ORIGINAL ARTICLE

COMPARISON OF MULTIPLE ARTIFICIAL INTELLIGENCE-MODELS' PREDICTIVE POWER IN DETECTING DELAYED ANALGESIA IN EMERGENCY DEPARTMENT PATIENTS

Ayesha Saeed, Muhammad Moeed Azwar Bhatti, Zubia Razzaq*, Zainub Saeed**

Department of Accident and Emergency, *Physiology, HITEC-Institute of Medical Sciences, Taxila, **COMSATS University, Attock, Pakistan

Background: Artificial intelligence (AI) has significant potential to enhance risk assessment by identifying patients at higher risk of delayed analgesia. The goal of this work was to create and validate AI models that predict the probability of delayed analgesia, and compare the predictive power of multiple AI models in detecting delayed analgesia in emergency department (ED) patients and avoiding longer patient stays. **Methods:** From Dec 2024 to Jan 2025, 300 adult patients with moderate to severe pain were studied in the Emergency Department of an academic facility teaching hospital. This retrospective observational study was collected and analysed retrospectively, with age, gender, triage category, triage pain score, and presentation during peak hours serving as input features. Five machine learning models were constructed and compared for their accuracy to forecast delayed analgesia. Important predictors were identified using SHAP (SHapley Additive exPlanations) values for the AI model with the highest accuracy. **Results:** Random Forest and J48 achieved 77% accuracy, with Random Forest having greater recall for delayed cases (Precision=0.71, Recall=0.84) for anticipating delayed analgesia. Naive Bayes and Logistic Regression had low recall for delayed cases, while MLP Neural Network demonstrated moderate predictive usefulness. Random Forest model had the best performance with the highest AUC [0.83 (95% CI: 0.75–0.90)] on ROC analysis. **Conclusion:** AI models were successfully implemented, Random Forest outperforming the others, for early identification of patients at risk of delayed analgesia. AI models can decrease unnecessary delays and improve pain management in ED.

Keywords: Acute pain management, Artificial Intelligence, Delayed analgesia, Emergency medicine, Machine Learning, Random Forest, Triage

Pak J Physiol 2026;22(1):52–7, **DOI:** https://doi.org/10.69656/pjp.v22i1.1931

INTRODUCTION

Delayed analgesia is generally defined as the time to get pain medication for patients presenting with painful conditions since their arrival in the emergency department. There is no single universal 'standard' definition for it, as the specific time frame can vary depending on the clinical context and institutional guidelines. The Royal College of Emergency Medicine advises giving adequate analgesia to patients arriving with moderate to severe pain within 15 minutes of presentation.[1] Often used as a quality indicator of patient care, emergency departments around the world struggle with the problem of delayed analgesia. More than a major percentage of patients experience delays longer than the first hour of ED visit, with a median time to analgesia exceeding two hours.[2,3] Lengthy hospital stays create lower patient satisfaction, and inadequacy of pain management is all linked to these delays.

Traditional triage methods have a reduced dynamic capability to identify individuals at risk of under treatment. The intricate correlation between patient demography, pain intensity, departmental workload, and understaffing in emergency department settings may not be sufficiently captured by traditional triage systems, which depend more on rule-based algorithms and clinician judgment. Because of this, patients with similar levels of pain may have quite diverse analgesic timeframes.

Machine learning (ML), a subset of AI, provides a clear benefit in emergency treatment by providing real-time risk stratification using routinely obtained triage data. The ML algorithms, in contrast to traditional statistical models, are capable of integrating high-dimensional data, modelling non-linear relationships, and adapting to varied patient strata. Because of these characteristics, machine learning is especially well-suited for dynamic yet time-sensitive environments such as the Emergency Department (ED), where quick, data-driven decisions may help enhance rather than replace clinical judgment.[4–7]

Literature supports the use of ML in acute care and pain management. AI-driven models have shown improved real-time pain management after caesarean delivery, better prediction of analgesic needs in postoperative populations, and precise identification of patients at risk for long-term opiate usage.[6,8–9] Explainability in AI is a crucial prerequisite for clinician trust, ethical use, and system acceptance of AI in clinical practice. Explainable AI, such as SHAP provides transparent, case-level, and population-level insights into model predictions. Explainable ML enhances clinical interpretability, facilitates auditability, and predicts potential bias by determining the relative contribution of individual input variables.[10] These important factors are highlighted in the STROBE-AI and JMIR-AI reporting standards. The use of explainable ML models to forecast delayed analgesia at the triage level is mainly unexplored. Bridging this gap could help identify high-risk patients early, promote focused care, and provide guidance for quality improvement initiatives.

