2-14-2022

Machine learning to predict sports-related concussion recovery using clinical data

Yan Chu
The University of Texas Health Science Center at Houston

Gregory Knell
The University of Texas Health Science Center at Houston

Riley P. Brayton
The University of Texas Health Science Center at Houston

Scott O. Burkhart
Children's Health Andrews Institute for Orthopaedics and Sports Medicine

Xiaoqian Jiang
The University of Texas Health Science Center at Houston

See next page for additional authors

Authors
Yan Chu, Gregory Knell, Riley P. Brayton, Scott O. Burkhart, Xiaoqian Jiang, and Shayan Shams

Annals of Physical and Rehabilitation Medicine 65 (2022) 101626

Available online at
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com

Original article

Machine learning to predict sports-related concussion recovery using clinical data

Yan Chu[a], Gregory Knell[b,c,d], Riley P. Brayton[b,c,d], Scott O. Burkhart[d], Xiaoqian Jiang[a], Shayan Shams[a,e,*]

[a] School of Biomedical Informatics, The University of Texas Health Science Center at Houston (UTHealth), Houston, TX, USA
[b] Department of Epidemiology, Human Genetics, and Environmental Sciences, The University of Texas Health Science Center at Houston (UTHealth), Dallas, TX, USA
[c] Children's Health and The University of Texas Health Science Center at Houston (UTHealth), Dallas, TX, USA
[d] Children's Health Andrews Institute for Orthopaedics and Sports Medicine, Plano, TX, USA
[e] Department of Applied Data Science, San Jose State University, San Jose, CA, USA

A R T I C L E I N F O

Article History:
Received 13 April 2021
Accepted 25 October 2021
Available online xxx

Keywords:
Machine learning
Brain concussion
Adolescent
Athletic injuries/rehabilitation
Vestibular function tests
Sport injuries

A B S T R A C T

Objectives: Sport-related concussions (SRCs) are a concern for high school athletes. Understanding factors contributing to SRC recovery time may improve clinical management. However, the complexity of the many clinical measures of concussion data precludes many traditional methods. This study aimed to answer the question, what is the utility of modeling clinical concussion data using machine-learning algorithms for predicting SRC recovery time and protracted recovery?
Methods: This was a retrospective case series of participants aged 8 to 18 years with a diagnosis of SRC. A 6-part measure was administered to assess pre-injury risk factors, initial injury severity, and post-concussion symptoms, including the Vestibular Ocular Motor Screening (VOMS) measure, King-Devick Test and C3 Logix Trails Test data. These measures were used to predict recovery time (days from injury to full medical clearance) and binary protracted recovery (recovery time > 21 days) according to several sex-stratified machine-learning models. The ability of the models to discriminate protracted recovery was compared to a human-driven model according to the area under the receiver operating characteristic curve (AUC).
Results: For 293 males (mean age 14.0 years) and 362 females (mean age 13.7 years), the median (interquartile range) time to recover from an SRC was 26 (18–39) and 21 (14–31) days, respectively. Among 9 machine-learning models trained, the gradient boosting on decision-tree algorithms achieved the best performance to predict recovery time and protracted recovery in males and females. The models' performance improved when VOMS data were used in conjunction with the King-Devick Test and C3 Logix Trails Test data. For males and females, the AUC was 0.84 and 0.78 versus 0.74 and 0.73, respectively, for statistical models for predicting protracted recovery.
Conclusions: Machine-learning models were able to manage the complexity of the vestibular-ocular motor system data. These results demonstrate the clinical utility of machine-learning models to inform prognostic evaluation for SRC recovery time and protracted recovery.

Introduction

Recently, increased attention has been given to the prevention, prognosis, treatment, and management of sport-related concussions (SRCs) occurring during childhood and adolescence [1–4]. This interest is in response to the growing body of evidence linking SRCs to short- and long-term health outcomes, including behavioural, academic, neurocognitive, and social deficits [5–9] stemming from persistent symptoms. Children and adolescents at high risk for persistent symptoms after an SRC are currently not able to be identified at the outset of the injury. This inability prevents the clinician from distinguishing between clinical and physiological recovery as well as prescribing a more personalized rehabilitation protocol for a safe and timely recovery. Despite growing interest in these areas, there are no tools or instruments available to estimate SRC recovery time.

Estimating SRC recovery time would be helpful to the clinician in several ways. First, the expected recovery time can be used to indicate any outlying clinical recoveries (i.e., those that recover before/after the estimated recovery time). It is important to identify individuals who have clinically recovered before the estimated recovery

* Corresponding author at: Department of Applied Data Science, San Jose State University, One Washington Sq, San Jose, CA 95192.
E-mail address: Shayan.Shams@sjsu.edu (S. Shams).

https://doi.org/10.1016/j.rehab.2021.101626

time, which may be an indication of clinical recovery rather than physiological recovery. There is evidence of athletes masking symptoms after an SRC to return to play because of the pressure they place on themselves or pressure from friends, teammates, or coaches [10]. Releasing an athlete to return to school or sport before physiological recovery may put the athlete at risk of subsequent injury, further delaying recovery, and possibly leading to long-term health problems. Second, the clinician can use the expected recovery time to develop personalized recovery protocols based on the expected recovery time. As has been seen in other areas of medicine [11–13], personalized treatment plans may be effective in improving patient outcomes. Finally, an accurate estimation of recovery time benefits the clinic and patient by eliminating unnecessary visits. Currently, people with concussion are scheduled for follow-up visits at standardized intervals until clinical recovery. Because of the high variability in SRC recovery time, this standardized approach may result in a delayed return to play for individuals with expedited recovery and unnecessary appointments for those with protracted or longer than usual recovery. This protracted recovery is clinically defined as > 21 elapsed days from date of injury to date of medical clearance to return to all activities.

However, estimating SRC recovery time remains a challenge. Several factors contribute to recovery time, including injury-specific factors (injury severity, post-acute injury symptoms [e.g., loss of consciousness]), pre-injury factors (concussion history, neurological condition[s] history, etc.), and individual/patient-specific factors (age, sex, neurodevelopmental and behavioral disorders, etc.). Female athletes in particular report more total symptoms, demonstrate more significant neurocognitive impairment, and report more symptoms during clinical evaluations after an SRC than do males [14].

