

Post-Traumatic Stress Disorder (PTSD) biomarker identification using a deep learning model

Sohini Banerjee¹, Mazen Istanbouli², Bernie J. Daigle, Jr.²

¹ Thomas Jefferson High School for Science and Technology, Alexandria, Virginia

² University of Memphis, Memphis, Tennessee

SUMMARY

Post-traumatic stress disorder (PTSD) is a neuropsychological disorder in which individuals struggle to recover from a traumatizing event. It affects a significant population, including COVID-19 patients, frontline health workers, and war veterans. Given biases associated with self-assessment and diagnosis of PTSD, researchers are actively searching for unbiased biological markers (biomarkers) for predicting PTSD status. The Systems Biology of PTSD Consortium has collected candidate biomarkers for PTSD using molecular and clinical measurements of male war veterans between the ages of 20 and 60. PTSD-positive and negative subjects were separated based on the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) Clinician Administered PTSD Scale (CAPS) scores, derived from structured interviews to measure an individual's abundance of symptoms, such as re-experiencing, flashbacks, and hyperarousal. CAPS scores higher than 40 were considered PTSD-positive and below 20 were considered negative. We created artificial neural network models to classify PTSD-positive and negative individuals based on metabolomics, micro-RNA (miRNA), protein expression, endocrine markers, and DNA methylation datasets. Model training involved 64 iterations of a Bayesian Hyperparameter Optimization algorithm with 5-fold cross-validation. Each model was calibrated based on cross-validation performance and variance across iterations and then fit to the entire respective dataset (76 PTSD-positive, 76 PTSD-negative). We applied the trained models to an independent validation cohort to assess accuracy on unseen datasets. The top performing datasets from the validation cohort based on classification accuracy were metabolomics (65.2%) and protein expression (61.8%). We anticipate the candidate biomarkers identified in this and future studies will assist with the diagnosis of PTSD.

INTRODUCTION

Post-traumatic stress disorder (PTSD) is a disorder characterized by the difficulties people encounter when recovering from a trauma (1). PTSD is a widespread condition, affecting many people including war veterans and COVID-19 patients. For example, one study examined 381 patients with SARS-CoV-2 and found that 30.2% of them suffered from PTSD (2). These patients were also diagnosed with depressive episodes and generalized anxiety disorder, both

of which share symptoms with PTSD (2). Additionally, about 3.5% of people in the United States (U.S.) are diagnosed with PTSD annually, and 1 in 11 individuals will suffer from PTSD at some point in their life (3). Another study examined a group of U.S. veterans from wars in Iraq and Afghanistan. Forty percent of these veterans had PTSD (4). Eighteen percent had subthreshold PTSD, which means the individuals did not meet all *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (DSM-IV), PTSD diagnostic requirements but still exhibited several symptoms of the disorder (4). Due to numerous stigmas associated with mental health, only half of U.S. veterans receive needed treatment for PTSD (5).

Researchers are actively searching for PTSD biomarkers to help better understand the biological underpinnings of the disorder. While there are studies that have focused on identifying biomarkers for PTSD, very few have analyzed a vast collection of datasets encompassing several omics levels such as metabolomics, proteomics, and genomics. In particular, the Department of Defense-funded Systems Biology of PTSD Consortium has collected clinical and molecular data from war veterans to develop a reproducible panel of blood-based biomarkers for PTSD (6). Using statistical tests and machine learning models, such as recursive feature elimination and support vector machine, the Consortium published a study in 2019 describing the isolation of 28 biomarkers spanning all collected datasets. The Consortium used these biomarkers to predict the PTSD status of veterans in an independent validation cohort with 81% accuracy (6). The study justifies binary classification of PTSD patients using the DSM-IV Clinician Administered PTSD Scale (CAPS) score by separating patients with moderate to extreme PTSD conditions from patients with no symptoms (6). Specifically, minimum CAPS score cutoffs of 40 for PTSD-positive patients and maximum cutoffs of 20 for PTSD-negative patients were chosen to create the two groups (6). Furthermore, one study found that insulin resistance was a key biomarker that corresponded to the severity of PTSD symptoms (7). Another study used the DSM-5 criteria to identify severity of PTSD symptoms and separated individuals into cases and controls (8). They focused on identifying biomarkers at the genetic level and concluded that an allele of *APOE2*, associated with reexperiencing, can help distinguish PTSD patients from controls (8). The study also found that complement system genes may serve as PTSD biomarkers (8).

