

Use of Machine Learning to Re-Assess Patterns of Multivariate Functional Recovery after Fluid Percussion Injury: Operation Brain Trauma Therapy

Hannah Radabaugh,¹ Jerry Bonnell,² Odelia Schwartz,² Dilip Sarkar,² W. Dalton Dietrich,¹ and Helen M. Bramlett^{1,3}

Abstract

Traumatic brain injury (TBI) is a leading cause of death and disability. Yet, despite immense research efforts, treatment options remain elusive. Translational failures in TBI are often attributed to the heterogeneity of the TBI population and limited methods to capture these individual variabilities. Advances in machine learning (ML) have the potential to further personalized treatment strategies and better inform translational research. However, the use of ML has yet to be widely assessed in pre-clinical neurotrauma research, where data are strictly limited in subject number. To better establish ML's feasibility, we utilized the fluid percussion injury (FPI) portion of the rich, rat data set collected by Operation Brain Trauma Therapy (OBTT), which tested multiple pharmacological treatments. Previous work has provided confidence that both unsupervised and supervised ML techniques can uncover useful insights from this OBTT pre-clinical research data set. As a proof-of-concept, we aimed to better evaluate the multi-variate recovery profiles afforded by the administration of nine different experimental therapies. We assessed supervised pairwise classifiers trained on a pre-processed data set that incorporated metrics from four feature groups to determine their ability to correctly identify specific drug treatments. In all but one of the possible pairwise combinations of minocycline, levetiracetam, erythropoietin, nicotinamide, and amantadine, the baseline was outperformed by one or more supervised classifiers, the exception being nicotinamide versus amantadine. Further, when the same methods were employed to assess different doses of the same treatment, the ML classifiers had greater difficulty in understanding which treatment each sample received. Our data serve as a critical first step toward identifying optimal treatments for specific subgroups of samples that are dependent on factors such as types and severity of traumatic injuries, as well as informing the prediction of therapeutic combinations that may lead to greater treatment effects than individual therapies.

Keywords: data analysis; machine learning; pharmacotherapy; traumatic brain injury

Introduction

WITH NEARLY 10 MILLION NEW CASES being reported annually, traumatic brain injury (TBI) is a significant health-care concern that results in important morbidity and mortality, along with care-associated costs surpassing billions of dollars each year.^{1–3} Despite the alarming incidence, there are relatively few effective strategies for treating patients with TBI due to the heterogeneity in injury cause, location, severity, and comorbidities observed in the patient population.^{4–6} To address this heterogeneity, it must first be acknowledged that new strategies

are needed to comprehensively understand and harness the variability that exists in TBI data.

A signature observation regarding the TBI literature is that treatments that are effective in some cases fail in others. This is true in both animal and clinical TBI studies and the range of efficacy can be significant.^{7–10} Tools such as machine learning (ML) that are capable of capturing non-linear dependencies, interactions across variables, and other patterns that fail to conform to the assumptions made by univariate analyses may be better oriented to enhance our understanding of data sets with significant heterogeneity.¹¹ Current approaches often rely solely on standard univariate analytics (e.g.,

¹Department of Neurological Surgery, University of Miami Miller School of Medicine, Miami, Florida, USA.

²Department of Computer Science, University of Miami College of Arts and Sciences, Miami, Florida, USA.

³Bruce W. Carter Department of Veterans Affairs Medical Center, Miami, Florida, USA.

analysis of variance [ANOVA]) that prioritize one metric at a time and are susceptible to experimenter-imposed parameters as well as errors in hypothesis testing. In contrast, multivariate techniques can complement our current analytics through their ability to capture patterns between measures that would otherwise be missed.^{11,12}

Although promising, due to the limited availability of high-quality data, the application of ML in the TBI field remains underexplored.^{13–15} That is, important questions such as “Will the patient live?” and “What variables determine whether they will live?” have been addressed, but more complex queries such as the degree of recovery and the efficacy of treatment strategies have yet to be investigated.^{16,17} Furthermore, application of these techniques in the pre-clinical arena where rodent data are generally strictly limited in subject number is even more scarce, given that ML systems work at maximum efficiency when given large amounts of high-quality data.^{13,15} However, previous work by Nielson and colleagues¹⁸ has demonstrated that data-driven techniques able to explore the syndromic space can generate novel insights from pre-clinical neurotrauma data sets.

To address this important gap, our collaborative team utilized the fluid percussion injury (FPI) portion of the rich, rat data set amassed by Operation Brain Trauma Therapy (OBTT). OBTT is a consortium study that operates under a novel experimental design for drug screening.¹⁹ Employing researchers across multiple national research centers, OBTT investigators produced a data set that includes three injury models (i.e., diffuse, focal, and penetrating type injuries), 12 therapies with doses determined by successful reports in the literature, and diverse metrics of recovery (i.e., physiology, biomarkers, motor, cognition, and histopathology).¹⁹ Further, previous OBTT reports have presented the hypothesis that pharmacotherapies afford different functional recovery effects depending on the injury, leading us to analyze one injury type in these proof-of-concept experiments.¹⁹ In these studies, to explore the potential of ML we utilized nine treatments tested in the FPI model, which represents the focal/diffuse type of injury in the OBTT data set.²⁰

The treatments analyzed included erythropoietin (EPO), levitiracetam (LEV), minocycline (MIN), nicotinamide (NIC), and amantadine (AMT). Two dosing regimens were evaluated by OBTT for each of these therapies, except MIN, for which a single treatment protocol that included bolus plus infusion was used. OBTT selected these treatments and determined dosing regimens by identifying reports in the literature of previous successes in affording functional recovery and evidence that they targeted a broad range of potential secondary injury mechanisms to treat TBI-related deficits. From a technical perspective, these particular treatments were chosen due to the number of features available and completeness of their corresponding data sets. That is, because the FPI portion was the only data set that included physiology features and the corresponding data sets had no missing values, we chose to begin our proof-of-concept ML experiments specifically with this portion of the OBTT data set.

Our prior work using these data provided confidence that the application of ML to the OBTT pre-clinical TBI data set could yield fruitful insights.^{21,22} Based on these findings, both unsupervised and supervised methods of learning were applied and assessed for their ability to uncover informative patterns and correlations within the data set.²¹ By utilizing a t-SNE visualization map followed by the unsupervised clustering technique known as k-means, we showed that ML has the ability to identify treatment effects.²¹ Further, these treatment effects were not found in the univariate analyses performed in the original study. We also assessed the

ability of over 60 supervised learning classifiers to predict an end-study metric of recovery (i.e., probe performance) given features from the first 7 days post-injury (i.e., physiology, biomarkers, and motor). Observations from these studies further corroborated the potential effectiveness of treatments deemed as insignificant across all metrics in the original statistical analyses.