Disruptions to the vestibular and/or ocular-motor systems may contribute to delayed SRC recovery [15–20]. The vestibular system is a network of organs in the inner ear that provides information on motion, position, balance, and spatial orientation [21]. This complex and delicate system provides direction for compensatory movements to maintain visual (vestibulo-ocular) and postural (vestibulospinal) balance control. Therefore, disruptions to the vestibular system often manifest as symptoms of dizziness, visual instability, and/or loss of balance. In addition to vestibular system impairment, an SRC often results in disruptions to the ocular-motor system. In fact, ocular-motor dysfunction is present in many neurological disorders and has been described as representing higher (dys)functions of the brain [22,23]. These dysfunctions may manifest as blurred vision, diplopia, impaired eye movement, difficulty reading, dizziness, headaches, ocular pain, and disrupted concentration. The complex multi-dimensional structure of the vestibular-ocular motor system may provide the level of sensitivity needed to identify concussions that will require longer recovery. However, from a clinical perspective, the complexity of the vestibular-ocular motor system limits the clinician's ability to deduce an estimated recovery time. Therefore, there is a critical need to identify tools that can overcome these complexities to produce meaningful results.

Machine learning (ML) has been proposed as an analytical tool to overcome these complexities and has been used in other areas of medicine [11,23], public health [24,25], and gene representation [26]. Walker and colleagues provided a proof-of-concept for ML approaches to predict concussion recovery time in a pediatric sample [27] and in other areas of sports injury [28]. The Walker et al. study did not use vestibular-ocular motor system testing data. To our knowledge, no study has used ML approaches to deduce data from vestibular-ocular motor system testing, in conjunction with other concussion evaluations, to predict SRC recovery time. Therefore, we do not know whether there is any clinical utility in ML algorithms of clinical concussion data for predicting SRC recovery time and protracted recovery. This study aimed to address this question.

We developed an ML framework to predict 1) SRC recovery time (days) and 2) protracted recovery (yes/no). To contextualize the ML model, we compared the results to traditional statistical models. Although ML approaches are important and provide an understanding of complex data not allowed by traditional, human-driven, statistical modeling approaches, one must contextualize and compare results from ML approaches with results from traditional approaches. The organization of paper follows STROBE guidelines.

Methods

Setting and participants

This study used a retrospective case series design. Cases were sampled from patients presenting to a pediatric specialty clinic in Plano, TX, USA from October 2017 to March 2020. Data were collected during clinical examinations occurring as part of the standard of care for treating pediatric concussion. Study inclusion criteria were age 8 to 18 years, participating in a sport at the time of injury, diagnosed with an SRC, and evaluated within 7 days after the initial date of injury. Exclusion criteria, documented by medical history, included any of the following: previous diagnosis of developmental delay, diagnosis of a comorbid neck or spine injury, previous diagnosis of congenital or acquired neurological defect not related to the concussion injury, and inability to understand the premise of the study due to language barriers. The research protocol was approved by the Institutional Review Boards at the University of Texas Southwestern Medical Center (UTSW) and the Committee for the Protection of Human Subjects at the University of Texas Health Science Center at Houston (UTHealth).

Measures

The vestibular-ocular motor system was assessed by using the Vestibular Ocular Motor Screening (VOMS) measure [21], King-Devick Test [29], and C3 Logix Trails Tests [30]. The tests were administered in the following order: 1) VOMS, 2) King-Devick Test, 3) C3 Logix Trails Tests, and are clinically independent (there is no sequential test administration based on prior test performance). The VOMS was specifically designed to detect concussion injuries by identifying the provocation of symptoms using a series of 5 tests (smooth pursuits, saccadic or rapid eye movements, near point convergence, vestibular ocular reflex, visual motion sensitivity) that promote the interaction of the vestibular and ocular motor systems. Before the administration of the VOMS measure, the clinician asked the patient to report his/her current severity (scale of 0 to 10) of the following concussion-related symptoms: headache, dizziness, nausea, fogginess. Then, after the clinician demonstrated the specific tasks for each test, the patient performed the test and was asked to report any provocation of symptoms occurring during or at the completion of the task. Symptom severity during or at the completion of each test was recorded by the clinician. Several summary estimates were generated from the VOMS measure. First, we computed the sum of the symptom scores across all tests and the sum of the test scores across all symptoms. Then, we computed the difference in test provoked symptom scores from baseline. We also computed the sums of the difference scores from baseline.. Each of the symptom scores and summary scores were reported and analyzed as discrete variables. Additionally, 3 binary (yes, no) variables were created to indicate a positive screen for delayed recovery by using the threshold of an increase in 2 points in symptom severity score from baseline for 1) ≥ 1 test, 2) ≥ 2 tests, and 3) ≥ 3 tests as informed by the VOMS scoring criteria proposed by Mucha et al. [21].

The King-Devick Test is a quick number-naming measure historically used as a sideline SRC assessment tool. For the current study, the total time and errors were recorded by using the same number-naming cards from the standard sideline administration [29].

The Trails A and B subtests from the C3 Logix system are a tablet-administered assessment of visual processing speed. Trails A and B tests require the test taker to draw a line between points with their fingers that are either a sequence of numbers or numbers and letters in an ascending progression. Both the sum and means of Trails A and B tests as well as the difference between scores for Trails A and B tests were calculated and used as performance measures [30].

The primary outcomes of interest were recovery time (days) and protracted recovery (yes/no). Recovery time was defined as the number of days from the patient-reported date of injury to the date of recovery, as determined by medical clearance to resume normal academic and athletic activities. Medical clearance was provided at the final clinic visit, which was scheduled when the patient experienced an asymptomatic response to unrestricted physical and cognitive activities for at least 3 consecutive days. Time in days was analyzed as a discrete variable. Additionally, a binary variable (yes/no) representing the presence of protracted recovery was included as an outcome and defined as recovery time > 21 days. A multi-class variable relating to recovery time in weeks was also included to assist the clinician with scheduling follow-up appointments. Five classification categories (i.e., 0−2 weeks, 2−3 weeks, 3−4 weeks, 4−5 weeks, and > 5 weeks) were defined for the multi-class variable.

Other participant characteristics (age [discrete, years] and sex) and risk factors for delayed recovery were collected as part of the patient's initial clinic visit. Participants were asked to report the sport they were participating in when the concussion occurred, loss of consciousness at the time of the injury, concussion history, migraine history, and acute presence of vision problems and amnesia. The sport the participant was participating in at the time of injury was classified as non-contact, contact, or collision, based on other previously defined criteria [31]. All other potential risk factors for protracted recovery (medical history of psychological problems, concussion, or migraine, acute symptoms after concussion including dizziness, headache, vision problems, amnesia, and loss of consciousness) were recorded and analyzed as binary (yes/no) variables.

All variables included in the modeling are defined and detailed in Appendix Table A.

Data analysis

All variables were assessed for missing data and normality as appropriate. The analytic sample did not contain missing data for any numeric variables. Categorical data with missing data were analyzed as an additional "missing" category. Data were evaluated by Student *t*-test, Kruskal-Wallis test, and chi-square test for heterogeneity to determine any statistical differences in mean, median, and proportion estimates, respectively.

