# **Title Page**

**Title –** **Development and Validation of Machine Learning Models for Predictive Functional Outcomes in Traumatic Brain Injury: Insights from an Institutional Dataset and Deployment of a Clinical Tool**

**Authors:**

1. **Dr. Vinay Suresh, MBBS¹|#**Email: [dr.vinay.neuro@gmail.com](mailto:dr.vinay.neuro@gmail.com)
2. **Dr. Ankur Bajaj, MS, MCh (Neurosurgery)²|#**  
   Email: [ankurbajaj@kgmcindia.edu](mailto:ankurbajaj@kgmcindia.edu)
3. **Dr. Suhrudh Panchawagh, MBBS**³**|#**  
   Email: [suhrudp@gmail.com](mailto:suhrudp@gmail.com)
4. **Dr. Bhavik Bansal, MBBS⁴|^**  
   Email: [bansalbhavik@aiims.edu](mailto:bansalbhavik@aiims.edu)
5. **Dr. Sonit Vasipalli, MBBS⁵|^**  
   Email: [sonitsaiv@gmail.com](mailto:sonitsaiv@gmail.com)
6. **Dr. Devansh Mishra, MBBS¹**  
   Email: [devansh.mishra@kgmcindia.edu](mailto:devansh.mishra@kgmcindia.edu)
7. **Ms. Sahajmeet Kaur, BTech**⁶  
   Email: [sahajmeetkaur1308@gmail.com](mailto:sahajmeetkaur1308@gmail.com)
8. **Dr. Amogh Verma, MBBS⁷|\***  
   Email: [amoghverma2000@gmail.com](mailto:amoghverma2000@gmail.com)
9. **Dr. Rodrigue Ndabashinze, MD**⁸|\*  
   Email: [rodrigue5151@students.mu.ac.ke](mailto:rodrigue5151@students.mu.ac.ke)   
   ORCID: 0009-0001-6921-7828  
   *Corresponding author*
10. **Prof. Somil Jaiswal, MS, MCh (Neurosurgery)²**  
    Email: [dr.somil26@gmail.com](mailto:dr.somil26@gmail.com)
11. **Prof. B.K. Ojha, MS, MCh (Neurosurgery)²**  
    Email: [bkojha@rediffmail.com](mailto:bkojha@rediffmail.com)
12. **Dr. Vivek Sanker, MBBS9Email:** [viveksanker@gmail.com](mailto:viveksanker@gmail.com)
13. **Dr. Atman Desai, MD9  
    Email:** [atman@stanford.edu](mailto:atman@stanford.edu)
14. **Dr. Harminder Singh, MD, MBA9  
    Email:** [harman@stanford.edu](mailto:harman@stanford.edu)

## **Affiliations:**

¹ King George's Medical University, Lucknow, India  
² Department of Neurosurgery, King George's Medical University, Lucknow, India  
³ Smt. Kashibai Navale Medical College and General Hospital, Pune, India  
⁴ All India Institute of Medical Sciences, Delhi, India  
⁵ Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, India  
⁶ Indian Institute of Technology, Indore, India  
⁷ Rama Medical College Hospital and Research Centre, Hapur, India  
⁸ University of Burundi, Bujumbura, Burundi  
9 Department of Neurosurgery, Stanford University, CA, USA

**#Vinay Suresh, Ankur Bajaj and Suhrudh Panchawagh contributed equally to this work as co-first authors.**  
**^Bhavik Bansal and Sonit Vasipalli contributed equally to this work as co-second authors.**

**\*Corresponding Author:**

**Dr. Rodrigue Ndabashinze, MD**  
University of Burundi, Bujumbura, Burundi  
Email: [rodrigue5151@students.mu.ac.ke](mailto:rodrigue5151@students.mu.ac.ke)   
ORCID: 0009-0001-6921-7828

# **Statements and Declarations**

* **Declaration of Funding:** An **Intramural Seed Grant** (budget: **20,000 INR**) was allotted for this study to **Dr. Ankur Bajaj** by **The Research Cell of King George’s Medical University, Lucknow, India** (Reference number: **1120/R.Cell-16; Dated: 24/12/2016**).
* **Declaration of Financial Interests:** The authors declare **no financial interests** relevant to this study.
* **Declaration of Non-Financial Interests:** The authors declare **no non-financial interests** relevant to this study.
* **Ethical Approval:** Ethical approval for the study was obtained from the **Institutional Ethics Committee of** King George’s Medical University **(IEC Ref No. 5057/Ethics/R.Cell-17).**
* **Clinical Trial Registration Details/Number:** Not applicable, as this study **does not report a clinical trial**.
* **Research Registry Number:** Not applicable.
* **Human Ethics and Consent to Participate Declarations:** This study was conducted using a retrospective institutional dataset, and no direct patient recruitment occurred. Ethical approval was obtained from the Institutional Ethics Committee (IEC Ref No. 5057/Ethics/R.Cell-17), which waived the requirement for informed consent due to the retrospective nature of the study. All data were anonymized prior to analysis to ensure patient confidentiality, in accordance with institutional and regulatory guidelines.
* **Generative AI Use Statement:** Generative AI tools, including **Paperpal and ChatGPT-4o**, were utilized solely for **language refinement, grammar enhancement, and stylistic refinements**. These tools had **no role** in conceptualization, data analysis, interpretation of results, or substantive content development. All intellectual contributions, data analysis, and scientific interpretations remain **solely** the work of the authors. The final manuscript was critically reviewed and edited to ensure accuracy and originality. The authors take **full responsibility** for the integrity of the work presented.
* **Data Availability Statement:** This study utilized an **institutional dataset**, for which appropriate **ethical approval** was obtained. Due to **institutional policies**, the dataset **will not be made publicly available**. However, the **analysis code** used in this study can be accessed from the corresponding author upon reasonable request.

