Archives of Academic Emergency Medicine. 2025; 13(1): e60

OPEN ACCESS

ORIGINAL RESEARCH

Machine Learning Models for Predicting Abnormal Brain CT Scan Findings in Mild Traumatic Brain Injury Patients

Amirmohammad Toloui[1], Amir Ghaffari Jolfayi[2]**, Hamed Zarei[1], Arash Ansarian[1], Amir Azimi[1,2], Seyed Mohammad Forouzannia[1], Rosita Khatamian Oskooi[3], Gholamreza Faridaalaee[4], Shayan Roshdi Dizaji[1], Seyed Ali Forouzannia[1], Seyedeh Niloufar Rafiei Alavi[1], Mohammadreza Alizadeh[1], Hadis Najafimehr[5], Saeed Safari[6,7], Alireza Baratloo[8], Mostafa Hosseini[5], Mahmoud Yousefifard[1]***

1. Physiology Research Center, Iran University of Medical Sciences, Tehran, Iran.

2. Cardiovascular Research Center, Rajaie Cardiovascular Institute, Tehran, Iran.

3. Emergency Medicine Department, Birjand University of Medical Sciences, Birjand, Iran.

4. Emergency and Trauma Care Research Center, Tabriz University of Medical Sciences, Tabriz, IR, Iran.

5. Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Poursina Ave. Enghelab Ave., Tehran, Iran.

6. Mens' Health and Reproductive Health Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

7. Emergency Department, Shohadaye Tajrish Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

8. Department of Emergency Medicine, Gold Coast University Hospital, Southport, QLD, Australia

Received: April 2025; Accepted: Ma 2025; Published online: 28 June 2025

Abstract: Introduction: Traumatic Brain Injury (TBI) is one of the leading causes of mortality and severe disability worldwide. This study aimed to develop and optimize machine learning (ML) algorithms to predict abnormal brain computed tomography (CT) scans in patients with mild TBI. **Methods:** In this retrospective analyses, the outcome was dichotomized into normal or abnormal CT scans, and univariate analyses were employed for feature selection. Then SMOTE was applied to address class imbalance. The dataset was split 80:20 for training/testing, and multiple ML algorithms were evaluated using accuracy, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). SHAP analysis was used to interpret feature contributions. **Results:** The data included 424 patients with an average age of 40.3 ± 19.1 years (76.65% male). Abnormal brain CT scan findings were more common in older males, patients with lower Glasgow Coma Scale (GCS) scores, suspected fractures, hematomas, and visible injuries above the clavicle. Among the ML models, XGBoost performed best (AUC 0.9611, accuracy 0.8937), followed by Random Forest, while Naive Bayes showed high recall but poor specificity. SHAP analysis highlighted that lower GCS scores, decreased SpO2 levels, and tachypnea were strong predictors of abnormal brain CT findings. **Conclusion:** XGBoost and Random Forest achieved high predictive accuracy, sensitivity, and specificity. GCS, SpO_2, and respiratory rate were key predictors. These models may reduce unnecessary CT scans and optimize resource use. Further multicenter validation is needed to confirm their clinical utility.

Keywords: Machine learning; Brain injuries; Glasgow coma scale

Cite this article as: Toloui A, Ghaffari Jolfayi A, Zarei H, et al. Machine Learning Models for Predicting Abnormal Brain CT Scan Findings in Mild Traumatic Brain Injury Patients. Arch Acad Emerg Med. 2025; 13(1): e60. https://doi.org/10.22037/aaemj.v13i1.2709.

1. Introduction

Traumatic Brain Injury (TBI) is one of the leading causes of mortality and severe disability worldwide, particularly in low- and middle-income countries (1). In 2019, nearly 49 million TBI patients were identified globally, with 27 million new cases recorded in the same year (2). Not only does TBI have the highest incidence rate among common neurological disorders worldwide, but this rate has also been increasing in recent years due to the rising number of motor vehicle accidents in developing countries (3).

According to the World Health Organization (WHO), mild TBI (mTBI) refers to acute, non-penetrating brain injuries caused by external mechanical forces, characterized by a Glasgow Coma Scale (GCS) score of 13-15 thirty minutes after injury, and at least one of the following: impaired mental state at the time of injury, post-injury unconsciousness lasting less than 30 minutes, amnesia lasting less than 24 hours, and tempo-

* **Corresponding Author:** Mahmoud Yousefifard; Physiology Research Center, Iran University of Medical Sciences, Tehran, Iran; Phone/Fax: +982186704771; Email: yousefifard20@gmail.com, ORCID: https://orcid.org/0000-0001-5181-4985.