This study aimed to develop and validate explainable AI models to predict delayed analgesia among ED patients using triage data. The secondary objective was to compare the predictive performance of five machine learning algorithms—Random Forest, Multi-layer Perceptron (MLP), J48 Decision Tree, Naive Bayes, and Logistic Regression, with reference to delayed analgesia.

METHODOLOGY

A total of 300 patients' data were analysed in a retrospective analytical study from emergency medical records. Data of one patient was deleted who had mild pain (pain score 3/10). Random selection was used to include patients (≥18 years old) with moderate to severe pain (Numeric Rating Scale ≥4/10) in the emergency department of a teaching hospital, which receives more than 80,000 ED visits annually. Patients who were triaged for emergency resuscitation or had missing values were excluded from the final sample by list-wise deletion.

The sample size was determined based on the stated accuracy (72%) of the reference AI model[5] with a confidence interval of 95%, a study power of 80%, and a ±5% margin of error. Our study included 300 patients, a number close to the calculated size (n=310) and deemed adequate. As per the literature review, no AI model had predicted delayed analgesia in emergency rooms previously; the authors relied on a post-operative patient study[5] for sample computation.

Input variables (age, gender, initial pain score, triage category, and presentation during peak hours) were documented at triage, and the time to analgesia was reported at the patient's bedside. Features were selected while considering their influence over the promptness of acute pain management in prior literature, such as pain severity[11], triage level[12], overcrowding[13], along with patient demographics[14]. The pain score (4–10 for moderate to severe) was reported using the Numeric Rating Scale (NRS), a freely available, frequently used 11-point pain assessment tool that rates pain on a scale of 0 to 10 (0 as no pain, 1–3 labelled as mild, 4–6 as moderate, and 7–10 as severe pain out of 10).[15] The triage category had been assigned according to the Emergency Severity Index (ESI, Version 4), a validated, five-level triage method (from Level 1= most urgent, e.g., unresponsive, to Level 5= least urgent, e.g., history/exam only), freely available for non-commercial research and clinical use.[16,17] Trained staff were deployed for triaging, and the senior triage person was responsible for double-checking the category to avoid personal bias. Presentation during peak hours was documented as nominal data. Time to analgesia of more than 15 minutes since arrival was defined as delayed analgesia, and it was reported as a primary outcome.[1]

The data file of 299 patients in SPSS-25 was exported to Python 3 (Colab) for building and validating AI models. Continuous variables were reported as Mean±SD, while skewed data distributions were represented as medians (IQR). Categorical variables were reported as frequencies with proportions. All tests were two-tailed with a significance level of $p<0.05$. Statistical analysis was done on IBM SPSS-25.[18] AI/ML models were created in Python (Version 3.8) using the Google Colaboratory environment,[19,20] both of which are freely available for research and academic purposes. Data was cleaned by removing invalid or missing values, numerically encoding binary and categorical characteristics, and standardizing input data formats. Categorical variables were transformed using one-hot encoding. To address class imbalance, the Synthetic Minority Over sampling Technique (SMOTE)[21] was applied resulting in a balanced training set of records. Feature scaling was applied only to the input of the Multi-Layer Perceptron (MLP) model.

Five supervised machine learning algorithms were trained using Python's scikit-learn[22] and imbalanced-learn libraries[23] and a fixed random state for reproducibility: Random Forest, J48 (decision tree), Logistic Regression, MLP, and Naive Bayes. Data was split using a 70/30 stratified split for training and testing. Models were evaluated on the holdout test set. The area under the receiver operating characteristic curve (AUC/ROC) was computed for all models to assess the individual model's discriminative power.

