patient should be recommended for additional screening. The algorithm then used the data from the screening to make a final assessment. This algorithmic strategy of using the initial information to rule out cases and request more information on the remaining cases is a good way to efficiently collect and process data. The clinician using the pipeline is given the ability to override the algorithm and take additional steps, if necessary, based on their insight regarding the case. The implementation of this pipeline is a proper example of user-focused design, as while the algorithm offers recommendations the clinician using the pipeline provides the final diagnosis. This pipeline gives both direct non-prescriptive advice and provides logic to justify its advice.

Researchers from University College London reflect on the various challenges that researchers face when trying to use medical data for biodata analysis (Denaxas et al., 2016). Many of these challenges surround issues regarding organization and permissions. Such issues can be resolved by an effective pipeline. When data is well organized and tags have been properly applied, pipelines are more efficient and easier to use. This paper shows, through its discussion of current biodata issues, that many users want a system that can anonymize data securely and store that data in a manner conducive to collaboration.
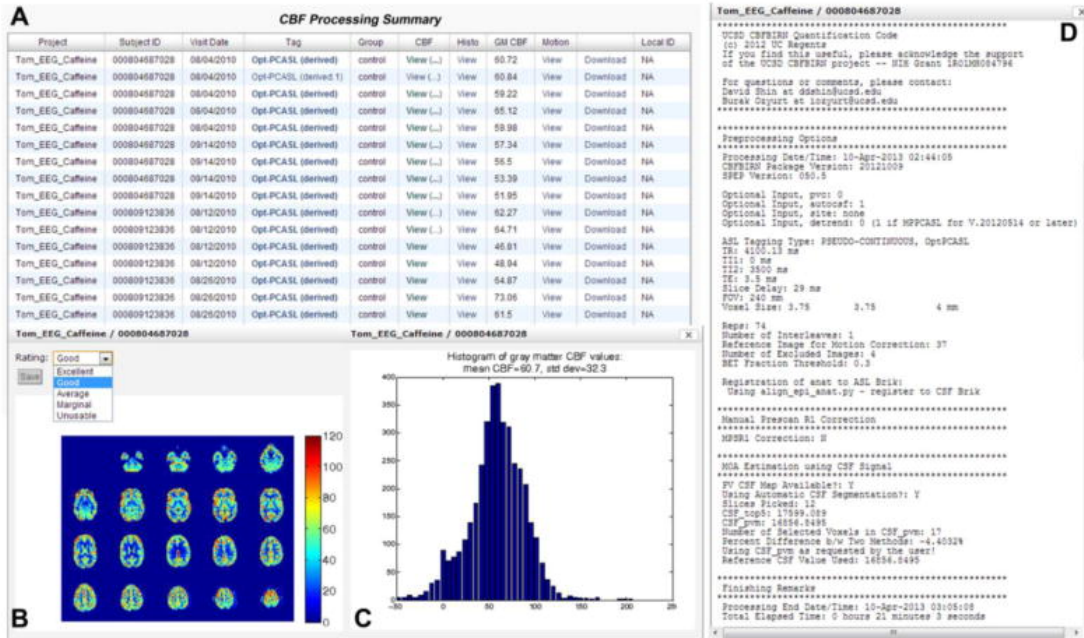
Figure 1: The CBFBIRN data visualization screen (Shin et al., 2016)

The Cerebral Blood Flow Biomedical Informatics Research Network is another example of a pipeline (Shin et al., 2016). The CBFBIRN is a highly specific and technical pipeline. The display for this pipeline uses many small windows to compartmentalize various data visualizations and explanations of the current data. This style provides a useful visual representation of stored and analyzed multimodal data.
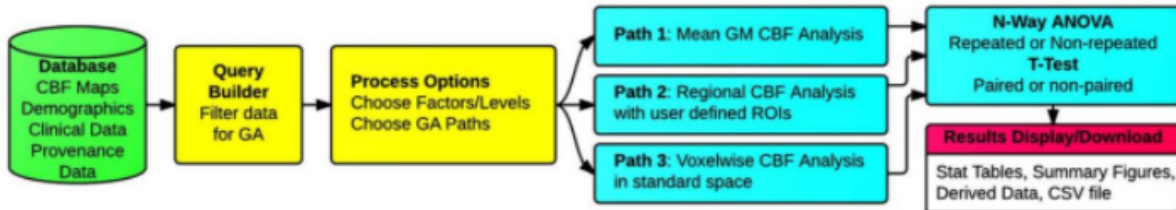
Figure 2: CBFBIRN data flow (Shin et al., 2016)

This pipeline's database implementation allows for data flow and is consistent with the current industry database architecture and structure. In particular, this pipeline provides helpful features such as a query builder and a method to choose and display analyses. The CBFBIRN system also provides data visualization and database tools. However, the system lacks rigorous analysis capabilities. Significantly expanding the analytical scope of this system was noted as a strong possibility for exploration.

2.2.1: Current Bio-data Analysis Tools

There are few biomedical data analysis tools on the market that suit the specific needs of polytrauma research. To better understand what is needed by researchers, one can view the

current software that is used in the healthcare and biomedical research fields. The current market can be viewed in three sections: machine learning and algorithms, toolboxes, and pipelines for med-tech creation.

2.2.1.1: Machine Learning in Medical Diagnosis & Research

Approximately 10 percent of patient deaths and 6 to 17 percent of hospital complications can be attributed to diagnostic errors according to The Institute of Medicine at the National Academies of Science, Engineering, and Medicine (Faggella, 2020). The researchers attribute these errors to inefficient collaboration and integration of health and information technologies, gaps in communication among clinicians, patients, and their families, and a healthcare system that does not adequately support the diagnostic process. A review by Johns Hopkins researchers of 25 years of malpractice claims payouts in the U.S. found that diagnostic error claims amounted to a sum of 38.8 billion dollars (Faggella, 2020). With the goal of reducing fatal diagnostic error, many researchers and companies are leveraging machine learning and artificial intelligence to improve medical diagnostics (Faggella, 2020).

Researchers predict a 44.9% growth for medical AI in the next five years (*10 Ways to Use Machine Learning for Medical Diagnosis | Fayrix*, n.d.). There are many fields in which the use of machine learning for medical diagnosis could be beneficial. Some of those fields are oncology, pathology, dermatology, genetics and genomics, and mental health. In oncology, machine learning could be used to aid in the detection of malignant tumors or in the classification of whether a given tumor is benign or malignant. In pathology, machine learning could be used to improve the precision of blood and culture analysis by using automated tissue and cell quantification, mapping disease cells and flagging areas of interest, creating tumor staging paradigms (models), or increasing the speed of profile screening. In dermatology, machine learning could be used to separate melanomas from benign skin lesions or track changes in skin moles to help detect skin conditions early. In genetics and genomics, machine learning is often used in preventive genetics and could be used to improve the efficiency of gene editing for purposes such as protecting a fetus from a mutation or fighting genetic diseases such as Cystic Fibrosis and Huffington Disease (*10 Ways to Use Machine Learning for Medical Diagnosis | Fayrix*, n.d.). In mental health, machine learning could be used to personalize cognitive-behavioral therapy (CBT) through the use of chatbots and virtual therapists, identify people or groups with a high risk of suicide who can then be provided with support, and create early detection systems which help diagnose individuals with a variety of mental disorders included but not limited to depression, anxiety, and PTSD (*10 Ways to Use Machine Learning for Medical Diagnosis | Fayrix*, n.d.).

2.2.1.2: Algorithms in Medical Diagnosis & Research

There are many machine learning algorithms being used in the medical field, including those that aid in the diagnoses of different diseases and conditions such as Random Forest, Naïve Bayes, deep neural networks, and support vector machine (SVM) classifiers.

Researchers used the classification algorithms Random Forest and Naïve Bayes to predict whether or not a given patient had a particular disease. They trained these algorithms with disease datasets for diabetes, coronary heart disease, and cancer. After the algorithms were trained, they calculated the performance of each model that they produced on the test data compared to the existing results. The Naïve Bayes classifier makes its model under the assumption that all features are independent. The Naïve Bayes classification network had an accuracy of 74.46%, 82.35%, and 63.74% for diabetes, coronary heart disease, and cancer respectively (Jackins et al., 2021). The Random Forest model had an accuracy of 74.03%, 83.85%, and 92.40% for diabetes, coronary heart disease, and cancer respectively (Jackins et al., 2021). The models performed similarly for diagnosing diabetes and coronary heart disease, with Random Forest performing slightly better. However, for diagnosing cancer Random Forest performed significantly better than Naïve Bayes with a difference in accuracy of 28.66%. Overall, both the Random Forest algorithm and the Naïve Bayes algorithm were able to correctly diagnose diseases the majority of the time (Jackins et al., 2021).

DeepGene is an advanced deep neural network-based classifier whose aim is to accurately classify somatic point mutation-based cancer (Yuan et al., 2016). DeepGene first implements clustered gene filtering (CGF), focusing the gene data by mutation occurrence frequency, which filters out the majority of irrelevant genes from the dataset. Next, an indexed sparsity reduction (ISR) is applied to the dataset which suppresses the impact of data sparsity by converting the gene data into indexes of its non-zero elements. Finally, the data is fed into a deep neural network classifier which extracts high-level features from the data to produce a model that classifies cancer type. The testing accuracy of DeepGene was 24% higher than three other widely adopted classifiers that share the same goal (Yuan et al., 2016).

Researchers have developed a support vector machine to be used for disease diagnosis. The performance of support vector machines depends on the kernel function that it is supplied with. Possible forms for the kernel functions are linear, polynomial, and radial basis functions (Tania & Chandra Shill, 2019). However, it is difficult to achieve high accuracy and performance when using only one kernel function. To combat this, researchers created a hybrid kernel function that combines linear and radial basis functions along with data preprocessing algorithms which reduce the noisiness of the data. This hybrid kernel function allows for increased learning ability and generalization performance. The proposed support vector machine using the hybrid kernel was tested on two breast cancer datasets and one heart disease dataset. It achieved accuracies of 98.57%, 99.46%, and 98.36% with 100% sensitivity for the databases respectively (Tania & Chandra Shill, 2019). These are higher accuracies than ordinary (non-hybrid) kernel functions and other existing methods for support vector machines (Tania & Chandra Shill, 2019).

A chatbot using machine learning algorithms has been developed by Babylon Health with the goal of preventing and diagnosing disease. The chatbot compares the reported symptoms that it receives from the user against a database of diseases, patient history, and patient circumstances to recommend an appropriate course of action. Courses of action may range from recommending

a visit to the pharmacy for over-the-counter medication to recommending a visit to the hospital or a call to an emergency hotline. The chatbot also integrates information such as heart rate and cholesterol level from wearable devices when such information is available. The chatbot was tested by patients from two hospitals in Essex, U.K. Both hospitals reported a reduction in patient waiting time following the release of the chatbot. However, it is unclear how much wait time was reduced and if such a reduction can be attributed to the chatbot (Faggella, 2020).

2.2.1.3: Toolboxes in Medical Diagnosis & Research

Each day there is a greater need for tools that will make the storage, handling, and analysis of big biomedical data easier and more accessible. These different biomarkers and bio-signals have no meaning in isolation, which is why many researchers and clinicians turn to digital toolboxes to handle this data.

A toolbox is a package of functions and classes that provide tools for a specific topic. Many toolboxes are open-source. Due to this, they have become popular in the biomedical and healthcare world for their ability to compute, integrate and analyze data. However, flaws, missing aspects, and complicated user interfaces often cause limited reliability on these toolboxes. Holistically, an ideal toolbox must incorporate different aspects of data gathering, handling, and analyzing while providing a user-friendly interface that researchers and clinicians can grasp with ease.

There are many distinct open-source toolboxes that target specific conditions and data. An example open-source toolbox is the Neurophysiological Biomarker Toolbox (NBT) which is used for the computation and integration of neurophysiological biomarkers. It aims to make biomarker research easier at all levels. From having raw data and cleaning it, to calculating biomarkers and performing advanced statistics. This toolbox has evolved from only handling biomarkers based on EEG or MEG signals, to supporting almost any type of biomarker data.

2.2.2: Pipelines in Medical Diagnosis & Research: GILBERT

2.2.2.1: Form and Function

Greatest Instrument for Leveraging Biomedical Electronic Resources Together (GILBERT) is a biomedical data pipeline that can store and analyze a wide variety of biomedical data that was created as a MQP project. It consists of a web application hosted in Heroku that uses a Django Python-based framework and stores its data in PostgreSQL. GILBERT implements four key functionalities: the uploader, study analysis, inventory, and the user system (Bernardo et al., 2021).

As seen in Figure 3 below, this web application starts by ingesting datasets from different studies to promote data sharing while maintaining patient privacy standards. The study data is populated into a centralized database that works in conjunction with different physiological data signals including discrete, analog, and digital. The front end gives researchers and clinicians the ability to query data across one or many studies. From here, the user has additional control over the analytical method and desired output. These outputs include descriptive metrics about the data, including analysis-ready datasets, and a variety of data visualizations (Bernardo et al.,

2021).
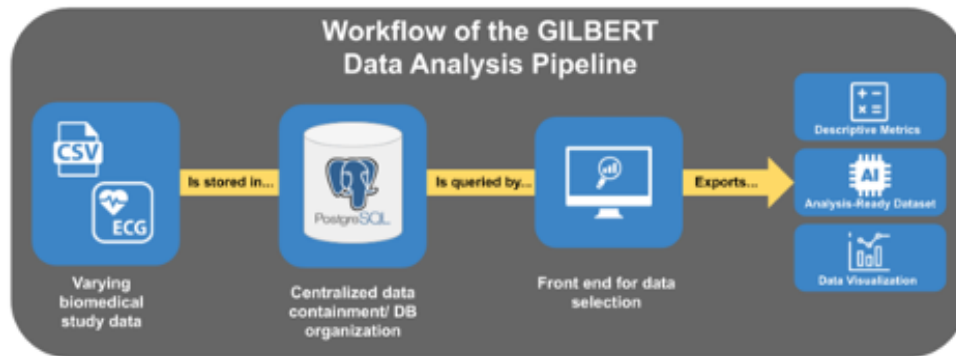
Figure 3. GILBERT data analysis pipeline for biomedical signals

GILBERT was created as there are few centralized options for storing the vast amounts of data collected in healthcare and biomedical research fields. It fulfills its purpose as a hub for researchers to upload, access, filter, and analyze biomedical data across multiple studies. This web application has the potential to be an effective tool for researchers who need a centralized data pipeline to store and assess a variety of biomedical data (Bernardo et al., 2021).

2.2.2.2: GILBERT's Shortcomings

GILBERT, as explained above, has various functions that make it a potential skeleton for a polytrauma research pipeline. However, GILBERT falls short in areas including its database design, data analysis, user interface, study management, and the uploader feature.

The GILBERT database design currently uses a relational database and the creation of tables when various data types are uploaded. Another limitation of the current GILBERT database design is the rigidity needed between datasets when conducting cross-study analysis. All datasets must have the same format to be analyzed with one another. Finally, the GILBERT database is not prepped to contain various kinds of data such as neural data or audio data (Bernardo et al., 2021)**.**

GILBERT's options for data analysis are limited. The current options include the ability to view the data as either a scatter plot or histogram and use one filter at a time over the total dataset rather than a specified range. GILBERT is also not capable of analyzing datasets often used in polytrauma research such as audio, visual, and neural datasets. Improvements could be made to allow for machine learning analysis techniques, better filter options, and better variety in analyzable datasets. If this could be done then GILBERT would have more value for polytrauma researchers (Bernardo et al., 2021).

The design of GILBERT is well done; however, improvements are needed for the user interface (UI). UI issues include a lack of consistency in button placement between pages, difficulty accessing the help manual, and insufficient descriptions of the terminology. Improving GILBERT's UI would make the system more easily accessible and desirable to users (Bernardo et al., 2021).

GILBERT does not have the option to edit or delete uploaded data. This means that metadata, study groups, subjects, data categories, and attributes cannot be modified. This could make research work difficult as new data may create new insights and the deletion or addition of new parameters and conditions. Adding the ability to edit study management data from the front-end would improve the usability of the pipeline (Bernardo et al., 2021).

Finally, GILBERT's uploading feature is functional but has some issues. For example, it does not detect duplicate files well, nor does it support data types other than CSV. Creating a better system for the detection of duplicate files and supporting more data types such as neural, video, audio, and log data would improve GILBERT's ease of use in general as well as for polytrauma research. (Bernardo et al., 2021).

GILBERT offers well-built foundations of a biomedical data pipeline for polytrauma research. However, its database design, data analysis, user interface, study management, and uploader tools are not perfect. By understanding where GILBERT falls short, one can determine where to begin modifying it to improve its functionality for polytrauma research.

2.3: Agile Methodology

Agile is an iterative approach to software development and project management that features short feedback cycles and continuous improvement (WiA). The Agile method aims to: satisfy the customer, allow flexibility, deliver constant updates, nurture collaboration and conversation, focus on software development, keep a constant pace and attention to technicalities, remain simple, have self-organizing teams and allow the team to reflect and pivot if needed (*Principles behind the Agile Manifesto*, n.d.). There are many Agile frameworks, the most popular being Scrum and Kanban (Atlassian, n.d.-a).

2.3.1: Scrum Meetings

Scrum is based on continuous learning and improvement. Teams using scrum work in set intervals called sprints, which could be 2 to 4 weeks long. The goal of scrum is to have a valuable increment of work at the end of each sprint that you can learn from and improve upon. Members on the team have set roles such as scrum master, product owner, and the development team. The scrum master helps the team use the scrum principles. The product owner helps the team prioritize the work that is done, manages the product backlog, and advocates for the customer. The development team chooses what work to do, delivers the increments of software, and demonstrates accountability.

