Feasibility of Using Machine Learning for Clinical Decision Support to Optimize Transfusion Practices in Trauma Care

By

Michael R. Kolesnikov, MSN

A DISSERTATION

Presented to the Department of Medical Informatics and Clinical Epidemiology and the Oregon Health & Science University School of Medicine

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Health and Clinical Informatics

Dissertation Defense Date: **April 24, 2025**

APRIL 2025

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the PhD dissertation of

Michael R. Kolesnikov, MSN

has been approved.

Vishnu Mohan, MD, MBI – Advisor/Mentor

Karen Eden, PhD – Chair

Steven Bedrick, PhD – Member, ML expert

Chris Hoekstra, DPT, MBI, PhD – Member, qualitative research expert

---

Laszlo Kiraly, MD – Member, trauma clinical expert

Nicole Weiskopf, PhD - Examiner

Table of Contents

Dedication

To my family: my wife, Yana, my daughter, Anais, and my son, Alan.
I could not have done this without your love and support.

"Love is not love
Which alters when it alteration finds,
Or bends with the remover to remove:
O no! it is an ever-fixed mark
That looks on tempests and is never shaken..."
— William Shakespeare, Sonnet 116

Acknowledgements

I would like to express my deepest gratitude to my adviser, Dr. Vishnu Mohan, for his unwavering guidance and support in so many aspects of my life, both professional and personal.

I am also sincerely grateful to my dissertation committee members: Drs. Steven Bedrick, Karen Eden, Laszlo Kiraly, Chris Hoekstra, and Nicole Weiskopf for generously sharing their expertise and for their continued encouragement throughout my research journey.

Special thanks go to the Trauma Service team, including Adrian Ramos and Amanda Staudt, for helping me adjust my clinical schedule to accommodate the demands of graduate study, and to Raina Villanueva and Rebecca Auer for their instrumental help in securing critical resources that enabled this research.

This was a data-intensive project, and I greatly appreciate the responsiveness and efficiency of Dawn Brand and Siran Abtin, who handled my numerous data requests with patience and precision.

I am also profoundly thankful to the OHSU Department of Radiology, and especially to Dr. Alexander Guimaraes and Guy Maresh, for their assistance in obtaining chest and pelvic X-rays—an essential component of this work.

To the faculty and staff of the Department of Medical Informatics and Clinical Epidemiology, thank you for your ongoing support, collaboration, and mentorship. Your commitment made these seven years not only educational, but genuinely enjoyable.

Finally, I extend my deepest thanks to my family, especially my wife and children, who made many sacrifices during my pursuit of a PhD. Your love, support, and belief in my dream carried me through each phase of this journey.

Disclaimer

The research reported in this publication used computational infrastructure supported by the Office of Research Infrastructure Programs, Office of the Director, of the National Institutes of Health under Award Number S10OD034224. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Chapter 1: Introduction

Early trauma research during World War I laid the foundation for our understanding of blood loss and its physiological consequences. Using then-novel methods such as hemoglobin concentration and blood pressure measurements, investigators were able to correlate substantial hemorrhage with changes in circulatory parameters. In one of the earliest documented efforts, Robertson and Bock observed that a systolic blood pressure below 95 mm Hg corresponded to an estimated blood volume reduction of approximately 30%, while a pressure of 80 mm Hg or lower was associated with a reduction of 40% or more—thresholds that today align with what is classified as Class III and Class IV hypovolemic shock, respectively[1].

The discovery of the four human blood types by Karl Landsteiner in 1901[2,3], followed by Ludwik Hirszfeld's work in 1904—including the development of a transfusion compatibility chart[4]—paved the way for a scientific and systematic approach to blood transfusion. These foundational insights made it possible to safely administer blood to patients in acute hypovolemic shock, transforming transfusion from a risky, empirical intervention into a lifesaving, evidence-based therapy.

Today, the rising complexity of trauma care has highlighted the critical need for innovative tools to support clinical decision-making, particularly in the management of blood transfusions. Trauma-related hemorrhagic shock remains a leading cause of mortality in patients under the age of 44, accounting for nearly 50,000 deaths annually in United States[5] and half of all deaths within the first 24 hours of hospital admission[6]. Effective transfusion practices can significantly impact survival outcomes, yet existing methods for determining the necessity and volume of blood transfusions often rely on subjective judgment or simplistic scoring systems, which may lead to suboptimal outcomes[6].

Advancements in machine learning (ML) present an opportunity to revolutionize decision-making in trauma care. ML algorithms have the potential to process complex, multimodal datasets, identify patterns, and provide evidence-based predictions to guide clinicians. Specifically, ML can aid in determining ICU admissions and determine mortality in adult ICU using a logistic regression model, based on age, sex, time of the day[7]. Also, some models show the ability to predict the need for transfusion, initiating massive transfusion protocols (MTPs) with greater sensitivity than the existing scores, such as Assessment of Blood Consumption (ABC) and Revised Assessment of Bleeding and Transfusion (RABT) scoring systems[8] . These capabilities can reduce waste, minimize adverse reactions, and improve overall patient outcomes.

This dissertation explores the feasibility of applying ML techniques to optimize transfusion practices in trauma care. By leveraging data from electronic health records (EHR), imaging studies, and physiological parameters, this work aims to develop robust clinical decision support (CDS) tools. These tools will not only enhance the accuracy of transfusion decisions but also address the gaps in existing scoring systems and the limitations of manual approaches.

Problem Statement

Despite the availability of massive transfusion protocols and existing scoring systems, several challenges persist in trauma care:

1. **Delayed or Inaccurate Transfusion Decisions:** Current methods often fail to promptly identify patients requiring transfusions, resulting in delays that can exacerbate tissue ischemia and increase mortality risk[9].

2. **Over-Resuscitation:** Excessive transfusion practices lead to increased risks of complications, including acute lung injury, infection, and multi-organ failure, while also straining blood bank resources[10].
3. **Lack of Practical and Usable Tools:** Many existing scoring systems are impractical for real-time use due to their reliance on variables unavailable in emergency settings or their complexity in high-stress environments[11].

Research Aims

This dissertation seeks to address these challenges through four primary aims:

Aim 1: Exploration of Clinical Decision-Making: understand the decision-making process related to blood transfusion with following sub-aims:

- SA 1A: how the clinical decision-making related to identifying patients in need of MT is done
- SA 1B: What information/variables are used in this decision
- SA 1C: Evaluate the statistical significance of variables used or available at the time of decision making

Aim 2: Build ML models to identify patients in need of transfusion:

- Design and evaluate ML models, including Multilayer Perceptron (MLP), convolutional neural networks (CNNs), and recurrent neural networks (RNNs), to predict transfusion needs and optimize resource allocation.
- Integrate multimodal data, such as physiological signals, imaging studies, and EHR-derived variables, and use it in a Fusion model to identify patients in need of transfusion.

Aim 3: Build ML models predicting admissions and/or blood demand from system's level:

- Use environmental variables, such as weather data, duration of daylight, and day of the week to build an ML model capable of predicting trauma admissions and/or blood demand from an institutional perspective.

Aim 4: Evaluation of models for accuracy of prediction:

- Assess the effectiveness of developed tools using metrics such as sensitivity, specificity, and area under the receiver operating characteristic curve (AUROC).

Significance of Study

Optimizing blood transfusion practices through ML-based CDS tools has the potential to transform trauma care. These tools can:

- Improve survival rates by enabling timely and precise transfusion decisions.
- Reduce complications associated with over-transfusion.
- Alleviate the burden on blood banks through efficient resource utilization and proper planning.

Moreover, this research contributes to the broader field of biomedical informatics by demonstrating the applicability of ML in high-stakes, time-sensitive clinical scenarios. By addressing the challenges of current transfusion practices, this work aims to set a new standard for data-driven decision support in trauma care.

Chapter 2: Background

Hypovolemic shock and blood transfusion

The history of blood transfusion can be broadly divided into three conceptual periods: romantic, experimental, and scientific. During the romantic period, blood was often viewed as a mystical substance with supernatural properties. In Euripides' *Medea* (431 BCE), for instance, the replacement of an old man's blood with that of a younger man is said to restore youth and vigor[12]. In Shakespeare’s *King Lear* (1606), blood symbolized kinship and loyalty, believed to forge deep familial bonds[13].

Echoes of these beliefs persisted into the 20th century. When in 1919 Herszfeld published an epidemiological study of prevalence of blood types across the nations[14], in 1930s Nazi propaganda, hijacked this research to justify that certain blood types were linked to supposed degeneracy or superiority. More recently, blood-type diets have claimed—without scientific basis[15,16]—that nutritional needs and weight loss strategies should be tailored according to blood type.

The experimental period of transfusion began in 1667, when Jean-Baptiste Denys in France and Richard Lower in England independently attempted blood transfusion using calves’, lamb’s, and sheep’s blood in mentally ill patients, influenced by romantic-era beliefs that blood could cure mental illness, as reviewed by Giangrande[17]. While these patients survived, other experiments resulted in fatal outcomes and led to a ban on transfusion research for over a century. In the 19th century, however, transfusions were again attempted for a wide range of ailments from anemia and tuberculosis to melancholy and "female hysteria"[18], though with inconsistent success and scientific understanding.

The scientific era began in earnest in the 20th century, following landmark discoveries by Karl Landsteiner (1901)[2,3] and Ludwik Hirszfeld (1904), who identified the ABO blood groups and developed the first compatibility charts. Also, an important step was the invention of the citrate method of transfusion in 1915 by Richard Lewisohn, which allowed blood to be anticoagulated using sodium citrate and prevented clot formation during transfusion[19]. These advances made transfusions significantly safer. There were multiple advancements in blood storage and transfusion medicine following 1915. These included the development of techniques to separate plasma, platelets, and red blood cells for storage and transportation[17], enabling more targeted and efficient use of blood products[20,21]. The ability to transfuse large quantities of separated blood components necessitated the creation of massive transfusion protocols (MTPs)[22] and decision-support systems ranging from clinical scores to algorithmic alerts to recognize when these protocols should be activated. In a separate study, we demonstrated that, when meticulously and appropriately applied, MTPs can achieve mortality outcomes equivalent to those of whole blood transfusion[23]. As the understanding of hemorrhagic shock evolved, blood transfusion transitioned from a speculative intervention to a foundational, life-saving component of modern trauma care.

Hemorrhage remains the leading cause of preventable death in trauma patients, with the greatest risk occurring within the first six hours of injury[24]. Despite advances in trauma systems and resuscitation techniques, including the development of damage control resuscitation and the implementation of MTPs, mortality from hemorrhagic shock remains substantial[24-26].Timely identification of patients who require early blood transfusion is essential to reduce mortality, yet traditional clinical scoring systems, such as the Assessment of Blood Consumption (ABC) score and the Trauma-Associated

Severe Hemorrhage (TASH) score, have notable limitations in both sensitivity and specificity[5,9,22,27].

Massive transfusion has historically been defined as the administration of ten or more units of red blood cells (RBCs) within 24 hours[6]. However, more recent literature proposes the Critical Administration Threshold (CAT)[27], defined as the transfusion of three or more units of RBCs within the first hour, as a more timely and sensitive marker for identifying patients in need of aggressive resuscitation[27]. Unlike the classic massive transfusion definition, CAT reduces survival bias and better correlates with early hemorrhage severity and mortality risk[27].

Numerous studies have proposed clinical prediction models to identify patients at risk for massive transfusion. However, many of these models rely on static, manually entered parameters and fail to incorporate continuously updated data or advanced analytic techniques[5]. Moreover, activation of MTPs based on these models has been shown to result in both under-triage, where patients in need are missed, and over-triage, which can lead to blood product waste and strain on hospital resources[28].