** The first and second authors made equal contributions.

rary focal neurological deficits not requiring surgical intervention (4).
While the majority of mTBI patients experience no complications, some may develop clinical symptoms or complications, such as a need for neurosurgical interventions or hospitalization, temporary or permanent motor deficits, or even death (5). Nearly half of mTBI patients do not return to their pre-injury health status within six months of the injury (3). Moreover, cognitive, emotional, behavioral, and physical complications persist in 1-20% of these patients for more than a year post-injury (6-8).
Given the importance of complications and the complexity of diagnosing TBIs, brain computed tomography (CT) scans are considered the primary imaging modality and are used alongside comprehensive neurological evaluations to detect hidden injuries. Emergency brain CT scan studies have shown that around 90% of patients with a minor head injury and a GCS score of 13-15 showed no acute pathological findings (9, 10). In light of that, concerns have been raised about the high rate of unnecessary exposure to ionizing radiation in these patients (11).
Various diagnostic tools have been developed to identify high-risk TBI patients using the patients' demographic, clinical, and injury-related variables to enhance diagnostic efficiency, reduce unnecessary radiation exposure, and prevent resource wastage and overcrowding in emergency departments (EDs) (12-16). Variables such as age, headache, skull and neck fractures, impaired consciousness, amnesia, nausea, vomiting, and seizures are used in these scoring systems. Despite the high sensitivity of these clinical tools, studies have shown that their low specificity has limited their effectiveness in reducing the number of unnecessary CT scans performed in the EDs (17, 18).
In the context of TBI prognostication, recent studies have indicated that machine learning (ML) offers more reliable results than traditional statistical methods (19). Despite these promising results in short- and long-term outcome prediction, the full potential of ML in predicting brain CT scan outcomes has yet to be investigated and their clinical adoption is hindered by a lack of interpretability (20-23). The integration of explainability techniques, such as SHapley Additive exPlanations (SHAP), can elucidate how individual features contribute to model predictions, thereby enhancing clinician trust and facilitating informed decision-making (24). This study aimed to develop and optimize ML algorithms to predict abnormal brain CT scans in mTBI patients using physiological parameters and clinical evaluations.

2. Methods

2.1. Study design and Setting

In this retrospective analysis, we utilized data from a previously published study by the current authors (MY and AT), which included 424 mTBI patients. The data from these patients were prospectively collected from the EDs of three university hospitals in Iran between December 2021 and September 2022 (25).
This study was conducted in compliance with ethical guidelines and received approval from the relevant institutional review board (IRB) under approval number: IR.IUMS.REC.1401.1032. Due to its retrospective design, informed consent was waived for the analysis of anonymized data, with all patient information de-identified to safeguard privacy and confidentiality. Data were extracted from electronic medical records in adherence to regulations to ensure the security of sensitive information. The method used for this study is summarized in Flowchart 1.

2.2. Data Variables

The dataset included a comprehensive collection of the demographic, clinical, and physical examination data from mTBI cases. Demographic variables encompassed age and sex, and the clinical parameters included GCS (26), heart rate, systolic and diastolic blood pressure (SBP and DBP), oxygen saturation (SpO2), and respiratory rate. Additionally, trauma-related features such as the mechanism of injury, suspected depressed or skull base fractures, frequent nausea (occurring more than once), visible injuries above the clavicle, hematomas, and focal neurological deficits were recorded. Coagulopathy was recorded based on the patient's medical history.

2.3. Primary Outcome

The outcome variable was defined as a binary classification of normal versus abnormal brain CT scan findings. Cases with brain CT scan indicators such as skull fractures, contusions, subdural hematomas (SDH), epidural hematomas (EDH), intracerebral hemorrhage (ICH), intraventricular hemorrhage (IVH), edema, midline shift, diffuse axonal injury (DAI), subarachnoid hemorrhage (SAH), subgaleal hematoma, or pneumocephalus were classified as Abnormal CT scans. Conversely, cases without any of these abnormalities were classified as having normal CT scan findings, ensuring a clear distinction between the two outcome categories.

2.4. Univariate Analyses

Univariate analyses were performed to assess the relationship between various clinical and demographic variables and abnormal CT scan findings. Continuous variables were analyzed using the appropriate statistical tests, with the t-test applied for normally distributed data and the Mann-Whitney U test used for non-normally distributed data. Categorical variables were evaluated using the chi-square test to determine significant associations. These statistical methods were employed to identify individual predictors and potential confounders, ensuring a comprehensive preliminary analysis before proceeding to ML modeling. A p-value of less than 0.05 was considered statistically significant for all tests.

2.5. Data Preprocessing for ML

1. Feature Selection

Feature selection was performed by identifying variables that exhibited statistically significant differences between the abnormal brain CT scan group and the normal group through univariate analyses. Clinically relevant features were selected based on their association with the outcomes of interest. To enhance the model's efficiency and reduce redundancy, highly correlated variables were removed or combined, prioritizing those offering the most comprehensive representation of underlying factors. This systematic approach ensured the retention of only the most relevant features, contributing to improved predictive performance and model interpretability.

2. Normalization

Continuous variables were standardized using z-score normalization, which scales features to have a mean of 0 and a standard deviation of 1. This step ensures comparability across variables and prevents features with larger numerical ranges from dominating the learning process.

3. Data Imbalance Handling

To address data imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was employed to augment the minority class (Abnormal CT scan). SMOTE generates synthetic samples for the minority class by interpolating between a selected data point and one of its k-nearest neighbors (k-NN), based on the equation: Eq1. $xnew = x_i+\lambda(x_i^*-x_i)$

x_i selected sample in the minority class, x_i^* one of its k-NN, λ random number between 0 and 1.

2.6. Train/Test Splitting

The dataset was partitioned into training (80%) and testing (20%) subsets to evaluate model performance. The training set was used for model development, including hyperparameter tuning through a grid search approach. Key parameters, such as learning rate, maximum tree depth, and regularization, were systematically optimized based on Area Under the Receiver Operating Characteristic Curve (AUC-ROC) and F1-score using the training data.

2.7. ML Models

To identify the most effective predictive model, a variety of ML algorithms were tested, including Logistic Regression, Support Vector Machine (SVM), Random Forest Classifier, Gradient Boosting Classifier, Decision Tree Classifier, Gaussian Naive Bayes, Multi-Layer Perceptron Neural Network, and k-NN. Additionally, eXtreme Gradient Boosting (XGBoost) was incorporated due to its high classification performance and explainability, making it particularly suitable for clinical applications by providing insights into feature importance.