2.3.2: User Stories

User stories are ways to explain the use of a software feature as used by the end user to "articulate how a software feature will provide value to the customer." (Atlassian, n.d.-b). User stories themselves are not features, but rather end goals of the end user. The end users are internal and external customers and colleagues who will depend on your software product. User stories are used by Agile frameworks to facilitate and monitor the workflow of teams. User stories can be grouped together as an epic and themselves be grouped together as initiatives. User stories keeps the product and design team focused on the user's needs and encourages teamwork

and creative solutions for solving an end goal. User stories are expressed as, "persona + need + purpose." (Atlassian, n.d.-b). The persona is who the product is being built for. The need is the end user's intent and end-goal. The purpose is how this need satisfies the end-goal of the system.

2.3.3: Sprints

The velocity of a sprint is the number of user stories that are completed. This is useful to know when planning future sprints. The product backlog is a list of features, requirements, enhancements, and fixes that act as a "To Do" list for the team which is maintained by the product owner or product manager. The sprint backlog is the items, user stories, or bug fixes that have been selected for implementation during the current sprint by the team. An increment is the usable end-product of a sprint.

Teams have a series of meetings every sprint such as organizing the backlog, sprint planning, daily scrum or stand up, and sprint retrospective. It is not required for teams to complete all of these meetings, as they can seem repetitive. Organizing the backlog is run by the product owner who uses feedback from users and the development team to keep the list clean and properly prioritizing. Sprint planning meetings happen at the beginning of a sprint and are where the team decides what the goal of the sprint is and what user stories will be added to the sprint to complete this goal. The daily scrum or stand up is a short daily meeting that usually happens in the morning where every team member describes what they completed the previous day, what the plan to complete today, and if they are facing any obstacles. Sprint retrospectives happen after a sprint where the team comes together to discuss what did and did not work well during the sprint with the goal of improving the next sprint.

2.3.4: Kanban

Kanban allows teams to match the amount of work in progress to the team's capacity. Kanban teams revolve around a Kanban board which they use to visualize their work and optimize the flow of work. A basic Kanban board has sections for to do, in progress, and done. Work items are represented as their own cards on the board. Cards have a brief description of the job, how long it is estimated to take, and other valuable technical information. Moving these cards around the board allows team members to track their work progress and quickly identify any blockers or dependencies.

Kanban allows for planning flexibility as the team is focused only on work that is actively in progress. When a team member completes a work item, they take a new one off the top of the backlog. The order and contents of the backlog are managed by the product owner such that the most important work is at the top. Cycle time is the time that it takes for a unit of work to travel through the team's workflow from start to finish. Kanban teams aim to optimize their cycle time by spreading knowledge among the team to prevent bottlenecks in productivity. Limiting the amount of work in progress is another key to optimizing the cycle time as the more context switching between different tasks can lead to them taking longer to complete. Teams often use control charts and cumulative flow diagrams to check on their workflow. A control

chart displays the cycle time for each item and a rolling average for the team. A cumulative flow diagram displays the number of items in each state as a function of time.

2.4: Stakeholder Analysis

Stakeholders are the groups and people that will, directly and indirectly, be affected by the development of the product (Yock et al., 2015). In the case of the creation of a polytrauma data pipeline, the major stakeholders include researchers, the polytraumatized communities, and healthcare diagnosticians.

The researcher would be the stakeholder directly working with the polytrauma pipeline. The researcher needs analysis tools, supported data types, functional features, and a usable product. Researchers such as Dr. Jean King and Dr. Adam Lammert expressed their needs for the pipeline and explained how they would use the system. If supplied with the correct tools and features, the researchers will be more likely to use the pipeline, improve their data analysis, and ultimately produce more data regarding polytrauma.

Mental health carers and diagnosticians such as family, social workers, and psychiatrists will be affected by the work developed by the researchers using the polytrauma pipeline. These people will have access to a more recent and complex understanding of symptoms and neural bases for polytrauma's effect on populations. The new information will allow mental health carers to better monitor symptoms and treat the patient. For diagnosticians, the new information regarding polytrauma can be used to better diagnose polytrauma using non-invasive methods.

The improved diagnostics, symptom monitoring, and polytrauma treatment based on research developed using the polytrauma pipeline would directly impact polytraumatized lives. The polytraumatized population would have access to more functional diagnostic tools and thus methods to increase their quality of life.

3: Design Process & Implementation

3.1: Design Process

The design process for BATS was based on the previous system, GILBERT, and its team's goals for the future. In order to create a functioning system based off of GILBERT we had to determine the scope of our project, the stakeholders, how the system should work, and the required front and backend functionalities.

3.1.1: Determining Project Scope

Before creating the system, we had to determine what form of data we wanted to focus on. We began researching diseased states of interest to us while cooperating with our preferred data types.

We used a heat map to better understand the various team members' interests and preferences towards various diseased states and types of data (Table 1). The heat maps were created in Google Sheets. The rows represented each of the team members and the columns represented the datatypes we had access to. Each member rated their interest in a datatype from 0 to 3 where 0 was least interested and 3 represented most interested. The three data types with the highest sum would be used in the document.

Preference	Neural	Speech	ECG	Gait	Log data	Facial data	Parkinsons	MDD	TBI	Alzheimers	Fatigue	PTSD	GAD
Claire	3	2	2	1	1	1	3	2	2	2	2	3	
Lauren	2	3	2	1	1	3	2	3	2	2	2	3	3
Nathan	2	2	2	2	2	3	2	3	3	2	2	2	3
Carlos	2	1	1	1	1	2	2	3	3	2	1	3	3
Results	9	8	7	5	5	9	9	11	10	8	7	11	

Table 1: Heat map of biosignal and disorder preferences

The same methodology was used to determine which data types were used in the analysis of various biomedical diseased states where each row represented the presence of a disease state and the columns represented the data types. This heat map was used to determine which data types could be used for the research and diagnosis of different conditions.

After discussing the options, we confirmed our interest in neural and speech data focusing on polytrauma analysis. We then spoke with researchers in fields related to polytrauma who work with speech and neural data within their respective scopes.

3.1.2: Stakeholder Interviews

To better understand the need in the research and clinical world, we spoke with various stakeholders. Our first conversation was with Anna Bean, a counselor with the Student Development and Counseling Center (SDCC) at WPI. The goal of the conversation was to better understand the current diagnostic tools and methods being used for diagnosing various conditions and to learn more about counselor-patient relations. Next, we spoke with stakeholders, researchers who focus on modeling and diagnosing polytrauma's comorbidities using speech or neural data. We had two goals for these conversations. We were seeking expertise from these researchers regarding their research and experience as they are at the forefront of their fields. We also wanted to understand what tools these researchers were using for data analysis, and any issues they commonly ran into when using them.

By speaking with stakeholders and potential users of the polytrauma pipeline, we were able to determine their needs and define the following needs statement: a way to improve data analysis and data storage for polytrauma researchers to facilitate multifaceted polytrauma research.

3.1.3: Design Matrices

During our interviews with speech and neural researchers, we asked them what systems their labs currently used for data analysis. We asked them what they liked and disliked about their current system(s). These questions allowed us to determine what aspects of analysis toolkits were most important to users. Using this information, we were able to decide what features we should add to our toolkit.

3.1.4: Feature Selection - Pairwise Comparison

To determine the importance of the toolkit functionalities for each data type (neural and speech)*, we completed a Pairwise Comparison Chart (PCC). The functionalities were each placed in a row and column. The two were compared to each other in terms of importance. The functionalities in the rows were ranked on a scale from 0 to 1, 0 being less important than the functionality in the column, 0.5 being the row and column functionalities are equally important, 1 being more important than the functionality in the column. These rankings were then summed and the total values were used as weights when calculating the value of various toolkits for future analysis.

	Open Source	Easy Structure	Visualization	Noise Reduction	Python	Java	Weighting
Open Source	x	1	1	1	1	1	5
Easy Structure	0	x	1	1	0	0	2
Visualization	0	0	x	0	0	0	0
Noise Reduction	0	0	1	x	0	0	1
Python	0	1	1	1	x	0	3
Java	0	1	1	1	1	x	4

Table 2: Pairwise Comparison Chart of features

3.1.5: Source Selection - Pugh Analysis

Once the features were selected and ranked via importance, we completed market analysis research on current open-source algorithm toolkits that were available to us. These algorithm toolkits would later be used as the foundations for our toolkits. We chose the best toolkits for the scope of our project by completing a Pugh Analysis of the various toolkits to see how they compared against each other. For the various features, the toolkits were compared to a baseline analysis tool used in research. If the toolkit functioned better, equal, or worse than the baseline for a specific feature, it received a score of 1, 0, or -1 respectively. These scores were then multiplied by the feature weights determined in the pairwise analysis to give each toolkit a final score. The toolkit with the highest score was the one we used to connect to our system.

	Weight factor	Baseline (HOMER)	MNE-NIRS	Huppert	NAP	NIRS-SPM
Open Source	6	0	1	1	0	0
Easy Structure	3	0	0	1	-1	1
Visualization	1	0	1	0	0	0
Noise Reduction	2	0	1	1	0	0
Python	4	0	1	-1	0	-1
Java	5	0	0	0	0	-1
Total			6	5	-6	0

Table 3: Pugh Analysis of available fNIRS analysis toolkits

	Weight factor	Baseline (Audacity)	IIRJ	Opensmile	pyAudioAnalysis	Covarep
Open Source	6	0	1	1	0	0
Easy Structure	3	0	1	1	-1	1
Visualization	1	0	1	0	0	0
Noise Reduction	2	0	1	1	0	0
Python	4	0	0	-1	1	-1
Java	5	0	1	-1	0	-1
Total			14	1	-4	0

Table 4: Pugh Analysis of available speech data analysis toolkits

The Pairwise Comparison Charts and Pugh Analysis allowed us to determine that the best open source toolkit systems to implement in BATS were MNE-NIRS and IIRJ for neural and speech analysis respectively.

3.2: Implementation

Our implementation was divided into two sections, the algorithms and the system. This was done so that the algorithms team did not have to wait for UI work to be completed to implement the desired algorithms.

3.2.1: Algorithm Development

Before linking the algorithms to the BATS system, the team used the stakeholder interviews to determine what tools the system needed and what open-source systems had them available. After selecting an open-source toolkit via pugh analysis, we removed any excess code to reduce the size of our BATS system and avoid difficulty connecting the backend to AWS. Finally, the "trimmed down" toolkits were connected to the BATS AWS system via jar files.

3.2.1.1: Trim Down

We only wanted to utilize certain functions of the selected toolkits. To do this, we "trimmed down" a large portion of the toolkits to reduce their size within the BATS system. The trim down process meant removing any functions and extra code that were not needed for the algorithms to work with the BATS system. Once trimmed down, the toolkits were turned into jar files that would be read and used by the BATS system.

3.2.1.2: Jar Packages

Jar packages were condensed versions of the created toolkits. These jar packages were how we allowed BATS to access and run the algorithms in our various toolkits. They could be a maximum of fifteen megabytes to work with the BATS system. We did not have enough time to fully connect and test all of our toolkits to BATS, so we created a proof-of-concept jar. This

proof-of-concept jar package included the subject of our toolkits and algorithms that were fully tested. We connected this jar package to BATS to prove that toolkits could be run within the system. However, we required more time to get the rest of the toolkits and algorithms to work harmoniously within BATS.

3.2.1.3: Version Control

While the toolboxes were being modified they were kept in a separate GitHub branch within our repository which was used for the development of toolboxes and algorithms. When a new version was created with the intention of being used by the AWS branch, it was turned into a compressed jar file and then uploaded to google drive with a descriptive name. This jar file was then incorporated into the main AWS branch of our repository.

3.2.1.4: Testing and Debugging

The analysis tools were primarily tested using mainline tests of their core functions that allowed us to determine how each tool worked individually. These tests involved running an algorithm with sample data and seeing if it functioned properly. Once the algorithm worked, it was saved with a demo file and later shown as proof of concept during a team meeting.

3.2.2: Front and Backend

The BATS AWS system consists of a front and back-end. The front-end is hosted on AWS and is accessible through the following link: https://batsmqp.s3.us-east-2.amazonaws.com/presentation/mainPage.html.

The back-end was made using Eclipse and uploaded to AWS lambda as zip files to be used by the front-end. The team chose to use Amazon Web Services (AWS) to host the BATS website and database. The team uses AWS S3 to store its HTML, JavaScript, and CSS code. We use AWS lambda to store our lambda functions, AWS API Gateway to deploy our API, and AWS RDS to host our relational database. We created our database tables using MySQL Workbench and we edit our database based on user interaction using JDBC as accessed through our lambda functions. Our database schema is in Figure 8.

3.2.2.1: Version Control

A first version of BATS was created using a flowchart made in LucidChart. This allowed us to better understand how the pages and features would interact with each other (Figure 4). This was then used to create a low fidelity prototype of the system's front end to ensure that the system followed heuristic guidelines for user friendliness (Nielsen, 1944; Figure 5). With these created, and evaluated using user stories and other evaluation methods, we were able to begin the development of the coded BATS system.
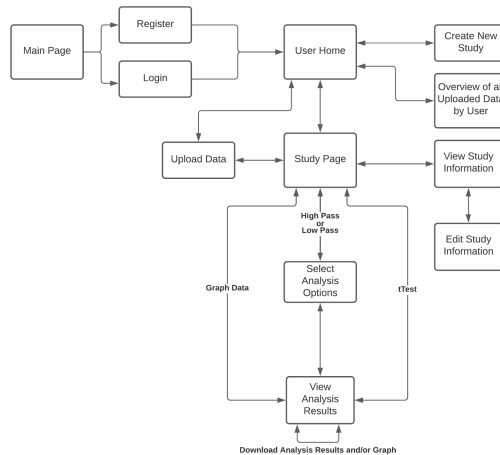
Figure 4: Flow diagram of possible movement for users between pages in the system
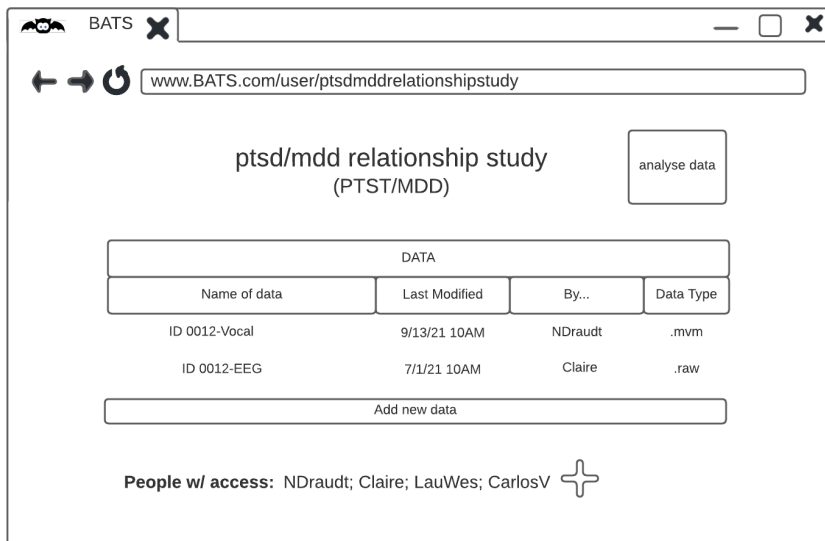
Figure 5: Low fidelity prototype of BATS Study Page

To ensure that we always had a working version of the system, we kept various versions of the system in different branches of our GitHub repository. We kept each other updated regarding the working status of each branch. This version control for the toolkits we were developing allowed the AWS branch developers to use either reliable, older versions of the algorithm branch or newer versions with extra features at any time. In turn, the AWS branch was the main production branch on GitHub, with constant new pushes that contained the current jar operating inside the system. This jar was frequently changed during development to test new features created by the algorithms team.

3.2.2.2: AWS and Programming

We used SwaggerHub to create our API. We made post requests for each of our lambda functions where each post request receives the corresponding request and response models. Each request and response model has different components depending on what the lambda function

takes in and gives out. We exported our API from SwaggerHub as a YAML unresolved file which was then uploaded to AWS API Gateway.

Our lambda functions were created in Java using an Eclipse IDE. We had classes for each of our object models as well as HTTP requests and responses. The lambda functions interact with our database using our DAO that connects with the database via JDBC. Once locally tested using JUnit test cases, the lambda function was uploaded to the AWS S3 bucket as a zip file using Eclipse's AWS lambda function uploader. In AWS Lambda, the lambda function then had its timeout changed to 5 minutes and its VPC and environmental variables set up. The lambda function was then connected to AWS API Gateway and CORS was enabled. All of this ensured that the functions could be called and used by the front-end. After the lambda function was tested in AWS Lambda and AWS API Gateway we deployed the API.

We made all of the required pages for our website in html using Bootstrap, a front-end open-source toolkit. We used CSS to customize our pages and make them more user-friendly. We used JavaScript to connect our lambda functions to our front-end. Once connected, the lambda functions were called as needed. This may involve displaying the requested data or performing a requested action. Next, we tested the current system all together.

3.2.2.3: Testing & Debugging

Testing of the front and back ends occurred first individually, and then together. As features were added to the back-end, they were tested using local JUnit test cases. These test cases checked that features were running correctly and included edge cases. If a test failed, we began debugging the feature in concern to determine and fix the cause of the issue. Once all tests passed, the back-end feature was uploaded to AWS Lambda to be used by the front-end. The lambda functions were also tested in AWS Lambda and AWS API Gateway given test JSON data. Once the local and AWS tests all passed, the lambda functions were hooked up to the front-end which then was tested. We tested the front-end manually, by interacting with the system as a user. We attempted all user actions that we could think of and confirmed that the system worked as anticipated. If there were any issues, we knew to look for bugs within the front-end, because the back-end functionality had already been thoroughly tested. If for some reason an unforeseen use case appears which cannot be properly handled by the back-end lambda function, the lambda function was updated to resolve the issue. and debugging was a complicated process that has had a few stages. The first stage was the individual stage. The AWS system was being tested through a series of tests that allowed the system to establish if the code was running correctly.

