**Archives of Academic Emergency Medicine. 2025; 13(1): e41**

ORIGINAL RESEARCH

# Predicting the Presence of Traumatic Chest Injuries Using Machine Learning Algorithm

**Mohammadhossein Vazirizadeh-mahabadii[1,2], Amir Ghaffari Jolfayi[3], Mostafa Hosseini[4], Mobina Yarahmadi[2], Hamed Zarei[2], Mohsen Masoodi[1], Arash Sarveazad[1,5]*, Mahmoud Yousefifard[2]†**

1. Colorectal Research Center, Iran University of Medical Sciences, Tehran, Iran
2. Physiology Research Center, Iran University of Medical Sciences, Tehran, Iran
3. Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran
4. Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran
5. Nursing Care Research Center, Iran University of Medical Sciences, Tehran, Iran

*Received: January 2025; Accepted: February 2025; Published online: 17 March 2025*

**Abstract: Introduction:** Various tools have been developed to determine the priority of radiography in trauma patients. This study aimed to investigate the role of machine learning models in predicting chest injuries following multiple trauma. **Methods:** We used the database of a comprehensive cross-sectional survey conducted in 2015. Eight machine learning models were developed using demographic characteristics, physical exam findings, and radiologic results of 2860 patients. **Results:** Area under the receiver operating characteristic curve (AUC) was greater than 0.96 in Random Forest, Gradient Boosting, XGBoost, Decision Tree, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), and Neural Network models. The random forest model, XGBoost and Gradient Boosting had the highest accuracy (0.99). Sensitivity was also highest in the Gradient Boosting, XGBoost and KNN models (0.99). The specificity of all of the models in predicting chest radiography outcomes of multiple trauma patients was higher than 0.97, except for logistic regression and SVM (0.912 and 0.885 respectively). **Conclusion:** Our study highlights the strong potential of machine learning models, especially Random Forest and Gradient Boosting, in predicting chest trauma outcomes with high accuracy and sensitivity.

**Keywords:** Multiple trauma; Thoracic injuries; Machine learning algorithms; Radiography, thoracic; Detection algorithms; Lung injury

Cite this article as: Vazirizadeh-mahabadi M, Ghaffari Jolfayi A, Hosseini M, et al. Predicting the Presence of Traumatic Chest Injuries Using Machine Learning Algorithm. Arch Acad Emerg Med. 2025; 13(1): e41. https://doi.org/10.22037/aaemj.v13i1.2512.

## 1. Introduction

Injury-related trauma is the primary cause of death, hospitalization, and disability globally (1). It accounts for 10% of deaths worldwide and is responsible for 90% of fatalities in Low and Middle-Income Countries (2, 3). Multiple trauma describes injuries that affect more than two anatomical regions or organs, with at least one injury being potentially life-threatening. In various studies, multiple trauma is typically defined by an Injury Severity Score (ISS) above 16 and the involvement of significant injuries (with an Abbreviated Injury Scale score greater than 3) in at least two body areas (4).

Thoracic trauma is a critical concern, varying widely in severity. Severe thoracic trauma often leads to adverse outcomes such as tracheostomy, mechanical ventilation, and oxygen therapy (5).

Radiographic assessments are essential in evaluating thoracic trauma patients. According to the Advanced Trauma Life Support (ATLS) protocol, all trauma patients should undergo chest radiography (6). However, studies show that pathologic findings in trauma patients are present in only 10% of chest X-rays(7) . Unnecessary radiographs can increased healthcare costs and delayed emergency care. As a result, recent studies have explored tools to prioritize radiography in trauma patients, including the Thoracic Injury Rule out Criteria (TIRC) (8) and National Emergency X-ray Utilization Study (NEXUS) Chest(9) . Despite their high sensitivity (approximately 98%), these tools have limited specificity (around 60%), leading to continued concerns about unnecessary radiation exposure and healthcare costs. Therefore, a more accurate tool to guide clinical decisions on radiography is needed (10).

---

*Corresponding Author: Arash Sarveazad; Colorectal Research Center, Rasoul-e-Akram Hospital, Nyaiesh Ave., Tehran, Iran Tel/Fax: +982166554790, Email: Arashsarveazad@gmail.com, ORCID: https://orcid.org/0000-0001-9273-1940.

†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.

In recent years, machine learning algorithms have been developed to assist physicians in making more accurate diagnoses and predicting injury severity. Natural language processing (NLP) has enabled the interpretation of information by machine learning algorithms to predict outcomes based on reference standards (11). The use of machine learning algorithms for diagnosing traumatic chest injuries has shown promising results. For example, A deep learning-based computer-aided diagnostic (DL-CAD) system was developed to improve diagnosis of acute rib fractures in chest trauma patients (12). The implementation of DL-CAD significantly improved the diagnostic sensitivity, specificity, and positive predictive value for both interns and attending radiologists, leading to increased diagnostic confidence and reduced image reading time. Additionally, Gipson et al. assessed a deep convolutional neural network, Annalise.ai, for detecting traumatic injuries on supine chest radiographs (13). The artificial intelligence (AI) program performed comparably to radiologists, particularly in identifying pneumothorax and segmental collapse, highlighting the potential of machine learning algorithms to enhance the efficiency of clinical procedures (14).

Despite advancements in machine learning techniques, few large-scale studies have investigated their role in trauma care, and some existing studies suffer from small sample sizes and inconsistent results (13, 15, 16). Furthermore, while numerous machine learning protocols exist, only a few have been thoroughly examined. This study aimed to explore the potential role of machine learning models in predicting the outcomes for multiple trauma patients regarding intrathoracic injuries.

