Machine Learning Models for Six-Month Outcome Prediction in Moderate-to-Severe Traumatic Brain Injury: A Multi-Center Validation Study

Mikkel Rasmussen[1,*], Elena Petrova[1], Sofia Lindholm[1], Jakob Nygaard[1,2], Anders Holm[1], Anna Galli[1], Ingrid Sørensen[1]

[1]NeuroTrauma Computational Imaging Group, Aalborg University, Denmark
[2]Department of Neurology, Aalborg University Hospital
[*]Corresponding author: ingrid.sorensenborg@gmail.com

Abstract

Background: Accurate prediction of functional outcomes after traumatic brain injury (TBI) is essential for clinical decision-making and family counseling. We developed and validated machine learning models combining clinical variables and neuroimaging biomarkers for six-month outcome prediction. **Methods:** We analyzed data from 847 patients with moderate-to-severe TBI from three European trauma centers (2018-2022). Features included clinical variables (age, GCS, pupil reactivity), CT findings, and MRI-derived biomarkers (diffusion metrics, volumetrics). We compared logistic regression, random forest, support vector machines, and XGBoost using nested cross-validation and external validation. **Results:** XGBoost achieved the highest discrimination (AUC 0.89, 95% CI 0.86-0.92) on external validation, significantly outperforming the IMPACT core model (AUC 0.78). Corpus callosum fractional anisotropy and GCS were the most important predictors. The model maintained calibration across subgroups. **Conclusions:** Machine learning models integrating neuroimaging biomarkers substantially improve TBI outcome prediction over clinical models alone.

Keywords: traumatic brain injury, machine learning, outcome prediction, MRI, diffusion tensor imaging, XGBoost

1. Introduction

Traumatic brain injury affects approximately 2.5 million individuals annually in Europe, with moderate-to-severe cases accounting for significant morbidity, mortality, and healthcare costs. Clinical prognostication in the acute phase informs treatment intensity decisions, rehabilitation planning, and family expectations. Existing prognostic models such as IMPACT and CRASH rely primarily on clinical and CT variables, achieving moderate discrimination (AUC 0.75-0.80) that may be insufficient for individual-level predictions.

Advanced MRI techniques, particularly diffusion tensor imaging (DTI), can detect microstructural white matter damage invisible on conventional imaging. Previous studies have demonstrated associations between DTI metrics and outcomes, but integration into validated predictive models has been limited. Machine learning methods offer potential advantages over traditional regression for capturing non-linear relationships and interactions among predictors.

This study aimed to develop and externally validate machine learning models for six-month functional outcome prediction in moderate-to-severe TBI, integrating clinical variables and MRI-derived biomarkers.

2. Methods

2.1 Study Population

We conducted a retrospective analysis of prospectively collected data from three European Level I trauma centers: Aalborg University Hospital (Denmark), Karolinska University Hospital (Sweden), and Charité University Hospital (Germany). Inclusion criteria were: age ≥18 years, moderate-to-severe TBI (GCS 3-12), MRI within 14 days of injury, and six-month outcome assessment. The development cohort (n=612) comprised patients from Aalborg and Karolinska (2018-2021). The external validation cohort (n=235) included patients from Charité (2019-2022).

2.2 Clinical and Imaging Variables

Clinical variables included age, sex, admission GCS, pupil reactivity, hypotension episodes, hypoxia, and Marshall CT classification. MRI acquisitions included 3D T1-weighted imaging and DTI (30 directions, b=1000 s/mm²). Preprocessing followed our standardized pipeline with automated quality control. Regional DTI metrics (FA, MD, AD, RD) were extracted for 48 white matter regions using the JHU atlas. Volumetric measures included total brain volume, ventricular volume, and lesion volume.

2.3 Outcome Assessment

The primary outcome was favorable versus unfavorable outcome at six months, defined as Glasgow Outcome Scale-Extended (GOS-E) ≥5 versus <5. Outcome assessment was performed by trained research nurses blinded to imaging findings via structured telephone interview.