In our study, we used the same datasets used by the Consortium (6). However, our analysis evaluated the ability of deep neural networks (DNNs) to accurately predict PTSD status based on individual clinical and molecular datasets. A neural network is a computational network of nodes (neurons) loosely modeled after the brain (9). The network weights are values assigned to edges between nodes of consecutive layers, which are multiplied by the values of

the nodes from the incoming layer (9). To get the value of each node in the network, the model evaluates an activation function applied to the sum of all the nodes' inputs multiplied by their corresponding weights (9). The resulting value determines the impact this node will have on values of nodes in subsequent layers (9). Nodes in the first layer of a neural network simply take on the values of the training data for a given dataset (9). Each network may have varying numbers of layers, and the final layer has a single node, which in our study represents the binary output of PTSD-positive or negative (9). In the process of training, the model quantifies error by calculating the difference between the true output (case/control status) and the output of the neural network (9). The neural network then modifies its weights to reduce error and repeats this process many times (**Figure 1**) (9). A neural network functions by finding the optimal combination of features that can accurately predict the output variable, and if a feature is not statistically significant between the two categories of patients, it is unlikely to positively contribute to the neural network. Including features without significance may negatively impact the neural network by creating noise, as the values of the feature do not substantially differ between PTSD-positive and negative patients.

One study compared deep neural networks to standard machine learning models, including support vector machines and logistic regression, and found that DNNs tended to perform classification with higher accuracy and were more effective in utilizing large amounts of training data (10). Another study examined the binary classification accuracy for drugs and non-drugs and found that support vector machines and artificial neural networks had comparable performance at 82% and 80% accuracy, respectively (11). There are various applications of deep learning, such as natural language processing, speech recognition, and bioinformatics, all of which aim to train a model on a large quantity of input data to make predictions about an occurrence (12). Along with some inconsistencies in the literature regarding the classification power of neural networks, their use is not as clearly documented as other machine learning techniques in the context of biological conditions. Thus, we used neural networks in this study to further investigate their classification performance and contribute findings to the goal of accurately diagnosing PTSD. We hypothesized that a neural network would improve classification accuracy of an unseen cohort of PTSD war veterans. Our model used the metabolomics dataset to predict PTSD at roughly 65.2% accuracy, nearing the highest accuracy in the field when individual datasets are used. A DNN can aid with PTSD diagnosis, as it is based on the levels of biological markers and thus, not subject to stigmas surrounding mental health.

RESULTS

Model Performance

We performed two phases of testing with two independent cohorts to evaluate the performance of neural networks on an unseen cohort of PTSD veterans: Original Testing involved training and testing on the Original Biomarkers cohort, and Validation Testing involved training on Original Biomarkers and testing on Validation Biomarkers. For Original Testing and Validation Testing, the test accuracy for classifying PTSD cases and controls was generally lower than the cross-validation accuracy, which was the average

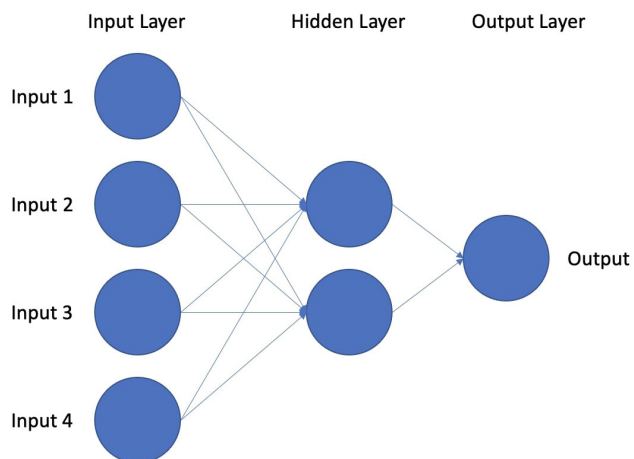


Figure 1: Sample neural network diagram with one internal layer. The input layer represents the data from a single dataset. The arrows represent the weights in the network, with each arrow having a different weight that connects one node to another node in the next layer. The output layer represents the final value of the network, which is used to calculate the error by finding the difference between this value and the true output.

training accuracy. Additionally, results of Original Testing showed higher sensitivities, the percentage of true positives, than Validation Testing, except for Endocrine Blood, which showed a sensitivity of 0. In contrast, the specificities, the percentage of true negatives, of both Original Testing and Validation Testing showed no consistent differences, with some datasets showing specificities close to 1 and others close to 0. In terms of individual datasets, Methylation Zymo-Probe, miRNA Deplete, and miRNA Exosome datasets had consistent sensitivities for binary PTSD classification, while the Methylation Zymo-Gene and Methylation Zymo-Probe datasets had consistent specificities across both testing methods. We did not find any statistically significant features based on the one-way ANOVA test in the Endocrine Blood and Metabolomics Metabolon datasets. Thus, we excluded them from subsequent testing because the features were unlikely to classify PTSD effectively if they were not statistically significant between PTSD cases and controls. The Metabolomics UCSF dataset had the highest test accuracy (65.2%) in Validation Testing, while Protein ELISA and Protein SRM had similar performances (61.8%). In comparison, the miRNA datasets, miRNA Deplete (47.1%), miRNA Exosome (50%), and miRNA Plasma (58.8%) had lower PTSD classification accuracies. Finally, the methylation datasets performed similarly, with Methylation Zymo-Gene classifying PTSD cases and controls at 56.7% and Methylation Zymo-Probe classifying at 55.2% (**Table 1**).