Each algorithm was optimized using specific configurations to enhance its performance. XGBoost, a powerful ensemble learning algorithm, was configured to disable label encoding and optimize evaluation through the log-loss metric. Random Forest employed the entropy criterion to improve split quality and enhance classification accuracy. Logistic Regression utilized L2 regularization, a solver designed for efficiency, and a high iteration limit to ensure thorough training. The Neural Network relied on multiple interconnected layers and extensive iterations to achieve effective learning. SVM applied a linear kernel for decision boundaries and provided probability estimates for its predictions. Naive Bayes, a probabilistic model, leveraged the Gaussian distribution for continuous data. k-NN classified data points based on their five nearest neighbors, and Decision Trees utilized non-parametric splitting techniques to classify without requiring assumptions about data distribution. This comprehensive evaluation of diverse models ensured a robust comparison to identify the most suitable approach for the given classification task.

2.8. Model Evaluation Metrics

The performance of the models was assessed using several key metrics, including accuracy, balanced accuracy, precision, recall, F1-score ($F1=(2\times(Precision\times Recall))/(Precision+Recall)$), and AUC-ROC. Accuracy offers a general measure of correctness but may be misleading in imbalanced datasets. F1-score is more informative in such scenarios, as it balances sensitivity and precision. AUC-ROC was selected for its robustness in evaluating the model's ability to distinguish between classes across different thresholds. Confusion matrices and ROC curves were generated to further analyze each model's predictive capability, providing a comprehensive visualization of their classification performance and ability to distinguish between classes.

2.9. Explainability and SHAP Analysis

To enhance the interpretability of our best-performing model, we employed Shapley Additive Explanations (SHAP), an approach that quantifies the contribution of each feature to individual predictions. The analysis was implemented using Python's SHAP library with the Tree Explainer, optimized for tree-based models. The SHAP Feature Importance Plot ranked features by their mean absolute SHAP values, revealing the most influential variables in predicting outcomes. Additionally, the SHAP Summary Plot illustrated how variations in feature values (e.g., high vs. low) influenced predictions, with color gradients indicating the magnitude and direction of impact.

Furthermore, we collaborated with clinicians to assess face validity, confirming that high-impact features aligned with established medical knowledge. Together, these methods not only improved transparency in our model's decision-making but also highlighted clinically actionable factors, ensuring that the results could be meaningfully applied in real-world healthcare scenarios (27).

2.10. Software and Tools

The study was conducted using Python 3.9, leveraging a range of libraries to support model implementation, evaluation, and analysis. The scikit-learn library was utilized for building and evaluating ML models, while the XGBoost library was employed for gradient boosting and explainable modeling. To address class imbalance, the imbalanced-learn library, specifically the SMOTE technique, was applied. SHAP (SHapley Additive exPlanations) was used to enhance model interpretability by analyzing feature importance. Additionally, matplotlib and seaborn were used to create visualizations, enabling a clear and effective presentation of the results. This combination of tools ensured a comprehensive and robust workflow for the study.

3. Results

3.1. Demographic Characteristics

In the present study, we enrolled a total of 424 mTBI patients who underwent brain CT scans, with an average age of 40.3 ± 19.1 years. Out of these patients, 345 had normal brain CT scans, while 79 showed abnormalities. The majority of patients, 381 (89.86%), presented with a GCS score of 15.

3.2. Univariate Analyses

The demographic and physical examination characteristics of the study population revealed significant differences between patients with normal and abnormal brain CT scans. The age of patients with abnormal CT scans was significantly higher compared to those with normal CT scans ($P = 0.03$). Similarly, the proportion of males was higher in the abnormal CT group (86.08%) compared to the normal CT group (74.49%, $P = 0.02$). GCS scores also differed notably, with a higher percentage of patients in the abnormal CT group presenting with lower GCS scores (e.g., 11.39% with GCS 13 and 20.25% with GCS 14) compared to the normal CT group ($P < 0.01$). Mechanisms of trauma varied significantly; for example, motorcycle collisions were the most frequent cause in the abnormal group (36.71%), while motor vehicle collisions were less prominent in this group compared to the normal CT group ($P = 0.002$).
In terms of clinical findings, patients with abnormal CT scans showed higher rates of suspected depressed fractures (16.46%), suspected basilar fractures (20.25%), and hematomas (16.46%) compared to those with normal CT scans (all $P < 0.01$). The percentage of patients with visible injuries above the clavicle was notably higher in the abnormal CT group (65.82%) compared to the normal group (51.01%, $P = 0.017$). Other parameters, such as heart rate, SBP, DBP, SpO2, and respiratory rate, showed no statistically significant differences between the two groups. In evaluating correlations, we removed variables with the highest degree of correlation: SBP and DBP were combined and represented as mean arterial pressure (MAP) due to their strong intercorrelation (Figure 1). MAP was significantly lower in patients with abnormal findings in brain CT scans ($P = 0.022$). These findings highlight the association between clinical indicators, demographic factors, and the likelihood of abnormal brain CT findings in trauma patients.

3.3. Evaluation Metrics

Table 2 presents a comparison of ML models based on their performance metrics, including AUC, accuracy, recall (sensitivity), specificity, precision (positive predictive value, PPV), negative predictive value (NPV), and F1-score. Among the models, XGBoost demonstrates the highest overall performance based on its AUC of 0.9611. This model showed an accuracy of 0.8937 and an F1-score of 0.8878. It also balances recall (0.8878) and specificity (0.8991), effectively distinguishing between positive and negative cases. Random Forest follows closely with an AUC of 0.9587 and comparable recall and specificity values, but a slightly higher precision and F1-score compared to XGBoost. These results suggest that ensemble methods like XGBoost and Random Forest are the most robust for this dataset (Table 2 and Figure 2).
Other models, such as the Neural Network and k-NN, show moderate performance, with Neural Network achieving an AUC of 0.8953 and an F1-score of 0.8195. However, these models lag behind XGBoost and Random Forest in terms of accuracy and specificity. Naive Bayes, despite having the highest recall (0.9898), exhibits poor specificity (0.0917) and precision (0.4949), indicating a high rate of false positives. Gradient Boosting, Logistic Regression, and SVM exhibit the lowest overall performance, with Gradient Boosting showing particularly low specificity (0.4404) and precision (0.5734). Overall, the results underscore the superiority of ensemble techniques like XGBoost and Random Forest for achieving high predictive performance (Table 2 and Figure 2).