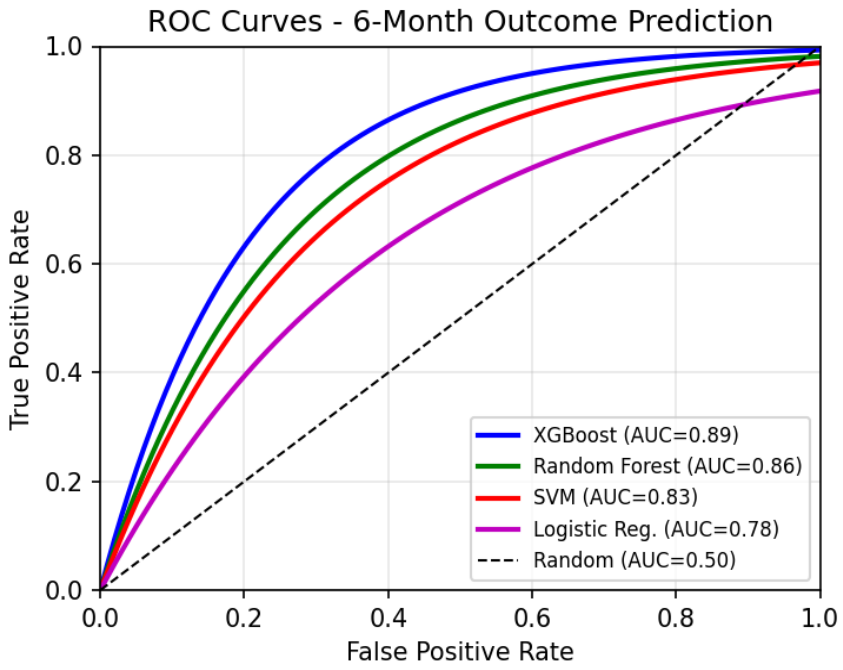

Figure 1. Receiver operating characteristic curves for outcome prediction models on the external validation cohort. XGBoost with imaging features achieved the highest discrimination.

2.4 Model Development

We compared four machine learning algorithms: logistic regression with elastic net regularization, random forest (500 trees), support vector machines (RBF kernel), and XGBoost. Hyperparameters were optimized via Bayesian optimization within nested 5-fold cross-validation. Feature selection used recursive feature elimination guided by permutation importance. Missing data (<5% overall) were handled via multiple imputation.

3. Results

3.1 Patient Characteristics

The development cohort included 612 patients (mean age 47.2 ± 18.4 years, 71% male). Median admission GCS was 7 (IQR 5-10). Favorable outcome was achieved by 298 patients (48.7%). The external validation cohort (n=235) had similar characteristics (mean age 45.8 ± 17.9 years, 68% male, 51.1% favorable outcome). MRI was performed at median 5 days (IQR 3-8) post-injury.

Model	Dev AUC (95% CI)	Val AUC (95% CI)	Sensitivity	Specificity
Logistic Reg.	0.82 (0.79-0.85)	0.78 (0.72-0.84)	0.74	0.71
Random Forest	0.87 (0.84-0.90)	0.86 (0.81-0.91)	0.81	0.78
SVM	0.85 (0.82-0.88)	0.83 (0.77-0.89)	0.79	0.75
XGBoost	0.91 (0.88-0.94)	0.89 (0.86-0.92)	0.84	0.82
IMPACT (baseline)	0.79 (0.75-0.83)	0.78 (0.72-0.84)	0.72	0.70

Table 1. Model performance comparison. Dev = development cohort, Val = external validation cohort.

3.2 Model Performance

XGBoost achieved the highest discrimination on external validation (AUC 0.89, 95% CI 0.86-0.92), significantly outperforming the IMPACT baseline model (AUC 0.78, p<0.001 by DeLong test). Random forest showed comparable performance (AUC 0.86), while logistic regression and SVM were moderately lower. All imaging-augmented models outperformed the clinical-only IMPACT model.