Feature Importance

We observed that most datasets had between 2–11 important features selected for prediction of PTSD on the Validation Biomarkers data. However, all three miRNA datasets (Deplete, Exosome, Plasma) used up to 20 features to predict PTSD status in the Validation Biomarkers data (**Table 2**). Some datasets that required fewer features for PTSD prediction showed a clear importance of one or two features, whereas the feature importance for the miRNA datasets were

Original Testing				
Dataset	CV Accuracy	Test Accuracy	Sensitivity	Specificity
Endocrine Blood	0.609	0.44	0	1
Metabolomics UCSF	0.652	0.44	1	0.067
Methylation Zymo-Gene	0.773	0.702	0.565	0.833
Methylation Zymo-Probe	0.727	0.617	0.69	0.5
miRNA Deplete	0.696	0.62	0.739	0.519
miRNA Exosome	0.696	0.68	0.667	0.7
miRNA Plasma	0.696	0.74	0.667	0.826
Protein ELISA	0.696	0.633	0.667	0.591
Protein SRM	0.652	0.673	0.905	0.5
Validation Testing				
Dataset	CV Accuracy	Test Accuracy	Sensitivity	Specificity
Endocrine Blood	0.606	0.571	0.7	0.455
Metabolomics UCSF	0.697	0.652	0.4	0.846
Methylation Zymo-Gene	0.71	0.567	0.138	0.895
Methylation Zymo-Probe	0.71	0.552	0.621	0.5
miRNA Deplete	0.667	0.471	0.586	0.385
miRNA Exosome	0.697	0.5	0.517	0.487
miRNA Plasma	0.697	0.588	0.034	1
Protein ELISA	0.667	0.618	0.345	0.821
Protein SRM	0.697	0.618	0.207	0.923

Table 1: Neural network classification performance summary for Original and Validation Testing. Original Testing and Validation Testing cross-validation (CV) accuracy, test accuracy, sensitivity, and specificity are listed for each subdataset.

more evenly distributed (**Figure 2-3**). The Endocrine Blood dataset was narrowed down from 13 features to 2, and the Metabolomics dataset from over 30 biomarkers to 6. Similarly, the protein datasets, ELISA and SRM, respectively went from 8 to 3 and 85 to 5 biomarkers in Validation Testing. Specifically, Neuropeptide Y (88%) in Endocrine Blood; insulin (45%) and c45_l_3p8p (38%) in Metabolomics UCSF; MDC1 (43%) and APOE (24%) in Methylation Zymo-Gene; Complement Factor H (CFH, 55%) in the Protein enzyme-linked immunosorbent assay (ELISA) dataset; and C4BPB (46%) and ACTC1 (40%) in the Protein selected reaction monitoring (SRM) dataset accounted for the majority of relative feature importance in each testing dataset (**Table 3**). In the miRNA datasets, no feature accounted for more than 10% of the relative importance when predicting PTSD. We determined the log-fold change of PTSD biomarkers by calculating the difference in concentration between the Original Biomarkers and Validation Biomarkers cohort, relative to the feature's average concentration in Original Biomarkers. The average log-fold change of CFH between PTSD cases and controls is

0.366 in the Original Biomarkers and 0.208 in the Validation Biomarkers cohorts, which means people with PTSD have slightly higher concentrations of CFH on average. In contrast, the average log-fold change of C4BPB is -0.245 in Original Biomarkers and -0.092 in Validation Biomarkers, indicating that PTSD cases have slightly lower concentrations of C4BPB on average. Finally, we identified the candidate biomarker hsa-miR-192-5p from the miRNA Plasma dataset, which is known to target *NUAK1* (6, 13). The average log-fold change of hsa-miR-192-5p is -0.046 in Original Biomarkers and -0.018 in Validation Biomarkers, indicating that PTSD cases have slightly lower levels of this miRNA on average.

We began our analysis with approximately 1,300 molecular features collected across all the datasets. After removing statistically insignificant features for each dataset using the one-way ANOVA test, our neural network models further reduced the number of features in each dataset to predict PTSD status as accurately as possible. Large datasets, such as miRNA Deplete, miRNA Exosome, and miRNA Plasma, showed a significant feature reduction of 94–96% of the

Dataset	Subdataset	Original Features	Selected Features (Original Testing)	Selected Features (Validation Testing)
Endocrine	Endocrine Blood	13	1	2
	Endocrine Urine	20	N/A	N/A
Metabolomics	Metabolomics Metabolon	244	N/A	N/A
	Metabolomics UCSF	34	3	6
Methylation	Methylation Zymo-Gene	11	6	7
	Methylation Zymo-Probe	22	7	11
miRNA	miRNA Deplete	459	20	20
	miRNA Exosome	350	20	20
	miRNA Plasma	334	20	20
Protein	Protein ELISA	8	3	3
	Protein SRM	85	6	5

Table 2: Original vs. selected features. The number of original features as well as the number of selected features in Original Testing and Validation Testing are listed for each subdataset.