Therefore, based on our previous findings, this study aimed to further explore whether the administration of different pharmacotherapies afford different multi-variate profiles of functional recovery. We addressed the following questions: (1) “Can binary classifiers significantly outperform a baseline learner to identify which treatment each rat received given metrics from clinically relevant feature groups?” (i.e. physiology, biomarkers, motor, and cognition); (2) “Can the same classifiers discriminate between doses of the same therapy?”; and (3) “Does the ability to distinguish between treatments depend entirely upon an individual feature group?” Given the promising results from our prior studies, addressing these research questions could reveal a series of novel insights with pre-clinical, clinical, and technical implications at a time when advancing the TBI research field beyond the current therapeutic stalemate is imperative.

Methods

Data source

All data were collected by the researchers carrying out the studies using the FPI model in OBTT.^{19,23–25} Briefly, nine treatments with 10–15 rats per group were selected for analyses based upon their completeness and the inclusion of physiology features. These treatments included EPO (low dose; 5000 IU/kg), EPO (high dose; 10,000 IU/kg), LEV (low dose; 54 mg/kg), LEV (high dose; 170 mg/kg), MIN (30 mg/kg +2 mg/kg each hour for 72 h), NIC (low dose; 50 mg/kg 15 min and 24 h post-surgery), NIC (high dose; 500 mg/kg 15 min and 24 h post-surgery), AMT (low dose; 10 mg/kg 15 min and for 21 days post-surgery), and AMT (high dose; 45 mg/kg 15 min and for 21 days post-surgery). These doses were tested based on the published literature in which similar drugs were evaluated in TBI models. All studies that generated the data used in the current analysis were approved and supervised by the Institutional Animal Care and Use Committee of the University of Miami. From a computational perspective, this portion of the OBTT data set is tabular (flat file), complete (no missing values), and pure (minimal noise). Table 1 shows the measurements we determined to be appropriate for use in our ML models. Although a portion of these data are currently unpublished, the entirety of the data set will be made available for analysis in a public repository following the publication of the original studies by the OBTT investigators.

Complete surgical and laboratory methods have been previously described.^{19,20} Briefly, physiological metrics were recorded before and after the administration of a moderate FPI (2.0–2.2 atm). Biomarker assessments of UCH-L1 and glial fibrillary acidic protein (GFAP) serum levels were conducted at 4 h and 24 h post-FPI.²⁶ The exploratory cylinder task to assess forelimb use was conducted 7 days after surgery. The Morris water maze task to assess learning and memory was conducted over days 14–18 post-FPI and was followed by a probe trial to assess memory retention.

Data pre-processing and feature engineering

The complete list of features included as well as the processing steps that each metric underwent can be found in Table 1. Briefly, our curated data set included a total of 92 rat subjects and nine features across three feature groups. For all metrics, we applied

TABLE 1. COMPREHENSIVE DESCRIPTION OF THE CURATED DATA SET

Category	Metric	Engineering
Physiology (15 min before and after surgery)	PCO ₂	1. Post-surgery – pre-surgery 2. Z-score normalization
	PO ₂	1. Post-surgery – pre-surgery 2. Z-score normalization
	MAP	1. Post-surgery – pre-surgery 2. Z-score normalization
Motor, cylinder (before and 7 days post-injury)	# left forelimb placements	1. Post-injury – baseline 2. Z-score normalization
	# right forelimb placements	1. Post-injury – baseline 2. Z-score normalization
Biomarker (4 h and 24 h post-injury)	UCH-L1	1. 24 h – 4 h 2. Z-score normalization
	GFAP	1. 24 h – 4 h 2. Z-score normalization
Cognition (13–18 days post-injury)	Learning and memory	1. Day 1 latency average 2. Day 4 latency average 3. Day 1 – Day 4 4. Z-score normalization
	Memory retention (probe)	1. Percentage of time spent on target 2. Z-score normalization

The dataset utilized in each of the analyses includes four feature groups (column 1) and metrics (column 2) engineered from the raw data measures (column 3).

GFAP, glial fibrillary acidic protein; MAP, mean arterial pressure.

z-score normalization to avoid introducing arbitrary weights in the form of scaling. Feature engineering techniques included change over time, averaging performance across trials, and assessing the percentage of time spent successfully completing the assessment. These processing steps were applied where appropriate with the goal of accurately informing the machine of the relevant information provided by each metric without imposing redundancy (Table 1).

Machine learning

Supervised learning classifiers from the scikit-learn toolkit in Python were trained and evaluated for their performance on this data set.²⁷ The techniques chosen included a baseline dummy classifier (DUMB), classification and regression tree (CART), random forest (RF), K-Nearest Neighbors (KNN), linear discriminant analysis (LDA), multi-layer perceptron (MLP), naïve-bayes (NB), and support vector machines (SVM). The default settings suggested by the toolkit were utilized.

In these studies, we defined our input as metrics from all feature groups collected prior to euthanasia (i.e., physiology, biomarkers, motor, and cognition) to define our multi-variate, recovery space. We then aimed to understand if the administration of different therapies resulted in different underlying profiles in this space. Therefore, the output class is defined as administered therapy on the premise that, if the machine can accurately predict which treatment each sample received, identifiable patterns and correlations unique to each treatment must exist in this space. In total, the findings from more than 112 supervised learning classifiers are described in this article.

Pairwise comparisons

The ML classifiers were trained on 14 separate data sets that represent each of the potential combinations of different pharmacotherapies (Fig. 1). These classifiers were assessed for their ability to learn which treatment each subject received given the defined input metrics.

In each case where the comparison was between drugs of different identity, the higher dose was utilized when applicable. All pairwise comparisons of EPO, LEV, MIN, NIC, and AMT were assessed. Further, instances where two doses were available in the original data set were tested for their ability to be distinguished from each other.