Calculation

ML modeling

Because of the nature of the data and impact of categorical variables on ML models, we used CatBoost-based [32] modeling to predict 1) SRC recovery time (days) and 2) protracted recovery by using premorbid and post-acute injury risk factors and VOMS, King-Devick Test and C3 Logix Trails Test results as predictive variables. All variables were included as potential predictive variables in model training because the CatBoost algorithm automatically selects significant variables whose contribution is evaluated by the SHapley Additive exPlanations (SHAP) [33] value. The dataset was split into proportions of 65%, 15%, and 20% for training, validation, and testing, respectively. The maximum number of trees selected was 4000 to avoid overfitting based on the size of the dataset and number of predictive variables. For females, the depth of model selected was 8 because a relatively deeper structure was required to learn complicated variable information as compared with a depth of 6 for males. Because of the non-linearity and complexity of our models and to provide insight into the models' predictions, SHAP values were used to estimate the average marginal contributions of each variable across all permutations and to quantify variable importance.

We compared the performance of the CatBoost model with other commonly used models, including decision tree [34], elastic net [35], random forest [36], XGBoost [37], and TabNet [38]. The performance of the models was compared by prediction error using the root mean square error (RMSE). All results were generated by 5-folder cross-validation. The ability of the models to discriminate individuals who would have protracted recovery from normal recovery was further evaluated by using the area under receiver operating characteristic (ROC) curve (AUC).

Statistical modeling

The utility of the ML algorithms was evaluated against a human-driven model building approach (zero-truncated negative binomial regression models). Zero-truncated negative binomial regression models were used to account for 1) the data structure of the outcome (number of days to recover) inherently lacking zeros (i.e., a patient will take at least 1 day to recover from the concussion, therefore requiring a zero-truncated model) and 2) dispersion of the count data (over-dispersed) requiring a negative binomial. First, the datasets for each sex were randomly split into training (75%) and validation (25%) datasets. With the training dataset, a forward and backward stepwise approach was used with variable p-value inclusion threshold set at 0.25 at step one (bivariable models) and 0.05 at step two to establish the main-effects multivariable model. Each main-effect model variable was tested for interaction with age, given the possibility of a certain level of cognitive development required to understand or answer the questions accurately. Tests for collinearity between variables were performed along with post-hoc analyses of model fit and tests for the appropriateness of the zero-truncated negative binomial model selection. The prediction model was then applied to the validation dataset, and specificity, accuracy and ROC curves were used to determine the model's ability to discriminate individuals who would have a protracted recovery from those who would have a normal recovery. ROC curves were also generated to compare the discriminatory abilities of the ML and human-driven models.

Results

Sample characteristics

A total of 1109 patients presented to the study site between October 2017 and March 2020; 930 met the inclusion criteria (age 8−18 years, participating in a sport at the time of injury, diagnosed with an SRC and evaluated within 7 days after the initial date of injury). A further 275 were excluded based on the exclusion criteria (diagnosed developmental delay, comorbid neck or spine injuries, or congenital or acquired neurological defect or injury). The analytic sample was 655 (362 males and 293 females). Table 1 details participant characteristics. The mean (SD) age was 13.7 (2.4) and 14.0 (2.2) years, respectively. Males most frequently participated in a collision sport at the time of injury (n = 236, 65%), and females most frequently participated in a contact sport (n = 208, 71%). Acute post-injury risk factors for protracted recovery with the highest prevalence were headache, dizziness, and vision problems for both males and females. Appendix Fig. A depicts the distribution and descriptive statistics of recovery time. Males and females significantly differed in recovery time (median 21 days [14−31] and 26 days [18−39], p<0.001).

Table 1
Descriptive statistics of patients undergoing sport-related concussion recovery treatment in a pediatric clinic setting, 2017–2020.

Characteristics	Male, 362 (55)	Female, 293 (45)	p-value
Recovery time (days), median (IQR)	21 (14–31)	26 (18–39)	<0.001
Age (years), Mean (STD)	13.7 (2.4)	14.0 (2.2)	0.089
8–13	163 (45)	104 (35)	
14–18	199 (55)	189 (64)	
Sport			<0.001
Non-contact	5 (1)	27 (9)	
Contact	88 (24)	208 (71)	
Collision	236 (66)	11 (4)	
Missing	33 (9)	47 (16)	
Presence of risk factors			
History of psychological problem	2 (1)	8 (3)	0.034
History of concussions	34 (9)	38 (13)	0.099
History of migraines	17 (5)	17 (6)	0.727
Acute dizziness	104 (29)	88 (30)	0.425
Acute headache	118 (33)	120 (41)	0.025
Acute vision problems	76 (21)	55 (19)	0.061
Acute amnesia	41 (11)	26 (9)	0.140
Loss of consciousness	15 (4)	9 (3)	0.421

Data are n (%) unless indicated.
IQR, interquartile range; STD, Standard Deviation.

Recovery time prediction

Table 2A compares CatBoost, XGBoost, Elastic Net, TabNet, and Extra Trees Regressor in predicting SRC recovery time for males and females [39]. Fine-tuned CatBoost prediction outperformed other ML models with regard to the RMSE. This finding was consistent with and without the King-Devick Test and C3 Logix Trails Tests.

Protracted recovery classification

Appendix Fig. B presents the ROC curves for discriminating protracted recovery (recovery time > 21 days) and normal recovery. For males and females, the AUC was 0.84 and 0.78, respectively. To quantify the contribution of the King-Devick Test and C3 Logix Trails Test variables to predicting protracted recovery and to illustrate the importance of comprehensive tests for patients, we ran post-hoc tests of the CatBoost model on the same dataset without King-Devick Test and C3 Logix Trails Test variables. The AUCs were 0.78 and 0.76, respectively.

SHAP summary plots of the prediction models used for classification of protracted recovery are shown in Fig. 1. The most significant variable for males was the VOMS near point convergence test difference score–dizziness (con_ddiff), which is the difference between the dizziness symptom score at baseline and after completing the near point convergence test. Additionally, the change in VOMS saccades-horizontal score (sach1), visual motion sensitivity test difference score–headache (vms_hadiff), and vestibular ocular reflex-vertical test difference score–headache (vorv_hadiff) also played crucial roles for males. The most significant variable for females was the near point convergence test difference score–dizziness (con_hadiff). Smooth pursuits test difference score–headache (pur_hadiff), King-Devick Test cards 2 error (kdc2_e), and saccades-horizontal score (sach1) also highly contributed to the classification accuracy for females. Although these variables most significantly contributed to the models, 11 total variables each for males and females showed relevance to the model output (Fig. 1). For variable details, please refer to the variable codebook (Appendix Table A).

We also compared our proposed ML model for predicting protracted recovery with several other commonly used methods, including boosting methods, deep learning methods, and linear models. Table 2B compares the models, illustrating that fine-tuned CatBoost outperformed other ML models for protracted recovery prediction.

Table 2
(**A**) Comparison with other machine-learning models for predicting sport-related concussion recovery time stratified by sex based on the root mean square error (RMSE).