Artificial intelligence and machine learning

Artificial intelligence (AI) involves the development of systems capable of perceiving their environment, reasoning, and taking actions to achieve defined objectives. A central concept in AI is the “rational agent,” which acts within its environment to maximize the likelihood of achieving a goal[29]. In surgical practice, artificial intelligence beyond the scope of machine learning has been applied through techniques such as genetic algorithms to optimize operating room scheduling and resource allocation. One such example from 2020 demonstrated the use of a hybrid genetic algorithm to improve operating room efficiency and throughput[30].

Machine learning (ML), a subfield of AI, can be broadly categorized into supervised and unsupervised learning: while supervised models require labeled data to learn mappings between inputs and outputs, unsupervised models discover hidden patterns or relationships without predefined labels. In contrast to supervised models, which require known outputs, unsupervised models are ideal for exploratory tasks such as clustering or association rule discovery. A notable example comes from Xiang et al., who used association rule mining, an unsupervised learning method, to analyze nearly 880,000 transfusion transactions across four hospital blood banks[31]. Their model identified non-obvious patterns of blood product wastage, such as increased loss during evening shifts or at smaller sites, enabling targeted interventions for quality improvement. This study highlights how unsupervised learning can uncover hidden inefficiencies in blood product handling and inform data-driven policy changes in transfusion medicine.

Supervised ML includes a wide range of algorithms that learn from labeled datasets, where each input is paired with a known outcome. The goal is to train a model that can accurately predict the correct label for new, unseen inputs. For example, Mitterecker et al. applied various ML models, including neural networks, logistic regression, random forests, and gradient boosting to admission hemoglobin levels, diagnosis, age, and additional variables to predict transfusion needs in non-trauma patients. Their models achieved an exceptionally high predictive performance with ROC-AUC values ranging from 0.963 to 0.966[28]. Similarly, recurrent neural networks (RNNs) have been employed to forecast platelet demand in hospitals, resulting in more efficient blood product utilization and estimated cost savings of approximately $250,000 per year[32].

Recent work has demonstrated the potential of machine learning (ML) models to improve prediction accuracy for massive transfusion requirements. These models can integrate complex, multimodal data, including vital signs, laboratory values, imaging findings, and EHR data to generate real-time, patient-specific risk predictions[5]. For example, the BloodNav-MLM developed by Benjamin et al. sought to identify patients at risk of CAT+ status (defined as receiving ≥3 units of RBCs within 1 hour of arrival) by stratifying model inputs into four tiers based on clinical availability: Tier 1 (on arrival), Tier 2 (early adjuncts like Focused Assessment with sonography in Trauma (FAST)), Tier 3 (rapid labs), and Tier 4 (comprehensive labs)[5]. They tested several machine learning algorithms—including logistic regression, random forests, k-nearest neighbors, and CatBoost—using a tree-based optimization pipeline in Python. After hyperparameter tuning with Optuna and adjusting for class imbalance, model performance was evaluated via AUC, F1 score, and Matthews Correlation Coefficient. Each model was benchmarked against the ABC score, and performance improved incrementally with each tier of added data. SHAP values and calibration curves were also used to assess interpretability and reliability of predictions.

Fusion models have emerged as an effective architecture to integrate multimodal clinical data, such as structured tabular inputs and medical images. A recent review identified 17 studies utilizing data fusion in medicine, with 4 of them implementing joint fusion strategies[33]. Joint fusion propagates loss back to the modality-specific feature extractors, thereby optimizing the entire model during training and enhancing predictive accuracy. For this dissertation, joint fusion was selected for its ability to capture cross-modal interactions in an end-to-end trainable framework.

Additionally, DenseNet-121, a convolutional neural network architecture, has previously been used by Hashmi et al. to classify pneumonia on chest radiographs. Their approach incorporated data augmentation techniques to address class imbalance and improve model generalizability[34].

Despite the promise of ML in transfusion prediction, clinical implementation remains limited. Barriers include variability in data availability across institutions, lack of integration with clinical workflows, and clinician trust in algorithmic recommendations. Nonetheless, the growing body of literature supports ML as a viable approach to augment clinical decision-making and improve transfusion practices in trauma care[35].

Model acceptance in clinical settings, however, depends not only on technical performance but also on clinician trust and the availability of supporting infrastructure. Trust in ML models is influenced by explainability, integration with workflow, and perceived reliability[35]. This dissertation seeks to build on this momentum by developing, validating, and evaluating ML-based decision support tools tailored for real-time transfusion guidance.

Chapter 3: Factors Impacting Trauma Bay Transfusion: A Mixed Method Study

Background

Although TBI is the leading cause of traumatic death overall[36], bleeding secondary to trauma is the leading cause of preventable death in the trauma patient population[26] and general population between ages of 1 and 44[9]. Additionally, massive blood loss accounts for half of hospital deaths in the first 24 hours[6]. Hence, timely recognition of hypovolemic shock in an incoming patient is one of the most important tasks the trauma team faces daily. There were 3-4 million patients in the United States who received blood transfusions in 2013[37] and approximately 11 million units of Red Blood Cells (RBC) were transfused in 2017[38]. Massive Transfusion Protocols (MTP) has been developed for proper administration of blood components during significant blood loss and have been associated with improved survival in injured solders and civilian population[6]. Currently there are approximately 24 different scoring systems which serve[9] as Clinical Decision Support (CDS) tools that help to identify a need for a massive blood transfusion in trauma patients[22]; however, they are seldom used in clinical practice, because of impracticality. Frequently the variables requested by the scoring system are not available in the first hour, and often they are not recognized by clinicians as important at all (such as weight and height of the patient). Also, the relationship between variables, such as age and systolic blood pressure, is ignored by these scoring tools. Even the simplest scoring systems used to determine a need for Massive Transfusion (MT), such as Shock Index (SI) call for a division of a double-digit (heart rate) by a triple-digit number (systolic blood pressure), which is difficult to carry out in high stress situations. Therefore, instead of calculating this index precisely, trauma surgeons often just directly compare pulse and systolic blood pressure and act when the values approach each other. This mixed-method study investigates decision-making related to blood transfusion by attending trauma surgeons for the purpose of finding which variables correlate with the amount of blood transfused. This analysis can potentially aid trauma centers in better preparing for incoming patients and presents an opportunity to optimize the clinical decision-making process.

Methods

Qualitative Analysis

This was a mixed-methodology study. First, we performed a qualitative study, then we performed a regression correlation analysis using OHSU data collected in Epic to evaluate variables against blood transfusion volumes. The qualitative part was a phenomenological study related to decision making pertinent to blood transfusion during trauma activation and consisted of an online questionnaire and semi-structured interviews. The participants were chosen by a convenience sampling: current or recently retired attending trauma surgeons currently or in the recent past working in OHSU, who were routinely making decisions regarding blood transfusion in the trauma bay. Then, the specific variables discovered in the qualitative part were evaluated by using regression-correlation analysis.

Instrument Development

The 12-item questionnaire was developed specifically for this study to better understand how trauma attendings make transfusion decisions in the trauma bay. It was shaped by a targeted literature review on clinical decision-making, informal observations of trauma resuscitations, and discussions with trauma surgeons and members of my dissertation advisory committee (DAC). The instrument was designed to elicit both structured and open-ended responses, reflecting the multifaceted nature of clinical decision-making under uncertainty.

The goal was to create a tool that captures both the structured and intuitive elements of decision-making under pressure. It was designed to explore how clinicians rely on heuristic reasoning, such as the mechanism of injury, gestalt impressions, or vital sign thresholds, as well as more analytical approaches, including the use of formal transfusion scoring systems like the ABC score, Shock Index, Emergency Transfusion Score (ETS), and Trauma Associated Severe Hemorrhage Score (TASHS). In addition, the questionnaire aimed to uncover perceived gaps in information, such as what clinicians wish they had access to during resuscitation, and to encourage reflection on past decisions, particularly instances where transfusion choices were later reconsidered.

The questionnaire included 12 items, which was enough to cover key areas, but short enough to respect the limited time clinicians have. Skip logic helped to reduce burden, and a mix of multiple-choice, ranking, and open-ended questions allowed both numerical analysis and free-text insights.

Before distributing the survey, I asked my DAC, including a trauma attending, internal medicine physician, and a qualitative research expert for feedback. They helped to refine the wording and ensured the questions were clear, relevant, and not overly technical. Because the tool was designed for exploratory insight rather than scale validation, I didn't conduct formal psychometric testing. Still, the items were grounded in decision-making theory and reflected real-world experience. Please see appendix 1 for the questionnaire.

To complement the questionnaire findings and deepen our understanding of clinical reasoning in transfusion decision-making, we developed an 11-item semi-structured interview guide. The guide was informed by the themes emerging from the survey responses, as well as the literature review on blood transfusion. It was designed to explore how trauma attendings identify the need for transfusion, what clinical cues or data variables they rely upon, or wished they had at the time the decision was made. Like the questionnaire, the interview guide paid extra attention to episodes, when the initial decision to transfuse or not transfuse were contradicting to the patient's course in the trauma bay, trying to unearth potential pitfalls in initial decision-making, which could be compensated by the proposed tools. The eleven items were enough to cover all necessary items without overwhelming the surgeons. Please see the Appendix 2 for complete interview guide.

Data Collection

The participants, via e-mails and social media messengers, were invited to answer the 12-item questionnaire via Qualtrics. There were 13 attending trauma surgeons who participated in the initial questionnaire.

Next, using the above described eleven-items interview guide we conducted semi-structured interviews. The interviews lasted from 15-45 minutes, were recorded in WebEx, transcribed by me, then uploaded into NVIVO.

Coding and Analysis

We employed a variable-oriented, thematic analysis to examine patterns of transfusion decision-making across participating attending surgeons. Interview transcripts were analyzed using a combination of provisional and descriptive coding. Provisional codes were derived from literature review and survey themes, such as use of heuristics, CDS tools, and specific variable commonly used during trauma resuscitation or incorporated into MT prediction tools. Descriptive codes were captured during the interviews, such as use of specific variables, which were not preemptively identified, such as importance of pulse pressure, or role of emotional intelligence during resuscitation.

Variable-oriented strategy was employed for the themes' analysis. Themes were semantically identified and iteratively refined during subsequent cycles. The first cycle used provisional codes to assign quotes to known or expected variables. The second cycle went through the interviews using descriptive coding, and third and fourth cycles were used to consolidate themes. Emphasis was placed at finding not only common, but also unique features, which could be employed in machine learning models in the future. Data saturation was reached after eight interviews, with no new themes emerging in the following three; the interview process was therefore concluded after eleven participants.

Quantitative Analysis

The quantitative data was collected in OHSU using Epic from the period between January 1, 2014, and April, 2024 and consisted of trauma patients 15 years and older, both directly admitted after a trauma activation and transferred from another facilities. The simple regression correlation analysis and logistic regression was done between variables of interest and amount of blood transfused using RStudio 2024.12.0[39], statistical R language version 4.4.2 (2024-10-31)[40], additional packages included packman[41], RefManageR[42], rbibutils[43], xlsx[44], janitor[45], purrr[46], broom[47], readxl[48], readr[49], reshape2[50], psych[51], lubridate[52], tidyverse[53], gtsummary[54], flextable[55], viridis[56], ggplot2[57], tidycmprsk[58], moderndive[59], survminer[60], aod[61], chron[62], sjPlot[63], jtools[64], gridExtra[65], openxlsx[66], data.table[67], cardx[68], dplyr[69], gt[70], webshot[71], and officer[72].

Results:

There were 13 attending trauma surgeons who completed the questionnaire and 11 participated in the interviews, with 8 surgeons participating in both. All participants were active or recently retired trauma surgeons. The variables identified in the qualitative process were statistically tested against a 10-years' worth of retrospective observational patient data from the same level 1 trauma center, where the surgeons were currently practicing. The patient dataset consisted of 33,824 patients, of whom 3,453 received blood transfusion and 30,371 did not. The patient data is summarized in the tables below.