## 2. Methods

### 2.1. Data source and study population

The present study used data from a comprehensive cross-sectional survey conducted in 2015 (8). In this survey, clinical presentations and chest radiograms of 2905 multiple trauma patients aged over 15 years referring to 6 educational hospitals in Iran were evaluated. The data mentioned above were gathered by two emergency medicine specialists.

### 2.2. Inclusion and exclusion criteria

In the mentioned study, all multiple trauma patients admitted to the hospitals during the study period were included. clinical presentations and chest radiograms of multiple trauma patients referring to 6 educational hospitals in Iran were evaluated.

The exclusion criteria were age under 15 years, penetrating trauma, and patients who were unwilling to participate.

### 2.3. Data structure

Collected data consisted of demographic characteristics of the patients (age, gender, mechanism of trauma) and physical examination findings. Physical examination included vital signs (systolic and diastolic blood pressure, respiratory rate, pulse rate), peripheral capillary oxygen saturation, mental status (Glasgow coma scale (GCS)<15), presence of dyspnea, painful distracting injuries, thoracic pain and tenderness, upper abdominal tenderness, crepitation in palpitation, chest deformity, chest abrasion or wound, diminished pulmonary sounds, and subcutaneous emphysema.

### 2.4. Dataset and preprocessing

The dataset used in this study comprised clinical and diagnostic data sourced from an Excel file. Initial preprocessing involved identifying and removing columns related to imaging findings including hemo-mediastinum, mediastinal widening, lung contusion, and diaphragm injury, to prevent bias.

### 2.5. Data labeling

All patients underwent lateral and posterior-anterior X-Rays. Chest CT-Scan was performed for 3.2% of the patients by the demand of other specialists. Final outcome in these patients were based on CT-Scan findings. Radiograms were interpreted by another emergency medicine specialist blinded to the study and by a radiologist in 3% of cases. The radiogram outcomes included pneumothorax, hemothorax, pulmonary contusion, diaphragm injuries, hemo-mediastinum, mediastinal widening, subcutaneous emphysema, and fractures of the ribs, sternum, clavicles, and scapula. The presence of any of the mentioned X-ray findings was considered a positive test result in this study.

The final diagnosis was made based on a positive finding from chest X-ray radiography or CT scan in ambiguous cases. The inter-rater agreement was reported as high as 98%.

### 2.6. Data imbalance handling

To address potential class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. SMOTE generates synthetic examples of the minority class to achieve a balanced distribution in the training dataset.

### 2.7. Data normalization

Subsequently, all remaining features were normalized using the StandardScaler function from the sklearn.preprocessing module. This normalization step ensures that all numerical features are standardized to a common scale, mitigating the impact of differing magnitudes on model training.

### 2.8. Model development, training, and testing

A diverse set of machine learning algorithms was utilized for predictive modeling. These included Random Forest, Gradient Boosting, XGBoost, Decision Tree, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), and Neural Network models. Each model was trained on the preprocessed and normalized dataset to predict the target variable.

### 2.9. Evaluation model metrics

Model evaluation focused on key performance metrics such as accuracy, balanced accuracy, precision, recall, F1 score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Diagnostic values including accuracy, sensitivity, specificity, and area under the ROC curve were calculated with a 95% confidence interval for each machine learning model, separately. These metrics collectively assess the model's predictive capability and generalization to unseen data.

### 2.10. Feature extraction and importance

Feature importance analysis was conducted using both Random Forest and XGBoost models. By averaging the importance of features derived from these models, the study identified the top features that contributed significantly to prediction accuracy. This step facilitated insights into the underlying relationships between input features and the target variable.

### 2.11. Ablation study

An ablation study was conducted to investigate the impact of individual features on the performance of the predictive model. The study systematically removed each feature from the dataset and evaluated the resulting changes in model performance metrics. The dataset used clinical and diagnostic information sourced from an Excel file, initially preprocessed by removing non-contributory features. Following preprocessing, the dataset underwent normalization using the Standard Scaler function to standardize numerical features and was subsequently balanced using the Synthetic Minority Over-sampling Technique (SMOTE). A base Random Forest Classifier model was trained on the original dataset to establish baseline performance metrics, including accuracy, balanced accuracy, precision, recall, F1 score, and AUC-ROC. Subsequently, for each feature in the dataset, an ablation study was conducted by excluding the feature, renormalizing the dataset, and training a new Random Forest Classifier model.

### 2.12. Statistical analysis

All data processing, model training, and performance evaluation were performed using Python programming language (version 3.15). Libraries such as pandas, numpy, scikit-learn and matplotlib were instrumental in implementing various analytical tasks. The computational environment utilized for this study was a standard personal computer equipped with an Intel Core i7 processor and 16 GB of RAM.

### 2.13. Visual representation

The findings of the ablation study were visually presented through a series of plots, each showing the performance difference for a specific metric when individual features were removed from the dataset. This visual analysis was held by R language (Version 4.3.3) in R studio.

### 2.14. Ethical considerations

The study was meticulously conducted following the approval of the Ethics Committee of Iran University of Medical Sciences (Code: IR.IUMS.REC.1402.877). Stringent measures were implemented to safeguard patient confidentiality and privacy throughout the research process.