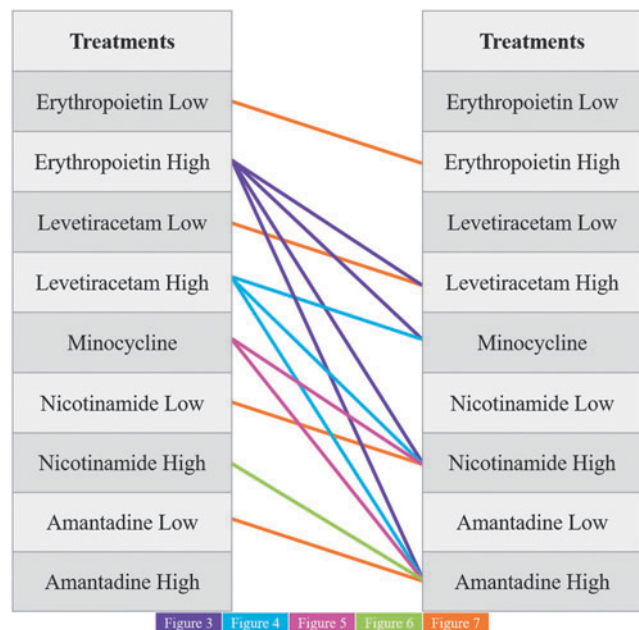


FIG. 1. An image visualizing each of the pairwise comparisons that were completed in the following analyses. Note that for each of the comparisons between pharmacotherapies of separate identity, the higher dose was used. The lower-dose data sets were only utilized for the intra-therapy comparisons where two doses were available (EPO, LEV, NIC, and AMT). AMT, amantadine; EPO, erythropoietin; LEV, levetiracetam; MIN, minocycline; NIC, nicotinamide.

		Actual class: as determined by reference	
		Actually positive (1)	Actually negative (0)
Predicted class: as called by algorithm	Predicted positive (1)	True positive	False positive (false alarm)
	Predicted negative (0)	False negative (missed case)	True negative

FIG. 2. The layout of a confusion matrix as well as the values associated with the precision, recall, and F1-score metrics of model performance.

Validation of models

To validate our ML classifiers, we used a stratified k-fold cross validation approach, with k set to 10. The data set was sectioned into 10 folds and the model was then trained with 9 of the folds and tested with the remaining 1 fold (i.e., the model was tested on examples it had never seen before). Further, each fold was used as the testing fold in separate runs. This was repeated until each fold had acted as the test fold and the mean performance \pm standard error reported with a 95% confidence interval. To eliminate the chance for model-initialization bias, cross-validation was repeated 20 times and the mean testing accuracy was reported. Estimation statistics were used in place of standard hypothesis testing and the subsequent generation of a *p*-value given that the core assumption of independence is violated in these multivariate analyses. Additionally, the stratified version of k-fold cross validation ensured that the balance of each class was represented equally across all folds, which is important when working with an imbalanced class attribute, so the model could learn patterns of all classes that it attempted to distinguish between.²⁷

Principal component analysis

To reveal the internal structure of the data in a way that best explains the variance throughout and determine the importance of employing multivariate analyses, we utilized the principal component analysis (PCA) algorithm from the scikit-learn library in Python. This analysis was applied to determine whether individual features or a combination of metrics were contributing substantially to the variance in our data set.

Leave one feature group out assessment

To determine whether the classifier's ability to distinguish between treatments was driven largely by individual feature groups, we repeated each of the first 10 comparisons with the individual feature groups removed and assessed the prediction accuracy of the model.

Reporting performance

Classification accuracy. The primary form of performance assessment reported is the classification (prediction) accuracy achieved. This was calculated as the number of correct predictions relative to the total number of attempts and was reported as a percentage.

Confusion matrices. In instances where prediction accuracy was not notably better than the baseline classifier, we further assessed model performance by visualizing its predictions in a confusion matrix, which comprises a table in which each section presents a unique value corresponding to correct positive and negative predictions as well as false-positives and false-negatives for each class identity (Fig. 2). This identifies whether a model is routinely mistaking one class for another.

Precision, recall, and F1 scores. In the instances where prediction accuracy failed to exceed baseline performance in any of the tested models, we report the precision, recall, and corresponding F1 score achieved by each classifier. The precision was calculated as the number of positive predictions divided by the number of predicted positive class values, thus providing a measure of exactness (a representative sample is visualized in corresponding confusion matrices). The recall, on the other hand, was calculated as the number of true positives divided by the number of false-negatives, thus providing a measure of completeness. These two values were then used to define the balance between the two metrics in the form of an F1 score.

Results

Many models were successful in discriminating between drugs of different identity

When comparing between drugs of different identities, we observed several interesting findings. In all experiments except one comparison, nearly all classifiers achieved high prediction accuracy, with the best performing classifiers for each comparison

TABLE 2. LIST OF VALUES CORRESPONDING TO THE BAR GRAPHS IN FIGURES 3–6

	<i>DUMB</i>	<i>RF</i>	<i>MLP</i>	<i>LDA</i>	<i>CART</i>	<i>NB</i>	<i>SVM</i>
EPO vs. LEV	0.50 \pm 0.21	0.88 \pm 0.12	0.87 \pm 0.14	0.92 \pm 0.11	0.73 \pm 0.21	0.87 \pm 0.14	0.85 \pm 0.15
EPO vs. MIN	0.53 \pm 0.28	1.0 \pm 0.00	0.95 \pm 0.10	0.90 \pm 0.14	0.93 \pm 0.9	0.92 \pm 0.11	0.92 \pm 0.11
EPO vs. NIC	0.40 \pm 0.24	0.70 \pm 0.22	0.80 \pm 0.16	0.70 \pm 0.22	0.90 \pm 0.13	0.65 \pm 0.21	0.60 \pm 0.20
EPO vs. AMT	0.55 \pm 0.23	0.82 \pm 0.21	0.88 \pm 0.12	0.78 \pm 0.18	0.58 \pm 0.18	0.68 \pm 0.20	0.70 \pm 0.20
LEV vs. MIN	0.50 \pm 0.19	0.93 \pm 0.10	1.0 \pm 0.00	0.92 \pm 0.11	0.92 \pm 0.11	0.95 \pm 0.10	0.90 \pm 0.20
LEV vs. NIC	0.60 \pm 0.20	0.92 \pm 0.11	0.87 \pm 0.14	0.85 \pm 0.15	0.88 \pm 0.12	0.92 \pm 0.11	0.90 \pm 0.13
LEV vs. AMT	0.53 \pm 0.28	0.62 \pm 0.21	0.83 \pm 0.15	0.95 \pm 0.10	0.67 \pm 0.21	0.73 \pm 0.15	0.81 \pm 0.17
MIN vs. NIC	0.53 \pm 0.21	0.88 \pm 0.12	0.90 \pm 0.13	0.92 \pm 0.11	0.85 \pm 0.12	0.87 \pm 0.14	0.90 \pm 0.13
MIN vs. AMT	0.53 \pm 0.19	0.85 \pm 0.13	0.88 \pm 0.13	0.90 \pm 0.13	0.83 \pm 0.15	0.88 \pm 0.13	0.98 \pm 0.05
NIC vs. AMT	0.58 \pm 0.18	0.73 \pm 0.18	0.47 \pm 0.19	0.72 \pm 0.16	0.52 \pm 0.26	0.52 \pm 0.19	0.73 \pm 0.15

Top performing classifiers are noted in bold for each pairwise comparison.