*****Table 1: Population Demographics by Blood Transfusion Status*****

Characteristic	N	NOT Transfused N = 30,371[1]	Transfused N = 3,453[1]	p-value[2]
Age	33,824	47.55 (26.55, 68.61)	54.06 (32.72, 71.02)	<0.001
Gender	31,369			0.4
Female		9,578 / 27,916 (34%)	1,218 / 3,453 (35%)	
Male		18,330 / 27,916 (66%)	2,235 / 3,453 (65%)	
Unknown		8 / 27,916 (<0.1%)	0 / 3,453 (0%)	
Missing		2,455	0	
ISS	33,697	8.00 (2.00, 13.00)	17.00 (9.00, 26.00)	<0.001
Missing		116	11	
Mechanism of Injury	33,692			<0.001
Blunt		27,872 / 30,248 (92%)	3,006 / 3,444 (87%)	
Penetrating		2,376 / 30,248 (7.9%)	438 / 3,444 (13%)	
Missing		123	9	

Characteristic	N	NOT Transfused N = 30,371[1]	Transfused N = 3,453[1]	p-value[2]
Mortality	31,306			<0.001
Alive		26,045 / 27,853 (94%)	2,832 / 3,453 (82%)	
Deceased		1,808 / 27,853 (6.5%)	621 / 3,453 (18%)	
Missing		2,518	0	

[1]Median (Q1, Q3); n / N (%)

[2]Wilcoxon rank sum test; Fisher's exact test; Pearson's Chi-squared test

*****Table 2: Mechanism of Injury by Blood Transfusion Status*****

Characteristic	N	NOT Transfused N = 30,371[1]	Transfused N = 3,453[1]	p-value[2]
Misc. Blunt	33,824	154 / 30,371 (0.5%)	7 / 3,453 (0.2%)	0.014
Cyclist	33,824	333 / 30,371 (1.1%)	18 / 3,453 (0.5%)	0.002
Fall	33,824	1,432 / 30,371 (4.7%)	122 / 3,453 (3.5%)	0.002
Ground-Level Fall	33,824	446 / 30,371 (1.5%)	41 / 3,453 (1.2%)	0.2
GSW	33,824	95 / 30,371 (0.3%)	29 / 3,453 (0.8%)	<0.001
Motorcycle Collision	33,824	264 / 30,371 (0.9%)	42 / 3,453 (1.2%)	0.041
Motor Vehicle Collision	33,824	1,184 / 30,371 (3.9%)	133 / 3,453 (3.9%)	0.9

Characteristic	N	NOT Transfused N = 30,371[1]	Transfused N = 3,453[1]	p-value[2]
Pedestrian	33,824	274 / 30,371 (0.9%)	65 / 3,453 (1.9%)	<0.001
Misc. Penetrating	33,824	182 / 30,371 (0.6%)	30 / 3,453 (0.9%)	0.057
Misc. Trauma	33,824	381 / 30,371 (1.3%)	18 / 3,453 (0.5%)	<0.001

[1]n / N (%)

[2]Pearson's Chi-squared test

Table 3: Vital Signs by Blood Transfusion Status

Characteristic	N	NOT Transfused N = 30,371[1]	Transfused N = 3,453[1]	p-value[2]
Heart Rate	30,887	88.00 (75.00, 102.00)	94.00 (78.00, 111.00)	<0.001
Missing		2,908	29	
Resp Rate	3,369	3.00 (1.00, 9.00)	7.00 (2.00, 14.00)	<0.001
Missing		28,160	2,295	
Systolic BP	30,757	139.00 (124.00, 155.00)	127.00 (110.00, 145.00)	<0.001
Missing		3,013	54	
Diastolic BP	30,757	85.00 (74.00, 97.00)	79.00 (65.00, 95.00)	<0.001
Missing		3,013	54	
Pulse Pressure	30,757	53.00 (41.00, 66.00)	44.00 (31.00, 59.00)	<0.001
Missing		3,013	54	
GCS	31,804	15.00 (14.00, 15.00)	15.00 (11.00, 15.00)	<0.001
Missing		1,785	235	
Shock Index	30,644	0.63 (0.51, 0.76)	0.74 (0.59, 0.95)	<0.001
Missing		3,120	60	

Characteristic	N	NOT Transfused N = 30,371[1]	Transfused N = 3,453[1]	p-value[2]

[1]Median (Q1, Q3)

[2]Wilcoxon rank sum test

Table 4: Labs and X-rays by Blood Transfusion Status

Characteristic	N	NOT Transfused N = 30,371[1]	Transfused N = 3,453[1]	p-value[2]
Hemoglobin	29,553	13.40 (12.20, 14.50)	11.40 (9.30, 13.00)	<0.001
Missing		4,209	62	
Hematocrit	29,564	39.80 (36.30, 42.80)	34.10 (28.30, 38.80)	<0.001
Missing		4,202	58	
Lactate	23,958	1.60 (1.00, 2.40)	2.40 (1.50, 3.90)	<0.001
Missing		8,921	945	
Base Excess	26,962	1.60 (1.00, 2.40)	2.50 (1.50, 4.10)	<0.001
Missing		6,462	400	
CXR taken	32,795	10,374 (35%)	1,981 (61%)	<0.001
Missing		822	207	
Pelvic X-ray taken	32,795	3,552 (12%)	909 (28%)	<0.001
Missing		822	207	
Both X-rays taken	32,795	2,412 (8.2%)	685 (21%)	<0.001
Missing		822	207	

[1]Median (Q1, Q3); n (%)

[2]Wilcoxon rank sum test; Pearson's Chi-squared test

Use of CDS tools:

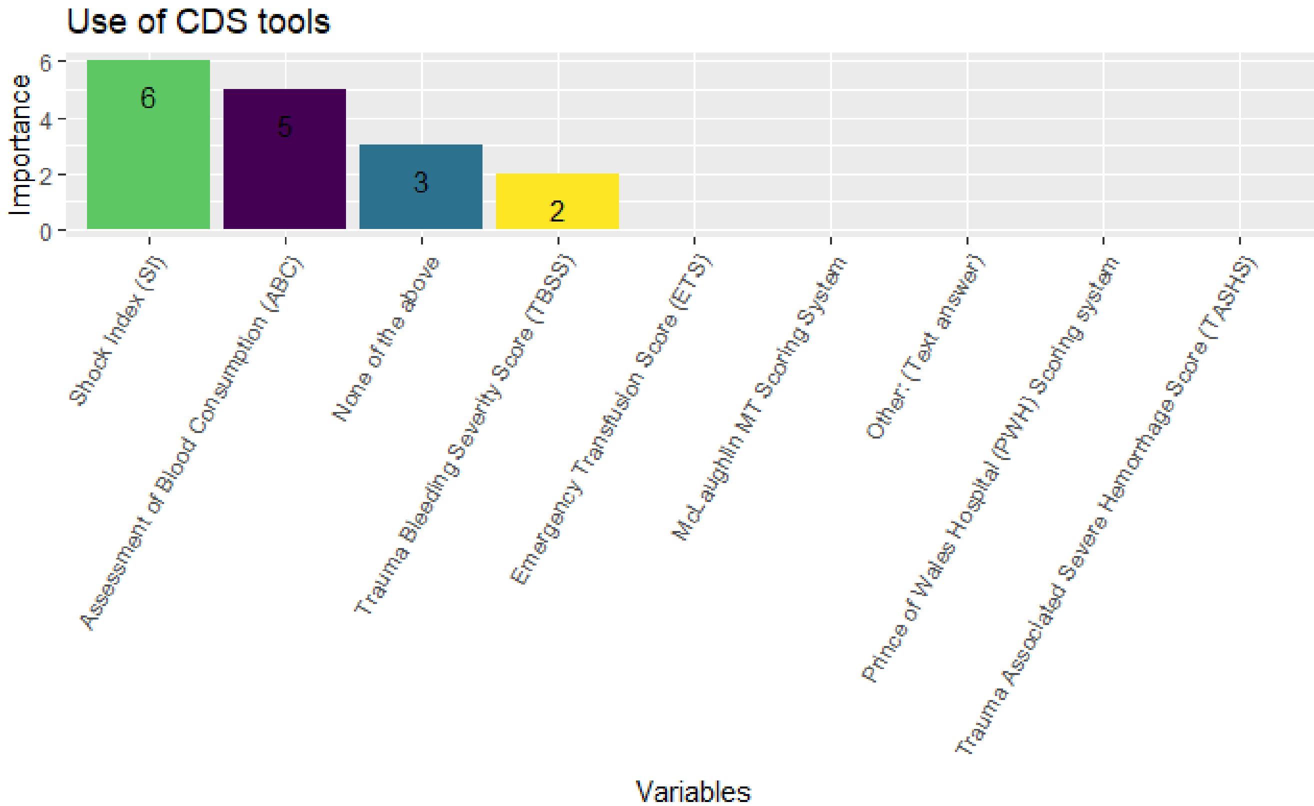


Figure 1: Use of CDS tools by trauma attendings as reported via Qualtrics questionnaire

One of the questions in the initial questionnaire was whether surgeons use heuristic model of decision making, rely solely on CDS, or use a combination of those two. None of the participants relied solely on a CDS tool, most participants (69.23%) chose a heuristic model, but 30.77% said they use a combination of heuristics and a CDS tool. Among the CDS tools 37.5% said they used Shock Index (SI) and Assessment of Blood Consumption (ABC) was a close second with 31.25% participants using it. However, during our interviews one surgeon said they occasionally had a dedicated resource to calculate ABC score for all hypotensive patients, while the rest of surgeons said they do not rely on any specific CDS tools during the resuscitation for decision making, but they look at the same data points utilized in those tools:

"We had the trauma research coordinators who would actually calculate ABC score for all the hypotensive patients… but we still rely on the actual vitals for bleeding: hypotension… we would just transfuse empirically. So yeah, sometimes we used ABC score, but I think that was mostly to get you in tune to use the specific data [points] to decide…" DM

"For me, it's driven by the vital signs and mechanism of injury, how the patient appears in front of me. So, if they're hypotensive and tachycardic, you could [say I use vital signs], but I typically don't. I don't use the shock index, but I look and see if their heart rate is higher than the systolic blood pressure, then that's a good indication of the poor shock index." PV

"There are like a half a dozen tools that exist to try to predict the need for the transfusion, right? And I think we just don't really use any of them… I don't remember the last time I calculated ABC score." MC

The ranking of variables:

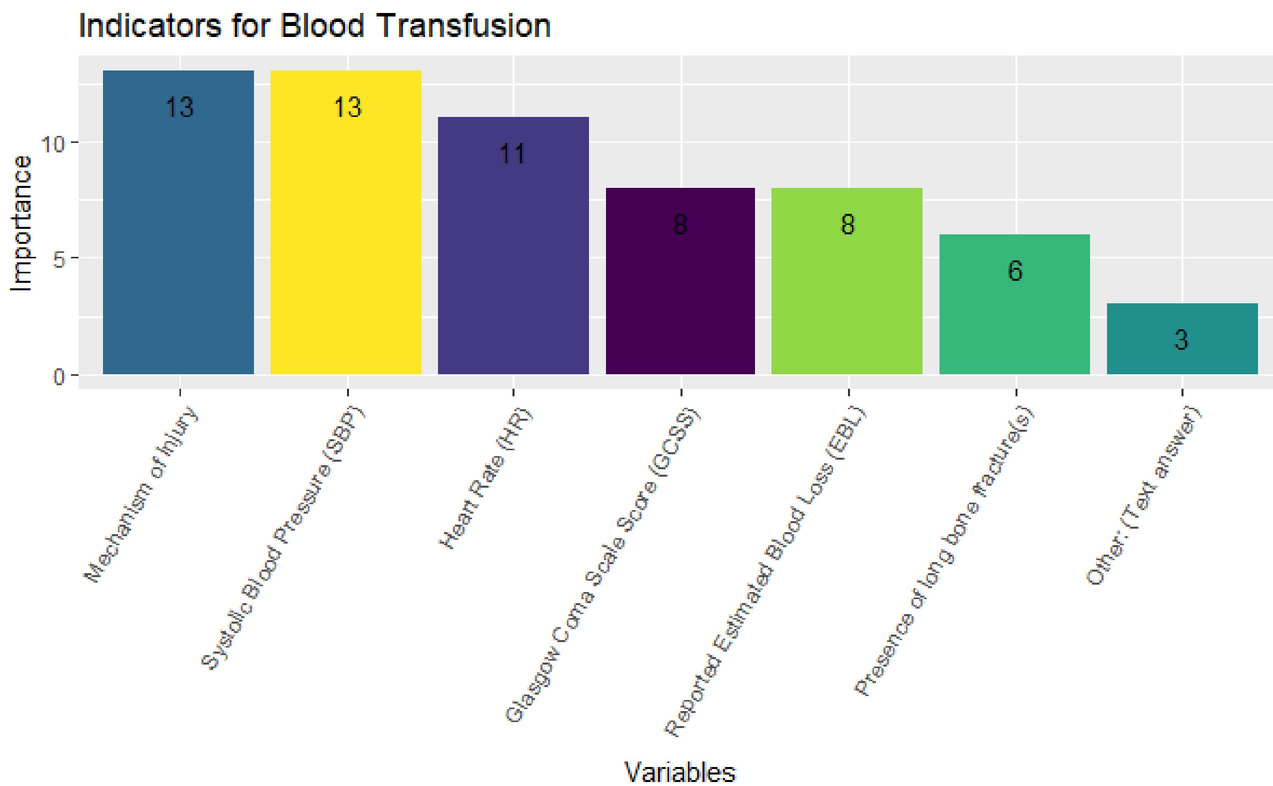


Figure 2: Indicators for transfusion

In our on-line questionnaire, all trauma surgeons were asked what variables they are using when deciding whether to transfuse a patient. The graph above shows the frequency of selecting each variable.

Mechanism of injury

The mechanism of injury was emphasized by all participants as one of two most informative data points employed in early identification of need of a blood transfusion:

"Question: and with [trauma activation] page you typically get the information about mechanism of injury. How important is that?

Answer: It's the second most important thing. If it's a low-risk mechanism my concern or thought that they might need a blood transfusion is going to be lower, but if they have vital signs that suggest shock, even with a low-risk mechanism, I'm concerned." KB

While ABC score gives a point for penetrating injury only, all surgeons emphasized that significant blunt injury also raised their level of alertness, as energy transfer could cause significant internal injuries and bleeding:

"I think the mechanism does make a difference: …it's the energy it delivers to the patient's body, so if they're hit by a train [as appose to a car], that's more likely to be associated with more fractures and therefore more hemorrhage, more liver, spleen, and kidney injury…" RM

It also appears our current perception of penetrating injury, as more likely to require a blood transfusion, causes a surprise for a trauma surgeon, when a blunt injury patient requires a transfusion:

"Question: Did you ever have a patient who you didn't expect to require a blood transfusion, but upon arrival you say, "this patient definitely needs a transfusion"?

Answer: I would say it's probably more likely in a blunt trauma situation [because] I'm maybe more inclined to think a penetrating or stab injury needs [a blood transfusion]." DM

Hypotension

Not surprisingly, hypotension was one of two main indicators of clinical status that requires a blood transfusion.

"…if somebody has a hypotension and I have a moderate to high clinical suspicion for bleeding, I will give them blood… I would certainly [give blood to] somebody who's hypotensive and tachycardic." HH

Our regression correlation analysis confirms the importance of hypotension for transfusion decision in trauma patients:

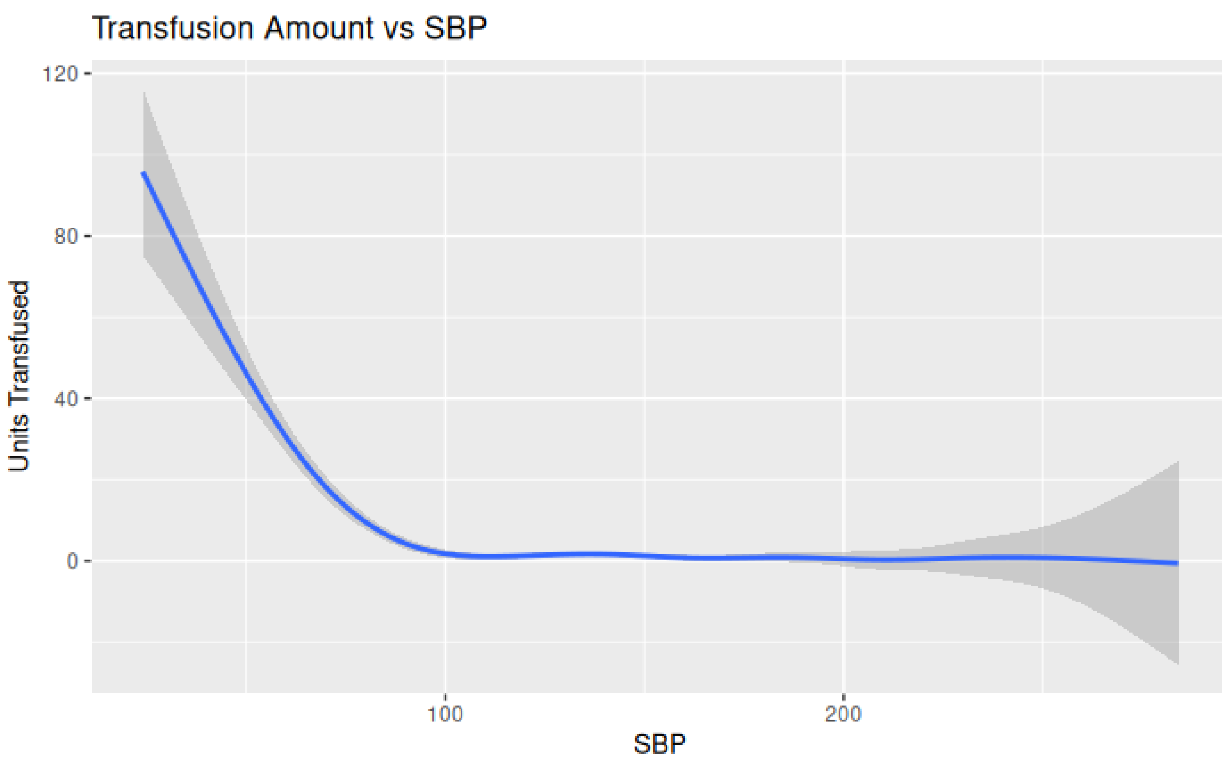


Figure 3Transfusion Amount vs SBP

Patients' age

When asked directly, most of the surgeons initially said they did not consider a patient's age as a factor in the diagnostic process, however, upon further probing they revealed that age is an important co-variate that helps to evaluate blood pressure in the context of clinical picture:

"An 80-year-old person, who has hypertension, blood pressure of 90 is much more concerning to me than somebody who's 18 and has a normal BMI and is otherwise healthy, because they are much closer to what I would guess their normal blood pressure to be than an 80-year-old." HH

Age brings up another dimension: comorbidities and how they affect the patient's presentation:

"The older people will fall and will be hypotensive for a number of reasons that may or may not be bleeding." TH

"In older patients I don't look at the heart rate if they're on a beta blocker" AC

Another take home point is that body reserves and ability to compensate to trauma is diminished at older age:

"Another confounding issue is that elderly people, I think, are more susceptible to less blood loss. So, the kind of the rules of the game change. In the older patients, I think, the blood pressure [is more sensitive and specific] than pulse rate…" RM

Although age does not correlate directly with the amount of blood transfused to patients, it is still an important consideration in medical decision making.

Tachycardia

While tachycardia was named as the third most helpful data point, it was always mentioned as a secondary factor, inferior or conditional to hypotension:

"I think systolic blood pressure was probably number one [factor in deciding on blood transfusion], and then maybe heart rate, but heart rate as, you know, when it's really

high, it can be from drug use, or pain, or anxiety. So, not as often the heart rate [had a significance] to be honest." DM

The graph below shows that the heart rate alone was not a very good indicator in making the decision regarding blood transfusion.

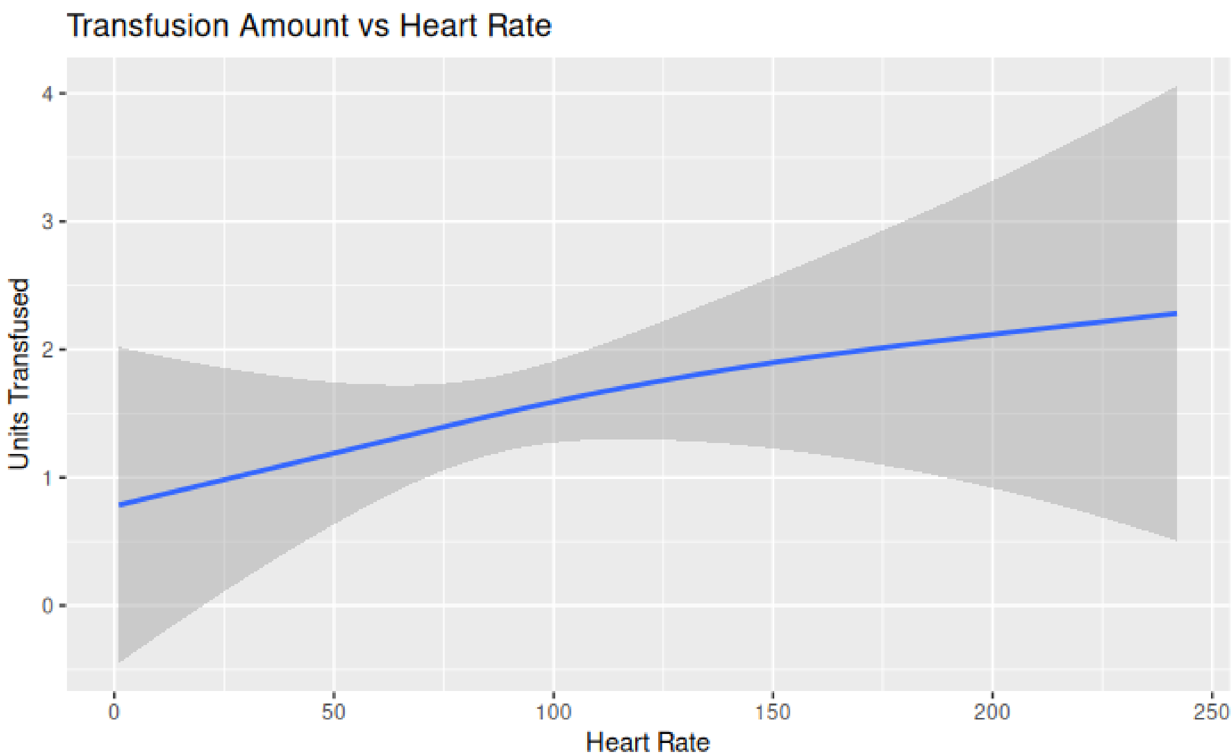


Figure 4Transfusion amount vs Heart Rate

However, the correlation between shock index and units transfused appears to be a better strategy for clinical decision making.