# **Word count: 7260 words.**

# **CRediT Author Statement**

* **Vinay Suresh (V.S.):** Conceptualization, Investigation, Methodology, Data Curation, Supervision, Methodology, Validation, Writing—Original Draft, Writing—Review & Editing.
* **Ankur Bajaj (A.B.):** Supervision, Methodology, Validation, Writing—Review & Editing.
* **Suhrudh Panchawagh (S.P.):** Investigation, Data Analysis, Data Curation, Visualization, Writing—Review.
* **Bhavik Bansal (B.B.):** Data Curation, Visualization, Writing—Original Draft, Writing—Review & Editing.
* **Sonit Vasipalli (S.V.):** Data Curation, Visualization, Writing—Original Draft, Writing—Review & Editing.
* **Devansh Mishra (D.M.):** Data Curation, Writing—Original Draft, Writing—Review & Editing.
* **Sahajmeet Kaur (S.K.):** Data Curation, Writing—Original Draft, Writing—Review & Editing.
* **Amogh Verma (A.V.):** Data Curation, Writing—Original Draft, Validation, Writing—Review & Editing.
* **Rodrigue Ndabashinze (R. N.):** Writing—Review & Editing.
* **Vivek Sanker (V. S):** Supervision, Writing—Review & Editing.
* **Somil Jaiswal (S.J.):** Supervision, Writing—Review & Editing.
* **B.K. Ojha (B.O.):** Supervision, Writing—Review & Editing.
* **Atman Desai (A.D.):** Writing—Review & Editing.
* **Harminder Singh (H.S.):** Writing—Review & Editing.

# **Acknowledgements**

The authors sincerely thank **Research Peer Network (RPN)** for their support and collaboration in this research. Additionally, we extend our gratitude to **Medicos In Research, Nautanwa, UP 273164, India,** for their invaluable guidance throughout the development of this manuscript. Their contributions played a crucial role in shaping this study.

**Total number of figures and tables:**

* **Main manuscript –** 3 figures and 2 tables.
* **In supplementary files –** 10 figures and 2 tables.

**Additional declarations:**

We the authors confirm the following –

1. The manuscript complies with all instructions to authors.

2. The authorship requirements have been met and the final manuscript was approved by all authors.

3. This manuscript has not been published elsewhere and is not under consideration by another journal

4. Any checklist that is required is attached and uploaded with the submission.

### 

**Manuscript**

**Title –** **Development and Validation of Machine Learning Models for Predictive Functional Outcomes in Traumatic Brain Injury: Insights from an Institutional Dataset and Deployment of a Clinical Tool**

**ABSTRACT**

Traumatic brain injury (TBI) presents a significant global health burden, with existing prognostic models often limited by linear assumptions and inadequate adaptability to heterogeneous clinical presentations. We developed and and validated machine learning (ML)-based predictive models for survival and functional recovery, utilizing an institutional dataset of 2,073 TBI patients. We employed logistic regression, random forest, support vector machine (SVM), and XGBoost algorithms, with performance evaluated via AUC-ROC, accuracy, and precision metrics. The random forest model demonstrated superior predictive performance (AUC-ROC 0.98 for survival; 0.88 for functional outcomes). SHAP analysis identified key predictors, including Rotterdam CT Grade, Glasgow Coma Scale, and intracranial hemorrhage volume. A deployable risk-scoring system was created for clinical integration. Despite limitations of retrospective design and single-center data, our findings highlight the potential of ML in TBI prognostication, emphasizing the need for prospective validation and broader outcome measures to enhance clinical applicability.

#### **Keywords:** Traumatic Brain Injury; Machine Learning; Outcome Prediction; SHAP Analysis; Clinical Risk Scoring

#### **Manuscript**

#### **Introduction**

Traumatic brain injury (TBI) is a critical global health issue, contributing significantly to morbidity, mortality, and long-term disability. The Global Burden of Disease (GBD) 2019 study estimated over 27 million new TBI cases globally each year, with nearly 49 million individuals living with its long-term effects and over 7 million years lived with disability (YLDs).[1,2](https://www.zotero.org/google-docs/?vrPSVd) Accurate prediction of TBI outcomes is critical for informing treatment strategies, optimizing resource allocation, and guiding prognostic discussions. Despite advancements in acute care, predicting outcomes in TBI remains a formidable challenge due to its heterogeneity and multifactorial nature[3](https://www.zotero.org/google-docs/?AGsoVc) and a highly variable clinical spectrum ranging from mild cases with full recovery to severe injuries resulting in death or lifelong impairment.[2](https://www.zotero.org/google-docs/?8wACf5)

Traditionally, TBI prognostication has relied on clinical scoring systems and statistical models. The Glasgow Coma Scale (GCS) is one of the most widely used tools for assessing the severity of brain injury due to its simplicity and broad adoption.[4](https://www.zotero.org/google-docs/?uy0BVy) However, it has notable limitations, including reduced utility in intubated patients and limited ability to predict long-term outcomes.[5](https://www.zotero.org/google-docs/?xsLmOn) Alternative bedside scoring systems, such as the Full Outline of UnResponsiveness (FOUR) score, have been proposed to address some of these shortcomings. However, these provide only a partial picture, often failing to capture the nuances required for personalized risk stratification or adapting to evolving patient characteristics.

General critical care scoring systems, such as the Acute Physiology and Chronic Health Evaluation II (APACHE-II) and the Simplified Acute Physiology Score II (SAPS-II), are also commonly used in TBI patients.[6,7](https://www.zotero.org/google-docs/?ISAmCO) These scores integrate physiological parameters and comorbidities to provide a broad assessment of illness severity but lack specificity for TBI.