Method	RMSE	
	Males	Females
Mean Baseline Regression Pipeline	14.43	13.09
Extra Trees Regressor w/ Imputer	14.73	12.74
Elastic Net Regressor w/ Imputer	16.29	12.44
XGBoost Regressor	15.77	13.43
Random Forest	14.18	12.51
Linear Regression w/ Imputer	15.97	14.47
TabNet Regressor	14.62	13.93
CatBoost without KD and C3 variables	**13.80**	**12.08**
CatBoost with KD and C3 variables	**12.84**	**10.83**

(**B**). Comparison with other machine-learning models for predicting protracted sport-related concussion recovery stratified by sex based on AUC .

Method	AUC	
	Males	Females
Mode Baseline Binary Classification	0.50	0.50
Extra Trees Classifier w/ Imputer	0.70	0.59
Elastic Net Classifier w/ Imputer	0.50	0.50
XGBoost	0.64	0.56
Random Forest	0.70	0.62
Logistic Regression Classifier	0.56	0.55
TabNet	0.74	0.70
Zero-truncated models	0.70	0.67
CatBoost without KD and C3 variables	**0.78**	**0.72**
CatBoost	**0.84**	**0.78**

KD, King-Devick Test; C3, C3 Logix Trails Tests.
AUC, area under the receiver operating characteristic curve; KD, King-Devick Test; C3, C3 Logix Trails Tests.

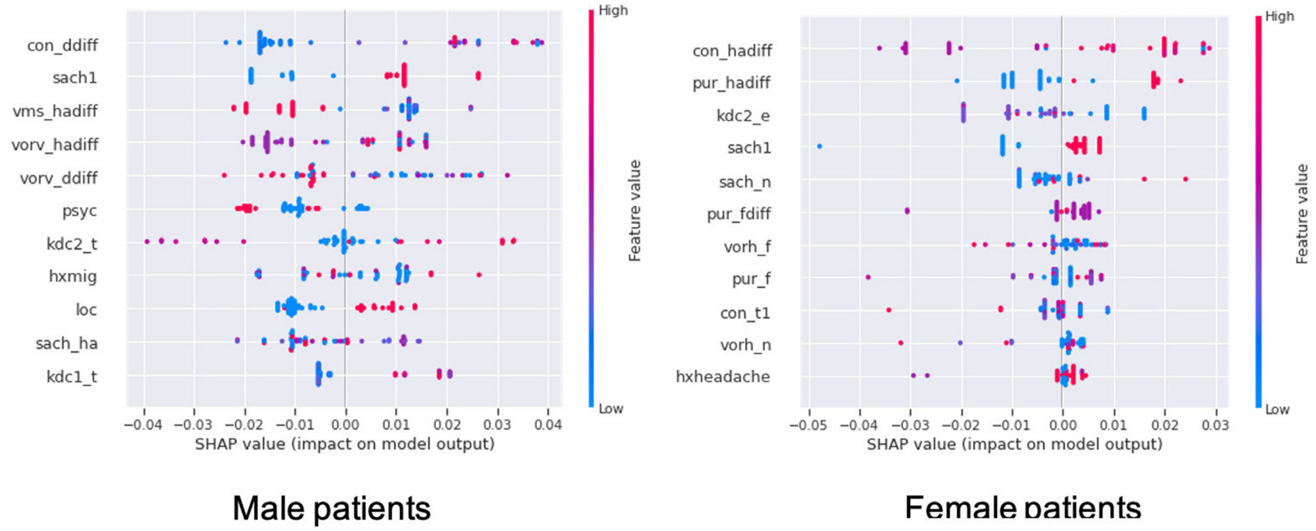

Fig. 1. SHapley Additive exPlanations (SHAP) summary plots of the prediction models used for classifying protracted recovery from a sports-related concussion by sex.

con_ddiff, difference from baseline on dizziness symptom scores during Vestibular Ocular Motor Screening (VOMS) convergence test; con_hadiff, difference from baseline on headache symptom scores during VOMS convergence test; con_t1, VOMS convergence test time 1; hxheadache, history of headaches; hxmig, history of migraines; kdc2_e, King-Devick card 2 errors; kdc1_t, King-Devick card 1 time; kdc2_t, King-Devick card 2 time; loc, loss of consciousness; psyc, history of psychological problems; pur_fdiff, difference from baseline on fogginess symptom scores during VOMS smooth pursuits test; pur_hadiff, difference in headache symptom scores on VOMS smooth pursuits test; sach1, sum of all baseline symptom scores on VOMS horizontal saccades test scores; sach_ha, headache symptom score during VOMS horizontal saccades test; sach_n, nausea symptom score on VOMS horizontal saccades test; vms_hadiff, difference from baseline on headache symptom score during VOMS visual motion sensitivity test; vorh_n, nausea symptom score during VOMS horizontal vestibular ocular reflex test; vorv_ddiff, difference from baseline on dizziness symptom score during VOMS vertical vestibular ocular reflex test; vorv_hadiff, difference from baseline on headache symptom score during VOMS vertical vestibular ocular reflex test.

Notes: The summary plot combines feature importance with feature effects. The variables are ordered according to the corresponding variable importance, and each point on the plot is the Shapley value for a variable and an instance. The color represents the value of the variable from low to high. *Left*: SHAP summary plot for males, which illustrates the significance of con-ddiff, sach1, and vms_hadiff. *Right:* SHAP summary plot for females, which illustrates the significance of con-hadiff, pur_hadiff, and kdc2_e. For variable details, please refer to the variable codebook (Appendix Table A).

For males, the prediction accuracy and precision of CatBoost models were 75% and 83% and 71% and 76% for females.

Multi-classification

To assist the clinician with scheduling follow-up appointments, we conducted multi-classification of recovery time in weeks. We manually split recovery time into 5 classes, 0–2 weeks, 2–3 weeks, 3–4 weeks, 4–5 weeks, > 5 weeks. A summary of the distribution of patient recovery time (weeks) classification is in Appendix Table B. Appendix Fig. C shows the ROC curve and corresponding AUC values for each recovery time class. For 0–2 weeks, 3–4 weeks, 4–5 weeks, and > 5 weeks, the model performed adequately, whereas for 2–3 weeks, the model performed inadequately because it was undermined by noisy input features.

Statistical analysis for predicting recovery time and protracted recovery

The human driven zero-truncated models predicting recovery time are in Appendix Table C. The effect of each mutually adjusted predictor on recovery time can be estimated by taking the exponent of the beta coefficient. For example, among males, the incidence rate ratio for the number of concussions previously experienced (hxconcquat) can be interpreted as follows: for each additional concussion experienced previously, the recovery time (days) rate would be expected to increase by a factor of 1.19 days (p<0.001), holding all other factors constant. Convergence +1 score change on any symptoms (con1), vestibular ocular reflex horizontal test fogginess symptom score (vorh_f) and convergence test dizziness symptom score (con_d) most significantly aided in prediction in the traditional statistical model for males, whereas King-Devick Test cards 1 error (kdc1_e), visual motion sensitivity test fogginess symptom score (vms_f) and visual motion sensitivity+1 score change on any symptoms (vms1) significantly aided in prediction for females (p<0.001).

