https://doi.org/10.47183/mes.2026-438

MACHINE LEARNING MODELS FOR STRATIFYING TRAUMATIC BRAIN INJURY SEVERITY BASED ON INTERLEUKIN-6 TEMPORAL DYNAMICS

Evgenii I. Balakin[1], Kseniya A. Yurku[1], Kristina A. Malsagova[2]✉, Tatiana V. Butkova[2], Anna L. Kaysheva[2], Vasiliy I. Pustovoit[1]

[1] State Research Center — Burnasyan Federal Medical Biophysical Center of Federal Medical and Biological Agency, Moscow, Russia
[2] Institute of Biomedical Chemistry, Moscow, Russia

Introduction. Traumatic brain injury (TBI) remains a serious healthcare problem due to its heterogeneity and unpredictable outcomes. In TBI, secondary inflammatory cascades are frequently associated with the level of interleukin-6 (IL-6). However, its diagnostic and prognostic value at various time points after injury requires further investigation in athletes.
Objective. To evaluate the diagnostic accuracy and prognostic significance of IL-6 in stratifying TBI severity at various time points after injury with the use of machine learning models.
Materials and methods. A prospective cohort study was conducted among 89 male athletes, with a mean age of 28.7 ± 5.3 years. These athletes were actively engaged in contact sports (mixed martial arts and kudo) and had documented evidence of concussion or mild to moderate brain contusion. Sequential blood plasma collection was performed at 3, 6, 12, and 24 h post-injury. IL-6 concentration was determined using validated enzyme-linked immunosorbent assay (ELISA) protocols. In order to classify injury severity, models were trained using gradient boosting (XGBoost), logistic regression, and random forest algorithms. The accuracy of model responses was evaluated using ROC analysis. To identify the most significant time points, the SHAP (SHapley Additive exPlanations) feature importance method was applied. For patient prediction and stratification, a model based on a long short-term memory (LSTM) recurrent neural network was implemented. Statistical validation used the Kruskal–Wallis H-test ($H = 31.77$; $p < 0.001$) and Spearman's correlation analysis ($r_s = 0.81$; $p < 0.001$).
Results. The greatest discriminatory ability of IL-6 concentrations for moderate brain contusion was noted at 6 and 12 h. The XGBoost model achieved an area under the curve of 0.92 [95% CI: 0.88; 0.96], with a sensitivity of 87% and a specificity of 84%. The SHAP analysis revealed that IL-6 values at 6 and 12 h had the greatest impact on the model's predictions. Logistic regression and random forest yielded areas under the curve of 0.84 and 0.88, respectively. The identified diagnostic window between 6 and 12 h post-injury coincides with the peak of neuroinflammatory activity.
Conclusions. The level of IL-6 measured within 6–12 h after TBI represents a specific biomarker for early stratification of injury severity. The integration of explainable machine learning approaches provides robust and clinically relevant decision support in neurotraumatology.

Keywords: IL-6; traumatic brain injury; machine learning; biomarker; long short-term memory; LSTM recurrent neural network; logistic regression; XGBoost

For citation: Balakin E.I., Yurku K.A., Malsagova K.A., Butkova T.V., Kaysheva A.L, Pustovoit V.I. Machine learning models for stratifying traumatic brain injury severity based on interleukin-6 temporal dynamics. *Extreme Medicine.* 2026;28(2):267–276. https://doi.org/10.47183/mes.2026-438

Funding: the work was performed within the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021–2030) (No. 125020701771-5).

Compliance with the ethical principles: the study was approved by the Ethics Committee of the Burnasyan Federal Medical Biophysical Center (minutes No. 40 dated November 18, 2020). All participants signed voluntary informed consent to participate in the study. The personal and biomedical data were processed in compliance with the General Data Protection Regulation (GDPR), including the anonymization of identifying information and the encryption of databases. All procedures complied with the provisions of GOST ISO/IEC 27001 regarding information security.

Potential conflict of interest: the authors declare no conflict of interest.

✉ Kristina. A. Malsagova kristina.malsagova86@gmail.com

Received: 17 Dec. 2025 **Revised:** 13 Feb.2026 **Accepted:** 4 Mar. 2026 **Online first:** 27 Apr. 2026

УДК 616.831:577.25:004.8

МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ ДЛЯ СТРАТИФИКАЦИИ ТЯЖЕСТИ ЧЕРЕПНО-МОЗГОВОЙ ТРАВМЫ НА ОСНОВЕ ВРЕМЕННОЙ ДИНАМИКИ ИНТЕРЛЕЙКИНА-6

Е.И. Балакин[1], К.А. Юрку[1], К.А. Мальсагова[2]✉, Т.В. Буткова[2], А.Л. Кайшева[2], В.И. Пустовойт[1]

[1] Государственный научный центр Российской Федерации — Федеральный медицинский биофизический центр им. А.И. Бурназяна, Москва, Россия
[2] Научно-исследовательский институт биомедицинской химии им. В.Н. Ореховича, Москва, Россия

Введение. Черепно-мозговая травма (ЧМТ) остается серьезной проблемой в области здравоохранения в связи с ее гетерогенностью и непредсказуемостью исходов. Среди молекулярных маркеров интерлейкин-6 (ИЛ-6) часто ассоциируется с каскадами вторичных воспалительных реакций. Однако его диагностическая и прогностическая ценность в различные временные точки после травмы остается недостаточно изученной у спортсменов.
Цель. Оценка диагностической точности и прогностической значимости ИЛ-6 при стратификации тяжести ЧМТ в различные временные точки после травмы с использованием моделей машинного обучения.
Материалы и методы. Проведено проспективное когортное наблюдение 89 мужчин-спортсменов, активно занимающихся контактными видами спорта (смешанные единоборства и кудо), средний возраст 28,7 ± 5,3 года, с документированным подтверждением сотрясения мозга, ушибом мозга легкой и средней степени тяжести. Осуществляли последовательный забор плазмы крови через 3, 6, 12 и 24 ч после травмы. Концентрация ИЛ-6 определена с помощью валидированных протоколов для иммуноферментного

анализа. Для решения задач классификации тяжести травмы были обучены модели на основе алгоритмов градиентного бустинга (XGBoost), логистической регрессии и случайного леса. Точность ответа моделей оценивалась с помощью ROC-анализа. Для определения наиболее значимых временных точек применялся метод оценки важности признаков SHAP (SHapley Additive exPlanations). Для прогнозирования и стратификации пациентов была реализована модель на основе рекуррентной нейронной сети с долгой краткосрочной памятью (LSTM). Статистическая валидация включала *H*-критерий Краскела – Уоллиса ($H = 31{,}77$; $p < 0{,}001$) и корреляционный анализ Спирмена ($r_s = 0{,}81$; $p < 0{,}001$).

