

ScholarWorks@GSU

Identifying Posttraumatic Stress Disorder and its Symptoms: a Diffusion Tensor Imaging Machine Learning Study

Authors	Haller, Olivia
Citation	Haller, Olivia. "Identifying Posttraumatic Stress Disorder and its Symptoms: a Diffusion Tensor Imaging Machine Learning Study." 2022. Thesis, Georgia State University. https://doi.org/10.57709/32366568
DOI	https://doi.org/10.57709/32366568
Download date	2026-03-09 02:24:20
Link to Item	https://hdl.handle.net/20.500.14694/13527

Identifying Posttraumatic Stress Disorder and its Symptoms: a Diffusion Tensor Imaging
Machine Learning Study

by

Olivia C. Haller

Under the Direction of Tricia Z. King, Ph.D.

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2022

ABSTRACT

Posttraumatic stress disorder (PTSD) impacts millions of Americans annually. Altered white matter microstructure may be a potential diagnostic biomarker for PTSD. White matter microstructural differences in persons with PTSD have been studied using machine learning, a method uniquely suited for biological datasets. This study examined the utility of white matter tracts in classifying persons with and without PTSD and predicting PTSD symptom cluster severity amongst trauma-exposed Black American women. Fractional anisotropy of 53 white matter tracts served as input features. Current PTSD presence was estimated using the Clinician-Administered PTSD Scale. Symptom cluster scores were calculated using the PTSD Symptom Scales. Only the random forest model demonstrated above-chance accuracy (58.88%) when classifying persons with and without PTSD. Regression models for symptom scores failed to show positive R-values. Results show a minimal signal for white matter microstructure and suggest a restricted set of white matter tracts is relevant to PTSD presence.

INDEX WORDS: Posttraumatic stress disorder, White matter, Machine learning, Tractography, Neuroimaging,

Copyright by
Olivia Claire Haller
2022

Identifying Posttraumatic Stress Disorder and Its Symptoms: A Diffusion Tensor Imaging
Machine Learning Study

by

Olivia C. Haller

Committee Chair: Tricia King

Committee: Negar Fani

Jessica Turner

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2022

DEDICATION

This proposal is dedicated to those who continue to support me through my academic career. In particular, I want to thank my family and friends who have helped me through challenges and successes.

ACKNOWLEDGEMENTS

This Master's thesis would not have been possible without the assistance of several notable individuals. Firstly, thank you to Dr. Tricia King for assisting me through each step of this process. As my mentor and committee chair, Dr. King has been vital in the conceptualization of several study ideas and the completion of the final product. This study would also not be possible without Dr. Negar Fani. I sincerely appreciate Dr. Fani's willingness to share her data, guide the project, and introduce me to her wonderful team. I am also thankful to Dr. Jessica Turner for her guidance on the methodological considerations of this thesis project and for connecting me with experts at the TReNDS center, including Dr. Sergey Plis. Thank you to Dr. Plis for reviewing my analyses and providing suggestions for refining methods. Thank you to the members of the King Lab, in particular Eric Semmel, Rella Kautiainen, Holly Aleksonis, Stephanie Steinberg, and Jordan Pincus, who provided advice and support throughout this process. I would also like to thank members of the FANs lab and the Grady Trauma Project for the collection and management of the data. Lastly, I am grateful for the participants that dedicated their time and energy to this research. They are the cornerstone of this entire project. This work is supported by the National Institute of Mental Health (MH101380, MH071537, MH094757, MH119603) and the Georgia State University Brains and Behavior Graduate Fellowship awarded to Olivia Haller.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	v
LIST OF TABLES	ix
LIST OF FIGURES	x
1 INTRODUCTION	1
1.1 PTSD Diagnosis and Symptom Clusters in Trauma-Exposed Adults.....	3
1.2 PTSD in Black American Women	5
1.3 Diffusion Tensor Imaging Metrics: Scalars & Tractography	7
1.4 White Matter Microstructure Related to PTSD.....	9
1.5 Machine Learning in Imaging Research.....	11
<i>1.5.1 Linear Support Vector Machine.....</i>	<i>13</i>
<i>1.5.2 Radial Basis Function Support Vector Machine.....</i>	<i>14</i>
<i>1.5.3 Random Forest.....</i>	<i>14</i>
<i>1.5.4 Multilayer Perceptron</i>	<i>14</i>
<i>1.5.5 Linear Regression</i>	<i>15</i>
<i>1.5.6 Least Absolute Shrinkage and Selection Operator</i>	<i>16</i>
<i>1.5.7 Ridge Regression.....</i>	<i>16</i>
<i>1.5.8 Elastic Net Regression</i>	<i>17</i>
1.6 Machine Learning Models in PTSD	17
1.7 Aims of the Proposed Study	20

1.7.1	<i>Specific Aim 1.A</i>	21
1.7.2	<i>Specific Aim 1.B</i>	21
1.7.3	<i>Specific Aim 2</i>	21
1.7.4	<i>Exploratory Aim 3</i>	22
2	METHODS	22
2.1	Participants	22
2.2	Procedure	24
2.3	Assessment of PTSD	25
2.3.1	<i>Clinician-Administered PTSD Scale (CAPS)</i>	25
2.3.2	<i>The Modified PTSD Symptom Scale (PSS)</i>	25
2.4	MRI Data Acquisition & Preprocessing	26
2.5	Tractography	27
2.6	MRI Data Harmonization	29
2.7	Machine Learning	30
2.7.1	<i>Analyses for Specific Aim 1</i>	31
2.7.2	<i>Analyses for Specific Aim 2</i>	32
2.7.3	<i>Analyses for Specific Aim 3</i>	32
3	RESULTS	33
3.1	Demographics	33
3.2	PTSD Assessment	34

3.3	Results for Specific Aim 1.....	35
3.4	Results for Specific Aim 2.....	37
3.5	Results for Specific Aim 3.....	38
3.5.1	<i>Comparison of Classification Models</i>	38
3.5.2	<i>Comparison of Regression Models</i>	40
4	DISCUSSION.....	41
4.1	PTSD Diagnostic Classification Based on White Matter Tracts.....	41
4.2	PTSD Symptom Severity Predication Based on White Matter Tracts	45
4.3	Comparison of Machine Learning Models	47
4.3.1	<i>Relative Performance of Classification Models</i>	47
4.3.2	<i>Relative Performance of Regression Model</i>	48
4.4	Strengths and Weaknesses.....	50
4.5	Conclusions & Future Directions	51
	REFERENCES.....	54

LIST OF TABLES

<i>Table 1. Characteristics of Total TEA sample</i>	23
<i>Table 2. List of White Matter Tracts</i>	28
<i>Table 3. PTSD Symptoms</i>	34
<i>Table 4. Classification Models</i>	35
<i>Table 5. Linear SVM Feature Importance Key at 10-folds</i>	36
<i>Table 6. Regression Models</i>	37
<i>Table 7. Random Forest Feature Importance Key</i>	39

LIST OF FIGURES

<i>Figure 1. Performance and Feature Importance of Linear SVM Classification at 10-folds</i>	
.....	36
<i>Figure 2. Performance of Classifiers and Random Forest Feature Importance at 20-folds</i>	
.....	39
<i>Figure 3. Regression Model Performance</i>	41

1 INTRODUCTION

Approximately 80% of Americans are exposed to at least one traumatic event in their lifetime (Milanak et al., 2019; Overstreet et al., 2017). In a study of 34,676 participants from 20 countries, about 4% of those exposed to traumatic events were diagnosed with current posttraumatic stress disorder (PTSD) when controlling for frequency of exposure (Liu et al., 2017; Shalev et al., 2017). Given that not everyone who is exposed to trauma is diagnosed with PTSD, there is continued interest in better understanding the disease process. In particular, there is a concerted effort to identify PTSD biomarkers. These biomarkers may assist in signaling what biological systems are at play in persons with PTSD and those at risk, along with improved diagnostics (Brewin et al., 2017; Shalev et al., 2017). However, identifying potential biomarkers can be difficult due to the high-dimensional nature of biological datasets. When data is highly dimensional, traditional multivariate and univariate methods are unable to accommodate the large number of features in said dataset within a single model. Therefore, other forms of statistical analysis, such as machine learning, should be explored.

Machine learning refers to a subset of artificial intelligence with the broad goal of building learning algorithms. Diagnostic machine learning methods are a relatively new field of study in psychology that is gaining interest, especially in the realm of biomarkers. There are several advantages to using machine learning methods to identify biomarkers (Orrù et al., 2012). Firstly, machine learning models utilize a multivariate approach that can identify patterns and trends in datasets with many features. Biological datasets can consist of an overwhelming number of variables that are not leveraged in standard univariate analysis but are used in a multivariate approach. Second, the shape of the model can be determined by the resulting algorithm rather than relying on an a priori hypothesis. This characteristic of machine learning

can be particularly advantageous when the model's shape is relatively unknown, as is often the case with biomarkers. Lastly, machine learning characterizes patterns at the individual level as opposed to the group level. Investigation at the individual level represents a significant advantage in clinical settings where personalized health decisions must be made, sometimes on a time-limited basis (Schultebrasucks & Chang, 2021). These characteristics of machine learning make it a useful statistical approach for PTSD researchers seeking to identify potential biomarkers, with the ultimate goal of improved clinical decision-making.

White matter microstructure has been previously investigated as a diagnostic and prognostic marker for PTSD using machine learning methods (Ramos-Lima et al., 2020). However, extant research appears to have utilized majority male, non-Black populations. Black American women have unique environmental experiences that impact the types of traumas they can experience, their health-seeking behaviors, their symptom profile, course, and more (Moody & Lewis, 2019). These experiences may also impact biological variables, including white matter microstructure. It is vital that these experiences are incorporated into diagnostic models to ensure models are applicable to real-world contexts. Further, there is no extant machine learning literature discussing the predictive accuracy of white matter microstructure on the severity of symptom clusters. Evaluating accuracy of white matter models on symptom clusters are a crucial step toward refining the neuropathophysiology of PTSD and evaluating the clinical utility of current diagnostic models.

Therefore, this study seeks to test the accuracy of white matter microstructure in identifying PTSD presence and symptom clusters in a population of trauma-exposed Black American women. This study will begin by providing relevant background on PTSD diagnostics and symptom clusters, the characteristics of PTSD in Black American women that are shared and

unique when compared to other demographic groups, methods for estimating white matter microstructure, white matter findings in PTSD, white matter microstructure findings that relate to the experiences of Black American women, machine learning methodology, and machine learning applications within PTSD populations. After establishing relevant background, the paper will review the specific aims of this study.

1.1 PTSD Diagnosis and Symptom Clusters in Trauma-Exposed Adults

A necessary criterion for PTSD is exposure to a traumatic event. Trauma exposures are conceived as life-threatening, sexually violent, or seriously injurious to the individual. However, diagnostic criteria for PTSD also consider witnessing an event, learning of an event experienced by a friend or family member, or repeated significant exposure to details of a traumatic event(s) to be sufficient trauma exposure. Type of trauma can be associated with the presence and severity of PTSD symptoms, underscoring the need to incorporate individuals with varied traumas into diagnostic research. For example, interpersonal traumatic events are associated with increased negative mood and cognitions (Overstreet et al., 2017), and sexual trauma is associated with greater symptom severity than other traumas in American veterans (Jakob et al., 2017). Factors that influence risk for certain traumas include sex/gender, race/ethnicity, occupation, location, socioeconomic status, and more. Similarly, environmental, developmental, and genetic influences play an important role in the development of PTSD symptoms (Ogle et al., 2013; Stein et al., 2002).

In addition to trauma exposure, the 5th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) includes criteria for four symptom clusters: reexperiencing, avoidance, negative cognitions and mood, and hyperarousal. Criteria for each of the clusters must be met for a PTSD diagnosis. While validated diagnostic criteria consistently require

exposure to a traumatic event and similar symptom criteria, there are some distinguishing aspects across criteria. Compared to the DSM-5, three of the symptom clusters (reexperiencing, avoidance, and hyperarousal) are included in the DSM-IV. The DSM-IV criteria do not include a negative cognitions and mood cluster, although some aspects of negative mood/cognitions are captured in the avoidance and emotional numbing cluster on the DSM-IV. The International Classification of Diseases, 11th Revision (ICD-11) criteria, similar to the DSM-IV criteria, does not include a separate cluster for negative mood/cognitions. However, unlike the DSM-IV, the ICD-11 does not include symptoms to capture negative mood/cognitions in any clusters. In addition to the symptom clusters, the DSM-5 includes an option to specify a dissociative subtype and three supplementary symptoms, which is inconsistent with the DSM-IV or ICD-11.

Reexperiencing, avoidance and emotional numbing, and hyperarousal are the primary symptom clusters that make up the DSM-IV PTSD criteria. Re-experiencing symptoms are the intrusive reliving of the traumatic event(s); this can consist of "flashbacks" or unwanted memories. Avoidance refers to changes in behavior that evade reminders of the event, including certain places or persons. The DSM-IV also includes feelings of detachment, loss of interest, and restricted affective range within the avoidance cluster. When referring to the DSM-IV criteria, the avoidance cluster will be referenced as the avoidance/numbing cluster to highlight this difference. Lastly, hyperarousal symptoms include irritation, hypervigilance, and concentration and sleep disturbances.

PTSD is a heterogeneous disorder both in terms of potential precipitating trauma types and resulting symptoms. In response to the complexity of factors that predict trauma and PTSD, researchers should think critically about the individuals being included in trauma research and their context. Trauma exposure and risk for PTSD are unequally distributed across populations

(Kessler et al., 2017). In American culture, historical systems of sexism and racism are two forms of oppression that currently place Black women at increased risk of exposure to traumatic events and development of PTSD. The following section will discuss PTSD in Black American women, including aspects that are shared with other identities and unique experiences stemming from race/ethnicity and gender/sex intersections.

1.2 PTSD in Black American Women

Please note that there are several qualifications to the findings reported in this section. Firstly, the findings discussed are frequently studied within a framework that explicitly or implicitly assumes Black women are a non-normative sample and that White men are the appropriate baseline. This assumption stems from racist conventions ingrained in academia, which promote an "othering" of non-White Americans. Black American women have unique experiences that must be recognized, but these experiences are not to be understood as "non-normative"—that is, not generalizable to other trauma-exposed populations. Secondly, the development of PTSD is shaped by a lifetime of experiences. It is important to recognize that in discussions of racial/ethnic and sex/gender differences, there is a larger current and historical environmental context that drive these differences and that there is no biological basis for race. Further, trauma exposure modifies and is modified by racial/ethnic, sex/gender, sexual identity, socioeconomic status, and other inequalities. An intersectionality perspective is best for understanding how unique systems of oppression interact and impact the historical and future experiences of each Black woman. Please see (Crenshaw, 1989) for the first use of intersectionality. For a review on intersectionality in health, see (Bowleg, 2012). Extant research on intersectionality in Black women exposed to trauma will be reported, which primarily focuses on race/ethnic and sex/gender intersections.

Female sex/gender and Black race are both associated with an increased risk of PTSD compared to male sex/gender and other races/ethnicities (Asnaani & Hall-Clark, 2017; Christiansen & Berke, 2020; Pineles et al., 2017). However, as demonstrated by Crenshaw (1989), Black women do not experience discrimination because they are Black and female, but because they are Black women. Research studies like the Grady Trauma Project (GTP), which focus on the experiences of Black Americans, are integral in understanding how the intersectional identities of Black American women relate to PTSD. The GTP is an ongoing, large-scale study that has examined risk factors and comorbidities of PTSD in a metropolitan population of Black Americans in Atlanta, GA, USA. Seminal work with GTP participants found that for women, lifetime rates of PTSD were 47.8% and that 86.1% of female participants reported experiencing a significant traumatic event in their lifetime (Gillespie et al., 2009). More recently, in a GTP cohort of 7,430 African American women, 91% of participants reported being exposed to at least one traumatic event, and 32.3% met criteria for PTSD (Gluck et al., 2021). Both GTP studies underscore the inequitable distribution of trauma and PTSD among Black women relative to international rates (Kessler et al., 2017).