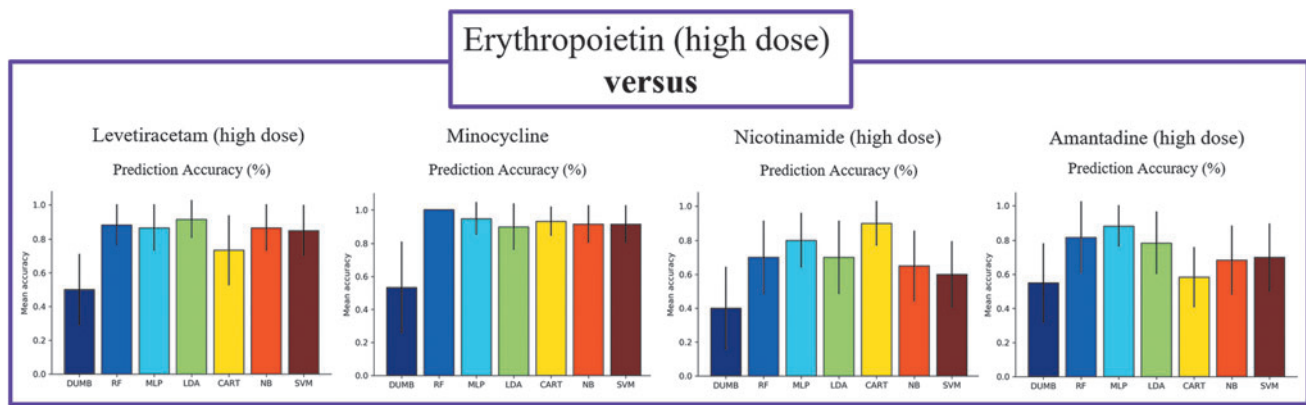


FIG. 3. Comparisons between the high dose of EPO and each of the other drugs tested in the analyses. The leftmost panel is EPO versus LEV. The next panel to the right is EPO versus MIN, the next panel is EPO versus NIC, and the final rightmost panel is EPO versus AMT. Average prediction accuracy across 20 runs with a stratified 10-fold cross validation are visualized. Error bars present a 95% confidence interval. AMT, amantadine; EPO, erythropoietin; LEV, levetiracetam; MIN, minocycline; NIC, nicotinamide.

ranging from 88% in the EPO versus AMT study to 100% in the LEV versus MIN study (Table 2, Figs. 3–5). This indicates that EPO, LEV, and MIN are all distinguishable from each of the other drugs in our analysis. This was not observed comparing NIC versus AMT, where the best performing classifier was the SVM with a 73% accuracy that was not notably better than the baseline (Table 2, Fig. 6). Furthermore, the LDA model was the best performing classifier in 3 of the 10 comparisons, making it the most successful model of the set followed by the RF, MLP, and SVM models, which each performed best in 2 of the 10 pairwise comparisons. Although there were many instances where multiple classifiers were successful in completing the assigned task, this was not always the case. For example, in the comparison between LEV and AMT, the LDA model appeared to perform better than the other classifiers, thus highlighting the importance of training multiple types of classifiers for each comparison.

Machine learning models tested experienced difficulty in distinguishing between doses of the same drug

When comparing doses of the same drug, we noted that the models in each study achieved lower prediction accuracies than those observed when distinguishing between different drugs. It is striking that these models did not outperform the baseline classifier in terms of prediction accuracy (Fig. 7; corresponding data shown in Supplementary Table S1).

When analyzing the precision, recall, and F1-scores for these comparisons, however, we noted that some models outperformed others in terms of exactness and completeness. Specifically, in the NIC dosing study, the MLP model achieved the highest values in terms of precision and recall despite not achieving the highest prediction accuracy. When comparing doses of AMT, closer examination revealed that the MLP model reported very low prediction and recall values relative to the other models. Utilizing these alternate forms of model assessment provides additional opportunities to compare models where prediction accuracy is relatively similar across all classifiers.

From a clinical perspective, however, we assessed our models for instances of imbalance between metrics. For example, if a classifier reported high precision but low recall, it could potentially report a low global accuracy score while limiting false-negatives. This would provide a unique advantage beyond global prediction accuracy in the clinical application of these techniques, where it is imperative to identify all positive cases. Of note, there were no instances where we observed this type of imbalance.

Findings reported required the application of multivariate techniques

The application of PCA highlighted a combination of individual metrics in each of the first four principal components, which accounted for 83.4% of the variance (Fig. 8, left panel). We further

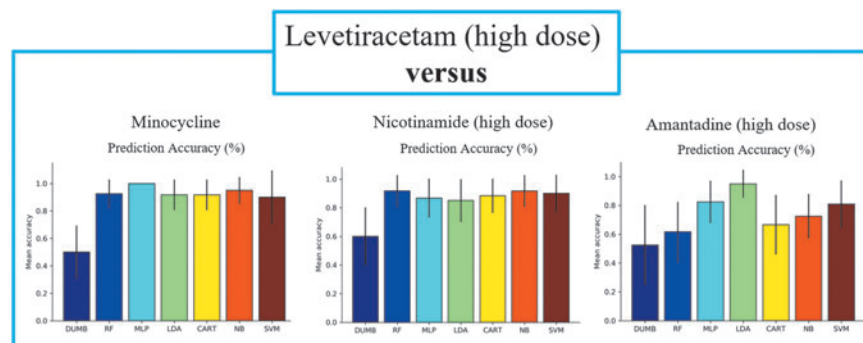


FIG. 4. Comparisons between the high dose of LEV and each of the other drugs tested in the analyses excluding the EPO comparison shown in Figure 3. Average prediction accuracy across 20 runs with a stratified 10-fold cross validation are visualized. Error bars present a 95% confidence interval. EPO, erythropoietin; LEV, levetiracetam.

