



Evaluation of Machine Learning Techniques to Predict the Likelihood of Mental Health Conditions Following a First mTBI

Filip Dabek^{1,2}, Peter Hoover¹, Kendra Jorgensen-Wagers³, Tim Wu¹ and Jesus J. Caban^{1*}

¹ National Intrepid Center of Excellence (NICoE), Bethesda, MD, United States, ² Computer Science Department, University of Maryland, Baltimore, MD, United States, ³ Landstuhl Regional Medical Center, Landstuhl, Germany

OPEN ACCESS

Edited by:

Robert David Stevens,
Johns Hopkins University,
United States

Reviewed by:

Vangelis P. Oikonomou,
Centre for Research and Technology
Hellas (CERTH), Greece
Sezen Vatansever,
Icahn School of Medicine at Mount
Sinai, United States

*Correspondence:

Jesus J. Caban
jesus.j.caban.civ@mail.mil

Specialty section:

This article was submitted to
Neurotrauma,
a section of the journal
Frontiers in Neurology

Received: 02 September 2021

Accepted: 09 December 2021

Published: 02 February 2022

Citation:

Dabek F, Hoover P,
Jorgensen-Wagers K, Wu T and
Caban JJ (2022) Evaluation of
Machine Learning Techniques to
Predict the Likelihood of Mental Health
Conditions Following a First mTBI.
Front. Neurol. 12:769819.
doi: 10.3389/fneur.2021.769819

Objective: Limited research has evaluated the utility of machine learning models and longitudinal data from electronic health records (EHR) to forecast mental health outcomes following a traumatic brain injury (TBI). The objective of this study is to assess various data science and machine learning techniques and determine their efficacy in forecasting mental health (MH) conditions among active duty Service Members (SMs) following a first diagnosis of mild traumatic brain injury (mTBI).

Materials and Methods: Patient demographics and encounter metadata of 35,451 active duty SMs who have sustained an initial mTBI, as documented within the EHR, were obtained. All encounter records from a year prior and post index mTBI date were collected. Patient demographics, ICD-9-CM and ICD-10 codes, enhanced diagnostic related groups, and other risk factors estimated from the year prior to index mTBI were utilized to develop a feature vector representative of each patient. To embed temporal information into the feature vector, various window configurations were devised. Finally, the presence or absence of mental health conditions post mTBI index date were used as the outcomes variable for the models.

Results: When evaluating the machine learning models, neural network techniques showed the best overall performance in identifying patients with new or persistent mental health conditions post mTBI. Various window configurations were tested and results show that dividing the observation window into three distinct date windows [−365 : −30, −30 : 0, 0 : 14] provided the best performance. Overall, the models described in this paper identified the likelihood of developing MH conditions at [14 : 90] days post-mTBI with an accuracy of 88.2%, an AUC of 0.82, and AUC-PR of 0.66.

Discussion: Through the development and evaluation of different machine learning models we have validated the feasibility of designing algorithms to forecast the likelihood of developing mental health conditions after the first mTBI. Patient attributes including demographics, symptomatology, and other known risk factors proved to be effective features to employ when training ML models for mTBI patients. When patient attributes and features are estimated at different time window, the overall performance increase illustrating the importance of embedding temporal information into the models.

The addition of temporal information not only improved model performance, but also increased interpretability and clinical utility.

Conclusion: Predictive analytics can be a valuable tool for understanding the effects of mTBI, particularly when identifying those individuals at risk of negative outcomes. The translation of these models from retrospective study into real-world validation models is imperative in the mitigation of negative outcomes with appropriate and timely interventions.

Keywords: mild traumatic brain injury (mTBI), mental health, machine learning, data science, predictive modeling, forecasting

1. INTRODUCTION

Previous estimates suggest that 15–22% of all Service Members (SMs) have sustained a mild traumatic brain injury (mTBI) (1, 2). Many SMs develop persistent symptoms such as headaches, sleep disturbances, cognitive deficits, as well as changes in mood and behavior (3–5). These symptoms can often be further compounded due to the environment in which these injuries were sustained, particularly among combat-related mTBIs in which co-morbid conditions such as posttraumatic stress disorder (PTSD) and other mental health (MH) conditions are prevalent (6–8). As such, SMs and Veterans who have sustained an mTBI are at greater risk for developing MH conditions (8).

Despite best practices for treating SMs and Veterans with mTBI (9), prognosticating recovery from mTBI remains challenging. Multiple factors can affect symptom progression, including characteristics of the injury itself (10–12), as well as premorbid health conditions (10, 13, 14). Research has shown that the presence of co-morbid conditions, particularly mental health conditions, can often delay and further complicate the recovery of mTBI (15, 16). As such, the ability to identify individuals at risk for developing certain symptoms and anticipating their needs can greatly improve an individuals' outcome. Having the appropriate and timely interventions can assist in a faster resolution of symptoms and a quicker recovery (13).

In the field of data science for healthcare, the development of predictive models focuses on creating algorithms and devising methods that can be used to predict the likelihood of occurrence or recurrence of a particular event, such as a specific clinical diagnosis or other negative outcomes. During the last decade there has been a significant increase in popularity of developing tools for predicting outcomes at the level of the individual patient (17). Unfortunately, most models rely on regression techniques which result in a regression formula that often simplifies the forecast into a general risk score factor. Regression techniques often are sensitive to outliers and the data must be independent. Given those limitations and the continued growth of standardized clinical data for mTBI patients, additional research is needed to assess the applicability of advanced machine learning (ML) techniques into predicting the probability of obtaining a mental health diagnosis.

This paper evaluates various methods and techniques in an attempt to anticipate the outcomes of mTBI. By leveraging machine learning models and longitudinal EHR data, we hope to assist clinicians in identifying those individuals at risk for negative outcomes.

2. BACKGROUND AND SIGNIFICANCE

Healthcare providers have long relied on data to understand the conditions and prognosis of a patient. The increased availability of health-related digital data has allowed for the exploration and evaluation of clinically-relevant issues on a much broader scale. Researchers and clinicians are no longer limited by finite sample sizes, often a constraint in many traditional cohort studies. The ability to analyze larger sample populations enables for more generalized results, due to diverse populations and settings, along with improved statistical power (18, 19). This has important implications when devising policies and establishing evidence-based clinical practices (20).

Through the widespread adoption of EHR systems, the rapid growth of healthcare data, and the steps taken by many hospital organizations to integrate different analytical tools within their clinical workflow; providers and administrators can now have greater insight into the etiologies of disease and subsequent outcomes (21). Currently, many analytical tools that have been integrated into the clinical workflow and are used to forecast specific clinical events such as hospital readmission, cardiovascular conditions, cost of the patient, fraud, and negative outcomes within well-defined conditions (22, 23).

Within TBI research, the development of such forecasting tools have furthered our understanding of the various clinical presentations of concussive events and have helped with the evaluation of treatment efficacy (24). However, much existing research has focused on estimating the importance of *risk factors* and their association with particular outcomes (25–27).

Through the utilization of longitudinal EHR data and the advancement of data science, we now have the ability to expand upon previous techniques and incorporate a richer set of clinical characteristics. Recently, there has been an increased popularity in the development of these clinical predictive models to assist clinicians in anticipating outcomes for an individual patient (28, 29). Data science enables the use of structured and unstructured

EHR data to develop a large set of features that models can use to perform classifications tasks (17). This increased flexibility is particularly important when considering the temporal aspect of events given that the time and sequence of events can play a vital role when evaluating patient outcomes. Recent research has stressed the importance of temporal events in predictive analytics (30, 31). Furthermore, data science techniques such as feature selection techniques can estimate and even select particular variables based upon predictive power. Dependent upon the specific task and its intention, the relevancy of features can greatly vary, and we can rely upon machine learning and data science techniques to assist in this evaluation.