4: BATS

4.1: Navigation

The landing page of BATS is the Main Page, as seen in Figure 6, where users can choose to login or register with the system. Once users are logged into the system, they are brought to the User Home page, as seen in Figure 7. BATS is centered around the idea of studies, where studies are projects that researchers are working on. On the User Home page, users see a list of all of their studies. To access a study, users must press on the row of the corresponding study in the study list. This takes users to the Study Page where users can view the data associated with the current study, graph data, analyze data, and view and edit the information associated with the current study.

The navigation bar is available for users from within any page in BATS. The navigation bar can take users to the User Home page, the About page, the View Data page, or the Help & Support page. The About Page displays information about what BATS is and the team members who made it possible. The Help & Support page provides instructions for how to create and access an account, create and view a study, upload and view data, graph data, and analyze data. There is also a logout button in the navigation bar which logs users out of the system and takes them to the Main Page.
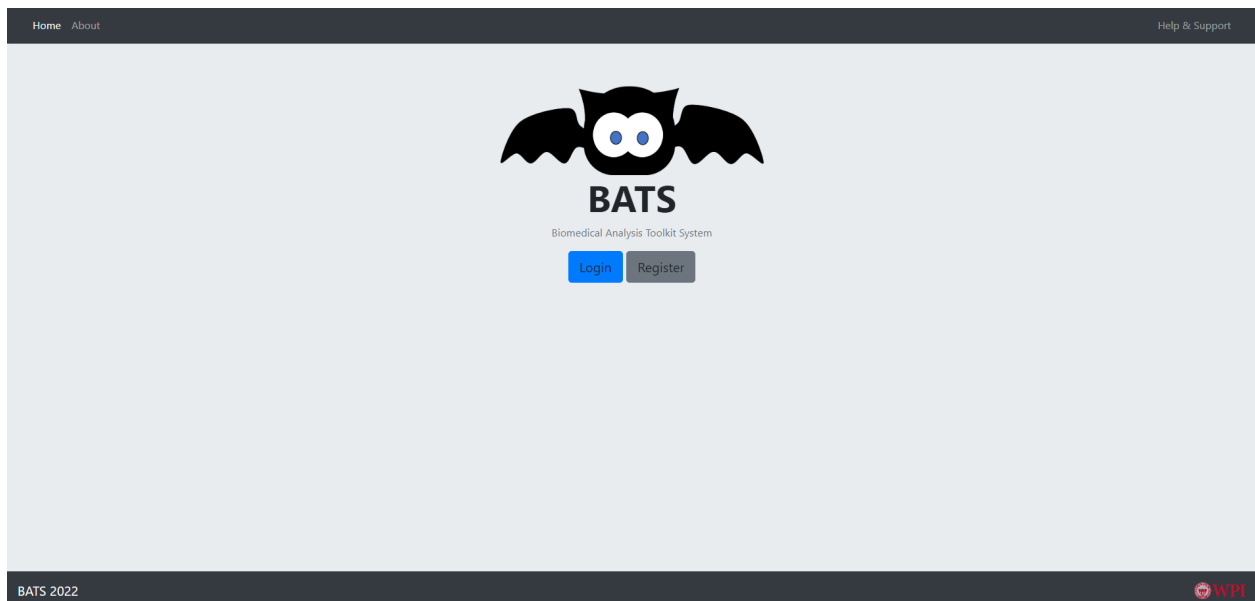
Figure 6: Screenshot of BATS Main Page

Figure 7: Screenshot of User Home Page

4.2: Database Schema

The database kept track of registered users, studies, and documents. Registered users are documented in the AuthUser table. Each user had an associated user ID, username, password, and email. Studies were documented in the Study table. Each study had an associated study ID, study name, study description, institutions involved, study short name, study contact, study notes, visibility status, IRB approval status, start date, end date, last modified date, and user ID. The user ID associated with a study is that of the user who created the study. Documents were documented in the Document table. Each document had an associated document ID, filename, name (given by the user), dataType (neural, audio, log, or other), file extension, and file content. The StudyDocument table kept track of what studies were associated with what documents using the corresponding IDs. The AuthUserDocument table kept track of what users were associated with what documents, also using the corresponding IDs. A visualization of the database schema can be seen below in Figure 8.
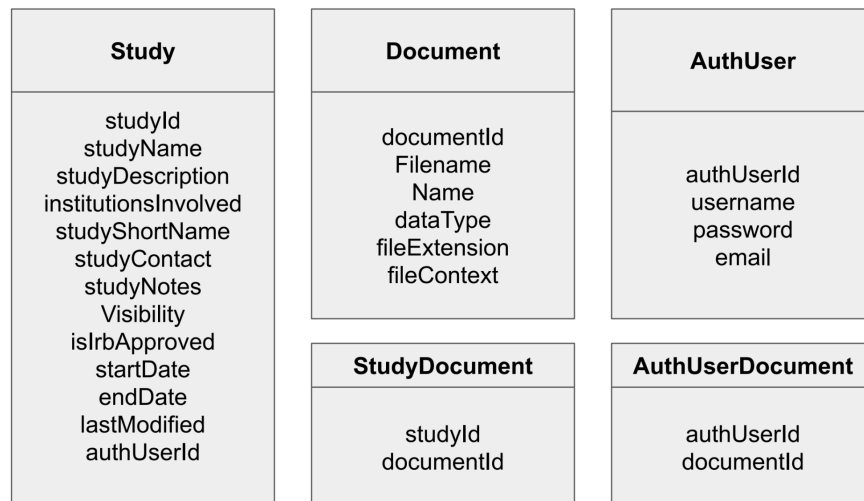
Figure 8: Visualization of database schema

4.3: User System

BATS has a variety of features available for users. These features include the ability to login or register, create a study, upload data, view study data, view study information, edit study information, graph data, analyze data, and view all user data.

4.3.1: Login & Register

New users can create an account from the Register page. To register, users must enter an email, username, password, and password confirmation. If the entered username is already in use, the user will be prompted to enter a different username. If the entered passwords do not match, the user will be prompted to re-enter the passwords. Existing users can access their accounts through the Login page. To login, users must enter their username and password. If there is no account in the system with the entered username, the user will be informed that the entered username is not registered with our system. If the wrong password is entered, the user will be informed that the password is incorrect. Once users have registered or logged in, they will be directed to the User Home page.
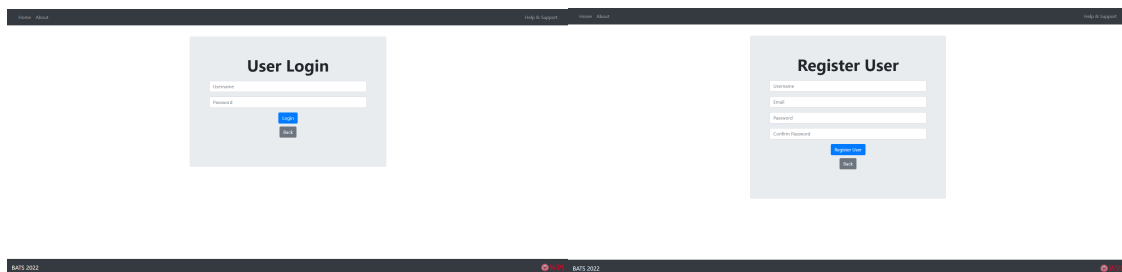
Figure 9: Screenshot of Login & Register pages respectively

4.3.2: Create Study

Users can create a new study on the New Study page, accessible from the User Home page (Figure 10). Users can enter a study name, short study name, abstract, institutions involved, study contact, and study notes as well as whether the study is IRB approved and visible. A study name is required to create a new study. The other information is optional and can be added to the study at a later time. Once users have created the new study, they will be directed to the User Home page.

Figure 10: Screenshot of Create Study Page

4.3.3: Upload Data

Users can upload data to a study from the User Home page or the Study Page. Users can select a data file from their computer, enter a name for the data, select a data type, and select a study to add the data to (Figure 11). The system accepts data files of type CSV, WAV, or NIRS. The data type options to select from include neural data, speech data, log data, and others. If the user has chosen to add new data from the Study Page, then the study to add the data to is preselected to the study the user is viewing.

Figure 11: Screenshot of Upload Data Page

4.3.4: View Study

Users can view a study on the Study Page, accessible from the User Home page (Figure 12). Viewing a study means that users can view the data associated with the study, graph and analyze this data, and access study details. The data is shown in a list and displays the name of the data, the filename, the file extension, and the data type that was selected by the user when it

was uploaded. Users can choose an analysis tool, from the available toolkits, to run on any of the uploaded data. This includes graphing CSV data as well as running a high pass filter, low pass filter, or tTest on CSV data.
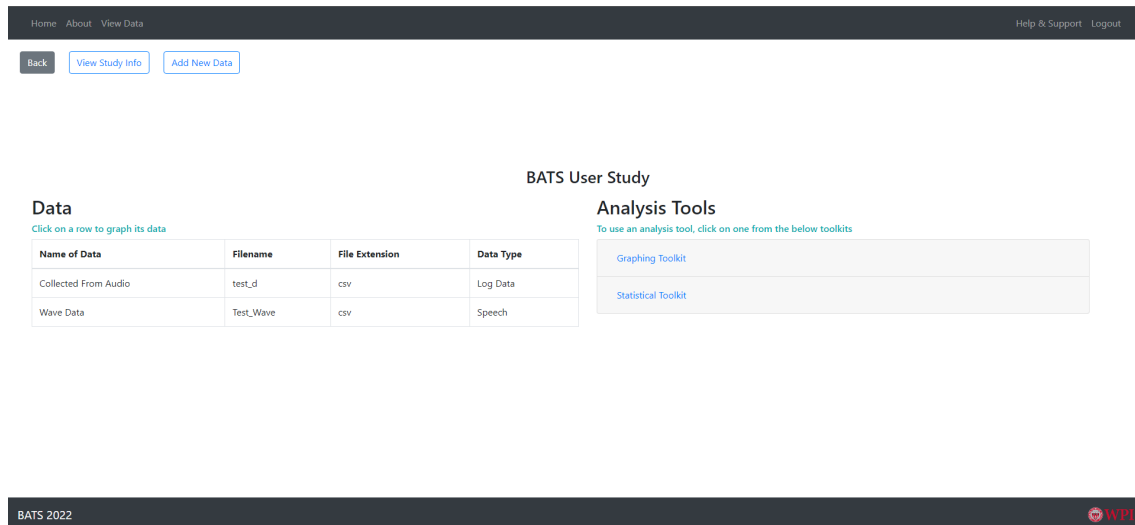
Figure 12: Screenshot of View Study Page

4.3.5: View and Edit Study Information

Users can view study details on the Study Info page, accessible from the Study Page (Figure 13). The study details available include the study name, study short name, study abstract, institutions involved, study contact, study notes, the start and end date of the study, and whether the study is IRB approved and visible. Users can edit any of the study details on the Edit Study page, accessible from the Study Info page. All of the study details are able to be modified except for the start date that is set to the date the study was created in the system.
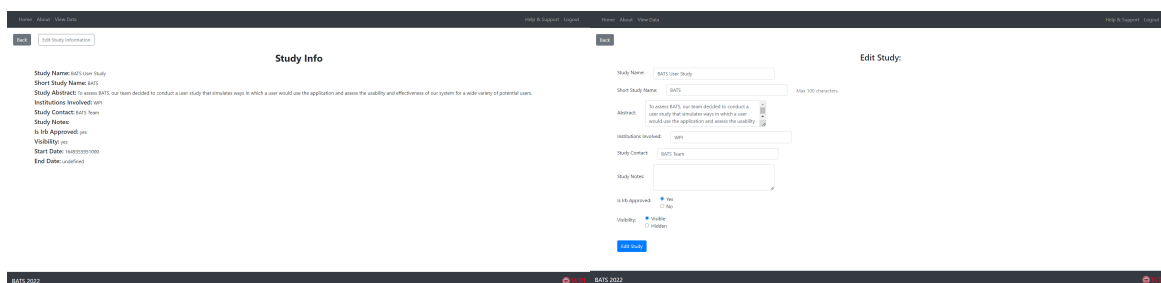
Figure 13: Screenshot of Study Information and Edit Study Information Page respectively

4.3.6: Graph and Analyze Data

Users can graph and analyze data from the Study Page (Figure 14). To graph data, users can press on a row of data or select the Graph CSV analysis tool. If users select the Graph CSV analysis tool, they will be brought to another page to select the data to be graphed. Users can see and download the resulting graph from the Analysis Results page to which they are brought. Similarly, users can run our other analysis tools by selecting the analysis tool's row from an

available toolkit. Users then choose the day to run the analysis tool on and if additional information is required, they are brought to the Options page. Additional information is required to run a high or low pass filter. The results of the analysis are displayed on the Analysis Results page and can be downloaded by the user. This page also shows the analysis tool and data that were selected.
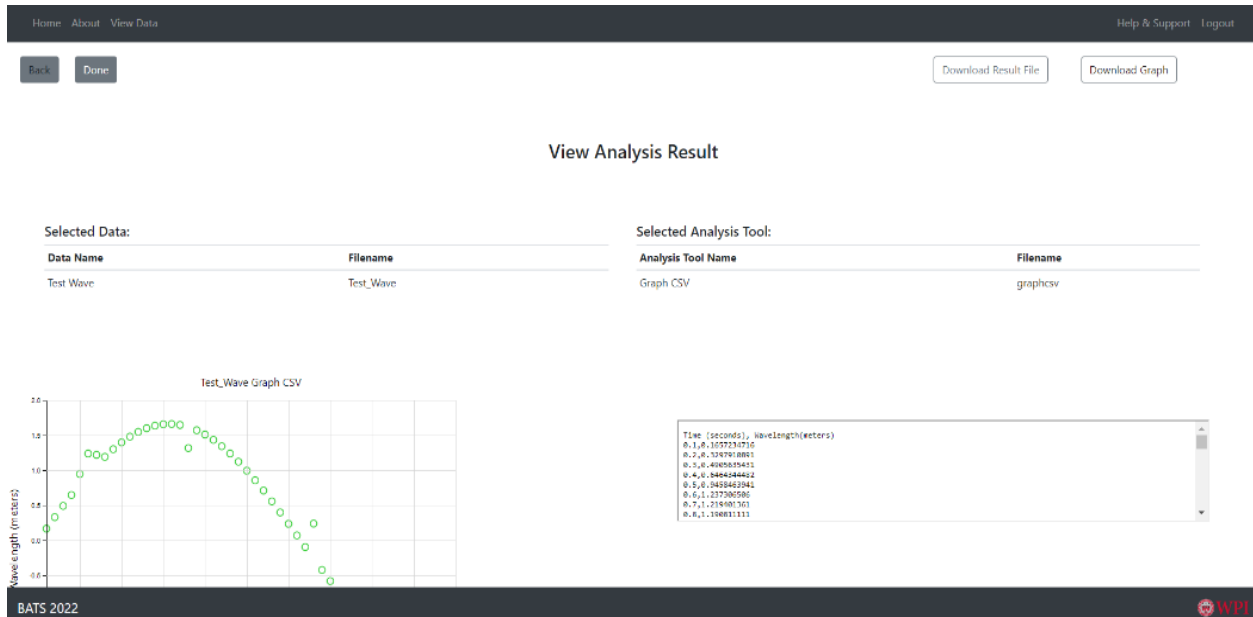
Figure 14: Screenshot of Analysis Results Page

4.3.7: View All User Data

Users can view all of their uploaded data from the View Data page (Figure 15). This page is accessible from the navigation bar which is available to logged in users. The data is a culmination of data associated with all of the user's studies. It is shown in a list containing the name of the data, filename, and data type.

Figure 15: Screenshot of View All Data Page

4.4: Analysis Tools

4.4.1: Graphing Toolkit

Graph CSV Data

This can be used to graph a CSV file. The graph produced is a PNG scatter plot where the first column is the X-axis and the second column is the Y-axis. The graph and the CSV data are downloadable. Graphing a CSV is a useful feature as many data types can be easily converted into CSVs.

High Pass Filter

This can be used to perform a high pass filter on the first two columns of a CSV where the first column is the X-axis and the second column is the Y-axis. This minimizes all sine-shaped waves under a certain length, allowing for larger waves or shapes to be more prominent. This is desirable as it removes interference from other waves. The user has the option to choose a cutoff frequency and order for the filter, which are given default values, as these are two of the most common considerations.

Low Pass Filter

This can be used to perform a low pass filter on the first two columns of a CSV where the first column is the X-axis and the second column is the Y-axis. This minimizes all sine-shaped waves over a certain length, allowing for smaller waves or shapes to be more prominent. This is desirable as it removes interference from other waves. The user has the option to choose a cutoff frequency and order for the filter, which are given default values, as these are two of the most common considerations.

4.4.2: Statistical Toolkit

The user is able to perform a non-paired t test between the first and second columns of a CSV where the column order is interchangeable. The system will produce degrees of freedom, t

value, and p value. The system is also capable of paired and strict size t tests. However, these are not currently accessible in the web version.