3.4. SHAP Analysis

In interpreting the ML model, SHAP analysis revealed that a GCS score of less than 15 significantly increases the likelihood of abnormal CT scan findings. Additionally, lower SpO2 levels and tachypnea are strongly associated with abnormal results. The analysis also indicated that the risk of abnormal findings increases with age.
Bradycardia and low blood pressure were identified as factors associated with a higher probability of abnormal findings. Furthermore, visible injuries above the clavicle and hematomas were also found to be strongly associated with abnormal CT scan results.
In our model, some features, like nausea and coagulopathy, contributed minimally to the model's prediction performance.

4. Discussion

Here, we built and compared multiple ML models to predict abnormal brain CT findings in mTBI patients using easily available clinical and demographic features. Our results showed that among 424 enrolled patients, 18.63% had abnormal CT findings, with significant associations observed between abnormal scans and factors such as older age, male sex, lower GCS scores, visible injuries above the clavicle, suspected skull fractures, and hematomas. The XGBoost and Random Forest models demonstrated the highest predictive performance, outperforming other ML algorithms, including Neural Networks and k-NN. SHAP analysis further revealed that SpO, GCS, and respiratory rate were the most influential predictors of CT abnormalities. These findings highlight the potential of ML-based models to improve clinical decision-making by identifying high-risk mTBI patients while minimizing unnecessary CT scans, particularly in resource-limited settings.

Our findings align with the study by Terabe et al., which also explored the applicability of ML models in predicting CT abnormalities in mTBI patients (20). Their study, conducted in a Brazilian ED, evaluated multiple ML models, with the XGBoost algorithm demonstrating the highest sensitivity. Similar to our study, Terabe et al. emphasized the importance of reducing unnecessary CT scans in resource-limited settings, noting that only 4.62% of patients who underwent imaging had positive findings. The feature selection process in their study prioritized clinical variables commonly collected in Brazilian healthcare services, mirroring our approach of using easily accessible clinical and demographic parameters. However, while both studies identified ML models as valuable tools for CT scan prioritization, key differences exist. Our study identified SpO, GCS, and respiratory rate as the most influential features, whereas Terabe et al. emphasized trauma mechanism, GCS, amnesia, and dizziness. Additionally, while we employed SMOTE to handle class imbalance, their study used the Random Over-Sampling Examples (ROSE) method. These differences highlight the need for validation across diverse populations and healthcare settings to ensure model generalizability and effectiveness.

Our findings further corroborate prior research demonstrating the effectiveness of ML models, particularly XGBoost, in predicting outcomes for TBI patients. For instance, a study by Lee et al. developed an XGBoost model to predict the progression of intracerebral hemorrhage in patients with mild to moderate TBI, achieving an AUC of 0.913, indicating high predictive accuracy (28). Similarly, research by Cao et al. employed an XGBoost-powered Cox regression model to predict in-hospital mortality among severe TBI patients, demonstrating outstanding predictive ability (29). These studies, along with our own, underscore the potential of ML approaches to enhance diagnostic and prognostic accuracy in TBI management. Our study contributes to this growing body of evidence by demonstrating that the XGBoost model effectively predicts abnormal brain CT findings in mTBI patients, thereby supporting the integration of ML-based tools into clinical practice to improve decision-making and patient outcomes.

Despite promising results, further multicenter validation and prospective studies are needed to assess the real-world applicability of these models and ensure their reliability across diverse populations and geographic regions. Future research should also explore alternative techniques for handling class imbalances and integrating ML models into clinical workflows to enhance decision-support systems. Additionally, Additionally, developing real-world integration strategies, such as embedding models into electronic health records or decision-support platforms, is crucial for successful clinical adoption. The implementation of such predictive tools could optimize resource utilization, reduce radiation exposure, and improve patient outcomes in emergency care settings.

5. Limitations

Although we addressed data imbalance by employing SMOTE, this method has inherent limitations. SMOTE generates synthetic data points by interpolating between existing minority-class samples rather than introducing entirely new observations. As a result, it cannot account for rare but clinically significant cases, potentially leading to overfitting and reduced generalizability when applied to real-world, highly imbalanced datasets. Additionally, SMOTE does not consider the underlying pathophysiology of TBI progression, which may limit the model's ability to differentiate between subtle but clinically important variations. SMOTE's adaptability to different algorithms and straightforward implementation contributed to its success (30). However, SMOTE also posed certain drawbacks, such as higher computational demands, possible noise amplification, and diminished efficacy in high-dimensional data. Furthermore, its influence on model performance can differ depending on the dataset. Studies indicate that SMOTE should be applied cautiously, emphasizing the importance of a thorough assessment of its real-world applicability (31-33).

6. Conclusions

Our results demonstrated that XGBoost and Random Forest outperformed other ML algorithms, achieving high predictive accuracy, sensitivity, and specificity. SHAP analysis identified GCS, SpO, and respiratory rate as the most influential predictors of brain CT abnormalities in mTBI. These findings highlight the potential of ML-based models to improve clinical decision-making, particularly in emergency settings, by identifying high-risk mTBI patients while minimizing unnecessary CT scans.