As the quantity and complexity of the healthcare data collected for patients continues to increase, healthcare providers must embrace modern techniques that can assist with understanding a patients' condition and help evaluate the likelihood of various outcomes. This paper evaluates the utility of machine learning models and longitudinal EHR data as a mechanism to assist clinicians in identifying individuals at risk for developing mental health conditions. Patient demographics, encounter metadata, and data science algorithms are used to create clinical prediction models that can provide insight into potential needs of the patients. Throughout the paper, various supervised machine learning models were employed including logistic regression, support vector machines (SVMs), and neural network (NN) to develop predictive models capable of identifying patients at increased risk. In addition, multi-dimensional feature vectors and optimized observation windows were used to assist providers at identifying patients at different time intervals. This work serves as the foundation of some prospective work planned to validate the use of data science in TBI.

3. OBJECTIVE

The objective of this study was to evaluate the utility of machine learning models and longitudinal EHR data to predict the likelihood of developing mental health (MH) conditions following the first diagnosis of mTBI. Electronic health record metadata, healthcare utilization, and preexisting conditions were utilized to generate a unique description of every patient, train different models, and perform evaluation of different approaches to determine the likelihood of developing mental health condition 1 year following injury.

4. MATERIALS AND METHODS

4.1. Data Source

Direct care records were accessed through the Comprehensive Ambulatory Provider Encounter Record (CAPER) data file within the Military Health System Data Repository (MDR). For each encounter, metadata was extracted, which included International Classification of Diseases (ICD-9 & 10) diagnostic codes, provider types, procedural terminology codes, clinic codes, and demographics.

As a retrospective study, a waiver of documentation of informed consent was requested and approved for this study by

the Walter Reed Institutional Review Board under IRB protocol #374953. All identifiable data was removed prior to analysis.

4.2. Sample Population

Our sample population included active duty SMs who had sustained an mTBI, as diagnosed by a healthcare provider adhering to VA/DoD criteria (9). In order to qualify for the study, each patient's direct care encounters had to be documented in the military EHR between 2005 and 2018. The earliest date of mTBI diagnosis for each patient was defined as the index date.

To guarantee a complete longitudinal dataset, only SMs with encounter data greater than 1 year prior to and 1 year post-injury were included. To ensure that the date of diagnosis accurately reflects the date of injury, patients were excluded whose initial mTBI was defined as a *personal history* of TBI (e.g., ICD-10-CM code Z87.820, ICD-9-CM code V15.52), a diagnostic category used to reflect a previous TBI regardless of when it occurred. Furthermore, to establish that results were not confounded by subsequent injuries, patients were excluded who had a more severe diagnosis of TBI up to 1 year after initial injury date.

The dataset consisted of 35,451 active duty SMs whose first mTBI was documented by a healthcare provider within the MHS, initially amounting to 4,901,840 direct care outpatient clinical encounters. Utilizing only encounters 365 days before and after mTBI diagnosis date, 1,369,740 encounter records were assessed. **Figure 1** illustrates the process and inclusion criteria for determining the sample population.

4.3. Feature Vectors

In data science and machine learning, *features* are individual independent variables directly extracted or derived from the raw data that are used to describe the unique characteristics of the object under consideration. A *feature vector* is the n-dimensional vector of independent features that is used to describe a given object.

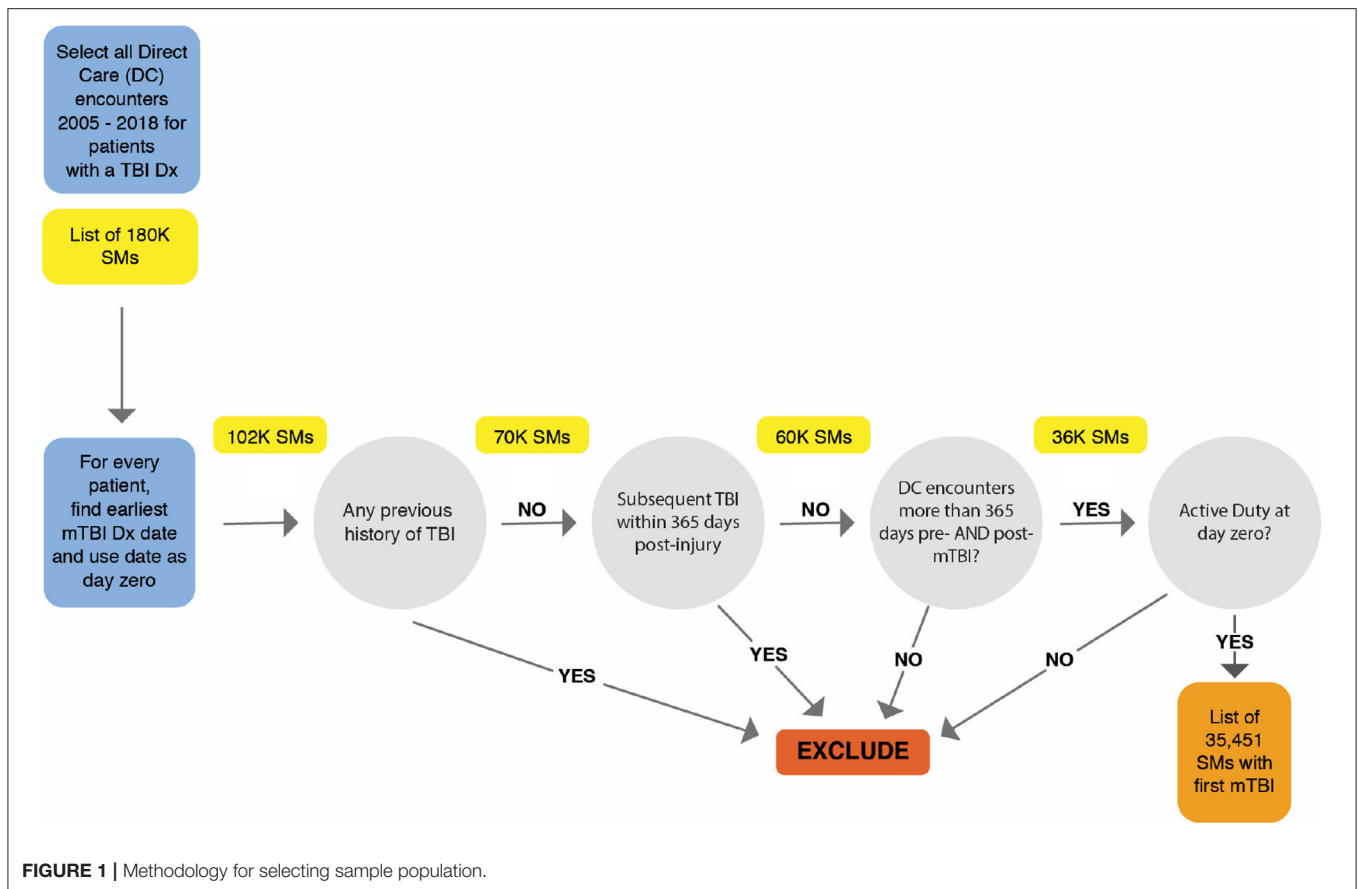
One of the key components of designing ML models is using domain knowledge to extract features. For this particular effort, with our sample population defined, the attributes that were included within the final feature vectors can be categorized into three domains: *demographics*, *symptomatology*, and *other risk factors*. Those feature vectors were generated only from the data from the observation period.