## 3. Results

### 3.1. Patient characteristics

In 2015, 2,905 multiple trauma patients were referred to six educational hospitals in Iran. Of these, 45 patients were excluded due to missing important values and the presence of outlier features. The mean age was 33.52± 15.45 years and 73.1% of the patients were male. The most common mechanism of trauma was pedestrian accidents (38.3%) followed by car accident (30.8%), motorcycle accident (12.5%), over ground falls (9.4%), and fall from over 3-meter height (4.6%). 10.3% of the patients presented with altered mental status (GCS < 15). In the physical examination, 5.8% exhibited decreased pulmonary sounds. Pathologic radiologic findings were observed in 18% of the patients. Table 1 provides a comprehensive review of the clinical symptoms and pathological findings used in data processing and analysis.

### 3.2. Machine learning models

Table 2 compares the diagnostic accuracy of various machine learning models. The Random Forest and Gradient Boosting models demonstrated the highest performance, achieving near-perfect metrics in terms of AUC, accuracy, and balanced accuracy. The Decision Tree and SVM models showed weak performance compared to others. Logistic Regression and SVM had the lowest accuracy among all models, indicating their relative ineffectiveness in this study. Overall, the advanced ensemble models outperformed the simpler models, showcasing their superior diagnostic capabilities. Figure 1 demonstrated their ROC curve compared to each other.

### 3.3. Feature importance analysis

Feature importance analysis using both Random Forest and XGBoost models identified the top ten most significant features. Chest pain emerged as the most significant feature, followed by dyspnea, chest tenderness, respiratory rate, and Oxygen saturation among the top features (Table 3 and Figure 2).

### 3.4. Ablation study

The ablation study revealed the impact of removing specific features on the model's performance metrics. Removing chest pain and GCS had the most significant negative impact on accuracy, balanced accuracy, precision, recall, and F1 score, highlighting their importance in model performance. Conversely, features such as systolic and diastolic blood pressure, chest sounds and crepitation had negligible impact when removed, suggesting they are less crucial to the model.

Most features had minimal impact on the AUC-ROC, but the removal of chest pain and GCS caused the largest drop, underscoring their critical predictive value (Supplementary table 1, Figure 3 and Figure 4).

## 4. Discussion

This study explored eight machine learning models to predict outcomes for multiple trauma patients, with each demonstrating robust predictive capabilities. Among these, the Random Forest and Gradient Boosting models exhibited the best performance, achieving near-perfect metrics in AUC, accuracy, and balanced accuracy. In contrast, the logistic regression and SVM models underperformed, displaying the lowest accuracy, thereby indicating their relative ineffectiveness in this study. The high sensitivity of the top-performing models highlights their potential as valuable tools in predicting imaging outcomes in nearly all chest trauma cases. Moreover, the satisfactory specificity of these models suggests their utility in prioritizing patients who require imaging.

Our findings on efficacy of machine learning models align with those of Kondori et al. (17), who analyzed 1,000 samples in 2021 using various models, including SVM, logistic regression, Naïve Bayes, decision tree, multilayer perceptron, random forest, and KNN. In their study, the decision tree model achieved the highest accuracy (87%). By comparison, our study observed the highest accuracy in the random forest model (99%), followed closely by the decision tree (98%). Notably, our dataset was more extensive, gathered from six hospitals and nearly three times larger than the dataset in Kondori's study, which included patients from only two hospitals.

Machine learning is a novel field in traumatic imaging, though its application in chest trauma remains underexplored. While some studies have developed machine learning algorithms to predict outcomes such as ICU admissions and hospital length of stay, few have focused specifically on chest trauma. For instance, Staziaki et al. (18) analyzed 840 adult patients admitted to a Level 1 trauma center, finding that SVM and artificial neural network models performed well in predicting ICU admission and extended hospital stays. Another study examined the effectiveness of the XGBoost model in predicting early death after severe trauma, with high accuracy, sensitivity, and satisfactory PPV, specificity, and NPV (19).

In our study, the exclusion of chest pain and GCS as features significantly reduced the models' accuracy, balanced accuracy, precision, recall, and F1 score, underscoring their critical role in model performance. Conversely, the removal of features such as systolic and diastolic blood pressure, chest sounds, and crepitation had minimal impact, suggesting these are less crucial to the models. While most features had little effect on AUC-ROC, chest pain and GCS were notably the most impactful, reaffirming its importance as a predictive variable. Previous research, such as that by Shukla et al., has similarly identified chest pain as the most common symptom in blunt chest trauma. (20).

This study has several strengths, including the large and diverse dataset extracted from six hospitals with different trauma systems, and the analysis conducted by experienced data scientists. It is worth noting that while machine learning models typically rely on a vast array of initial variables, our models utilized all available and practical variables obtainable before an X-ray, such as demographic characteristics, patient signs and symptoms, and routine physical examination data. However, the study also has limitations. Radiologic assessment was primarily conducted using X-rays (96.8% of cases), despite CT scans being the gold standard for chest trauma evaluation.

In recent years, numerous studies have been conducted to evaluate the accuracy of various rule-out criteria tools for chest trauma (21-23). Among these, machine learning models have emerged as relatively newer tools with the same objective. Although there is existing literature on machine learning in trauma (24, 25), few studies have specifically evaluated the radiologic outcomes of chest trauma patients. Future multi-centric studies in varied settings are necessary to validate these results. Additionally, prospective studies are needed to assess the real-world impact of machine learning models in clinical practice, particularly in prioritizing patients for imaging. Further research could also explore integrating machine learning models with other diagnostic tools or clinical decision support systems to enhance accuracy and efficiency in trauma diagnosis.