In addition to increased risk of exposure to time-limited traumatic events, Black women experience prolonged exposure to gendered racism. Gendered racism is a term used to characterize the unique form of oppression experienced by persons that are simultaneously racially and sexually marginalized, which does not equate to sexism or racism operating individually (Essed, 1991; Spates et al., 2020). Just as racial trauma can generate traumatic stress and PTSD responses in Black men (Paradies, 2006; Pieterse et al., 2012), increased exposure to gendered racism was related to an increased risk of PTSD and more severe symptoms in a sample of GTP women (Mekawi, Carter, et al., 2021). Importantly, traumatic exposures to

discrimination do not need to be violent or blatant in nature to generate PTSD symptoms. For example, gendered racial microaggressions are positively associated with traumatic stress in Black women (Moody & Lewis, 2019). There is some evidence that mechanisms contributing to PTSD in Black women can stem from the distinct roles that Black women are socialized to adopt (Spates et al., 2020). In particular, the Strong Black Women Schema can contribute to poor emotional regulation by discouraging emotional intercommunication, self-care, and social support (Mekawi, Watson-Singleton, et al., 2021). Some work with GTP participants has suggested that emotional dysregulation may be a unique risk for more complex PTSD symptoms (Powers et al., 2015; Powers et al., 2017). These results underscore the importance of focusing on mechanisms contributing to PTSD in populations with the highest burden. The remainder of this introduction will focus on neuroimaging findings and analysis, starting with an overview of metrics for measuring white matter microstructure.

1.3 Diffusion Tensor Imaging Metrics: Scalars & Tractography

Diffusion tensor imaging (DTI) is an established method of estimating white matter microstructure. DTI is a form of structural imaging that measures the movement of water molecules in brain tissue. DTI is a useful imaging tool because it can measure microstructural changes that are not easily quantified in traditional structural magnetic resonance imaging (MRI) protocols. DTI data can be analyzed in a number of ways, but the most common metric is fractional anisotropy (FA). FA measures the degree of anisotropy in the diffusion ellipsoid. Low anisotropy represents isotropic or unconstrained diffusion, and high anisotropy indicates diffusion in a single direction. It is generally accepted that high anisotropy is indicative of healthy white matter integrity because intact myelin and parallel fibers constrain the flow of water molecules in one direction. However, it should be noted that there are a number of factors

that modify anisotropy in healthy axons. For example, crossing fibers in a single voxel will result in a lower anisotropy value even if all the fibers captured in that voxel are intact. Other factors that affect anisotropy include axon diameter and axon maturity (Sampaio-Baptista & Johansen-Berg, 2017).

DTI tractography takes advantage of the anatomical structure of white matter fibers to provide nuanced information about the structural connections between anatomical regions. Tractography is a method of region of interest selection that uses water molecules to define anatomical white matter tracts. DTI tractography has strengths in that it can provide additional functional and structural information compared to other DTI methods. Tractography data does not need to be spatially normalized for individuals to be compared, which can be especially useful in populations one might expect to have subtle differences in white matter microstructure (Takemura et al., 2016). However, just like with other DTI methods, tractography cannot fully account for issues such as crossing fibers or “kissing” fibers.

Tractography data is traditionally measured in two ways, referred to as deterministic and probabilistic tractography. In deterministic tractography, white matter fibers are defined based on a combination of traditional anatomical information and probability pathways. Probabilistic tractography does not attempt to recreate any classical pathways but instead uses the probability within each voxel to identify the most likely pathway between two regions (Kuceyeski et al., 2011). While both probabilistic and deterministic methods have their value, deterministic tractography will be used in this study. Deterministic methods were selected for two primary reasons. First, multi-fiber deterministic methods appear to be more accurate at reconstructing tract connectivity, given that they are less likely to include false-positive fibers (Sarwar et al.,

2019). Additionally, deterministic methods are less computationally intensive, making them ideal for research replication and clinical application (Takemura et al., 2016).

1.4 White Matter Microstructure Related to PTSD

Another advantage of DTI tractography is that it is conceptually straightforward to relate white matter tracts with functional connections. Relationships between white matter microstructure and functional networks are useful as research has made significant advances in identifying altered functional pathways in PTSD and relating these networks to symptoms. Notable functional systems include the default mode network, the salience network, and the central executive network (Akiki et al., 2017). Anatomical regions within these networks are predominately frontal and subcortical, including regions like the prefrontal cortex, amygdala, insula, and hippocampus. Many of these functionally pertinent regions are physically connected via the cingulum bundle, a white matter tract that passes through frontal cortices and subcortical nuclei (Bubb et al., 2018).

In individuals with PTSD, altered FA in the cingulum bundle is perhaps the most consistent DTI finding (Siehl et al., 2018). However, the direction of the finding is inconsistent, with some studies finding increased FA and a similar number of studies finding decreased FA when compared to trauma-exposed controls and non-exposed controls (Kunimatsu et al., 2020). Differences in findings are likely due to the nature of FA measurement and methodological considerations. Other regions of altered FA include the thalamic radiations, corpus callosum, uncinate fasciculus, superior longitudinal fasciculus, superior fronto-occipital fasciculus, and the inferior fronto-occipital fasciculus (Dennis et al., 2019; Kunimatsu et al., 2020). The widespread nature of FA changes and the variation in the direction of these changes underscores the diffuse pathological impact of PTSD but makes it difficult to identify meaningful patterns of change. A

recent meta-analysis by Ju et al. (2020) found a general pattern of decreased FA in anterior anatomical regions and increased FA in visual processing areas in PTSD across studies.

Increased FA values in visual processing tracts may relate to flashback symptoms or other forms of involuntary reexperiencing. Decreased FA in anterior regions agrees with reported functional dysconnectivity between frontal and subcortical structures reported in other work (Akiki et al., 2017).

A variety of socio-environmental experiences that disproportionately impact Black Americans are associated with unique patterns of white matter integrity. Racial discrimination is associated with variance in white matter microstructure (lower FA throughout white matter pathways) in trauma-exposed Black American women (Fani et al., 2022; Okeke et al., 2022). Black Americans are more likely to grow up with lower socioeconomic status compared to white Americans (Williams & Collins, 1995). Childhood socioeconomic status is positively related to FA values throughout development (Dufford et al., 2020). Black Americans are more likely to be born prematurely in comparison to their white counterparts (York et al., 2010). Preterm birth is associated with abnormal white matter development and reductions in white matter microstructure (Rogers et al., 2016). To the best of our knowledge, there is no literature that examines white matter microstructure in PTSD on the function of race. However, PTSD and DTI studies have examined Black women as a population. In a sample of Black American women from the GTP, decreased FA was found in the posterior cingulum of women with PTSD compared to trauma-exposed controls (Fani et al., 2012). This finding is consistent with the broader body of work that highlights the cingulum as a region of interest in PTSD. Please note that the qualifications described above regarding research findings in Black American women also apply to this paragraph.

Despite the explosion of DTI-based research studies in past years, notable limitations exist in the literature. The heterogeneity, both in terms of location and type of changes to white matter microstructure in trauma-exposed controls and persons with PTSD, makes it difficult to interpret research findings in a cohesive narrative. The reported FA findings suggest that, allowing for a few regions of interest (e.g., cingulum), there is not a clear pattern of differences between persons with and without PTSD when examining a select number of tracts or combining findings via meta-analyses.

1.5 Machine Learning in Imaging Research

Supervised machine learning models refer to a subset of machine learning algorithms that use labeled data to create a model. Labeled data is used in the training stage, during which the algorithm derives its function from the predefined data set. The algorithm is then given unlabeled data, referred to as the test dataset. The function derived during the testing phase is applied to the unlabeled data. The success of the final model is evaluated by its ability to classify or predict the outcome variables based during the testing phase (Kourou et al., 2015; Sirsat et al., 2020; Zhu et al., 2020).

There are three main advantages to using machine learning methods for investigating PTSD outcomes in trauma-exposed controls. Firstly, the majority of machine learning models utilize a multivariate approach. A machine learning model can use as many features as the user wishes to include (although there are important methodological issues with using too many features). When studying imaging data, it is difficult to examine each variable without substantially reducing the power of the analysis due to multiple comparisons. Machine learning models are capable of identifying patterns and trends with a larger number of features, which

make them appropriate for looking for variables with small marginal effects (Kriegskorte & Bandettini 2007).

Secondly, a machine learning approach can be useful when prior exploratory analysis has not enabled proper determination of the shape of the underlying model. As previously established, the relationship between white matter changes and PTSD is still debated in the general population, and even less is known about these relationships in Black women. Therefore, the predictive statements that can be made for this population are limited, including the shape of the relationship. In some machine learning approaches, the shape of the model is determined by the resulting algorithm, which reduces potential biases that occur when testing a hypothesized model. This is not to say that this approach is always appropriate. Indeed, testing hypothesized models lies at the core of psychological science. However, when there is only a limited foundation to generate a model, machine learning can serve as a useful method of avoiding bias by deriving the model from the available data.

Thirdly, pattern recognition allows characterization at the level of the individual. Standard univariate analysis is only able to characterize activation based on group differences, which becomes a significant disadvantage in a clinical setting where characterization needs to be done at the individual level (Orrù et al., 2012). A unique utility of supervised machine learning is that, assuming the model has been successfully trained, one can take individual data and predict the desired variable without the need for additional group-level analysis. This characteristic makes pattern recognition potentially more practical for the clinical setting where patients are assessed individually rather than being coalesced into groups (Stephan et al., 2017). Theoretically, a machine learning algorithm could identify relevant neurobiological features of a psychological disorder. These features could then be gathered for an individual and inputted into

the trained algorithm to guide diagnostic decisions. In the same vein, a machine learning algorithm could also assist in prognostic decisions, such as recommending certain treatments. An algorithm could be trained to predict treatment response based on a valid set of predictors and could then be used to predict individual response (Schultebrucks & Chang, 2021). While this paper does not serve to propose a clinically applicable algorithm, it does assist in evaluating the feasibility of this technique with trauma-exposed adults (TEAs).

There are a variety of machine learning models that can be used to examine data. Below are descriptions of some models that are popular with psychology researchers that use biological data and that we plan to use in this study.

1.5.1 Linear Support Vector Machine

Linear support vector machines (SVMs) are some of the most widely used machine learning models and have gained popularity in the biomedical sciences over the past decade. SVMs work by fitting an optimal hyperplane, or a subspace that is one dimension less than the considered space, between two sets of data points, maximizing the margin between groups within the feature space. The decision boundary is learned during the training phase, where support vectors, or training samples that lie on the hyperplane's borders, define that margin. The resulting model is then used to predict the group of the test data. One advantage of SVMs is that the feature weights are easily obtained. Feature weights are often used to calculate importance scores. Importance scores are thought to represent features that contribute to the success of the model. In this study, these scores will be used to identify white matter tracts that are contributing heavily to the model.

1.5.2 Radial Basis Function Support Vector Machine

Radial basis function SVMs operate similarly to linear SVMs, wherein the model uses support vectors to maximize the margin between classes using an optimal hyperplane. However, unlike linear SVMs, which work best for datasets that are already linearly separable, radial basis function SVMs use a kernel or a function that transforms a non-linear dataset into a linear-separable dataset using a higher-dimensional space. Radial basis function kernels are one of the most popular kernels due to their similarity to the Gaussian distribution, and they are often the “go-to” kernel for SVMs in non-linear datasets. One disadvantage to radial basis function SVMs is that they can be poor at generalization due to overfitting. This is especially true in small datasets, where the transformation into high-dimensional space can result in a hyperplane that is too specific to the training data.

1.5.3 Random Forest

Random forest is a classification method that works by constructing thousands of decision trees. These trees are grown based on the labeled input features and the output of the training data. Each tree is grown using a bootstrap sample of the training data and is composed of a split node and a leaf node. The split node sends data points to the left or right subtrees by thresholding one of the input features. The leaf node stores the incoming sample statistic. Random forest can be robust to overfitting and ranks feature importance similar to linear SVMs. However, random forests are complex and are therefore harder to interpret than linear SVMs. They also typically take more computation time than other methods.

1.5.4 Multilayer Perceptron

Multilayer perceptron is a neural network that stacks multiple perceptrons, or multiple machine learning algorithms, together. Each perceptron projects a set of inputs to a

multidimensional representation of the previous layer. Thus, as the number of layers increases, the network is better able to model increasingly complex structures. The output projection of the input data is compared to the desired label of training data using a binary cross-entropy cost function. The method then uses a learning algorithm to minimize the binary cross-entropy. When using biological data, multilayer perceptrons can be more successful than other machines because they allow for increased complexity. However, this complexity comes at the cost of reduced interpretability of findings and increased risk of overfitting. These costs make it less optimal for answering targeted questions in small datasets. Additionally, multilayer perceptrons are considered a “black box” method where the components of the algorithm cannot be derived (Ulloa Cerna, 2016). As a result, values such as feature weights are not observable, meaning it is not possible to identify which features contribute to the model.

1.5.5 Linear Regression

Linear regression is the simplest form of regression algorithm and is, therefore, a popular approach in psychological research. However, linear regression is not robust to outliers and is prone to overfitting. Linear regression is weak to these issues because there is no penalty to reduce the influence of outliers, nor is there a penalty for multicollinearity. For example, if the model identifies one feature as being particularly important during the training phase, the model may place a large weight on the feature. Then, when the model is applied to the testing data, there will be lower testing accuracy compared to the training accuracy. A benefit of linear regression is that it has straightforward interpretability. Additionally, its popularity can make it an informative comparison to other more complex models.

1.5.6 Least Absolute Shrinkage and Selection Operator

Least absolute shrinkage and selection operator (LASSO) regression is based on linear regression and applies the absolute value penalization, otherwise referred to as the L1-norm regularization, to the ordinary least squares loss function. L1-norm regularization aims to minimize the sum of the absolute value of the regression coefficients by setting most coefficients to zero and retaining one random feature among the correlated ones. Therefore, LASSO models can only select a maximum of $N-1$ features in the final model, where N is the sample size. For some datasets with many contributing features and very few samples, this can be a problem as the penalty will scale back the number of features considerably. The benefit of LASSO regression is the restriction of features creates an algorithm that optimizes the predictors and reduces the model complexity. These benefits are particularly important in models with a large amount of noise relative to signal.

1.5.7 Ridge Regression

Just like LASSO, ridge regression is a form of linear regression with an added penalty function. Unlike the L1-norm regularization that penalizes the model for the sum of the absolute value of the weights, ridge regression penalizes the model for the sum of the squared value of the weights. This penalty function, called the squared penalization or the L2-norm regularization, can shrink the regression coefficients and create a more even distribution because extreme weights are more heavily penalized compared to normal weights. As a result, ridge regressions typically have a better fit than other models. However, ridge regression works best when all inputted features are considered informative to the model, which is not always the case with biological datasets, and can be hard to predict when looking at novel models.

1.5.8 Elastic Net Regression

Elastic net builds on both LASSO and ridge regression by including both penalty functions in the ordinary least squares loss function. These dual penalizations protect against the risk of multicollinearity when using highly related features, which is often the case in biological samples. As with other regressors, a regularization parameter λ is used to control the trade-off between the prediction error on the training data and regularization. In addition, a mixing parameter α is used to control the relative weighting of the penalty functions. Model comparisons in neuroimaging datasets indicate that elastic net is optimal for small datasets with high dimensionality (Jollans et al., 2019).

1.6 Machine Learning Models in PTSD

As noted previously, machine learning methods represent a significant opportunity for improved research methodology in PTSD studies. Ramos-Lima et al. (2020) published a recent review of 49 articles using machine learning to investigate trauma-related disorders. They note that 41 of the reported articles showed good accuracy of greater than 80%, supporting the use of this technique to answer trauma-related questions. The majority of the reported studies investigated diagnostic and prognostic-related questions using behavioral and health data. Only 13 studies utilized neuroimaging data, and of those, two used DTI FA for prognostic questions. None of the studies included in the review utilized DTI data for diagnostics (Ramos-Lima et al., 2020). To the best of this author's knowledge, only one study has used diffusion data to make diagnostic classifications for PTSD. This study included individuals exposed to an earthquake in China (Suo et al., 2022). Two groups made of 77 participants with PTSD and 76 demographically-matched trauma-exposed controls were included in the study. White matter microstructure was measured using a fiber quantification approach, wherein diffusion metrics

were quantified at 100 equally-spaced points along 20 deterministically defined tracts. Mean accuracy of FA when predicting individuals with and without PTSD was 64.3%. Feature importance was not reported.