4.3.1. Demographics Features

Although various demographic attributes were available within the EHR, due to collinearity among variables, our final dataset included demographics specific to age, gender, and branch of service (BoS). Existing literature has supported the utility of BoS, as different branches experience, diagnose, and/or treat TBIs differently (32).

4.3.2. Symptomatology

To indicate and quantify preexisting conditions, ICD-9-CM and ICD-10 codes associated with common mTBI symptoms and complaints (33–35), were obtained. These codes were then grouped into Enhanced Diagnostic Related Groups (EDRGs) based on related symptomatology. For example, the Anxiety



EDRG included codes such as “300.02-Generalized Anxiety Disorder,” “300.01-Panic Disorder,” and “F41.9-Anxiety Disorder, unspecified.” To support the grouping and classification process, the Expanded Diagnostic Clusters (EDC) used in the John Hopkins Adjusted Clinical Groups (ACG) model (36), were employed. To evaluate symptom presence, EDRGs were then mapped to related encounter-level diagnostic codes and the frequency for each EDRG was then computed. Encounter records within the year prior to index mTBI were utilized to obtain these counts.

EDRGs related to mental health (e.g., anxiety, depression, adjustment disorder, etc.) were extracted from encounter records within the predictive timeframe to define the outcome variable of presence or not of MH conditions following mTBI. For a complete list of diagnosis codes, see **Supplementary Table 1**.

4.3.3. Risk Factors Features

A critical component of feature engineering is the process of using domain knowledge to extract features. Previous literature has identified key risk factors for developing mental health conditions (37). As a result, diagnosis codes specific to suicide attempts/ideation and substance abuse disorders were obtained from the Clinical Classifications Software (CCS) mapping created by the Agency for Healthcare Research and Quality (AHRQ) (38, 39). Encounter records within the year prior to index mTBI were used to identify those with preexisting risk factors. The

frequency of these risk factors were then quantified and included as part of the feature vector.

Once the different sub-components of our feature vectors were generated, they were combined into an n-dimensional feature vector:

$$P_1 = \{D_1, D_2, D_3, \dots\} \cup \{S_1, S_2, S_3, \dots\} \cup \{R_1, R_2, R_3, \dots\}$$

where P_1 is a sample patient, D_i the features associated with demographics, S_i features associated with symptomatology, and R_i features associated with risk factors. For a complete list of the different codes that were used for the different EDRGs, see **Supplementary Table 2**.

4.4. Window Configurations

As discussed, we recognize the temporal significance of preexisting symptoms and the specific sequence these conditions happen. Conditions present 1 year pre-mTBI might hold different clinical relevance than a condition present at time of injury. For a model to better learn of these temporal associations and identify their significance, we sought to build and devise various observation window configurations.

Based upon clinical knowledge, we first divided the entire observation period into distinct intervals, referring to them as “window configurations.” In the past, we had found that 1-month (30-day) intervals were most effective for splitting EHR

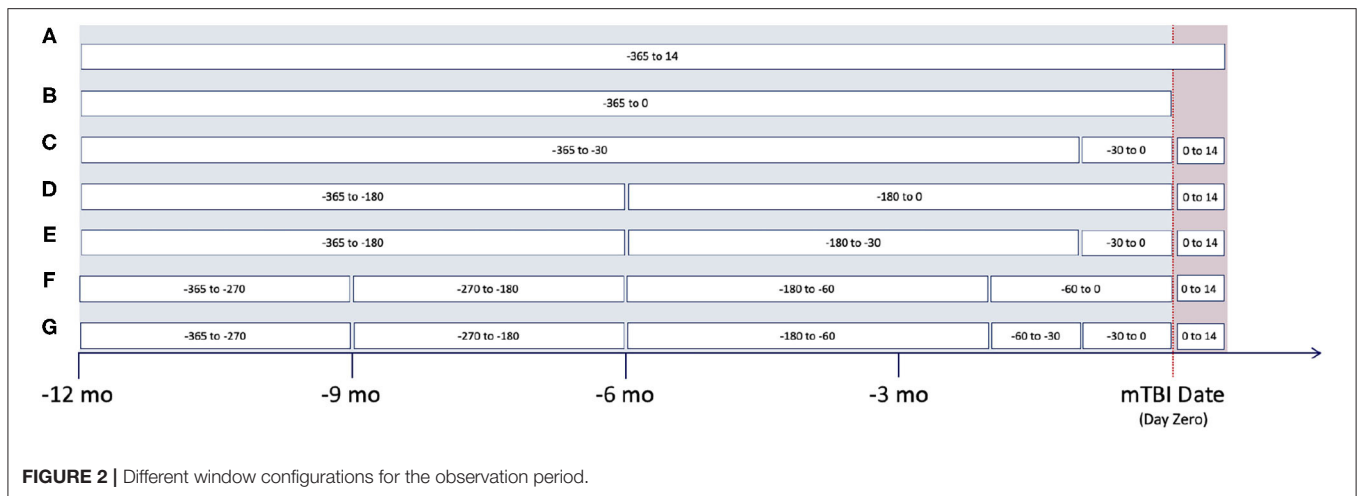


FIGURE 2 | Different window configurations for the observation period.

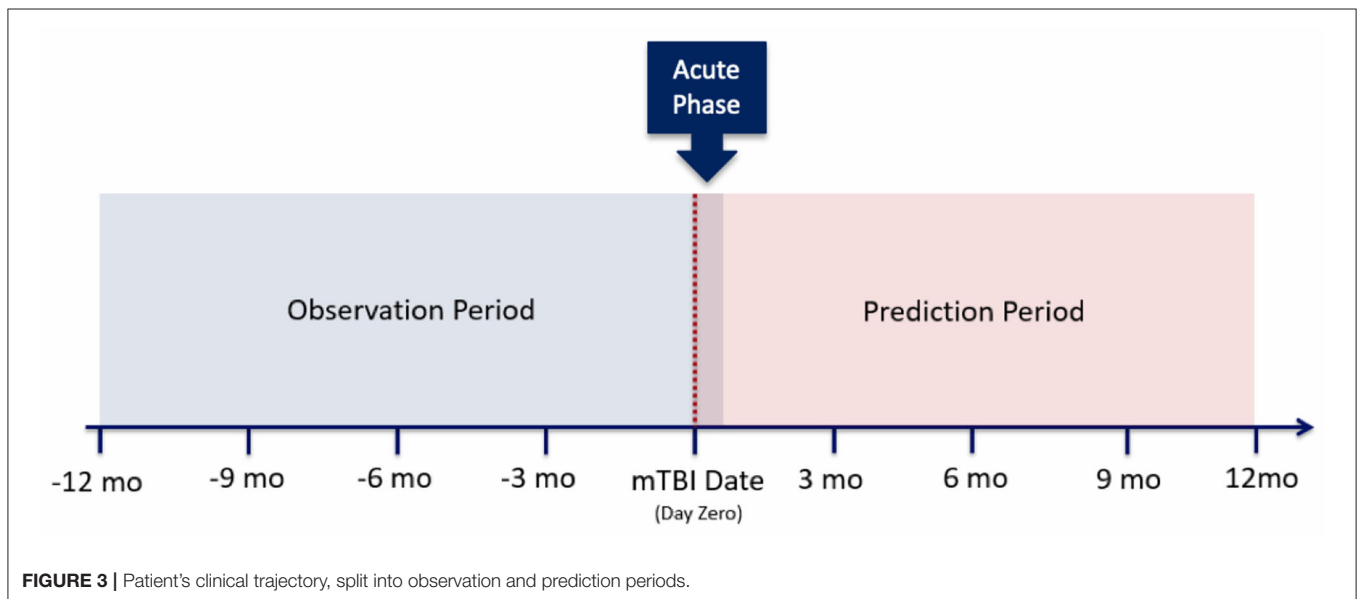


FIGURE 3 | Patient's clinical trajectory, split into observation and prediction periods.

data (40). However, as historical EHR data can be relatively sparse, various other window configurations were created. These adaptive configurations grouped observation windows with sparse clinical data, typically those time periods at greatest distance from mTBI date. **Figure 2** provides an illustration of the various window configurations, A through G, that we employed for this particular project.