## 5. Limitations

Despite the promising results, our study has several limitations. First, we used a dataset from a single cross-sectional survey conducted in 2015, which may limit the generalizability of our findings to more recent trauma populations with evolving clinical practices. Also, our models were developed and validated on the same dataset, and external validation on independent datasets is needed to confirm their real-world applicability. Second, most radiologic assessments were based on chest X-rays rather than CT scans. However, although CT scans are more accurate, they are less frequently used. Finally, clinical implementation and integration into decision-making workflows require further prospective validation to assess their ability to predict patient outcomes and healthcare efficiency.

## 6. Conclusions

Trauma imposes a significant burden on emergency departments, particularly in high-demand situations. In our study, we explored the role of machine learning algorithms in predicting outcomes for chest trauma patients, with all eight models achieving impressive results, each exceeding an AUC of 0.96. These findings suggest that machine learning models can serve as highly sensitive tools to assist physicians in predicting outcomes prior to X-rays or CT scans for multiple trauma patients, ultimately aiding in more efficient and accurate clinical decision-making.

## 7. Declarations

### 7.1. Acknowledgments

None.

### 7.2. Author Contribution

Study design: MY, AS; Data management: MV, MYa, MY, AG; Analysis: MY, MV, AG; Interpretation: All authors; Drafting: MV, MYa, AG, HZ; Revised: All authors. All authors read and approved the final version.

### 7.3. Funding/Support

This study was supported by Iran University of Medical Sciences (Grant number:1402-3-75-25080).

### 7.4. Conflict of interest

The authors declare no conflicts of interest.

### 7.5. Using artificial intelligence chatbots

No artificial intelligence chatbots were used in the preparation of this manuscript.

### 7.6. Data availability and source codes

Data and python source codes will be provided by the corresponding author on reasonable request.

## References

1. Osifo OD, Iribhogbe PE, Ugiagbe EE. Epidemiology and pattern of paediatric and adolescent trauma deaths in a level 1 trauma centre in Benin city, Nigeria. Injury. 2012;43(11):1861-4.
2. Yadollahi M. A study of mortality risk factors among trauma referrals to trauma center, Shiraz, Iran, 2017. Chin J Traumatol. 2019;22(4):212-8.
3. Norman R, Matzopoulos R, Groenewald P, Bradshaw D. The high burden of injuries in South Africa. Bull World Health Organ. 2007;85(9):695-702.
4. Anghele M, Marina V, Anghele A-D, Moscu C-A, Dragomir L. Negative Factors Influencing Multiple-Trauma Patients. Clinics and Practice. 2024;14(4):1562-70.
5. Dennis BM, Bellister SA, Guillamondegui OD. Thoracic trauma. Surgical Clinics. 2017;97(5):1047-64.
6. Kool DR, Blickman JG. Advanced Trauma Life Support. ABCDE from a radiological point of view. Emerg Radiol. 2007;14(3):135-41.
7. Ong D, Cheung M, Cuenca P, Schauer S. Clinical Utility of Routine Chest X-Rays During the Initial Stabilization of Trauma Patients. South Med J. 2019;112(1):55-9.
8. Safari S, Yousefifard M, Baikpour M, Rahimi-Movaghar V, Abiri S, Falaki M, et al. Validation of thoracic injury rule out criteria as a decision instrument for screening of chest radiography in blunt thoracic trauma. J Clin Orthop Trauma. 2016;7(2):95-100.
9. Rodriguez RM, Anglin D, Langdorf MI, Baumann BM, Hendey GW, Bradley RN, et al. NEXUS chest: validation of a decision instrument for selective chest imaging in blunt trauma. JAMA Surg. 2013;148(10):940-6.
10. Safari S, Radfar F, Baratloo A. Thoracic injury rule out criteria and NEXUS chest in predicting the risk of traumatic intra-thoracic injuries: A diagnostic accuracy study. Injury. 2018;49(5):959-62.
11. Kulshrestha S, Dligach D, Joyce C, Gonzalez R, O'Rourke AP, Glazer JM, et al. Comparison and interpretability of machine learning models to predict severity of chest injury. JAMIA Open. 2021;4(1):ooab015.
12. Tan H, Xu H, Yu N, Yu Y, Duan H, Fan Q, et al. The value of deep learning-based computer aided diagnostic system in improving diagnostic performance of rib fractures in acute blunt trauma. BMC Med Imaging. 2023;23(1):55.
13. Gipson J, Tang V, Seah J, Kavnoudias H, Zia A, Lee R, et al. Diagnostic accuracy of a commercially available deep-learning algorithm in supine chest radiographs following trauma. Br J Radiol. 2022;95(1134):20210979.
14. Artificial Intelligence-enabled Automated Medical Prediction and Diagnosis in Trauma Patients. Dans: CRC Press eBooks; 2022. p. 135 45.
15. Lyu WH, Xia F, Zhou CS, Huang M, Ding WW, Zhang S, et al. [Application of deep learning-based chest CT auxiliary diagnosis system in emergency trauma patients]. Zhonghua Yi Xue Za Zhi. 2021;101(7):481-6. [Article in Chinese]
16. Wang S, Wu D, Ye L, Chen Z, Zhan Y, Li Y. Assessment of automatic rib fracture detection on chest CT using a deep learning algorithm. Eur Radiol. 2023 Mar;33(3):1824-1834. doi: 10.1007/s00330-022-09156-w. Epub 2022 Oct 10. PMID: 36214848.
17. Shahverdi Kondori M, Malek H. Determining the Need for Computed Tomography Scan Following Blunt Chest Trauma through Machine Learning Approaches. Arch Acad Emerg Med. 2021;9(1):e15.
18. Staziaki PV, Wu D, Rayan JC, Santo IDO, Nan F, Maybury A, et al. Machine learning combining CT findings and clinical parameters improves prediction of length of stay and ICU admission in torso trauma. Eur Radiol. 2021;31(7):5434-41.
19. Lee KC, Lin TC, Chiang HF, Horng GJ, Hsu CC, Wu NC, et al. Predicting outcomes after trauma: Prognostic model development based on admission features through machine learning. Medicine (Baltimore). 2021;100(49):e27753.
20. Shukla V, Pandey M, Sahu SK, Singh A, Khare A, Jeswani M, Bisen J, Lahariya CP. Clinicoepidemiological Study of Traumatic Chest Injuries in a Tertiary Care Center. Int J Sci Stud 2020;8(2):12-17.
21. Asgarzadeh S, Feizi B, Sarabandi F, Asgarzadeh M. Thoracic Injury Rule out Criteria in Prediction of Trau-