Результаты. Концентрации ИЛ-6 через 6 и 12 ч продемонстрировали наибольшую дискриминационную способность при ушибе мозга средней степени тяжести. Модель XGBoost достигла AUC 0,92 [95% ДИ: 0,88; 0,96], чувствительности 87% и специфичности 84%. Анализ SHAP показал, что значения ИЛ-6 через 6 и 12 ч оказали наибольшее влияние на прогнозы модели. Логистическая регрессия и случайный лес дали AUC 0,84 и 0,88 соответственно. Выявленное диагностическое окно между 6 и 12 ч после травмы совпадает с пиком нейровоспалительной активности.

Выводы. Уровень ИЛ-6, измеренный в течение 6–12 ч после ЧМТ, представляет собой специфический биомаркер для ранней стратификации тяжести травмы. Интеграция объяснимых подходов машинного обучения обеспечивает надежную и клинически значимую поддержку принятия решений в нейротравматологии.

Ключевые слова: ИЛ-6; черепно-мозговая травма; машинное обучение; биомаркер; рекуррентная нейронная сеть с долгой краткосрочной памятью; логистическая регрессия; XGBoost

Для цитирования: Балакин Е.И., Юрку К.А., Мальсагова К.А., Буткова Т.В., Кайшева А.Л., Пустовойт В.И. Модели машинного обучения для стратификации тяжести черепно-мозговой травмы на основе временной динамики интерлейкина-6. *Экстремальная биомедицина*. 2026;28(2):267–276. https://doi.org/10.47183/mes.2026-438

Финансирование: работа выполнена в рамках Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021–2030 гг.) (№ 125020701771-5).

Соответствие принципам этики: исследование одобрено этическим комитетом Федерального медицинского биофизического центра им. А.И. Бурназяна ФМБА России (протокол № 40 от 18.11.2020). Все участники подписали добровольное информированное согласие на участие в исследовании. Обработка персональных и биомедицинских данных осуществлялась с соблюдением норм Европейского регламента по защите персональных данных (General Data Protection Regulation, GDPR), включая анонимизацию идентифицирующей информации и шифрование баз данных. Все процедуры соответствовали положениям ГОСТ ISO/IEC 27001 в части информационной безопасности.

Потенциальный конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

✉ Мальсагова Кристина Ахмедовна kristina.malsagova86@gmail.com

Статья поступила: 17.12.2025 **После доработки:** 13.02.2026 **Принята к публикации:** 04.03.2026 **Online first:** 27.04.2026

INTRODUCTION

Traumatic brain injury (TBI) represents a global medical and socioeconomic problem that causes disability and premature mortality in millions of patients every year. According to the World Health Organization, up to 69 million people worldwide annually sustain a TBI of varying severity, with the majority of cases occurring in the working-age population (15–44 years). The diagnosis and stratification of TBI remain clinically challenging, especially in the absence of obvious neuroimaging changes and due to the variable clinical picture in the subacute and intermediate periods [1–3].

Biochemical markers are currently not used for diagnosis in Russian neurotraumatology. According to clinical guidelines[1,2], TBI is diagnosed by assessing the level of consciousness on the Glasgow Coma Scale (GCS) and analyzing data from computed tomography or magnetic resonance imaging. However, visual diagnostic methods are expensive and not always accessible, as they require patients to visit specialized centers. Also, existing clinical assessment scales, including the GCS, have limited sensitivity to hidden neuronal tissue damage.

As a result, a growing interest is being shown in objective laboratory markers capable of reflecting the molecular and cellular processes behind neurotrauma. Among such biomarkers, particular attention is paid to interleukin-6 (IL-6), a pro-inflammatory cytokine that plays a key role in the pathogenesis of neuroinflammation and the activation of secondary injury cascades [4, 5]. Several studies show that IL-6 plasma levels correlate with TBI severity, brain edema development, and disease outcomes, especially within the first 24 h post-injury [6–8]. For example, the study by Hergenroeder et al. was one of the first works to demonstrate a correlation between serum IL-6 level and TBI severity [9]. Although these early studies used no machine learning tools, they laid the foundation for incorporating IL-6 into predictive models for TBI severity and outcomes [9].

However, in order to use IL-6 as a stratification tool in clinical practice, it is necessary to take into account its temporal dynamics in relation to other parameters [9, 10]. In this context, machine learning methods offer unique opportunities for constructing highly accurate predictive models capable of identifying nonlinear dependencies and determining the significance of individual time points. In particular, gradient boosting (XGBoost) and recurrent neural networks (LSTM), adapted for the analysis of time-series biomarkers, not only enhance

[1] Clinical guidelines “Mild Traumatic Brain Injury,” 2016. URL: https://ruans.org/Text/Guidelines/mild_head_injury.pdf
[2] Clinical guidelines “Concussion,” 2022. URL: https://legalacts.ru/doc/klinicheskie-rekomendatsii-sotrjasenie-golovnogo-mozga-utv-minzdravom-rossii/

diagnostic accuracy but also ensure the interpretability of decisions through the use of SHAP (SHapley Additive exPlanations) analysis [11–13].

The present study addresses the development and validation of machine learning models for stratifying TBI severity based on serial IL-6 measurements, with a focus on their prognostic significance and potential for clinical integration into early diagnostic algorithms.

The study aims to develop and interpret a machine learning model for stratifying TBI severity based on the temporal dynamics of IL-6 plasma concentrations.

MATERIALS AND METHODS

Study Design and Population

The present study is a prospective cohort observation conducted to investigate the dynamics of IL-6 and its role in stratifying TBI severity in the early stages after traumatic impact. The study included 89 male athletes, with a mean age of 28.7 ± 5.3 years, who were actively engaged in contact sports (mixed martial arts and kudo). Inclusion criteria were as follows: a traumatic brain injury (confirmed clinically and through medical history), the absence of concomitant somatic pathology, and written informed consent to participate in the study.