As noted within **Figure 2**, several of these configurations utilize the first 14 days post-mTBI, often referred to as the “acute phase.” We chose to include this window as the first 14 days post-TBI are considered the crucial period of treatment for understanding the recovery and long term trajectory of a patient (41). Therefore, our models leveraged data between $[-365 : 14]$ days, with respect to mTBI date, for the prediction of events between $[14 : 365]$ (**Figure 3**).

It's important to note that by adding different configuration windows, the feature vector now is estimated for each window, thus resulting in vectors like

$$P_1 = \{D_1, D_2, D_3, \dots\} \cup \{S_1^{w1}, S_2^{w1}, S_3^{w1}, \dots\} \cup \{S_1^{w2}, S_2^{w2}, S_3^{w2}, \dots\} \cup \dots \cup \{R_1^{w1}, R_2^{w1}, R_3^{w1}, \dots\} \cup \{R_1^{w2}, R_2^{w2}, R_3^{w2}, \dots\}$$

where P_1 is a sample patient, D_i the features associated with demographics, S_i^{w1} features associated with symptomatology for window 1, and R_i^{w2} features associated with risk factors for window 2.

As detailed above, each patient within the dataset had attributes specific to three main categories (demographics, risk factors, and symptoms). Demographics included three distinct attributes: age, gender, and branch of service. Risk factors included two attributes: substance abuse and suicide attempts/ideation. Lastly, symptoms included 14 attributes, each representing a distinct symptom. As such, each patient had 19 attributes by default. The number of attributes may increase depending upon the number of window configurations used

within the model, as symptoms and risk factors were calculated with respect to the observation windows. For example, if three observation windows were used (Window Configuration C), patients would then have 51 attributes; 14 symptoms and two risk factors for each individual window.

4.5. Models and Metrics

For this study, various supervised machine learning models were employed. Supervised algorithms are those that learn from “labeled” data and once trained are tested and used to predict the classification of “unlabeled” data. As each model type has its own benefits and limitations, it is important to evaluate and compare different models and their performances. For completeness, logistic regression, support vector machines (SVM), random forests, elastic nets, adaptive boosting (AdaBoost), and neural networks were all employed. To compensate for class imbalance, Synthetic Minority Over-sampling Technique (SMOTE) (42), was also employed on the training set. These techniques were used in combination to the different model types.

To test and perform a thorough evaluation of the different models, we split the original dataset into two groups using a 80:20 split: *training* and *testing*. The *training* set was then split (training/validation) to be used as input variables to create the model and estimate parameters. The *testing* dataset was used to validate, compare, and optimize the different models.

To assist in the tuning of hyper-parameters among the varying models, grid search was employed on the training data for the non-neural network approaches. For the neural networks, the Adam optimizer was used along with its default learning rate. The network consisted of an input with its dimension in respect to the size of the corresponding window size, a hidden layer of size 19 with a tanh activation function, and then a single output node that used a sigmoid activation function. The network was trained using the binary cross entropy loss.

The receiver operating characteristic (ROC) curve, the area under curve (AUC), and the area under the precision-recall curve (AUC-PR) were utilized for comparison and evaluation. The ROC curve is created by plotting the true positive rate (i.e., probability that an actual positive case will test positive) against the false positive rate (i.e., when the truth is negative, but the model predicts a positive) at various threshold settings. The precision-recall curve summarizes the trade-off between optimizing for precision (i.e., positive predictive value) against recall (i.e., sensitivity). As it is difficult for a model to perform well under both metrics, the precision and recall values are often inversely proportionate.

The AUC-PR metric summarizes this tradeoff into a singular value for easy model comparison. Of note, the AUC-PR metric is a different measurement than the AUC, has been found to be more informative in the context of imbalanced binary classification tasks (43, 44), and will typically have smaller values ranging from about 0.2 to 0.5 (45, 46).

The utilization of ROC, AUC, and AUC-PR is important to validate models with class imbalance limitations—when the dataset has an unequal distribution of classes in the training dataset. Class imbalance is common in healthcare given that often

TABLE 1 | Summary of patient demographics and prevalence of preexisting conditions ($n = 35,451$).

		<i>n</i>	%
Gender	Male	29,736	83.9
	Female	5,715	16.1
Age	17–24	16,890	47.6
	25–34	12,816	36.2
	35–44	4,665	13.2
	45+	1,080	3.0
Service branch	Army	19,778	55.8
	Air force	6,204	17.5
	Marine corps	4,396	12.4
	Navy	4,491	12.7
	Other	582	1.6
Rank	Cadet	1,148	3.2
	Enlisted, Junior	17,988	50.7
	Enlisted, Senior	12,401	35
	Officer, Junior	2,131	6.0
	Officer, Senior	1,021	2.9
	Officer, Warrant	322	0.9
	Unknown	440	1.2
Preexisting conditions	Anxiety	2,551	7.2
	Appetite	3,005	8.5
	Audiology	1,745	4.9
	Balance/Dizziness	1,289	3.6
	Cognitive	1,972	5.6
	Depression	2,401	6.8
	Fatigue	828	2.3
	Headaches	7,871	22.2
	Musculoskeletal	21,531	60.7
	Neurology	3,338	9.4
	Psychology, Other	7,201	20.3
	PTSD	1,955	5.5
	Sleep	3,718	10.5
	Substance abuse	2,008	5.6
Suicide ideation/Attempt	324	0.9	
Vision	1,160	3.3	

there will be a majority of cases that are negative and a minority group that are positive for a particular condition.

For the development of feature vectors and the construction of models, Python 3.6.6 was used (47). Accompanying packages including Pandas 0.24.4, Numpy 1.15, and Imbalanced-learn 0.4.3 were used for the manipulation, cleaning, and resampling of data (48–50). Scikit-learn 0.20.1, Keras 2.2.4, SciPy 1.1.0, and Matplotlib 3.4.1 were utilized to create models and evaluate performance (51–54).

5. RESULTS

5.1. Demographics

For this analysis, a total of 35,451 SMs met the inclusion/exclusion criteria (Table 1). This resulted in a sample population of 29,736 men (83.9%), with the majority between the

TABLE 2 | Proportion of service members with mental health conditions.

	Remitting	New onset	Persistent	Not diagnosed
Anxiety	1,105 (3.1%)	2,141 (6%)	1,446 (4.1%)	30,759 (86.8%)
Depression	1,117 (3.2%)	1,959 (5.5%)	1,284 (3.6%)	31,091 (87.7%)
Psychology, Other	3,256 (9.2%)	4,786 (13.5%)	3,945 (11.1%)	23,464 (66.2%)
PTSD	548 (1.5%)	1,536 (4.3%)	1,407 (4%)	31,960 (90.2%)
Substance abuse	1,037 (2.9%)	1,660 (4.7%)	971 (2.7%)	31,783 (89.7%)
Suicide ideation/Attempt	286 (0.8%)	401 (1.1%)	38 (0.1%)	34,726 (98.0%)

Remitting, conditions present only prior to mTBI; New Onset, newly diagnosed conditions; Persistent, conditions present before and after mTBI.