matic Intra-thoracic Injuries; a Validation Study. Emerg (Tehran). 2017;5(1):e27.

22. Yousefifard M, Hosseini M, Parvizi MR. Pediatric Thoracic Injury Rule out Criteria (pTIRC) in Diagnosis of Very Low Risk Children for Traumatic Intrathoracic Injuries; a Diagnostic Accuracy Study. Arch Acad Emerg Med. 2020;8(1):e7.
23. Vazirizadeh-Mahabadi M, Yarahmadi M. Canadian C-spine Rule versus NEXUS in Screening of Clinically Important Traumatic Cervical Spine Injuries; a systematic review and meta-analysis. Arch Acad Emerg Med. 2023;11(1):e5.
24. Peng HT, Siddiqui MM, Rhind SG, Zhang J, da Luz LT, Beckett A. Artificial intelligence and machine learning for hemorrhagic trauma care. Military Medical Research. 2023;10(1):6.
25. Zhang T, Nikouline A, Lightfoot D, Nolan B. Machine Learning in the Prediction of Trauma Outcomes: A Systematic Review. Annals of Emergency Medicine. 2022;80(5):440-55.

**Table 1:** Baseline characteristics of the studied patients

| **Characteristics** | **Value** |
|---|---|
| **Demographic and background information** | |
| **Age (year)** | |
| <60 | 2608 (91.2) |
| ≥60 | 252 (8.8) |
| **Sex** | |
| Male | 2090 (73.1) |
| Female | 770 (26.9) |
| **Mechanism of trauma*** | |
| Motorcycle accident | 357 (12.5) |
| Pedestrian accidents | 1094 (38.3) |
| Car accident | 881 (30.8) |
| Fall over 3m | 132 (4.6) |
| Over ground fall | 269 (9.4) |
| Other | 127 (4.4) |
| **Physical examination findings** | |
| Glasgow Coma Scale < 15 | 296 (10.3) |
| Dyspnea | 443 (15.5) |
| Distracting pain | 1145 (40) |
| Thoracic skin abrasion | 447 (15.6) |
| Chest tenderness | 564 (19.7) |
| Chest deformity | 50 (1.7) |
| Crepitation | 122 (4.3) |
| Abdominal tenderness | 439 (15.3) |
| Decrease in pulmonary sounds | 166 (5.8) |
| Chest pain | 721 (25.2) |
| Subcutaneous emphysema | 74 (2.6) |
| **Vital signs** | |
| Heart rate (/minute) | 97.73± 15.20 |
| Systolic blood pressure (mmHg) | 114.35± 19.21 |
| Diastolic blood pressure (mmHg) | 74.00± 10.55 |
| Respiratory rate (/minute) | 14.08± 2.20 |
| Peripheral capillary oxygen saturation (%) | 97.22± 2.06 |
| **Imaging findings** | |
| Pathologic radiologic findings | 514 (18) |
| Pneumothorax | 134 (4.7) |
| Hemothorax | 111 (3.9) |
| Rib fracture | 217 (7.6) |
| Other radiologic findings | 152 (5.3) |
| Hemo-mediastinum | 1 (0.00) |
| Mediastinal widening | 35 (1.2) |
| Lung contusion | 172 (6.0) |
| Diaphragm injury | 14 (0.5) |
| Emphysema in radiography | 103 (3.6) |

Data are presented as mean ± standard deviation or frequency (%). *Motorcycle accident: including motorcycle to caraccident and motorcycle to motorcycle accident. Pedestrian: including accidents of pedestrians with motorcycles or cars. Car accidents: including car-to-car accidents and car rollovers.

**Table 2:** Diagnostic accuracy of machine learning models

| Model | AUC | Accuracy | Balanced Accuracy | Precision | F1 Score | LR+ | LR- | Sensitivity/ Recall | Specificity |
|---|---|---|---|---|---|---|---|---|---|
| Random Forest | 0.999 | 0.99 | 0.99 | 0.991 | 0.99 | 118.684 | 0.011 | 0.989 | 0.992 |
| HistGradient Boosting | 0.999 | 0.994 | 0.994 | 0.991 | 0.993 | 119.474 | 0.004 | 0.996 | 0.992 |
| XGBoost | 0.997 | 0.991 | 0.992 | 0.985 | 0.991 | 68.421 | 0.002 | 0.998 | 0.985 |
| Decision Tree | 0.986 | 0.986 | 0.986 | 0.987 | 0.986 | 78.772 | 0.016 | 0.985 | 0.988 |
| SVM | 0.963 | 0.913 | 0.914 | 0.887 | 0.914 | 8.23 | 0.064 | 0.943 | 0.885 |
| Logistic Regression | 0.969 | 0.917 | 0.917 | 0.909 | 0.915 | 10.526 | 0.087 | 0.921 | 0.912 |
| K-Nearest Neighbor | 0.996 | 0.982 | 0.982 | 0.972 | 0.982 | 36.599 | 0.009 | 0.991 | 0.973 |
| Neural Network | 0.996 | 0.981 | 0.981 | 0.978 | 0.98 | 47.158 | 0.018 | 0.982 | 0.979 |

AUC: Area Under the Curve; LR: Likelihood Ratio; HistGradient Boosting: Histogram-based Gradient Boosting; XGBoost: Extreme Gradient Boosting; SVM: Support Vector Machine.