To reflect performance on clinically significant instances, class-wise metrics such as accuracy, precision, recall, F1-score, confusion matrices, macro, and weighted averages were explicitly examined for class 1 (delayed analgesia) and class 0 (timely analgesia). SHAP analysis was performed to identify the input variables with the highest prediction probabilities.[24]

Ethical approval for this study was obtained from the Institutional Review Board/Ethical Review Board, which waived consent for the study. The study workflow is shown in Figure-1.

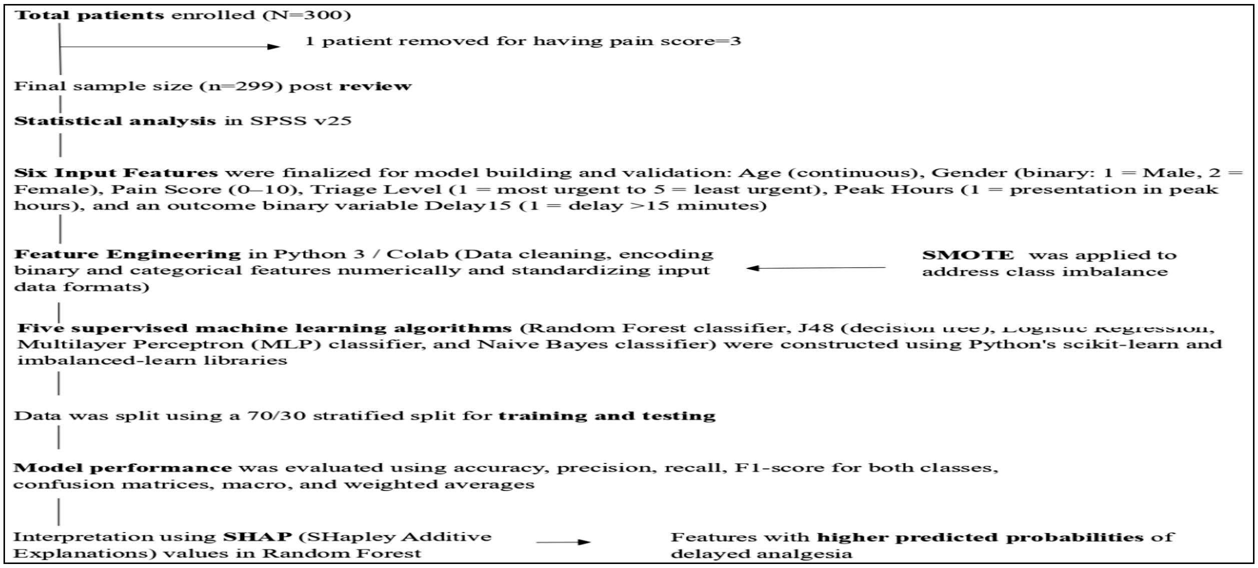

Figure-1: The study workflow

RESULTS

The study involved 299 individuals, with 24.7% experiencing delayed analgesia. The cohort's median age was 38 years (IQR 21), with a male predominance, and no significant difference in age between the two groups (p=0.341). Severe pain was reported by most of the patients, with similar rates in both groups (82.4% *vs* 81.3%, p=0.899). The majority of patients (68.2%) reported during peak hours. Critical triage was allocated to 45.5% of the total cohort, with similar distributions between the delayed and timely groups (40.5% *vs* 47.1%, p=0.325). The delayed group had slightly higher mean age, larger male preponderance, and a higher proportion of presentations during peak hours than the timely group, but pain scores and triage distributions were similar across both groups (Table-1).

The confusion matrix analysis added to our understanding of each model's classification behaviour and offered information on its classification performance. Random Forest properly recognized 55 true positives and 49 true negatives, with 12 false positives and 19 false negatives, demonstrating reliable discrimination between the two classes. J48 Decision Tree performed similarly, with more false negatives (n=17). While Naive Bayes had the highest number of true positives (61), it also had the highest number of false positives (n=26) resulting in lower overall specificity. In contrast, MLP and Logistic Regression provided more balanced but less accurate outputs, with more false negatives (24 and 31, respectively) which may restrict their clinical reliability.