Two studies noted by Ramos-Lima and colleagues (2020) have used DTI for prognostic PTSD models. Li et al. (2016) studied Chinese emergency department patients with mild traumatic brain injury who were evaluated for PTSD at 1 and 6 months after injury. FA and mean diffusivity (MD) were measured using tract-based spatial statistics within three days of injury. They found good predictive accuracy for MD of 75.56% but reported that FA did not discriminate above chance. A second study (Wang et al., 2016) used 33 emergency department patients exposed to a traffic collision (nationality, race, and/or ethnicity unreported). Cortical thickness and deterministic tractography (FA and MD scalars) were combined, and three SVM kernels were tested. Predictive accuracies for linear, polynomial and radial basis function SVM were 63.45%, 68.32%, and 79.86% respectively. Radial basis function SVM also evaluated features independently and found that the predictive accuracies for FA and MD were 69.54% and 54.69% respectively. Since the publication of the review by Ramos-Lima and colleagues (2020), this author was unable to find any new studies investigating PTSD in adults using DTI data. To the best of this author's knowledge, there is no available literature on diagnostic or prognostic DTI models that specifically include Black American women. Although some studies do not report enough demographic data to understand their samples.

No existing machine learning research has used DTI data to investigate PTSD symptom clusters, but several recent studies have used functional magnetic resonance imaging (fMRI) data. The most relevant of these used least-angle regression to predict the severity of intrusion, avoidance, cognition/mood, and arousal/reactivity symptoms using functional connectivity data

(Zandvakili et al., 2020). R^2 values for intrusion, avoidance, cognition/mood, and arousal/reactivity symptoms was 0.33, 0.23, -0.01 , and 0.06, respectively. Differential strengths in relationships suggest that connectivity may play a greater role in intrusion and avoidance symptoms compared to cognition/mood and arousal/reactivity symptoms in PTSD. In a similar vein, two recent articles by Nicholson and colleagues (2019; 2020) use fMRI data to classify healthy controls, persons with PTSD, and dissociative subtype PTSD. All models showed good to excellent accuracy, ranging from 63% - 92% across studies.

Across existing PTSD studies that incorporate white matter scalars and machine learning, there are some notable gaps. Firstly, there is no existing literature that has asked these questions about Black American women. As outlined above, Black American women are more likely to have unique environmental experiences (e.g., gendered racism) that may impact clinical factors such as trauma type and symptom severity. Black American women may also have experiences that can uniquely affect their white matter microstructure. Therefore, to better understand the generalizability of findings and to increase representation in psychological science, Black American women should be included in research. Secondly, the current literature has yet to utilize machine learning regression to examine the relationship between white matter microstructure and PTSD symptom clusters. Given that white matter microstructure may be more strongly related to certain subtypes of symptoms, it may be useful to examine these relationships. Third, the existing body of work has primarily focused on classification models using an SVM approach. Incorporation of additional models may demonstrate improved classification and regression strength, which in turn may indicate characteristics of the dataset. For example, prior prognostic work showed improved accuracy when using a radial basis function SVM which suggests a non-linear distribution of data (Wang et al., 2016). A non-linear

distribution indicates rather than data being separable on a simple lower-order dimension (e.g., high FA and low FA), the data is best separated on a higher-order dimension.

1.7 Aims of the Proposed Study

The prevalence and impact of PTSD in Black American women presents a significant challenge to mental health professionals. Effective biomarkers may provide greater insight into the disease process of PTSD, which may assist in more accurate diagnosis and treatment. The characteristics of machine learning make it an advantageous statistical approach for PTSD researchers seeking to develop better diagnostic approaches using biological information.

White matter microstructure has been previously investigated as a diagnostic and prognostic marker for PTSD using machine learning methods with promising results (Ramos-Lima et al., 2020). However, prior research appears to have focused utilized majority male, non-Black populations. Black American women are more susceptible to unique environmental experiences that may impact their psychological and biological profiles (Moody & Lewis, 2019). It is vital that these experiences are incorporated into diagnostic models to ensure models are applicable to real-world contexts. Further, there is no extant machine learning literature discussing the predictive accuracy of white matter microstructure on the severity of symptom clusters. Evaluating the accuracy of white matter models on symptom clusters are may assist in further refining pathological questions about PTSD. Lastly, there is limited work comparing the accuracy of different machine learning models in PTSD populations when using white matter as a predictive measure.

Therefore, this proposal seeks to test the accuracy of white matter microstructure in identifying PTSD presence and symptom clusters in a population of trauma-exposed Black

American women using the DSM-5 diagnostic criteria. It also seeks to compare performance across different machine learning models.

1.7.1 Specific Aim 1.A: Examine the accuracy of FA using a linear SVM paradigm, based on the model's accuracy in discriminating between TEA PTSD- and TEA PTSD+ individuals in a population of Black American women.

Hypothesis 1.A: Linear SVM was selected as the primary model due to its popularity in prior work. Prior work using FA has reported an accuracy of 64.3%. We hypothesize that FA will show good to excellent accuracy (70%-89%).

1.7.2 Specific Aim 1.B: To investigate which brain regions or patterns of brain regions contribute the most to the success of the linear SVM model by investigating the feature importance scores of white matter tracts.

Hypothesis 1.B: White matter tracts associated with the prefrontal and limbic cortex will have the highest feature weights in all models based on existing literature. In particular, we expect cingulum tracts to have the greatest feature weight over all other white matter tracts.

1.7.3 Specific Aim 2: Examine the relationship between FA and PTSD symptom clusters (reexperiencing, avoidance and emotional numbing, and hyperarousal) in TEA Black women using an elastic net model.

Hypothesis 2: Elastic net was selected as the primary model due to its popularity in machine learning neuroimaging research and its more conservative nature. We hypothesize that models will show moderate to good fit ($R^2=0.5-0.6$) for all symptom clusters.

1.7.4 Exploratory Aim 3: Compare the accuracy of common machine learning models for each of the previously described specific aims. Multilayer perceptron, radial basis function SVM, and random forest will be compared to linear SVM performance. Linear regression, LASSO, and ridge regression will be compared to elastic net performance.

Hypothesis 3: Linear SVM and elastic net were chosen to be the primary models for this study because they are appropriate for the data and are frequently used in the bioimaging field. For classifiers, multilayer perceptron will likely have higher accuracy but will be less interpretable. Random forest may also be a good fit for the dataset and will likely perform similarly to linear SVM. Radial basis function SVM will perform the worst due to overfitting. For regression, most models will perform similarly. However, linear regression will be the worst due to overfitting. We hypothesize that models will show moderate to good results ($R^2=0.5-0.6$) for all symptom clusters.

2 METHODS

2.1 Participants

Data for this study was collected as part of the GTP. As previously discussed, the GTP involves several collaborating studies focused on risk and resilience in trauma-related disorders (MH101380, MH071537, MH094757; see Gillespie et al., 2009 for more detail on GTP methodology). Study procedures were reviewed and approved by The Institutional Review Board of Emory University and Grady Hospital Research Oversight Committee. Participants were recruited in the waiting rooms of non-psychiatric medical clinics at a publicly funded, non-profit hospital in inner-city Atlanta, Georgia. Participants were considered eligible for GTP if they were at least 18 years old, understood English, and were willing and able to provide informed

consent. Demographics of the sample can be found in Table 1. Participants (100% female) had a mean age of 39.28 (SD = 21.14; range: 18-62) at the time of study participation.

Table 1. Characteristics of Total TEA sample

	TEA (N=174)
	%(n)
Education	
< 12 years	13% (23)
High School/G.E.D.	32.7% (57)
Some college/technical school	29.9% (52)
College/technical school degree	19% (35)
Graduate degree	2.9% (5)
Employment (% Employed)	33.3% (58)
Income	
\$0 – 249	11.4% (19)
\$250 – 499	11.4% (19)
\$500 – 999	29.9% (50)
\$1000 – 1999	28.7% (48)
\$2000+	18.6% (31)
Substance Abuse (% Yes)	9.5% (16)
Birth Control (% Yes)	12.1% (21)
Handedness (% Left-Handed)	
Ethnicity	
Hispanic/Latinx	1.1% (2)
Non-Hispanic/Latinx	98.3% (171)
Race	
Black American	97.7% (170)
Mixed Race	1.1% (2)
Other	1.2% (2)
M (SD)	
Age (years)	39.28 (21.14)
TEI Lifetime Frequency	2.46 (1.24)
CTQ	
Sexual Abuse	8.68 (5.56)
Physical Abuse	8.14 (3.83)
Physical Neglect	6.59 (2.85)
Emotional Abuse	8.98 (4.43)
Emotional Neglect	9.18 (4.61)
BDI Total Score	12.75 (11.09)

2.2 Procedure

After providing informed consent, all participants were asked to complete several measures to assess trauma exposure, stress, current behavioral and psychological health, and PTSD. Of the GTP participants that reported trauma exposure and reported either no/few PTSD symptoms or elevated/clinically significant PTSD symptoms, a subset completed additional procedures, including an MRI scan and a psychodiagnostic interview. The psychodiagnostic interview consisted of the Clinician-Administered PTSD Scale (Weathers et al., 2018); participants also completed the Structured Clinical Interview for DSM-IV (SCID-DSMIV; First et al., 1995) if the participant was seen between 2005 and June 2012 or the Mini-International Neuropsychiatric Interview (MINI; Sheehan et al., 1998) if the participant was screened from June 2012 – October 2016. The Traumatic Events Inventory (TEI) was used to detail the frequency and type of trauma experienced by the participants (Schwartz et al., 2006; Schwartz et al., 2005). The Childhood Trauma Questionnaire (CTQ) was used to measure childhood trauma exposure (Bernstein, 1998; Bernstein et al., 2003). The Beck Depression Inventory-II (BDI-II; Beck et al., 1996) was used to measure current depression symptoms on the day of the MRI scan. Participants were asked to avoid caffeine intake on the day of the MRI scan. Participants were excluded if they had any current neurological conditions, current bipolar disorder, current substance or alcohol dependence or intoxication, or current psychosis. Participants were also excluded if they reported MRI contraindications (e.g., metal implants, pacemaker). In total, 174 participants received MRI scans.

2.3 Assessment of PTSD

2.3.1 Clinician-Administered PTSD Scale (CAPS)

The CAPS-5 was used to assign PTSD diagnostic status (current and lifetime) in a subset of participants (N=136). The CAPS-5 is a semi-structured interview assessment meant to diagnose current PTSD based on the DSM-5 criteria (Weathers et al., 2018). The CAPS has been psychometrically validated in both civilian and veteran populations with good to excellent reliability and validity (Bovin et al., 2016; Pupo et al., 2011). Additionally, interrater reliability for diagnosis of PTSD in a subsample of Black American women from the GTP has been shown to be good in prior work ($k = 0.83$; Powers et al., 2017). The presence of PTSD symptoms was assessed by an interviewer, who scored each criterion from 0 (absent) to 4 (extreme/incapacitating) based on the frequency and severity of each criterion. In addition to exposure to criterion A traumatic event, individuals needed to have at least one threshold criterion B symptom (reexperiencing), one threshold criterion C symptom (avoidance), two criterion D symptoms (negative cognitions and mood), and two criterion E symptoms (reactivity and arousal), with duration longer than one month and functional impairment present to be diagnosed with PTSD. The CAPS-5 scores on a severity continuum with higher scores indicating greater frequency and severity of symptoms, but for the purposes of this study, TEAs were examined dichotomously as either PTSD+, or PTSD- for the past month.

2.3.2 The Modified PTSD Symptom Scale (PSS)

The PSS was used to examine the severity of PTSD symptoms in the past two weeks as measured on the day of the participants' MRI scan (Falsetti et al., 1993; Foa & Tolin, 2000). The PSS is a self-report measure designed to assess the severity of posttraumatic symptom clusters, as defined in the DSM-IV (i.e., reexperiencing, avoidance/emotional numbing, hyperarousal).

The PSS has been shown to have good reliability and validity (Coffey et al., 1998) and high internal consistency in a previous study using a subset of the study population ($\alpha = 0.92$; Powers et al., 2017). The PSS was selected as the primary measure for symptom severity because it has good convergent and concurrent validity with the CAPS and SCID in female populations (Santiago Papini, 2014). Additionally, the interview form of the PSS correlates strongly with the CAPS, Posttraumatic Diagnostic Scale for DSM-5, and the PTSD Checklist, Specific Version (Foa et al., 2016). The PSS also allowed for the inclusion of the maximum sample size in our statistical analysis ($n=166$). To better align with the current DSM-5 criteria, the avoidance/numbness subscale was split into two subscales to create a total of four subscales. Prior work in a subset of this sample has demonstrated good internal consistency for the four subscale approach (hyperarousal: 5 items, $\alpha = 0.79$; avoidance: 2 items, $\alpha = 0.72$; numbness: 3 items, $\alpha = 0.77$; reexperiencing: 5 items; $\alpha = 0.83$; (Mekawi et al., 2020). The distributions of scores were checked for skew and kurtosis and found to be within normal limits. A normal distribution means that the scores on the PSS and CAPS are not biased towards higher or lower values, nor are they abnormally clustered towards central or outlying values. This means the sample is not abnormally representative of highly symptomatic or asymptomatic individuals.

2.4 MRI Data Acquisition & Preprocessing

Participants were scanned on one of two research-dedicated Siemens 3-Tesla TIM-Trio scanners (12 channel head coil) located at Emory University Hospital. Diffusion-weighted images were acquired similarly across both sites, with maximum gradient strength of 40mTm^{-1} with the following parameters: 39 x 2.5 mm thick axial slices, matrix = 128 x 128, field of view (FOV) = 220 x 220 mm, voxel size = 1.72 x 1.72 x 2.5 mm. Diffusion weighting was isotropically distributed along 60 directions using a b-value of 1,000 s/mm^2 . Four normalization

images, with no diffusion encoding ($b=0$), were acquired and averaged for each direction using linear rigid-body registration (FLIRT; (Jenkinson et al., 2002; Jenkinson & Smith, 2001). All diffusion-weighted image processing and analysis were conducted using FMRIB Software Library (FSL version 4.1; www.fmrib.ox.ac.uk/fsl; (Woolrich et al., 2009).

Correction for head motion and eddy current distortion was performed for data from each participant using an automated affine registration algorithm. Both diffusion-weighted and T1 images were skull-stripped using the FSL brain extraction tool (Jenkinson et al., 2005; Smith, 2002).

2.5 Tractography

Deterministic tractography was completed using DSI Studio (<http://dsi-studio.labsolver.org>). A deterministic fiber tracking algorithm (Yeh et al., 2013) was used with augmented tracking strategies (Yeh, 2020b) to improve reproducibility. The anatomy prior of a tractography atlas (Yeh et al., 2018) was used to map a total of 55 tracts with a distance tolerance of 18 (mm). A seeding region was placed at the track region, as indicated by the tractography atlas. A region of interest was placed at the track tolerance region (39,48,33) with a volume size of $2.4e+06$ mm cubic. The track-to-voxel ratio was set to 2. The anisotropy threshold was randomly selected. The angular threshold was randomly selected from 15 degrees to 90 degrees. The step size was randomly selected from 0.5 voxels to 1.5 voxels. Tracks with a length shorter than 30 or longer than 300 mm were discarded. Topology-informed pruning (Yeh et al., 2019) was applied to the tractography with 32 iteration(s) to remove false connections. A list of all the tracts that were initially generated can be found in Table 2.

All subjects with MRI data (N=174) were initially included in tractography analysis. Following tractography, the left and right corticobulbar tract was removed due to poor tract

generation (46.3% missing), possibly due to the large turning angle of the tract (Yeh, 2020a). The remaining tracts resulted in 53 features. Four subjects were removed who had more than seven tracts fail to generate as this was one standard deviation above the norm ($M= 1.44$, $SD = 6.02$). Average FA values were extracted for the left and right tract where or segment if applicable. Prior to harmonization, any missing FA values were replaced using the feature implementation tool in sci-kit learn (<https://scikit-learn.org/stable/index.html>). A univariate imputation algorithm was used to replace missing FA values with the average FA value for each tract.