The participants were stratified into three study groups by the clinical severity of TBI: a group with cerebral concussion (n = 43), a group with mild brain contusion (n = 35), and a group with moderate brain contusion (n = 11). This classification was based on the GCS score, duration of unconsciousness, and the presence of focal neurological symptoms.

Biological sample collection and analysis

For subsequent analysis of IL-6 level, venous blood was collected from the cubital vein into vacuum tubes containing ethylenediaminetetraacetic acid (K2EDTA) as an anticoagulant (VACUETTE® K2EDTA, Greiner Bio-One, Austria) at 3, 6, 12, and 24 h post-injury (t3, t6, t12, and t24). Blood samples were processed according to centrifugation protocols (3000 g; 10 min at 4°C), followed by plasma aliquoting and storage at –80°C until analysis.

IL-6 concentrations were determined using an enzyme-linked immunosorbent assay (ELISA) with certified diagnostic kits (Accquant® HS-IL-6 ELISA, FineTest, China), manufactured in accordance with ISO 13485:2016 requirements. Laboratory analysis was performed at an accredited laboratory meeting the criteria of GOST ISO 15189:2015. Calibration curves were constructed based on standard samples, and quality control was ensured by using internal and external standards, which guaranteed an inter-assay variability of less than 10%.

Each IL-6 value was encoded as a time profile, enabling the use of the biomarker's dynamic characteristics in subsequent modeling. Missing values were assessed for their missingness mechanism and handled using multiple imputation algorithms. The complete dataset was normalized before being input into the machine learning models.

Definition of endpoints

In the stratification model, the primary clinical outcome variable (dependent variable) is TBI severity, determined based on a set of clinical criteria, including GCS scores, the presence and duration of unconsciousness, retrograde amnesia, as well as the severity of focal neurological symptoms. The participants were categorized into three diagnostic groups: cerebral concussion, mild brain contusion, and moderate brain contusion. The classification of injury severity was carried out in a neurological hospital setting according to standardized diagnostic protocols.

For the purpose of analysis, the categorical outcome was recoded into a binary variable: "non-severe TBI" (concussion + mild brain contusion) versus "severe TBI" (moderate brain contusion), with subsequent boundary validation based on predictive sensitivity modeling.

Machine Learning Tools

In order to implement a predictive tool for stratifying TBI severity based on time-series IL-6 concentrations, a combination of machine learning algorithms was applied. The primary model was built using XGBoost (Extreme Gradient Boosting) on tabular data, which included IL-6 values at four time points (3, 6, 12, and 24 h post-injury). The model hyperparameters were optimized using 5-fold cross-validation.

The interpretability of the model was assessed via the SHAP method, which provided a means to quantify the contribution of each time-specific IL-6 feature to the final classification. The greatest significance in predicting severe TBI outcomes was demonstrated by the IL-6 levels at 6 and 12 h post-injury (SHAP-value ≥ 0.37), which correlated with the phase of systemic inflammatory response.

In addition, an LSTM (Long Short-Term Memory) recurrent neural network was built for analysis of IL-6 time sequences (t3–t24) in order to model biomarker dynamics and predict outcomes. The input tensor had the shape [n, 4, 1], where n is the number of patients. The model was trained using a supervised classification scheme, with a stratified split of the dataset into training and testing sets in an 80:20 ratio. The model's performance metrics were accuracy, F1-score, and area under the ROC curve. For model validation, a confusion matrix was also used to calculate sensitivity and specificity.

The machine learning methods used in this work expand and complement the capabilities of classical statistical tools. Classical methods (Kruskal–Wallis, Mann–Whitney, and ROC analysis) were employed to identify intergroup differences and assess the diagnostic significance of individual parameters. In contrast, machine learning algorithms (XGBoost, LSTM, and SHAP) were used to build accurate predictive models for individual patients in order to identify complex nonlinear relationships and temporal patterns in the analyzed parameters that cannot be detected using classical statistical models. The machine learning results were interpreted via the SHAP method, which provides a deeper and more intuitive interpretation of the contribution of predictors compared to standard regression coefficients. The combination of statistical approaches provided a means to identify significant differences and the diagnostic value of markers, while also ensuring an accurate, personalized outcome prediction.

Statistical analysis

All quantitative data were tested for normality of distribution using the Shapiro–Wilk test. Since a deviation from normal distribution was detected ($p \leq 0.05$), non-parametric methods were used to assess differences between groups: the Kruskal–Wallis test for comparing multiple groups and the Mann–Whitney test with Bonferroni correction for pairwise comparisons. Spearman's rank correlation coefficient (r_s) was used to assess the strength of association between variables.

The diagnostic accuracy of IL-6 biomarkers was assessed using area under the curve analysis (ROC analysis). In order to calculate sensitivity, specificity, and Youden's index (J = Se + Sp – 1, where Se is sensitivity and Sp is specificity), we used values corresponding to the optimal threshold criterion, which is determined by the maximum Youden's index. The significance of differences between ROC curves was evaluated employing the DeLong test. The interpretation of AUC (area under the curve) values was carried out in accordance with recommendations for measuring the accuracy of diagnostic systems [14]: AUC 0.7–0.8 — fair accuracy; 0.8–0.9 — good; >0.9 — excellent.

In order to develop a model for stratifying the severity of traumatic brain injury, gradient boosting (XGBoost) was applied. The training and test sets were formed in a 70:30 ratio with stratification (stratified split) to ensure balanced representation of classes. Model hyperparameters were selected using stratified 5-fold cross-validation, in which the training set was sequentially divided into five equal subsets while maintaining class proportions in each. In each iteration, four subsets were used for model training, and one subset was used for quality assessment. Hyperparameter optimization was performed by minimizing the log-loss. The quality of the final model was evaluated on the test set using the following metrics: area under the ROC curve (AUC-ROC), precision, recall, and F1-score. The contributions of features to model predictions were interpreted via the SHAP method in its specialized implementation for tree ensembles (TreeSHAP).

For temporal prediction and patient stratification, a model based on an LSTM recurrent neural network was also implemented. Time series were generated from four measurement points (3, 6, 12, and 24 h) and were normalized using the z-score scale. Training was performed using the Adam optimizer and the categorical cross-entropy loss function. Early stopping, which relies on the loss value on the validation set, was used to combat overfitting. The accuracy of the model was compared with baseline algorithms (logistic regression and XGBoost).