Logistic regression and multivariable statistical models have been foundational in clinical research but are inherently limited in their ability to handle high-dimensional, non-linear data.[8,9](https://www.zotero.org/google-docs/?T0bgwP) Efforts to advance outcome prediction, including the Corticosteroid Randomization after Significant Head Injury (CRASH) and International Mission for Prognosis and Clinical Trials in Traumatic Brain Injury (IMPACT) studies, have led to the development of more sophisticated prognostic models. These models integrate clinical, imaging, and laboratory data but still rely on linear assumptions that may oversimplify the complex interactions underlying TBI pathophysiology.

Machine learning (ML) has emerged as a transformative tool for addressing the limitations of traditional approaches. ML algorithms excel at analyzing high-dimensional data with large p (features) and large n (sample size), uncovering non-linear relationships, and generating accurate predictions. Models such as random forests, support vector machines (SVM), and neural networks are particularly adept at handling non-linear interactions, feature redundancy, and missing data—common challenges in clinical datasets.[10](https://www.zotero.org/google-docs/?Olxi9Q) In the context of TBI, ML models have demonstrated superior performance in predicting outcomes such as mortality, functional recovery, and complications.[11](https://www.zotero.org/google-docs/?L6YDrR) Moreover, advancements in explainable AI (XAI) frameworks have improved the interpretability of these models, allowing clinicians to better understand the contributions of individual features to the predictions, thereby addressing concerns about the “black-box” nature of ML.[12,13](https://www.zotero.org/google-docs/?Xhu42u)

In this study, we utilized an institutional dataset of TBI patients to develop and validate ML-based models for survival and functional recovery, as measured by the Glasgow Outcome Scale (GOS). Our objectives included creating a clinically interpretable risk-scoring system based on feature contributions to facilitate integration into bedside decision-making.

**Methodology**

#### **Study Design and Dataset**

This study aimed to develop and validate predictive models for **survival** and **functional outcomes** (Glasgow Outcome Scale) after traumatic brain injury (TBI), using clinical, radiological, and demographic data. Ethical approval for the study was obtained from the Institutional Ethics Committee (IEC Ref No. 5057/Ethics/R.Cell-17). A formal study protocol was not prepared or published for this study. We adhered to the TRIPOD+ AI Statement guidelines for reporting clinical prediction models using regression or machine learning methods, as part of the EQUATOR Network's initiative for improving research reporting standards **(Table S2)**.[14](https://www.zotero.org/google-docs/?wqghTc)

Data were retrospectively collected from the Neurosurgery Department of a North Indian apex tertiary care hospital, including in-patient admissions from 2016 to 2019. Patients and the public were not involved in the design, conduct, or reporting of this study. An initial pool of 4700+ patients was considered for analysis. The diagnosis of TBI was made based on relevant ICD-10 codes (S01.7, S01.8, S01.9, S02.0, S02.1, S02.7, S02.8, S02.9, S06.0–S06.9, S07.1, S07.8, S07.9, S09.7, S09.8, S09.9, T02.0, T04.0, T06.0, T90.2, T90.5, T90.8, T90.9). The patients were managed according to the *National Guidelines for the Management of Traumatic Brain Injury* by the Neurotrauma Society of India.[15](https://www.zotero.org/google-docs/?tfGeKF) The dataset included information on the mechanism of injury, neurological findings, imaging results, and patient demographics. Outcome measures were based on survival status and the GOS score recorded three months after discharge.

The features considered in the analysis spanned clinical, radiological, and demographic domains. These included:

* **Demographics**: age, sex, alcohol consumption, time to first doctor, time in the trauma center.
* **Clinical presentation**: headache, seizures, vomiting, amnesia, CSF rhinorrhea, GCS components (GCS\_E, GCS\_V, GCS\_M), speech abnormalities, and pre-existing neurological deficiency.
* **Radiological findings**: Marshall CT scan, Rotterdam CT grade, intracranial hemorrhage volume, X-ray abnormalities.
* **Functional and motor parameters**: eye vision and power of limbs.

Categorical variables were coded as 0 or 1, ensuring consistency across the dataset. The Glasgow Outcome Scale classes were grouped into three categories based on the severity of recovery:

**Functional Outcome Prediction Based on GOS:**

* **GOS class 1, 2**: Poor recovery (Low Risk)
* **GOS class 3**: Moderate recovery (Intermediate Risk)
* **GOS class 4, 5**: Good recovery (High Risk)

#### **Data Preprocessing**

Preprocessing was conducted to address missing data, feature selection, class imbalance, and scaling. Variables and records with more than 10% missing data were excluded to maintain dataset integrity. For the remaining variables, missing values were imputed using **multiple imputation by chained equations (MICE)**, with 10 iterations to ensure convergence. During imputation, categorical variables were temporarily encoded numerically and subsequently decoded back to their original form. Irrelevant features were excluded based on exploratory data analysis and domain expertise.

To address class imbalance in the outcomes, oversampling was performed using **KMeansSMOTE**, which combines k-means clustering with synthetic minority oversampling to generate synthetic samples for the minority class. Sensitivity analyses were conducted through random undersampling to ensure robustness. Finally, feature scaling was performed using **RobustScaler**, which normalizes data based on the median and interquartile range, minimizing the impact of outliers.

After pre-processing the dataset, a total of 2,073 patients were included in the development of the ML-based prediction models. The sample size was sufficient to ensure training and validation of the models, addressing potential class imbalances and allowing reliable estimates of performance metrics.