The ROC curves for all five models revealed that all classifiers outperformed chance (diagonal reference line, AUC=0.5), (Figure-2). The Random Forest model consistently demonstrated the greatest divergence from the reference line, particularly in the higher sensitivity range, resulting in the best AUC (0.83). The J48 decision tree (AUC=0.80) and MLP (AUC=0.79) both performed well, with ROC curves trailing the Random Forest. Naive Bayes (AUC=0.76) demonstrated moderate separation, whereas logistic regression (AUC=0.72) had the lowest curve elevation, indicating a lesser discriminative capacity to identify complicated predictor associations.

The Random Forest classifier demonstrated the best overall performance, with an accuracy of 77% (95% CI: 69–84), an F1 score of 0.77 for delayed instances, and an AUC of 0.83 (95% CI: 0.75–0.90), indicating a balanced sensitivity and specificity. The J48 Decision Tree similarly demonstrated significant predictive abilities with slightly higher precision for non-delayed cases than Random Forest, but marginally lower recall for delayed cases. In comparison, the Naive Bayes model performed less well, primarily due to a lower recall for delayed cases, despite achieving adequate precision. This result shows that Naive Bayes underperformed in detecting patients at risk of delayed analgesia, which is crucial in a therapeutic setting. The MLP demonstrated moderate predictive performance, with balanced but modest precision and recall. Finally, logistic regression showed the lowest predictive accuracy and AUC, with a lower recall for delayed cases (Table-2).

SHAP analysis on the top-performing AI model (Random Forests) revealed the mean absolute SHAP values of age (0.074568) and pain score (0.074568) as the two most significant parameters influencing the model's output (Figure-3). A higher score for these features indicated a greater probability of a delay, suggesting that older patients with higher pain scores are more likely to experience delayed pain alleviation.

Table-1: Baseline characteristics of study cohort and stratified comparison of delayed and timely analgesia [n (%)]

Characteristics	Overall (n=299)	Delayed Analgesia (n=74, 24.7%)	Timely Analgesia (n=225, 75.3%)	*p*
Age (Years) [Median (IQR)]	38 (21)	38 (21.5)	38 (21)	0.341
Gender				
Male	218 (72.9)	58 (78.4)	159 (70.7)	0.197
Female	81 (27.1)	16 (21.6)	66 (29.3)	
Pain Score				
Severe	245 (81.9)	61 (82.4)	183 (81.34)	0.899
Moderate	54 (18.1)	13 (17.56)	42 (18.67)	
Presentation Time				
Peak	204 (68.2)	45 (60.8)	158 (70.2)	0.133
Non-peak	95 (31.8)	29 (39.2)	67 (29.8)	
Triage Level				
Critical	136 (45.5)	30 (40.5)	106 (47.1)	0.325
Non-critical	163 (54.5)	44 (59.5)	119 (52.9)	

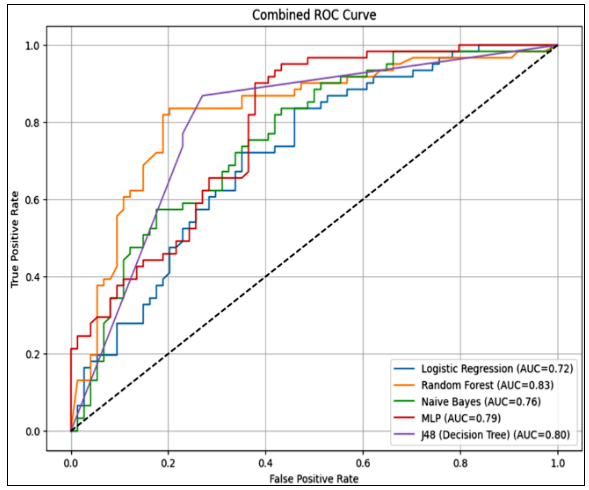

Figure-2: Combined receiver operating characteristic (ROC) curves for five machine learning models