4.4.3: Audio Toolkit (Prototype Toolkit)

The audio toolkit is capable of looking at tone, phrase length, and other speech factors. It uses pyAudioAnalysis as a foundation, and has been trimmed down to the features we are utilizing. Using Jython, a python-java interface, we were able to integrate the audio toolkit into our local toolkit system. However, due to space and time restraints the audio toolkit is not available in the web-based version of BATS. Figure 16 is a screenshot of an example graph that the audio toolkit can create through various filters.
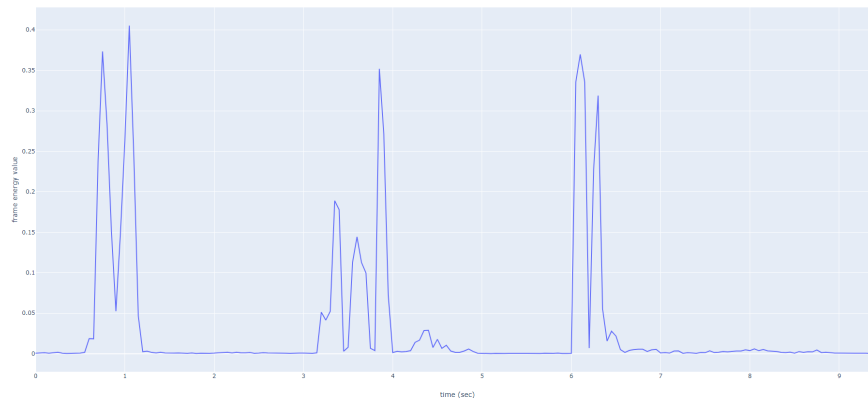
Figure 16: Displays a frame energy value over time
(Consult Appendix B for more information and how to access these files.)

4.4.4: Neural Toolkit (Prototype Toolkit)

Using MNE-NIRS as a foundation, we were able to create a trimmed neural toolkit capable of preprocessing and reading fNIRS files for content and more. Using Jython, a python-java interface, we were able to integrate the neural toolkit into our local toolkit system. However, due to space and time restraints the neural toolkit is not available in the web-based version of BATS. Figure 17 is a screenshot of an example graph that the neural toolkit can create through various filters.
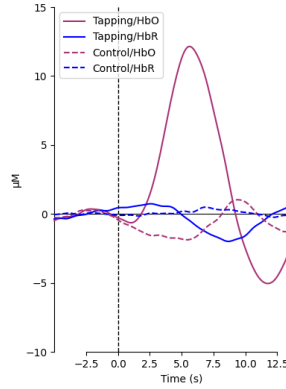
Figure 17: Displays an enhancement algorithm comparison for a single waveform, showing different ways the signal can be interpreted by the system
(Consult Appendix B for more information and how to access these files.)

5: Methods

This chapter discusses the steps taken to create the BATS system. The methodology is divided into three sections: planning, implementation, and testing. These sections allow us to examine the various aspects of the system creation in a relatively chronological order.

5.1: Planning via Agile Methodology

The agile methodology was used to promote efficient work and the creation of various deliverables throughout the terms. Our team used Microsoft Planner to organize and edit our backlog of user stories and tasks. We chose to use Microsoft Planner for our backlog because our email alias also used Microsoft and we preferred to keep our organization within the same system. Our sprints were two weeks each. At the beginning of each sprint we would have a meeting for our sprint retrospective and sprint planning. This included discussing how the previous sprint had gone, our progress and any areas for improvement, as well as deciding and assigning to team members what user stories and tasks should be completed during that sprint. We timed the end of our sprints to align with a weekly meeting that we had with our advisors. This allowed us to show them our progress and incorporate their suggestions and advice into our next sprint.

## 5.2: User Testing

When creating BATS, we had to make sure that it was easy to operate for everyday users of the system. To ensure this, we conducted a user study to assess BATS's usability. IRB approval was received from the WPI Institutional Review Board (IRB) on February 9, 2022. Once we received IRB approval, we conducted a think-aloud and semi-structured interview with 17 WPI students regarding the use of BATS. We then analyzed various aspects of their interaction with the system. This section explains how we gathered participants for our user study, the tasks they completed, and the data we analyzed from their participation.

5.2.1: Recruitment & Scheduling

To encourage participation, we kept our evaluation to 30 minutes in length. Additionally, our team decided to give each person who completed the study an entry into a raffle for a chance to win one of four $40 Amazon gift cards. Due to the ongoing COVID-19 pandemic and accessibility to both participants and our team of evaluators, we elected to conduct the evaluations virtually. We chose to use Zoom as our evaluation platform since it offered numerous advantages such as familiarity to students, screen sharing abilities, and meeting recording options.

We sent documents containing the details of our proposed user study to the Internal Review Board (IRB) at WPI for review exemption. The IRB determined that our study was exempt from review due to the minimal risk nature of our proposed study. We then contacted potential participants via email (Appendix F).

We determined that potential participants would be undergraduate and graduate students at WPI. An initial email was sent to the email aliases of WPI Biomedical Engineering Graduate

and Undergraduate Students as well as WPI Computer Science Graduate and Undergraduate Students. The email contained an interest survey that was made using Qualtrics (Appendix G). The survey collected basic information such as their major, education level, and email address. The survey also asked potential participants about any previous experience they might have using data analysis tools. These questions helped us better understand the participants' background and gauge their knowledge in the field. Survey respondents needed to be over the age of 18 and specify that they were interested in being in our user study to participate in the evaluation. The email also contained a user-testing schedule where participants were instructed to select a 30-minute time slot to participate in the study.

For each user study, one team member acted as the facilitator and another as the notetaker. The facilitator read the script to the participant as they underwent the evaluation. The notetaker wrote down every comment made by the user and any trouble the user experienced with the system during the evaluation. The notetaker also listed any possible solutions to problems that they thought of at the moment. In addition, notes were taken on the participant's behavior, feedback, and responses to the questions asked after each scenario was completed.

We sent each interested participant a confirmation email and prompted them to virtually sign an IRB informed consent form prior to the evaluation to help them more fully understand the study. Additionally, we sent each participant an Outlook calendar event invite to confirm the date and time they had chosen to participate in our study.

On the day of the participant's scheduled user test, a team member sent them another email with a link to the Zoom meeting which would host their evaluation, a PDF document with details regarding the user study and the steps they would follow during the evaluation, and a CSV file to be used during one of the scenarios. Additionally, we reminded the participant to read over the IRB informed consent form and virtually sign if they had not already done so.

5.2.3: Procedure

Our user study simulates ways in which a user would use the application. To minimize variation across evaluations, our team developed a testing protocol script to guide participants through a subset of the features available in the application (Appendix H). Our team also developed a PDF document labeled BATS Testing Details (Appendix I) that was sent to each participant before their evaluation. This document contained the link to access BATS, the scenarios to be carried out, and important information to be used for the completion of each scenario.

The script and testing details documents separated the subset of features to be tested into three different scenarios that the user would have to complete. The scenarios were the following:

- Scenario 1: Creating an account, creating a study, and uploading the provided data.
- Scenario 2: Graphing the provided data.
- Scenario 3: Analyzing the data with a specific analysis tool and downloading the results.

Before the user test began, we asked participants to complete the tasks using the think-aloud technique. The *think-aloud* technique is a tool used by front-end designers to ensure that

everyday users know how to use the system they are developing. We wanted to see if the current design and setup of BATS made sense to the everyday user. Ideally, as the participant worked through the tasks, we wanted them to talk to themselves, telling us everything that was going through their mind relating to the task at hand. Participants were encouraged to view the BATS Testing Details pdf of the Help and Support page if they had trouble completing their tasks rather than ask the facilitator questions.

Once each task was completed, the participant was asked a series of questions regarding the features they just interacted with to help us determine their usability. At the end of each user test, we thanked the user for participating and asked for final feedback. Once all of the user tests were completed, the team contacted the four winners of the raffle and thanked every participant for their participation via email.

5.2.4: Analysis

After all the scheduled user tests were completed, the team extracted quantitative data from each user test. We looked over the recordings of each user test, and recorded the following data:

Recorded Data	Description
Time Taken to Complete Each Scenario	For each scenario we recorded the time stamps for the start and end of each scenario.
Number of Failed Tasks per Scenario	Scenario One: 3 total tasks Scenario Two: 1 total task Scenario Three: 1 total task
Number of Instructions Document References	Recorded each time the participant referenced the BATStestingdetails.pdf
Number of Help Page References	Recorded each time the participant referenced the Help & Support page.
Number of Verbal Questions	Recorded the number of times per scenario a participant clearly asked a question directly to the facilitator.
Number of Filled Silences	Filled silences are the filler words and sounds that people use, good examples are "errr", "like", "umm", or other vocalizations that do not add to the value of the sentence.
User Friendliness Score	Score given to the system by the participant. Scale from 1 to 5, 1 being least user friendly and 5 most user friendly.
Comments on the System's Flow Efficiency	Recorded the response each participant gave to the system's flow efficiency question.
Issue Frequency	During the completion of tasks, our team recorded issues or confusions participants encountered. The frequency in which each issue appeared was totaled and recorded for all participants.

Table 5: List of collected data points from the user testing recordings

For clarification on some of the data collected from the user testing, the following sections have been created, explaining how, what, and why these data points were collected.

5.2.4.1: Filled Silences

Filled silences are defined as the small vocalizations made by a person such as "like", "ummm", "errrr", and "let's see" that do not add any value or context to the sentence. Previous research has found that when cognitive load increases within a person, the number of filled silences increases as well (Defrancq & Plevoets, 2018, ; Lin et al., 2021). Hence, to understand how difficult the system was for users to navigate, we determined the average number of filled silences per minute per task. The higher the rate, the more difficult it was for users to navigate.

5.2.4.2: Common Themes

The common themes and issues that arose during the user studies were collected and analyzed for their frequency. We noted if multiple users brought up an aspect of the interface with which they had difficulty with and what it was. These issues were then condensed into a shorter list encompassing all of the issues that were brought up. The more times an issue was brought up, the higher the frequency of concern. The frequency of issues would give us a better understanding of the most common troubles users were facing with our system.

5.2.4.3: Analysis Methods

The goal of our analysis was to determine the general usability of BATS using a variety of metrics. These metrics included: time to complete the scenarios, number of filled silences, numbers of questions asked to the facilitator, user ranking of the system, and total negative score.

The participant trials were also divided into two groups, group 1 or prototype group and group 2 or final version group. Group 1 used a more rudimentary version of BATS when completing their user testing. Group 2 used the final version of BATS which included more options than the group 1 had during their user testing, such as simpler data uploading and more navigation options. The groups are compared against each other to view how the edits made to BATS affect the usability of the system. Along with testing between groups, the various collected metrics will be compared against one another and will be statically analyzed to determine the correlation between the two metrics. All of these analyses will allow the group to determine how the users worked with the system.

We used various analysis methods to view the collected data. For ease of understanding and readability, most analysis relied on graphs that would aid in the detailed understanding of the data and what it meant in terms of the usability of the BATS system.

6: Results

This section presents the data and information collected from the user study we conducted with BATS. The participants completed an interest survey and a think-aloud with intermittent semi-structured interviews.

6.1: Participant Demographics

The interest survey was used to collect background information about the participants before they began the think-aloud. In particular, we were interested in whether participants had experience analyzing data or using analysis tools. A total of 17 participants completed the interest survey and user study. Sixteen of these participants were undergraduate students and the seventeenth participant is a WPI alumni who works in industry (Figure 18.a).

Figure 18.a: Population distribution of participants' role in academia	Figure 18.b: Population distribution of participants' source of experience
Figure 18.c: Histogram of the experience participants had with data analysis tools (n=18)	Figure 18.d: Histogram of participants' experience in data analysis methods (n= 27)
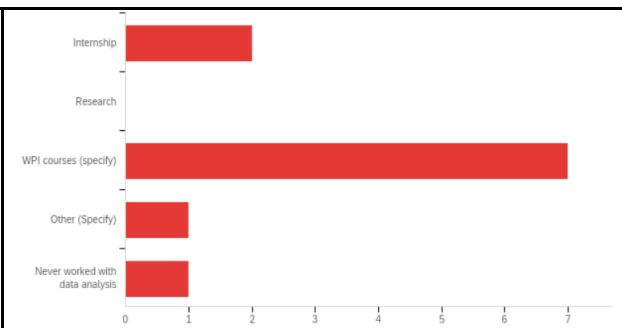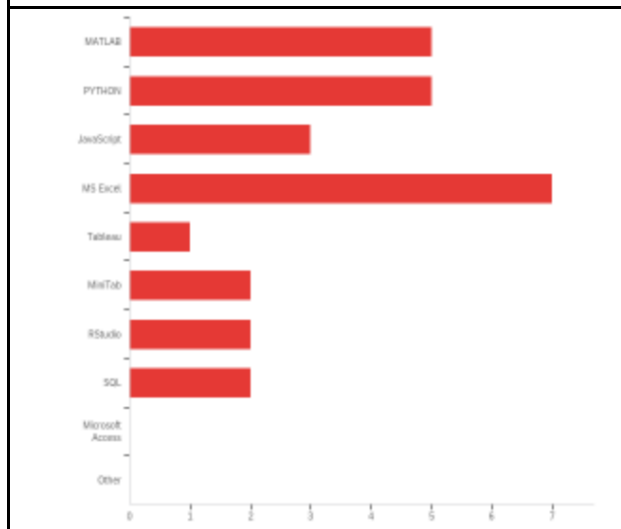
Figure 18: Histograms presenting the variability in participants' data analysis experience

Of the 17 participants, 52.94% had no experience with data analysis, 35.29% had minimal experience, and 11.76% had prior experience with data analysis. Of those with prior

experience, the majority had used MS Excel, Matlab and Python (Figure 17.c). 44.44% and 38.89% of participants with prior experience confirmed that they relied on graphical and statistical analysis methods respectively (Figure 17.d). The remaining 16.67% of participants with prior experience had used machine learning and frequency transformation methods for data analysis. 63.64% of participants gained their data analysis knowledge from offered WPI courses, 18.18% of participants from internship experience (Figure 17.b). These results show the diversity in participants' data analysis background and knowledge.

6.2: Results of the Think-Alouds and Interviews

Once the user study was completed, the team developed various metrics to analyze users' interactions with BATS. These metrics included: user ranking, time taken to complete each task, number of failures to complete a task (without direct instruction), number of questions asked, number of references to the instructions, and number of filled silences. The team then rewatched the video recordings of all participant tests and noted the values of each metric. This data was then condensed into a large table (Table 6).

Participant ID	Time Stamps (s)						Task Completed: Number of Fails			
	Scenario 1 Start	Scenario 1 End	Scenario 2 Start	Scenario 2 End	Scenario 3 Start	Scenario 3 End	Scenario 1	Scenario 2	Scenario 3	Total Fails

Table 6.a: First third of metrics noted in during the reviewing of participant think alouds

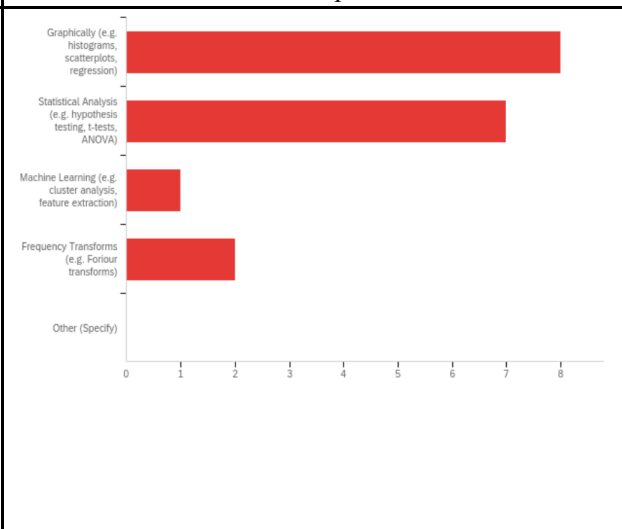| Num. of BATStestingdetails.pdf Ref.				Num. of Help page Ref.			
Scenario 1	Scenario 2	Scenario 3	Total Bats Ref.	Scenario 1	Scenario 2	Scenario 3	Total Help Page

Table 6.b: Second third of metrics noted in during the reviewing of participant think alouds

Num. of Verbal Q				Num. Filled Silences		
Scenario 1	Scenario 2	Scenario 3	Total Verbal Q	Scenario 1	Scenario 2	Scenario 3

Table 6.c: Final third of metrics noted in during the reviewing of participant think alouds

The metrics data that was collected by the team was then graphed and compared using two methods. Firstly, the various metrics were combined over all three scenarios to attain a more comprehensive view of visible trends in user experience. Secondly, the participants' results were divided into a prototype group and a final version group based on the version of BATS used during their think-aloud.

6.3: Holistic View

We observed how the participants viewed and interacted with BATS. The metrics our team used to quantify user experience of BATS were: total time, total negative score, user friendliness score, and average filled silences per minute. We used these metrics to perform statistical analyses of the group.