#### **Model Development**

Separate predictive models were developed for survival and functional outcomes (GOS). The preprocessed dataset was split into training and testing subsets in a 70:30 ratio, with stratified sampling to preserve class distributions. The development and evaluation datasets were derived from the same source population and healthcare setting. Both datasets used identical eligibility criteria, outcome definitions, and predictor variables. Therefore, no systematic differences exist between the development and evaluation datasets.

Four machine learning algorithms were employed: **Logistic Regression**, **Random Forest**, **Support Vector Machine (SVM)**, and **XGBoost**. Each model was trained and evaluated independently to optimize performance and generalizability **(Table S1)**.

#### **Risk Scoring System**

Feature importance was evaluated using SHapley Additive exPlanations (SHAP), derived from the XGBoost model. SHAP values quantify the contribution of each feature to the prediction. We pushed our model to a GitHub repository and used Streamlit to deploy and use our model in clinical settings. This approach enabled patient stratification based on individual risk levels, providing interpretable and clinically actionable insights.

#### **Model Evaluation**

The models' performance was assessed using **accuracy**, **precision**, **recall** **(sensitivity)**, **specificity**, **F1-score**, and the **area** **under the receiver operating** **characteristic curve (AUROC)**. Bootstrap resampling with 2,000 iterations was employed to compute 95% confidence intervals for each metric. Confusion matrices were generated to visualize classification performance, offering a granular view of model efficacy.

**Results**

#### **Demographics and Patient Characteristics**

This study initially collected data from 4,700 patients with traumatic brain injury (TBI) across diverse clinical settings **(Table 1)**. Of these, 3,196 (68%) were between 30 and 50 years of age, underscoring the burden of TBI on the working-age population. After data preprocessing, 2,073 patients remained, with a mean age of 42.5 years (±15.7) and an overall age range of 18 to 85 years. The distribution by sex revealed a male-to-female ratio of 3:1, with males comprising 75% of the cohort. This aligns with established epidemiological patterns, where males are disproportionately affected due to high-risk activities such as driving, construction work, and other occupational hazards. Road traffic accidents emerged as the leading mechanism of injury, accounting for 60% of cases, followed by falls (25%), which were more prevalent among older adults, and assaults (15%), reflecting interpersonal violence. Patient outcomes were assessed using the Glasgow Outcome Scale (GOS), where 61% of patients achieved favorable outcomes (GOS 5), while 39% experienced poor outcomes (GOS 1–4), including a 5% mortality rate (GOS 1). These findings emphasize the multifaceted nature of TBI, influenced by demographic, occupational, and societal factors.

### **Survival Model**

The survival model demonstrated excellent predictive performance across various machine-learning algorithms. The Random Forest model outperformed others, achieving an AUC-ROC of 0.9800, an accuracy of 94.76%, and a precision of 98.51%, with high recall (90.91%) and specificity (98.62%). The confusion matrix and performance metrics **(Figure 1)** revealed minimal misclassifications, highlighting the model's robustness in predicting survival outcomes across different patient groups. The normalized scores for survival oversampling **(Figure S1)** provided insights into the distribution of model inputs and their scaling. SHAP analysis identified Rotterdam CT Grade, Glasgow Coma Scale (GCS), and Intracranial Hemorrhage Volume as the most significant predictors, emphasizing the importance of radiological and neurological factors.

The XGBoost model exhibited comparable performance, with an AUC-ROC of 0.9825, an accuracy of 94.48%, and a precision of 96.54%, alongside recall and specificity values of 92.29% and 96.69%. The SHAP bar and beeswarm plots collage **(Figure 3)** further illustrate the model's reliability. The ROC curve for survival oversampling **(Figure S2)** demonstrated the trade-off between sensitivity and specificity. On the undersampled dataset, XGBoost demonstrated an AUC-ROC of 0.9402 and an accuracy of 89.76%, indicating a slight drop in performance due to reduced data diversity. The Support Vector Machine (SVM) achieved an AUC-ROC of 0.9469, with sensitivity and specificity values of 91.46% and 88.40%, respectively. While slightly less effective than Random Forest and XGBoost, it demonstrated consistent classification of true positives. Logistic Regression, serving as the baseline model, achieved an AUC-ROC of 0.9433 and an accuracy of 87.45%.

**Functional Outcome Model**

For predicting functional outcomes, where Glasgow Outcome Scale (GOS) scores were dichotomized as GOS 5 (favorable outcome) and GOS 0–4 (unfavorable outcome) at three months post-injury or discharge, the Random Forest model again excelled. It achieved an AUC-ROC of 0.88, an accuracy of 74%, and an F1 Score of 0.81, with Pulse, Age, and Rotterdam CT Grade emerging as critical predictors. The confusion matrix and performance metrics **(Figure 2)** highlight the model's predictive capacity. The normalized scores for GOS oversampling **(Figure S3)** further validated the model's input distributions.

The XGBoost model performed similarly, achieving an AUC-ROC of 0.88, an accuracy of 73.5%, and an F1 Score of 0.81. The SHAP bar plots for Classes 1, 2, and Class 3 **(Figure Figure S8, Figure S9 and Figure S10)** provide insights into feature importance. The ROC curve for GOS oversampling **(Figure S4)** demonstrated the sensitivity-specificity trade-offs. On the undersampled dataset, XGBoost achieved an AUC-ROC of 0.84 and an accuracy of 69.8%, reflecting a moderate decrease in performance. Logistic Regression, while interpretable, showed restricted performance with an AUC-ROC of 0.83 and an accuracy of 68%.