All statistical calculations and visualizations were performed in the R environment (version 4.3.2) using the following packages: tidyverse, ggplot2, pROC, xgboost, SHAPforxgboost, keras, caret, and rstatix. Differences were considered statistically significant at $p \leq 0.05$, with adjustments for multiple comparisons where appropriate.

RESULTS AND DISCUSSION

A time profile analysis of IL-6 concentrations (t3, t6, t12, and t24) revealed significant intergroup differences depending on the clinical severity of TBI. Figure 1 presents the distributions of IL-6 values, indicating medians and interquartile ranges for the three cohorts: cerebral concussion, mild brain contusion, and moderate brain contusion.

The most pronounced differences between groups were noted at 6 and 12 h post-injury. According to the Kruskal–Wallis test, a statistically significant difference in IL-6 concentrations was observed at these time points between the three severity categories ($p < 0.001$). Conversely, the 3-h and 24-h time points did not reach the level of statistical significance after the Bonferroni correction ($p > 0.05$).

In pairwise comparisons (Mann–Whitney test), IL-6 concentrations in moderate TBI patients exceeded the corresponding values in patients with mild brain contusion and concussion by an average of 1.8–2.3 times at 6 and 12 h post-injury, indicating the presence of a diagnostically relevant window of inflammatory activity within 3–12 h after trauma.

Spearman's correlation analysis revealed a positive relationship between IL-6 level and severity in these time intervals ($r_s = 0.81$; $p < 0.001$). The most informative indicators were found in samples obtained at 6 and 12 h post-injury, which is consistent with the phase of systemic inflammatory response and justifies their inclusion in prognostic stratification algorithms (Fig. 2).

In order to develop a tool for stratifying TBI severity, an XGBoost model was built using IL-6 concentration

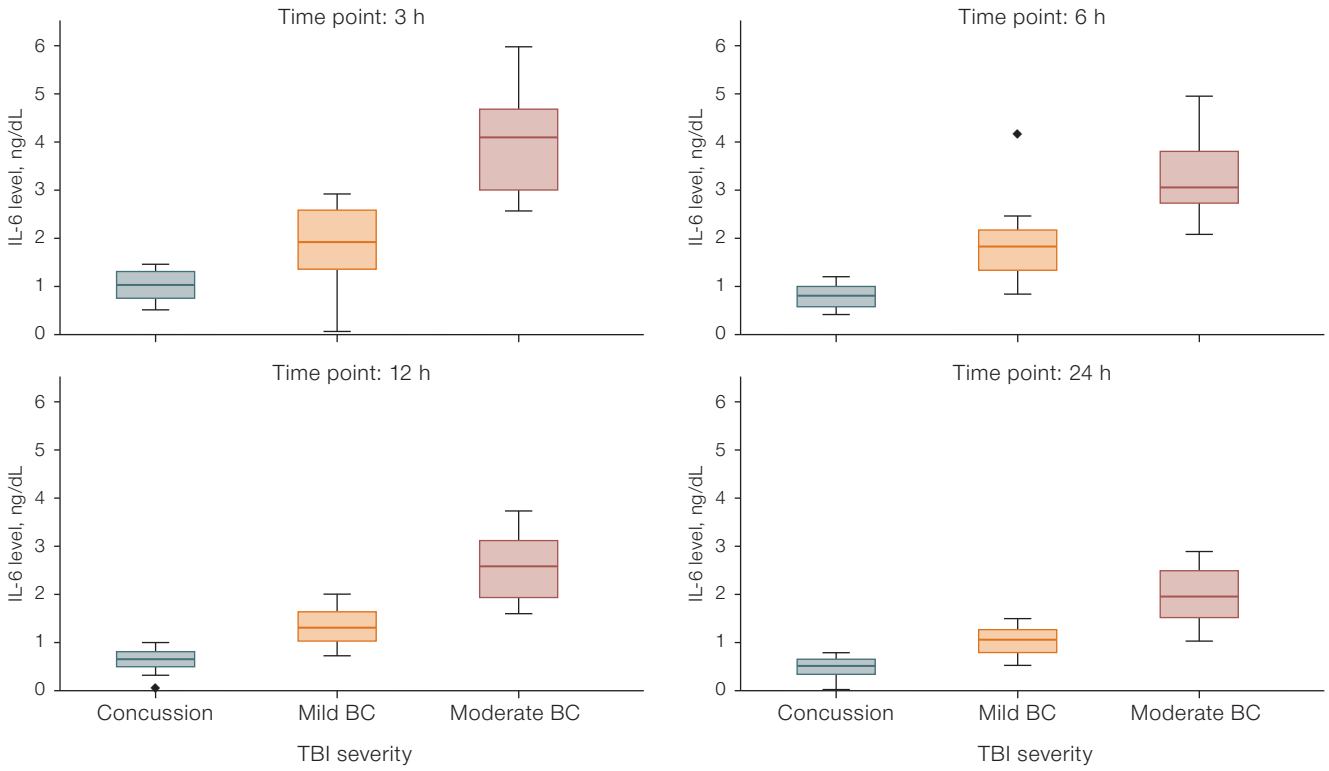

Figure prepared by the authors based on their own data

Fig. 1. Interleukin-6 level distribution in groups with concussion, mild brain contusion, and moderate brain contusion at 3, 6, 12, and 24 h post-injury: ⬥ — outliers beyond the interquartile ranges (i.e., observations in the dataset that significantly deviate from the other values); BC — brain contusion; TBI — traumatic brain injury

values at four time points as input parameters: 3, 6, 12, and 24 h post-injury.

Each row in the graph corresponds to one time point of IL-6 measurement (3, 6, 12, and 24 h post-injury), and the horizontal axis represents the SHAP value, i.e., the quantitative contribution of each feature to the final model prediction. Points to the right of the zero line (positive SHAP values) indicate an increased risk of being classified as severe TBI, while points to the left indicate a decreased risk. The color scale (from yellow to purple) represents the effect on the model response: yellow points correspond to samples with a low effect on the response, purple points to those with a significant effect. The pronounced shift in the distribution of points for IL-6_6h (IL6_t6) and IL-6_12h (IL6_t12) into the positive region (to the right), with maximum SHAP values ≥0.37, demonstrates a significant contribution and correlates with the phase of systemic inflammatory response in TBI.