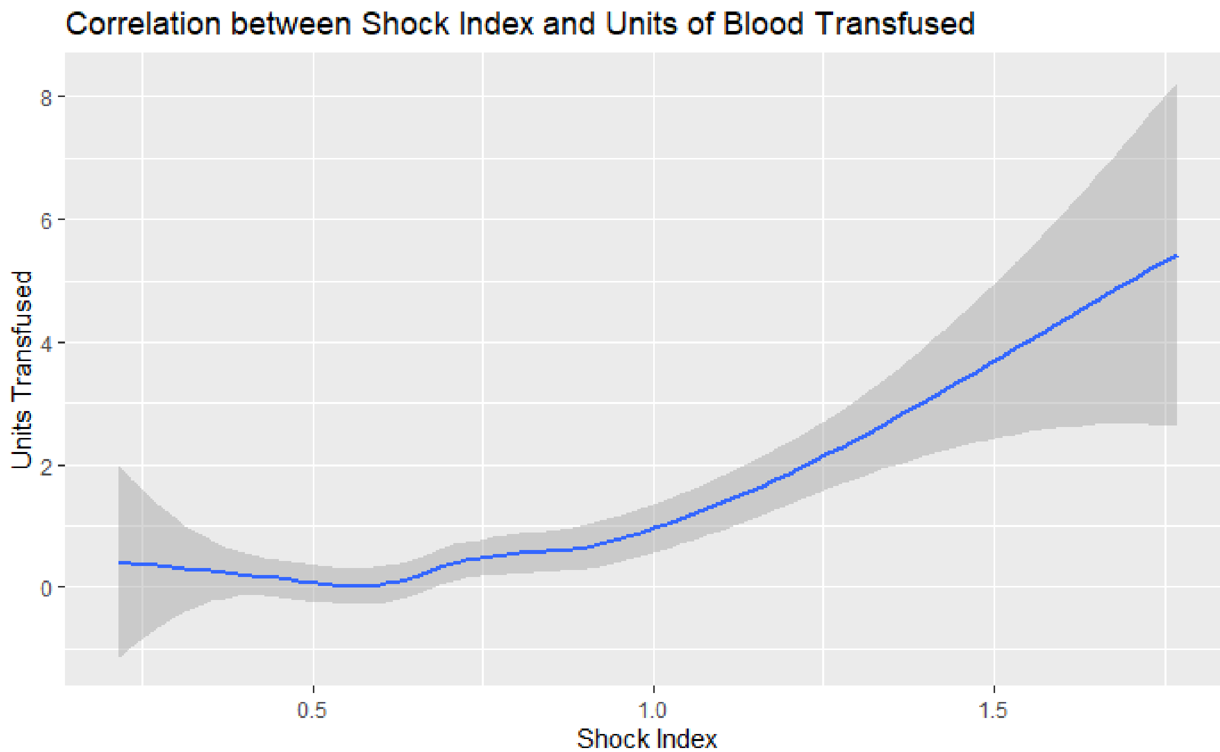


Figure 5: Transfusion amount vs SI

Glasgow Coma Scale Score/Mental status

Altered mental status in trauma patients is often a sign of systemic hypoperfusion. Many surgeons reported low GCS score or diminished mental status as one of the important factors in early identification of shock:

"Yeah, I think [the information from the initial page, such as] blood pressure, pulse pressure, declining GCS help predict [the need for blood transfusion] … [declining GCS score] is a marker of [global hypoperfusion], so, declining GCS depending on mechanism makes me very concerned about [need for a] blood transfusion." MS

In our quantitative analysis, the GCS score appears to have a significant correlation with the amount of blood transfused and we suggest it should be incorporated into decision making.

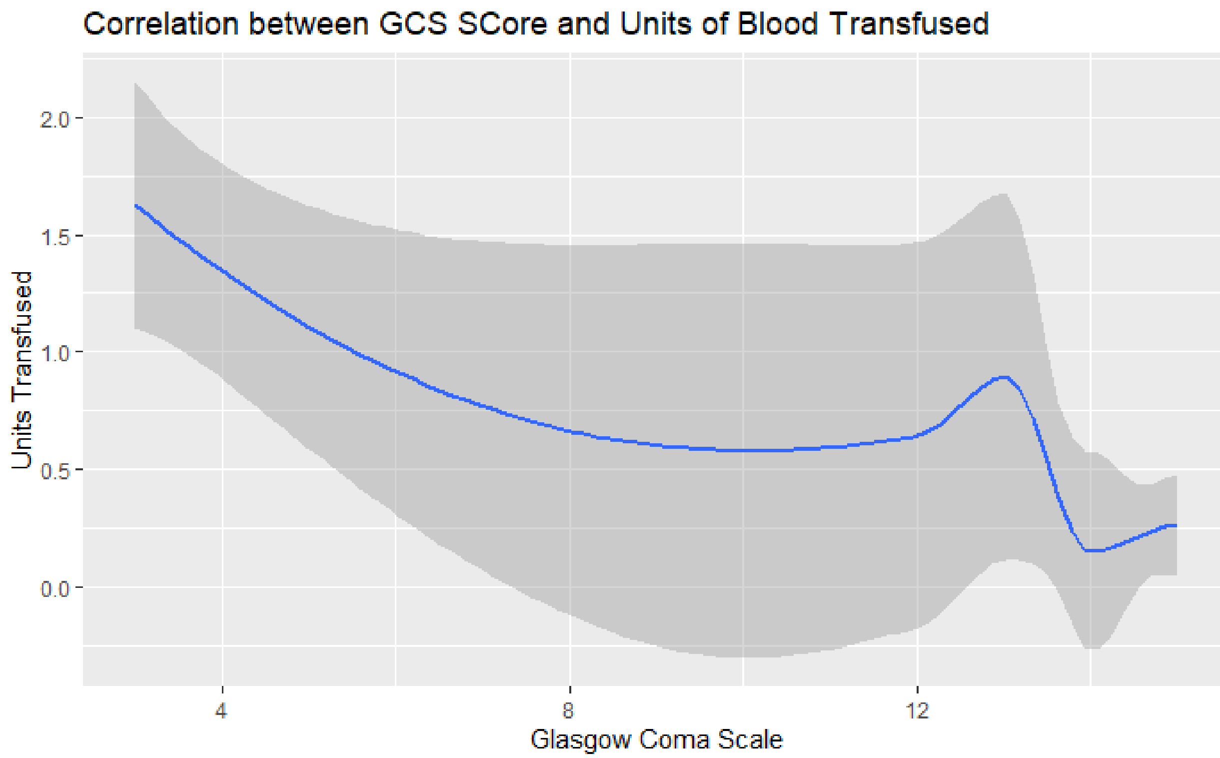


Figure 6: Transfusion amount vs GCS

Pulse pressure

Pulse pressure is known to narrow during hypovolemic shock [73], as a part of compensatory mechanism to decreased blood volume. Some trauma surgeons pay close attention to changes in pulse pressure.

"I would say, [...] a really narrow pulse pressure or really even a wide pulse pressure can be an earlier sign [of hypovolemic shock], so, [...] it might make me think [about blood transfusion] ..." DM.

The graph below shows that narrow pulse pressure (below 37.5) is associated with increased volume of blood transfusion in the first 24 hours, but overall, we consider it as a weak indicator for transfusion.

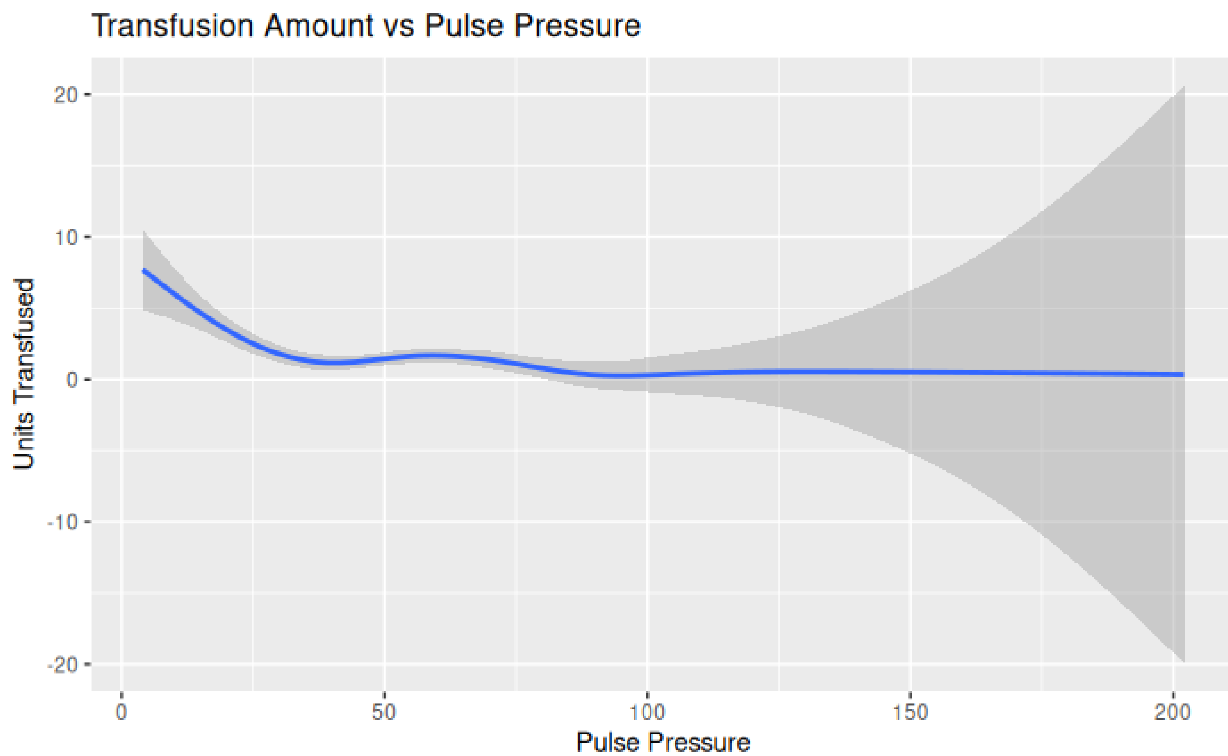


Figure 7: Transfusion amount vs Pulse Pressure

Patient's presentation

While most of the CDS tools are concentrated on getting as much objective or measurable data into the medical decision making, surgeons report their decision making to be more complex and requiring more vigorous approach:

"You know, part of the problem is the way that I assess the patients. The [objective] data itself isn't sufficient to decide how sick the patient is. Blood pressure and heart rate are [important], and it'd be helpful, but it's not the entire clinical picture. If that was all we needed to decide what we were going to do with the patient, sure, but there's more that goes into the assessment..." PV

Many surgeons said they will order blood or make other medical decisions, based on patient's appearance "*pale […] diaphoretic and clammy… in combination with a very bad mechanism." PV*, and they also use emotional intelligence, when assessing patients:

“You can tell how concerned the paramedics are, so that's the first thing.