### **Sensitivity Analysis**

Sensitivity analysis was conducted using the undersampled dataset to validate the robustness of the models trained on oversampled data. For survival predictions, the Random Forest model maintained strong performance with an AUC-ROC of 0.9460, an accuracy of 89.24%, and a recall of 88.17%, with minimal misclassifications observed in the confusion matrix (Figure 1). Predictors such as Rotterdam CT Grade, GCS, and Intracranial Hemorrhage Volume retained their importance. The ROC curve for survival undersampling **(Figure S5)** demonstrated the model's ability to balance sensitivity and specificity. The SVM model achieved an AUC-ROC of 0.9155 and an accuracy of 85.10%.

For functional outcomes, Random Forest achieved an AUC-ROC of 0.82 and an accuracy of 70% on the undersampled dataset, with Pulse, Age, and GCS remaining dominant predictors. Slightly higher misclassification rates compared to the oversampled dataset were observed **(Figure 2)**. The performance metrics and confusion matrix for GOS undersampling **(Figure S6 and S7)** provide additional validation.

### **Comparative Analysis**

A comparative analysis of model performances is summarized in **Table 2**. Ensemble methods, particularly Random Forest and XGBoost, consistently outperformed others across both survival and functional outcome predictions. The Random Forest model achieved the highest AUC-ROC values for survival (0.9800) and functional outcome predictions (0.88), demonstrating superior accuracy, precision, and recall across datasets. In contrast, SVM and Logistic Regression provided adequate performance but lacked the complexity to match ensemble methods in addressing the multifaceted nature of TBI datasets.

#### **SHAP Analysis and TBI Risk Scoring**

SHAP (SHapley Additive exPlanations) analysis provided invaluable insights into the feature importance and interpretability of the machine learning models. The global SHAP analysis emphasized the role of radiological features, such as Rotterdam CT Grade and Intracranial Hemorrhage Volume, in predicting survival outcomes. In contrast, physiological markers, particularly Pulse, were identified as key determinants for predicting Glasgow Outcome Scale (GOS) outcomes.

Local SHAP analysis offered individualized interpretations, revealing how specific features influenced predictions for each patient. For example, higher Rotterdam CT Grades were consistently associated with lower survival probabilities, while better GCS scores were linked to improved recovery outcomes. These local explanations provide personalized risk profiles for patients, enhancing the clinical utility of the models.

Visual representations of the SHAP results, including bar plots for GOS Classes 1 and 2 and Class 3 **(Figure S8, Figure S9, and Figure S10)** and the SHAP beeswarm plots **(Figure 3)**, offer intuitive insights into feature importance and the distribution of their impacts on predictions.

To facilitate clinical interpretability and bedside decision-making, we made a deployable model using Streamlit and pushed the model to a GitHub repository.[16](https://www.zotero.org/google-docs/?1DiZHE)

#### **Discussion:**

The present study represents the largest dataset from India focused on developing predictive models for key outcomes, such as survival and functional status (measured by the Glasgow Outcome Scale, GOS) at three months post-discharge in patients with TBI using non-linear machine learning models.

The field of TBI prognosis has seen significant advancements with the application of both traditional and machine learning models. Landmark studies like the CRASH and IMPACT models have been pivotal in shaping our understanding of TBI prognostication. The CRASH model, developed on a dataset exceeding 10,000 patients, incorporates predictors such as age, GCS, pupillary reactivity, and extracranial injuries, focusing on short-term (14-day mortality) and long-term (6-month unfavorable outcomes) predictions.[17](https://www.zotero.org/google-docs/?YX2qyq) Similarly, the IMPACT model builds on these predictors and includes CT classification, hypoxia, and hypotension to predict mortality and unfavorable outcomes at six months.[18](https://www.zotero.org/google-docs/?4IvpM3) These models have undergone extensive external validation and remain widely cited for their clinical utility, particularly in settings with resource constraints. [19](https://www.zotero.org/google-docs/?RcRoUK) However, limitations like reliance on logistic regression models and binary outcome predictions restrict their sensitivity to subtle variations in TBI recovery. Recent ML-based models have demonstrated superior performance with accuracy exceeding 80%, outperforming traditional methods by leveraging complex data interactions and optimizing for clinically relevant predictions.[20,21](https://www.zotero.org/google-docs/?Tud2Rl)

Our analysis included a comprehensive cohort of TBI patients, capturing detailed demographic, clinical, physical examination, laboratory, and imaging data. Leveraging this dataset, we aimed to build predictive models using advanced machine learning techniques, including Random Forests (RF), Support Vector Machines (SVMs), and XGBoost, and benchmarked their performance against traditional logistic regression models.

Our findings underscore the superiority of ensemble machine learning approaches—specifically Random Forests and XGBoost—which can account for complex, non-linear relationships among predictor variables. In our study, these models consistently outperformed conventional logistic regression in predicting both survival and functional outcomes at three months. By incorporating non-linearities and interactions that might be overlooked by simpler models, these methods demonstrate the potential of advanced analytics to enhance prognostic accuracy in TBI management, as evidenced by recent ML-based predictive models in literature. Importantly, we also developed streamlined prediction scoring tools that can be deployed in clinical settings, enabling clinicians to make rapid and informed decisions at the bedside. These simplified tools, grounded in key predictors, ensure model transparency, provide an opportunity to improve clinical assessments, guide triage decisions, and ultimately, influence patient outcomes. It allows for individual patient predictions through the HTML tool by inputting patient data. For batch data, the feature weights from our models (provided in the supplementary file) can be applied. Poor-quality or missing data should be handled using imputation techniques, with excessive missing values potentially excluded. Synthetic data generation, such as kMeansSMOTE, can address class imbalance.