A feature importance analysis based on the reduction in classification accuracy and the Gini index showed that the IL-6 values obtained at 6 and 12 h post-injury made the maximum contribution to the classification of severe TBI forms (Fig. 3). These time points were also identified as key in the non-parametric analysis ($H = 31.77$; $p < 0.001$ and $H = 28.91$; $p < 0.001$), indicating the reproducibility of the result across independent methodological platforms. The greatest contribution to classification is made by the IL-6 values recorded at 3 and 6 h post-injury, which confirms their diagnostic significance (Fig. 3).

The diagnostic accuracy of the model, measured by the AUC-ROC, amounted to 0.92 in a multi-class approach, indicating high prediction reliability. The SHAP analysis confirmed that at 6 and 12 h post-injury, the IL-6 values had a decisive influence on the classification of moderate TBI cases. For predicting the moderate form, an IL-6 threshold of >12.5 pg/mL was determined at these time points.

The modeling results underscore the prognostic significance of inflammatory response within the window between 6 and 12 h after injury. In clinical diagnostic algorithms, these biomarkers could enhance the sensitivity and specificity of stratification models and, consequently, increase the proportion of complications detected at early stages. The interpretability is ensured by means of SHAP visualization (Fig. 2), which provides grounds for considering the model to be clinically relevant and suitable for validation on external cohorts.

A comparison of predictive capabilities possessed by the machine learning algorithms revealed the superiority of non-linear ensemble models in classifying TBI severity based on time-series IL-6 concentrations. Within a

Figure prepared by the authors based on their own data

Fig. 2. SHAP analysis of the contribution of interleukin-6 time points to the XGBoost model for traumatic brain injury stratification: the color scale represents the interleukin-6 level, from low (yellow) to high (purple)

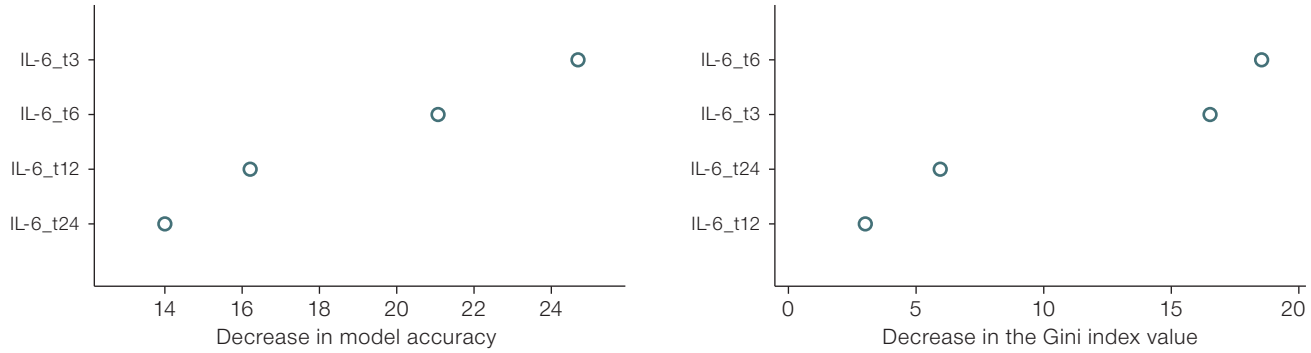

Figure prepared by the authors based on their own data

Fig. 3. Feature importance plots calculated based on the Random Forest algorithm: the left panel shows a decrease in model accuracy when the corresponding feature is excluded; the right panel shows a decrease in the Gini index value

stratified dataset built on the levels of neuroinflammatory response, four measurement points were used: 3, 6, 12, and 24 h.

The gradient boosting model (XGBoost) demonstrated the highest diagnostic accuracy (AUC = 0.92), which surpassed that of random forest (AUC = 0.88) and regularized logistic regression (AUC = 0.84). The differences between the models were statistically confirmed by the DeLong test (χ^2 = 12.41; p = 0.002). The optimal sensitivity and specificity for XGBoost reached 87% and 84%, respectively, at the maximum Youden's index.

The most significant contribution to the model output was made by IL-6 concentrations obtained at 6 and 12 h post-injury, which is consistent with the correlation analysis and non-parametric comparisons. These time points form a diagnostically sensitive window for stratifying TBI severity.

Summary data on the accuracy metrics of all models are presented in the table.

The presented data support the application of a comprehensive approach, which is based on multi-time IL-6 profiles, to stratifying neurotrauma severity in athletes from contact sports.

In order to analyze the IL-6 concentration time series and classify TBI severity, an LSTM model was implemented (Fig. 3). The input data consisted of sequences of IL-6 values obtained at 3, 6, 12, and 24 h post-injury. The model architecture was adapted to the tensor dimensions [n, 4, 1], where n is the number of participants.

The training was conducted using a supervised classification principle with an 80:20 dataset split, while

Table. Model accuracy metrics and the contribution of IL-6 sampling time to the classification of traumatic brain injury severity

Model	AUC [95% CI]	Sensitivity (%)	Specificity (%)	Youden's index	Most significant IL-6 sampling time, h
XGBoost	0.92 [0.87–0.96]	87	84	0.71	6; 12
Random Forest	0.88 [0.82–0.93]	82	80	0.62	6
Logistic Regression	0.84 [0.77–0.89]	78	76	0.54	12

Table compiled by the authors based on their own data

Note: AUC — Area Under the Curve; XGBoost (eXtreme Gradient Boosting) — machine learning model based on gradient boosting algorithms; IL-6 — interleukin-6.

maintaining the stratified distribution by severity levels. The model showed an average classification accuracy of 84.5%, and the F1-score for the group with moderate brain contusion reached 0.91. This indicates a high ability of the model to correctly identify the most severe cases, taking into account the temporal dynamics of the inflammatory biomarker.

A sensitivity analysis to feature exclusion revealed the key role of IL-6 levels measured at 6 h post-injury. Removal of this feature led to an AUC decrease of 0.08, confirming its critical importance in prediction. The visualized model outputs showed that the 6–12 h temporal pattern makes the greatest contribution to group differentiation, correlating with the previously identified window of systemic inflammatory response.

The obtained results suggest that recurrent neural networks are a promising tool for interpreting temporal biomolecular signals in TBI stratification. Such models can be implemented with the use of sequential biomarkers and standardized time points, which ensures not only high accuracy but also the clinical applicability of the algorithm.