Table-2: Comparison of the predictive performance of AI models

Models	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	AUC/ROC (95% CI)
Random Forest	77 (69–84)	74.3 (63–83)	80.3 (69–89)	82 (71–90)	72 (60–82)	**0.83** (0.75–0.90)
J48 Decision Tree	77 (69–84)	77.0 (66–86)	77.0 (65–86)	80.2 (69–89)	73.4 (61–83)	0.80 (0.72–0.88)
Naive Bayes	71 (63–79)	82.4 (72–90)	57.3 (45–69)	70.1 (59–80)	72.9 (59–83)	0.76 (0.67–0.84)
MLP	67 (59–75)	67.5 (55–78)	67.2 (55–78)	71.4 (59–82)	63 (51–74)	0.79 (0.70–0.87)
Logistic Regression	64 (55–72)	60.5 (48–72)	72.1 (60–82)	71.6 (59–82)	58.6 (46–71)	0.72 (0.63–0.81)

PPV=Positive predictive value, NPV=Negative predictive value

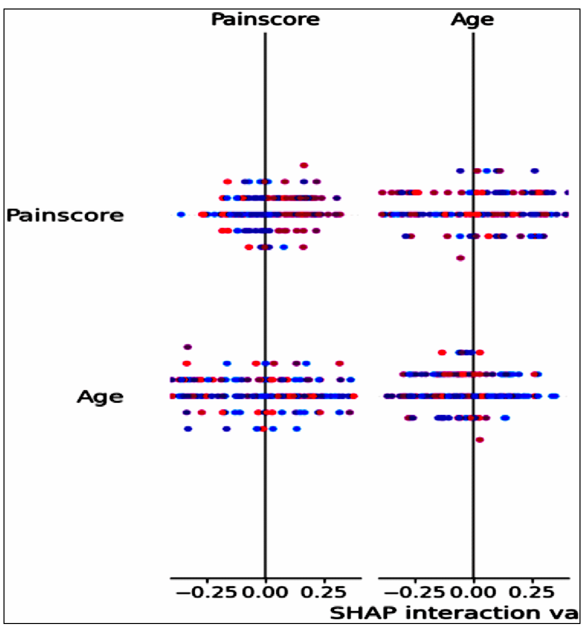

Figure-3: Beeswarm plot for SHAP analysis of top predictors

DISCUSSION

Predicting delayed analgesia entails identifying high-risk individuals based on clinical criteria. While AI is working in medical predictive and risk stratification domains, this study is among the initial ones to demonstrate its efficacy in predicting delayed analgesia in the emergency departments using quick triage data, —a previously unexplored area. The study employed machine learning techniques, including Random Forest, Naive Bayes, Multilayer Perceptron, J48 Decision Tree, and Logistic Regression, to predict delayed analgesia delivery in emergency department patients by examining complex interactions.

The comparative analysis of the models suggested that ensemble and decision-tree models (Random Forest, J48) handled structured triage data precisely, demonstrating their clinical utility in predicting delayed analgesia in the emergency department. The Random Forest model had the greatest AUC and balanced classification metrics. It was able to identify both delayed and non-delayed analgesia cases at a rate of 77% (the highest predictive accuracy among all five models). From a clinical standpoint, Random Forest and J48 provided the most dependable trade-off between sensitivity and specificity, ensuring the identification of at-risk individuals without over classification. Previous literature has also supported the efficacy of Random Forest in a systematic review of fifteen studies, with the decision-tree-based boosting algorithms in 5 studies with AUC ranging from 0.81 to 0.66 as the best performing models.[9]

SHAP analysis on the Random Forest revealed that older patients with higher pain scores were more likely to experience delayed pain relief.

Bloom *et al*[25] and Platts-Mills *et al*[26] have also made similar observations, supporting the idea that SHAP-informed AI tools can help risk classification at ED triage and provide a more comprehensive understanding of the model.