TABLE 3 | Feature importance derived from model performance with iterative addition of features where D are demographics, R are risk factors, and S are symptoms.

Feature	Logistic regression			SVM			Neural network		
	ACC	AUC	AUC-PR	ACC	AUC	AUC-PR	ACC	AUC	AUC-PR
D	68.0	0.51	0.33	67.7	0.5	0.32	68.0	0.61	0.40
D + R	69.6	0.53	0.36	69.8	0.54	0.36	69.8	0.65	0.49
D + R + S	74.9	0.62	0.46	76.8	0.68	0.50	77.0	0.75	0.65

ACC, Accuracy; AUC, Area Under Curve; AUC-PR, Area Under Curve Precision Recall.

ages of 17 and 24 (47.6%). Examining military characteristics, over half of the sample population was from the United States Army (55.4%), while 50.7% were classified as Junior Enlisted. Chi-square was performed to validate that sample population is representative of the active duty TBI population.

In the evaluation of preexisting conditions, musculoskeletal were most prominent among this population, (60.7%), followed by headaches (22.2%). Examining the MH-related conditions, 6.8% of the population were diagnosed with depression with 20.3% diagnosed with a Psychological, Other condition. Regarding the outcome variable, 32.1% of the sample population were diagnosed with at least one MH condition within the year following the index mTBI (Table 2).

5.2. Feature Importance Ranking

In machine learning, feature importance ranking is the process that measures the contribution of individual features into the model. As part of the initial model development stages, we first evaluated the impact of embedding different feature vectors. We trained models to predict the year post-mTBI [14 : 365] and incrementally added features (e.g., demographics, symptoms, and risk factors). Table 3 contains the results for models (SVM, neural network), with the addition of the logistic regression as a baseline comparison. Assessing the logistic regression model, for example, utilizing only patient demographics resulted in an accuracy of 68.0%, an AUC of 0.51, and an AUC-PR of 0.33. By adding risk factors as input features, we notice an improvement in model performance; as noted by small increase in accuracy, AUC, and AUC-PR. However, when adding the patient symptomatology to the models, performance significantly improved to 74.9%. Through the iterative addition of features, the incremental predictive power of each could be

seen through the boosts in performance metrics across these model types.

5.3. Model Types and Variations

With the inclusion of all feature vectors, we evaluated the performance of the different model types and variations previously outlined. The model types varied and we found the best performing for comparison to be logistic regression, SVM, and neural network models. Table 3 details the original results utilizing the full feature vectors (demographics, risk factors, symptoms) against these models. With these models, we attempted to employ SMOTE to assist with the class imbalance. However, improvement in model performances were not observed.

First, we can see that after using all the features logistic regression performs with a 74.9% accuracy and 0.46 AUC-PR. By leveraging more advanced techniques such as SVMs, we can see an improvement to 76.8% accuracy and the 0.50 AUC-PR. Finally, when we apply state-of-the-art techniques like neural networks we can see an improvement to 77% accuracy and 0.65 AUC-PR.

With regards to computational costs, the logistic regression and SVM were consistently trained within 13–14 s, irrespective of the number of observation windows. AdaBoost and the random forest trained within 1 s while the neural network took approximately 23 s to train a single observational window (window configuration A). When training the maximum number of windows, the neural network took 27.4 s.

5.4. Observation Window Configurations

To evaluate the effectiveness of window configurations and its impact on performance, we ran each configuration from Figure 2 against the various model types. The results from the logistic regression, SVM, and neural network are provided within

TABLE 4 | Model performance on different window configurations for the observation period.

Configuration	Logistic regression			SVM			Neural network		
	ACC	AUC	AUC-PR	ACC	AUC	AUC-PR	ACC	AUC	AUC-PR
A	74.9	0.62	0.46	76.8	0.68	0.50	76.7	0.75	0.65
B	73.8	0.60	0.44	75.7	0.66	0.48	75.3	0.73	0.62
C	77.4	0.67	0.52	78.5	0.71	0.54	78.2	0.78	0.70
D	76.8	0.65	0.50	77.4	0.68	0.51	77.4	0.76	0.67
E	77.0	0.66	0.50	77.2	0.68	0.50	77.4	0.76	0.67
F	74.8	0.62	0.46	75.7	0.65	0.48	76.1	0.74	0.64
G	75.0	0.62	0.46	75.6	0.65	0.47	75.9	0.74	0.64

ACC, Accuracy; AUC, Area Under Curve; AUC-PR, Area Under Curve Precision Recall.

Table 4. When comparing configuration A [−365 : 14] and B [−365 : 0], we see the importance of including the acute phase, 14 days post-mTBI, within our models; as noted by the differences in accuracies, AUC, and AUC-PR values. This underscores the impact of the acute phase on the trajectory of a patient as well as on the ability to accurately predict long-term patient outcomes. Evaluating **Table 4**, configuration C [−360 : −30, −30 : 0, 0 : 14], provided the best performance. The accuracy, AUC, and AUC-PR substantially increased compared to configurations A and B. It appears that isolating those 30 days prior to mTBI seemed to provide the models with more significant information.

Dividing the observation period further, configurations D and E, the results of the models began to degrade. We can see that as the number of windows increases, configurations F and G, the performance of our models continues to drop. The sparseness of the data might have had an impact on reductions in model performance. Therefore, configuration C was chosen as the optimal window configuration among our model types. **Figure 4** includes the AUC and AUC-PR curves for this configuration against all the different model types. It becomes apparent that many of these model types performed similarly, with the exception of the neural network which outperformed the other model types substantially.

5.5. Prediction Timeframes

While we were able to develop models that predict the likelihood of mental health symptoms within the year post-mTBI, we recognize the utility of more finite prediction windows; predicting the likelihood of symptoms within smaller time frames. As such, our prediction period was divided into a set of smaller windows, including the first 3 months post-mTBI [14 : 90], the following 6 months [90 : 180], and the last 6 months [180 : 365] (**Figure 5**). Utilizing the various model types, along with the observation window configuration C, we predicted the likelihood of MH conditions within each interval.

Table 5 provides the results from these models. We can see that those intervals closest to the mTBI date, [14 : 90], obtained the best performances. The further from mTBI date, the lower the model accuracies. **Figure 6** details the AUC and AUC-PR curves for predicting the onset of MH conditions within the [14 : 90] prediction window. Again, while many of the model types seemed to perform similarly, the neural network models performance

stood out, with an accuracy of 88.2, an AUC of 0.82, and an AUC-PR of 0.66.

6. DISCUSSION

Through the development of feature vectors and the configuration of observation windows, this study assessed the capability of using EHR data to predict those at risk for negative mTBI outcomes. Due to the nature of the data, considerations were needed when devising our feature vectors. We first evaluated the impact of incrementally adding various variables. The inclusion of patient demographics, risk factors, and pre-morbid conditions increased model performance. In an attempt to provide temporal significance, we examined the benefit of dividing our observation period into distinct windows. After iterating through multiple configurations, dividing the observation period into three distinct windows provided the best performance.

With our best performing model, leveraging a neural network, we further attempted to predict the diagnosis of any mental health condition within distinct outcome windows. Encoding the observation windows into intervals allowed the models to assess and weigh the significance of each vector. Oftentimes those intervals closest to the mTBI date held greatest significance. This was particularly true when assessing those prediction intervals closest to the mTBI date, which exhibited the greatest accuracies.