7. Declarations

7.1. Acknowledgments

None.

7.2. Authors' Contributions

Study design: MY, MH, SS, GFA, AB

Data gathering: AT, SMF, RKO, GFA, SRD, SAF, SNRA, MA, HN, AB, MY

Analysis: AT, AGJ, HZ, Amir Azimi, Arash Ansarian, MY
Drafting: AT, AGJ, HZ, Amir Azimi, Arash Ansarian
Critically revised: all authors
All authors read and approved the final version of the manuscript.

7.3. Conflict of Interest

The authors declare that they have no conflict of interest.

7.4. Data Availability

All data included in this study are available from the corresponding author upon email request.

7.5. Using Artificial Intelligence Chatbots

In this study, we utilized ChatGPT to assist in refining and enhancing the clarity of the written content. The chatbot was not used to generate original scientific ideas or interpretations, but rather to improve the structure, grammar, and readability of text already drafted by the authors. All substantive contributions and analyses were performed independently by the research team.

References

1. Maas AI, Menon DK, Adelson PD, Andelic N, Bell MJ, Belli A, et al. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. The Lancet Neurology. 2017;16(12):987-1048.
2. 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.
3. Maas AI, Menon DK, Manley GT, Abrams M, Åkerlund C, Andelic N, et al. Traumatic brain injury: progress and challenges in prevention, clinical care, and research. The Lancet Neurology. 2022;21(11):1004-60.
4. Holm L, Cassidy JD, Carroll LJ, Borg J. Summary of the WHO collaborating centre for neurotrauma task force on mild traumatic brain injury. Journal of rehabilitation medicine. 2005;37(3):137-41.
5. Vos PE, Alekseenko Y, Battistin L, Ehler E, Gerstenbrand F, Muresanu DF, et al. Mild traumatic brain injury. European journal of neurology. 2012;19(2):191-8.
6. Obiano KS, Singh R, Dawson J. Post-concussion symptoms 1-year after traumatic brain injury: using the Rivermead Post-concussion Questionnaire to identify predictors of severity. Brain injury. 2022;36(12-14):1323-30.
7. Schneider AL, Huie JR, Boscardin WJ, Nelson L, Barber JK, Yaffe K, et al. Cognitive outcome 1 year after mild traumatic brain injury: results from the TRACK-TBI study. Neurology. 2022;98(12):e1248-e61.
8. Theadom A, Parag V, Dowell T, McPherson K, Starkey N, Barker-Collo S, et al. Persistent problems 1 year after mild traumatic brain injury: a longitudinal population study in New Zealand. British Journal of General Practice. 2016;66(642):e16-e23.
9. Zadegan SA, Rahimi-Movaghar V. Indications for Brain Computed Tomography Scan After Mild Traumatic Brain Injury. Traumatic Brain Injury: IntechOpen; 2014.
10. Smits M, Dippel DW, de Haan GG, Dekker HM, Vos PE, Kool DR, et al. External validation of the Canadian CT Head Rule and the New Orleans Criteria for CT scanning in patients with minor head injury. Jama. 2005;294(12):1519-25.
11. Jones LA, Morley EJ, Grant WD, Wojcik SM, Paolo WF. Adherence to head computed tomography guidelines for mild traumatic brain injury. Western Journal of Emergency Medicine. 2014;15(4):459.
12. Stiell IG, Wells GA, Vandemheen K, Clement C, Lesiuk H, Laupacis A, et al. The Canadian CT Head Rule for patients with minor head injury. The Lancet. 2001;357(9266):1391-6.
13. Haydel MJ, Preston CA, Mills TJ, Luber S, Blaudeau E, DeBlieux PM. Indications for computed tomography in patients with minor head injury. New England Journal of Medicine. 2000;343(2):100-5.
14. guideline NG232 N. Head Injury: assessment and early management. 2023.
15. Mower WR, Hoffman JR, Herbert M, Wolfson AB, Pollack Jr CV, Zucker MI, et al. Developing a decision instrument to guide computed tomographic imaging of blunt head injury patients. Journal of Trauma and Acute Care Surgery. 2005;59(4):954-9.
16. Servadei F, Teasdale G, Merry G. Defining acute mild head injury in adults: a proposal based on prognostic factors, diagnosis, and management. Journal of neurotrauma. 2001;18(7):657-64.
17. Sakkas A, Weiß C, Wilde F, Ebeling M, Scheurer M, Thiele OC, et al. Justification of Indication for Cranial CT Imaging after Mild Traumatic Brain Injury According to the Current National Guidelines. Diagnostics. 2023;13(11):1826.
18. Żyluk A. Indications for CT scanning in minor head injuries: A review. Neurologia i neurochirurgia polska. 2015;49(1):52-7.
19. 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.
20. Terabe ML, Massago M, Iora PH, Hernandes Rocha TA, de Souza JVP, Huo L, et al. Applicability of machine learning technique in the screening of patients with mild traumatic brain injury. Plos one. 2023;18(8):e0290721.
21. Klement W, Wilk S, Michalowski W, Farion KJ, Osmond MH, Verter V. Predicting the need for CT imaging in children with minor head injury using an ensemble of Naive Bayes classifiers. Artificial intelligence in medicine. 2012;54(3):163-70.
22. Miyagawa T, Saga M, Sasaki M, Shimizu M, Yamaura

A. Statistical and machine learning approaches to predict the necessity for computed tomography in children with mild traumatic brain injury. Plos one. 2023;18(1):e0278562.