Then if the patient looks distressed, that's the second thing and that's really before they're at the Trauma Bay… If the patient is really pale... So, [you register] all of that clinical information about the patient [before] you start to get the vital signs and heart rate.” KB

Also, a patient’s appearance plays a significant role to rule out the need for a blood transfusion:

“Patient comes in and they're joking with the prehospital personnel… If the patient has their legs crossed, if the patient has their feet crossed, they don't need blood. They're going to be fine.” KB

Blood test results

When trauma surgeons are in doubt, they are looking for additional information. In these situations, lactate and base deficit were found to be most useful:

“So, lactate update is helpful. Base deficit is helpful. I think hemoglobin hematocrit are actually not as useful in hyperacute patient.” JF

Although most of the trauma surgeons said hemoglobin and hematocrit (H&H) have low sensitivity, as their values are lagging behind the blood loss and onset of hemorrhagic shock, but they have high specificity, so low H&H will call for a transfusion:

“So, if you have a patient, who has a low hemoglobin or a low hematocrit, you are more likely to transfuse. But I wouldn't necessarily not transfuse because the hemoglobin or hematocrit are normal.” JF

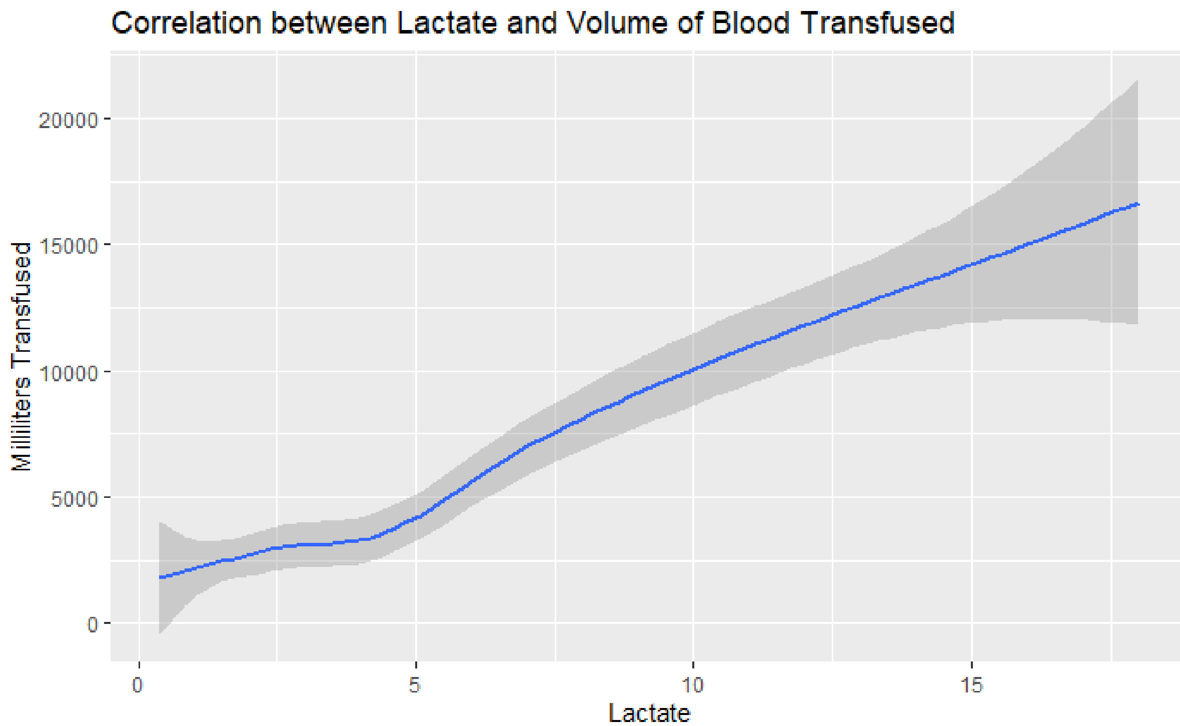


Figure 8: Lactate vs Volume of Transfusion

Surgeon's wish list

Many surgeons reported that initial report of hypotension and tachycardia were frequently inaccurate, so they "*…[wish for greater] reliability of information in the prehospital setting" JF*

"I have to say through the years back, there were a lot of patients who were hypotensive, or reported hypotensive in the field, who weren't hypotensive." RM

As trauma activation comes with a price tag, lower systolic blood pressure contributes to false full trauma activation, which misappropriates resources, increases cost of healthcare, and may also contribute to catastrophic health expenditures, which disproportionately affect people with lower income [74]. To mitigate that and avoid higher level of activation, surgeons asked for regular updates from the field.

"Question: What information from the field could help you to identify the patients in need

of blood transfusion?

Answer: The indicators would be blood pressure, heart rate … and the trajectory of those things, so whether they were declining or improving...” LK

Reports about wound location also got some attention and could be an area for improvement:

“…probably the most vague is when you have a patient with a penetrating injury… I think a… photo [would be] more helpful than somebody trying to describe it in 2 or 3 words.” PV

Many surgeons also reported they take into consideration the interventions that were done in the field and especially what hemodynamic response patient had:

“Question: What helps you to make this prediction [need for a blood transfusion] based on the information from the ambulance?

Answer: If they tell me there's gross blood loss at the scene, and whether the patient responds to the resuscitation prior to the arrival. Or resuscitated, but they are still hypotensive, unable to get a blood pressure, things like that.” AC

Imaging studies

On the use of imaging studies, the surgeons were unanimous to give blood to somewhat unstable patient if the also see a positive FAST or a specific injury on the plain films. A positive FAST had a high specificity, but not as high sensitivity. Hemothorax or open book pelvic fracture increased their inclination to administer blood for an unstable patient.

“So, in imaging studies: a positive FAST in an unstable patient, an open book pelvic fracture in an unstable patient, a large presumed hemothorax in unstable patient will get blood, but in the absence of the hemodynamic criteria I am unlikely to give blood…” JF

A tension pneumothorax or a spinal fracture could reveal hypotension without need for a blood transfusion:

"…certainly, other things that cause hypotension, like a tension pneumothorax would be picked up by a chest x-ray and maybe a reason why they're hypotensive, but don't actually need blood." HH

"… let's say [you found on imaging] a spine fracture. It might lead you to give less blood, because you can attribute the hypotension to neurogenic shock." DM

By the time patient was taken to the CT, the need for blood transfusion has been ruled out and that study typically did not change that decision:

"It is pretty rare that a CT scan will change my mind about whether a patient needs blood." KB

Discussion

Clinical decision-making can be analytical, heuristic, or hybrid, combining the intuitive and analytical approaches, and all three are well represented in transfusion-related decision-making. From the first moments of patient's arrival to the Trauma Bay, the attending trauma surgeons first evaluate patients with non-analytical clues, such as patient appearance, paramedic's level of concern, visible blood, and interventions done in the field, even before they start to collect objective information about level of hypotension relative to tachycardia. This is known as "thin-slicing" or the "augenblick" model of reasoning and was described in 1992 by Ambady and Rosenthal[75]. At this level the surgical team weeds out those patients who do not need blood transfusion and those who are in shock and need it urgently.

Typically, this decision is done with minimal amount of information and represents System 1[76] or "fast thinking", which comes from experience, requires no time for analysis and is a result of residency training and experience as a trauma surgeon. However, this is a highly subjective, biased, and error-prone decision-making process. One way to improve the quality of decisions made at this level would be augmenting System 1 thinking with analytical data through rigorous control of information provided to the decision-makers[75]. Many surgeons have noticed that they either look at the same variables as in certain CDS tools, such as SI, or sometimes are forced to pay attention to those variables by auxiliary personnel, who collect the data to calculate ABC score. One of the sources of errors during non-analytical reasoning is insufficient information[77]. Since System 1 thinking is associated with visual perception[77], in order to improve the decision-making at this stage, we suggest using data visualization in the Trauma Bay, so the pertinent information will be available to the team immediately and visualized. Recent articles point out that System 1 thinking is affected by subconscious biases, such as anchoring, base-rate neglect, and even emotional polarization towards patient[78] and suggest using of checklists to mitigate the effects. Using AI tools could also help to faster and more reliably identify patients in need of blood transfusion, as they have higher sensitivity and specificity. We did member-checking with attending trauma surgeons, and they confirmed that the graph below of two simulated patients would help them in the decision-making process. In this graph we combined and color-coded systolic and diastolic blood pressure, heart rate, GCS, respiration rate and added SI X 100, as presenting shock index was more informative, then presenting just heart rate and systolic blood pressure.

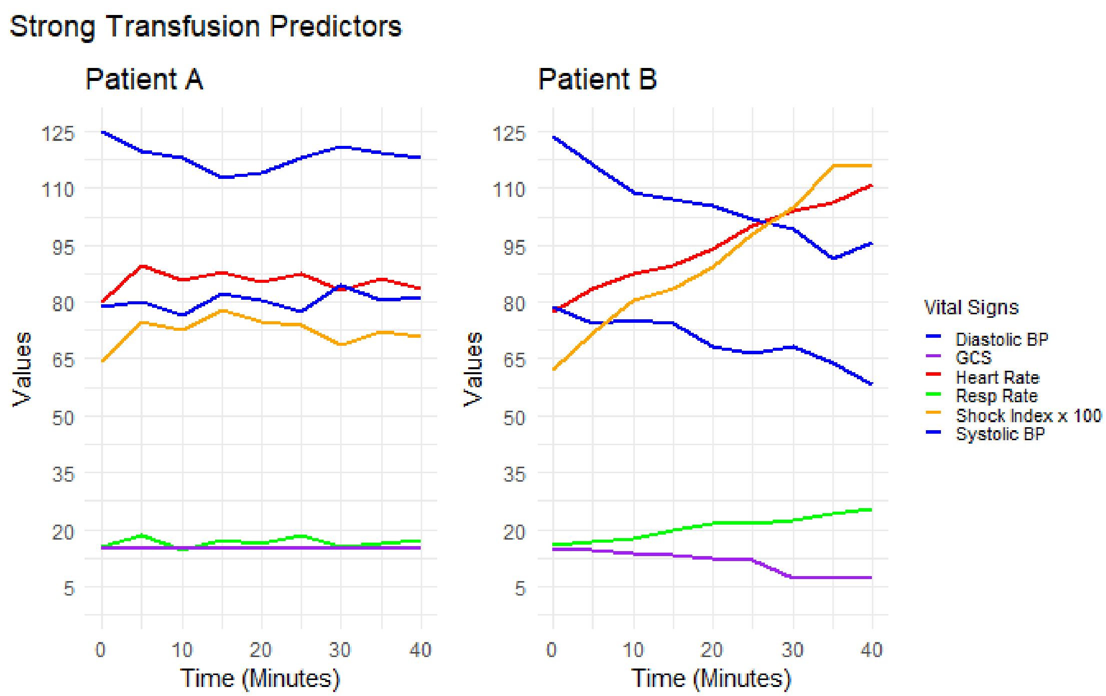


Figure 9: Stable vs Declining patient

The next stage is System 2 thinking process and can be characterized as analytical or "slow thinking". At this level attending surgeons collect objective data, such as vital signs and physical exam, until they reach a conclusion about the need for a blood transfusion and the need for emergent surgical interventions. The main objective of our research was to find out what variables are mostly correlated with decisions regarding blood transfusion and whether there is a room for improvement in medical decision making in the first hour by providing this information even before the patient arrives to the Trauma Bay and whether there is a room for a clinical decision support tool. While interviewing the surgeons and evaluating variables we noted some of the variables to be more informative than others. Perhaps the most informative for the surgeons was not a single data point, but the hemodynamics of the patient's vital signs over time and after interventions, that helped the surgeons to classify if the patient needs none, *"a little bit, or a lotta bit of blood" (MC)* as well as whether the patient responds to the therapeutic intervention and whether that response is sustained.

Conclusion

The decision-making process regarding blood transfusion of the trauma victim that happens in the Trauma Bay often relies on the System 1 thinking process and quality of the decisions may depend on subconscious biases, incomplete information, and looking at suboptimal variables to determine the need of blood transfusion.

This study has certain limitations. Since most surgeons are from one center, the results may be skewed to a certain practice pattern. While blood transfusion was the main endpoint, we cannot confidently say that all blood transfusions were necessary or lifesaving. Surgeons' recall of their own decision making may not be a reflection of their actual practice pattern, so the results might be affected by a recall bias. We also had to resort to convenience sampling, which may not represent the full diversity of the decision-making done in the trauma bay.

Some of the easy fixes include using data visualization and use of checklists to communicate pertinent variables to the whole trauma team during resuscitation in the trauma bay. As recent research shows, digitalized checklists show good performance and time saving[79], which would be really precious in trauma bay.

Among the important variables, we identify mechanism of injury, age, GCS score, respiration rate, systolic blood pressure and heart rate, which were also confirmed as factors playing major role in administration of blood in pre-hospital settings[80]. Tachycardia alone was a poor indicator of blood transfusion, so we recommend using SI. Other important variables were the location of penetrating injuries, lactate, and results of FAST exam.