Table 2. List of White Matter Tracts

Tract	Segments
Arcuate Fasciculus	
Cingulum	Frontal Parahippocampal Frontal Parietal Parahippocampal Parietal Parahippocampal Parolfactory
Extreme Capsule	
Frontal Aslant	
Inferior Fronto-Occipital Fasciculus	
Inferior Longitudinal Fasciculus	
Middle Longitudinal Fasciculus	
Parietal Aslant	
Superior Longitudinal Fasciculus	Segment 1 Segment 2 Segment 3
Uncinate Fasciculus	
Vertical Occipital Fasciculus	
Corticospinal Tract	
Thalamic Radiations	Anterior Posterior Superior

Anterior Commissure	
	Tapetum
	Forceps Major
Corpus Callosum	Forceps Minor
	Body
Fornix	
Optic Radiations	
Medial Lemniscus	

2.6 MRI Data Harmonization

The combined association test with generalized additive models (ComBat-GAM; now referred to as neuroHarmonize) was used to integrate images from both MR scanners (Pomponio et al., 2020). neuroHarmonize is an adaption of ComBat, which was originally proposed to remove batch effects in genomics data (Johnson et al., 2007). ComBat has been previously modified for DTI data (Fortin et al., 2017). In DTI, ComBat works as a voxel-wise harmonization approach that regresses select covariates to harmonize across data sets. Site-specific parameters included in the regression equation include the additive site effects, which are assumed to have normal prior distributions, and multiplicative site effect, which is assumed to be inverse gamma distributions (Pinto et al., 2020). Given the wide age range included in our sample, neuroHarmonize was chosen because it is designed for harmonization of lifespan datasets. neuroHarmonize builds on ComBat by using GAMs to capture non-linearities in age-related differences in brain structure, including age, gender, and intracranial volume. neuroHarmonize has been shown to preserve biological differences in white matter across age ranges and sample size better than linear and quadratic models (Pomponio et al., 2020). Age was included as a covariate in the neuroHarmonize for this study. Intracranial volume was not included as a covariate because the images were normalized to a shared template during preprocessing.

2.7 Machine Learning

All machine learning models were generated using the Polyssifier toolbox (<https://github.com/alvarouc/polyssifier>) implemented in the scikit-learn Python library (Pedregosa et al., 2011). Polyssifier is a publicly available toolbox that is able to run multiple classification approaches simultaneously. Findings were validated with k -fold cross-validation at the common levels of 5, 10, and 20 folds (Gareth et al., 2013; Kuhn & Johnson, 2013). Prior to algorithm generation, features were standardized using the StandardScaler in scikit-learn to ensure a normalized distribution.

This study did not rely on null-hypothesis significance testing; instead, model validation tested the "fit" of the model. The fit of a machine learning model refers to a model's ability to align with training and testing data. In other words, it refers to the algorithm's ability to generalize to novel data. When a model underfits, it will have high training and high testing error. Underfit models limit generalizability because they make assumptions about the data shape that are unsupported by the training data and do not accurately predict novel data. Underfitting can be addressed in any sample by increasing the algorithm's complexity and is not usually a significant risk in small clinical samples (Vabalas et al., 2019). Overfit models will have an extremely low training error but a high testing error. Overfitting results from a model that is too rigid to the shape of the training data, typically caused by the incorporation of noise in the training sample that provides no discriminative ability in novel data.

Overfitting is influenced by the size of the sample, the number of features included, the statistical method, and the validation procedure. Overfitting is the primary risk for clinical samples with biological features because it is most common in small samples with a high number of features. Ideal sample size in machine learning is a debated topic, and there is currently no

prescriptive sample size of models. The sample used in this study is considered moderately sized for algorithms utilizing finite clinical samples (Vabalas et al., 2019), but it is undeniable that fit exponentially improves as samples increase in size (Cui & Gong, 2018). Therefore, the following steps were employed to reduce the risk of overfitting. Firstly, the number of features was reduced from 90 white matter tracts to 53 based on their relevance in the literature. Models with a high number of features increase the potential noise and the risk of overfitting (Dietterich, 1995; Roelofs et al., 2019). The study only included pathologically validated tracts in the cortex. Secondly, linear SVM and elastic net were selected as primary models because they are robust to overfitting in small samples (Jollans et al., 2019; Ulloa Cerna, 2016). Test/training validation, where the sample is split into a testing and training group, is considered the gold standard of validation to avoid overfitting because the validation or testing group is completely novel to the machine trained with the testing data. *k*-fold cross-validation was used for this sample because it is thought to provide a more accurate assessment of model performance in small datasets (Liu, 2017). In *k*-fold cross validation the samples are divided into *k* sets, using *k* - 1 sets for training and one for testing. The average of *k* scores for the testing score is then calculated. Polyssifier calculates the average score for each combination of parameters and the combination that provides the highest score on the training set is used on the testing set.

2.7.1 Analyses for Specific Aim 1

The goal of this aim is to classify TEA PTSD + and TEA PTSD- individuals based on average FA of 53 white matter tracts. A linear SVM classifier was used as the machine learning model. The overall performance of the model was assessed by using a receiver operator curve which plots the sensitivity of the model against the specificity resulting in an accuracy estimate.

Mean accuracies of the models are reported. The ten most discriminating features (tracts) were calculated using feature importance scores.

2.7.2 Analyses for Specific Aim 2

In this aim, symptom cluster scores for reexperiencing, avoidance, numbing, and hyperarousal were examined continuously using elastic net regression because there is no empirically derived cut-off for clinical severity of symptom clusters on the PSS. One model was generated for each symptom cluster resulting in a total of four models. The fit of the final models are reported using the resulting mean square error (MSE) and the R^2 value. High R^2 and low MSE indicate a good fit. Note that R^2 calculations in these models can result in negative values. A negative value occurs when the goodness of fit for the cross-validated data is less than a zero-slope fit.

2.7.3 Analyses for Specific Aim 3

There are a number of machine learning methods that can be applied to a given dataset. Some models may be a more appropriate fit than others, so it is useful to compare the accuracy of multiple algorithms. A previously noted Polyssifier can be used to cross-validate multiple techniques. The accuracy of compatible classifiers and regression models were compared for the models tested in the previous aims. Multilayer perceptron, radial basis function SVM, and random forest models were generated to compare to the linear SVM test in Aim 1. For the regression task described in Aim 2, linear regression, LASSO, and ridge regression were compared. All comparison models were chosen based on their frequency of use in the relevant literature, appropriateness to the sample, and applicability to the research questions (Cui & Gong, 2018; Jollans et al., 2019; Vabalas et al., 2019). Accuracy values were calculated for all classification models. Additionally, the ten most discriminating features were calculated for the

random forest and linear SVM models (feature weights are incalculable for all other classification models). Mean square error and the R^2 scores were calculated for all regression models.

3 RESULTS

3.1 Demographics

Clinical and demographic characteristics of the sample by group are available in Table 3. Further analysis was conducted to identify potential confounds between TEA PTSD+ and TEA PTSD- participants (classified using the CAPS). Chi-squared tests of independence and independent t-tests revealed that groups did not differ based on education, employment, income level, substance abuse history, birth control use, handedness, and age ($p > .05$). Groups also did not significantly differ on the frequency of lifetime trauma exposure (excluding child abuse) as estimated on the TEI. The groups also did not significantly differ based on childhood sexual, physical, or emotional abuse as measured using the CTQ. Groups did differ on current depression symptoms as measured via the BDI ($t = -4.56$, $p < .001$, $d = .84$). Presence of comorbid depression symptoms is common in PTSD (Brewin et al., 2017; Zandvakili et al., 2020). Additionally, the BDI has been previously correlated with PTSD measures (Arbisi et al., 2012; Pupo et al., 2011), and in our sample, the BDI was significantly correlated with scores on the PSS ($p < .05$). Groups did not differ on history of depression treatment or suicide attempt.

Table 3. Demographic Data of Trauma Exposed Adults

	PTSD+ (n=45)	PTSD- (n=91)	<i>p</i>-value
Education			0.07
< 12 years	11.1% (5)	14.3% (13)	
High School/G.E.D.	28.9% (13)	34.1% (31)	
Some college/technical school	37.8% (17)	36.2% (33)	
College/technical school degree	15.6% (7)	16.5% (15)	

Graduate degree	4.4% (2)	3.3% (3)	
Employment (% Employed)	26.6% (12)	38.5% (35)	0.37
Income			0.71
\$0 – 249	13.3% (6)	9.9% (9)	
\$250 – 499	8.9% (4)	8.8% (8)	
\$500 – 999	24.4% (11)	35.2% (32)	
\$1000 – 1999	31.1% (14)	23.1% (21)	
\$2000+	20% (9)	18.7% (17)	
Substance Abuse (% Yes)	9.5% (4)	11.24% (10)	0.77
Birth Control (% Yes)	15.6% (7)	11% (10)	0.61
Handedness (% Left-Handed)	4.4% (2)	6.6% (6)	0.54
Depression treatment	22% (10)	21% (19)	0.83
Suicide Attempt	18% (8)	10% (9)	0.181
	PTSD+	PTSD-	p-value
Age (years)	40.36 (12.85)	39.77 (11.64)	0.78
TEI Lifetime Frequency	2.48 (1.00)	2.36 (1.16)	0.56
CTQ			
Sexual Abuse	9.29 (5.64)	8.29 (5.46)	0.32
Physical Abuse	8.11 (3.45)	8.26 (4.03)	0.84
Physical Neglect	6.18 (1.83)	6.57 (2.89)	0.41
Emotional Abuse	10.00 (5.01)	8.87 (4.12)	0.16
Emotional Neglect	10.04 (4.59)	9.13 (4.92)	0.29
BDI Total Score	19.48 (10.85)	10.23 (11.05)	<.001

3.2 PTSD Assessment

There were 136 participants who completed the CAPS and 166 participants who completed the PSS. Both the PSS and the CAPS were completed by 134 participants. Of those that completed both the CAPS and the PSS, the TEA PTSD+ group had significantly higher scores across all PSS subscales ($p < .05$). Descriptive data of the symptom measures is presented in Table 3.

Table 3. PTSD Symptoms

TEA (n=136)					
% (n)					
CAPS PTSD Present	33.6% (47)				
TEA (n=166)		PTSD+	PTSD-	p-value	d

	M (SD)	(n=45)	(n=89)		
PSS Total	14.14 (12.40)	24.71 (11.54)	8.91 (9.72)	<.001	1.53
PSS Reexperiencing	3.66 (3.72)	6.67 (3.82)	2.11 (2.72)	<.001	1.46
PSS Avoidance	5.55 (5.59)	9.78 (5.35)	3.27 (4.43)	<.001	1.37
PSS Hyperarousal	4.93 (4.18)	8.27 (3.38)	3.53 (3.89)	<.001	1.27
PSS Numbing	2.12 (2.60)	3.98 (2.90)	1.15 (1.99)	<.001	1.21

3.3 Results for Specific Aim 1

Classification accuracy was below chance for the linear SVM at all folds (see Table 4). Mean accuracy was highest at 10-folds (47.11%). Results show somewhat higher training accuracy compared to testing accuracy, which indicates some overfitting (Figure 1). Feature importance was generated for each input feature in the classification model. The ten most discriminating features for the linear SVM with the highest accuracy, the 10-fold model, are shown in Figure 1 and Table 5. Table 5 provides a legend that matches each feature index number with its respective white matter tract. In the 10-fold linear SVM, the right vertical occipital fasciculus and the right uncinate fasciculus contributed the most positive discriminating power to the model. Additionally, the third segment of the right superior longitudinal fasciculus showed some negative discriminating power. All other features did not contribute heavily to the model. However, given that the linear SVM did not discriminate about chance, feature importance should be interpreted with caution.

Table 4. Classification Models

Model	k-fold	Accuracy M (SD)
Linear SVM	5-folds	44.94% (6.74%)
	10-folds	47.11% (7.78%)
	20-folds	42.13% (10.91%)
Radial basis function SVM	5-folds	46.67% (13.3%)
	10-folds	44.67% (15.74%)
	20-folds	47.96% (23.90%)

Multilayer Perceptron	5-folds	46.17% (9.18%)
	10-folds	40.44% (10.62%)
	20-folds	42.88% (23.90%)
Random Forest	5-folds	56.20% (12.10%)
	10-folds	52.72 (15.1%)
	20-folds	58.88% (22.77%)

Figure 1. Performance and Feature Importance of Linear SVM Classification at 10-folds

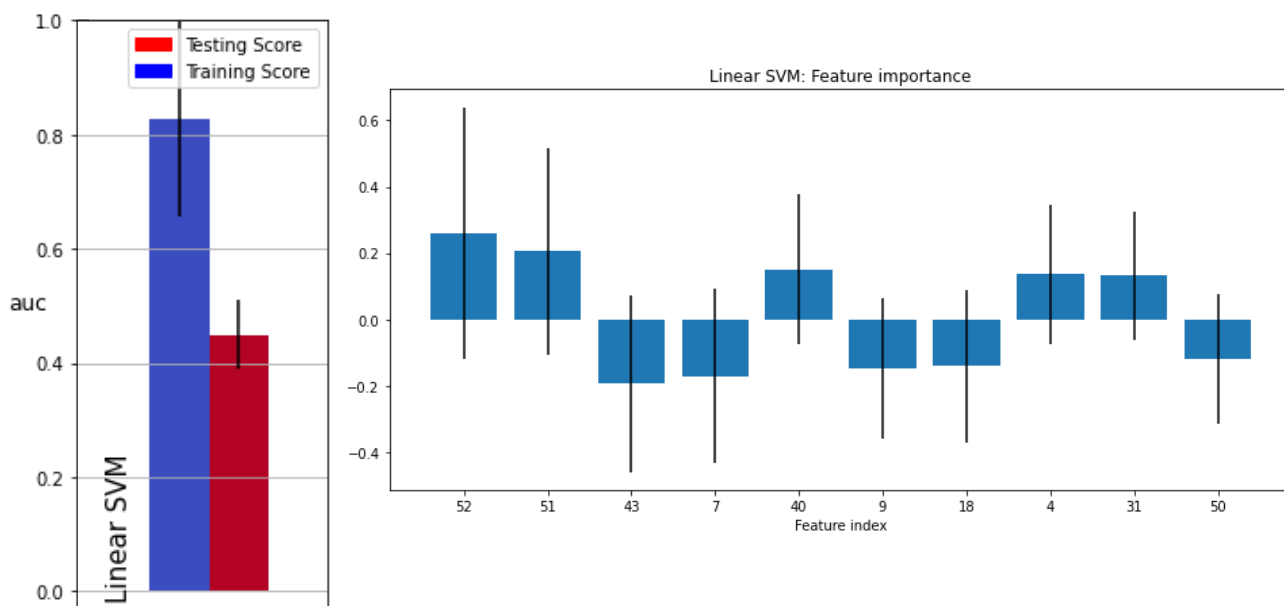


Table 5. Linear SVM Feature Importance Key at 10-folds

ML algorithm	Feature number	White Matter Tract
Linear SVM	52	Right Vertical Occipital Fasciculus
	51	Right Uncinate Fasciculus
	43	Right Superior Longitudinal Fasciculus 3
	7	Right Frontal Parietal Cingulum
	40	Left Superior Longitudinal Fasciculus 2
	9	Left Parahippocampal Parietal Cingulum
	18	Left Corticospinal Tract
	4	Left Frontal Parahippocampal Cingulum
	31	Right Medial Lemniscus
	50	Left Uncinate Fasciculus

3.4 Results for Specific Aim 2

Mean square error and R^2 values suggest poor predictive ability for all elastic net models (see Table 6). The models that examined scores on the numbness subscale of the PSS had the lowest MSE because this scale has the lowest maximum score (max=9). R^2 results suggest that models examining hyperarousal and avoidance scores were the most successful for elastic net. The elastic net models with the greatest R^2 for all subscales were at five-folds. No elastic net models had positive R^2 values. Therefore, feature importance for these models was not reported.