■ ELISA Neuropeptide Y (NPY) ■ ELISA Cortisol

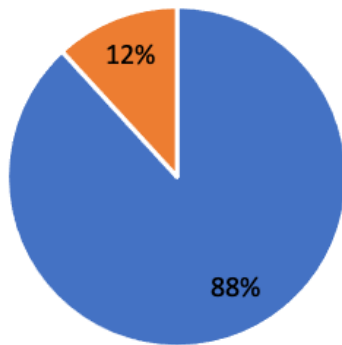


Figure 2: Endocrine Blood feature importance distribution. The pie chart shows the relative contribution of the selected features. The dataset had 13 total features.

■ C4BPB ■ ACTC1 ■ APOC4 ■ CA1 ■ PRG4

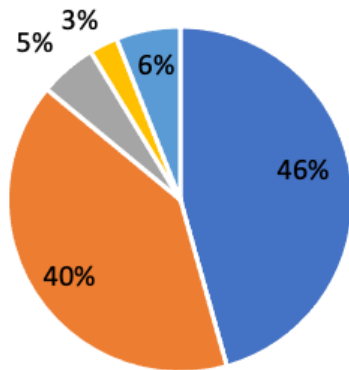


Figure 3: Protein SRM feature importance distribution. The pie chart shows the relative contribution of the selected features. The dataset had 85 total features in the neural network.

original count. In contrast, small datasets like Methylation Zymo-Gene, Methylation Zymo-Probe, and Protein ELISA showed a smaller percentage reduction of features at approximately 32-55% of the original feature count.

DISCUSSION

As expected, the observed test accuracy was generally lower than the cross-validation accuracy because the model was evaluated on unseen data during testing. Additionally, most of the test accuracies in Validation Testing were lower compared to those of Original Testing because the model was tested on the Validation Biomarkers cohort, which consists of an entirely separate group of veterans compared to the Original Biomarkers cohort. Compared to the Consortium, which used the same datasets to predict PTSD status in the Validation Biomarkers cohort at approximately 81% accuracy, our neural networks did not perform as well, with the highest observed classification performance for the validation cohort at 65.2% accuracy using the Metabolomics UCSF dataset (6). This is not unexpected, as our study considered each

Feature	Dataset	Relative %
ELISA Neuropeptide Y (NPY)	Endocrine Blood	88
Complement Factor H	Protein ELISA	55
C4BPB	Protein SRM	46
insulin	Metabolomics UCSF	45
MDC1	Methylation Zymo-Gene	43
ACTC1	Protein SRM	40
c45_1_3p8p	Metabolomics UCSF	38

Table 3: Top 7 most important features for Validation Testing. Validation Testing is the process of training using five-fold cross-validation on the Original Biomarkers cohort and testing on the Validation Biomarkers cohort. Relative % represents the importance of each feature relative to its own dataset.

dataset individually, whereas the Consortium combined features across all the datasets to predict PTSD status, giving such a model higher predictive power (6). However, when comparing the performance of the individual datasets in the Consortium, our performance was comparable (6). For example, our miRNA Plasma dataset performed at 58.8% accuracy, whereas the Consortium achieved roughly 50% accuracy with all miRNA datasets combined (6). Additionally, both the Protein ELISA and Protein SRM datasets in our study performed at 61.8% accuracy, while the Consortium showed lower than 60% accuracy even when the researchers combined both protein datasets (6). Methylation datasets in both studies performed similarly, achieving approximately 55% accuracy (6).

The endocrine data measure the concentration of various hormones in the blood and urine. The metabolomics data measure substances used in metabolism, which is the collection of all chemical processes that convert organic material into energy (14). ELISA Neuropeptide Y (NPY) holds the highest rank in terms of relative importance at 88% in the Endocrine Blood dataset. NPY is a peptide made of 36 amino acids and is frequently involved in stress regulation (15, 16). One study found an abnormal level of NPY in a rat depression model, while another found an association between NPY and stress and anxiety levels, which is notable because depression, anxiety, and stress are all symptoms of PTSD (16, 17). This prior research corroborates our results, suggesting that NPY can be used as a biomarker to predict PTSD status.

Insulin had the highest importance in the Metabolomics UCSF dataset. The primary function of insulin is to allow cells to absorb glucose, or sugar, from the bloodstream (18). Several studies that searched for biomarkers for PTSD found that insulin resistance was greater in individuals with PTSD compared to controls (6, 7). One study found that, along with insulin resistance, PTSD patients had higher levels of blood glucose (19). While this finding does not indicate insulin causes PTSD or results from PTSD, as there may be a third variable involved that affects both the presence of PTSD and insulin levels, insulin can be used as a correlation factor to distinguish PTSD cases from controls. Finally, elevated blood sugar frequently causes obesity, depression, and anxiety, all of which are frequently associated with PTSD (20). These findings support our results suggesting that insulin levels (and insulin resistance) are good indicators of PTSD presence and severity.