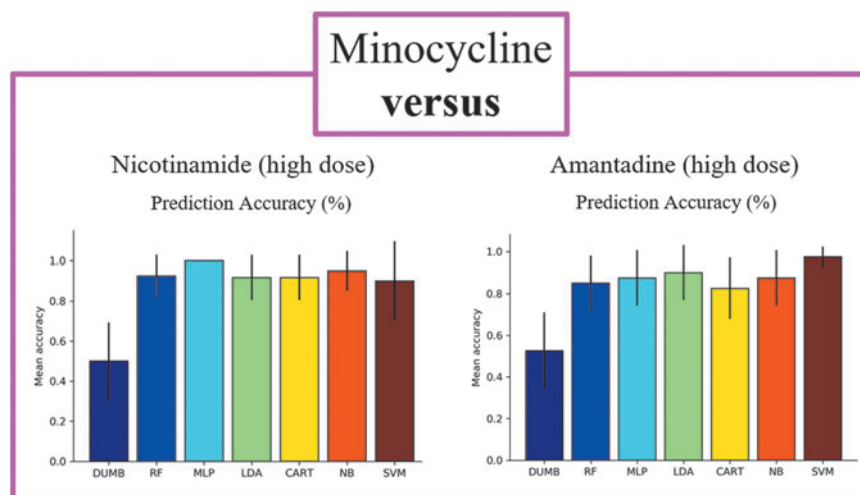


FIG. 5. Comparisons between the high dose MIN and each of the other drugs tested in the analyses excluding comparisons shown in the earlier figures. The top row of figures visualizes average prediction accuracy across 20 runs with a stratified 10-fold cross validation. Error bars present a 95% confidence interval. MIN, minocycline.

assessed the impact that each feature group has on model performance, by consistently removing a set of metrics associated with one feature group (e.g., physiology, biomarkers, motor and cognition). Each time that we removed a set of metrics, there was a notable reduction in prediction accuracy for every model included in the study. However, these reductions were not specific to a single feature group (Fig. 8, right panel). Figure 8 shows a representative image of these observations for the EPO versus LEV comparison.

Discussion

The research teams involved in the original OBTT consortium continue to report novel findings and amass a rich, rat data set.¹⁹ Here, we utilized a portion of these previously collected data to

explore functional recovery following TBI from a multi-variate perspective and assess what impacts the administration of several pharmacotherapies had on the patterns and correlations present within this high-dimensional space. The results presented here demonstrate promise from two major viewpoints. First, from a technical standpoint, and given our previously reported findings, we have demonstrated with confidence that ML methods typically reserved for data sets much larger than those seen in rat research can uncover fruitful insights in this setting. Further, our data have strong biological and potential clinical implications in their ability to highlight that different pharmacotherapies result in identifiably separate profiles. Whether the profiles identified by this exciting new approach reflect clinically meaningful differences in outcome clearly merits further exploration.

With technical implications in mind, to better understand the importance of their integration it is important to note how the methods we employed differ from and complement the conventional univariate analyses. Although TBI researchers are often well-trained in statistical applications, the field has lagged behind in its incorporation of new analytical methodologies developed in other data-driven fields such as ML. Briefly, ML and typical statistics feature considerable overlap in their goal of providing a mathematical description of a data set and its underlying implications. However, the two approaches differ in their incorporation of hypotheses.¹¹ Where statistics are hypothesis-driven, ML models tend to learn their hypotheses from the data during the training process. That is, statistical modeling begins with a series of assumptions regarding the distribution of the data and the parameters that likely gave rise to the observations. These hypotheses are then accepted or rejected in a predefined manner.

In ML, “hypothesis-free” testing focuses on prediction of new instances following a training period on past data without imposing experimenter-defined hypotheses or parameters.¹¹ Given the complexity of identifying patterns between metrics in a high-dimensional data set, however, these methods often fall short in that there is no direct explanation of what variables impacted the predictive capabilities gained during the training process. In this way, univariate statistics are strong in their ability to provide a clear picture of how subject groups differ from one another in each data

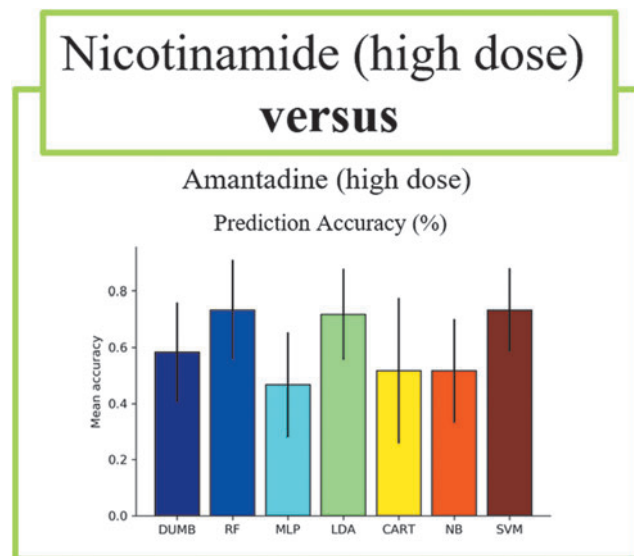


FIG. 6. Comparisons between high-dose NIC and high-dose AMT. Average prediction accuracy across 20 runs with a stratified 10-fold cross validation are visualized. Error bars present a 95% confidence interval. AMT, amantadine; NIC, nicotinamide.