23. Molaei S, Korley FK, Soroushmehr SR, Falk H, Sair H, Ward K, et al., editors. A machine learning based approach for identifying traumatic brain injury patients for whom a head CT scan can be avoided. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2016: IEEE.
24. Pan Y, Wei M, Jin M, Liang Y, Yi T, Tu J, et al. An interpretable machine learning model based on optimal feature selection for identifying CT abnormalities in patients with mild traumatic brain injury. eClinicalMedicine. 2025;82.
25. Forouzannia SM, Najafimehr H, Oskooi RK, Faridaalaee G, Dizaji SR, Toloui A, et al. Clinical decision rules in predicting computed tomography scan findings and need for neurosurgical intervention in mild traumatic brain injury: a prospective observational study. European Journal of Trauma and Emergency Surgery. 2024;50(4):1199-207.
26. Jain S, Iverson LM. Glasgow Coma Scale. StatPearls. Treasure Island (FL): StatPearls Publishing Copyright © 2025, StatPearls Publishing LLC.; 2025.
27. Marcílio WE, Eler DM, editors. From explanations to feature selection: assessing SHAP values as feature selection mechanism. 2020 33rd SIBGRAPI conference on Graphics, Patterns and Images (SIBGRAPI); 2020: Ieee.
28. Lee HS, Kim JH, Son J, Park H, Choi J. Machine learning models for predicting early hemorrhage progression in traumatic brain injury. Scientific reports. 2024;14(1):11690.
29. 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 [Internet]. 2023; 13(9).
30. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research. 2002;16:321-57.
31. Fernández A, Garcia S, Herrera F, Chawla NV. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. Journal of artificial intelligence research. 2018;61:863-905.
32. Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. BMC bioinformatics. 2013;14:1-16.
33. Pradipta GA, Wardoyo R, Musdholifah A, Sanjaya INH, Ismail M, editors. SMOTE for handling imbalanced data problem: A review. 2021 sixth international conference on informatics and computing (ICIC); 2021: IEEE.

Table 1: Comparing the baseline characteristics and clinical findings between cases with and without abnormal bran computed tomography findings

Variables	Total N=424	Normal N=345	Abnormal N=79	P value
Age (years)				
Median (IQR)	35 (25-52)	34 (25-50)	41 (25-65)	0.03
Sex				
Male	325 (76.65)	257 (74.49)	68 (86.08)	0.028
Glasgow Coma Scale				
13	16 (3.77)	7 (2.03)	9 (11.39)	< 0.01
14	27 (6.37)	11 (3.19)	16 (20.25)	
15	381 (89.86)	327 (94.78)	54 (68.35)	
Mechanism of trauma				
Fall from less than 1 meter	84 (19.81)	67 (19.42)	17 (21.52)	0.002
Direct blunt trauma	57 (13.44)	48 (13.91)	9 (11.39)	
Bicycle collision	4 (0.94)	4 (1.16)	0 (0)	
Fall from more than 1 meter	33 (7.78)	27 (7.83)	6 (7.59)	
Pedestrian struck	35 (8.25)	25 (7.25)	10 (12.66)	
Motorcycle collision	96 (22.64)	67 (19.42)	29 (36.71)	
Motor vehicle collision	59 (13.92)	54 (15.65)	5 (6.33)	
Others	56 (13.21)	53 (15.36)	3 (3.8)	
Coagulopathy				
Number (%)	15 (3.5)	13 (3.77)	2 (2.53)	0.747
Vital signs				
Heart Rate (/minute)	82 (80-88)	82 (80-88)	84 (80-89)	0.438
SBP (mmHg)	120 (110-125)	120 (112-125)	115 (110-125)	0.087
DBP (mmHg)	75 (70-80)	77 (70-80)	75 (70-80)	0.053
MAP (mmHg)	101 (96.45-106)	101.2 (97-106)	98 (94-105)	0.022
SpO2 (%)	96 (95-97)	96 (96-97)	96 (95-97)	0.519
Respiratory Rate (/minute)	16 (14-18)	16 (14-18)	17 (16-18)	0.094
Skull fracture				
Suspected Depressed	37 (8.73)	24 (6.96)	13 (16.46)	0.007
Suspected Basilar	26 (6.13)	10 (2.90)	16 (20.25)	< 0.001
Clinical sign and symptom				
Frequent Nausea1	14 (3.3)	11 (3.19)	3 (3.8)	0.785
Visible Injuries Above the Clavicle	228 (53.77)	176 (51.01)	52 (65.82)	0.017
Focal Neurologic Deficit	9 (2.12)	8 (2.32)	1 (1.27)	0.558
Hematoma	27 (6.37)	14 (4.06)	13 (16.46)	< 0.001

Values are presented as median (IQR) or N (%). 1Frequent nausea is considered as nausea more than 1 time.
CT: computed tomography, SBP: systolic blood pressure, DBP: diastolic blood pressure, SpO2: oxygen saturation,
RR: respiratory rate, FND: focal neurological deficit.

Table 2: Performance metrics comparison of machine learning models

Model	AUC	Accuracy	Recall	Specificity	Precision (PPV)	NPV	F1-score
eXtreme Gradient Boosting	0.9611	0.8937	0.8878	0.8991	0.8878	0.8991	0.8878
Random Forest	0.9587	0.9034	0.9082	0.8991	0.8900	0.9159	0.8990
Multi-Layer Perceptron Neural Network	0.8953	0.8213	0.8571	0.7890	0.7850	0.8600	0.8195
k-Nearest Neighbors	0.8697	0.7729	0.8571	0.6972	0.7179	0.8444	0.7814
Decision Tree	0.8486	0.8454	0.9082	0.7890	0.7946	0.9053	0.8476
Naive Bayes	0.7820	0.5169	0.9898	0.0917	0.4949	0.9091	0.6599
Gradient Boosting	0.7764	0.6280	0.8367	0.4404	0.5734	0.7500	0.6805
Logistic Regression	0.7650	0.6957	0.7041	0.6881	0.6699	0.7212	0.6866
Support Vector Machine	0.7622	0.7295	0.7449	0.7156	0.7019	0.7573	0.7228

AUC: area under the curve; PPV: positive predictive value; NPV: negative predictive value.