The study results indicate a high prognostic value of IL-6 levels in stratifying TBI severity, particularly within the 6–12 h post-injury interval. This aligns with the established pathophysiological phases of neuroinflammatory response, during which IL-6 participates in cascade activation, leading to secondary neuronal and glial damage.

A comparison with the results of a prospective TRACK-TBI study shows the comparable significance of IL-6, glial fibrillary acidic protein (GFAP), and ubiquitin carboxy-terminal hydrolase L1 (UCHL1) in predicting outcomes in mild to moderate TBI [15]. Unlike GFAP and UCHL1, which primarily reflect structural damage to astrocytes and axons, IL-6 helps characterize the systemic inflammatory response, especially in early post-injury stages [4, 5, 7, 16]. Papa et al. show that GFAP/UCHL1 levels can predict the presence of intracranial abnormalities on computed tomography scans but do not cover the dynamic spectrum of immune activation [17–19].

In this study, the IL-6 levels at 6 and 12 h post-injury demonstrated the highest prognostic significance (SHAP value ≥ 0.37; $p < 0.001$), making a key contribution to the XGBoost model (Figs. 2, 3). This is consistent with the concept of "systemic vulnerability interval" described by Yue et al. [20], according to which early cytokine cascade activation is crucial in determining prognosis and injury outcomes.

The performed comparative analysis of the algorithms indicates the advantage of non-linear models, including XGBoost (AUC = 0.92) and LSTM (F1 = 0.91 for moderate brain contusion), over logistic regression and random forest (Table). The model interpretation through SHAP analysis confirms the key role of the 6 and 12-h time points in classifying severity. This makes IL-6 not only diagnostically significant but also a clinically applicable biomarker for monitoring in the subacute period.

The presence of a 6–12 h diagnostic window is also reflected in the visual distributions of IL-6 concentrations (Fig. 1), where statistically significant differences were observed at these time points between all values in the severity groups ($H = 31.77$; $p < 0.001$ and $H = 28.91$; $p < 0.001$, respectively). Thus, these time points can be considered critical in developing laboratory monitoring protocols and making clinical decisions.

Noteworthy is that, unlike most existing studies, the present modeling was based on a strictly stratified cohort with serial measurements, which increased the accuracy of predictions. However, the use of only one biomarker (IL-6) and the relatively small size of the subgroup with moderate brain contusion ($n = 11$) still impose limitations, requiring subsequent external validation on independent samples.

The conducted study confirms that the post-TBI interval between 6 and 12 h has the highest diagnostic value in assessing injury severity based on plasma IL-6 concentrations. It was during this period that the subgroup of patients with moderate brain contusion exhibited the maximum median IL-6 levels, which significantly exceeded those in the groups with mild injury and concussion (Fig. 1). The statistical significance of differences at 6 and 12 h post-injury was confirmed by the non-parametric Kruskal–Wallis test: $H = 31.77$;

$p < 0.001$ for the 6 h time point and $H = 28.91$; $p < 0.001$ for the 12 h time point.

The prognostic value of these time points was further confirmed within the construction of machine learning models. According to gradient boosting (XGBoost), IL-6 concentrations at 6 and 12 h post-injury had the highest importance values both in terms of the decrease in model accuracy and the Gini index (Fig. 3). Using SHAP analysis to interpret the model, we found that these two temporal features had the most significant effect on the model's predictions (maximum SHAP values ≥ 0.37). This approach ensured a high level of algorithm explainability, which is critical for application in clinical practice.

Along with XGBoost, the LSTM model adapted to the time-series biomarker data demonstrated high accuracy. Despite the small sample size ($n = 89$), it achieved an F1-score of 0.91 in the classification of TBI severity. This indicates the potential of LSTM neural networks in multiclass stratification and monitoring.

However, the methodology has several limitations. First, the study relies exclusively on IL-6 levels, without the use of other inflammatory, glial, and axonal biomarkers. This limits the potential for constructing multi-biomarker models, whose promise has been demonstrated in large cohort studies. Second, the sample size ($n = 89$) does not allow for robust external validation of the models, especially within subgroups where patients with moderate brain contusion account for less than 15% of the total cohort. Finally, the IL-6 levels were measured using only one analytical method (ELISA), which limits data reproducibility in interlaboratory settings.

Nevertheless, plasma IL-6 levels possess high diagnostic accuracy in predicting TBI severity, particularly within the diagnostically sensitive post-injury window between 6 and 12 h. This conclusion is confirmed not only via statistical methods but also by interpretable machine learning models, which provides a foundation for incorporating IL-6 into digital protocols for personalized monitoring and decision-making in sports medicine and military medical practice.

Clinical and regulatory aspects of implementing intelligent models in neurocritical care

Current research on TBI identification extensively explores machine learning approaches [21, 22], and machine learning algorithms are actively used to create predictive models for TBI patient outcomes. For instance, in the study by Cao et al. [23], XGBoost and SHAP were used to develop and validate a predictive model for in-hospital mortality in patients with isolated severe TBI and to identify the most influential predictors. This model proved to be highly accurate in predicting in-hospital mortality within the first five days based on injury severity, admission GCS score, and patient age.

For the prediction of GCS score dynamics in TBI patients, Nayebi et al. propose a model based on a deep recurrent neural network (including LSTM). This study showed that training a model on time-series data from TBI patients can be informative and improve prediction accuracy. The most important features were derived from the recurrent neural network model; their values show a clear trend for favorable and unfavorable outcomes. This study demonstrates the amount of information that can be obtained from time-series data for more accurate TBI prediction [24].

Using artificial intelligence algorithms to stratify TBI severity opens up prospects for integrating intelligent clinical decision support systems (CDSS) into the practice of neurocritical care. Validated models built on data regarding IL-6 concentrations and other biomarkers provide a means to rapidly assess the likelihood of developing adverse neurological outcomes and determining injury severity in the early stages post-incident.

In order to ensure regulatory compatibility of the proposed solutions, compliance with international standards for the reliability and interpretability of AI/ML approaches is of particular importance. Specifically, the structure of the developed algorithms and their logic align with the principles of ISO/IEC 23053:2022. This standard provides a framework for describing AI systems incorporating machine learning, with provisions for ensuring transparency, robustness, and ethical acceptability of digital solutions. Adherence to such a framework is especially critical when algorithms are applied in contexts of clinical uncertainty and under life-threatening conditions.