Low Dose versus High Dose

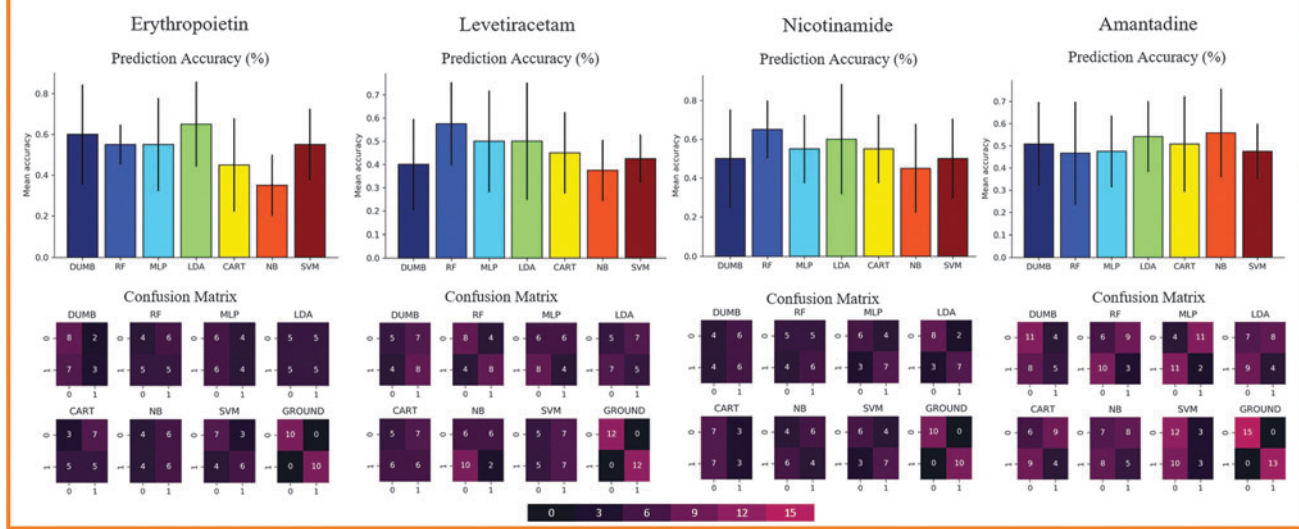


FIG. 7. Comparisons between the high and low dose of the same drug therapy for each treatment included in the analyses, excluding MIN. The top row of figures visualizes average prediction accuracy across 20 runs with a stratified 10-fold cross validation. Error bars present a 95% confidence interval. Bottom rows show a representative confusion matrix visualized with a heat map that corresponds to how many subjects fell into each category. MIN, minocycline.

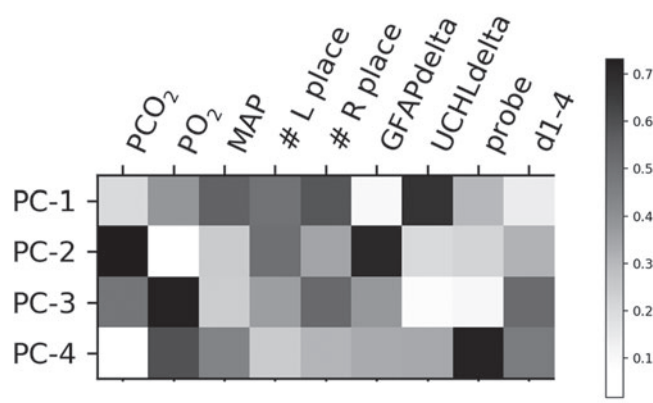
metric. Both multivariate and univariate metrics have the potential to provide useful insights in the TBI field and, when integrated, may better orient researchers to highlight the individual variabilities that will forge a path toward defining efficacious and/or optimal treatment strategies for each subject. This claim is supported by our findings, given the premise that exploring the high-dimensional functional recovery space yielded novel insights that would have not been discoverable without the application of multivariate analytics.

Regarding biological implications, the findings detailed above lead us to conclude that different therapies have distinct recovery profiles in the form of underlying, multivariate patterns and cor-

relations that allow for supervised classification models to identify which treatment each subject received. In the case of pairwise comparisons between drugs of completely different identity (e.g., EPO vs. LEV), these findings support our hypotheses.

Based on pathobiology, TBI presents a very complex combination of multiple underlying pathomechanisms that give rise to the observed deficits.²⁸ The multitude of mechanisms at work present many potential therapeutic targets and thus the literature reporting efficacious treatments includes many types of therapies. This lays the foundation of our hypothesis that addressing separate pathomechanisms may give rise to differential effects in the functional recovery space.

Principal component analysis



Leave one feature group out

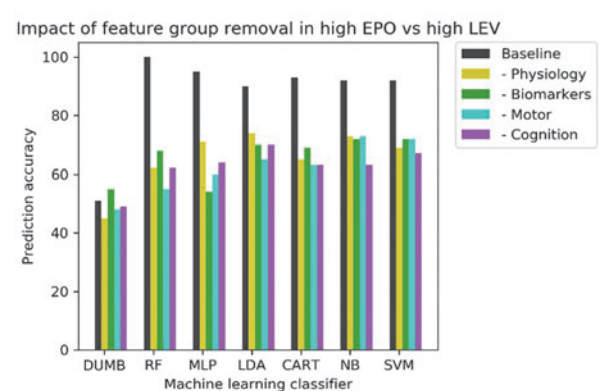


FIG. 8. The left panel visualizes a representative principal component analysis (PCA) showing the first four principle components (PCs) in our data set, which account for 83.4% of the variance. PCA gives rise to a complex mixture of original features and the heat map pictured corresponds to the relative impact of each feature in giving rise to the PC. The right panel shows a representative example of the impact on classification accuracy following the removal of individual feature groups. EPO, erythropoietin; GFAP, glial fibrillary acidic protein; LEV, levetiracetam; MAP, mean arterial pressure.

It is notable that our findings were less robust in the case of NIC versus AMT. Although each of the tested therapies in OBTT were carefully selected through a comprehensive literature search based on the quality of study design, methods of scientific rigor, and clinical relevance, the effective doses across models were not always consistent. That is, the mechanisms of action for the two therapies are purported to be vastly different. AMT therapy has been shown to augment dopamine systems and serve as a partial *N*-Methyl-D-aspartate (NMDA) receptor antagonist, whereas NIC acts on a number of core cellular processes such as the inhibition of poly adenosine diphosphate ribose (ADP-ribose) polymerase-1, the depletion of nicotinamide adenine dinucleotide phosphate (NADPH), and the potential amelioration of oxidative stress.^{9,29} However, in the case of FPI, both NIC and AMT have shown efficacy at doses higher than those used by OBTT—given that the dosing protocols were crafted prior to a more recent report.²⁹ Benefits were not observed post-FPI from AMT until a dose of 135 mg/kg was reached. However, the maximum dose in our data set was 45 mg/kg.¹⁹ Furthermore, similar findings are suggested for NIC.²⁹

Although many positive results have been achieved in the frame of the focal, controlled cortical impact (CCI) injury with the doses tested in OBTT, in FPI, there is only one study that reported modest functional benefits from the 500 mg/kg dosing regimen.^{30,31} Given these results, it is possible that NIC and AMT could not be distinguished from one another in our studies. Our ML studies thus support the likelihood that unique dosing regimens are needed to confer efficacy across models in pre-clinical studies, mirroring emerging approaches guiding precision medicine.^{31,32} Ongoing studies are exploring whether similar experiments using data from the CCI portion of the OBTT data set will echo what we have seen in the FPI data set or if new patterns will be observed with changing the injury model and severity.

In the experiments where we assessed the machine's ability to discriminate between two doses of the same drug, the classifiers exhibited difficulty with the task. Originally, we expected the machine to be successful in discriminating doses due to the fact that dosing is a critical component of optimal treatment administration.^{32,33} We expected the machine to identify the intricacies of dose-response better than typical univariate studies, given that ML methodologies have been previously applied to accomplish this task.³⁴ That is, in our studies, the effects afforded by different doses of the same drug do not appear to create differences in the multivariate space as distinguishable from one another, as do two completely different drugs. It is important to note that, although this is true for the therapies we examined, it may not be universally true in that certain drugs generate very different responses at different levels, which may lead to more robust multivariate pattern differences.³² A potential reason for the lack of differences observed in these experiments is that OBTT selected doses to maximize the chance of success, rather than examine the full spectrum of efficacy and toxicity.