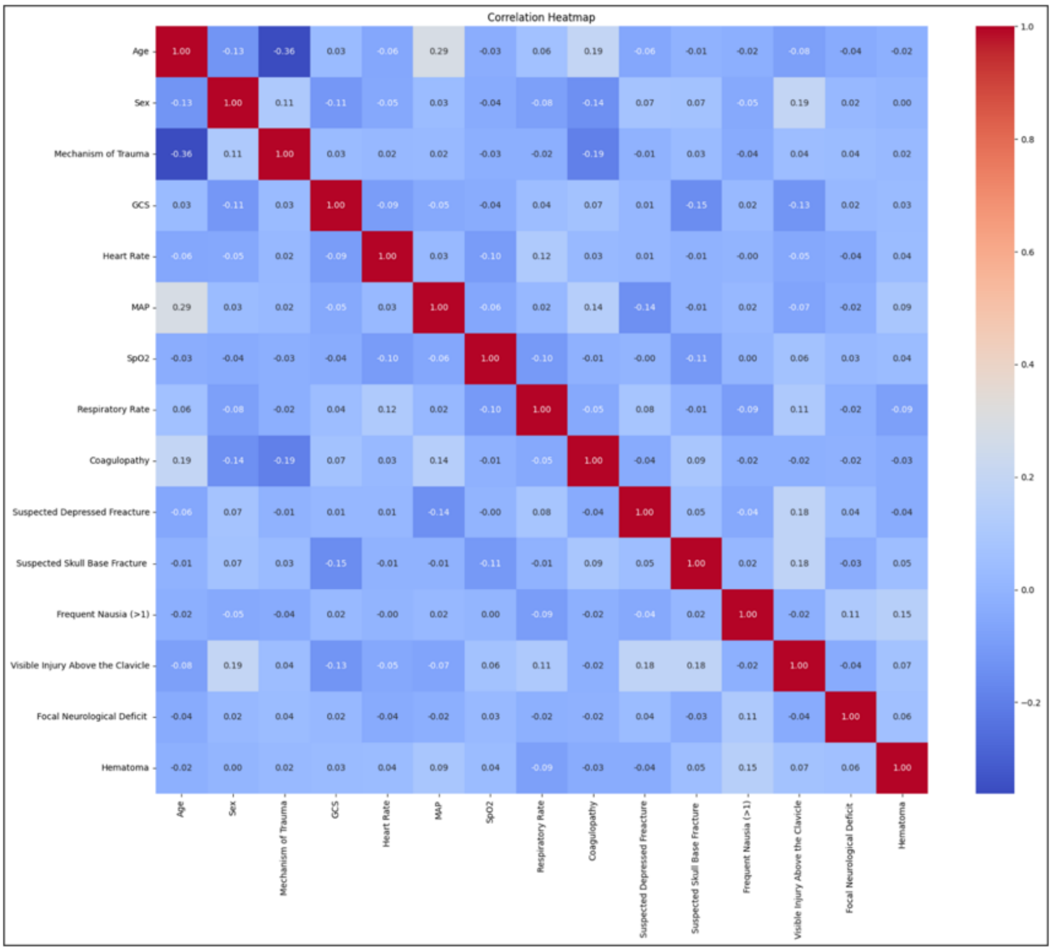


Figure 1: Correlation of clinical and demographic variables.