These variables can be clearly communicated as a graph, so data visualization can take advantage of System 1's activation of visual perception[77], which happens during "thin-slicing"[75] and also present not as a single data point but rather representing the hemodynamic trends of the patient's vital signs over time and patient's response to interventions. Additionally, since ML prediction methods are not inferior to conventional methods[81], and most models achieve maintaining similar specificity to conventional methods[8], we propose that AI tools could assist in more rapidly and reliably identifying patients in need of blood transfusion.

As many trauma surgeons suggested, providing photographic evidence of lacerations and wounds from penetrating injuries by pre-hospital personnel could be another easy fix. This could greatly increase trauma team's understanding of specific underlying anatomical structures being involved and could help receiving surgeons to estimate possible blood loss.

A significantly more challenging intervention would be to provide hemodynamics' updates of vital signs in real time to keep the trauma team prepared for the incoming patient. This task may present a technological challenge as it requires establishing a secure up-to-minute connectivity to deliver the data directly to the receiving team, as well as a mechanism for receiving team to obtain and view this information, while the patient is being transported. Certain areas may not have cellular networks available to reliably transmit this data.

Providing information about interventions done by the prehospital team in real time is another important area for potential improvement. If this could be done simultaneously with the vital signs update, this could help the trauma surgeons to judge whether the patient falls into the category of "responders", especially if they receive blood enroute.

Another potential area for improvement could be providing the level of concern by paramedics on a Likert scale. This could help the trauma surgeons to incorporate professional judgement of prehospital teams into their decision about blood transfusion and further management of the patient.

Finally, we suggest investigating the feasibility of obtaining lactate and base deficit levels in pre-hospital setting and including it in the report. Our regression correlation analysis shows a strong association between the values of lactate and amount of blood transfused.

Chapter 4: Predicting the Need for Massive Blood Transfusion Using Machine Learning

Use of ML for Clinical Decision Support

Introduction

In one of the earliest documented efforts, in 1919, Robertson and Bock observed that a systolic blood pressure below 95 mm Hg corresponded to an estimated blood volume reduction of approximately 30%, while a pressure of 80 mm Hg or lower was associated with a reduction of 40% or more, a thresholds that today align with what is classified as Class III and Class IV hypovolemic shock, respectively[1]. This discovery not only defined the mechanism of hemorrhagic shock but also suggested blood transfusions as treatment for it. Advances in blood storage and component therapy led to the development of Massive Transfusion (MT) protocols, which were designed to improve survival in severely injured trauma patients.

While hemorrhagic shock remains one of the main causes of preventable death[24], there were multiple tools created to reliably identify patients in need of transfusion. Recent reviews have concluded that the optimal initiation, composition, and monitoring of MTs remain unclear[6]. Clinical practice guideline for massive transfusion scoring system, after an extensive literature review, has identified and recommended Trauma Associated Severe Hemorrhage (TASH) score, and Assessment of Blood Consumption (ABC) score as level B (moderate) recommendation for practice, while Shock Index (SI), McLaughlin Score, Prince of Wales Hospital/Rainer (PWH/Rainer), Traumatic Bleeding Severity Score (TBSS), and Vandromme Score as level C recommendation for practice. Emergency Transfusion Score (ETS) was not recommended based on limited ability to predict the need for MT[22]. Another attempt to review 24 scoring system by El-Menyar and co-authors revealed that many scoring systems are complex and require difficult calculations and blood tests results that are not available at the time of the decision-making and overall trauma centers expertise played the major role in identifying the need for MT and was the leading driving force in preventing mortality[11].

Objective

We aimed to develop and evaluate a joint fusion machine learning model that combines clinical data and imaging to predict the need for massive transfusion in trauma patients at the time of arrival. The goal is to leverage data immediately available upon presentation. such as vital signs and point-of-care laboratory results and enhance predictive performance by integrating it with features extracted from chest X-rays.

Machine Learning in Healthcare

Machine learning, a subset of artificial intelligence, involves the use of algorithms that enable computers to learn from and make decisions based on data. In healthcare, ML has shown remarkable promise in diagnosing diseases, predicting patient outcomes, and personalizing treatment plans. The ability of ML models to process vast amounts of data and identify patterns that may be imperceptible to human clinicians is particularly beneficial in high-stakes situations such as MBT.

Previously machine learning and deep learning has been used in surgery for a variety of tasks from dynamic multi-outcome predictions after injury[82], predicting prolonged ICU stay[83], predicting mortality[84-86], detection of bacteremia[87], need for ICU admission in patients with truncal GSW[7], predicting complications of blood transfusion[88], and effect of plasma transfusion on perioperative complications[88].

Recently, there were also several attempts to use machine learning to identify patients in need of transfusion. Mitterecker and coauthors[28] used a wide variety of ML methods, including neural networks (NNs), logistic regression (LR), random forests (RFs)[89], and gradient boosting (GB) trees, using over 131,041 unique patients database, but excluding all patients, who's hemoglobin at admission was missing[28]. Another attempt was done using a machine learning decision tree algorithm: classification and regression tree (CRT) and eXtreme gradient boosting (XGBoost)[90] on a total of 1371 trauma patients, showing it is not inferior to traditional LR method[28]. A very creative and interesting methodology was developed by Benjamin and co-authors, where they have reviewed what variables are available in trauma patients and developed a four-tier system, starting with pre-hospital data and including new data, such as admission vitals, FAST, point of care lab values, and later lab values, as they becomes available, which resulted in a progressively more accurate predictions[5]. Finally, a work by Strickland and co-authors attempting to predict massive transfusion (MT) using only variables available at the time of the initial trauma assessment, using support vector machines (SVMs)[91], naive Bayes, boosting techniques such as XGBoost and AdaBoost[92], and neural networks, reporting 0.55 for sensitivity, 0.83 specificity, 0.14 PPV, and 0.23 NPV[8].

New methods identifying patients in need of MT have been continuously developing. Several methods employed ML to identify MT were taking into consideration the anatomical location of the injury. One of them is a prehospital scoring system developed by Israeli physicians, which takes into account not only mechanism of injury and vital signs, but also anatomical location of the injury and employed a range of ML algorithms, including LR, Lasso regression, RF, and XGBoost[9]. The other was development of a field AI tool to predict MT in GSW population using Dirichlet deep neural network (DNN-IAD), also was taking into consideration GSW anatomical location and patient information available in field[93]. The incorporation of anatomical injury location into ML models was a novel advancement and inspired our consideration of integrating radiologic imaging with tabular clinical data.

Using two different types of data, specifically imaging and electronic health records data, has been successful in medicine using fusion. Following Huang's definition[33], there are three major types of fusion: early, when we are fusing two similar modalities coming from different sources, and process them through the same model, late fusion, when two different modalities go through two different model and the resulting probabilities are combined in the end, and joint fusion, where two different types of data are modalities are going through two different models, but then, instead of going to the activation the features are joint in the fusion layer and only after that they go through the activation[33]. The joint fusion method has an important advantage, as the cost function is propagated all the way to original models through the back propagation method, making final predictions more accurate.

Algorithms Used in Predictive Modeling

After reviewing the available methodologies, we decided to make this experiment in three stages:

1. We employed several ML protocols, MLP, XGBoost, LR, and RF to make predictions based on tabular data and evaluated their performance using ROC-AUC.
2. Next, we used CNN[94], to make predictions based on X-ray images and analyze the performance of CNN on ROC-AUC after experimenting with multiple pre-trained models, we have selected DenseNet-121[95]. It is a well-established architecture, has been previously used for pneumonia detection[34], originally trained on a large dataset of chest X-rays, proved itself in ability to extract complex features from medical images, including subtle patterns in the lungs and thoracic cavity.
3. The interpretability of the DenseNet-121 classification was achieved by using Gradient-weighted Class Activation Mapping (Grad-CAM), a proven method to visualize a model's successes as well as its failures[96]. Previously Grad-CAM was used to assist in classification of pneumonia cases[97], COVID-19 classification on chest-X-rays[97-99], and to enhance explainability of various deep learning methods[99-103].

4. We will use joint fusion to extract features from tabular data using MLP and images using DenseNet-121, joining them together at the features level. Previously an MLP-CNN fusion model has been successfully used in fundamental physics for search for mono-Higgs signals in bb final states[104].
5. We evaluated predictions by employing Youden's index to determine the optimal balance between sensitivity and specificity[105]. Youden's index identifies the optimal threshold for binary diagnostic tests and has been extensively used for this purpose[105,106]. Furthermore, Youden's index and the ROC-AUC are both important and correlated measures of diagnostic accuracy[107,108]. Using both indices jointly offers a more reliable and comprehensive evaluation than relying on a single metric[106,108]. This joint approach is recommended for medical research, particularly in the evaluation and comparison of binary diagnostic outcomes[107,108].

Data Collection and Preprocessing

The success of ML models hinges on the quality and quantity of data. For predicting MT needs, data was collected from trauma registry and electronic health records (EHRs), including patient demographics, vital signs, laboratory results, and chest X-rays. Preprocessing steps involve cleaning the data to remove inconsistencies, obvious errors, removing records with missing values, and normalizing the data to ensure uniformity. Feature selection is crucial to identifying the most relevant variables that influence the need for MT and was addressed by separate mixed-methodology research.

The resulting patient population represents trauma victims for the period between January 01, 2014 and April 15, 2024. It consists of trauma patients 15 years and older, both directly admitted after a trauma activation and transferred from another facilities. 3,453 received blood transfusion and 30,371 who did not. The simple regression correlation analysis and logistic regression was done between variables of interest and amount of blood transfused using RStudio/2024.12.0[39], statistical R language version 4.4.2 (2024-10-31)[40], additional packages included packman[41], RefManageR[42], rbibutils[43], xlsx[44], janitor[45], purrr[46], broom[47], readxl[48], readr[49], reshape2[50], psych[51], lubridate[52], tidyverse[53], gtsummary[70], flextable[55], viridis[56], ggplot2[57], tidycmprsk[58], moderndive[59], survminer[60], aod[61], chron[62], sjPlot[63], jtools[64], gridExtra[65], openxlsx[66], data.table[67], cardx[68], dplyr[69], gt[70], webshot[71], and officer[72].

Results

See Tables 1-4 for Population Demographics, Mechanism of Injury, Vital Signs, Labs and X-rays by Blood Transfusion Status.

Data Handling and Feature Engineering

Feature engineering involves creating new features from existing data to improve model predictions. To improve the models' performance in predicting MT prediction, we constructed the following synthetic variables:

- Abbreviated ABC Score calculations: ABC score gives one point for each of the following conditions: penetrating mechanism of injury, systolic blood pressure under 90 mm HG, heart rate over 120 bpm, and positive results of focused abdominal ultrasound (FAST) exam. Since we did not have the results of FAST exam, we created binary variables for each of the aspects of ABC, except for FAST results. We kept both: the individual synthetic variables and abbreviated ABC score in the models.

- Shock Index (SI): we calculated SI as pulse divided by systolic blood pressure and added it to the list of variables given to the models.

As shown in the preceding tables, the dataset was highly imbalanced. Only 3,246 or 10% out of our 32,795 patient population received any blood product transfusion during first 24 hours, 948 (~3%) received critical administration of transfusion (CAT), 323 (~1%) received CAT6 or early MTP, which was defined as 6 units or more within first 4 hours after admission, and 435 (~1.32%) received true MT, which was defined as 10 or more units within 24 hours after admission. For this study we selected MT group as our class of interest, which made our dataset extremely unbalanced with 98.7% being in the non-MT class.