Through these practices, we can better understand how EHR data can be leveraged to predict clinical outcomes. Though careful consideration is needed, the development of these tools and their deployment within clinical settings can provide great benefit to patients and clinicians alike. By understanding ones' risk of developing certain conditions or adverse outcomes, clinicians can provide prompt and timely interventions. This can aid in the mitigation of potential negative outcomes and the alteration of outcome trajectories. Furthermore, understanding which factors contribute to certain outcomes can help better understand the development and progression of disease.

With the utilization of EHR data, we can look forward to deploying and verifying our models within clinical settings. As this study leveraged a relatively clean dataset, further inclusion of feature vectors should be explored. Additional EHR data,

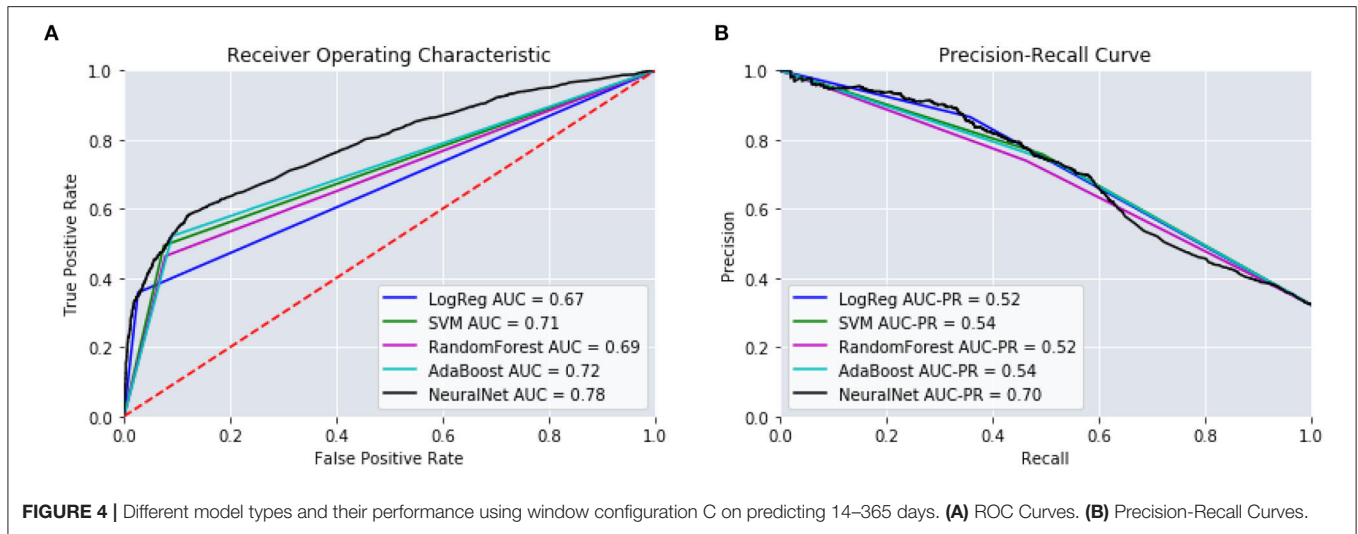


FIGURE 4 | Different model types and their performance using window configuration C on predicting 14–365 days. **(A)** ROC Curves. **(B)** Precision-Recall Curves.

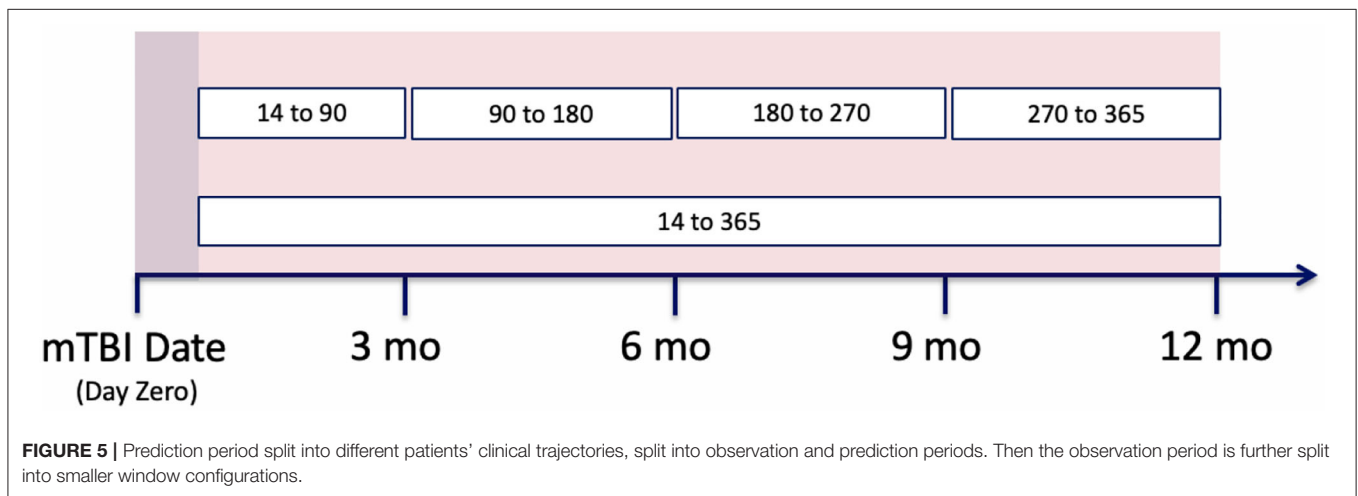


FIGURE 5 | Prediction period split into different patients' clinical trajectories, split into observation and prediction periods. Then the observation period is further split into smaller window configurations.

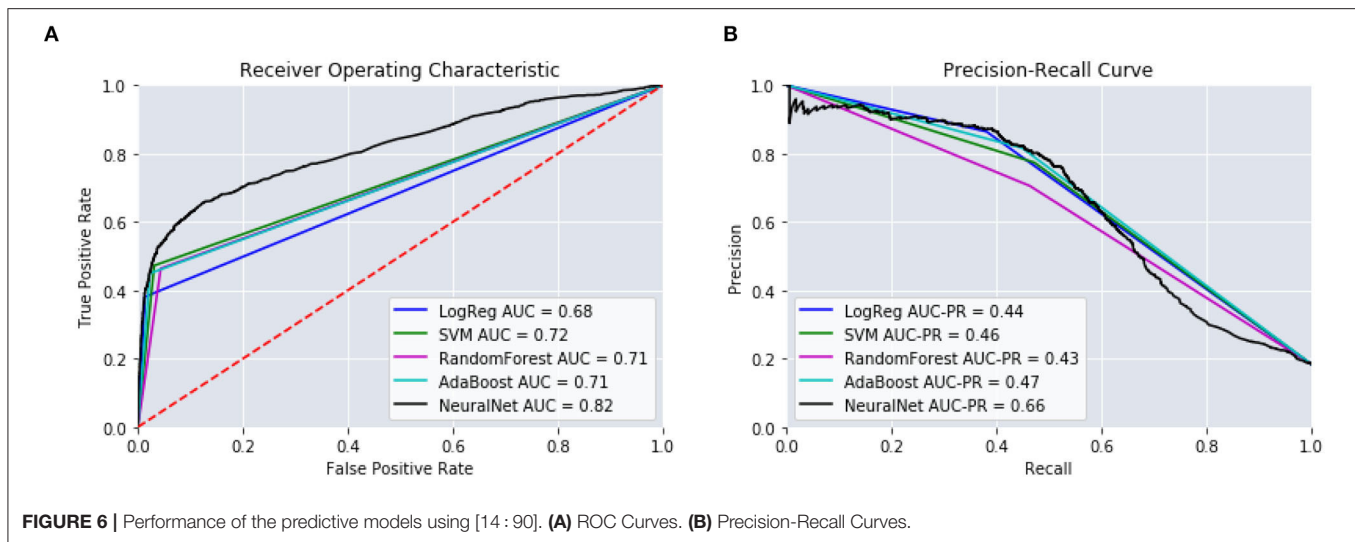
TABLE 5 | Effect of splitting the prediction period into windows based on clinical insight.