Figure 19.a: Population distribution of participants' time taken to complete the three scenarios	Figure 19.b: Population distribution of participants' total negative score
Figure 19.c: Population distribution of participants' user-friendliness score	Figure 19.d: Population distribution of participants' average filled silences per minute
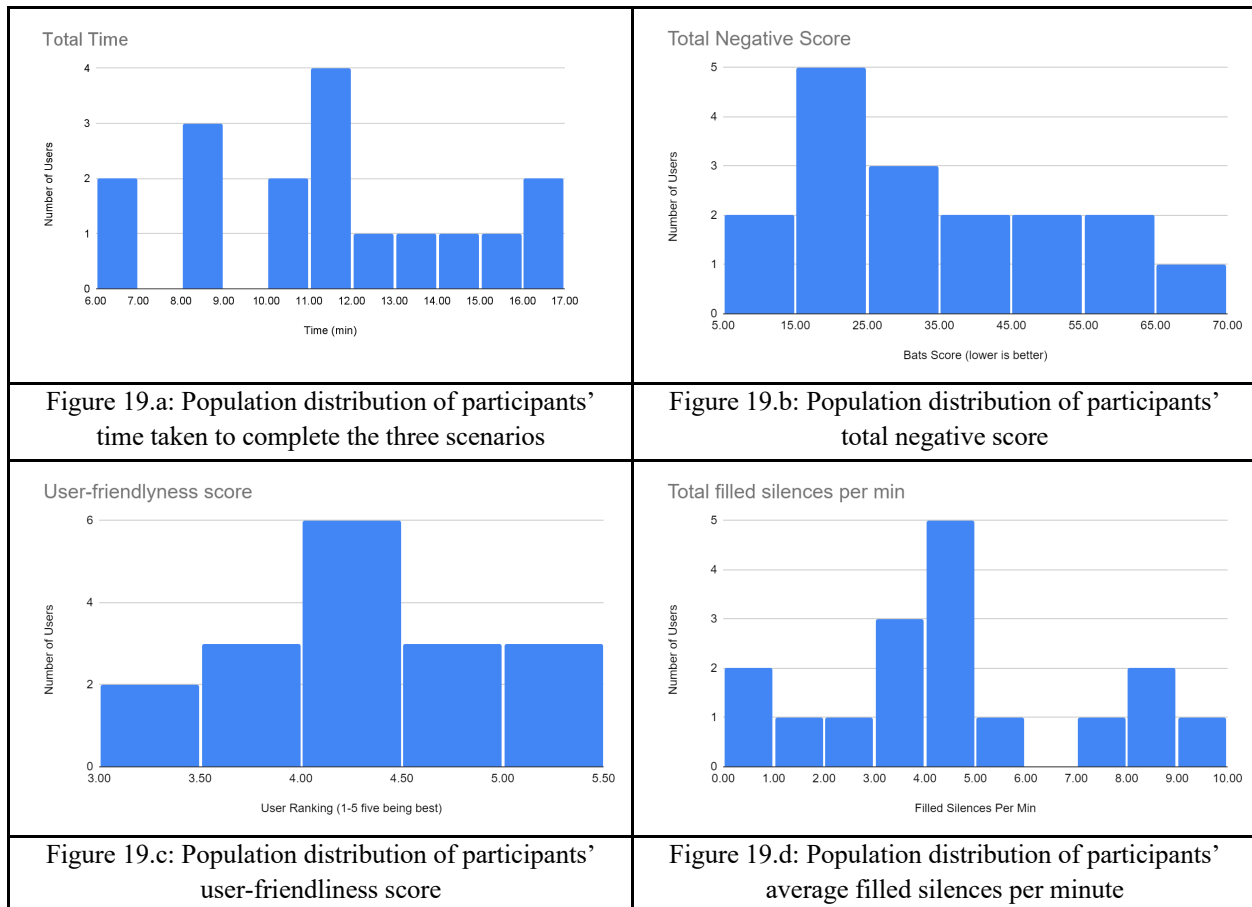
Figure 19: Population distribution of participants for each metric

Most of the participants took between 11 to 12 minutes to complete the three scenarios (Figure 18.a). The total negative score, used in Figure 18.b, was calculated using the following equation:

Negative Score = time taken + 4*number of questions asked + 10*number of tasks failed

The total negative score represents the correctness and ease of use of the system. A high total negative score indicates that the user had difficulty using the system (Figure 18.b). Figure 18.c shows the self-described, quantifiable user ranking given to the system by each participant after completing the scenarios. The users could choose a score from a range between 1 and 5, with one being not user friendly and five being very user friendly. Figure 18.d shows a cluster of users that averaged between 3 to 5 filled silences per minute. Filled silences were counted by the team as moments where the user would fill silence with various sounds such as "umm", "err", "uhh". These metrics, viewed over all scenarios, allow the team to examine user interactions with the system.

6.4: Results for the Prototype and Final Version Group

Another way we studied the users' experience was comparing participants who used the prototype and the final version of BATS. The prototype group, denoted with red, were the first

four participants to run the user study; they completed their user tests on a prototype version of BATS. The final version group, denoted with blue, completed their user tests on the final version of BATS. The final version of BATS included important navigational changes to the system. We compared the two groups using the four metrics described in the previous section. These comparisons allow us to see changes to BATS, in response to initial problems the prototype group experienced, which aided the usability of the system. The prototype and final version groups were compared to each other in various scenarios. The scenarios attempt to determine if differences in user-friendliness could be found between the two groups (Figure 19 & Figure 20).
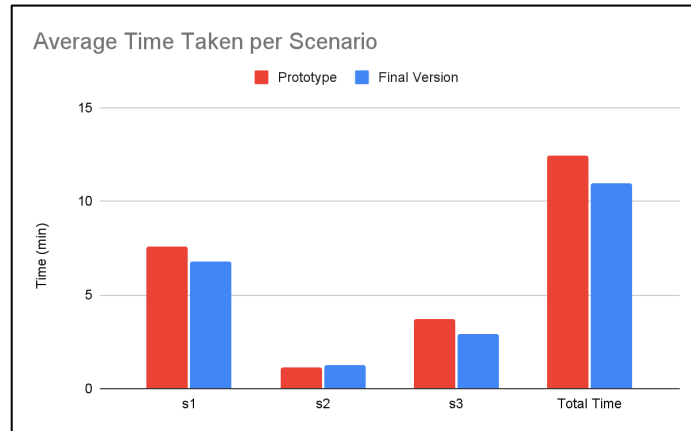
Figure 20: Histogram representing the average time taken per task for the prototype group (red) and final version group (blue)
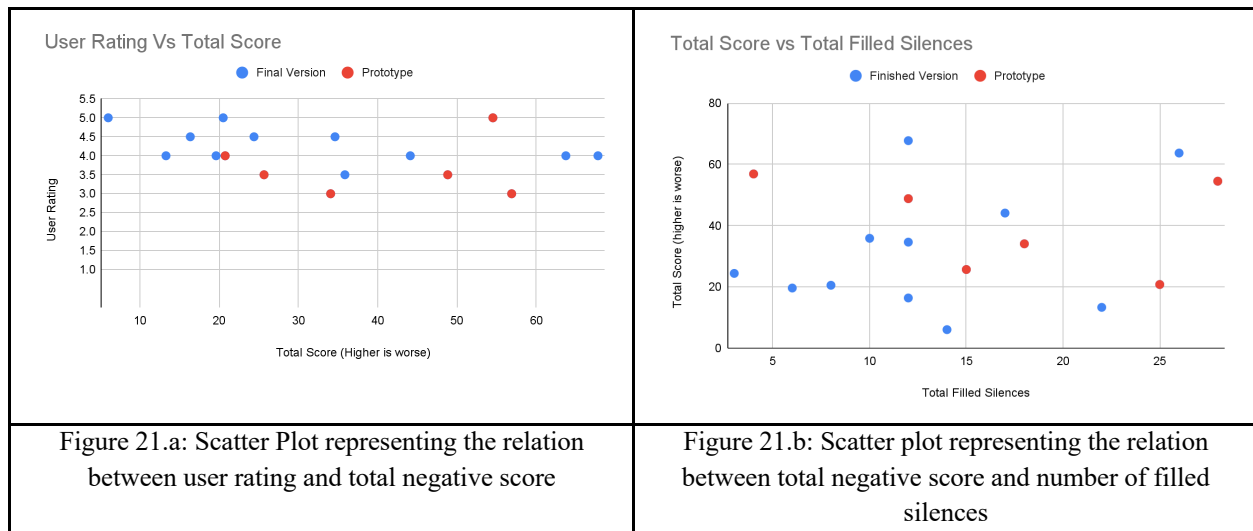
Figure 21.a: Scatter Plot representing the relation between user rating and total negative score	Figure 21.b: Scatter plot representing the relation between total negative score and number of filled silences

Figure 21.c: Scatter plot representing the relation between total negative score and filled silences per minute

Figure 21: Scatter plots showing the various comparison results between the prototype group (red) and the final version group (blue)

Figure 20.a shows evidence of correlation between user rating and total negative score No participants ranked the system below a 3 in the prototype group or a 3.5 in the final version group. The highest user rating in both groups was 5. Total negative score maximums were 56.9 and 67.7, the minimums were 20.7 and 6.01 for the prototype and the final version groups respectively. Figure 20.b shows a slight positive correlation between total negative score and total filled silences for both groups. Filled silences maximums were 28 and 26, the minimums were 4 and 3 for the prototype and the final version groups respectively. Figure 20.c is a graph with filled silences per minute to total negative score to control for time. Filled silences per minute maximums were 9.90 and 8.70, the minimums were 0.469 and 0.495 for the prototype group and the final version group respectively. Filled silences per minute show similarly strong correlation to total filled silences and have less of a chance of correlation being based on a confounding variable.

7: Discussion

This section discusses the evaluation and statistical analysis that we gathered throughout our system's development and user study. The idea behind this evaluation overview and statistical analysis is to properly understand how optimal BATS functions as an online biosignal data management and analysis tool. The features discussed in this section are those that were within the system during the completion of the user study.

As explained in the results section, a total of 17 users tested our system. During the first 6 user tests, we noticed that users were encountering similar problems and confusing tasks. We made changes to the system to fix these problems. The major changes were creating a new way for users to graph and analyze data and allowing users to upload data from the User Home page. This replaced the old way to graph and analyze data and added another way for users to upload data. In this section, it is important to note the distinction between the results of the prototype system and the improved system. We will refer to the group that tested the prototype system as the "prototype group" and the group that tested the improved system as the "final version group".

7.1: Survey Evaluation

As mentioned in our methods section, we collected responses from Biomedical Engineering and Computer Science Graduate and Undergraduate Students at WPI. At least half of the respondents reported experience in data analysis through courses at WPI and internship experience. This experience was gained through an array of well-known data management, storage, and analysis software such as MicroSoft Excel, MATLAB, and Python. The many different methods and uses of data storage, organization, and analysis indicates the need for a tool that can accomplish these functions in a single application with greater effectiveness and system fluidity.

7.2: Statistical Analysis and Discussion

We had two areas of statistical interest regarding our user study data. First, what correlations among the chosen metrics are statistically significant. Second, quantification of the differences between the prototype group and the final version group. Our results show multiple cases of reasonable significance indicating the validity of our metrics and improvement in system usability between prototype and final versions. We discuss the pertinent results here, as well as limitations of our collected data.

7.2.1: Total Negative Score vs. User Ranking

As shown in Figure 20.a in the results, there appears to be a negative correlation between the total negative score and user rating. To investigate this, we performed three negative correlation tests. The results between total negative score and user ranking were $r(15)=-0.328$ $p=0.0995$ for the whole population, $r(4)=0.113$ $p=0.5847$ for the prototype group, and $r(9)=-0.511$ $p=0.0539$ for the final version group. The strongest p-value is from the final version group. This is likely because there is less correlation between these two metrics when all of the users have uniform experience, such as with the prototype, as this would be a cluster and not a line,

even if there is a correlation. Those who had a harder time with the system ranked it as less user-friendly, resulting in a correlation between total negative score and user ranking. This suggests that our total negative score metric works as intended. The prototype group is not expected to have a relationship with total negative score, as the prototype they tested had more issues than the final version. This affected the prototype groups' experience with the system and corresponding scores.

A correlation between user score and total negative score would indicate that users were generally ranking the system worse when they failed to use the system correctly and vice versa. This would be a desirable correlation because it would indicate that users were able to criticize and identify issues within the system. Users that rank the system regardless of success or failure, may have had difficulty understanding the system and as such provide less constructive commentary on the system.

The correlation between user score and total negative score indicates that total Negative score is a successful metric of failure to use the system effectively. This confirms that a total negative score can predict users' difficulty with the system independent of their direct opinion. As such, we will use a total negative score in comparison with other metrics to help view the usability of BATS, such as comparing it to filled silences per minute.

7.2.2: Total Negative Score vs. Filled Silences per Minute

Another important correlation is the total negative score and filled silences per minute. A positive correlation would indicate that the more failures, questions, and time the user takes, the more often they produce a filled silence. Filled silences are often an indicator of increased cognitive load (Defrancq & Plevoets, 2018, ; Lin et al., 2021). They may suggest that a user is unsure about how to use a system

We performed Pearson's linear r-tests for total negative score vs. filled silences per minute on the entire study population. We removed two outliers from the population regarding filled silences per minute. These included a participant who used no filled silences and a participant that used an immense amount of filled silences per minute. These participants were outside the 95% normal curve of filled silences per minute. It is likely that their amount of filled silences per minute was due to individual speech patterns and was not representative of cognitive load due to interactions with the system.

With the removal of these outliers, the correlation between total negative score and filled silences per minute improved. Results for the whole population went from $r(15)=-0.0046$ $p=0.507$ to $r(13)=0.344$ $p=0.105$ and results from the final version group went from $r(9)=0.0160$ $p=0.481$ to $r(8)=0.509$ $p=0.0663$. Results from the prototype group did not improve much, in part due to its reduced sample size.

While not statistically significant, correlation between total negative score and filled silences per minute is strong for the given sample size. Due to the strong positive correlation within the 95th percentile normal range, we can conclude that filled silences per minute is a useful metric to view the cognitive load needed to use the system. As established in the methods that filled silences indicate cognitive load, any web system should hope to cause minimal

cognitive load when navigating between pages. Based on these measures, lower cognitive load is a desirable trait and we are pleased that this correlation reinforces this interpretation.

7.2.3: Difference Between the Prototype and Final Version Group

We ran several paired t-tests between the prototype group and the final version group. A t-test was run for each metric to view the differences in metric value between the prototype and final version group. We did not expect large differences between the metrics of total time taken, total negative score, total filled silences, and filled silences per minute. These metrics had too wide a variance within a small sample size to have a significant effect. In this we were largely correct, the lowest p-value among the metrics belonged to filled silences per minute which was an inconsequential 0.249 with $t(15)=1.2$. If we had a larger sample size, we may be more confident that these results are not due to low degrees of freedom. Despite their low p-values, each metric had a difference in averages between groups that were considerable percentages of their value (Table 7). These differences show that users in the final version group had a better experience with the system. These four metrics reflect more poorly on user experience the higher their value, and each of them decreases in the final version.

User Metric Averages	Prototype Group	Final Version Group	Difference	Percentage Difference (of Final)
Total Time Taken	12.46	10.93	1.52	13.91
Total Negative score	40.12	31.48	8.64	27.45
Total Filled Silences	17	12.90	4.09	31.69
Filled Silences per min	5.56	3.92	1.64	41.73
User Score	3.66	4.27	0.60	14.18

Table 7: Displays the Means of our metrics and comparisons between Group 1 (prototype) and Group 2 (final)

User score, which is our most controlled metric, was our only metric where a larger value designated a better user experience. Its results were far more significant than that of the other metrics with a $t(15) =-2.07$ and $p=0.057$. This is just on the far edge of 0.05 but is a promising result for the small sample size. As the metric most tied to user experience and its low variance, we hypothesized that user score would be the most significant metric.

Despite not being statistically significant, these results are still promising. With the user study, we were able to see marked improvement in all devised categories. These metrics of improvement between the prototype and final system versions indicate a strong standing for the

usability of our product. This is especially true with the average user score out of five for the final version being 4.27 compared to 3.66 for the prototype version.

8: Future Work

8.1: Interface Improvements and Expansion

This section will cover changes we would have liked to make based on these insights, as well as features that we did not have time to include. Future work on BATS could involve improvements to the existing interface and an expansion of the system and its features. This could include making BATS more aesthetically pleasing and adding the ability to create a biomarker, upload custom analysis tools, revisit previous analysis results, and share studies and data between users. These improvements and features were unable to be added to the BATS system due to our time constraints. We had to decide what features of our system were the most important for usability and focus on those essentials.

For the purposes of BATS, a biomarker consists of a pattern found in a person's biodata that can be recognized by the system as an indicator of some condition or circumstance in the body. For example, a user might run a t-test on a certain data type against a standard dataset to retrieve a confidence interval regarding the biomarker's presence.

The analysis tools currently available for BATS users are those which we have provided. These built-in analysis tools might not satisfy all a user's biosignal analysis needs. To remedy this, future work could add the ability for users to upload custom analysis tools to BATS. Custom analysis tools would be algorithms that the user would be able to run on uploaded data. These algorithms could be written in Java, Python, or MATLAB. Security checks would need to be run on the analysis tools before they are uploaded to the system to prevent transferring malicious code.

When users perform analysis on data using BATS, they are brought to the analysis results page. This page displays the selected analysis tool and data to be run and the analysis results. These results may be composed of a combination of text, CSV data, or graph(s). Currently, to re-run an analysis a user would have to re-select the desired analysis tool and data. Adding the ability to revisit previous analysis results would allow users to save the results of the analysis as well as the selected analysis tool and data. This feature would allow users to revisit saved analyses in the future, keep track of the analyses that they have run, and more easily access said analyses.

Users are currently only able to view studies that they have created and data that they have uploaded. However, researchers for whom our system is designed tend to work in teams. Adding the ability to share studies and data between users would allow BATS to be a collaborative platform. Users with whom studies or data are shared would have permissions set by the study or data creator. Permission options may include viewing, commenting, and editing.

We would also like to add more analysis tools to the system for users. We have made and tested more analysis tools than are currently available on our interface. However, we had issues incorporating these analysis tools due to the large libraries required which made our files too big for AWS lambda functions. We also hoped to develop and integrate a toolkit for facial data, stress indicator data, and log data. These toolkits proved to be beyond the scope of our project. We did not have enough time to complete all of the originally wanted toolkits (audio, neural,

facial, stress indicators, and log data), so we decided to focus on making the audio and neural toolkits as useful as possible.

8.2: Algorithms

Currently, the BATS website is able to run only a small portion of the desired biosignal analysis tools for audio and neural data. Ways to improve BATS in future implementations would be to expand on the currently available toolkits for audio and neural data analysis. The neural and audio toolkits are available to run locally and are not integrated into the web-accessible version of BATS. In the future, we would like for the audio and neural toolkits to be available in BATS.

8.2.1: Audio Toolkit

The audio toolkit is based on two open source systems, Jython and PyAudioAnalysis, that were combined to enable the system to examine various types of voice data. The back-end of BATS was made in Java. We had difficulty finding language compatible existing work in the audio analysis space, where research was mostly done in streamlined computation friendly languages such as Matlab Scripts, or Python. Therefore, we made the decision to integrate a python interpreter, using Jython, into the system to enable these existing toolkits to be integrated into the system.