Furthermore, the ability of the traditional models to discriminate protracted recovery from normal recovery was evaluated by the ROC curve in the validation dataset. The traditional model demonstrated an AUC of 0.74 for males and 0.73 for females (Appendix Fig. B), whereas the CatBoost models produced an AUC of 0.84 for males and 0.78 for females (Table 2B).

Discussion

As compared with the human-driven statistical model, the CatBoost-based ML model showed higher predictive and discriminative ability, which indicates that the CatBoost method was more accurate in identifying both males and females that would experience a normal or protracted recovery after an SRC. Additionally, the CatBoost method produced a more parsimonious model than the human-driven model for both females and males, with 11 features versus 25 for females, and 11 versus 27 for males [40]. Furthermore, we evaluated the ability of our ML algorithm to predict week-level recovery times to demonstrate its potential utility in optimizing patient follow-up appointment scheduling. Among the ML algorithms, the CatBoost-based model outperformed other related models. The CatBoost algorithm was able to overcome limitations present in other ML methods that use boosting tree and numeric inputs, which are limited because of their assumption on continuous variable space and the number of patients involved in the dataset, thereby undermining the robustness of the algorithms. Other methods, based on an attention mechanism and neural networks, are limited by high correlation among predictive variables that conflict with the orthogonality assumption. Comparatively, the CatBoost method overcomes both these limitations by using ordered Boosting and random permutation. The CatBoost method does not feature issues related to overfitting and can handle categorical data input while remaining comparatively efficient with several variables in the optimal model.

Few other studies have evaluated the utility of ML to predict SRC recovery, which was our primary aim; however, these findings generally support other analyses of the association between vestibular-ocular motor evaluation and concussion recovery. Previously, we [18] found that VOMS test domains and VOMS test thresholds were associated ($p<0.05$) with SRC recovery time in days, but the VOMS test performed poorly (AUC=0.66 for males; 0.56 for females) as a tool to discriminate participants who would recover normally and those who would have a delayed recovery. We concluded that despite the VOMS association with recovery time, as has also been found by others [15,17,19,20], the VOMS does not appear to have sufficient ability to identify delayed SRC recovery in a pediatric sample as a stand-alone prognostic tool. Hence, this was the catalyst for the current study, which complemented the VOMS testing of the vestibular-ocular motor system with the King Devick test of ocular motor speed and C3 Trails Tests of cognitive functioning.

Study limitation

The first limitation is that ML methods are inherently large-sample procedures. Although the analytic sample (n = 655) in the current study was sufficiently powered to produce reliable estimates using both the ML and human-driven statistical methods, a larger analytic sample would better account for individual differences and random error, thereby further refining the reliability and predictive ability of the models. However, ML methods hold assumptions, including orthogonality, normality, or independence among predictive variables, which could be a challenge when collecting more data. Furthermore, unmeasured factors may be contributing to recovery time, such as behaviours occurring during the recovery phase (levels of physical activity, sleep, cognitive strain, and vestibular-ocular motor system rehabilitation, etc.). Of note, this study was designed to predict SRC recovery time and protracted recovery based on pre-morbid and post-acute injury factors alone. As such, given the strong influence of these other factors on recovery time, the predictive ability of this model may be limited. Further research is necessary to validate the predictive model provided in settings outside of the original data collection site. The predictive model sample was collected at a specialized concussion center and results may differ when compared to those from emergency medicine and primary care settings. Finally, multi-classification of the number of weeks for SRC recovery is limited by the number of patients involved in the dataset. In the future, better prediction accuracy in multi-classification is expected to be achieved with the availability of more data.

Conclusions

Clarification around clinical recovery is important for medical providers managing SRC in pediatric patients. Prediction models may benefit providers by informing SRC recovery time thereby allowing for specialized education and specific therapeutic interventions, and identifying those who may benefit from referrals to additional specialty medical providers. Accurate prognosis upon the initial visit is also important in setting realistic expectations for patients and may contribute to patient compliance with prescribed treatment plans. This is a promising tool to support clinicians and patients in predicting SRC recovery by eliminating unnecessary follow-up appointments in patients with protracted recovery.

Author contributions

YC, GK and SS were responsible for the investigation, conceptualization and methodology. SS and XJ provided supervision and resources for analysis. YC, GK, SS contributed to formal analysis and methodology. SB is responsible for data curation and clinical validation. YC, GK, SB, RB, XJ and SS drafted the paper. All authors reviewed and conducted the final approval of the version to be published. All authors agreed to submit the report for publication.

Conflict of interest

None declared.

Funding

This work was supported in part by the Cancer Prevention and Research Institute of Texas (CPRIT) [RR180012 and RP200526]. The content is solely the responsibility of the authors, and the study sponsor was not involved in the collection, analysis or interpretation of the data, writing of the report, or decision to submit the article for publication. XJ is CPRIT Scholar in Cancer Research (RR180012) and was supported in part by the Christopher Sarofim Family Professorship, UT Stars award, UTHealth startup, the National Institutes of Health (award nos. R01AG066749, R01GM114612 and U01TR002062), and the National Science Foundation (RAPID no. 2027790).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.rehab.2021.101626.

Appendix

Fig. A, Fig. B and Fig. C.