The Protein enzyme-linked immunosorbent assay (ELISA) and selected reaction monitoring (SRM) data measure protein

concentrations. Complement Factor H (CFH) is the highest-ranking feature from the Protein ELISA dataset with a relative importance of 55%. CFH is a component of the human immune system ensuring that the complement system is activated when foreign organisms, such as viruses and bacteria, are detected (21). One study measured the distributions in genotype and allele frequencies of various complement factors, including CFH, in PTSD cases and control patients (8). However, they did not find a statistically significant difference in CFH in any genotype or allele frequencies between the two groups (8).

At 46%, *C4BPB* had the highest relative importance for predicting PTSD status in the Protein SRM dataset. While previous studies have not found a significant role for *C4BPB* in response to trauma, the known function of this gene involves immunity and negative regulation of the complement system (22). Considering previous studies described a relationship between PTSD and the activation of complement genes, our identification of *C4BPB* further implicates the role of the complement system in the PTSD phenotype (8). The directional change of *C4BPB* is opposite that of CFH, which makes sense given that CFH activates the complement system, while *C4BPB* inhibits it. Specifically, *C4BPB* levels are lower in PTSD cases compared to controls, while CFH levels are higher because the complement system is activated.

The miRNA data measure levels of micro-RNA, a type of non-coding RNA that is transcribed from DNA and primarily works to regulate gene expression (23). The miRNA Plasma data consist of miRNAs isolated from blood plasma. The miRNA Exosome data represent plasma miRNAs found within exosomes, which are vesicles that carry different substances around the cell, while miRNA Deplete data contain the miRNAs remaining after depleting plasma of the exosomes (24). In the miRNA datasets, our models consistently selected 20 features for classification, and thus each feature had lower relative importance than the features in the Endocrine and Protein datasets. We identified many biomarkers in these datasets that overlapped those found by the Consortium (6). Given that most miRNAs inhibit expression of their targets, we would expect the levels of *NUAK1* to increase in PTSD subjects when *hsa-miR-192-5p* decreased. One study investigated *NUAK1* for PTSD patients and found a positive log-fold change for this gene, indicating that *NUAK1* was expressed at higher levels in PTSD cases than controls (13).

We did not calculate the feature importance distributions for Original Testing, as we preferred to focus on biomarkers that can effectively predict PTSD status in an independent cohort of veterans. At the transcriptomic level, we found over 60 biomarkers in the miRNA datasets that contributed relatively equally to binary classification of PTSD. However, we observed a different pattern for biomarkers from protein, metabolomics, and endocrine datasets, which had a few features with disproportionately high contributions to the neural network's PTSD classification.

Overall, the number of markers selected indicates that a relatively large number of biomarkers have high importance in binary classification at the transcriptomic level, while there are significantly fewer from other categories. The methylation data measure the levels of DNA methylation throughout the genome, which collectively provide an epigenetic mechanism for controlling gene expression (25). In our study, transcriptomic biomarkers were miRNA and methylation, which had several more variables compared to

other datasets like endocrine or protein expression, and thus were likely to have more unique combinations of genes that were expressed, allowing for improved PTSD classification.

There are some limitations of our study, particularly concerning the applicability of the results to other populations. The Consortium collected clinical and molecular data from male war veterans, meaning our results are unlikely to generalize well to women and individuals with non-combat-related trauma (6). In addition, the Original Biomarkers and Validation Biomarkers cohorts had 165 and 67 samples, respectively. A larger cohort would have been preferred to give the neural network more training data, possibly allowing to classify PTSD on an unseen cohort more effectively. Finally, we evaluated the neural network classification performance on individual datasets. However, our methods cannot take into consideration that there may be dependencies between potential biomarkers across different datasets. Thus, an integrative algorithm that encompasses information from all the datasets, such as Integrative Network Fusion (INF) (26), could be useful. The INF algorithm has been tested as a multi-omics method on cancer genome datasets and could provide a better PTSD prediction model by simultaneously using data from multiple datasets (26).

This study allowed us to down-select from over 1,300 clinical and molecular features from war veterans to approximately 100 biomarkers that were most predictive for PTSD status. In many of our analyzed datasets, we identified a small number of features having disproportionately high importance for classification. Our results support the understanding that complex neurological diseases including psychiatric disorders are influenced by multiple biological factors. Thus, we would expect a combination of features to be necessary to predict PTSD status with the highest accuracy. We anticipate that our deep learning approach and the candidate biomarkers identified in this study will be useful for assisting medical practitioners with the diagnosis of PTSD in the future.

MATERIALS AND METHODS

Datasets

We analyzed two cohorts of the Systems Biology of PTSD Consortium's molecular and clinical measurements of male veterans between the ages 20–60 years (6). Researchers collected measurements for each dataset from two independent cohorts of male war veterans from Operation Iraqi Freedom or Operation Enduring Freedom: the "Original Biomarkers" cohort, consisting of 165 samples, and the "Validation Biomarkers" cohort, consisting of 67 samples (6). The Consortium classified the subjects in each cohort as PTSD-positive or negative based on their DSM-IV CAPS scores, derived from a structured interview to measure a patient's abundance of symptoms, such as re-experiencing, flashbacks, and hyperarousal (6). Researchers marked individuals with CAPS scores >40 as PTSD-positive ("cases") and <20 as PTSD-negative ("controls") (6).