**Table 3:** Top 10 important features, which are extracted from using both Random Forest and XGBoost models

| Feature | Random Forest | XGBoost | Average |
|---|---|---|---|
| Chest pain | 0.235217 | 0.37151 | 0.303363 |
| Dyspnea | 0.111621 | 0.313063 | 0.212342 |
| Chest tenderness | 0.120489 | 0.022761 | 0.071625 |
| RR | 0.111543 | 0.026628 | 0.069086 |
| SPo2 | 0.102818 | 0.013382 | 0.0581 |
| Abrasion chest | 0.026314 | 0.082643 | 0.054479 |
| SBP | 0.065695 | 0.032849 | 0.049272 |
| DBP | 0.037199 | 0.019822 | 0.028511 |
| Mechanism of trauma | 0.036064 | 0.016223 | 0.026144 |
| Abdominal tenderness | 0.014884 | 0.030698 | 0.022791 |

RR: Respiratory Rate; SpO2: Peripheral Capillary Oxygen Saturation; SBP: Systolic Blood Pressure;
DBP: Diastolic Blood Pressure.

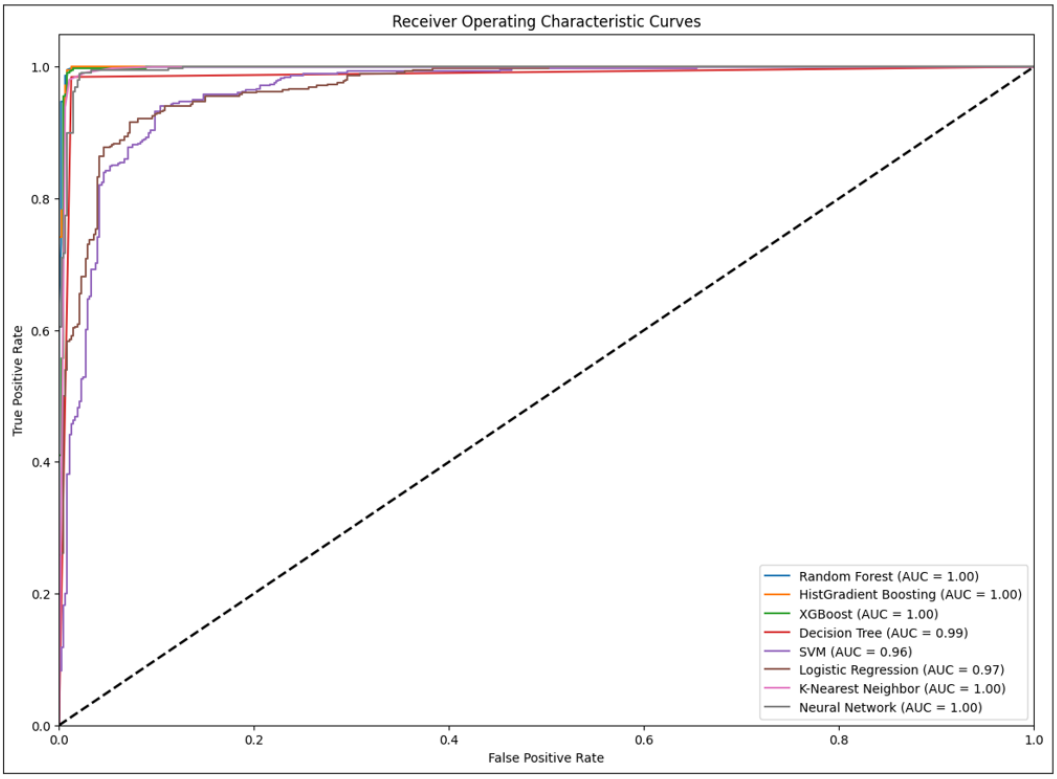

**Figure 1:** Area under the curve of the machine learning models in predicting the presence of intrathoracic injuries following blunt chest trauma. AUC: Area Under the Curve; HistGradient Boosting: Histogram-based Gradient Boosting; XGBoost: Extreme Gradient Boosting; SVM: Support Vector Machine