Table 6. Regression Models

Model	<i>k</i>-fold	Mean Square Error M (SD)	R^2 M (SD)
Numbness			
Linear Regression	5-folds	10.23 (2.60)	-0.66 (0.47)
	10-folds	9.52 (4.51)	-0.71(0.73)
	20-folds	9.69 (4.40)	-2.47 (5.27)
Ridge Regression	5-folds	9.80 (2.46)	-0.58 (0.39)
	10-folds	4.02 (0.29)	-0.67 (0.71)
	20-folds	9.47 (4.39)	-2.36 (5.06)
LASSO Regression	5-folds	6.90 (2.05)	-0.08 (0.09)
	10-folds	6.89 (3.22)	-0.20 (0.37)
	20-folds	6.87 (3.88)	-1.04 (2.79)
Elastic Net	5-folds	6.92 (2.04)	- 0.08 (0.82)
	10-folds	6.89 (3.22)	-0.20 (0.37)
	20-folds	6.87 (3.88)	-1.23 (3.08)
Hyperarousal			
Linear Regression	5-folds	31.26 (5.48)	-0.89 (0.30)
	10-folds	26.40 (8.84)	-0.60 (0.32)
	20-folds	26.84 (9.75)	-1.20 (1.64)
Ridge Regression	5-folds	29.78 (5.36)	-0.78 (0.25)
	10-folds	25.61 (8.57)	-0.55 (0.30)
	20-folds	26.14 (9.51)	-1.14 (1.59)
LASSO Regression	5-folds	17.80 (3.37)	-0.06 (0.05)
	10-folds	17.68 (6.27)	-0.04 (0.41)
	20-folds	17.72 (7.54)	-0.33 (0.83)
Elastic Net	5-folds	17.80 (3.37)	-0.06 (0.05)
	10-folds	19.49 (6.98)	-0.16 (0.11)
	20-folds	17.72 (7.54)	-0.54 (1.08)
Avoidance			

Linear Regression	5-folds	54.48 (14.84)	-0.84 (0.31)
	10-folds	49.38 (19.37)	-0.77 (0.43)
	20-folds	50.03 (21.42)	-3.02 (7.26)
Ridge Regression	5-folds	52.11 (13.76)	-0.76 (0.27)
	10-folds	48.14 (18.94)	-0.72 (0.42)
	20-folds	48.89 (20.96)	-2.94 (7.16)
LASSO Regression	5-folds	32.27 (5.39)	-0.01 (0.11)
	10-folds	32.10 (9.57)	-0.16 (0.24)
	20-folds	31.78 (13.09)	-1.69 (5.55)
Elastic Net	5-folds	37.30 (8.04)	-0.01 (0.11)
	10-folds	32.07 (9.55)	-0.16 (0.24)
	20-folds	31.78 (13.09)	-1.88 (5.50)
Reexperiencing			
Linear Regression	5-folds	25.83 (8.88)	-0.98 (0.35)
	10-folds	22.52 (7.13)	-0.82 (0.48)
	20-folds	21.22 (9.27)	-1.42 (1.97)
Ridge Regression	5-folds	24.62 (8.88)	-0.88 (0.34)
	10-folds	13.69 (0.50)	-0.76 (0.46)
	20-folds	20.72 (9.14)	-1.34 (1.86)
LASSO Regression	5-folds	14.21 (3.65)	-0.11 (0.14)
	10-folds	14.03 (4.64)	-0.10 (0.11)
	20-folds	13.99 (5.96)	-0.57 (1.22)
Elastic Net	5-folds	14.21 (3.65)	-0.11 (0.14)
	10-folds	14.17 (4.63)	-0.12 (0.14)
	20-folds	15.67 (7.12)	-0.70 (1.21)

3.5 Results for Specific Aim 3

3.5.1 Comparison of Classification Models

Across classification models, only one model demonstrated accuracy above chance. The random forest model achieved a mean accuracy of 58.88% at 20-folds. However, both random forest and multilayer perceptron models had a training accuracy of 100%. Such a high accuracy suggests overfitting. Therefore, results should be interpreted with caution. Radial basis function SVM demonstrated some underfitting with greater testing accuracy compared to training accuracy. Table 2 and Figure 2 show the accuracy output for all four models at 20-folds.

Feature importance scores for the most discriminating features for the random forest model at 20-folds are plotted in Figure 2. A key for the random forest feature numbers is located in Table 7. In the random forest model, the feature importance for each feature is low, suggesting that there was no feature that contributed heavily towards discriminating between PTSD+ and PTSD- participants. The white matter tract with the greatest feature importance score was the left optic radiation.

Figure 2. Performance of Classifiers and Random Forest Feature Importance at 20-folds

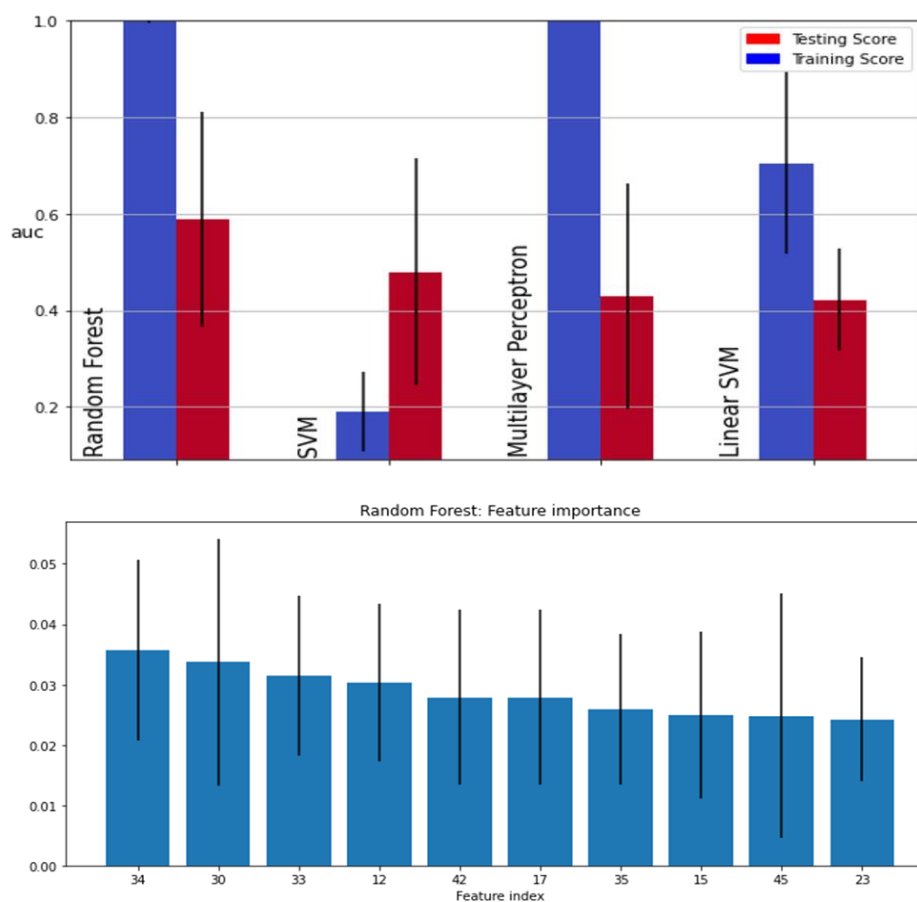


Table 7. Random Forest Feature Importance Key

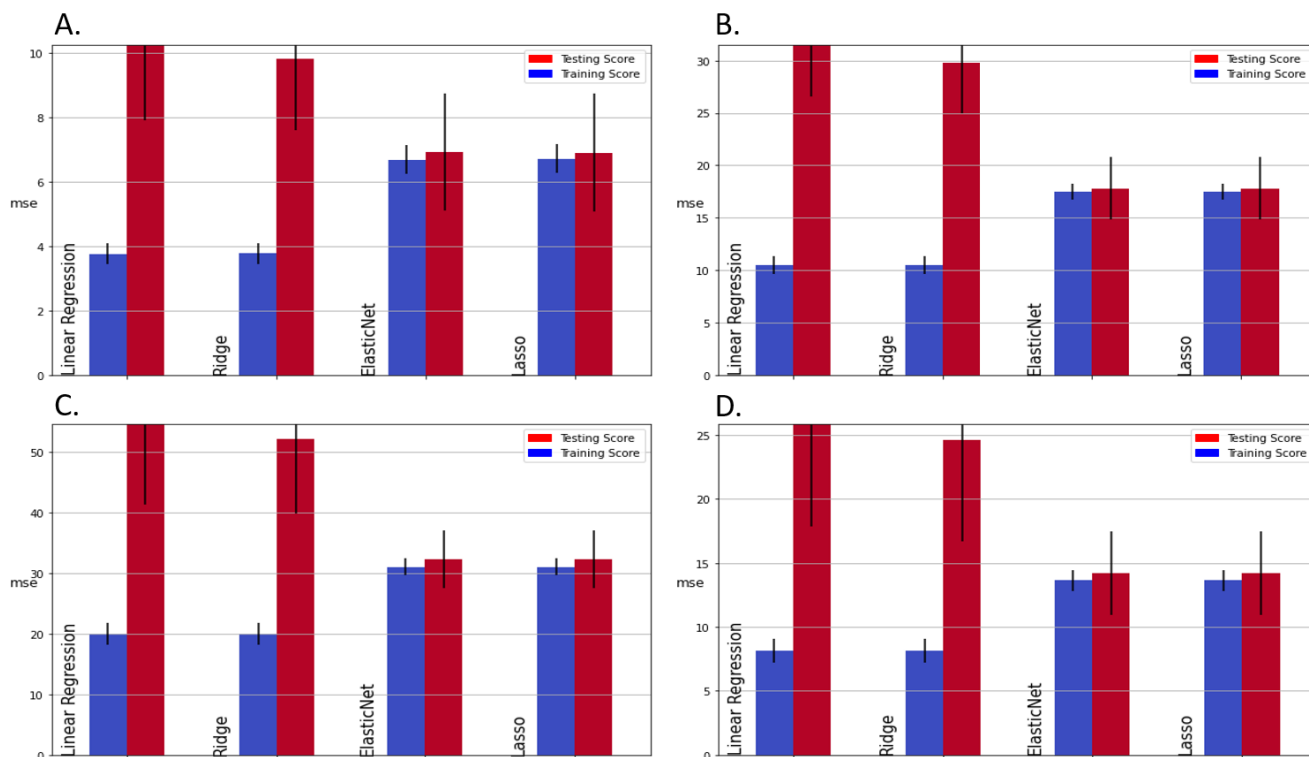
ML algorithm	Feature number	White Matter Tract
Random Forest	34	Left Optic Radiation

30	Left Medial Lemniscus
33	Right Middle Longitudinal Fasciculus
12	Left Parolfactory Cingulum
42	Left Superior Longitudinal Fasciculus 3
17	Tapetum of the Corpus Callosum
35	Right Optic Radiation
15	Forceps Major of the Corpus Callosum
45	Right Anterior Thalamic Radiation
23	Right Fornix

3.5.2 Comparison of Regression Models

For regression models, no model demonstrated a positive R^2 value (see Table 6). Based on R^2 values, models predicting avoidance subscale scores had the best performance, with the LASSO regression model showing the least negative values. As was the case for the elastic net models, LASSO performed the best at 5-folds for the avoidance subscale. LASSO regression was also the most successful at predicting the numbness, hyperarousal, and reexperiencing subscales with the best R^2 outcome at either 5-folds or 10-folds. Figure 3 shows the comparative performance of the regression models for all subscales at 5-folds. Bar graphs are scaled by mean square error to provide a better visual comparison. As demonstrated in Figure 3, linear regression and ridge regression models were both prone to overfitting. Results indicate a beneficial impact of the L1-norm regularization on model fit. Feature importance for compatible models is not reported because R^2 values were all negative.

Figure 3. Regression Model Performance



Note. A.) Numbness B.) Hyperarousal C.) Avoidance D.) Reexperiencing

4 DISCUSSION

4.1 PTSD Diagnostic Classification Based on White Matter Tracts

The purpose of Aim 1 was to examine the accuracy of FA from 53 white matter tracts in discriminating between TEA PTSD- and TEA PTSD+ individuals in a population of Black American women. The initial hypothesis was that results would show excellent predictive accuracy (70%-89%). However, the accuracies for the linear SVM models in this study failed to score above chance. This outcome is somewhat inconsistent with one study; Suo and colleagues (2022) did report above chance accuracy for their SVM using FA values, but their mean accuracy was relatively modest (64.3%). As such, it may be that white matter microstructure has limited classification utility for PTSD samples and that natural statistical variability can account

for differential results. Beyond regular statistical variability, there are several areas in which these studies differ.

The sample used in the current study and in the Suo et al., (2022) study had several unique aspects. The prior study included 153 participants who survived an earthquake in Sichuan. Our study included 136 American participants who were exposed to any DSM-defined criterion A trauma. Greater variability in the trauma type and history of trauma in our sample may have generated more noise in our features which would have masked the signal in our results. Additionally, as noted in the introduction, Black American women are more likely to experience some socially determined events (e.g., adverse childhood events, premature birth, race, and gender-based discrimination) that may significantly impact white matter microstructure development and thus mask PTSD-specific abnormalities. In the same vein, factors such as income level and education history may also have varied – especially given that these samples were recruited from two unique countries. Further, the Suo study was split between 77 persons with PTSD and 76 demographically matched controls. In our sample, approximately one-third ($n=45$) of the participants were diagnosed with PTSD. Therefore, despite similar total sample sizes, the pool of PTSD+ participants in this study was notably smaller, which may have contributed to lower accuracy values.

In addition to characteristics of the study sample, characteristics of the study features may have played a role in findings. PTSD diagnosis was determined in this study using the CAPS. The prior study also relied on the CAPS. However, in addition to a CAPS score ≥ 50 , inclusion criteria included a score ≥ 35 on the PTSD checklist-Civilian Version (Weathers et al., 1994). The addition of a secondary measure may have resulted in a more stringently defined PTSD group, which may have resulted in more predictable patterns of FA. Additionally, this study

included 53 white matter tracts or tract segments, and FA was calculated as the mean value of the entire tract. The prior study included 20 tracts but quantified FA along each tract at 100 equally-spaced points. This method increases the input features to 2,000 (relative to 53 in our study), which can increase accuracy, assuming a limited amount of noise.

A secondary goal of Aim 1 was to examine the features with the most importance in the linear SVM. Although results should be considered exploratory at most, given the low predictive accuracy of the linear SVM models, it appears the hypothesis for this aim was partially met. Regions of the cingulum bundle, uncinate fasciculus, and superior longitudinal fasciculus were among those tracts with the greatest feature weights. These tracts are in line with expectations that white matter tracts associated with the prefrontal and limbic cortex would have the highest feature weights in linear models. The right uncinate fasciculus, which had the second greatest positive feature importance score, connects several regions in the limbic system to the inferior frontal lobe. The differential FA in the uncinate fasciculus between persons with PTSD and trauma-exposed controls has been reported when measured with tract-based spatial statistics and voxel-based analysis (Kunimatsu et al., 2020), although this is not a consistent finding (Dennis et al., 2019).

Notably, the right uncinate had a positive feature score, but the left uncinate had a negative score. In our classification model, negative feature importance scores predict the PTSD-group, whereas positive scores predict the PTSD+ group. A similar pattern was seen for the superior longitudinal fasciculus and the cingulum. In this SVM model, feature importance scores were calculated based on coefficient weights, indicating a potential pattern wherein right hemisphere tracts were more related to PTSD and left hemisphere tracts were more related to controls. Given that the only prior diagnostic SVM model from Suo and colleagues (2022) does

not report feature importance scores, it is unclear to what extent this finding corresponds to the machine learning literature. In the broader literature, there has been some suggestion that structural changes in the right hemisphere are related to stress modulation in PTSD (Ocklenburg et al., 2016; Schore, 2002; Wolf et al., 2016). However, right-hemisphere-specific differences in white matter have not been consistently reported (Fani et al., 2012; Olson et al., 2017). Future work may wish to investigate hemispheric differences further.