The collected datasets for both cohorts include Endocrine Blood, Endocrine Urine, Metabolomics Metabolon, Metabolomics UCSF, Methylation Zymo-Gene, Methylation Zymo-Probe, miRNA Deplete, miRNA Exosome, miRNA Plasma, Protein ELISA, and Protein SRM (6). The Metabolon Corporation collected the Metabolomics Metabolon data, and researchers at the University of California, San Francisco

collected the Metabolomics UCSF data. For each dataset from both cohorts, we applied the Python Standard scaler to preprocess features to a mean of 0 and variance of 1. We also shuffled each dataset before training and testing models to avoid any predisposed biases in the ordering of samples.

Original Testing

To begin, we randomly split each dataset from the Original Biomarkers cohort into a training and testing set, 70% and 30%, respectively. Using an R program to conduct a one-way ANOVA test to identify statistically significant differences between the means of cases and controls in the training set, we removed features from each dataset whose p -values were > 0.05 . We then trained classification models for each dataset on the training set data and evaluated on the testing set data.

Validation Testing

To understand the ability of our models to classify data from an independent cohort, we also used each of the entire Original Biomarkers datasets as training sets and the Validation Biomarkers datasets as test sets. As above, we wrote an R program to conduct a one-way ANOVA test on each training dataset and removed features with p -values > 0.05 before performing classifier training and testing.

Artificial Neural Network

We trained each model using 64 iterations of a 5-fold cross-validation procedure on the training set before evaluating on the test set. Cross-validation provides an estimate of the performance of a machine learning model on a training dataset. To perform a 5-fold cross-validation, we divided each training dataset into 5 folds (sections) and trained 5 different classifiers, each using a unique set of 4 folds for training and the remaining fold for validation (27). We determined the performance of each iteration of cross-validation by averaging the performances of all 5 validation folds of the training set (27). Importantly, the models used for both Original Testing and Validation Testing were not given the testing data before the final evaluation. We used the Adaptive Movement Estimation Optimizer and Binary Cross-Entropy Loss functions to train the model. Additionally, we used three types of activation functions: swish, ReLu, and sigmoid (**Table 4**) (28, 29). To build and apply the network, we used the Python (3.7.3) programming language with the packages NumPy (1.18.5), Pandas (1.1.0), Keras (2.4.3), scikit-learn (0.23.2), scikit-optimize (0.8.1), and Google Tensorflow (2.3.0) (30).

We used the Bayesian Hyperparameter Optimization algorithm to modify the weights and select optimal model hyperparameters, which are pre-determined characteristics of a neural network model. The algorithm was used to test various combinations of hyperparameters, including the number of internal layers and learning rate of the neural network, to identify the combination that predicted PTSD at the highest accuracy. We optimized the number of internal layers in the network (between 1 and 3) and the nodes per layer factor (between 0.1 and 1.0), which is the proportion of total nodes used in each internal layer. We also optimized the dropout rate (between 0.15 and 0.50), which determines the probability that a node is used in training, as some nodes are excluded to avoid overfitting the model to the training data (31). We used ReLu or swish for the internal activation functions

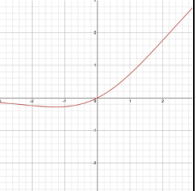
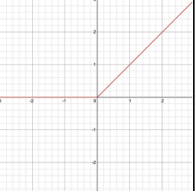

Swish		$f(x) = x * (1 + \exp(-x))^{-1}$
ReLu		$f(x) = x \{x > 0\}, 0 \{x \leq 0\}$
Sigmoid		$f(x) = 1 / (1 + e^{-x})$

Table 4: Equations and graphs for the swish, ReLu, and sigmoid activation functions. Based on the value of a single input node in the neural network, the equations determine the new value that will be passed on to the subsequent layer of the network, which represents the sum of the outputs of multiple input nodes.

and sigmoid for the final activation function. The number of epochs (between 25 and 100) and mini-batch size (between 16 and 64) were optimized as well. Finally, we optimized the learning rate (between 0.001 and 0.100) that determines the fraction of the weights to modify during training (32).

Feature Importance

To select only the most important features, we designed the artificial neural network to include a maximum of 20 features to classify PTSD status. To calculate feature importance, we first selected the final model with the best performance on the Validation Biomarkers dataset. Then, we randomized each selected feature (features that were statistically significant at a p -value cutoff of 0.05 on the Original Biomarkers dataset) and calculated the difference in accuracy on the Validation set when the model included the current feature versus when it did not. We selected the top 20 features that produced the highest deviation from the initial accuracy and calculated the relative deviations to produce feature importance for each variable.