We identified imaging parameters like Rotterdam CT Grade and Intracranial Hemorrhage volume, clinical parameters like GCS and Pulse, and demographic parameters like age being key predictors in prognosticating TBI outcomes in our models. The consistent predictors across both traditional and ML-based studies—age, CT parameters, and GCS—underscore their critical role in TBI outcome determination. Age, as a surrogate for physiological resilience, has been consistently linked to worse outcomes in both survival and long-term recovery. [22,23](https://www.zotero.org/google-docs/?BhhO5V) GCS, particularly its motor component, reflects the severity of neurological impairment and remains one of the strongest predictors in both traditional and ML models. CT features, such as the condition of cisterns and the presence of subarachnoid hemorrhage, provide insights into structural brain injury and its prognostic implications. Our study’s findings align with this existing literature, reaffirming the centrality of these predictors in TBI risk stratification.

The association of clinical variables like GCS and pulse rate, combined with imaging markers such as the Rotterdam CT Grade, may serve as proxy indicators for elevated intracranial pressure. This aligns with the existing belief that links these parameters to the physiological response to increased intracranial pressure and secondary brain injury mechanisms. The GCS score may reflect the severity of neurological impairment, while elevated pulse rates may indicate autonomic dysregulation, commonly associated with severe TBI. Similarly, the Rotterdam CT Grade and hemorrhage volume provide structural insights into brain injury severity and its potential impact on intracranial dynamics, which may predict both survival and long-term functional status. While most studies, including ours, focus on survival and functional outcomes like the Glasgow Outcome Scale (GOS), a growing body of research highlights the importance of long-term cognitive and motor function, as well as quality of life (QoL). Recent studies also explored rehabilitation-specific milestones and identified key predictors like days to rehabilitation admission and advanced mobility activities, demonstrating the potential of ML in identifying tailored intervention strategies.[24](https://www.zotero.org/google-docs/?2zD2KT) These findings emphasize the need for future studies to expand beyond traditional outcomes, integrating holistic measures that capture the full spectrum of TBI recovery and enabling more comprehensive patient care pathways.

Our findings corroborate observations in the literature, where combinations of clinical, imaging, and demographic variables have been shown to improve prognostication in TBI. By incorporating these predictors into advanced predictive models, our study bridges the gap between established knowledge and the potential for enhanced clinical application, ultimately supporting better-informed decision-making in the management of TBI patients.

***Strengths and Limitations:***

This study demonstrates several strengths that contribute to its impact on traumatic brain injury (TBI) outcome prediction. It leverages a large dataset of 2,073 patients, one of the largest cohorts for TBI outcomes in India, ensuring robust statistical power. The use of advanced machine learning techniques, including Random Forest and XGBoost, combined with SHAP-based interpretability, addresses the limitations of traditional statistical models by capturing complex, non-linear interactions. Moreover, the development of a clinically usable risk-scoring tool bridges the gap between computational advancements and bedside decision-making, making the findings immediately relevant for real-world application. The integration of multidimensional predictors, encompassing demographic, clinical, radiological, and functional parameters, offers a comprehensive evaluation of factors influencing TBI outcomes, further enhancing the predictive accuracy and clinical utility. The incorporation of SHAP analysis provides transparency, enabling clinicians to understand and trust the predictive models by offering insights into feature importance.

Despite these strengths, the study is not without limitations. Its retrospective design may introduce inherent biases. Additionally, the single-center nature of the dataset limits external validity, as the findings may not fully apply to diverse healthcare settings and populations. While internal validation was robust, our inability to externally validate using independent datasets puts the model’s credibility in broader contexts into question. Although imputation techniques were employed to address missing data, variables with high proportions of missing values were excluded, potentially omitting critical predictors. Furthermore, the scope of outcomes was restricted to survival and functional recovery measured by the Glasgow Outcome Scale (GOS), overlooking other potentially important dimensions such as quality of life, cognitive deficits, and psychological well-being in this cohort of patients.

Furthermore, the issue of class imbalance in our dataset required the use of kMeansSMOTE to generate synthetic data. While this approach helps address the imbalance, it may influence the generalizability of the model. Additionally, the need for parameter tuning, such as the number of clusters in kMeansSMOTE, adds computational overhead. [Zhou and Sun et al.]

From a methodological perspective, while tree-based models like Random Forests and Gradient Boosted Trees (e.g., XGBoost) were chosen for their efficiency and ability to handle structured data, they may not match the predictive potential of more advanced deep sequential neural networks or convolutional neural network models. However, the use of tree-based models offers an important practical advantage: they are computationally less demanding and often better optimized for structured datasets, making them more accessible for real-world clinical applications. In contrast, artificial neural networks (ANNs), though potentially superior in performance, require significantly more computational resources due to their iterative optimization and dependence on multiple layers and features, making them less feasible in resource-constrained settings. Balancing these trade-offs underscores the need for continued refinement and validation of these models.

Future studies should incorporate multicenter datasets and prospective designs to enhance generalizability and mitigate biases. External validation with independent datasets and the inclusion of more comprehensive outcome measures would provide a holistic understanding of TBI recovery. Despite its limitations, this study represents a significant advance in TBI research by introducing clinically interpretable, AI-driven predictive tools. It challenges existing paradigms, fosters evidence-based decision-making, and sets the stage for integrating advanced analytics into routine clinical practice.