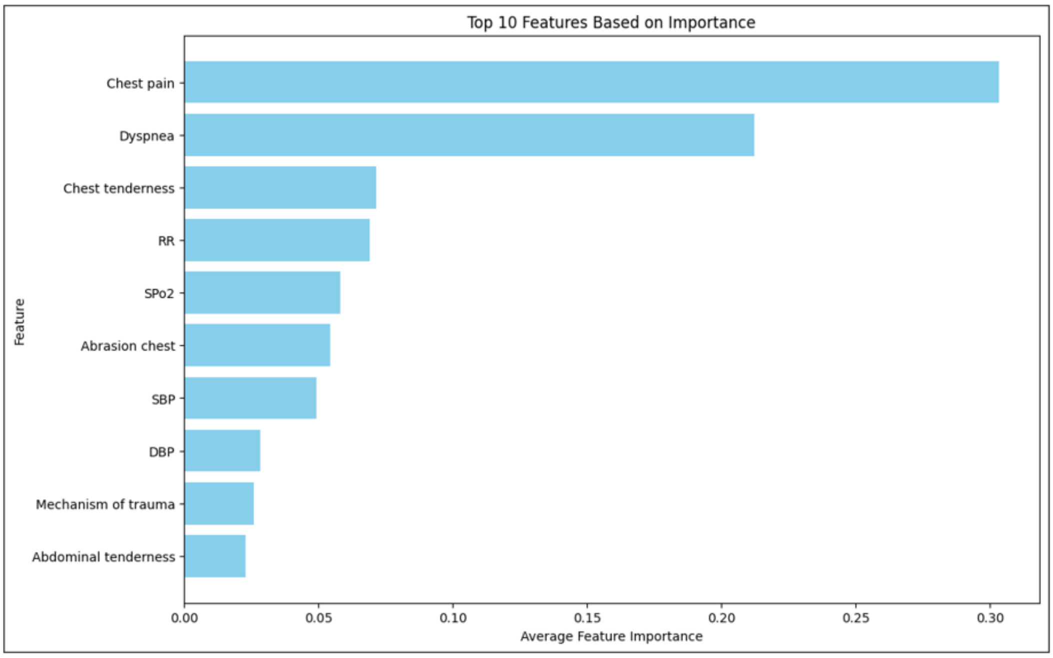

**Figure 2:** Top 10 important features, which are extracted from the machine learning study. RR: Respiratory Rate; SpO2: Peripheral Capillary Oxygen Saturation; SBP: Systolic Blood Pressure; DBP: Diastolic Blood Pressure.

**Supplementary table 1:** TEffect of ablation study findings on accuracy, balanced accuracy, precision, recall, F1 Score, and AUC-ROC differences

| Feature | Differences | | | | | |
|---|---|---|---|---|---|---|
| | Accuracy | Balanced Accuracy | Precision | Recall | F1 Score | AUC-ROC |
| Age | 0 | -0.000055 | 0.002159 | -0.002193 | -0.000017 | 0.000411 |
| Sex | -0.001068 | -0.001042 | -0.002164 | 0 | -0.001086 | -0.001012 |
| Trauma mechanism | 0 | 0 | 0 | 0 | 0 | -0.001272 |
| GCS | -0.003205 | -0.00318 | -0.004348 | -0.002193 | -0.003274 | -0.003024 |
| HR | -0.002137 | -0.002138 | -0.002188 | -0.002193 | -0.002191 | -0.001181 |
| SBP | 0 | 0 | 0 | 0 | 0 | 0.000009 |
| DBP | 0 | 0 | 0 | 0 | 0 | 0.000146 |
| RR | 0.002137 | 0.002083 | 0.004357 | 0 | 0.002179 | -0.000142 |
| SPo2 | 0 | -0.000055 | 0.002159 | -0.002193 | -0.000017 | -0.000135 |
| Dyspnea | -0.001068 | -0.001096 | -0.000019 | -0.002193 | -0.001105 | 0.000039 |
| Distract pain | 0.001068 | 0.001042 | 0.002174 | 0 | 0.001088 | 0.000199 |
| Chest abrasion | 0.001068 | 0.001042 | 0.002174 | 0 | 0.001088 | -0.00024 |
| Chest tenderness | 0 | 0.000055 | -0.002141 | 0.002193 | 0.000017 | -0.000183 |
| Chest deformity | 0 | 0 | 0 | 0 | 0 | -0.000311 |
| Crepitation | 0 | 0 | 0 | 0 | 0 | 0.000087 |
| Abdominal tenderness | -0.001068 | -0.001096 | -0.000019 | -0.002193 | -0.001105 | 0.000169 |
| Chest sounds | 0 | 0 | 0 | 0 | 0 | -0.000021 |
| Chest pain | -0.001068 | -0.001206 | 0.004328 | -0.006579 | -0.001144 | -0.001601 |
| Subcutaneous emphysema | -0.001068 | -0.001096 | -0.000019 | -0.002193 | -0.001105 | -0.000066 |

AUC-ROC: Area Under the Receiver Operating Characteristic Curve; RR: Respiratory Rate; SpO2: Peripheral Capillary OxygenSaturation; SBP: Systolic Blood Pressure; DBP: Diastolic Blood Pressure; HR: Heart Rate; GCS: Glasgow Coma Scale.