Additionally, another limitation of our approach was that we did not include the results of histology in our analyses. It is possible that the inclusion of these data could enhance the capabilities of our approach; however, following this successful proof-of-concept study, ongoing work is assessing the impact of these pharmacotherapies in the other injury models. Given that experimental injury types are often classified by their histological phenotypes (e.g., focal vs. focal/diffuse vs. penetrating) and our team had already separated rats by specific injury type, we felt the addition of histology would introduce redundancy and make the interpretation of multivariate models less clear in the setting where the research

question is focused on assessing functional improvements. Lastly, we did not include all 12 therapies tested by OBTT, given that some therapies were missing values due to an inability to complete the biomarker analyses and also because entire data features were missing for some groups. In our ongoing studies assessing the other injury types in the OBTT data set and for other groups employing these metrics, however, missing values are an important topic of concern, and methods for handling deletions must be carefully selected in ways that minimize loss. Our group is currently assessing a multitude of ways to handle these instances to better inform which may be best suited for application in similar data sets.

In conclusion, ML methods demonstrate vast potential for their ability to provide fruitful insights on multiple fronts from rodent data sets. The current studies have the ability to serve as an informative initial step toward identifying optimal treatments for specific TBI subgroups of samples as well as informing the prediction of therapeutic combinations that may lead to additive and/or synergistic benefits. Furthermore, in ongoing work, we are devising ways to incorporate a larger number of therapies into a single model to transpose the architecture to a clinical setting. This study presents important proof-of-concept findings that can encourage other groups to incorporate these techniques into their studies as a complement to the typical univariate analyses. In the future, we will expand upon the impact of these techniques by comparing the results presented here to the application of the same methodologies on a second, unique data set. Our group has found success in the application of both unsupervised and supervised modeling supporting the effectiveness of these methods for the neurotrauma setting. The integration of these techniques into the TBI field is both timely given recent advances in ML and necessary given the lack of translational success achieved to date.

Acknowledgments

We sincerely thank Dr. Patrick Kochanek and Dr. Stefania Mondello for their valuable guidance throughout the course of these studies and each of the OBTT scientists for their hard work in collecting the rich data set used in these analyses. We also thank our collaborators Dr. Lauren Shapiro, Dr. Lucina Uddin, and Zsuzsanna Nemeth for their insight regarding the integration of computational techniques into brain injury medicine.

Funding Information

These studies were funded by the National Institutes of Health (R01 NS042133), two Department of Defense funding mechanisms (DAMD W81HWH-14-2-0118 and W81XWH-10-1-0623), the University of Miami U-LINK program for interdisciplinary research, and the University of Miami Maytag Graduate Fellowship.

Author Disclosure Statement

No competing financial interests exist.

Supplementary Material

Supplementary Table S1

References

- Hyder, A.A., Wunderlich, C.A., Puvanachandra, P., Gururaj, G., and Kobusingye, O.C. (2007). The impact of traumatic brain injuries: a global perspective. *NeuroRehabilitation* 22, 341–353.
- Max, W., Mackenzie, E., and Rice, D. (1991). Head injuries: costs and consequences. *J. Head Trauma Rehabil.* 6, 76–91.