**References**

[1. Guan B, Anderson DB, Chen L, Feng S, Zhou H. Global, regional and national burden of traumatic brain injury and spinal cord injury, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *BMJ Open*. 2023;13(10):e075049. doi:10.1136/bmjopen-2023-075049](https://www.zotero.org/google-docs/?w60tu1)

[2. Maas AIR, Menon DK, Adelson PD, et al. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *Lancet Neurol*. 2017;16(12):987-1048. doi:10.1016/S1474-4422(17)30371-X](https://www.zotero.org/google-docs/?w60tu1)

[3. Stocchetti N, Carbonara M, Citerio G, et al. Severe traumatic brain injury: targeted management in the intensive care unit. *Lancet Neurol*. 2017;16(6):452-464. doi:10.1016/S1474-4422(17)30118-7](https://www.zotero.org/google-docs/?w60tu1)

[4. Teasdale G, Jennett B. ASSESSMENT OF COMA AND IMPAIRED CONSCIOUSNESS: A Practical Scale. *The Lancet*. 1974;304(7872):81-84. doi:10.1016/S0140-6736(74)91639-0](https://www.zotero.org/google-docs/?w60tu1)

[5. Teasdale G, Maas A, Lecky F, Manley G, Stocchetti N, Murray G. The Glasgow Coma Scale at 40 years: standing the test of time. *Lancet Neurol*. 2014;13(8):844-854. doi:10.1016/S1474-4422(14)70120-6](https://www.zotero.org/google-docs/?w60tu1)

[6. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13(10):818-829.](https://www.zotero.org/google-docs/?w60tu1)

[7. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA*. 1993;270(24):2957-2963. doi:10.1001/jama.270.24.2957](https://www.zotero.org/google-docs/?w60tu1)

[8. Statistics for High-Dimensional Data: Methods, Theory and Applications | SpringerLink. Accessed December 18, 2024. https://link.springer.com/book/10.1007/978-3-642-20192-9](https://www.zotero.org/google-docs/?w60tu1)

[9. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition | SpringerLink. Accessed December 18, 2024. https://link.springer.com/book/10.1007/978-0-387-84858-7](https://www.zotero.org/google-docs/?w60tu1)

[10. Chen JH, Asch SM. Machine Learning and Prediction in Medicine — Beyond the Peak of Inflated Expectations. *N Engl J Med*. 2017;376(26):2507. doi:10.1056/NEJMp1702071](https://www.zotero.org/google-docs/?w60tu1)

[11. Courville E, Kazim SF, Vellek J, et al. Machine learning algorithms for predicting outcomes of traumatic brain injury: A systematic review and meta-analysis. *Surg Neurol Int*. 2023;14:262. doi:10.25259/SNI\_312\_2023](https://www.zotero.org/google-docs/?w60tu1)

[12. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: *Advances in Neural Information Processing Systems*. Vol 30. Curran Associates, Inc.; 2017. Accessed November 24, 2024. https://papers.nips.cc/paper\_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html](https://www.zotero.org/google-docs/?w60tu1)

[13. Ribeiro MT, Singh S, Guestrin C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’16. Association for Computing Machinery; 2016:1135-1144. doi:10.1145/2939672.2939778](https://www.zotero.org/google-docs/?w60tu1)

[14. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378. doi:10.1136/bmj-2023-078378](https://www.zotero.org/google-docs/?w60tu1)

[15. NATIONAL GUIDELINES FOR THE MANAGEMENT OF TRAUMATIC BRAIN INJURY. Accessed February 11, 2025. https://neurosocietyindia.com/national-guidelines-for-the-management-of-traumatic-brain-injury/](https://www.zotero.org/google-docs/?w60tu1)

[16. Panchawagh S. suhrudp/tbi\_survival\_prediction. Published online February 9, 2025. Accessed February 11, 2025. https://github.com/suhrudp/tbi\_survival\_prediction](https://www.zotero.org/google-docs/?w60tu1)

[17. Collaborators MCT. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. *BMJ*. 2008;336(7641):425-429. doi:10.1136/bmj.39461.643438.25](https://www.zotero.org/google-docs/?w60tu1)

[18. Maas AIR, Marmarou A, Murray GD, Teasdale SGM, Steyerberg EW. Prognosis and clinical trial design in traumatic brain injury: the IMPACT study. *J Neurotrauma*. 2007;24(2):232-238. doi:10.1089/neu.2006.0024](https://www.zotero.org/google-docs/?w60tu1)

[19. Majdan M, Lingsma HF, Nieboer D, Mauritz W, Rusnak M, Steyerberg EW. Performance of IMPACT, CRASH and Nijmegen models in predicting six month outcome of patients with severe or moderate TBI: an external validation study. *Scand J Trauma Resusc Emerg Med*. 2014;22:68. doi:10.1186/s13049-014-0068-9](https://www.zotero.org/google-docs/?w60tu1)

[20. Adil SM, Elahi C, Patel DN, et al. Deep Learning to Predict Traumatic Brain Injury Outcomes in the Low-Resource Setting. *World Neurosurg*. 2022;164:e8-e16. doi:10.1016/j.wneu.2022.02.097](https://www.zotero.org/google-docs/?w60tu1)

[21. Khalili H, Rismani M, Nematollahi MA, et al. Prognosis prediction in traumatic brain injury patients using machine learning algorithms. *Sci Rep*. 2023;13(1):960. doi:10.1038/s41598-023-28188-w](https://www.zotero.org/google-docs/?w60tu1)

[22. Plata CM de la, Hart T, Hammond FM, et al. Impact of Age on Long-term Recovery From Traumatic Brain Injury. *Arch Phys Med Rehabil*. 2008;89(5):896. doi:10.1016/j.apmr.2007.12.030](https://www.zotero.org/google-docs/?w60tu1)

[23. Hukkelhoven CWPM, Steyerberg EW, Rampen AJJ, et al. Patient age and outcome following severe traumatic brain injury: an analysis of 5600 patients. *J Neurosurg*. 2003;99(4):666-673. doi:10.3171/jns.2003.99.4.0666](https://www.zotero.org/google-docs/?w60tu1)