Window	Logistic regression			SVM			Neural network		
	ACC	AUC	AUC-PR	ACC	AUC	AUC-PR	ACC	AUC	AUC-PR
14 to 365	85.32	0.59	0.28	85.6	0.62	0.31	86.1	0.78	0.53
14 to 90	87.5	0.68	0.44	87.7	0.72	0.46	88.2	0.82	0.66
90 to 180	89.0	0.56	0.19	89.1	0.58	0.21	89.3	0.78	0.44
180 to 270	83.8	0.56	0.24	84.2	0.60	0.28	84.6	0.75	0.46
270 to 365	80.9	0.54	0.25	81.5	0.58	0.28	80.8	0.72	0.46

ACC, Accuracy; AUC, Area Under Curve; AUC-PR, Area Under Curve Precision Recall.

including laboratory and radiology records could be considered. Furthermore, the inclusion of SMs with complex clinical histories is needed. Within the military population, it is not uncommon for SMs to experience repeated concussions, subsequent injuries, or other comorbidities. In order for these models to have clinical merit, they must be able to cater and adapt to various scopes and populations.

Furthermore, this paper detailed the retrospective application of these models. Future clinical application should be explored through prospective means. With the inclusion of additional features, applying our models prospectively will assist in further improving and validating our techniques. Clinicians could then verify such tools while enhancing clinical decision making.



7. CONCLUSION

Predictive analytics, specifically the use of neural networks, show promise in adopting data science models to identify the likelihood of developing mental health conditions following mTBI. These models can enable clinicians to not only identify at risk individuals, but to better anticipate patient needs and provide interventions to mitigate negative outcomes.

Translation of these models from retrospective constructs into real-world application is imperative. While this paper has demonstrated the technical feasibility of leveraging neural networks for TBI clinical application, on-going multi-site efforts are focusing on (i) optimizing the models to incorporate additional variables, (ii) implement interpretability to be able to incorporate the forecasting models into clinical ancillary applications, and (iii) use these retrospective data models into real-world prospective clinical environments.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the Defense Health Agency (DHA) requires an approved Data Sharing Agreement (DSA) with potential users to confirm that data from the Military Health System (MHS) be used or disclosed in compliance with Department of Defense (DoD) privacy and security regulations. Requests to access the datasets should be directed to Jesus Caban, jesus.j.caban.civ@mail.mil.

REFERENCES

- Brenner LA, Ivins BJ, Schwab K, Warden D, Nelson LA, Jaffee M, et al. Traumatic brain injury, posttraumatic stress disorder, and postconcussive symptom reporting among troops returning from Iraq. *J Head Trauma Rehabil.* (2010) 25:307–12. doi: 10.1097/HTR.0b013e3181cada03
- Farmer CM, Krull H, Concannon TW, Simmons M, Pillemer F, Ruder T, et al. Understanding treatment of mild traumatic brain injury in the military health system. *Rand Health Q.* (2017) 6:11. doi: 10.7249/RR844
- Okie S. Traumatic brain injury in the war zone. *N Engl J Med.* (2005) 352:2043–7. doi: 10.1056/NEJMp058102

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Walter Reed National Military Medical Center. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR'S NOTE

The views expressed in this article are those of the authors and do not necessarily reflect the official policy of the Department of Defense or the U.S. Government.

AUTHOR CONTRIBUTIONS

FD, PH, KJ-W, and JC developed methodology and outlined project. TW queried and extracted electronic healthcare record data. FD built and evaluated predictive models. FD and PH drafted manuscript. KJ-W and JC edited manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2021.769819/full#supplementary-material>

- Jaramillo CA, Eapen BC, McGeary CA, McGeary DD, Robinson J, Amuan M, et al. A cohort study examining headaches among veterans of Iraq and Afghanistan wars: Associations with traumatic brain injury, PTSD, and depression. *Headache.* (2016) 56:528–39. doi: 10.1111/head.12726
- Collen J, Orr N, Lettieri CJ, Carter K, Holley AB. Sleep disturbances among soldiers with combat-related traumatic brain injury. *Chest.* (2012) 142:622–30. doi: 10.1378/chest.11-1603
- Combs HL, Berry DT, Pape T, Babcock-Parziale J, Smith B, Schleenbaker R, et al. The effects of mild traumatic brain injury, post-traumatic stress disorder, and combined mild traumatic brain injury/post-traumatic