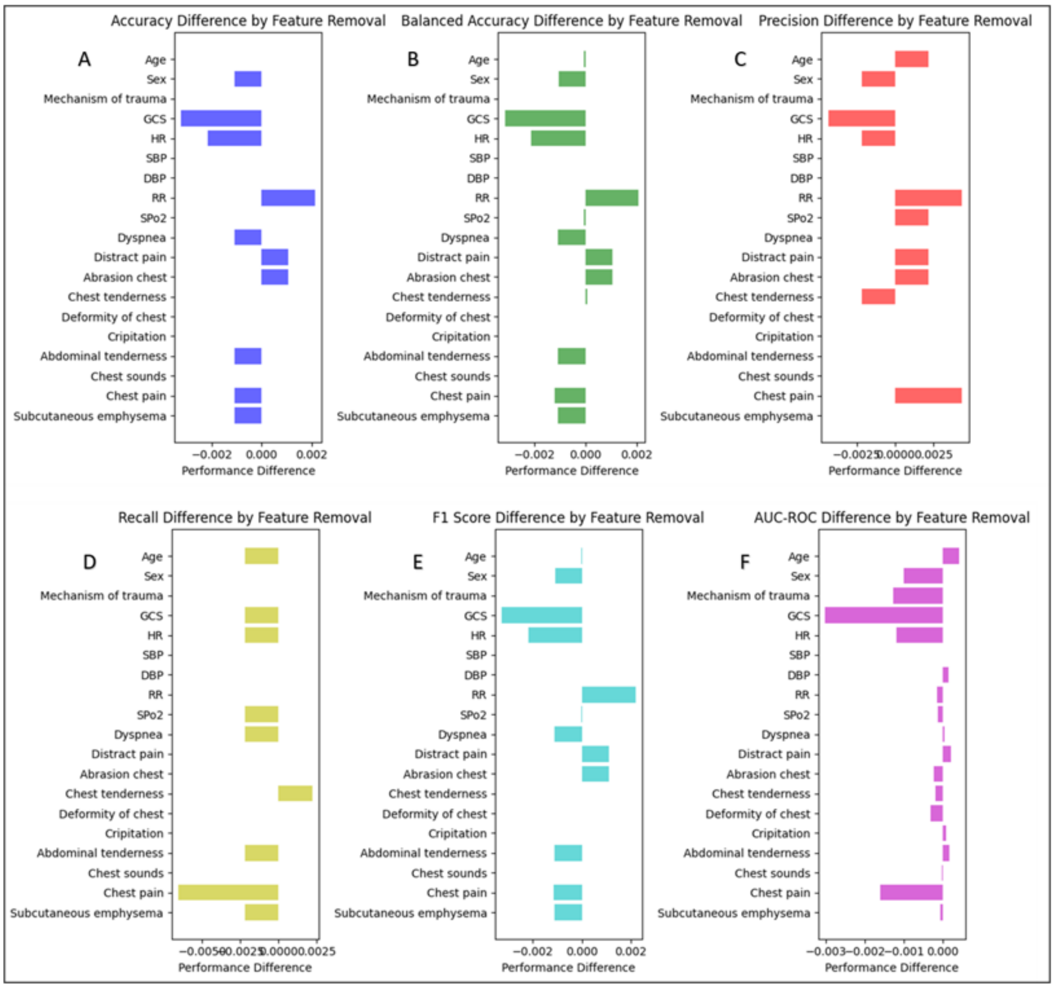

**Figure 3:** Ablation study findings after removal of each feature in A) accuracy, B) balanced accuracy, C) precision, D) recall, E) F1 Score, F) AUC-ROC differences. RR: Respiratory Rate; SpO2: Peripheral Capillary Oxygen Saturation; SBP: Systolic Blood Pressure; DBP: Diastolic Blood Pressure; HR: Heart Rate; GCS: Glasgow Coma Scale; AUC-ROC: Area Under the Receiver Operating Characteristic Curve.

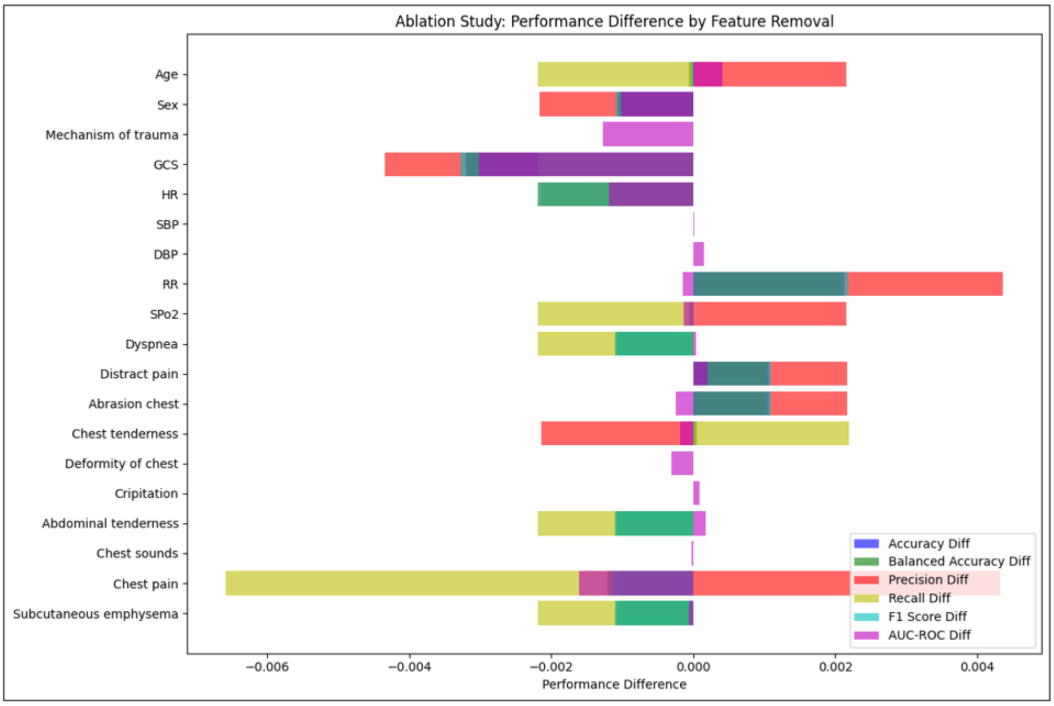

**Figure 4:** Cumulative summary of the ablation study's effect on evaluation metrics. AUC-ROC: Area Under the Receiver Operating Characteristic Curve; Diff: Difference; RR: Respiratory Rate; SpO2: Peripheral Capillary Oxygen Saturation; SBP: Systolic Blood Pressure; DBP: Diastolic Blood Pressure; HR: Heart Rate; GCS: Glasgow Coma Scale.