3. Selassie, A.W., Zaloshnja, E., Langlois, J.A., Miller, T., Jones, P., and Steiner, C. (2008). Incidence of long-term disability following traumatic brain injury hospitalization, United States, 2003. *J. Head Trauma Rehabil.* 23, 123–131.
4. Bramlett, H.M., and Dietrich, W.D. (2015). Long-term consequences of traumatic brain injury: current status of potential mechanisms of injury and neurological outcomes. *J. Neurotrauma* 32, 1834–1848.
5. Sosin, D.M., Sniezek, J.E., and Thurman, D.J. (1996). Incidence of mild and moderate brain injury in the United States, 1991. *Brain Inj.* 10, 47–54.
6. Rogers, J.M., and Read, C.A. (2007). Psychiatric comorbidity following traumatic brain injury. *Brain Inj.* 21, 1321–1333.
7. Hammond, F.M., Sherer, M., Malec, J.F., Zafonte, R.D., Dikmen, S., Bogner, J., Bell, K.R., Barber, J., and Temkin, N. (2018). Amantadine did not positively impact cognition in chronic traumatic brain injury: a multi-site, randomized, controlled trial. *J. Neurotrauma* 35, 2298–2305.
8. Ghalaenovi, H., Fattahi, A., Koohpayehzadeh, J., Khodadost, M., Fatahi, N., Taheri, M., Azimi, A., Rohani, S., and Rahatlou, H. (2018). The effects of amantadine on traumatic brain injury outcome: a double-blind, randomized, controlled, clinical trial. *Brain Inj.* 32, 1050–1055.
9. Wang, T., Huang, X.J., Van, K.C., Went, G.T., Nguyen, J.T., and Lyeth, B.G. (2014). Amantadine improves cognitive outcome and increases neuronal survival after fluid percussion traumatic brain injury in rats. *J. Neurotrauma* 31, 370–377.
10. Ghate, P.S., Bhanage, A., Sarkar, H., and Katkar, A. (2018). Efficacy of amantadine in improving cognitive dysfunction in adults with severe traumatic brain injury in Indian population: a pilot study. *Asian J. Neurosurg.* 13, 647–650.
11. Valletta, J.J., Torney, C., Kings, M., Thornton, A., and Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Anim. Behav.* 124, 203–220.
12. Friedman, J.H. (2007). The role of statistics in the data revolution? *Int. Stat. Rev.* 69, 5–10.
13. Agoston, D.V., and Langford, D. (2017). Big Data in traumatic brain injury: promise and challenges. *Concussion* 2, CNC44.
14. Sakai, K., and Yamada, K. (2019). Machine learning studies on major brain diseases: 5-year trends of 2014–2018. *Jpn. J. Radiol.* 37, 34–72.
15. Haefeli, J., Ferguson, A.R., Bingham, D., Orr, A., Won, S.J., Lam, T.I., Shi, J., Hawley, S., Liu, J., Swanson, R.A., and Massa, S.M. (2017). A data-driven approach for evaluating multi-modal therapy in traumatic brain injury. *Sci. Rep.* 7, 1–10.
16. Rau, C.-S., Kuo, P.-J., Chien, P.-C., Huang, C.-Y., Hsieh, H.-Y., and Hsieh, C.-H. (2018). Mortality prediction in patients with isolated moderate and severe traumatic brain injury using machine learning models. *PLoS One* 13, e0207192.
17. Christie, S.A., Conroy, A.S., Callcut, R.A., Hubbard, A.E., and Cohen, M.J. (2019). Dynamic multi-outcome prediction after injury: applying adaptive machine learning for precision medicine in trauma. *PLoS One* 14, e0213836.
18. Nielson, J.L., Paquette, J., Liu, A.W., Guandique, C.F., Tovar, C.A., Inoue, T., Irvine, K.A., Gensel, J.C., Kloke, J., Petrossian, T.C., Lum, P.Y., Carlsson, G.E., Manley, G.T., Young, W., Beattie, M.S., Bresnahan, J.C., and Ferguson, A.R. (2015). Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury. *Nat. Commun.* 6, 8581.
19. Kochanek, P.M., Bramlett, H.M., Dixon, C.E., Dietrich, W.D., Mondello, S., Wang, K.K.W., Hayes, R.L., Lafrenaye, A., Povlishock, J.T., Tortella, F.C., Poloyac, S.M., Empey, P., and Shear, D.A. (2018). Operation Brain Trauma Therapy: 2016 update. *Mil. Med.* 183, 303–312.
20. Blaya, M.O., Bramlett, H.M., Naidoo, J., Pieper, A.A., and Dietrich, W.D. (2014). Neuroprotective efficacy of a proneurogenic compound after traumatic brain injury. *J. Neurotrauma* 31, 476–486.
21. Radabaugh, H., Bonnell, J., Nemeth, Z., Uddin, L., Shapiro, L., Sarkar, D., Schwartz, O., Dietrich, W.D., and Bramlett, H.M. (2019). Probing the Operation Brain Trauma Therapy dataset using machine learning techniques. *J. Neurotrauma* 36, DBB-24. <http://doi.org/10.1089/neu.2019.29100.abstracts> (Last accessed January 6, 2021).
22. Radabaugh, H.L., Bonnell, J., Dietrich, W.D., Bramlett, H.M., Schwartz, O., and Sarkar, D. (2020). Development and evaluation of machine learning models for recovery prediction after treatment for traumatic brain injury, in: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. Institute of Electrical and Electronics Engineers Inc., pps. 2416–2420.
23. Bramlett, H.M., Dietrich, W.D., Dixon, C.E., Shear, D.A., Schmid, K.E., Mondello, S., Wang, K.K.W., Hayes, R.L., Povlishock, J.T., Tortella, F.C., and Kochanek, P.M. (2016). Erythropoietin treatment in traumatic brain injury: Operation Brain Trauma Therapy. *J. Neurotrauma* 33, 538–552.
24. Browning, M., Shear, D.A., Bramlett, H.M., Dixon, C.E., Mondello, S., Schmid, K.E., Poloyac, S.M., Dietrich, W.D., Hayes, R.L., Wang, K.K.W., Povlishock, J.T., Tortella, F.C., and Kochanek, P.M. (2016). Levetiracetam treatment in traumatic brain injury: Operation Brain Trauma Therapy. *J. Neurotrauma* 33, 581–594.
25. Kochanek, P., Dietrich, W.D., Shear, D.A., Bramlett, H.M., Mondello, S., Lafrenaye, A., Wang, K.K.W., Hayes, R.L., Gilsdorf, J., Povlishock, J.T., Poloyac, S.M., Empey, P.E., and Dixon, C.E. (2019). Exploring additional approaches to therapy ranking in operation brain trauma therapy. *J. Neurotrauma* 36, DBA-09. <http://doi.org/10.1089/neu.2019.29100.abstracts> (Last accessed January 6, 2021).
26. Mondello, S., Shear, D.A., Bramlett, H.M., Dixon, C.E., Schmid, K.E., Dietrich, W.D., Wang, K.K.W., Hayes, R.L., Glushakova, O., Catania, M., Richieri, S.P., Povlishock, J.T., Tortella, F.C., and Kochanek, P.M. (2016). Insight into pre-clinical models of traumatic brain injury using circulating brain damage biomarkers: Operation Brain Trauma Therapy. *J. Neurotrauma* 33, 595–605.
27. Raschka, S., and Mirjalili, V. (2017). Chapter 4: Building Good Training Sets—Data preprocessing, in: *Python Machine Learning*, 2nd ed. Packt Publishing, pps. 104–107.
28. Kline, A.E., Leary, J.B., Radabaugh, H.L., Cheng, J.P., and Bondi, C.O. (2016). Combination therapies for neurobehavioral and cognitive recovery after experimental traumatic brain injury: is more better? *Prog. Neurobiol.* 142, 45–67.
29. Haar, C.V., Peterson, T.C., Martens, K.M., and Hoane, M.R. (2013). A possible pharmacological strategy for nerve diseases the use of nicotinamide as a treatment for experimental traumatic brain injury and stroke: a review and evaluation. *Clin. Pharmacol. Biopharm.* S1, 005.
30. Quigley, A., Tan, A.A., and Hoane, M.R. (2009). The effects of hypertonic saline and nicotinamide on sensorimotor and cognitive function following cortical contusion injury in the rat. *Brain Res.* 1304, 138–148.
31. Hoane, M.R., Akstulewicz, S.L., and Toppen, J. (2003). Treatment with vitamin B3 improves functional recovery and reduces GFAP expression following traumatic brain injury in rats. *J. Neurotrauma* 20, 1189–1199.
32. Nortje, J., and Menon, D.K. (2004). Traumatic brain injury: physiology, mechanisms, and outcome. *Curr. Opin. Neurol.* 17, 711–718.
33. Okigbo, A.A., Helkowski, M.S., Royes, B.J., Bleimeister, I.H., Lam, T.R., Bao, G.C., Cheng, J.P., Bondi, C.O., and Kline, A.E. (2019). Dose-dependent neurorestorative effects of amantadine after cortical impact injury. *Neurosci. Lett.* 694, 69–73.
34. Linden, A., Yarnold, P.R., and Nallamothu, B.K. (2016). Using machine learning to model dose–response relationships. *J. Eval. Clin. Pract.* 22, 856–863.

Address correspondence to:

Helen M. Bramlett, PhD

Department of Neurological Surgery

University of Miami

Miller School of Medicine

1095 NW 14th Terrace

Miami, FL 33136

USA

E-mail: hbramlett@med.miami.edu