PyAudioAnalysis is an efficient and powerful system that we attempted to integrate with our system using the Python interpreter. We were able to get PyAudioAnalysis running in BATS through our Python interpreter. However, we experienced problems adding PyAudioAnalysis to the final system, as the files were too large to upload with our AWS Lambda functions. In addition, getting the Python interpreter to work properly required fine tuning of the machine's Python environment. We determined that getting the Python environment to work with the BATS website was out of scope at this point in the project. Had we chosen a different web host or coded the system in python, we may have been able to resolve these issues and add the PyAudioAnalysis to the final website. Unfortunately, we had to leave the audio toolkit as a prototype feature that is only able to run locally and not from the website.

The audio toolkit can provide tone analysis, tone visualization, and conversion between CSV and WAV file types. This would allow for the ability to categorize voices of polytrauma victims. This would have been a powerful addition to BATS, given the time to work on AWS integration and the funding for the necessary storage space. The code for this system is available in our Prototype GitHub Depository (Appendix B).

8.2.2: Neural Toolkit

The neural toolkit focuses on data preprocessing and data analysis. Preprocessing neural data in BATS would remove the need for external tools. Some important preprocessing algorithms to incorporate include: converting light intensity to hemoglobin, picks pair, additional epox picks pair, and negative correlation enhancement. Additional neural analysis methods to incorporate include: the creation of general linear models, comparison between triggers and trigger modifications, data filtering and cleaning abilities, removal of noisy channels, and

comparison of activity per channel per task among other algorithms. Neural visualization tools to incorporate include adding the option to view three-dimensional models of the participant's brain and its activations over time. This would allow for the user to have a better understanding of what portions of the brain are active during various tasks. The code for this system is available in our Prototype GitHub Repository (Appendix B).

9: Conclusion

BATS was created to provide a centralized, easy-to-access application for polytrauma researchers, communities, and diagnosticians. This system offers stakeholders a user-friendly, web-based biosignal data analysis pipeline that can upload, store, access, analyze, and visualize neural, speech, and diagnostic data.

We spoke with stakeholders to determine their needs and completed a market analysis to determine currently available tools for polytrauma research. The design process to create the system involved UI models such as flowcharts and low-fidelity prototypes and pairwise comparison and pugh analysis to determine the best toolkits to modify and implement. The implementation of BATS consists of two parts, the analysis tools, and the system. The analysis tools used by BATS were either coded by the team or found in a respectable open-source toolkit and modified to work within the system. The toolkits were condensed into a jar file and integrated into the backend of the system. BATS is hosted by AWS and is accessible through the following link: https://batsmqp.s3.us-east-2.amazonaws.com/presentation/mainPage.html.

Once developed, the team completed verification and validation tests in the form of user stories and user testing. We completed a think-aloud user study using three scenarios that tested account creation, data uploading, data analysis, data viewing, and data downloading. The user tests were conducted over Zoom by members of the team. seventeen participants of the user studies would have to complete the three scenarios, via zoom, in the presence of two members of the team. The meeting was recorded and analyzed along with the study interest survey results and notes taken during each evaluation. This allowed us to assess the usability and effectiveness of BATS for a wide variety of potential users.

BATS received an average usability score of 4.27 from the user study. The usability score indicates the comprehensiveness and unavailability of the system for the average user. However, BATS has room for improvement, including the full integration of the created analysis toolkits and the addition of further features to the system.

References

10 Ways to Use Machine Learning for Medical Diagnosis | Fayrix. (n.d.). Retrieved October 2, 2021, from https://fayrix.com/blog/ten-use-cases-of-machine-learning-for-medical-diagnosis

A Closer Look at Substance Use and Suicide. (n.d.). https://doi.org/10.1176/appi.ajp-rj.2018.130603

About AWS. (n.d.). Amazon Web Services, Inc. Retrieved October 13, 2021, from https://aws.amazon.com/about-aws/

Acute Pain vs. Chronic Pain: Differences & Causes. (n.d.). Cleveland Clinic. Retrieved September 26, 2021, from https://my.clevelandclinic.org/health/articles/12051-acute-vs-chronic-pain

Agorastos, A., Pittman, J. O. E., Angkaw, A. C., Nievergelt, C. M., Hansen, C. J., Aversa, L. H., Parisi, S. A., Barkauskas, D. A., & Baker, D. G. (2014). The cumulative effect of different childhood trauma types on self-reported symptoms of adult male depression and PTSD, substance abuse and health-related quality of life in a large active-duty military cohort. *Journal of Psychiatric Research*, *58*, 46–54. https://doi.org/10.1016/j.jpsychires.2014.07.014

Al-Fahoum, A. S., & Al-Fraihat, A. A. (2014). Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains. *ISRN Neuroscience*, *2014*, e730218. https://doi.org/10.1155/2014/730218

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders* (Vol. 4).

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Ashina, H., Al-Khazali, H. M., Iljazi, A., Ashina, S., Amin, F. M., Lipton, R. B., & Schytz, H. W. (2021). Psychiatric and cognitive comorbidities of persistent post-traumatic headache attributed to mild traumatic brain injury. *The Journal of Headache and Pain*, *22*(1), 83. https://doi.org/10.1186/s10194-021-01287-7

Atlassian. (n.d.-a). *Kanban vs Scrum*. Atlassian. Retrieved September 30, 2021, from https://www.atlassian.com/agile/kanban/kanban-vs-scrum

Atlassian. (n.d.-b). *Kanban—A brief introduction*. Atlassian. Retrieved October 2, 2021, from https://www.atlassian.com/agile/kanban

Atlassian. (n.d.-c). *Scrum—What it is, how it works, and why it's awesome*. Atlassian. Retrieved September 30, 2021, from https://www.atlassian.com/agile/scrum

Atlassian. (n.d.-d). *User Stories | Examples and Template*. Atlassian. Retrieved October 2, 2021, from https://www.atlassian.com/agile/project-management/user-stories

Atlassian. (n.d.-e). *What is Agile?* Atlassian. Retrieved September 30, 2021, from https://www.atlassian.com/agile

Austerlitz, H. (2003). *Data Acquisition Techniques Using PCs*. Academic Press.

Banks, S., & Kerns, R. (1996). Explaining high rates of depression in chronic pain: A diathesis-stress framework. *Psychological Bulletin*, *119*, 95–110. https://doi.org/10.1037/0033-2909.119.1.95

Bernardo, M., Dang, J., Markoski, A., McLean, B., Penzola, S., & Whittall, F. (2021). *GILBERT: A Reimagined Biomedical Data Storage and Analysis Pipeline*. Worcester Polytechnic Institute.

berndporr/iirj: An efficient IIR filter library written in JAVA. (n.d.). Retrieved April 24, 2022, from https://github.com/berndporr/iirj

Bibliography. (n.d.). Google Docs. Retrieved April 24, 2022, from https://docs.google.com/document/u/1/d/1bGyXP3tAHjmnCYgVnssalgY6kqCwW3n1sb Qg4LGOGc4/edit?usp=embed_facebook

Blakey, S. M., Wagner, H. R., Naylor, J., Brancu, M., Lane, I., Sallee, M., Kimbrel, N. A., & Elbogen, E. B. (2018). Chronic Pain, TBI, and PTSD in Military Veterans: A Link to Suicidal Ideation and Violent Impulses? *The Journal of Pain : Official Journal of the American Pain Society*, *19*(7), 797–806. https://doi.org/10.1016/j.jpain.2018.02.012

Blow, F. C., Brockmann, L. M., & Barry, K. L. (2004). Role of Alcohol in Late-Life Suicide. *Alcoholism: Clinical and Experimental Research*, *28*(s1), 48S-56S. https://doi.org/10.1111/j.1530-0277.2004.tb03603.x

Bryant, R. A., & Harvey, A. G. (2002). Delayed-onset posttraumatic stress disorder: A prospective evaluation. *The Australian and New Zealand Journal of Psychiatry*, *36*(2), 205–209. https://doi.org/10.1046/j.1440-1614.2002.01009.x

Byllesby, B. M., Charak, R., Durham, T. A., Wang, X., & Elhai, J. D. (2016). The Underlying Role of Negative Affect in the Association between PTSD, Major Depressive Disorder, and Generalized Anxiety Disorder. *Journal of Psychopathology and Behavioral Assessment*, *38*(4), 655–665. https://doi.org/10.1007/s10862-016-9555-9

Cheng, X., Wang, X., Ouyang, T., & Feng, Z. (2020). Advances in Emotion Recognition: Link to Depressive Disorder. In *Neurological and Mental Disorders*. IntechOpen. https://doi.org/10.5772/intechopen.92019

Chicco, D., & Rovelli, C. (2019). Computational prediction of diagnosis and feature selection on mesothelioma patient health records. *PLoS ONE*, *14*(1), e0208737. https://doi.org/10.1371/journal.pone.0208737

Chronic Pain. (n.d.). Retrieved September 26, 2021, from https://www.hopkinsmedicine.org/health/conditions-and-diseases/chronic-pain

Contractor, A. A., Elhai, J. D., Fine, T. H., Tamburrino, M. B., Cohen, G., Shirley, E., Chan, P. K., Liberzon, I., Galea, S., & Calabrese, J. R. (2015). Latent profile analyses of posttraumatic stress disorder, depression and generalized anxiety disorder symptoms in trauma-exposed soldiers. *Journal of Psychiatric Research*, *68*, 19–26. https://doi.org/10.1016/j.jpsychires.2015.05.014

Defrancq, B., & Plevoets, K. (2018). Over-uh-Load, Filled Pauses in Compounds as a Signal of Cognitive Load. In M. Russo, C. Bendazzoli, & B. Defrancq (Eds.), *Making Way in*

Corpus-based Interpreting Studies (pp. 43–64). Springer. https://doi.org/10.1007/978-981-10-6199-8_3

Denaxas, S. C., Asselbergs, F. W., & Moore, J. H. (2016). The tip of the iceberg: Challenges of accessing hospital electronic health record data for biological data mining. *BioData Mining*, *9*(1), 29. https://doi.org/10.1186/s13040-016-0109-1

Depression. (2021, September 13). World Health Organisation. https://www.who.int/news-room/fact-sheets/detail/depression

Donadel, D. G., Zortea, M., Torres, I. L. S., Fregni, F., & Caumo, W. (2021). The mapping of cortical activation by near-infrared spectroscopy might be a biomarker related to the severity of fibromyalgia symptoms. *Scientific Reports*, *11*(15754). https://doi.org/10.1038/s41598-021-94456-2

Dorn, S. (2015). *Digital Health: Hope, Hype, and Amara's Law*. https://doi.org/10.1053/j.gastro.2015.07.024

Elhai, J. D., Grubaugh, A. L., Kashdan, T. B., & Frueh, B. C. (2008). Empirical examination of a proposed refinement to DSM-IV posttraumatic stress disorder symptom criteria using the National Comorbidity Survey Replication data. *The Journal of Clinical Psychiatry*, *69*(4), 597–602. https://doi.org/10.4088/jcp.v69n0411

Esang, M., & Ahmed, S. (2018). A Closer Look at Substance Use and Suicide. *American Journal of Psychiatry Residents' Journal*, *13*(6), 6–8. https://doi.org/10.1176/appi.ajp-rj.2018.130603

Faggella, D. (2020). *Machine Learning for Medical Diagnostics—4 Current Applications*. Emerj. Retrieved October 2, 2021, from https://emerj.com/ai-sector-overviews/machine-learning-medical-diagnostics-4-current-applications/

Falkenberg, L., Zeckey, C., Mommsen, P., Winkelmann, M., Zelle, B. A., Panzica, M., Pape, H.-C., Krettek, C., & Probst, C. (2017). Long-term outcome in 324 polytrauma patients: What factors are associated with posttraumatic stress disorder and depressive disorder symptoms? *European Journal of Medical Research*, *22*, 44. https://doi.org/10.1186/s40001-017-0282-9

Finley, E. P., Bollinger, M., Noël, P. H., Amuan, M. E., Copeland, L. A., Pugh, J. A., Dassori, A., Palmer, R., Bryan, C., & Pugh, M. J. V. (2015). A national cohort study of the association between the polytrauma clinical triad and suicide-related behavior among US Veterans who served in Iraq and Afghanistan. *American Journal of Public Health*, *105*(2), 380–387. https://doi.org/10.2105/AJPH.2014.301957

First, M. B. (2005). Mutually Exclusive versus Co-Occurring Diagnostic Categories: The Challenge of Diagnostic Comorbidity. *Psychopathology*, *38*(4), 206–210. https://doi.org/10.1159/000086093

Galyamina, A. G., Kovalenko, I. L., Smagin, D. A., & Kudryavtseva, N. N. (2017). Interaction of Depression and Anxiety in the Development of Mixed Anxiety/Depression Disorder. Experimental Studies of the Mechanisms of Comorbidity (review). *Neuroscience and Behavioral Physiology*, *47*(6), 699–713. https://doi.org/10.1007/s11055-017-0458-3

Griffin, J. B. (1990). Substance Abuse. In H. K. Walker, W. D. Hall, & J. W. Hurst (Eds.),
　　　Clinical Methods: The History, Physical, and Laboratory Examinations (3rd ed.).
　　　Butterworths. http://www.ncbi.nlm.nih.gov/books/NBK319/

Honigfeld, L., Macary, S. J., & Grasso, D. J. (2017). A Clinical Care Algorithmic Toolkit for
　　　Promoting Screening and Next-Level Assessment of Pediatric Depression and Anxiety in
　　　Primary Care. *Journal of Pediatric Health Care*, *31*(3), e15–e23.
　　　https://doi.org/10.1016/j.pedhc.2017.01.008

Introduction · Bootstrap v5.1. (n.d.). Retrieved April 24, 2022, from
　　　https://getbootstrap.com/docs/5.1/getting-started/introduction/

Issue, I. V. H. (n.d.). *Polytrauma/TBI System of Care Home* [Program Homepage]. Retrieved
　　　September 26, 2021, from https://www.polytrauma.va.gov/

Jackins, V., Vimal, S., Kaliappan, M., & Lee, M. Y. (2021). AI-based smart prediction of clinical
　　　disease using random forest classifier and Naive Bayes. *The Journal of Supercomputing*,
　　　77(5), 5198–5219. https://doi.org/10.1007/s11227-020-03481-x

Java Software. (n.d.). Retrieved October 13, 2021, from https://www.oracle.com/java/

Jcavanagh@Unm.Edu, J. F. C. (2021a). *Rest* [Data set]. Openneuro.
　　　https://doi.org/10.18112/OPENNEURO.DS003478.V1.1.0

Jcavanagh@Unm.Edu, J. F. C. (2021b). *Rest* [Data set]. Openneuro.
　　　https://doi.org/10.18112/OPENNEURO.DS003478.V1.1.0

JDSP - Digital Signal Processing in Java. (n.d.). Retrieved April 24, 2022, from https://jdsp.dev/

Jha, R. M., & Shutter, L. (2017). Chapter 34—Neurologic complications of polytrauma. In E. F.
　　　M. Wijdicks & A. H. Kramer (Eds.), *Handbook of Clinical Neurology* (Vol. 141, pp.
　　　633–655). Elsevier. https://doi.org/10.1016/B978-0-444-63599-0.00034-X

Joseph Volpicelli, Geetha Balarman, Julie Hahn, Heather Wallace, & Donald Bux. (1999). The
　　　Role of Uncontrollable Trauma in the Developlment of PTSD and Alcohol Addiction.
　　　Alcohol Research & Health, *23*(4), 256–262.