A fundamental strength of this study is the use of accessible triage-level data, which renders the models as simple, explainable, and viable for real-time deployment in high-pressure emergency departments. Gabriel *et al*[4] created a neural network model to predict outpatient opioid refills after surgery, with an AUC of 0.75 and an accuracy of 73%, and our study, which relied entirely on triage-stage information, achieved a higher accuracy and AUC while producing fewer false negatives. Similarly, other studies with comparable model accuracy have found promising results in management pathways.[5,27] Our methodology was generalized across varied emergency patients and explicitly favoured strong recall (84%) to reduce analgesic delays—a critical goal in emergency medicine.

LIMITATIONS

The limited sample size of this study makes generalizability difficult, despite the use of cross-validation processes. The models were verified internally on a single-centre basis and excluded ED operational features such as staff-to-patient ratios, revealing departmental work intensity. These factors demonstrated the challenges of assessing time-to-analgesia in real-world emergency rooms.

CONCLUSION

The study shows the successful implementation of AI models to predict delayed analgesia in emergency department patients using easily accessible triage data. Our approach provides a preliminary, interpretable, and practical solution for early identification of patients at risk of delayed analgesia. Random Forest outperformed the other AI models in predicting delayed analgesia. However, external validation is recommended, and AI models can be integrated into clinical decision support systems to optimize patient care pathways, decrease unnecessary delays, and improve pain management in high-volume emergency settings.

REFERENCES

1. Paul Hunt JF. Management of acute pain in adults in the Emergency Department: Summary of recommendations. Royal College of Emergency Medicine; 2024.
2. Pines JM, Shofer FS, Isserman JA, Abbuhl SB, Mills AM. The effect of emergency department crowding on analgesia in patients with back pain in two hospitals. Acad Emerg Med 2010;17(3):276–83. doi:10.1111/j.1553-2712.2009.00676.x
3. Arendts G, Fry M. Factors associated with delay to opiate analgesia in Emergency Departments. J Pain 2006;7(9):682–6. doi:10.1016/j.jpain.2006.03.003.
4. Gabriel RA, Simpson S, Zhong W, Burton BN, Mehdipour S, Said ET. A neural network model using pain score patterns to predict the need for outpatient opioid refills following ambulatory surgery: algorithm development and validation. JMIR Perioper Med 2023;6:e40455. https://periop.jmir.org/2023/1/e40455/
5. Nair AA, Velagapudi MA, Lang JA, Behara L, Venigandla R, Velagapudi N, *et al*. Machine learning approach to predict postoperative opioid requirements in ambulatory surgery patients. PLoS One 2020;15(7):e0236833.
6. Matsangidou M, Liampas A, Pittara M, Pattichi CS, Zis P. Machine learning in pain medicine: An up-to-date systematic review. Pain Ther 2021;10(2):1067–84. doi:10.1007/s40122-021-00324-2
7. Lv S, Sun N, Hao C, Li J, Li Y. Development and validation of machine learning models for predicting post-cesarean pain and individualized pain management strategies: a multicenter study. BMC Anesthesiol 2025;25(1):170. doi:10.1186/s12871-025-03034-w
8. Tan CW, Koh JZ, Jin H, Han NR, Cheng SM, Ta AWA, *et al*. Machine learning approach to predict postoperative pain after spinal morphine administration during caesarean delivery. Heliyon 2024;10(23):e40602. doi:10.1016/j.heliyon.2024.e40602
9. Emam OS, Eldaly AS, Avila FR, Torres-Guzman RA, Maita KC, Garcia JP, *et al*. Machine learning algorithms predict long-term postoperative opioid misuse: A systematic review. Am Surg 2024;90(1):140–51. doi:10.1177/00031348231198112
10. Okada Y, Ning Y, Ong MEH. Explainable artificial intelligence in emergency medicine: an overview. Clin Exp Emerg Med 2023;10(4):354–62. doi:10.15441/ceem.23.145
11. Rampanjato RM, Florence M, Patrick NC, Finucane BT. Factors influencing pain management by nurses in emergency departments in Central Africa. Emerg Med J 2007;24(7):475–6. doi:10.1136/emj.2006.045815
12. Wang L, Song C, Bai Y, Huang X, Shi H, Pan J. Practice and reflection on the management mode of pain quality control in emergency pre-check and triage. Ann Palliat Med 2020;9(4):1879–85. doi:10.21037/apm-20-1108
13. Chen YW, Lee JH, Chiang CY, Yeh YN, Lin JC, Tsai MJ. Factors associated with delayed order-to-administration time in the emergency department: a retrospective analysis. BMC Emerg Med 2025;25(1):74. doi:10.1186/s12873-025-01229-5
14. Lautenbacher S, Peters JH, Heesen M, Scheel J, Kunz M. Age changes in pain perception: A systematic-review and meta-analysis of age effects on pain and tolerance thresholds. Neurosci Biobehav Rev 2017;75:104–13. doi:10.1016/j.neubiorev.2017.01.039
15. Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA. Studies with pain rating scales. Ann Rheum Dis 1978;37(4):378–81. doi:10.1136/ard.37.4.378
16. Medicine E. Index, Version 4. AHRQ Pub. No. 05-0046-2 May 2005; 2005. http://www.ahrq.gov/research/esi/
17. Ganjali R, Golmakani R, Ebrahimi M, Eslami S, Bolvardi E. Accuracy of the emergency department triage system using the emergency severity index for predicting patient outcome; A single center experience. Bull Emerg Trauma 2020;8(2):115–20. doi:10.30476/BEAT.2020.46452
18. Corp. I. IBM SPSS Statistics for Windows, Version 25.0.; 2017.
19. Foundation PS. Python Language Reference, version 3.8. https://www.python.org/
20. Llc G. Google Colaboratory. https://colab.research.google.com/?
21. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res 2002;16:321–57. doi:10.1613/JAIR.953
22. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, *et al*. Scikit-learn: Machine learning in python. J Mach Learn Res 2011(12):2825–30. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?source=post_page
23. Lematre G, Nogueira F, Aridas CK. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. J Mach Learn Res 2017;18:1–5. https://www.jmlr.org/papers/volume18/16-365/16-365.pdf