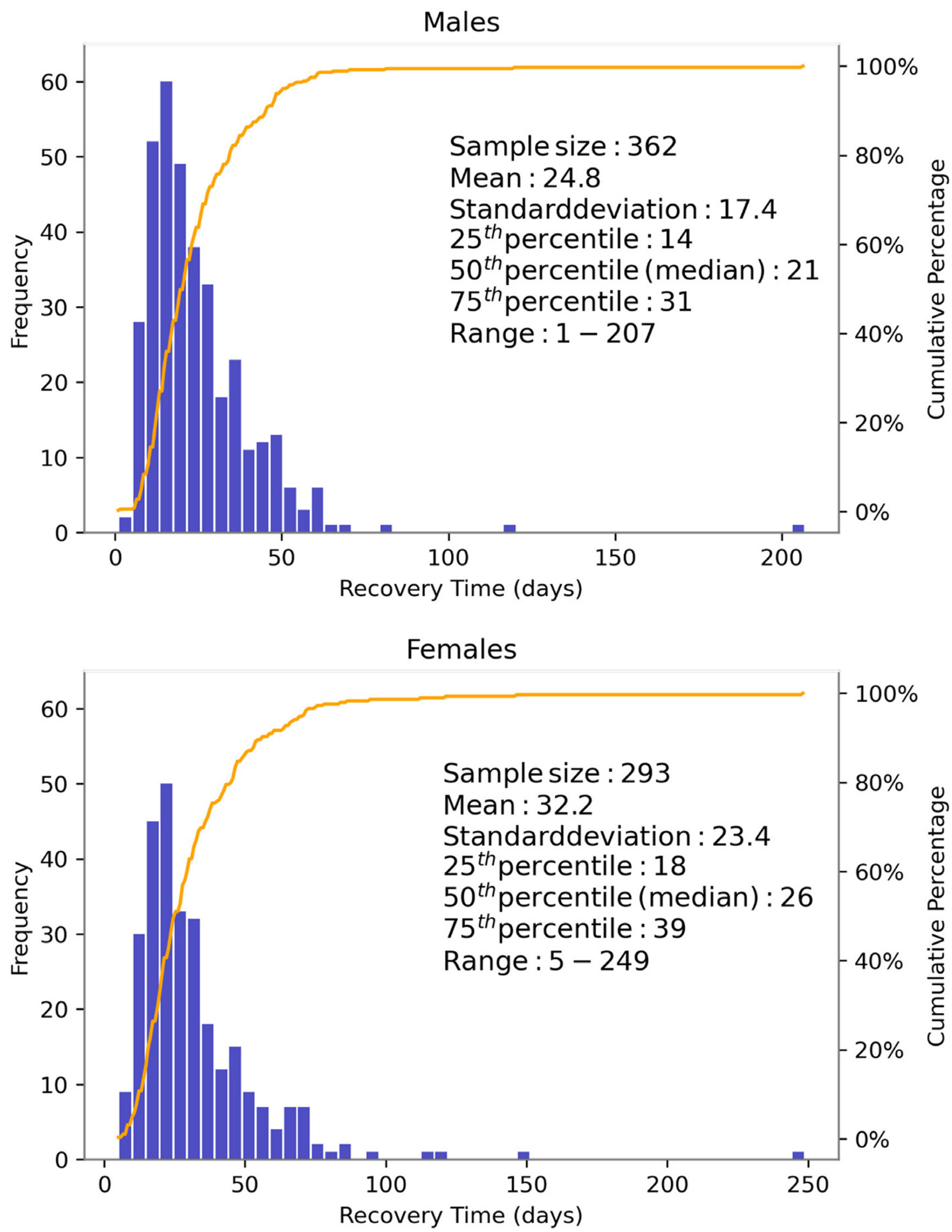

Fig. A. Distribution and descriptive statistics of days to recover from a sports-related concussion by sex among pediatric patients in our dataset.

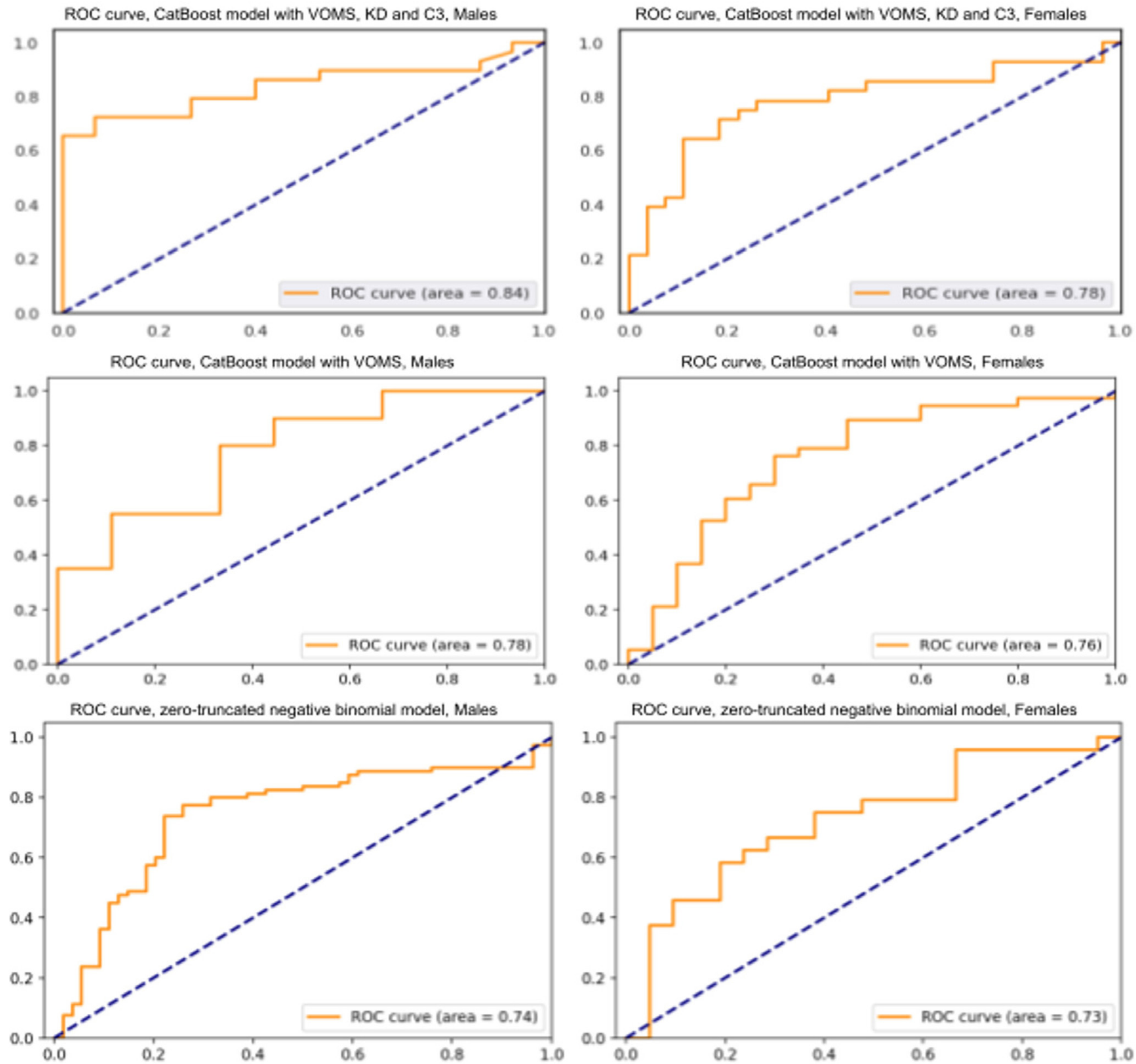

Fig. B. Area under the receiver operating characteristic curve for males and females with or without King-Devick Test and C3 Logix Trails Test variables using CatBoost and a human-driven statistical model. The area under the ROC curve increases in males and females from use of human-driven statistical model to use of machine-learning model CatBoost informed by VOMS variables and improves again when King-Devick Test and C3 Logix Trails Test variables are used in conjunction with VOMS variables. ROC, receiver operating characteristic; VOMS, Vestibular Ocular Motor Screening.