Each iteration of Bayesian Hyperparameter Optimization produced a different model, and we applied only those models whose cross-validation accuracies had a variance of < 0.01 to the testing data. To avoid model overfitting, we discarded each model with the highest training accuracy after satisfying the variance condition. We thus selected models for testing with training accuracies of approximately 90% of the highest observed (**Table 5**).

We measured cross-validation accuracy, test accuracy, sensitivity, and specificity to assess our model's classification performance. Cross-validation accuracy is the average proportion of PTSD cases classified correctly in Original Testing, while test accuracy is the proportion of PTSD cases classified correctly in Validation Testing. These metrics are based on using the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) counts for

Model Details				
Dataset	Number of Internal Layers	Nodes per Layer Factor	Internal Activation Function	Final Activation Function
Endocrine Blood	1	1	swish	sigmoid
Metabolomics UCSF	1	1	swish	sigmoid
Methylation Zymo-Gene	3	0.609	swish	sigmoid
Methylation Zymo-Probe	1	0.841	relu	sigmoid
miRNA Deplete	2	0.536	relu	sigmoid
miRNA Exosome	1	0.436	swish	sigmoid
miRNA Plasma	1	0.651	relu	sigmoid
Protein ELISA	1	1	swish	sigmoid
Protein SRM	1	0.888	swish	sigmoid

Table 5: Model details for Original and Validation Testing neural networks. The numbers of internal layers, nodes per layer factor, internal activation functions, and final activation functions are noted for the selected model for each dataset in Validation Testing.

each dataset. Accuracy is the fraction of samples identified correctly as PTSD-positive or negative. Sensitivity is the fraction of samples correctly identified as PTSD-positive, given that the patient tested positive (33). Specificity is the fraction of samples correctly identified as PTSD-negative, given that the patient tested negative (33).

ACKNOWLEDGEMENTS

Research was sponsored in part by the Army Research Laboratory and was accomplished under Grant Number W911NF-17-1-0069. We would like to thank all of the members of the Systems Biology of PTSD Consortium for generating and sharing the datasets used in this study, as well as for providing their expertise to support this work.

Received: July 9, 2021

Accepted: February 6, 2022

Published: September 23, 2022

REFERENCES

1. "Post-traumatic stress disorder (PTSD) - symptoms and causes." *Mayo Clinic*, www.mayoclinic.org/diseases-conditions/post-traumatic-stress-disorder/symptoms-causes/syc-20355967.
2. Janiri, Delfina, et al. "Posttraumatic stress disorder in patients after severe COVID-19 infection." *JAMA Psychiatry*, vol. 78, no. 5, May 2021, pp. 567–69, doi.org/10.1001/jamapsychiatry.2021.0109.
3. "What is posttraumatic stress disorder?" www.psychiatry.org/patients-families/ptsd/what-is-ptsd.
4. Jakupcak, Matthew, et al. "Anger, hostility, and aggression among Iraq and Afghanistan war veterans reporting PTSD and subthreshold PTSD." *Journal of Traumatic Stress*, vol. 20, no. 6, 2007, pp. 945–54, doi.org/10.1002/jts.20258.
5. Sharp, Marie-Louise, et al. "Stigma as a barrier to seeking health care among military personnel with mental health problems." *Epidemiologic Reviews*, vol. 37, 2015, pp. 144–62, doi.org/10.1093/epirev/mxu012.
6. Dean, Kelsey R., et al. "Multi-omic biomarker identification and validation for diagnosing warzone-related post-traumatic stress disorder." *Molecular Psychiatry*, vol. 25, no. 12, Dec. 2020, pp. 3337–49, doi.org/10.1038/s41380-019-0496-z.
7. "PTSD linked to insulin resistance and metabolic syndrome, early markers of heart disease." *American College of Cardiology*, 7 Mar. 2013, www.acc.org/about-acc/press-releases/2013/03/11/19/36/ptsd.
8. Anna, Boyajyan, et al. "Genetics of posttraumatic stress disorder — candidate genes and their implication in the disease-associated molecular pathomechanisms." *A Fresh Look at Anxiety Disorders*, 2015, doi.org/10.5772/60443.
9. Hardesty, Larry. "Explained: neural networks." *MIT News | Massachusetts Institute of Technology*, news.mit.edu/2017/explained-neural-networks-deep-learning-0414.
10. Hung, Chen-Ying, et al. "Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database." *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2017, July 2017, pp. 3110–13, doi.org/10.1109/EMBC.2017.8037515.
11. Byvatov, Evgeny, et al. "Comparison of support vector machine and artificial neural network systems for drug/nondrug classification." *Journal of Chemical Information and Computer Sciences*, vol. 43, no. 6, Nov. 2003, pp. 1882–89, doi.org/10.1021/ci0341161.
12. Najafabadi, Maryam M., et al. "Deep learning applications and challenges in big data analytics." *Journal of Big Data*, vol. 2, no. 1, Feb. 2015, p. 1, doi.org/10.1186/s40537-014-0007-7.
13. Bam, Marpe, et al. "Dysregulated immune system networks in war veterans with PTSD is an outcome of altered miRNA expression and DNA methylation." *Scientific Reports*, vol. 6, Aug. 2016, p. 31209, doi.org/10.1038/srep31209.
14. "Metabolism | definition, process, & biology." *Encyclopaedia Britannica*, www.britannica.com/science/metabolism.
15. "Human Neuropeptide Y (NPY) ELISA." *Millipore Sigma*, www.emdmillipore.com/US/en/product/Human-Neuropeptide-Y-NPY-ELISA,MM_NF-EZHNPY-25K.
16. Heilig, Markus. "The NPY system in stress, anxiety and depression." *Neuropeptides*, vol. 38, no. 4, Aug. 2004, pp. 213–24, doi.org/10.1016/j.npep.2004.05.002.
17. Caberlotto, Laura, et al. "Alterations in neuropeptide Y levels and Y1 binding sites in the flinders sensitive line rats, a genetic animal model of depression." *Neurosci-*