Kang, D., McAuley, J. H., Kassem, M. S., Gatt, J. M., & Gustin, S. M. (2019). What does the
　　　grey matter decrease in the medial prefrontal cortex reflect in people with chronic pain?
　　　European Journal of Pain, *23*(2), 203–219. https://doi.org/10.1002/ejp.1304

Keel, M., & Trentz, O. (2005). Pathophysiology of polytrauma. *Injury*, *36*(6), 691–709.
　　　https://doi.org/10.1016/j.injury.2004.12.037

Koffel, E., Krebs, E. E., Arbisi, P. A., Erbes, C. R., & Polusny, M. A. (2016a). The Unhappy
　　　Triad: Pain, Sleep Complaints, and Internalizing Symptoms. *Clinical Psychological
　　　Science*, *4*(1), 96–106. https://doi.org/10.1177/2167702615579342

Koffel, E., Krebs, E. E., Arbisi, P. A., Erbes, C. R., & Polusny, M. A. (2016b). The Unhappy
　　　Triad: Pain, Sleep Complaints, and Internalizing Symptoms. *Clinical Psychological
　　　Science : A Journal of the Association for Psychological Science*, *4*(1), 96–106.
　　　https://doi.org/10.1177/2167702615579342

Laughter, S., Khan, M., Banaag, A., Madsen, C., & Koehlmoos, T. P. (2021). Prevalence of
　　　Polytrauma Clinical Triad Among Active Duty Service Members. *Military Medicine*,

usab199. https://doi.org/10.1093/milmed/usab199

Lew, H. L., Otis, J. D., Tun, C., Kerns, R. D., Clark, M. E., & Cifu, D. X. (2009). Prevalence of chronic pain, posttraumatic stress disorder, and persistent postconcussive symptoms in OIF/OEF veterans: Polytrauma clinical triad. *Journal of Rehabilitation Research and Development*, *46*(6), 697–702. https://doi.org/10.1682/jrrd.2009.01.0006

Lin, Y., Xu, D., & Liang, J. (2021). Differentiating Interpreting Types: Connecting Complex Networks to Cognitive Complexity. *Frontiers in Psychology*, *12*, 590399. https://doi.org/10.3389/fpsyg.2021.590399

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 94–101. https://doi.org/10.1109/CVPRW.2010.5543262

Lyons, M., Kamachi, M., & Gyoba, J. (1998). *The Japanese Female Facial Expression (JAFFE) Dataset* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.3451524

Marmar, C. R., Brown, A. D., Qian, M., Laska, E., Siegel, C., Li, M., Abu-Amara, D., Tsiartas, A., Richey, C., Smith, J., Knoth, B., & Vergyri, D. (2019). Speech-based markers for posttraumatic stress disorder in US veterans. *Depression and Anxiety*, *36*(7), 607–616. https://doi.org/10.1002/da.22890

Mayo Clinic Staff. (2020, October 6). *Persistent post-concussive symptoms (Post-concussion syndrome)—Symptoms and causes*. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/post-concussion-syndrome/symptoms-causes/syc-20353352

MNE-NIRS — MNE-NIRS 0.2.1 documentation. (n.d.). Retrieved April 24, 2022, from https://mne.tools/mne-nirs/stable/index.html

Mundt, J. C., Vogel, A. P., Feltner, D. E., & Lenderking, W. R. (2012). Vocal Acoustic Biomarkers of Depression Severity and Treatment Response. *Biological Psychiatry*, *72*(7), 580–587. https://doi.org/10.1016/j.biopsych.2012.03.015

Munir, S., & Takov, V. (2021). Generalized Anxiety Disorder. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK441870/

MySQL :: About MySQL. (n.d.). Retrieved October 13, 2021, from https://www.mysql.com/about/

Naylor, B., Hesam-Shariati, N., McAuley, J. H., Boag, S., Newton-John, T., Rae, C. D., & Gustin, S. M. (2019). Reduced Glutamate in the Medial Prefrontal Cortex Is Associated With Emotional and Cognitive Dysregulation in People With Chronic Pain. *Frontiers in Neurology*, *10*, 1110. https://doi.org/10.3389/fneur.2019.01110

NIMH » Mental Illness. (n.d.). Retrieved October 11, 2021, from https://www.nimh.nih.gov/health/statistics/mental-illness

Noyes, R. (2001). Comorbidity in Generalized Anxiety Disorder. *Psychiatric Clinics of North America*, *24*(1), 41–55. https://doi.org/10.1016/S0193-953X(05)70205-7

Otte, C., Gold, S. M., Penninx, B. W., Pariante, C. M., Etkin, A., Fava, M., Mohr, D. C., &

Schatzberg, A. F. (2016a). Major depressive disorder. *Nature Reviews Disease Primers*, *2*(1), 1–20. https://doi.org/10.1038/nrdp.2016.65

Otte, C., Gold, S. M., Penninx, B. W., Pariante, C. M., Etkin, A., Fava, M., Mohr, D. C., & Schatzberg, A. F. (2016b). Major depressive disorder. *Nature Reviews Disease Primers*, *2*(1), 1–20. https://doi.org/10.1038/nrdp.2016.65

Pampouchidou, A., Simos, P. G., Marias, K., Meriaudeau, F., Yang, F., Pediaditis, M., & Tsiknakis, M. (2019). Automatic Assessment of Depression Based on Visual Cues: A Systematic Review. *IEEE Transactions on Affective Computing*, *10*(4), 445–470. https://doi.org/10.1109/TAFFC.2017.2724035

Papers with Code—JAFFE Dataset. (n.d.). Retrieved October 12, 2021, from https://paperswithcode.com/dataset/jaffe

Pediaditis, M., Giannakakis, G., Chiarugi, F., Manousos, D., Pampouchidou, A., Christinaki, E., Iatraki, G., Kazantzaki, E., Simos, P. G., Marias, K., & Tsiknakis, M. (2015). Extraction of facial features as indicators of stress and anxiety. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, *2015*, 3711–3714. https://doi.org/10.1109/EMBC.2015.7319199

Peixoto, C., Hyland, L., Buchanan, D. M., Langille, E., & Nahas, R. (2018a). The polytrauma clinical triad in patients with chronic pain after motor vehicle collision. *Journal of Pain Research*, *11*, 1927–1936. https://doi.org/10.2147/JPR.S165077

Peixoto, C., Hyland, L., Buchanan, D. M., Langille, E., & Nahas, R. (2018b). The polytrauma clinical triad in patients with chronic pain after motor vehicle collision. *Journal of Pain Research*, *11*, 1927–1936. https://doi.org/10.2147/JPR.S165077

Peixoto, C., Thompson-Hyland, L., Buchanan, D., Langille, E., & Nahas, R. (2018). The polytrauma clinical triad in patients with chronic pain after motor vehicle collision. *Journal of Pain Research*, *Volume 11*, 1927–1936. https://doi.org/10.2147/JPR.S165077

Permenter, C. M., Fernández-de Thomas, R. J., & Sherman, A. l. (2022). Postconcussive Syndrome. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK534786/

Price, M., Legrand, A. C., Brier, Z. M. F., & Hébert-Dufresne, L. (2019a). The symptoms at the center: Examining the comorbidity of posttraumatic stress disorder, generalized anxiety disorder, and depression with network analysis. *Journal of Psychiatric Research*, *109*, 52–58. https://doi.org/10.1016/j.jpsychires.2018.11.016

Price, M., Legrand, A. C., Brier, Z. M. F., & Hébert-Dufresne, L. (2019b). The symptoms at the center: Examining the comorbidity of posttraumatic stress disorder, generalized anxiety disorder, and depression with network analysis. *Journal of Psychiatric Research*, *109*, 52–58. https://doi.org/10.1016/j.jpsychires.2018.11.016

Price, M., & van Stolk-Cooke, K. (2015). Examination of the interrelations between the factors of PTSD, major depression, and generalized anxiety disorder in a heterogeneous trauma-exposed sample using DSM 5 criteria. *Journal of Affective Disorders*, *186*, 149–155.

https://doi.org/10.1016/j.jad.2015.06.012

Principles behind the Agile Manifesto. (n.d.). Retrieved September 30, 2021, from
https://agilemanifesto.org/principles.html

PTSD and comorbidity: 80% of PTSD patients have a comorbid disorder. (n.d.). Retrieved
September 26, 2021, from https://www.pharmaceutical-technology.com/comment/ptsd-
and-comorbidity/

Pugh, M. J. V., Finley, E. P., Copeland, L. A., Wang, C.-P., Noel, P. H., Amuan, M. E., Parsons,
H. M., Wells, M., Elizondo, B., & Pugh, J. A. (2014). Complex Comorbidity Clusters in
OEF/OIF Veterans: The Polytrauma Clinical Triad and Beyond. *Medical Care*, *52*(2),
172–181.

Reardon, A. F., Brief, D. J., Miller, M. W., & Keane, T. M. (2014). Assessment of PTSD and its
comorbidities in adults. In *Handbook of PTSD: Science and practice, 2nd ed* (pp. 369–
390). The Guilford Press.

Resources – Jeffrey Cohn. (n.d.). Retrieved October 12, 2021, from
https://www.jeffcohn.net/resources/

Sareen, J. (2014). Posttraumatic Stress Disorder in Adults: Impact, Comorbidity, Risk Factors,
and Treatment. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*,
59(9), 460. https://doi.org/10.1177/070674371405900902

Sheerin, C. M., Franke, L. M., Aggen, S. H., Amstadter, A. B., & Walker, W. C. (2018).
Evaluating the Contribution of EEG Power Profiles to Characterize and Discriminate
Posttraumatic Stress Symptom Factors in a Combat-Exposed Population. *Clinical EEG
and Neuroscience*, *49*(6), 379–387. https://doi.org/10.1177/1550059418767583

Smith, N. D. L., & Cottler, L. B. (2018). The Epidemiology of Post-Traumatic Stress Disorder
and Alcohol Use Disorder. *Alcohol Research : Current Reviews*, *39*(2), 113–120.

Soo Hyun Rhee, John K Hewitt, Michael C. Stallings, Robin P Corley, & Michael C. Neale.
(2004). The Validity of the Neale and Kendler Model-Fitting Approach in Examining the
Etiology of Comorbidity | SpringerLink. *Behavior Genetics*, *34*, 251–256.

*Speech-based markers for posttraumatic stress disorder in US veterans—Marmar—2019—
Depression and Anxiety—Wiley Online Library*. (n.d.). Retrieved October 4, 2021, from
https://onlinelibrary.wiley.com/doi/full/10.1002/da.22890

Sponheim, S. R., McGuire, K. A., Kang, S. S., Davenport, N. D., Aviyente, S., Bernat, E. M., &
Lim, K. O. (2011). Evidence of disrupted functional connectivity in the brain after
combat-related blast injury. *NeuroImage*, *54 Suppl 1*, S21-29.
https://doi.org/10.1016/j.neuroimage.2010.09.007

Tania, F. A., & Chandra Shill, P. (2019a). An Improved Support Vector Machine for Disease
Diagnosis. *2019 22nd International Conference on Computer and Information
Technology (ICCIT)*, 1–6. https://doi.org/10.1109/ICCIT48885.2019.9038369

Tania, F. A., & Chandra Shill, P. (2019b). An Improved Support Vector Machine for Disease
Diagnosis. *2019 22nd International Conference on Computer and Information
Technology (ICCIT)*, 1–6. https://doi.org/10.1109/ICCIT48885.2019.9038369

Torres, F. (n.d.). *What Is PTSD?* American Psychiatric Association. Retrieved September 26, 2021, from https://www.psychiatry.org/patients-families/ptsd/what-is-ptsd

Treede, R.-D., Rief, W., Barke, A., Aziz, Q., Bennett, M. I., Benoliel, R., Cohen, M., Evers, S., Finnerup, N. B., First, M. B., Giamberardino, M. A., Kaasa, S., Kosek, E., Lavand'homme, P., Nicholas, M., Perrot, S., Scholz, J., Schug, S., Smith, B. H., … Wang, S.-J. (2015). A classification of chronic pain for ICD-11. *Pain*, *156*(6), 1003–1007. https://doi.org/10.1097/j.pain.0000000000000160

Tyiannak/pyAudioAnalysis: Python Audio Analysis Library: Feature Extraction, Classification, Segmentation and Applications. (n.d.). Retrieved April 24, 2022, from https://github.com/tyiannak/pyAudioAnalysis

UML Class Diagram Tutorial. (n.d.). Retrieved October 12, 2021, from https://www.visual-paradigm.com/guide/uml-unified-modeling-language/uml-class-diagram-tutorial/

Using Functional Near-Infrared Spectroscopy (fNIRS) to Study Chronic Pain – TouchTomorrow. (n.d.). Retrieved October 4, 2021, from https://wp.wpi.edu/touchtomorrow2021/activity/using-functional-near-infrared-spectroscopy-fnirs-to-study-chronic-pain/

Vachon-Presseau, E., Roy, M., Woo, C.-W., Kunz, M., Martel, M.-O., Sullivan, M. J., Jackson, P. L., Wager, T. D., & Rainville, P. (2016). Multiple faces of pain: Effects of chronic pain on the brain regulation of facial expression. *Pain*, *157*(8), 1819–1830. https://doi.org/10.1097/j.pain.0000000000000587

What is Polytrauma? - Polytrauma/TBI System of Care. (n.d.). [General Information]. U.S. Department of Veterans Affairs. Retrieved September 26, 2021, from https://www.polytrauma.va.gov/understanding-tbi/definition-and-background.asp

Yiu, T. (2021, September 29). *Understanding Random Forest*. Medium. https://towardsdatascience.com/understanding-random-forest-58381e0602d2

Yock, P. G., Stefanos Zenios, Josh Makower, Todd J. Brinton, Uday N. Kumar, F. T. Jay Watkins, & Thomas M. Krummel (Eds.). (2015a). *Biodesign: The process of innovating medical technologies* (Second edition). Cambridge University Press.

Appendices

Appendix A: Diagnostic Background of Polytrauma using the DSM 5

The modern understanding of Psychological polytrauma represents all of these new ways of thinking about health very well. Every case of it is different, with different causes, relationships, and chronological orders between the different conditions. But in all these cases there are interactions and/or causations between these diseases that need to be accounted for in care and ways these subjects and patients can be left behind if not correctly understood.

The diagnostic criteria of these conditions is laid out in the table below:

PostTraumatic Stress Disorder	1: Exposure to actual or threatened death, serious injury, or sexual assault. 2: Intrusive or compulsive behaviors developing after/about the trauma. 3: Avoidance of stimuli associated with traumatic event. 4: Negative changes in thought or mood caused by the event. 5: More easily brought to anger or otherwise hypervigilant. 6: Any of these criteria are more significant if they go on for more than a month, cause a breaking down of existing relationships, or occur when sober.
Chronic Pain	Pain that continues for more than 3 months. Subtypes include primary (general type when no others fit), postsurgical, cancer, neuropathic (nervous system issue), headache, visceral (internal trauma), and musculoskeletal.
Major Depressive Disorder	1: Depressed mood most of the day. 2: Diminished interest in activities. 3: Significant weight loss or gain. 4: Insomnia or Hypersomnia 5: Psychomotor agitation or retardation most days. 6: Fatigue. 7: Strong feelings of worthlessness or unstructured guilt. 8: Diminished ability to concentrate and make decisions. 9: Thoughts of death or suicidal ideation.
Generalized Anxiety Disorder	1: Excessive anxiety and worry. 2: This worry is difficult or impossible to ignore or control. 3: This worry causes disturbances such as restlessness, fatigue, concentration issues, irritability, or muscle issues. 4: The severity of these individual issues on their social life.

Table A.1: Diagnostic Criteria (paraphrased from DSM and Chronic pain from NCBI (*Acute Pain vs. Chronic Pain*, n.d.))

These sorts of criteria are what define the treatment of these conditions, and the combination of them helps give a better view of what polytrauma is. It's not just a series of conditions, but a list of symptoms. Often it is not clear which symptoms cause what others, and often a symptom can be indicative of two of these conditions, but if a patient has enough of these symptoms seriously enough, then they fall under the polytrauma umbrella.

Appendix B: Toolkits not Present in Final Web App

There were some parts of the final algorithm development that were not able to be made part of the final web app because of file storage and time constraints. These systems worked based on a python interpretation system that was bulky but allowed flexibility and more advanced toolkit use. This appendix outlines how to access these systems and test them.

The following is the link to the github repository:

https://github.com/Nathan-Draudt/Prototype-Toolboxes

At this link there is the full codebase necessary to run the functions of these toolkits, as well as instructions on how it should be done, and demos one can run of useful features. This git repository contains overlap with our other repository, but we thought that a clean repository with only the information on these toolkits would help future work go smoothly, and better display the function of these subsystems.

Appendix C: Main Github Information

The following is the link to our main github repository for production of the BATS system: https://github.com/BATSMQP/PipelineMQP

The main production branches were "main" and "only_ttest". The "main" branch contains all of our web infrastructure, and "only_ttest" contains the section of code used with maven to make jar files of algorithms to feed AWS. The other branches were largely to preserve data when data was being removed from other branches, in order to preserve versions we wanted to reference, or mine for code at a later date.

Appendix D: Definitions

Polytrauma*: Appearance of a multitude of mental and physical comorbidities that arise due to a singular or multiple traumatic events*

Comorbidities*: Physical or mental illness or condition present with other illnesses or conditions*

PCS*: Post-concussion Syndrome*

PCT*: Polytraumatic Clinical Triad*

PTSD*: Post Traumatic Stress Disorder*

MDD*: Major Depressive Disorder*

GAD*: Generalized Anxiety Disorder*

DSM*: Diagnostic and Statistical Manual of Mental Disorders*

TBI*: Traumatic Brain Injury*

EEG*: Electroencephalogram*

fNIRS*: functional Near-Infrared Spectroscopy*

Mean*: The average value of a group of data as determined by the sum of all the data points divided by the number of them.*

Median*: The center data point of an odd amount of a data group or the average of the two center data points in an even amount.*

Mode*: The most commonly occurring data points in a group. There are also more advanced mode features such as mode areas (which three consecutive shoe sizes contain the most purchases) or mode rankings (what are the top 5 best selling shoe sizes).*

Standard Deviation*: A descriptor of how far apart the data points are from each other, helpful in comparing data sets in the same medium.*

Statistical Significance*: This would allow the testing of P-value and other essential metrics of significance based on an expected outcome for a group or a control group vs. an experimental group. It will recommend 0.05 as a baseline but will allow for other standards.*

Appendix E: User Study Consent Form

Informed Consent Agreement for Participation in a Research Study
User Testing

Investigators: Nathan Draudt(ndraudt@wpi.edu),

Claire Nicolas (canicolas@wpi.edu),

Carlos Velasquez (cavelasquez@wpi.edu),

Lauren Wasserman (lewasserman@wpi.edu)

Primary Contact Information: gr-pipelinemqp21@wpi.edu

Project Title: Development of a Data Management & Analysis Pipeline for Biomedical Signals (BATS)

Primary Contact Information: gr-pipelinemqp21@wpi.edu

Advisors: Adam Lammert (alammert@gmail.com), Mark Claypool (claypool@wpi.edu)

Introductions

Hello, we would like to thank you for meeting with us today. We would like to introduce our team members, which you can address on a first-name basis. Our team comprises Nathan Draudt, Claire Nicolas, Carlos Velasquez, and Lauren Wasserman, and we are all college seniors completing our Master Qualifying Project (MQP) at Worcester Polytechnic Institute. Today, you will be helping us improve our MQP project system.

Introduction to the format of this meeting:

You are being asked to participate in a user study. Before you agree, we wish to inform you about the project's purpose, the procedures, and the protocols we follow. Our goal here is to ensure that you are making a fully informed decision regarding your participation and that you feel comfortable participating.

Purpose of our project:

The purpose of this project is to create a pipeline, named Biosignal Analysis of Trauma System (BATS), for biomedical data to be used by researchers for the study and diagnosis of polytrauma. Polytrauma is the combination of various physical and mental traumas in a person. We will be focusing on the intersection of PTSD, depression, anxiety, post-concussive syndrome, and chronic pain. Using these interviews, our team will work towards creating an easy-to-use research tool focused on analyzing speech and neural data.