A key condition for the clinical acceptability of intelligent models is their interpretability by medical professionals. This study utilized explainable AI (XAI) methods, including SHAP value analysis, which enables a personalized assessment of the contribution of each feature (specifically, IL-6 levels at 6 or 12 h post-injury) to the final model output. This allows an expert physician to verify the result and make informed clinical decisions.

Noteworthy is that to be used in clinical settings, algorithms must be validated on independent samples. In this study, the generalizability of results was limited by the use of a single cohort. The transition to clinical application would require multicenter studies with external datasets to confirm the model's robustness across various populations and laboratory settings. Only with reproducible accuracy (AUC > 0.85) and a controlled error rate (as per ISO 14971 [8]) can such algorithms be included in regulated patient management protocols for TBI.

CONCLUSION

The study developed a predictive model that uses gradient boosting and recurrent neural networks to analyze temporal inflammatory biomarkers within the first 24 h after TBI.

The conducted time profile analysis of IL-6 during the post-traumatic period in TBI patients revealed the

significance of 6 and 12-h post-injury time points in clinical severity stratification. The measured IL-6 concentrations demonstrated a stable association with clinical assessment and significant differences between groups according to non-parametric analysis results ($H = 31.77$; $p < 0.001$) and correlation coefficients with neurological status scales ($r_s = 0.81$; $p < 0.001$).

The developed machine learning models — XGBoost and LSTM recurrent neural networks — provided high stratification accuracy, which was confirmed by the area under the ROC curve values (AUC > 0.90) and robustness to overfitting during cross-validation. The most informative features in these models were the IL-6 levels within the specified critical intervals.

The SHAP feature importance method helped determine the predictive contribution of each time marker in the model. This approach provided a means to preserve the clinical interpretability of the model and ensure the physician's participation in the decision-making process.

The scientific significance of this study lies in the identification of critical time points for IL-6 measurement (6 and 12 h), which could potentially enable the standardization of future research. The diagnostically significant IL-6 concentrations identified in the early period correlate with TBI outcomes, confirming the hypothesis about the critical role of an early and exaggerated neuroinflammatory response in the pathogenesis of secondary brain injury. This shifts the focus of attention to the very earliest post-traumatic processes. Moreover, the new data can be used as input parameters for creating machine learning models.

In practice, the obtained results can be used as follows: additional diagnostic methods alongside GCS assessment and CT data; a decision-making tool in the absence of radiological diagnostic methods; a criterion for personalized therapy; markers of recovery. Furthermore, the study results will allow patients to be classified not only by the GCS score but also by early inflammatory response in the testing of new neuroprotective agents, thereby increasing the precision of research.

Early TBI diagnosis provides a means to identify critical patients, determine patient management strategy, and predict the course of rehabilitation. In this study, we quantitatively assessed the contribution of each time-specific IL-6 measurement to the prediction of TBI severity with the use of machine learning algorithms (including SHAP analysis), which ensured the clinical interpretability of the model and its potential for integration into a personalized system for neurotrauma diagnosis.

The study results advance laboratory TBI diagnostics that incorporates machine learning. In the clinical settings of emergency departments, neuro-intensive care units, and neurosurgery departments, such approaches will enable early diagnosis and assist in medical decision-making.

Thus, the obtained data demonstrate the potential for integrating intelligent models into personalized medical care systems for TBI patients. The use of inflammatory molecular markers in combination with algorithmic analysis of time-series data allows for more accurate prediction of neurological outcomes and the development of individualized strategies for post-traumatic rehabilitation.

References

1. Butkova TV, Malsagova KA, Nakhod VI, Petrovskiy DV, Izotov AA, Balakin EI, et al. Candidate Molecular Biomarkers of Traumatic Brain Injury: A Systematic Review. *Biomolecules*. 2024;14(10):1283. https://doi.org/10.3390/biom14101283
2. VanItallie TB. Traumatic brain injury (TBI) in collision sports: Possible mechanisms of transformation into chronic traumatic encephalopathy (CTE). *Metabolism*. 2019;100S:153943. https://doi.org/10.1016/j.metabol.2019.07.007
3. Gerber KS, Alvarez G, Alamian A, Behar-Zusman V, Downs CA. Biomarkers of Neuroinflammation in Traumatic Brain Injury. *Clinical Nursing Research*. 2022;31(7):1203–18. https://doi.org/10.1177/10547738221107081
4. Ichwan K, Gazali S, Suherman S, Desiana D, Nurjannah N. Plasma interleukin 6 as an outcome predictor of traumatic brain injury patients. *Narra J*. 2023;3(3):e234. https://doi.org/10.52225/narra.v3i3.234
5. Li G, Liu H, He Y, Hu Z, Gu Y, Li Y, et al. Neurological Symptoms and Their Associations With Inflammatory Biomarkers in the Chronic Phase Following Traumatic Brain Injuries. *Frontiers in Psychiatry*. 2022;13:895852. https://doi.org/10.3389/fpsyt.2022.895852
6. Malik S, Alnaji O, Malik M, Gambale T, Farrokhyar F, Rathbone MP. Inflammatory cytokines associated with mild traumatic brain injury and clinical outcomes: a systematic review and meta-analysis. *Frontiers in Neurology*. 2023;14:1123407. https://doi.org/10.3389/fneur.2023.1123407
7. Ooi SZY, Spencer RJ, Hodgson M, Mehta S, Phillips NL, Preest G, et al. Interleukin-6 as a prognostic biomarker of clinical outcomes after traumatic brain injury: a systematic review. *Neurosurgical Review*. 2022;45:3035–54. https://doi.org/10.1007/s10143-022-01827-y
8. Yang J, Ran M, Li H, Lin Y, Ma K, Yang Y, et al. New insight into neurological degeneration: Inflammatory cytokines and blood-brain barrier. *Frontiers in Molecular Neuroscience*. 2022;15:1013933. https://doi.org/10.3389/fnmol.2022.1013933
9. Hergenroeder GW, Moore AN, McCoy JP, Samsel L, Ward NH, Clifton GL, et al. Serum IL-6: a candidate biomarker for intracranial pressure elevation following isolated traumatic brain injury. *Journal of Neuroinflammation*. 2010;7:19. https://doi.org/10.1186/1742-2094-7-19
10. Helmy A, Carpenter KLH, Menon DK, Pickard JD, Hutchinson PJA. The Cytokine Response to Human Traumatic Brain Injury: Temporal Profiles and Evidence for Cerebral Parenchymal Production. *Journal of Cerebral Blood Flow & Metabolism*. 2010;31(2):658–70. https://doi.org/10.1038/jcbfm.2010.142
11. Lundberg S, Lee S-I. A Unified Approach to Interpreting Model Predictions. NIPS'17: *Proceedings of the 31st International Conference on Neural Information Processing*