Contrary to expectations, the tract with the greatest feature importance score was the right vertical occipital fasciculus. The vertical occipital fasciculus is located in the occipital region and runs perpendicular to the inferior fronto-occipital and the inferior longitudinal fasciculi. The vertical occipital fasciculus is the only white matter tract that connects the dorsal and ventral visual streams and is thought to be involved in the integration of these streams (Yeatman et al., 2014). The positive sign of the vertical occipital fasciculus feature score indicates that this tract predicts the PTSD+ group rather than the PTSD- group. Interestingly, a meta-analysis from Ju and colleagues (2020) found a similar association between vision-related white matter microstructure and PTSD, although their primary tract of interest was the inferior fronto-occipital fasciculus. In their interpretation, the authors tentatively suggest that abnormal white matter microstructure in these regions may be related to reexperiencing and hyperarousal symptoms. A cursory review of the available feature importance scores for our regression models examining reexperiencing and hyperarousal symptoms did not suggest the vertical occipital fasciculus meaningfully contributed to symptom prediction in this study. However, given the weak accuracy scores in this sample, further exploration of the vertical occipital fasciculus within a more accurate model is merited.

4.2 PTSD Symptom Severity Prediction Based on White Matter Tracts

The purpose of Aim 2 was to examine the accuracy of elastic net models in predicting symptoms scores on the PSS. Four models were generated, each of which used participants' total scores on a symptom cluster (numbing, hyperarousal, reexperiencing, and avoidance) as the dependent variable. The initial hypothesis was that models would show moderate to good results ($R^2=0.5-0.6$) for all symptom clusters. However, the R^2 values for all models were negative, which supports the null hypothesis. When comparing the relative fit of the model for each symptom cluster, R^2 results indicate that hyperarousal and avoidance scores were the most successful. These findings were overall consistent with the findings from Aim 1, which suggests that white matter tracts had a weak signal for PTSD and its symptoms in this sample. This study is the first to examine the predictive utility of white matter tracts on PTSD symptom severity scores using a machine learning approach. As such, it is not possible to compare findings to those reported in other studies. However, Zandvakili and colleagues (2020) have examined a similar question using fMRI.

There are some key methodological differences between this paper and Zandvakili et al. (2020) that should be noted prior to any comparisons. The prior paper consisted of 50 participants, all of whom met DSM-5 criteria for PTSD. In contrast, our sample included participants with and without PTSD. It could be that having a mixed sample resulted in a weaker fit. However, a mixed sample can contribute by increasing the range in scores. Additionally, our sample of PTSD+ participants was similarly sized ($n=45$). Race and/or ethnicity of the Zandvakili et al. sample was not reported, and the majority of the sample was male. Differences in sex and possibly race may contribute to different findings. Another difference is that the Zandvakili et al. paper defined symptom clusters using the PTSD Checklist for DSM-5. In

comparison, the current study utilized a different symptom checklist based on older symptom clustering, which may impact model fit (Rangaprakash et al., 2017). Further, the input features for the previous study consisted of functionally connected regions of interest that were narrowed from 4950 dimensions of the connectivity matrix to 39 using principal component analysis. Functional connectivity may have different relationships with PTSD symptoms than white matter. Further, dimensionality reduction using principal component analysis is known to improve machine learning algorithms, although this improvement comes at the cost of reduced interpretability (Nielsen et al., 2020).

With regard to their findings, Zandvakili and colleagues (2020) report a mixed utility of functional connectivity data in predicting intrusion, avoidance, cognition/mood, and arousal/reactivity symptoms. Similar to our findings, results indicate a relatively better fit for avoidance scores. However, the authors found the strongest fit for intrusion symptoms (which are referred to as reexperiencing symptoms in this paper), which was not consistent with our results. Also, somewhat contrary to our findings, the authors reported a relatively weaker fit for arousal symptoms. In comparison, hyperarousal had a relatively better fit than other clusters in our sample – although the fit was still weak overall. The majority of models in the previous study had positive R^2 values, but one model was negative – similar to our models. The model for the cognition/mood cluster in the prior study had an R^2 value of -0.01 . In a similar vein, the results of our study show the weakest fit for the numbness cluster, which is most similar to the cognition/mood cluster used by Zandvakili and colleagues (2020). Both the numbness and the cognition/mood cluster focus on evaluating symptoms that have significant crossover with other DSM diagnoses (Brewin et al., 2017), which may explain the weakness of fit for these clusters.

4.3 Comparison of Machine Learning Models

The goal of this exploratory aim was to compare the fit of common machine learning models for each of the previously described specific aims. Multilayer perceptron, radial basis function SVM, and random forest were compared to linear SVM. Linear regression, LASSO, and ridge regression were compared to elastic net regression.

4.3.1 Relative Performance of Classification Models

Our hypothesis for the comparison of classification models was that multilayer perceptron would perform the best of all classifiers. We also anticipated that random forest and linear SVM would have good accuracy and perform similarly. We predicted that radial basis function SVM would perform the worst due to overfitting. Lastly, we anticipated that all models would perform above chance. Results only partially met our hypotheses. Contrary to our expectations, multilayer perceptron did not perform the best. Rather, it performed the worst across all folds. Results suggest that overfitting, which can be a problem with small samples, was a consistent issue with all algorithms. Multilayer perceptron can be sensitive to overfitting, and it appears that has been the case in our analyses (as shown in Figure 2). Overfitting also appeared to occur for the random forest, although the disparity between testing and training was not as large. Unlike our expectations, radial basis function SVM did not have difficulty with overfitting, but rather results suggest some underfitting with worse training accuracy compared to testing accuracy. An inconsistent pattern with fit suggests a high amount of noise among our features, which reinforces findings from our first aim that a large proportion of the white matter tracts are not contributing to the model. Additionally, performance improved for both multilayer perceptron and radial basis function SVM as folds increased. Folds were not increased above 20 because of the computational intensity, but it may be that the samples randomly selected in the

cross-validation process contributed to unusual fit properties, and that fit may improve at even higher folds.

Somewhat consistent with our hypotheses, random forest did perform similarly to linear SVM across folds. However, unlike we anticipated, results show that random forest had the best performance and was the only model to achieve accuracy above chance. At 20-folds, the random forest model achieved the highest mean accuracy of 58.88%. Feature importance scores for random forest at 20-fold indicate that there were no features that highly contributed to the model. However, it is notable that of those features, the tract with the greatest importance score was the left optic radiation. The optic radiations connect the lateral geniculate nucleus of the thalamus to the primary visual area, V1, and convey both dorsal and ventral visual information from the optic nerve. It is interesting that this result is similar to the feature importance results for the linear SVM model as the tract with the greatest score, the right vertical occipital fasciculus, also conveys dorsal and ventral visual information. Results may suggest the combined visual streams as being more implicated in PTSD diagnosis than previously thought. However, the optic radiation is a challenging tract to isolate using tractography approaches, so further work may be necessary to assure that this tract is related to PTSD (Schurr et al., 2018).

4.3.2 Relative Performance of Regression Model

Our hypothesis for the comparison of regression models was that most models would perform similarly. We anticipated that linear regression would be the worst due to overfitting. We also hypothesized that models would show moderate to good results ($R^2=0.5-0.6$) for all symptom clusters. Contrary to our hypotheses, there was a discrepancy between models wherein linear regression and ridge regression performed similarly, and LASSO and elastic net performed similarly across all symptom clusters. No models reached an R^2 value above zero. These results

were consistent with our results from previous aims, which suggests that there is minimal signal from these white matter tract metrics. Additionally, this strengthens the conclusion by demonstrating that our null result in Aim 2 was not due to our choice of model.

LASSO and elastic net models performed the best overall, and these results were fairly consistent across all folds and clusters. Given the relatively better fit of LASSO and elastic net models, these results suggest that the L1-norm regularization, a penalty function that is included in both these models, had a positive impact on the fit of the model. The L1-norm regularization function restricts the number of features in the model to reduce noise and model complexity. The positive impact of L1-norm regularization supports the conclusion that there is a high amount of noise in this sample. In other words, several of the white matter tracts are not contributing to the predictive ability of the sample. This conclusion is consistent with broad theories of the neuropathology of PTSD that suggests localized differences in white matter microstructure occur as opposed to whole-brain abnormalities.

We anticipated the weakest performance for the linear regression model, and our results support that hypothesis. Linear regression does not contain any penalty functions. Penalty functions typically optimize a model by encouraging a more linear shape. Therefore, most models benefit from these penalty functions unless a model naturally has a perfectly linear shape. As such, it is unlikely that our model is linear in nature. Somewhat contrary to expectations, ridge regression also performed relatively weak in comparison to other models. These results suggest that there was also a minimal positive effect of the L2-norm regularization penalty function. The L2-norm regularization works to create a more even distribution and minimize the impact of outliers. It is likely that our choice to normalize the distribution of the input features prior to running our models using StandardScaler minimized any impact of the L2-norm

regularization. Further, ridge regression works best when all inputted features are considered informative to the model, which, as noted previously, does not appear to be the case for this sample.

4.4 Strengths and Weaknesses

Findings for the proposed study should be considered within the context of its weaknesses. While this study is appropriately compared to other machine learning clinical neuroimaging samples, machine learning typically performs best at infinitely-sized datasets. Prior work in neuroimaging shows that accuracy generally improves with sample size irrespective of the chosen algorithm. In this sense, “bigger is always better” when it comes to machine learning samples. Therefore, for clinical neuroimaging samples that are naturally modest, a limited (as opposed to infinite) sample size is always a weakness. Nevertheless, as stated previously, our sample size is similarly-sized compared to prior studies and therefore is consistent with previous work. Secondly, this study uses cross-sectional data, which removes any ability to make causative statements about findings. Specifically, for reported feature weights, it is unclear if the predictive value for the most informative tracts extends before or after a TEA’s PTSD diagnosis. Lastly, as previously noted, prior work has found that cognitive mechanisms, such as emotional regulation, may function uniquely in Black American women with PTSD. This study focuses on PTSD presence and symptom severity but does not investigate cognitive mechanisms. Future work may wish to focus on cognitive mechanisms and white matter microstructure using a similar machine learning paradigm.

The proposed study carefully considers the existing PTSD literature and seeks to strengthen it at several key points. First, it uses advanced statistical methods, namely machine learning algorithms, which are statistically ideal for examining biological datasets with a large

number of features. Machine learning is a logical next step for neuroanatomical questions about trauma exposure and PTSD and has the added bonus of evaluating the clinical utility of features for individualized predictions. The weaker performance of linear regression in Aim 3 supports the idea that more advanced statistical methods than simple regressions can increase model performance. Additionally, Polyssifier is a publicly available tool that is easy to use. Across our review of the literature, model methods varied widely, which reduces the ability to make comparisons between studies. Regular use of a simple tool like Polyssifier in research may strengthen conclusions about high signal predictors by allowing for comparison to other Polyssifier-generated models. The study also uses neuroHarmonize, a validated approach to harmonize large imaging datasets by removing site-specific effects while preserving biological heterogeneity. Of the four preliminary studies that use DTI data to make diagnostic or prognostic predictions about PTSD, this study further expands the existing literature by focusing on Black American women. Black American women are historically underrepresented in research. While it is not entirely clear due to the lack of reported race or ethnic demographic data in the literature, it appears that this group has been underrepresented in prior PTSD and machine learning research. Our results show some differences compared to prior research that may be due to differences in samples. As such, results support the need for greater diversity in research samples and indicate that prior literature may not be fully generalizable.

4.5 Conclusions & Future Directions

This study also asks new questions by investigating symptoms severity and white matter using machine learning regression. While there are some weaknesses to the research, the proposed study takes every available opportunity to improve upon and further existing work and lays a foundation for future directions.

Firstly, results across models suggest that dimensionality reduction for our features would improve model performance. We've completed an exploratory analysis in which features were reduced from the original tracts to targeted tracts of interest were limited to the cingulum, uncinate fasciculus, superior longitudinal fasciculus, and corpus callosum with promising results. Principal component analysis may aid in feature refinement, as seen in other work (Zandvakili et al., 2020). Secondly, this study found some evidence to suggest that visual tracts, particularly tracts that convey both dorsal and ventral stream information, may be more implicated in PTSD than previously thought. Third, this study only examined the average FA values of each tract. However, other work has shown greater accuracy with other DTI scalars, such as MD, compared to FA (Li et al., 2016; Suo et al., 2022). Additionally, other DTI metrics, such as the span, surface area, or volume of tracts, have yet to be investigated and may contribute to model success. Lastly, some machine learning studies have found improved accuracy with multimodal models (Ramos-Lima et al., 2020). For example, Zhang and colleagues (2016) combined gray matter volume, amplitude of low-frequency fluctuations, and regional homogeneity and reported improved accuracy with multiple feature types over a single feature type. Medical data such as acute endocrine levels have also been reported to contribute to high accuracy models when used in combination with questionnaires (Ramos-Lima et al., 2022; Schultebrucks et al., 2021). Schultebrucks and colleagues reported average accuracies $\geq 78\%$ when using a combination of bloodwork, computerized neurocognitive testing, and symptom self-reports. It may be that white matter microstructure metrics may modestly contribute to greater accuracy along with a wider range of imaging, biomedical, and clinical features.

In conclusion, PTSD impacts millions annually. This study sought to move the field forward by investigating white matter microstructure through a machine learning approach. It

innovates on previous work by using more advanced methods and a unique sample. Our results tentatively suggest minimal signal from white matter tracts in predicting PTSD or symptom cluster scores. Future work may provide greater clarity about this study's findings.

REFERENCES

- Akiki, T. J., Averill, C. L., & Abdallah, C. G. (2017). A Network-Based Neurobiological Model of PTSD: Evidence From Structural and Functional Neuroimaging Studies. *Current Psychiatry Reports, 19*(11). <https://doi.org/10.1007/s11920-017-0840-4>
- Arbisi, P. A., Kaler, M. E., Kehle-Forbes, S. M., Erbes, C. R., Polusny, M. A., & Thuras, P. (2012, Dec). The predictive validity of the PTSD Checklist in a nonclinical sample of combat-exposed National Guard troops. *Psychol Assess, 24*(4), 1034-1040. <https://doi.org/10.1037/a0028014>
- Asnaani, A., & Hall-Clark, B. (2017, Apr). Recent developments in understanding ethnocultural and race differences in trauma exposure and PTSD. *Curr Opin Psychol, 14*, 96-101. <https://doi.org/10.1016/j.copsyc.2016.12.005>
- Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and-II in Psychiatric Outpatients. *Journal of Personality Assessment, 67*(3), 588-597. https://doi.org/10.1207/s15327752jpa6703_13
- Bernstein, D. P., Fink, L. (1998). *Childhood Trauma Questionnaire: A retrospective self-report manual*. The Psychological Corporation.
- Bernstein, D. P., Stein, J. A., Newcomb, M. D., Walker, E., Pogge, D., Ahluvalia, T., Stokes, J., Handelsman, L., Medrano, M., Desmond, D., & Zule, W. (2003, Feb). Development and

- validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse Negl*, 27(2), 169-190. [https://doi.org/10.1016/s0145-2134\(02\)00541-0](https://doi.org/10.1016/s0145-2134(02)00541-0)
- Bovin, M. J., Marx, B. P., Weathers, F. W., Gallagher, M. W., Rodriguez, P., Schnurr, P. P., & Keane, T. M. (2016, Nov). Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (PCL-5) in veterans. *Psychol Assess*, 28(11), 1379-1391. <https://doi.org/10.1037/pas0000254>
- Bowleg, L. (2012, Jul). The problem with the phrase women and minorities: intersectionality-an important theoretical framework for public health. *Am J Public Health*, 102(7), 1267-1273. <https://doi.org/10.2105/ajph.2012.300750>
- Brewin, C. R., Cloitre, M., Hyland, P., Shevlin, M., Maercker, A., Bryant, R. A., Humayun, A., Jones, L. M., Kagee, A., Rousseau, C., Somasundaram, D., Suzuki, Y., Wessely, S., van Ommeren, M., & Reed, G. M. (2017, Dec). A review of current evidence regarding the ICD-11 proposals for diagnosing PTSD and complex PTSD. *Clin Psychol Rev*, 58, 1-15. <https://doi.org/10.1016/j.cpr.2017.09.001>
- Bubb, E. J., Metzler-Baddeley, C., & Aggleton, J. P. (2018, 2018/09/01/). The cingulum bundle: Anatomy, function, and dysfunction. *Neuroscience & Biobehavioral Reviews*, 92, 104-127. <https://doi.org/https://doi.org/10.1016/j.neubiorev.2018.05.008>