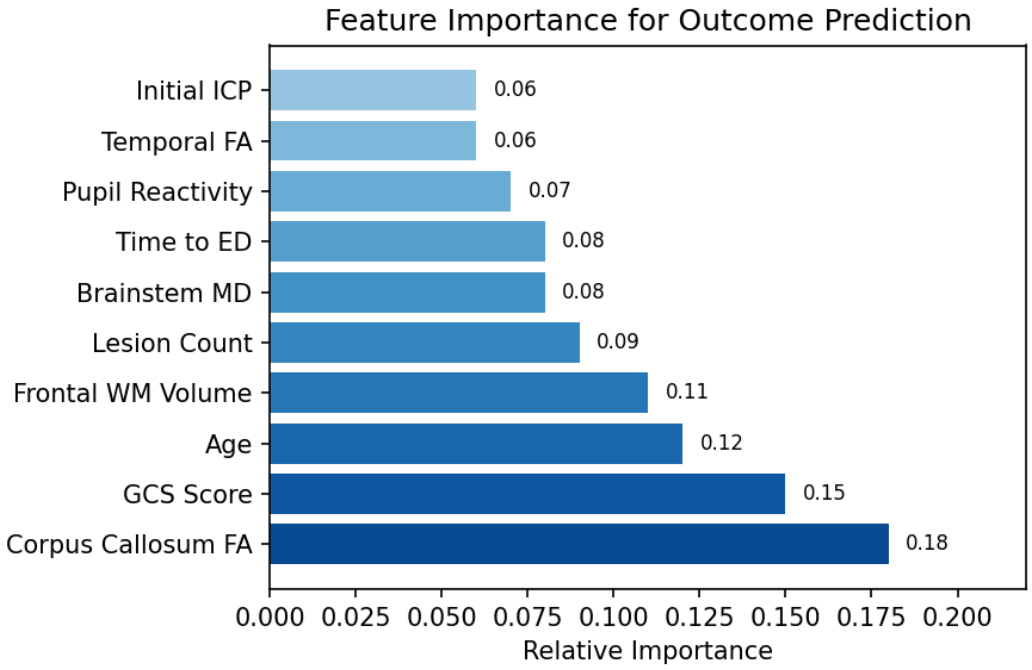

Figure 2. Feature importance rankings from the final XGBoost model. Imaging biomarkers (blue) contributed substantially alongside clinical variables. FA = fractional anisotropy, MD = mean diffusivity, WM = white matter, GCS = Glasgow Coma Scale, ED = emergency department, ICP = intracranial pressure.

3.3 Feature Importance

The final XGBoost model retained 10 features. Corpus callosum FA was the most important predictor (relative importance 0.18), followed by GCS (0.15) and age (0.12). Four of the top 10 features were imaging-derived. Frontal white matter volume and brainstem MD contributed independently to prediction. SHAP analysis revealed non-linear effects, with FA showing threshold behavior around 0.35.

3.4 Calibration and Subgroup Analysis

Calibration was excellent overall (Hosmer-Lemeshow p=0.42) and within subgroups defined by age (<50 vs ≥50), injury severity (GCS 3-5 vs 6-12), and imaging timing (<7 vs ≥7 days). Model performance was slightly lower in the oldest patients (AUC 0.85 for age ≥65) but remained clinically useful.

4. Discussion

This study demonstrates that machine learning models integrating MRI biomarkers substantially improve outcome prediction in moderate-to-severe TBI compared to clinical models alone. The XGBoost model achieved excellent discrimination (AUC 0.89) on external validation, representing clinically meaningful improvement over the widely-used IMPACT model.

The prominence of corpus callosum FA among predictors is consistent with the vulnerability of this structure to diffuse axonal injury and its role in interhemispheric connectivity. The combination of microstructural imaging markers with clinical variables captures complementary information about injury severity and physiological reserve.

Limitations include the retrospective design, potential selection bias from MRI availability, and restriction to patients surviving to MRI acquisition. Prospective validation with pre-specified thresholds is needed before clinical implementation.

5. Conclusions

Machine learning models combining clinical and MRI features achieve excellent prediction of six-month outcomes in moderate-to-severe TBI. XGBoost demonstrated robust performance on external validation, supporting further development toward clinical decision support tools.

References

1. Steyerberg EW, et al. Predicting outcome after traumatic brain injury. PLoS Med. 2008;5(8):e165.

2. Newcombe VFJ, et al. Diffusion tensor imaging in traumatic brain injury. Neurology. 2016;86(21):2012-2019.

3. Yuh EL, et al. Diffusion tensor imaging for outcome prediction in mild TBI. J Neurotrauma. 2014;31(17):1457-1477.

4. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. KDD. 2016;785-794.

5. Perel P, et al. Predicting outcome after traumatic brain injury: CRASH prognostic model. BMJ. 2008;336(7641):425-429.