Purpose of the study:

The purpose of this study is to understand how users would use BATS. This will allow us to see how we can make the system more user-friendly and potential hiccups we were not able to foresee. Ultimately, after seeing how you interact with the system, we will improve the system to fit the needs you described.

Risks to participants:

There are no anticipated, perceived, or known risks to you that we are aware of due to your participation in this think-aloud and semi-structured interview.

The task will be completed virtually, via zoom, to avoid any potential risk of exposing one another to COVID-19.

The test subjects will need to download a CSV file provided by the team. However, the team will test the link to the file that the subject needs to download to ensure that the subject does not encounter any risk when downloading the file.

Procedures to be followed:

Subjects will be recruited for this study via email. If interested to participate, Subjects will fill out and interest form and select a 30-minute time slot to. Researchers will then send the subject a confirmation email with the date and time selected for the completion of the study. The email will include the zoom link for virtual study and a CSV file to be used during the study.

During the 30 minute period, you will be asked to share your screen and turn on your camera. The interaction will be recorded. You will then be asked to complete three scenarios in a *think-aloud** manner where you will be interacting with the BATS system. The three scenarios will involve:

- creating an account, creating a study, uploading the given data,
- analyzing the data with the given tools
- downloading the analyzed dataset.

Between the scenarios, and at the end of the final scenario you will be asked a series of questions regarding your interaction with the system, you may ask any questions as well. After the final interview, you will be thanked, your anonymous code will be noted, data downloaded and you will be entered in the raffle to win one of the four $40 Amazon gift cards.

***Think Aloud:** A think-aloud is a tool used by front-end designers to ensure that everyday users know how to use the system they are developing. Here you will be asked to vocalize every thought you have*

pertaining to the scenario you are working through, this is a way to understand your thought process as you interact with the system.

Benefits to research participants and others: Benefits of this study for the researcher would be to have feedback from potential users on the built system, and use this to improve its usability. Benefits to the participant involve learning more about polytrauma and the analysis of speech and neural datasets.

Record keeping and confidentiality: We would like to receive your verbal permission to record this task for our records and to publish your responses. <u>Anonymity is always guaranteed.</u> Your responses will be given a unique anonymous code that will be stored with other participants and their anonymous coded data. We will not be sharing any recordings with members outside of the research group. If your data will be used, a transcript of the interaction will be placed in the appendix of the MQP report; instead of your name, we will use your given anonymous code. If we wish to use a quote that identifies you, we will ask for your permission, and you have the right to review any quotes or information before publication.

The answers to your questions today will be maintained in the investigators' passcode-protected computer only to review this team. No one else will have access to the answers. All unpublished data will be deleted upon completion of the project. "Records of your participation in this study will be held confidential so far as permitted by law. However, the study investigators, the sponsor or its designee, and, under certain circumstances, the Worcester Polytechnic Institute Institutional Review Board (WPI IRB) will be able to inspect and have access to confidential data that identify you by name. Any publication or presentation of the data will not identify you."

Compensation or treatment in the event of injury:

This research does not involve any risk of injury or harm. If any injury or harm does occur while participating in the study (ex: falling out of chair). Please assess your injuries and dial 911 as necessary. The task will be stopped early if you injure yourself. You do not give up any of your legal rights by signing this statement.

Compensation for participants:

All participants will be placed in a raffle to win one of four $40 amazon gift cards. The winners of the cards will be selected using a random number generator where each anonymous code will be associated with a number.

For more information about this research or about the rights of research participants, or in case of research-related injury, contact: Investigators contact information is seen above,

contact information for the IRB Manager (Ruth McKeogh, Tel. 508 831- 6699, Email: irb@wpi.edu), and the Human Protection Administrator (Gabriel Johnson, Tel. 508-831-4989, Email: gjohnson@wpi.edu).

Your participation in this research is voluntary. If you feel uncomfortable at any point and do not wish to continue the study, you have every right to do so. This will not result in any penalty to you or any loss of benefits to which you may otherwise be entitled. If you wish to reschedule your participation to another time, please alert the project investigators as soon as possible. The project investigators retain the right to cancel or postpone the task at any time they see fit.

Agreement: "I agree that I am 18 years of age or older, I have a computer with an internet browser, and I agree to participate in the studies described above."

Date: _____

Name of Participant:

_____

Signature of Participant:

_____

Date: _____

Signature of Person who explained this study:

_____

Appendix F: User Study Interest Email

<u>Sent to</u>: dl-be-undergrads@wpi.edu, dl-be-grads@wpi.edu, dl-cs-grads@wpi.edu, dl-cs-majors@wpi.edu

<u>Sent From:</u> cavelasquez@wpi.edu

<u>Subject:</u> WIN A 40$ AMAZON GIFTCARD- MQP User Testing

<u>Content:</u>

Hello,

We are an MQP team of CS and BME students. We designed and developed a pipeline for biomedical data to be used by researchers for the study and diagnosis of polytrauma.

We would really appreciate your participation in our 30-minute virtual user study (over Zoom) to evaluate the usability and functionality of our web-based application. <u>Every participant will have the opportunity to win one of four $40 Amazon Gift Cards.</u>

If interested, please fill out <u>this survey</u> (it should take two minutes or less) and select a time slot <u>here</u> that suits your schedule. Once filled in, you will be contacted by our team!

Thank you for your time, and if you have any further questions, please reach out to us at: gr-pipelinemqp21@wpi.edu

Best,
PipelineMQP

BATS MQP - User Testing Interest Form

Q1
BATS MQPHello! We are an MQP team of CS and BME students at WPI. We have designed and developed a pipeline for biomedical data to be used by researchers for the study and diagnosis of polytrauma. We are looking for participants to evaluate the usability and functionality of our application. This survey should take two minutes or less. Thank you for your time and we can be contacted with any further questions at gr-pipelinemqp21@wpi.edu!

Q2 Are you 18 years or older? (Yes/No)

◯ Yes (1)

◯ No (2)

Display This Question:

If Are you 18 years or older? (Yes/No) = No

Error Unfortunately, we cannot take participants under the age of 18. Thank you for your time! Please click "Next" to finish the survey.

Skip To: End of Survey If Unfortunately, we cannot take participants under the age of 18. Thank you for your time! Please c... Is Displayed

Display This Question:

If Are you interested in participating in a 30-min virtual user study of our data analysis pipeline? = No

Q13 Thank you for your time! Please click "Next" to finish the survey.

Display This Question:

Q7 Are you interested in participating in a 30-min virtual user study of our data analysis pipeline?

○ Yes (1)

○ No (2)

Q8 First Name

_____

Q9 Last Name

_____

Q10 WPI Email Address

_____

Q11 Major(s)

_____

Q12 Please select which of the following most closely applies to you:

◯ Undergraduate Student (1)

◯ Graduate Student (2)

◯ Faculty/Staff (3)

◯ Other (4) _____

Q14 Do you have experience in data analysis?

◯ Yes (1)

◯ No (2)

◯ Some (3)

Q11 In what environment have gained your data analysis experience? (check all that apply)

☐ Internship (1)

☐ Research (2)

☐ WPI courses (specify) (3)

_____

☐ Other (Specify) (4) _____

☐ Never worked with data analysis (5)

Q12 How did you analyze the data? (check all that apply)

☐ Graphically (e.g. histograms, scatterplots, regression) (1)

☐ Statistical Analysis (e.g. hypothesis testing, t-tests, ANOVA) (2)

☐ Machine Learning (e.g. cluster analysis, feature extraction) (3)

☐ Frequency Transforms (e.g. Foriour transforms) (4)

☐ Other (Specify) (5) _____

☐ None of the above (6)

Q13 Which of the following tools do you have experience with? (check all that apply)

☐ MATLAB (1)

☐ PYTHON (2)

☐ JavaScript (3)

☐ MS Excel (4)

☐ Tableau (5)

☐ MiniTab (6)

☐ RStudio (7)

☐   SQL (8)

☐   Microsoft Access (9)

☐   Other (10) _____

☐   None of the above (11)

Q17 Please follow this link to select a time slot to schedule your participation in our User Testing

Q15 Thank you for your time! Please click "Next" to finish the survey.

End of Block: BATS MQP - User Testing

Appendix H: User Study Script

<u>BATS Protocol</u>

As a research associate involved with this experiment, you are responsible for actively monitoring the subject, ensuring that all instructions and tasks detailed below are described clearly to the subject and that any questions are answered to the best of your ability. If at any time, the subject has reservations about participating or continuing to participate in this experiment, you should encourage them to withdraw. DO NOT ENCOURAGE PARTICIPATION! If a question arises that you are unable to answer, seek assistance from a senior member of the research staff. DO NOT GUESS. During the course of the study, you should minimize any unnecessary actions in the room, e.g. moving around the room, getting papers organized, extra keystrokes, etc. while subjects are completing tasks. Low-level noise and slight movements will impact subject focus. In addition, full attention should be devoted to active monitoring of the subject, website, and protocol.

This user study will take about 30 minutes, with some time for sharing feedback at the end. Users will only be using a subset of the features available in this application. You will be leading this user study and another member of the research staff should be taking notes. All data that we record from this should be kept anonymous.

During this evaluation, it would be helpful for the research team if you could urge the test subject to think out loud as the tasks are completed to understand the thought process. Please note, there is a help page attached to the website. Please refer the test subject to it if any confusion arises regarding the terminology and how to proceed forward.

EMERGENCY (Medical or Police Support): Dial 911

Subject Requirements
      Age 18+
      Comfortable speaking and reading English
      Has access to a computer with a camera, speaker, internet browser, and microphone

Protocol Summary
The protocol will be carried out by a single or team of Researchers over zoom.
1. When Subject Arrives - Study Overview, Consent Form
2. Introduce the think-aloud
3. Conduct the think-aloud and semi-structured interviews
4. Thank the user for their participation

Part 1. Subject's Arrival
Forms: Consent form (Should be sent to the participant)
      Think-aloud details document (Should be sent to the participant)

      Before subject arrival, ensure the subject has been sent all necessary documents and the link to the zoom where the research will be taking place. Join the meeting early, with camera and microphone on. When the subject arrives, say "*Hello, my name is _____ , and this is [facilitator's name]. We are the researcher guiding you through today's study. How are you? For today's study, I will be explaining the experiment by reading a script so that I do not miss any points and to guarantee that all participants have the same experiment. Do you have a cell phone with you today? Could you please silence it for the duration of the experiment?*"

Part 2. Introducing the experiment
      "*Today's study examines the usability of an online biomedical data signal analysis tool called Biomedical Analysis Toolkit System (BATS). If you choose to continue participating, You will only be using a subset of the features available in this application. I will be leading this user study and [facilitator's name] will be taking notes. All data that we record from this will be kept anonymous.*

      During this evaluation, it would be helpful for us if you could think out loud as you go about completing these tasks so we can understand your thought process but please be clear if you are actually asking us questions directly. Please note, there is a help page attached to the website. Please refer to it if you have any confusion on the terminology and how to proceed forward. Do you have any questions before we start?"

Do you know what a think-aloud is? If they do not know then stay the following
      [NO] *A think-aloud is a tool used by front-end designers to ensure that everyday users know how to use the system they are developing. In our case, we want to see if the current design and setup of BATS makes sense to the everyday user.*
      [YES] *proceed to next section*

If you choose to proceed, *we will give you three tasks to complete, you will be given a 30 minute time limit. Then, as you work your way through the task, we want you to talk to yourself, telling us everything that is going through your mind relating to the task at hand. For clarification:*

- *We are not interested in your secret thoughts, but we are interested in what you are thinking about regarding your tasks.*
- *You are not being tested, the system is, so if you encounter any trouble, that is the system's fault and not yours.*

Do you now understand what a think-aloud is:?

[YES] proceed to the next section

[NO] We will now send you a video of a person partaking in a think-aloud. Please watch this and ask us any questions you have: (LINK).

https://drive.google.com/file/d/1zxvAiryz7Do6YG7p2y7bsPU2rtvwmCjX/view?usp=sharing

As practice, count and describe the windows in your room using the think-aloud technique

Today we are collecting audio and visual data, therefore we will be recording this zoom meeting. If at any time you are uncomfortable, please let us know and we will stop the recording or end the task. Your data will be recorded anonymously with other participants with anonymous codes."

Provide the subject with a copy of the informed consent form. "*We sent you the in the consent form before the study. Have you had the chance to read it through?*

[NO]. Please read through it thoroughly and ask any questions you may have. [YES] proceed to next section

If you choose to continue, please sign the consent form through the qualtrics link you have been sent and which is in the BatsTestingDetails.pdf document"

Give the subject time to review the informed consent form. If they choose to sign the form, thank them for participating in the study. If they choose to no longer participate, thank them for their time and guide them out of the lab.

Part 3. Running the Think Aloud & Semi-Structured Interviews

After the participant has reviewed and signed the consent form, say *"Please open the document we sent you labeled BATSTestingDetails.pdf, please read it carefully." "Once you have read the document please open the BATS website. This is the system you will be testing. Please share your screen at this time and I will now begin recording the meeting."* Once they have shared their screen, and BATS is open, begin recording document *"You may now begin with scenario 1 as described in the BATSTestingDetails.pdf"* Turn off your camera, if the user has any questions during the scenario or takes them longer than 1-2 minutes, turn on your camera and ask them guiding questions. While they complete the task, take notes on their mouse movement, how often they make mistakes, or important details in how they interact with the system. Once they have completed the scenario.

Scenario 1: Create account, upload data, create a study
Scenario 2: graph data

Scenario 4: Analyze uploaded data, download analyzed data
"We will now commence an interview, please answer the following questions truthfully.
- How many times would you say that you were confused about how to accomplish a task while completing the scenarios?
    - (If the user does not fully elaborate:) Can you please elaborate more about why you were confused in each of those scenarios?
- How efficient did you find the flow of the interface to be?
- What aspects of the interface were the most user-friendly?
- What aspects of the interface were the least user-friendly?
SCENARIO 2
- How easy was the graph to read?
    - How clear was the labeling?

Scenario 3
- Did you have any difficulty running the algorithm?
    - How understandable was the algorithm selection menu?
    - Were there any aspects that confused you?
    - Did you understand why and where you added the requested inputs for the algorithm?
- Did you ever feel overwhelmed during your time using BATS?
    - Any of which due to the number of choices offered?
- How user-friendly did you find the interface on a scale of 1 to 5? 1 being the lowest and five being the highest."

Once you have completed Scenario 1, repeat Phase 2 for Scenarios 2 and 3.

Part 4. End of Experiment

At the end of the experiment, say "*Thank you for your time today, do you have any questions about the tasks you completed?*" Wait for a response and answer the questions to the best of your ability.

Appendix I: BATS Testing Details

Details Regarding the BATS User Studies

Do NOT copy and/or paste any information from this document into the BATS system during the user-study. This document is only for user reference.

Consent Form:

If you have not yet signed the consent form, please do so at this link:

https://wpi.qualtrics.com/jfe/form/SV_78svLlXlF7i4xfw

Who are you:

In this think-aloud, you are a part of a research team conducting a study. Here you are analyzing and modeling the speech patterns of individuals with Major Depressive Disorder (MDD). You are using the BATS system to store and analyze your data.

Your research:

Modeling of speech patterns of adults with Major Depressive Disorder.

By: Claire Nicolas - WPI & Nathan Draudt - WPI

In the study you have conducted, you have recorded 20 people with diagnosed MDD, and 15 without MDD, all reading the same paragraph from Armisted Maupin's *Tales of the City.* From this data, you have collected a .csv file representing the change in pitch overtime during the recording.

The goal of your research is to confirm that there is a speech pattern in people with diagnosed MDD and those without. Ultimately, you wish to create a model that can be used to better diagnose MDD in adults.

In order to do this, you must analyze all the data you have collected and look for similarities between participants with MDD and how their speech pattern differs from those without MDD.

This study is IRB approved. Other researchers involved in this study should be able to view its data.

Why You are using BATS:

You heard about the BATS system from your colleagues. After reviewing the system, you appreciate the online tool and want to try it out. Currently you need an easy to use system that will allow for easy filtering and analysis of the data you just finished collecting.

How to Access BATS:

Use this link:
https://batsmqp.s3.us-east-2.amazonaws.com/presentation/mainPage.html

Scenario 1: Create an Account, Upload Data, Create a Study

For scenario 1, we want for you to make an account with BATS, upload data and create a study where you will house and analyze the uploaded data from.

Creating an Account:

You may fill in the information any way that you wish. For the email, please put in the following email BatsTesting@gmail.com

Creating a Study:

Here, we want you to create a virtual study where you would be uploading and analyzing all the data you've collected for your MDD speech model.

Uploading Data:

Please upload the .csv file we sent to you in the email with this document. The .csv should be called Test_Wave.csv

Scenario 2: Graph Data

For scenario 2 you will guide yourself through the BATS system, use the graphing tools to view the waveform

Graphing Data:

You want to confirm that the data you uploaded in the last scenario is the correct one and not from the other study you are running. To make sure you have the right data uploaded, graph the data on BATS using the available graphing tools.

Scenario 3: Analyze Data, Download Data

For scenario 3 we want to see you find, navigate and use the given toolkits on the data you have recently uploaded. For this task it will be useful to begin on the **Study Page**. We then want you to download the analyzed file in the form of a .csv and the figure it produces (.png).

Analyzing data:

The data you currently have is a little messy and has lots of background noise. In order to clean the data to have better results in future analysis, you want to use one of the offered analysis tools.

From your previous experience, you know that a lowpass filter works wonderfully for filtering out background noise. Try cleaning the previously graphed data using one of the offered tools.

Downloading data:

Once you have successfully filtered the data, download the data from BATS, both the new .csv document and the new graph (.png)