- Christiansen, D. M., & Berke, E. T. (2020). Gender- and Sex-Based Contributors to Sex Differences in PTSD. *Current Psychiatry Reports*, 22(4). <https://doi.org/10.1007/s11920-020-1140-y>
- Coffey, S. F., Dansky, B. S., Falsetti, S. A., Saladin, M. E., & Brady, K. T. (1998). Screening for PTSD in a substance abuse sample: Psychometric properties of a modified version of the PTSD symptom scale self-report. *Journal of Traumatic Stress*, 11(2), 393-399. <https://doi.org/https://doi.org/10.1023/A:1024467507565>
- Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, 8, 139-168. <https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8>
- Cui, Z., & Gong, G. (2018). The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *NeuroImage*, 178, 622-637. <https://doi.org/10.1016/j.neuroimage.2018.06.001>
- Dennis, E. L., Disner, S. G., Fani, N., Salminen, L. E., Logue, M., Clarke, E. K., Haswell, C. C., Averill, C. L., Baugh, L. A., Bomyea, J., Bruce, S. E., Cha, J., Choi, K., Davenport, N. D., Densmore, M., Du Plessis, S., Forster, G. L., Frijling, J. L., Gonenc, A., Gruber, S., Grupe, D. W., Guenette, J. P., Hayes, J., Hofmann, D., Ipser, J., Jovanovic, T., Kelly, S., Kennis, M., Kinzel, P., Koch, S. B. J., Koerte, I., Koopowitz, S., Korgaonkar, M.,

Krystal, J., Lebois, L. A. M., Li, G., Magnotta, V. A., Manthey, A., May, G. J., Menefee, D. S., Nawijn, L., Nelson, S. M., Neufeld, R. W. J., Nitschke, J. B., O'Doherty, D., Peverill, M., Ressler, K. J., Roos, A., Sheridan, M. A., Sierk, A., Simmons, A., Simons, R. M., Simons, J. S., Stevens, J., Suarez-Jimenez, B., Sullivan, D. R., Théberge, J., Tran, J. K., Van Den Heuvel, L., Van Der Werff, S. J. A., Van Rooij, S. J. H., Van Zuiden, M., Velez, C., Verfaellie, M., Vermeiren, R. R. J. M., Wade, B. S. C., Wager, T., Walter, H., Winternitz, S., Wolff, J., York, G., Zhu, Y., Zhu, X., Abdallah, C. G., Bryant, R., Daniels, J. K., Davidson, R. J., Fercho, K. A., Franz, C., Geuze, E., Gordon, E. M., Kaufman, M. L., Kremen, W. S., Lagopoulos, J., Lanius, R. A., Lyons, M. J., McCauley, S. R., McGlinchey, R., McLaughlin, K. A., Milberg, W., Neria, Y., Olf, M., Seedat, S., Shenton, M., Sponheim, S. R., Stein, D. J., Stein, M. B., Straube, T., Tate, D. F., Van Der Wee, N. J. A., Veltman, D. J., Wang, L., Wilde, E. A., Thompson, P. M., Kochunov, P., Jahanshad, N., & Morey, R. A. (2019). Altered white matter microstructural organization in posttraumatic stress disorder across 3047 adults: results from the PGC-ENIGMA PTSD consortium. *Molecular Psychiatry*. <https://doi.org/10.1038/s41380-019-0631-x>

Dietterich, T. (1995). Overfitting and undercomputing in machine learning. *ACM computing surveys (CSUR)*, 27(3), 326-327.

Dufford, A. J., Evans, G. W., Dmitrieva, J., Swain, J. E., Liberzon, I., & Kim, P. (2020).

Prospective associations, longitudinal patterns of childhood socioeconomic status, and white matter organization in adulthood. *Human Brain Mapping*, 41(13), 3580-3593.

<https://doi.org/10.1002/hbm.25031>

Essed, P. (1991). *Understanding everyday racism: An interdisciplinary theory*. Sage Publications, Inc.

Falsetti, S. A., Resnick, H. S., Resick, P. A., & Kilpatrick, D. G. (1993). The modified PTSD symptom scale: a brief self-report measure of posttraumatic stress disorder. *The Behavior Therapist*.

Fani, N., Harnett, N. G., Bradley, B., Mekawi, Y., Powers, A., Stevens, J. S., Ressler, K. J., & Carter, S. E. (2022, 2022/02/01/). Racial Discrimination and White Matter Microstructure in Trauma-Exposed Black Women. *Biological Psychiatry*, *91*(3), 254-261.
<https://doi.org/https://doi.org/10.1016/j.biopsych.2021.08.011>

Fani, N., King, T. Z., Jovanovic, T., Glover, E. M., Bradley, B., Choi, K., Ely, T., Gutman, D. A., & Ressler, K. J. (2012). White Matter Integrity in Highly Traumatized Adults With and Without Post-Traumatic Stress Disorder. *Neuropsychopharmacology*, *37*(12), 2740-2746. <https://doi.org/10.1038/npp.2012.146>

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (1995). The structured clinical interview for DSM-III-R personality disorders (SCID-II). Part I: Description. *Journal of Personality disorders*, *9*(2), 83-91.

Foa, E. B., McLean, C. P., Zang, Y., Zhong, J., Rauch, S., Porter, K., Knowles, K., Powers, M. B., & Kauffman, B. Y. (2016). Psychometric properties of the Posttraumatic Stress Disorder Symptom Scale Interview for DSM–5 (PSSI–5). *Psychological Assessment*, 28(10), 1159-1165. <https://doi.org/10.1037/pas0000259>

10.1037/pas0000259.supp (Supplemental)

Foa, E. B., & Tolin, D. F. (2000). Comparison of the PTSD symptom scale–interview version and the clinician-administered PTSD scale. *Journal of Traumatic Stress*, 13(2), 181-191. <https://doi.org/https://doi.org/10.1023/A:1007781909213>

Fortin, J.-P., Parker, D., Tunç, B., Watanabe, T., Elliott, M. A., Ruparel, K., Roalf, D. R., Satterthwaite, T. D., Gur, R. C., Gur, R. E., Schultz, R. T., Verma, R., & Shinohara, R. T. (2017, 2017/11/01/). Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*, 161, 149-170. <https://doi.org/https://doi.org/10.1016/j.neuroimage.2017.08.047>

Gareth, J., Daniela, W., Trevor, H., & Robert, T. (2013). *An introduction to statistical learning: with applications in R*. Springer.

Gillespie, C. F., Bradley, B., Mercer, K., Smith, A. K., Conneely, K., Gapen, M., Weiss, T., Schwartz, A. C., Cubells, J. F., & Ressler, K. J. (2009). Trauma exposure and stress-related disorders in inner city primary care patients. *General Hospital Psychiatry*, 31(6), 505-514. <https://doi.org/10.1016/j.genhosppsych.2009.05.003>

Gluck, R. L., Hartzell, G. E., Dixon, H. D., Michopoulos, V., Powers, A., Stevens, J. S., Fani, N., Carter, S., Schwartz, A. C., Jovanovic, T., Ressler, K. J., Bradley, B., & Gillespie, C. F. (2021, 2021/05/15). Trauma exposure and stress-related disorders in a large, urban, predominantly African-American, female sample. *Archives of Women's Mental Health*. <https://doi.org/10.1007/s00737-021-01141-4>

Jakob, J. M. D., Lamp, K., Rauch, S. A. M., Smith, E. R., & Buchholz, K. R. (2017). The Impact of Trauma Type or Number of Traumatic Events on PTSD Diagnosis and Symptom Severity in Treatment Seeking Veterans. *The Journal of Nervous and Mental Disease*, 205(2). https://journals.lww.com/jonmd/Fulltext/2017/02000/The_Impact_of_Trauma_Type_or_Number_of_Traumatic.3.aspx

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002, Oct). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2), 825-841. [https://doi.org/10.1016/s1053-8119\(02\)91132-8](https://doi.org/10.1016/s1053-8119(02)91132-8)

Jenkinson, M., Pechaud, M., & Smith, S. (2005). *BET2: MR-based estimation of brain, skull and scalp surfaces* Eleventh Annual Meeting of the Organization for Human Brain Mapping,

- Jenkinson, M., & Smith, S. (2001, Jun). A global optimisation method for robust affine registration of brain images. *Med Image Anal*, 5(2), 143-156.
[https://doi.org/10.1016/s1361-8415\(01\)00036-6](https://doi.org/10.1016/s1361-8415(01)00036-6)
- Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 8(1), 118-127.
<https://doi.org/10.1093/biostatistics/kxj037>
- Jollans, L., Boyle, R., Artiges, E., Banaschewski, T., Desrivières, S., Grigis, A., Martinot, J.-L., Paus, T., Smolka, M. N., Walter, H., Schumann, G., Garavan, H., & Whelan, R. (2019). Quantifying performance of machine learning methods for neuroimaging data. *NeuroImage*, 199, 351-365. <https://doi.org/10.1016/j.neuroimage.2019.05.082>
- Ju, Y., Ou, W., Su, J., Averill, C. L., Liu, J., Wang, M., Wang, Z., Zhang, Y., Liu, B., Li, L., & Abdallah, C. G. (2020). White matter microstructural alterations in posttraumatic stress disorder: An ROI and whole-brain based meta-analysis. *Journal of Affective Disorders*, 266, 655-670. <https://doi.org/10.1016/j.jad.2020.01.047>
- Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Benjet, C., Bromet, E. J., Cardoso, G., Degenhardt, L., De Girolamo, G., Dinolova, R. V., Ferry, F., Florescu, S., Gureje, O., Haro, J. M., Huang, Y., Karam, E. G., Kawakami, N., Lee, S., Lepine, J.-P., Levinson, D., Navarro-Mateu, F., Pennell, B.-E., Piazza, M., Posada-Villa, J., Scott, K. M., Stein, D. J., Ten Have, M., Torres, Y., Viana, M. C., Petukhova, M. V., Sampson, N. A.,

- Zaslavsky, A. M., & Koenen, K. C. (2017). Trauma and PTSD in the WHO World Mental Health Surveys. *European Journal of Psychotraumatology*, 8(sup5), 1353383. <https://doi.org/10.1080/20008198.2017.1353383>
- Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, 13, 8-17. <https://doi.org/10.1016/j.csbj.2014.11.005>
- Kuceyeski, A., Maruta, J., Niogi, S. N., Ghajar, J., & Raj, A. (2011). The generation and validation of white matter connectivity importance maps. *NeuroImage*, 58(1), 109-121. <https://doi.org/10.1016/j.neuroimage.2011.05.087>
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26). Springer.
- Kunimatsu, A., Yasaka, K., Akai, H., Kunimatsu, N., & Abe, O. (2020). MRI findings in posttraumatic stress disorder. *Journal of Magnetic Resonance Imaging*, 52(2), 380-396. <https://doi.org/10.1002/jmri.26929>
- Li, L., Sun, G., Liu, K., Li, M., Li, B., Qian, S.-W., & Yu, L.-L. (2016). White Matter Changes in Posttraumatic Stress Disorder Following Mild Traumatic Brain Injury: A Prospective Longitudinal Diffusion Tensor Imaging Study. *Chinese Medical Journal*, 129(9), 1091-1099. <https://doi.org/10.4103/0366-6999.180518>

Liu, H., Petukhova, M. V., Sampson, N. A., Aguilar-Gaxiola, S., Alonso, J., Andrade, L. H., Bromet, E. J., De Girolamo, G., Haro, J. M., Hinkov, H., Kawakami, N., Koenen, K. C., Kovess-Masfety, V., Lee, S., Medina-Mora, M. E., Navarro-Mateu, F., O'Neill, S., Piazza, M., Posada-Villa, J., Scott, K. M., Shahly, V., Stein, D. J., Ten Have, M., Torres, Y., Gureje, O., Zaslavsky, A. M., & Kessler, R. C. (2017). Association of DSM-IV Posttraumatic Stress Disorder With Traumatic Experience Type and History in the World Health Organization World Mental Health Surveys. *JAMA Psychiatry*, *74*(3), 270.
<https://doi.org/10.1001/jamapsychiatry.2016.3783>

Liu, Y. H. (2017). *Python Machine Learning By Example*. Packt Publishing Ltd.

Mekawi, Y., Carter, S., Brown, B., Martinez de Andino, A., Fani, N., Michopoulos, V., & Powers, A. (2021, 2021/03/15). Interpersonal Trauma and Posttraumatic Stress Disorder among Black Women: Does Racial Discrimination Matter? *Journal of Trauma & Dissociation*, *22*(2), 154-169. <https://doi.org/10.1080/15299732.2020.1869098>

Mekawi, Y., Murphy, L., Munoz, A., Briscione, M., Tone, E. B., Norrholm, S. D., Jovanovic, T., Bradley, B., & Powers, A. (2020, 2020/02/01/). The role of negative affect in the association between attention bias to threat and posttraumatic stress: An eye-tracking study. *Psychiatry Research*, *284*, 112674.
<https://doi.org/https://doi.org/10.1016/j.psychres.2019.112674>

- Mekawi, Y., Watson-Singleton, N. N., Dixon, H. D., Fani, N., Michopoulos, V., & Powers, A. (2021). Validation of the difficulties with emotion regulation scale in a sample of trauma-exposed Black women. *Journal of Clinical Psychology, 77*(3), 587-606.
<https://doi.org/10.1002/jclp.23036>
- Milanak, M. E., Zuromski, K. L., Cero, I., Wilkerson, A. K., Resnick, H. S., & Kilpatrick, D. G. (2019). Traumatic Event Exposure, Posttraumatic Stress Disorder, and Sleep Disturbances in a National Sample of U.S. Adults. *Journal of Traumatic Stress, 32*(1), 14-22. <https://doi.org/10.1002/jts.22360>
- Moody, A. T., & Lewis, J. A. (2019). Gendered Racial Microaggressions and Traumatic Stress Symptoms Among Black Women. *Psychology of Women Quarterly, 43*(2), 201-214.
<https://doi.org/10.1177/0361684319828288>
- Nicholson, A. A., Densmore, M., McKinnon, M. C., Neufeld, R. W. J., Frewen, P. A., Théberge, J., Jetly, R., Richardson, J. D., & Lanius, R. A. (2019). Machine learning multivariate pattern analysis predicts classification of posttraumatic stress disorder and its dissociative subtype: a multimodal neuroimaging approach. *Psychological Medicine, 49*(12), 2049-2059. <https://doi.org/10.1017/S0033291718002866>
- Nicholson, A. A., Harricharan, S., Densmore, M., Neufeld, R. W. J., Ros, T., McKinnon, M. C., Frewen, P. A., Theberge, J., Jetly, R., Pedlar, D., & Lanius, R. A. (2020). Classifying heterogeneous presentations of PTSD via the default mode, central executive, and

saliency networks with machine learning [Article]. *Neuroimage-Clinical*, 27, 15, Article 102262. <https://doi.org/10.1016/j.nicl.2020.102262>

Nielsen, A. N., Barch, D. M., Petersen, S. E., Schlaggar, B. L., & Greene, D. J. (2020). Machine Learning With Neuroimaging: Evaluating Its Applications in Psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 5(8), 791-798. <https://doi.org/10.1016/j.bpsc.2019.11.007>

Ocklenburg, S., Korte, S. M., Peterburs, J., Wolf, O. T., & Güntürkün, O. (2016, 2016/10/01/). Stress and laterality – The comparative perspective. *Physiology & Behavior*, 164, 321-329. <https://doi.org/https://doi.org/10.1016/j.physbeh.2016.06.020>

Ogle, C. M., Rubin, D. C., & Siegler, I. C. (2013). The impact of the developmental timing of trauma exposure on PTSD symptoms and psychosocial functioning among older adults. *Developmental Psychology*, 49(11), 2191-2200. <https://doi.org/10.1037/a0031985>

Okeke, O., Elbasheir, A., Carter, S., Powers, A., Mekawi, Y., Gillespie, C. F., Schwartz, A., Bradley, B., & Fani, N. (2022, May 18). Indirect Effects of Racial Discrimination on Health Outcomes Through Prefrontal Cortical White Matter Integrity. *Biol Psychiatry Cogn Neurosci Neuroimaging*. <https://doi.org/10.1016/j.bpsc.2022.05.004>

Olson, E. A., Cui, J., Fukunaga, R., Nickerson, L. D., Rauch, S. L., & Rosso, I. M. (2017). Disruption of white matter structural integrity and connectivity in posttraumatic stress

disorder: A TBSS and tractography study. *Depression and Anxiety*, 34(5), 437-445.