24. Lundberg S, Lee SI. A unified approach to interpreting model predictions. arXiv; 2017. URL: https://arxiv.org/abs/1705.07874 doi:10.48550/ARXIV.1705.07874
25. Bloom B, Fritz CL, Gupta S, Pott J, Sjene I, Astin-Chamberlain R, *et al*. Older age and risk for delayed abdominal pain care in the emergency department? Eur J Emerg Med 2024;31(5):332–8.
26. Platts-Mills TF, Hunold KM, Weaver MA, Dickey RM, Fernandez AR, Fillingim RB, *et al*. Pain treatment for older adults during prehospital emergency care: Variations by patient gender and pain severity. J Pain 2013;14(9):966–74. doi:10.1016/j.jpain.2013.03.014
27. Anderson AB, Grazal CF, Balazs GC, Potter BK, Dickens JF, Forsberg JA. Can predictive modeling tools identify patients at high risk of prolonged opioid use after ACL reconstruction? Clin Orthop Relat Res 2020;478(7):00-1618. doi:10.1097/ CORR.0000000000001251

Address for Correspondence:

Dr Ayesha Saeed, Senior Registrar, Department of Emergency Medicine, HITEC-IMS, Taxila, Pakistan. **Cell:** +92-333-5974092
Email: ayesha73saeed@gmail.com

Received: 27 Jan 2026 **Reviewed:** 20 Mar 2026 **Accepted:** 31 Mar 2026

Contribution of Authors:

Authors approved the draft and are accountable in ensuring that questions related to accuracy or integrity of the work are duly investigated and resolved.

AS: Conceptualization, drafting initial manuscript, final proofreading
MMAB: Ideation, first drafting, final proofreading
ZR: Ideation, first drafting, final proofreading
ZS: Development and design of AI models, proof reading

Conflict of Interest: No conflict of interest is declared, **Funding:** None received from any agency