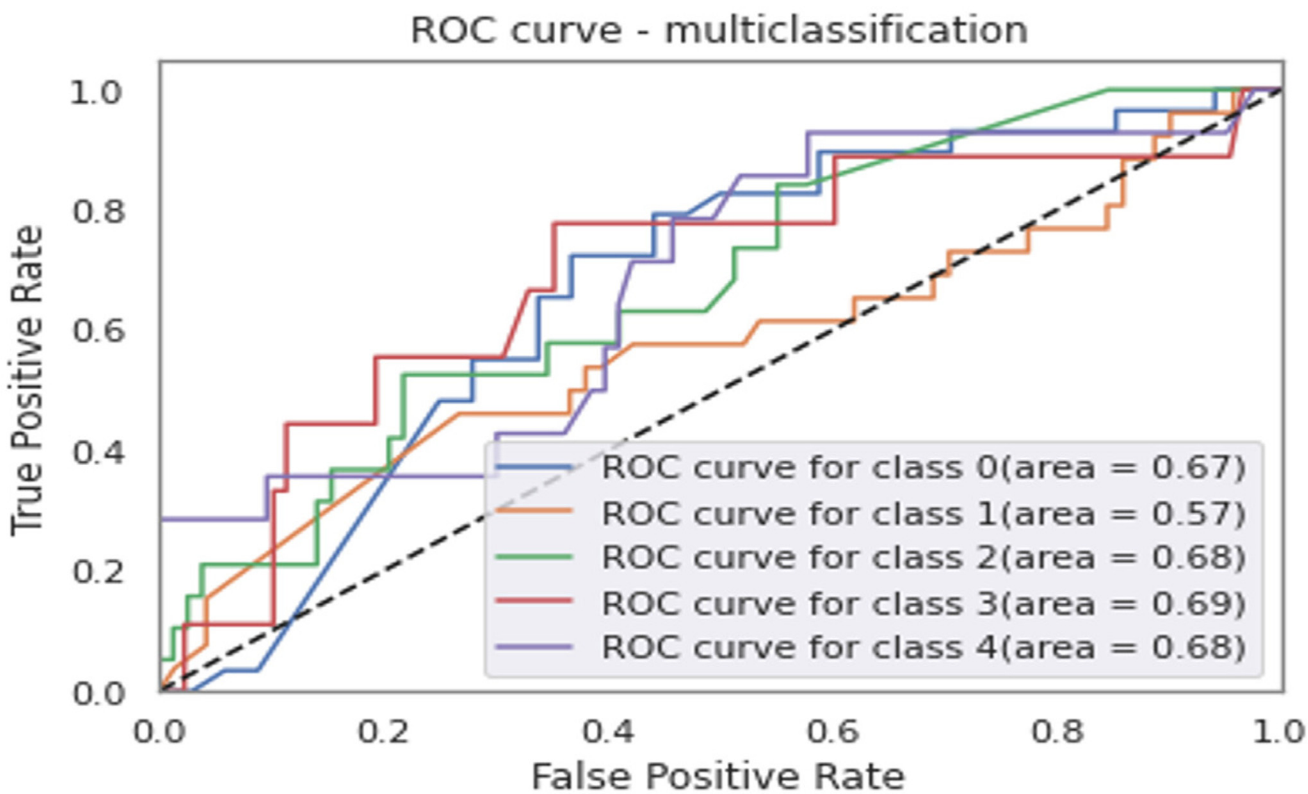

Fig. C. Area under the ROC curve for multi-class classification for all patients with King-Devick Test and C3 Logix Trails Test variables using CatBoost. Class 0, 0–2 weeks; 1, 2–3 weeks; 2, 3–4 weeks; 3, 4–5 weeks; and 4, > 5 weeks.

References

[1] Feddermann-Demont N, Echemendia RJ, Schneider KJ, Solomon GS, Hayden KA, Turner M, et al. What domains of clinical function should be assessed after sport-related concussion? A systematic review. Br J Sports Med 2017;51:903–18. doi: 10.1136/bjsports-2016-097403.

[2] Haider MN, Leddy JJ, Pavlesen S, Kluczynski M, Baker JG, Miecznikowski JC, et al. A systematic review of criteria used to define recovery from sport-related concussion in youth athletes. Br J Sports Med 2018;52:1179–90. doi: 10.1136/bjsports-2016-096551.

[3] Manley G, Gardner AJ, Schneider KJ, Guskiewicz KM, Bailes J, Cantu RC, et al. A systematic review of potential long-term effects of sport-related concussion. Br J Sports Med 2017;51:969–77. doi: 10.1136/bjsports-2017-097791.

[4] Schneider KJ, Leddy JJ, Guskiewicz KM, Seifert T, McCrea M, Silverberg ND, et al. Rest and treatment/rehabilitation following sport-related concussion: a systematic review. Br J Sports Med 2017;51:930–4. doi: 10.1136/bjsports-2016-097475.

[5] Junn C, Bell KR, Shenouda C, Hoffman JM. Symptoms of concussion and comorbid disorders. Curr Pain Headache Rep 2015;19:46. doi: 10.1007/s11916-015-0519-7.

[6] Karver CL, Wade SL, Cassedy A, Taylor HG, Stancin T, Yeates KO, et al. Age at injury and long-term behavior problems after traumatic brain injury in young children. Rehabil Psychol 2012;57:256–65. doi: 10.1037/a0029522.

[7] Knell G, Burkhart SO, Caze TJ, Polousky JD, Kohl 3rd HW, Messiah SE. Association between concussion history and factors relating to cognitive, behavioral, and emotional health among American high school athletes: a cross-sectional analysis. Am J Sports Med 2020;48:2534–43. doi: 10.1177/0363546520938776.

[8] McKinlay A, Dalrymple-Alford JC, Horwood LJ, Fergusson DM. Long term psychosocial outcomes after mild head injury in early childhood. J Neurol Neurosurg Psychiatry 2002;73:281–8. doi: 10.1136/jnnp.73.3.281.

[9] Sariaslan A, Sharp DJ, D'Onofrio BM, Larsson H, Fazel S. Long-term outcomes associated with traumatic brain injury in childhood and adolescence: a nationwide Swedish cohort study of a wide range of medical and social outcomes. PLoS Med 2016;13:e1002103. doi: 10.1371/journal.pmed.1002103.

[10] Valovich McLeod TC, Wagner AJ, Bacon CEW. Lived experiences of adolescent athletes following sport-related concussion. Orthop J Sports Med 2017;5:2325967117745033. doi: 10.1177/2325967117745033.