- stress disorder on returning veterans. *J Neurotrauma*. (2015) 32:956–66. doi: 10.1089/neu.2014.3585
7. Van der Naalt J, Timmerman M, de Konig M, van der Horn H, Scheenen M, Jacob B, et al. Early predictors of outcome after mild traumatic brain injury an observational cohort. *Lancet Neurol*. (2017) 16:532–40. doi: 10.1016/S1474-4422(17)30117-5
 8. Scofield DE, Proctor SP, Kardouni JR, Hill OT, McKinnon CJ. Risk factors for mild traumatic brain injury and subsequent post-traumatic stress disorder and mental health disorders among United States Army soldiers. *J Neurotrauma*. (2017) 34:3249–55. doi: 10.1089/neu.2017.5101
 9. Department of Veterans Affairs, Department of Defense. *VA/DoD Clinical Practice Guideline for the Management of Concussion-Mild Traumatic Brain Injury*. (2016).
 10. Lingsma H, Yue J, AIR M. Outcome prediction after mild and complicated mild traumatic brain injury: external validation of existing models and identifications of new predictors using the TRACK-TBI pilot study. *J Neurotrauma*. (2015) 32:83–94. doi: 10.1089/neu.2014.3384
 11. Meastas K, Sander A, Clark A. Preinjury coping, emotional functioning and quality of life following uncomplicated and complicated mild TBI. *J Head Trauma Rehabil*. (2014) 29:407–17. doi: 10.1097/HTR.0b013e31828654b4
 12. Mooney G, Speed J, Sheppard S. Factors related to recovery after mild traumatic brain injury. *Brain Injury*. (2005) 19:975–87. doi: 10.1080/02699050500110264
 13. Kontos AP, Kotwal RS, Elbin RJ, Lutz RH, Forsten RD, Benson PJ, et al. Residual Effects of Combat-Related Mild Traumatic Brain Injury. *J Neurotrauma*. (2013) 30:680–6. doi: 10.1089/neu.2012.2506
 14. Lagarde E, Salmi L, Holm H. Association of symptoms following mild traumatic brain injury with post-traumatic stress disorder vs post concussion syndrome. *JAMA Psychiatry*. (2014) 71:1032–40. doi: 10.1001/jamapsychiatry.2014.666
 15. Lange RT, Iverson GL, Rose A. Depression strongly influences postconcussion symptom reporting following mild traumatic brain injury. *J Head Trauma Rehabil*. (2011) 26:127–37. doi: 10.1097/HTR.0b013e3181e4622a
 16. Lange RT, Brickell TA, Kennedy JE, Bailie JM, Sills C, Asmussen S, et al. Factors influencing postconcussion and posttraumatic stress symptom reporting following military-related concurrent polytrauma and traumatic brain injury. *Arch Clin Neuropsychol*. (2014) 29:329–47. doi: 10.1093/arclin/acu013
 17. Kubben P, Dumontier M, Dekker A. Fundamentals of clinical data science. *Springer Nature*. (2019).
 18. Goldstein BA, Navar AM, Pencina MJ, Ioannidis J. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc*. (2017) 24:198–208. doi: 10.1093/jamia/ocw042
 19. Hubbard R. *Department of Biostatistics Epidemiology & Informatics*. Philadelphia, PA: University of Pennsylvania. (2019).
 20. Kamper SJ. Generalizability: linking evidence to practice. *J Orthopaedic Sports Phys Therapy*. (2020) 50:45–6. doi: 10.2519/jospt.2020.0701
 21. Atasoy H, Greenwood BN, McCullough JS. The digitization of patient care: a review of the effects of electronic health records on health care quality and utilization. *Annu Rev Public Health*. (2019) 40:487–500. doi: 10.1146/annurev-publhealth-040218-044206
 22. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Australia. (2015). p. 1721–30.
 23. Bauder R, Khoshgofaar TM, Seliya N. A survey on the state of healthcare upcoding fraud analysis and detection. *Health Serv Outcomes Res Methodol*. (2017) 17:31–55. doi: 10.1007/s10742-016-0154-8
 24. Collins MW, Kontos AP, Reynolds E, Murawski CD, Fu FH. A comprehensive, targeted approach to the clinical care of athletes following sport-related concussion. *Knee Surgery Sports Traumatol Arthrosc*. (2014) 22:235–46. doi: 10.1007/s00167-013-2791-6
 25. Lindberg DS, Prospero M, Bjarnadottir RI, Thomas J, Crane M, Chen Z, et al. Identification of important factors in an inpatient fall risk prediction model to improve the quality of care using EHR and electronic administrative data: a machine-learning approach. *Int J Med Inform*. (2020) 143:104272. doi: 10.1016/j.ijmedinf.2020.104272
 26. Mugisha C, Paik I. Pneumonia outcome prediction using structured and unstructured data from EHR. In: *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE (2020). p. 2640–46.
 27. Morawski K, Dvorkis Y, Monsen CB. Predicting hospitalizations from electronic health record data. *Am J Manag Care*. (2020) 26:e7–e13. doi: 10.37765/ajmc.2020.42147
 28. De Guise E, LeBlanc J, Feyz M, Lamoureux J, Greffou S. Prediction of behavioural and cognitive deficits in patients with traumatic brain injury at an acute rehabilitation setting. *Brain Injury*. (2017) 31:1061–8. doi: 10.1080/02699052.2017.1297485
 29. Rassovsky Y, Levi Y, Agranov E, Sela-Kaufman M, Sverdluk A, Vakil E. Predicting long-term outcome following traumatic brain injury (TBI). *J Clin Exp Neuropsychol*. (2015) 37:354–66. doi: 10.1080/13803395.2015.1015498
 30. Peng X, Long G, Shen T, Wang S, Jiang J, Zhang C. BiteNet: bidirectional temporal encoder network to predict medical outcomes. *arXiv [Preprint]*. arXiv:200913252. (2020) doi: 10.1109/ICDM50108.2020.00050
 31. Darabi S, Kachuee M, Fazeli S, Sarrafzadeh M. Taper: time-aware patient ehr representation. *IEEE J Biomed Health Inform*. (2020) 24:3268–75. doi: 10.1109/JBHI.2020.2984931
 32. Druvenga B. *Traumatic Brain Injuries*. (2017). Available online at: https://bocac.org/newsroom/traumatic-brain-injuries?category_key=at
 33. Prince C, Bruhns ME. Evaluation and treatment of mild traumatic brain injury: the role of neuropsychology. *Brain Sci*. (2017) 7:105. doi: 10.3390/brainsci7080105
 34. Belanger HG, Kretzmer T, Vanderploeg RD, French LM. Symptom complaints following combat-related traumatic brain injury: relationship to traumatic brain injury severity and posttraumatic stress disorder. *J Int Neuropsychol Soc*. (2010) 16:194–99. doi: 10.1017/S135561770990841
 35. Van der Horn H, Spikman J, Jacobs B, Van der Naalt J. Postconcussive complaints, anxiety and depression related to vocational outcome in minor to severe traumatic brain injury. *Arch Phys Med Rehabil*. (2013) 94:867–74. doi: 10.1016/j.apmr.2012.11.039
 36. Center H. *The Johns Hopkins ACG Case-Mix System Reference Manual Version 7.0*. Baltimore, MD: Health Services Research and Development Center, The Johns Hopkins University Bloomberg School of Public Health Baltimore (2005).
 37. Stein MB, Jain S, Giacino JT, Levin H, Dikmen S, Nelson LD, et al. Risk of posttraumatic stress disorder and major depression in civilian patients after mild traumatic brain injury: a TRACK-TBI study. *JAMA Psychiatry*. (2019) 76:249–58. doi: 10.1001/jamapsychiatry.2018.4288
 38. Agency for Healthcare Research Quality, Rockville, MD. *HCUP Clinical Classifications Software (CCS) for ICD-9-CM*. Healthcare Cost and Utilization Project (HCUP). (1999). Available online at: <https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp>
 39. Agency for Healthcare Research Quality, Rockville, MD. *Beta Clinical Classifications Software (CCS) for ICD-10-CM/PCS*. (2018). Available online at: <https://www.hcup-us.ahrq.gov/toolssoftware/ccs10/ccs10.jsp>
 40. Dabek F, Caban JJ. A neural network based model for predicting psychological conditions. In: Guo Y, Friston K, Aldo F, Hill S, Peng H, editors. *Brain Informatics and Health. BIH 2015. Lecture Notes in Computer Science, Vol. 9250*. Cham: Springer.
 41. Gravel J, D'Angelo A, Carrière B, Crevier L, Beauchamp MH, Chauny JM, et al. Interventions provided in the acute phase for mild traumatic brain injury: a systematic review. *Syst Rev*. (2013) 2:63. doi: 10.1186/2046-4053-2-63
 42. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. (2002) 16:321–57. doi: 10.1613/jair.953
 43. Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. In: *Proceedings of the 23rd International Conference on Machine Learning*. Pittsburgh, PA. (2006). p. 233–40.
 44. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*. (2015) 10:e0118432. doi: 10.1371/journal.pone.0118432

45. Choi E, Xiao C, Stewart W, Sun J. Mime: multilevel medical embedding of electronic health records for predictive healthcare. In: *Advances in Neural Information Processing Systems*. Montréal. (2018). p. 4547–57.
46. Kemp J, Rajkomar A, Dai AM. improved patient classification with language model pretraining over clinical notes. *arXiv [Preprint]*. arXiv:190903039. (2019).
47. Van Rossum G, Drake FL Jr. *Python Tutorial*. Amsterdam: Centrum voor Wiskunde en Informatica (1995).
48. McKinney, W., & others. Data structures for statistical computing in python. In: *Proceedings of the 9th Python in Science Conference*. SCIPY (2010). 51–6. Available Online at: <http://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf>
49. Harris CR, Millman KJ, van derWalt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature*. (2020) 585:357–62. doi: 10.1038/s41586-020-2649-2
50. Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *J Mach Learn Res*. (2017) 18:559–63.
51. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. (2011) 12:2825–30.
52. Chollet F, et al. Keras (2015). *GitHub*. Available online at: <https://github.com/fchollet/keras>
53. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SCIPY 1.0: fundamental algorithms for scientific computing in python. *Nat Meth*. (2020) 17:261–72.
54. Hunter JD. Matplotlib: a 2d graphics environment. *Comput Sci Eng*. (2007) 9:90–95. doi: 10.1109/MCSE.2007.55

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Dabek, Hoover, Jorgensen-Wagers, Wu and Caban. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.