- ence Letters, vol. 265, no. 3, Apr. 1999, pp. 191–94, doi.org/10.1016/S0304-3940(99)00234-7.
18. “Insulin: function and types.” *Medical News Today*, 21 Nov. 2018, www.medicalnewstoday.com/articles/323760.
 19. Biscaldi, Lauren. “Insulin resistance, biological markers of PTSD increased in military veterans.” *Psychiatry Advisor*, 13 July 2017, www.psychiatryadvisor.com/home/topics/anxiety/ptsd-trauma-and-stressor-related/insulin-resistance-biological-markers-of-ptsd-increased-in-military-veterans/.
 20. Rapaport, Lisa. “Successful PTSD treatment tied to lowered diabetes risk | Physician’s Weekly.” *Physician’s Weekly*, www.physiciansweekly.com/successful-ptsd-treatment-tied/.
 21. “CFH gene.” *MedlinePlus Genetics*, medlineplus.gov/genetics/gene/cfh/.
 22. Hair, Pamela S., et al. “Complement regulator C4BP binds to staphylococcus aureus and decreases opsonization.” *Molecular Immunology*, vol. 50, no. 4, Apr. 2012, pp. 253–61, doi.org/10.1016/j.molimm.2012.01.010.
 23. MacFarlane, Leigh-Ann, and Paul R. Murphy. “MicroRNA: biogenesis, function and role in cancer.” *Current Genomics*, vol. 11, no. 7, Nov. 2010, pp. 537–61, doi.org/10.2174/138920210793175895.
 24. Kalluri, Raghu, and Valerie S. LeBleu. “The biology, function, and biomedical applications of exosomes.” *Science*, vol. 367, no. 6478, Feb. 2020, doi.org/10.1126/science.aau6977.
 25. Moore, Lisa D., et al. “DNA methylation and its basic function.” *Neuropsychopharmacology*, vol. 38, no. 1, Jan. 2013, pp. 23–38, doi.org/10.1038/npp.2012.112.
 26. Chierici, Marco, et al. “Integrative Network Fusion: A Multi-Omics Approach in Molecular Profiling.” *Frontiers in Oncology*, vol. 10, 2020, p. 1065, doi.org/10.3389/fonc.2020.01065.
 27. “3.1. Cross-validation: evaluating estimator performance.” *Scikit-Learn 0.24.2 Documentation*, scikit-learn.org/stable/modules/cross_validation.html.
 28. Ye, Andre. “Swish: booting ReLU from the activation function throne.” *Medium*, 28 Aug. 2020, towardsdatascience.com/swish-booting-relu-from-the-activation-function-throne-78f87e5ab6eb.
 29. “Activation Functions.” *ML Glossary Documentation*, ml-cheatsheet.readthedocs.io/en/latest/activation_functions.html.
 30. “Daiglelab / deep_ptsd_biomarkers / ptsd_cross_validation.py” *Bitbucket*, bitbucket.org/daiglelab/deep_ptsd_biomarkers/src/master/ptsd_cross_validation.py.
 31. Brownlee, Jason. “A gentle introduction to dropout for regularizing deep neural networks.” *Machine Learning Mastery*, 2 Dec. 2018, machinelearningmastery.com/dropout-for-regularizing-deep-neural-networks/.
 32. Brownlee, Jason. “Recommendations for deep learning neural network practitioners.” *Machine Learning Mastery*, 24 Feb. 2019, machinelearningmastery.com/recommendations-for-deep-learning-neural-network-practitioners/.
 33. “Disease Screening - Statistics Teaching Tools.” *New York State Department of Health*, www.health.ny.gov/diseases/chronic/discreen.htm.

Copyright: © 2022 Banerjee, Istanbouli, and Daigle. All JEI articles are distributed under the attribution non-commercial, no derivative license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>). This means that anyone is free to share, copy and distribute an unaltered article for non-commercial purposes provided the original author and source is credited.