[11] Chen JH, Asch SM. Machine learning and prediction in medicine - beyond the peak of inflated expectations. N Engl J Med 2017;376:2507–9. doi: 10.1056/NEJMp1702071.

[12] Fisher AJ, Bosley HG, Fernandez KC, Reeves JW, Soyster PD, Diamond AE, et al. Open trial of a personalized modular treatment for mood and anxiety. Behav Res Ther 2019;116:69–79. doi: 10.1016/j.brat.2019.01.010.

[13] Quinn DK, Mayer AR, Master CL, Fann JR. Prolonged postconcussive symptoms. Am J Psychiatry 2018;175:103–11. doi: 10.1176/appi.ajp.2017.17020235.

[14] Covassin T, Savage JL, Bretzin AC, Fox ME. Sex differences in sport-related concussion long-term outcomes. Int J Psychophysiol 2018;132:9–13. doi: 10.1016/j.ijpsycho.2017.09.010.

[15] Anzalone AJ, Blueitt D, Case T, McGuffin T, Pollard K, Garrison JC, et al. A positive vestibular/Ocular Motor Screening (VOMS) is associated with increased recovery time after sports-related concussion in youth and adolescent athletes. Am J Sports Med 2017;45:474–9. doi: 10.1177/0363546516668624.

[16] Eagle SR, Puligilla A, Fazio-Sumrok V, Kegel N, Collins MW, Kontos AP. Association of time to initial clinic visit with prolonged recovery in pediatric patients with concussion. J Neurosurg Pediatr 2020;26:1–6. doi: 10.3171/2020.2.PEDS2025.

[17] Ellis MJ, Cordingley D, Vis S, Reimer K, Leiter J, Russell K. Vestibulo-ocular dysfunction in pediatric sports-related concussion. J Neurosurg Pediatr 2015;16:248–55. doi: 10.3171/2015.1.peds14524.

[18] Knell G, Caze T, Burkhart SO. Evaluation of the vestibular and ocular motor screening (VOMS) as a prognostic tool for protracted recovery following paediatric sports-related concussion. BMJ Open Sport Exerc Med 2021;7:e000970. doi: 10.1136/bmjsem-2020-000970.

[19] Sufrinko AM, Marchetti GF, Cohen PE, Elbin RJ, Re V, Kontos AP. Using acute performance on a comprehensive neurocognitive, vestibular, and ocular motor assessment battery to predict recovery duration after sport-related concussions. Am J Sports Med 2017;45:1187–94. doi: 10.1177/0363546516685061.

[20] Whitney SL, Eagle SR, Marchetti G, Mucha A, Collins MW, Kontos AP, et al. Association of acute vestibular/ocular motor screening scores to prolonged recovery in collegiate athletes following sport-related concussion. Brain Inj 2020;34:840–5. doi: 10.1080/02699052.2020.1755055.

[21] Mucha A, Collins MW, Elbin RJ, Furman JM, Troutman-Enseki C, DeWolf RM, et al. A Brief Vestibular/Ocular Motor Screening (VOMS) assessment to evaluate concussions: preliminary findings. Am J Sports Med 2014;42:2479–86. doi: 10.1177/0363546514543775.

[22] White OB, Fielding J. Cognition and eye movements. J Neuroophthalmol 2012;32:266–73. doi: 10.1097/wno.0b013e3182688230.

[23] Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019;380:1347–58. doi: 10.1056/NEJMra1814259.

[24] Huang T, Chu Y, Shams S, Kim Y, Annapragada AV, Subramanian D, et al. Population stratification enables modeling effects of reopening policies on mortality and hospitalization rates. J Biomed Inform 2021;119:103818. doi: 10.1016/j.jbi.2021.103818.

[25] Mooney SJ, Pejaver V. Big data in public health: terminology, machine learning, and privacy. Annu Rev Public Health 2018;39:95–112. doi: 10.1146/annurev-publhealth-040617-014208.

[26] Du J, Jia P, Dai Y, Tao C, Zhao Z, Zhi D. Gene2vec: distributed representation of genes based on co-expression. BMC Genomics 2019;20:82. doi: 10.1186/s12864-018-5370-x.

[27] Walker G, Kiefer A, Richards N, Klug E, Reed C, Silva P, et al. Machine learning for concussion recovery prognosis: a novel tool to empower proactive physician treatments. Pediatrics 2019;144 198–198. doi: 10.1542/peds.144.2_MeetingAbstract.198.
[28] Edouard P, Verhagen E, Navarro L. Machine learning analyses can be of interest to estimate the risk of injury in sports injury and rehabilitation. Ann Phys Rehabil Med 2020:101431.
[29] Galetta KM, Liu M, Leong DF, Ventura RE, Galetta SL, Balcer LJ. The King-Devick test of rapid number naming for concussion detection: meta-analysis and systematic review of the literature. Concussion 2016;1. doi: 10.2217/cnc.15.8.
[30] Simon M, Maerlender A, Metzger K, Decoster L, Hollingworth A, Valovich McLeod T. Reliability and concurrent validity of select C3 Logix test components. Dev Neuropsychol 2017;42:446–59. doi: 10.1080/87565641.2017.1383994.
[31] Meehan 3rd WP, Taylor AM, Berkner P, Sandstrom NJ, Peluso MW, Kurtz MM, et al. Division III collision sports are not associated with neurobehavioral quality of life. J Neurotrauma 2016;33:254–9. doi: 10.1089/neu.2015.3930.
[32] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. In: Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, editors. Advances in neural information processing systems, vol. 31. Curran Associates, Inc.; 2018.
[33] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Proceedings of the 31st international conference on neural information processing systems. Red Hook, NY, USA: Curran Associates Inc.; 2017. p. 4768–77.
[34] Lewis RJ. An Introduction to classification and regression tree (CART) analysis. 2000.
[35] Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Series B 2005;67:301–20. doi: 10.1111/j.1467-9868.2005.00503.x.
[36] Breiman L. Random forests. Mach Learn 2001;45:5–32. doi: 10.1023/A:1010933404324.
[37] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. New York, NY, USA: Association for Computing Machinery; 2016. p. 785–94. doi: 10.1145/2939672.2939785.
[38] Arik S.O., PfisterT. TabNet: attentive interpretable tabular learning. ArXiv [CsLG] 2019.
[39] He X, Zhao K, Chu X. AutoML: a survey of the state-of-the-art. Knowl Based Syst 2021;212:106622. doi: 10.1016/j.knosys.2020.106622.
[40] Vandekerckhove J, Matzke D, Wagenmakers E-J. Model comparison and the principle of parsimony. In: Busemeyer JR, Wang Z, Townsend JT, Eidels A, editors. The Oxford handbook of computational and mathematical psychology. Oxford University Press; 2015. doi: 10.1093/oxfordhb/9780199957996.013.14.