Systems. 2017;arXiv:1705.07874v2. https://doi.org/10.48550/ARXIV.1705.07874
12. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Scientific Reports*. 2018;8:6085. https://doi.org/10.1038/s41598-018-24271-9
13. Mekkodathil A, El-Menyar A, Naduvilekandy M, Rizoli S, Al-Thani H. Machine Learning Approach for the Prediction of In-Hospital Mortality in Traumatic Brain Injury Using Bio-Clinical Markers at Presentation to the Emergency Department. *Diagnostics.* 2023;13(15):2605. https://doi.org/10.3390/diagnostics13152605
14. Swets JA. Measuring the Accuracy of Diagnostic Systems. *Science.* 1988;240:(4857):1285–93. https://doi.org/10.1126/science.3287615
15. Gardner RC, Puccio AM, Korley FK, Wang KKW, Diaz-Arrastia R, Okonkwo DO, et al. Effects of age and time since injury on traumatic brain injury blood biomarkers: a TRACK-TBI study. *Brain Communications*. 2023;5(1):fcac316. https://doi.org/10.1093/braincomms/fcac316
16. Reyes J, Spitz G, Major BP, O'Brien WT, Giesler LP, Bain JWP, et al. Utility of Acute and Subacute Blood Biomarkers to Assist Diagnosis in CT-Negative Isolated Mild Traumatic Brain Injury. *Neurology*. 2023;101(20):e1992–2004. https://doi.org/10.1212/WNL.0000000000207881
17. Papa L, Brophy GM, Welch RD, Lewis LM, Braga CF, Tan CN, et al. Time Course and Diagnostic Accuracy of Glial and Neuronal Blood Biomarkers GFAP and UCH-L1 in a Large Cohort of Trauma Patients With and Without Mild Traumatic Brain Injury. *JAMA Neurology*. 2016;73(5):551–60. https://doi.org/10.1001/jamaneurol.2016.0039
18. Papa L, Ladde JG, O'Brien JF, Thundiyil JG, Tesar J, Leech S, et al. Evaluation of Glial and Neuronal Blood Biomarkers Compared With Clinical Decision Rules in Assessing the Need for Computed Tomography in Patients With Mild Traumatic Brain Injury. *JAMA Network Open*. 2022;5(3):e221302. https://doi.org/10.1001/jamanetworkopen.2022.1302
19. Papa L, McKinley WI, Valadka AB, Newman ZC, Nordgren RK, Pramuka PE, et al. Diagnostic Performance of GFAP, UCH-L1, and MAP-2 Within 30 and 60 Minutes of Traumatic Brain Injury. *JAMA Network Open*. 2024;7;(9):e2431115. https://doi.org/10.1001/jamanetworkopen.2024.31115
20. Yue JK, Kobeissy FH, Jain S, Sun X, Phelps RRL, Korley FK, et al. Neuroinflammatory Biomarkers for Traumatic Brain Injury Diagnosis and Prognosis: A TRACK-TBI Pilot Study. *Neurotrauma Reports*. 2023;4(1):171–83. https://doi.org/10.1089/neur.2022.0060
21. Malhotra AK, Shakil H, Smith CW, Huang YQ, Kwong JCC, Thorpe KE, et al. Predicting outcomes after moderate and severe traumatic brain injury using artificial intelligence: a systematic review. *NPJ Digital Medicine.* 2025;8:373. https://doi.org/10.1038/s41746-025-01714-y
22. Courville E, Kazim SF, Vellek J, Tarawneh O, Stack J, Roster K, et al. Machine learning algorithms for predicting outcomes of traumatic brain injury: A systematic review and meta-analysis. *Surgical Neurology International*. 2023;14:262. https://doi.org/10.25259/SNI_312_2023
23. Cao Y, Forssten MP, Sarani B, Montgomery S, Mohseni S. Development and Validation of an XGBoost-Algorithm-Powered Survival Model for Predicting In-Hospital Mortality Based on 545,388 Isolated Severe Traumatic Brain Injury Patients from the TQIP Database. *Journal of Personalized Medicine*. 2023;13(9):1401. https://doi.org/10.3390/jpm13091401
24. Nayebi A, Tipirneni S, Foreman B, Ratcliff J, Reddy CK, Subbian V. Recurrent Neural Network based Time-Series Modeling for Long-term Prognosis Following Acute Traumatic Brain Injury. *AMIA Annual Symposium Proceedings*. 2022;2021:900–9.

Authors' contributions. All authors confirm that their contributions meet the ICMJE criteria for authorship. The primary contributions are as follows: Evgenii I. Balakin — statistical analysis, development of the stratification model, writing and editing of the manuscript draft; Kseniya A. Yurku — collection of biological material, measurement of IL-6 concentration; Kristina A. Malsagova — writing and editing of the manuscript draft; Tatiana V. Butkova — material and technical support; Anna L. Kaysheva — conceptualization, data analysis; Vasiliy I. Pustovoit — research supervision, approval of the final manuscript.

Authors:

Evgenii I. Balakin, Cand. Sci. (Med.), ORCID: https://orcid.org/0000-0001-5545-135X
Kseniya A. Yurku, Cand. Sci. (Med.), ORCID: https://orcid.org/0000-0002-1973-1693
Kristina. A. Malsagova, Cand. Sci. (Biol.), ORCID: https://orcid.org/0000-0001-9404-1660
Tatiana V. Butkova, Cand. Sci. (Med.), ORCID: https://orcid.org/0000-0001-5111-3863
Anna L. Kaysheva, Dr. Sci. (Biol.), ORCID: https://orcid.org/0000-0003-4472-2016
Vasiliy I. Pustovoit, Dr. Sci. (Med.), ORCID: https://orcid.org/0000-0003-3396-5813