[24. Appiah Balaji NN, Beaulieu CL, Bogner J, Ning X. Traumatic Brain Injury Rehabilitation Outcome Prediction Using Machine Learning Methods. *Arch Rehabil Res Clin Transl*. 2023;5(4):100295. doi:10.1016/j.arrct.2023.100295](https://www.zotero.org/google-docs/?w60tu1)

**Titles and legend for figure, tables and supplementary files**

# **Figures and Tables with Legends**

## **Figures**

### **Figure 1: Confusion Matrices, ROC Curves, and Performance Metrics for Survival Prediction Models.**

**Legend –** Confusion matrices, receiver operating characteristic (ROC) curves, and performance metrics for different machine learning classifiers in predicting survival after traumatic brain injury (TBI). Classifiers include Logistic Regression, Support Vector Machines (SVM), Random Forest, and XGBoost. ROC curves illustrate model discrimination ability, while confusion matrices display classification accuracy.

### **Figure 2: Confusion Matrices, ROC Curves, and Performance Metrics for Functional Outcome Prediction Models.**

**Legend –** Comparative performance of machine learning classifiers in predicting functional outcomes post-TBI. Confusion matrices highlight classification accuracy across different Glasgow Outcome Scale (GOS) categories, while ROC curves depict classifier discrimination power. Performance metrics assess accuracy, precision, recall, specificity, and F1-score.

### **Figure 3: SHAP Summary Plots for Model Interpretability Across Survival and Functional Outcome Predictions.**

**Legend –** SHAP (Shapley Additive Explanations) summary plots depict the importance of individual features contributing to survival and functional outcome predictions in the trained machine learning models. Feature importance is ranked based on SHAP values, indicating their influence on model predictions.

## **Tables**

### **Table 1: Patient Demographics and Clinical Characteristics.**

**Legend –** Overview of demographic and clinical characteristics of the study cohort, including age distribution, gender ratio, mechanism of injury, vital parameters, neurological examination findings, and imaging results.

### **Table 2: Performance Comparison of Oversampled Machine Learning Models for Survival and Functional Outcome Prediction.**

**Legend –** Comparison of different machine learning models, including Logistic Regression, SVM, Random Forest, and XGBoost, in predicting survival and functional outcomes. Metrics include AUROC, accuracy, precision, recall, specificity, and F1-score.

# **Supplementary Figures and Tables**

## **Supplementary Figures**

### **Figure S1: Distribution of Normalized TBI Risk Scores in Survival Prediction (Oversampled Dataset).**

**Legend –** Histogram visualizing the distribution of normalized TBI risk scores for survival prediction in the oversampled dataset. The distribution provides insights into model-calibrated risk stratification.

### **Figure S2: ROC Curves for Machine Learning Models in Survival Prediction (Oversampled Dataset).**

**Legend –** ROC curves for survival prediction models trained on an oversampled dataset, illustrating classifier performance in distinguishing between survival and mortality.

### **Figure S3: Distribution of Normalized TBI Risk Scores in Functional Outcome Prediction (Oversampled Dataset).**

**Legend –** Histogram illustrating the distribution of normalized risk scores for functional outcome prediction, indicating how patients are stratified based on their recovery potential.

### **Figure S4: ROC Curves for Machine Learning Models in Functional Outcome Prediction (Oversampled Dataset).**

**Legend –** ROC curves for functional outcome prediction models trained on an oversampled dataset, demonstrating model classification efficacy across different functional recovery categories.

### **Figure S5: ROC Curves for Machine Learning Models in Survival Prediction (Undersampled Dataset).**

**Legend –** ROC curves for survival prediction models using an undersampled dataset, evaluating classifier robustness under different class balance conditions.

### **Figure S6: Performance Metrics for Functional Outcome Prediction (Undersampled Dataset).**

**Legend –** Bar chart comparing the performance of machine learning classifiers in predicting functional outcomes using an undersampled dataset. Metrics include accuracy, precision, recall, specificity, and F1-score.

### **Figure S7: Confusion Matrices for Functional Outcome Prediction (Undersampled Dataset).**

**Legend –** Confusion matrices displaying classification performance for functional outcome prediction using an undersampled dataset. The matrices show actual versus predicted classes for different models.

### **Figure S8: SHAP Summary Plot for Class 1 Predictions in Functional Outcome Model (Oversampled Dataset).**

**Legend –** SHAP summary plot highlighting the most influential features in predicting poor recovery (Class 1) in functional outcome models. Features are ranked by their impact on the prediction.

### **Figure S9: SHAP Summary Plot for Class 2 Predictions in Functional Outcome Model (Oversampled Dataset).**

**Legend –** SHAP summary plot displaying the top features influencing intermediate recovery (Class 2) predictions in functional outcome models. Feature importance is based on SHAP value contributions.

### **Figure S10: SHAP Summary Plot for Class 3 Predictions in Functional Outcome Model (Oversampled Dataset).**

**Legend –** SHAP summary plot for Class 3 predictions illustrating the most significant features contributing to good recovery (Class 3) in functional outcome models.

## **Supplementary Tables**

### **Table S1: Model Descriptions and Equations for Survival and Functional Outcome Predictions.**

**Legend –** Mathematical formulations and descriptions for the logistic regression, random forest, SVM, and XGBoost models used in survival and functional outcome prediction. Details include key predictor variables and model structures.

### **Table S2: TRIPOD + AI Checklist for Reporting Machine Learning-Based Clinical Prediction Models.**

**Legend –** TRIPOD + AI checklist outlining the study’s adherence to best practices for developing and validating machine learning-based clinical prediction models.