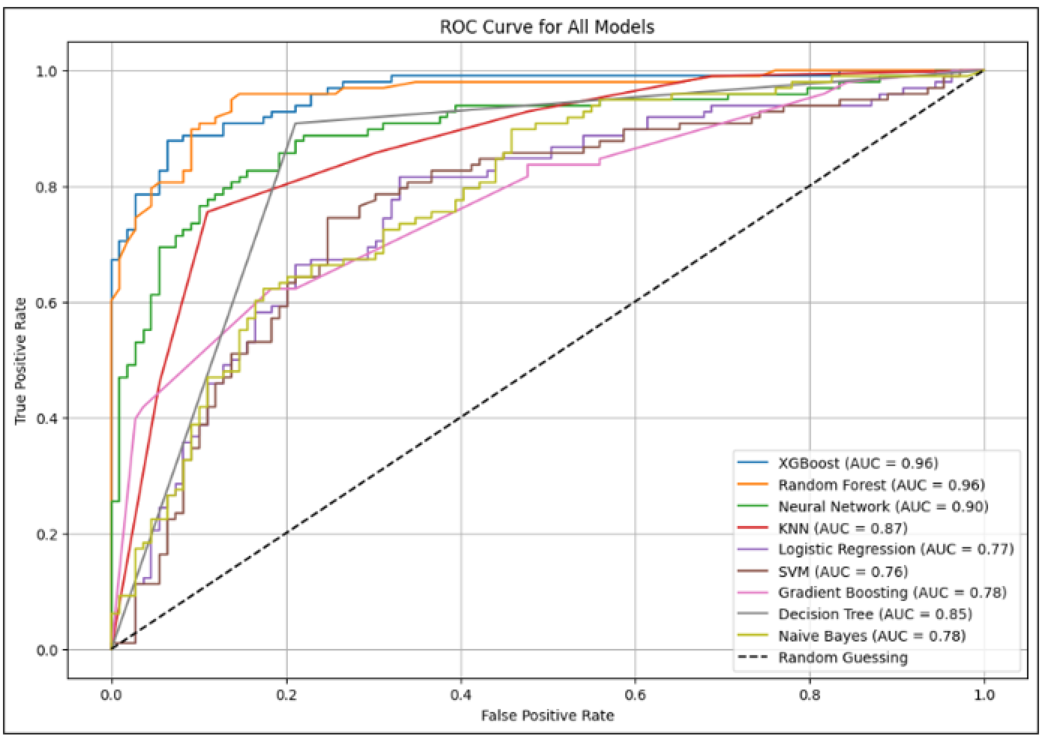


Figure 2: Area Under the Curve (AUC) for the Performance of Various Machine Learning Models.

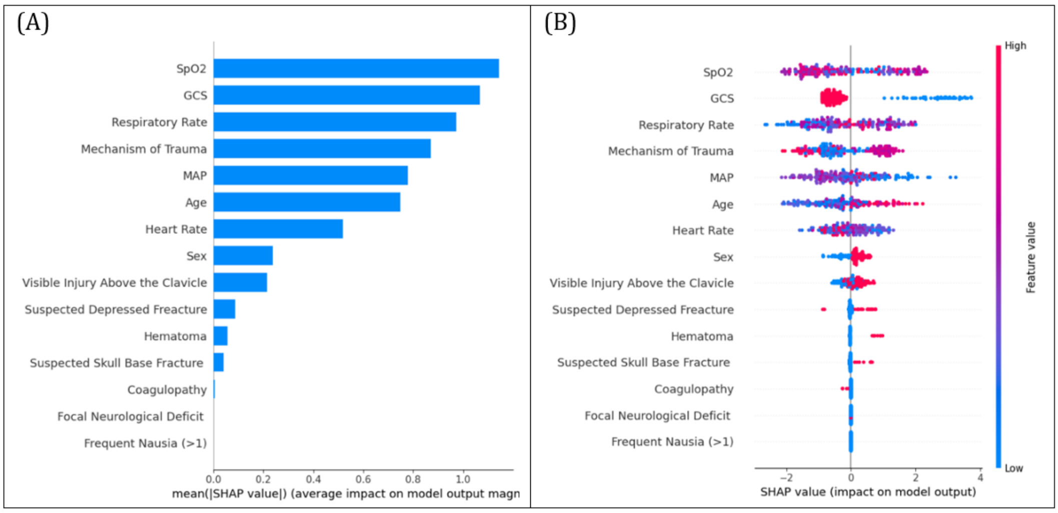
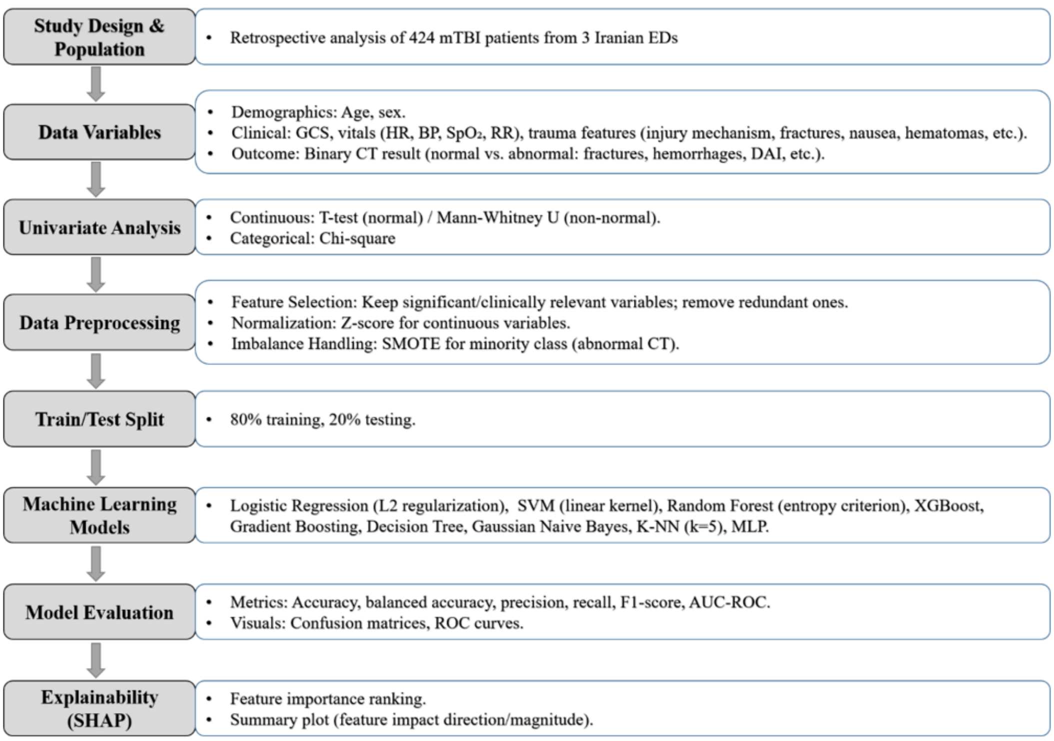


Figure 3: SHAP analysis of findings (A) The importance of each parameter in the final prediction of the model. (B) The influence of parameter variations on the model's predicted outcomes.



Flow chart 1: Flowchart of the machine learning pipeline for predicting abnormal brain computed tomography (CT) findings in mild traumatic brain injury (mTBI) patients.