<https://doi.org/10.1002/da.22615>

Orrù, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., & Mechelli, A. (2012, Apr). Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev*, 36(4), 1140-1152.

<https://doi.org/10.1016/j.neubiorev.2012.01.004>

Overstreet, C., Berenz, E. C., Kendler, K. S., Dick, D. M., & Amstadter, A. B. (2017). Predictors and mental health outcomes of potentially traumatic event exposure. *Psychiatry Research*, 247, 296-304.

<https://doi.org/10.1016/j.psychres.2016.10.047>

Paradies, Y. (2006). A systematic review of empirical research on self-reported racism and health. *International Journal of Epidemiology*, 35(4), 888-901.

<https://doi.org/10.1093/ije/dyl056>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, 12, 2825-2830.

Pieterse, A. L., Todd, N. R., Neville, H. A., & Carter, R. T. (2012, Jan). Perceived racism and mental health among Black American adults: a meta-analytic review. *J Couns Psychol*,

59(1), 1-9. <https://doi.org/10.1037/a0026208>

- Pineles, S. L., Arditte Hall, K. A., & Rasmusson, A. M. (2017). Gender and PTSD: different pathways to a similar phenotype. *Current Opinion in Psychology, 14*, 44-48. <https://doi.org/10.1016/j.copsyc.2016.11.002>
- Pinto, M. S., Paoletta, R., Billiet, T., Van Dyck, P., Guns, P.-J., Jeurissen, B., Ribbens, A., Den Dekker, A. J., & Sijbers, J. (2020). Harmonization of Brain Diffusion MRI: Concepts and Methods. *Frontiers in Neuroscience, 14*. <https://doi.org/10.3389/fnins.2020.00396>
- Pomponio, R., Erus, G., Habes, M., Doshi, J., Srinivasan, D., Mamourian, E., Bashyam, V., Nasrallah, I. M., Satterthwaite, T. D., Fan, Y., Launer, L. J., Masters, C. L., Maruff, P., Zhuo, C., Völzke, H., Johnson, S. C., Fripp, J., Koutsouleris, N., Wolf, D. H., Gur, R., Gur, R., Morris, J., Albert, M. S., Grabe, H. J., Resnick, S. M., Bryan, R. N., Wolk, D. A., Shinohara, R. T., Shou, H., & Davatzikos, C. (2020). Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *NeuroImage, 208*, 116450. <https://doi.org/10.1016/j.neuroimage.2019.116450>
- Powers, A., Cross, D., Fani, N., & Bradley, B. (2015). PTSD, emotion dysregulation, and dissociative symptoms in a highly traumatized sample. *Journal of Psychiatric Research, 61*, 174-179. <https://doi.org/10.1016/j.jpsychires.2014.12.011>
- Powers, A., Fani, N., Carter, S., Cross, D., Cloitre, M., & Bradley, B. (2017). Differential predictors of DSM-5 PTSD and ICD-11 complex PTSD among African American

women. *European Journal of Psychotraumatology*, 8(1), 1338914.

<https://doi.org/10.1080/20008198.2017.1338914>

Pupo, M. C., Jorge, M. R., Schoedl, A. F., Bressan, R. A., Andreoli, S. B., Mello, M. F., & Mari, J. D. J. (2011). The accuracy of the Clinician-Administered PTSD Scale (CAPS) to identify PTSD cases in victims of urban violence. *Psychiatry Research*, 185(1-2), 157-160. <https://doi.org/10.1016/j.psychres.2009.11.006>

Ramos-Lima, L. F., Waikamp, V., Antonelli-Salgado, T., Passos, I. C., & Freitas, L. H. M. (2020). The use of machine learning techniques in trauma-related disorders: a systematic review. *Journal of Psychiatric Research*, 121, 159-172.

<https://doi.org/10.1016/j.jpsychires.2019.12.001>

Ramos-Lima, L. F., Waikamp, V., Oliveira-Watanabe, T., Recamonde-Mendoza, M., Teche, S. P., Mello, M. F., Mello, A. F., & Freitas, L. H. M. (2022, 2022/05/01/). Identifying posttraumatic stress disorder staging from clinical and sociodemographic features: a proof-of-concept study using a machine learning approach. *Psychiatry Research*, 311,

114489. <https://doi.org/https://doi.org/10.1016/j.psychres.2022.114489>

Rangaprakash, D., Deshpande, G., Daniel, T. A., Goodman, A. M., Robinson, J. L., Salibi, N., Katz, J. S., Denney, T. S., & Dretsch, M. N. (2017). Compromised hippocampus-striatum pathway as a potential imaging biomarker of mild-traumatic brain injury and

posttraumatic stress disorder. *Human Brain Mapping*, 38(6), 2843-2864.

<https://doi.org/10.1002/hbm.23551>

Roelofs, R., Shankar, V., Recht, B., Fridovich-Keil, S., Hardt, M., Miller, J., & Schmidt, L. (2019). A meta-analysis of overfitting in machine learning. *Advances in Neural Information Processing Systems*, 32, 9179-9189.

Rogers, C. E., Smyser, T., Smyser, C. D., Shimony, J., Inder, T. E., & Neil, J. J. (2016). Regional white matter development in very preterm infants: perinatal predictors and early developmental outcomes. *Pediatric Research*, 79(1), 87-95.

<https://doi.org/10.1038/pr.2015.172>

Sampaio-Baptista, C., & Johansen-Berg, H. (2017, Dec 20). White Matter Plasticity in the Adult Brain. *Neuron*, 96(6), 1239-1251. <https://doi.org/10.1016/j.neuron.2017.11.026>

Santiago Papini, L. M. R. (2014). Psychometric Properties of the Modified Posttraumatic Stress Disorder Symptom Scale among Women with Posttraumatic Stress Disorder and Substance Use Disorders Receiving Outpatient Group Treatments. *Journal of Traumatic Stress Disorders & Treatment*, 04(01). <https://doi.org/10.4172/2324-8947.1000139>

Sarwar, T., Ramamohanarao, K., & Zalesky, A. (2019, Feb). Mapping connectomes with diffusion MRI: deterministic or probabilistic tractography? *Magn Reson Med*, 81(2), 1368-1384. <https://doi.org/10.1002/mrm.27471>

Schore, A. N. (2002). Dysregulation of the Right Brain: A Fundamental Mechanism of Traumatic Attachment and the Psychopathogenesis of Posttraumatic Stress Disorder. *Australian & New Zealand Journal of Psychiatry*, 36(1), 9-30.
<https://doi.org/10.1046/j.1440-1614.2002.00996.x>

Schultebrucks, K., & Chang, B. P. (2021). The opportunities and challenges of machine learning in the acute care setting for precision prevention of posttraumatic stress sequelae. *Experimental Neurology*, 336, 113526.
<https://doi.org/10.1016/j.expneurol.2020.113526>

Schultebrucks, K., Sijbrandij, M., Galatzer-Levy, I., Mouthaan, J., Olf, M., & van Zuiden, M. (2021, May). Forecasting individual risk for long-term Posttraumatic Stress Disorder in emergency medical settings using biomedical data: A machine learning multicenter cohort study. *Neurobiol Stress*, 14, 100297. <https://doi.org/10.1016/j.ynstr.2021.100297>

Schurr, R., Duan, Y., Norcia, A. M., Ogawa, S., Yeatman, J. D., & Mezer, A. A. (2018, 2018/11/01/). Tractography optimization using quantitative T1 mapping in the human optic radiation. *NeuroImage*, 181, 645-658.
<https://doi.org/https://doi.org/10.1016/j.neuroimage.2018.06.060>

- Schwartz, A. C., Bradley, R., Penza, K. M., Sexton, M., Jay, D., Haggard, P. J., Garlow, S. J., & Ressler, K. J. (2006). Pain Medication Use Among Patients With Posttraumatic Stress Disorder. *Psychosomatics*, *47*(2), 136-142. <https://doi.org/10.1176/appi.psy.47.2.136>
- Schwartz, A. C., Bradley, R. L., Sexton, M., Sherry, A., & Ressler, K. J. (2005). Posttraumatic Stress Disorder Among African Americans in an Inner City Mental Health Clinic. *Psychiatric Services*, *56*(2), 212-215. <https://doi.org/10.1176/appi.ps.56.2.212>
- Shalev, A., Liberzon, I., & Marmar, C. (2017, 2017 Jun 22 2018-09-14). Post-Traumatic Stress Disorder. *The New England Journal of Medicine*, *376*(25), 2459-2469. <https://doi.org/http://dx.doi.org/10.1056/NEJMra1612499>
- Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*, *59 Suppl 20*, 22-33;quiz 34-57.
- Siehl, S., King, J. A., Burgess, N., Flor, H., & Nees, F. (2018). Structural white matter changes in adults and children with posttraumatic stress disorder: A systematic review and meta-analysis. *NeuroImage: Clinical*, *19*, 581-598. <https://doi.org/10.1016/j.nicl.2018.05.013>

Sirsat, M. S., Fermé, E., & Câmara, J. (2020). Machine Learning for Brain Stroke: A Review.

Journal of Stroke and Cerebrovascular Diseases, 29(10), 105162.

<https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105162>

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, 17(3), 143-

155. <https://doi.org/10.1002/hbm.10062>

Spates, K., Evans, N. T., James, T. A., & Martinez, K. (2020). Gendered Racism in the Lives of Black Women: A Qualitative Exploration. *Journal of Black Psychology*, 46(8), 583-606.

<https://doi.org/10.1177/0095798420962257>

Stein, M. B., Jang, K. L., Taylor, S., Vernon, P. A., & Livesley, W. J. (2002). Genetic and Environmental Influences on Trauma Exposure and Posttraumatic Stress Disorder Symptoms: A Twin Study. *American Journal of Psychiatry*, 159(10), 1675-1681.

<https://doi.org/10.1176/appi.ajp.159.10.1675>

Suo, X., Lei, D., Li, W., Sun, H., Qin, K., Yang, J., Li, L., Kemp, G. J., & Gong, Q. (2022).

Psychoradiological abnormalities in treatment-naive noncomorbid patients with posttraumatic stress disorder. *Depression and Anxiety*, 39(1), 83-91.

<https://doi.org/10.1002/da.23226>

Takemura, H., Caiafa, C. F., Wandell, B. A., & Pestilli, F. (2016). Ensemble Tractography.

PLOS Computational Biology, 12(2), e1004692.

<https://doi.org/10.1371/journal.pcbi.1004692>

Ulloa Cerna, A. E. (2016). *Data Driven Sample Generator Model with Application to*

Classification. https://digitalrepository.unm.edu/math_etds/82

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm

validation with a limited sample size. *PLOS ONE*, 14(11), e0224365.

<https://doi.org/10.1371/journal.pone.0224365>

Wang, S., Hu, H., Su, S., Liu, L., Wang, Z., Wang, Q., & Shen, D. (2016). Combination of Grey

Matter and White Matter Features for Early Prediction of Posttraumatic Stress Disorder.

In (pp. 560-567). Springer International Publishing. [https://doi.org/10.1007/978-3-319-](https://doi.org/10.1007/978-3-319-46257-8_60)

[46257-8_60](https://doi.org/10.1007/978-3-319-46257-8_60)

Weathers, F. W., Bovin, M. J., Lee, D. J., Sloan, D. M., Schnurr, P. P., Kaloupek, D. G., Keane,

T. M., & Marx, B. P. (2018). The Clinician-Administered PTSD Scale for DSM-5

(CAPS-5): Development and initial psychometric evaluation in military veterans.

Psychological Assessment, 30(3), 383-395. <https://doi.org/10.1037/pas0000486>

Weathers, F. W., Litz, B., Herman, D., Juska, J., & Keane, T. (1994). PTSD checklist—civilian

version. *Journal of Occupational Health Psychology*.

Williams, D. R., & Collins, C. (1995). US Socioeconomic and Racial Differences in Health:

Patterns and Explanations. *Annual Review of Sociology*, 21(1), 349-386.

<https://doi.org/10.1146/annurev.so.21.080195.002025>

Wolf, E. J., Sadeh, N., Leritz, E. C., Logue, M. W., Stoop, T. B., McGlinchey, R., Milberg, W.,

& Miller, M. W. (2016). Posttraumatic Stress Disorder as a Catalyst for the Association

Between Metabolic Syndrome and Reduced Cortical Thickness. *Biological Psychiatry*,

80(5), 363-371. <https://doi.org/10.1016/j.biopsych.2015.11.023>

Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann,

C., Jenkinson, M., & Smith, S. M. (2009). Bayesian analysis of neuroimaging data in

FSL. *NeuroImage*, 45(1), S173-S186. <https://doi.org/10.1016/j.neuroimage.2008.10.055>

Yeatman, J. D., Weiner, K. S., Pestilli, F., Rokem, A., Mezer, A., & Wandell, B. A. (2014). The

vertical occipital fasciculus: A century of controversy resolved by in vivo measurements.

Proceedings of the National Academy of Sciences, 111(48), E5214-E5223.

<https://doi.org/doi:10.1073/pnas.1418503111>

Yeh, F.-C.[Frank Yeh]. (2020a). *There is something wrong to draw corticobular tract.* [Online

forum post]. DSI Studio Google Group. [https://groups.google.com/g/dsi-](https://groups.google.com/g/dsi-studio/c/xJpVF5yc0bM/m/6qsEpJT_AwAJ)

[studio/c/xJpVF5yc0bM/m/6qsEpJT_AwAJ](https://groups.google.com/g/dsi-studio/c/xJpVF5yc0bM/m/6qsEpJT_AwAJ)

- Yeh, F.-C. (2020b). Shape analysis of the human association pathways. *NeuroImage*, 223, 117329. <https://doi.org/10.1016/j.neuroimage.2020.117329>
- Yeh, F.-C., Panesar, S., Barrios, J., Fernandes, D., Abhinav, K., Meola, A., & Fernandez-Miranda, J. C. (2019). Automatic Removal of False Connections in Diffusion MRI Tractography Using Topology-Informed Pruning (TIP). *Neurotherapeutics*, 16(1), 52-58. <https://doi.org/10.1007/s13311-018-0663-y>
- Yeh, F.-C., Panesar, S., Fernandes, D., Meola, A., Yoshino, M., Fernandez-Miranda, J. C., Vettel, J. M., & Verstynen, T. (2018). Population-averaged atlas of the macroscale human structural connectome and its network topology. *NeuroImage*, 178, 57-68. <https://doi.org/10.1016/j.neuroimage.2018.05.027>
- Yeh, F.-C., Verstynen, T. D., Wang, Y., Fernández-Miranda, J. C., & Tseng, W.-Y. I. (2013). Deterministic Diffusion Fiber Tracking Improved by Quantitative Anisotropy. *PLOS ONE*, 8(11), e80713. <https://doi.org/10.1371/journal.pone.0080713>
- York, T. P., Strauss, J. F., Neale, M. C., & Eaves, L. J. (2010). Racial Differences in Genetic and Environmental Risk to Preterm Birth. *PLOS ONE*, 5(8), e12391. <https://doi.org/10.1371/journal.pone.0012391>
- Zandvakili, A., Barredo, J., Swearingen, H. R., Aiken, E. M., Berlow, Y. A., Greenberg, B. D., Carpenter, L. L., & Philip, N. S. (2020). Mapping PTSD symptoms to brain networks: a

machine learning study. *Translational Psychiatry*, *10*(1). <https://doi.org/10.1038/s41398-020-00879-2>

Zhang, Q., Wu, Q., Zhu, H., He, L., Huang, H., Zhang, J., & Zhang, W. (2016). Multimodal MRI-Based Classification of Trauma Survivors with and without Post-Traumatic Stress Disorder. *Frontiers in Neuroscience*, *10*. <https://doi.org/10.3389/fnins.2016.00292>

Zhu, W., Xie, L., Han, J., & Guo, X. (2020). The Application of Deep Learning in Cancer Prognosis Prediction. *Cancers*, *12*(3), 603. <https://doi.org/10.3390/cancers12030603>