There are several ways to overcome this challenge: undersample the majority class, which was not acceptable, because it will leave us with a database too small to train a deep learning model, oversample the minority class by copying the existing cases, which will contribute to overfitting, or use Synthetic Minority Oversampling Technique (SMOTE), which creates synthetic instances of minority class within the minority decision domain. This technique has been successfully used in a range of fields, where datasets are often imbalanced, such as fraud detection[109], machine fault classification in electronic engineering[110], in medicine for visual field progression in glaucoma[111], and predicting intracranial hemorrhage expansion in neurosurgery[112].

Naturally, after addressing the class imbalance in our tabular data we had to address it in the chest X-rays dataset. Similarly to tabular data, there are a few options to balance the x-ray dataset. Since convolutional neural networks require large datasets for effective training, undersampling was not feasible; instead, we explored data augmentation techniques. One of them was generating high quality synthetic images using Generative-Adversary Networks (GAN) similarly to what was used for data augmentation for detection of COVID-19 by Schaudt and co-authors, however we quickly realized this will require increasing scope of our research without improving classification performance[113]. Instead, we opted for a conventional augmentation pipeline using random horizontal flipping and small-angle rotations, which have previously been shown to improve model generalization in chest X-ray classification tasks, including those for COVID-19 pneumonia[114,115], and to improve learning for classification of pneumoconiosis[116].

For image preprocessing, all chest X-rays were converted to three-channel format by repeating grayscale channels and resized to 224×224 pixels to match DenseNet-121 input requirements. During training, data augmentation included random horizontal flipping and random rotation within ±10 degrees to simulate real-world variability common in trauma imaging. Finally, all images were normalized using dataset-specific mean (0.1462) and standard deviation (0.2761). Model Training and Validation

Once the data is preprocessed, several ML models are trained using historical patient data, including LR, RF, XGBoost, and MLP. Metrics such as accuracy, precision, recall, and the area under the receiver operating characteristic curve (AUC-ROC) are used to evaluate model effectiveness. And AUC-ROC showed similar performance between these models.

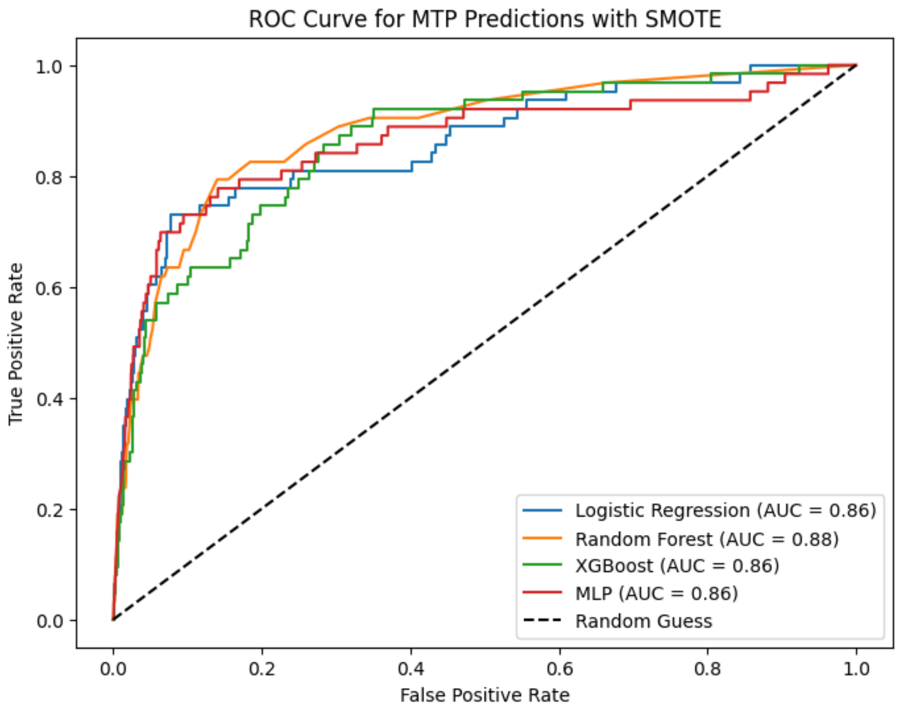


Figure 10: ROC-AUC for MT predictions with SMOTE

This proved that MLP performance is not inferior to other models and we started to experiment with different configurations of pre-trained CNN's. Among the models we looked at ResNet-18 and DenseNet-121, noticing later to perform better on our dataset.

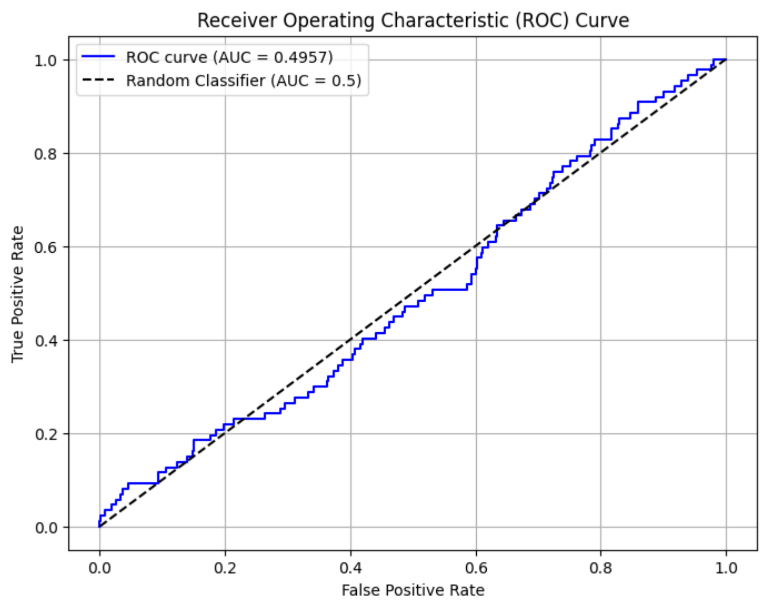


Figure 11: ROC-AUC for ResNet-18 model

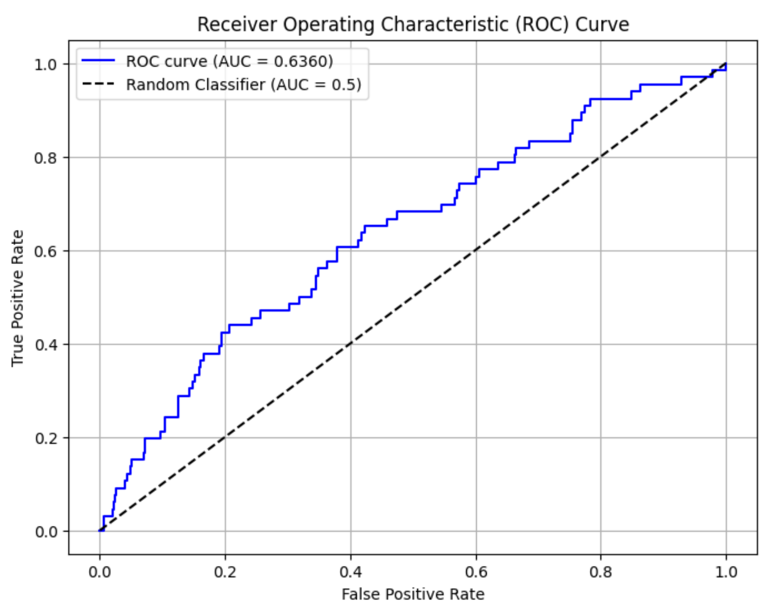


Figure 12: ROC-AUC for DenseNet-121 model

Advantages of the Joint Fusion Approach

The joint fusion architecture offers distinct advantages over early or late fusion models by enabling deeper integration of heterogeneous data types. Unlike early fusion, which requires modality alignment at the input level, or late fusion, which combines outputs after independent processing, joint fusion allows for the extraction of modality-specific features followed by their integration at an intermediate layer. This design enables the model to capture complex, complementary interactions between imaging and clinical data, leading to improved predictive accuracy. Furthermore, joint fusion supports end-to-end training with shared backpropagation across modalities, allowing the optimization process to influence both branches of the model synergistically. In our study, this approach yielded superior performance, particularly when paired with data augmentation and SMOTE, making it a compelling strategy for clinical decision support in trauma care.

The fusion model was implemented using PyTorch Lightning, with key hyperparameters including a learning rate of 0.0001, weight decay of 0.001, and a dropout rate of 0.2 applied at multiple layers to prevent overfitting. Class imbalance was addressed by incorporating a positive class weighting factor of 2.0 in the cross-entropy loss function. The model was optimized using the Adam optimizer.

After training and testing our fusion models, we observed improved generalization and predictive performance when data augmentation was applied, consistent with theoretical expectations. Models trained with augmented data demonstrated thresholds closer to 0.5, suggesting better calibration. However, the gain in sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) was modest.

The graph below shows sensitivity and specificity across thresholds for the fusion model without SMOTE or X-ray augmentation. The optimal threshold determined by Youden's Index (J-threshold) was 0.031, indicating poor calibration and overfitting to the training distribution.

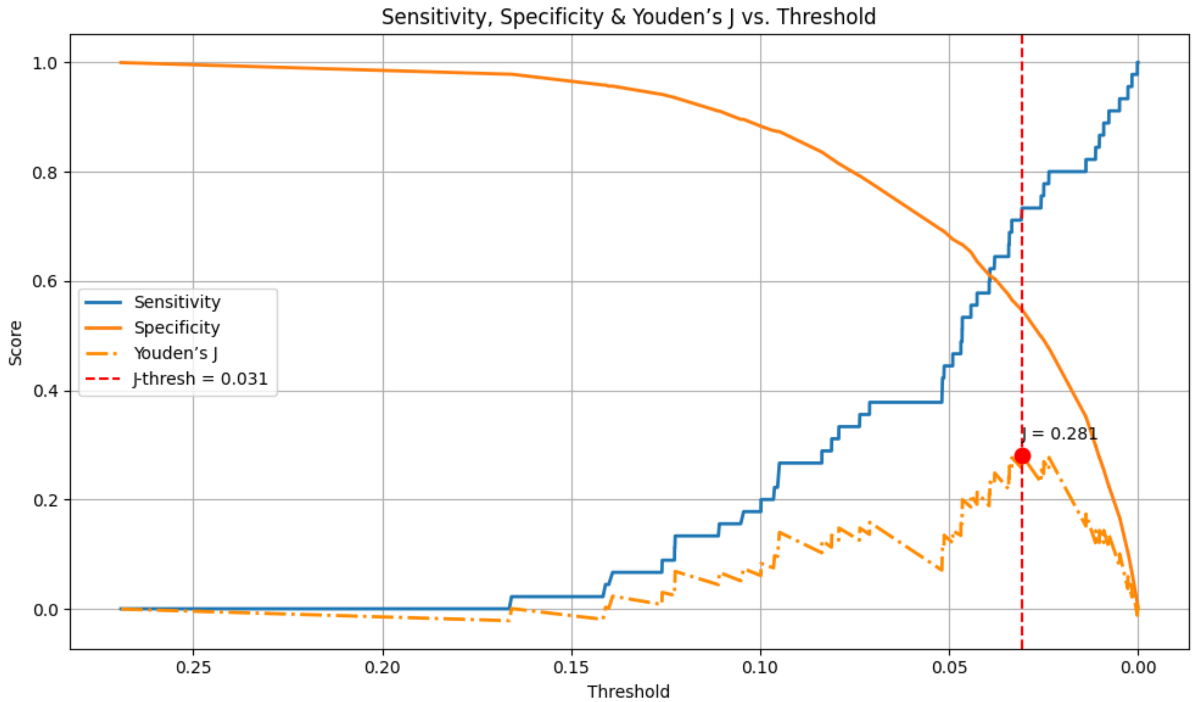


Figure 13: Sensitivity, Specificity, and Youden's J vs Threshold NON-Augmented Model

In contrast, the fusion model trained with both SMOTE and X-ray augmentation achieved a J-threshold of 0.59, suggesting better generalization and a more clinically meaningful threshold.

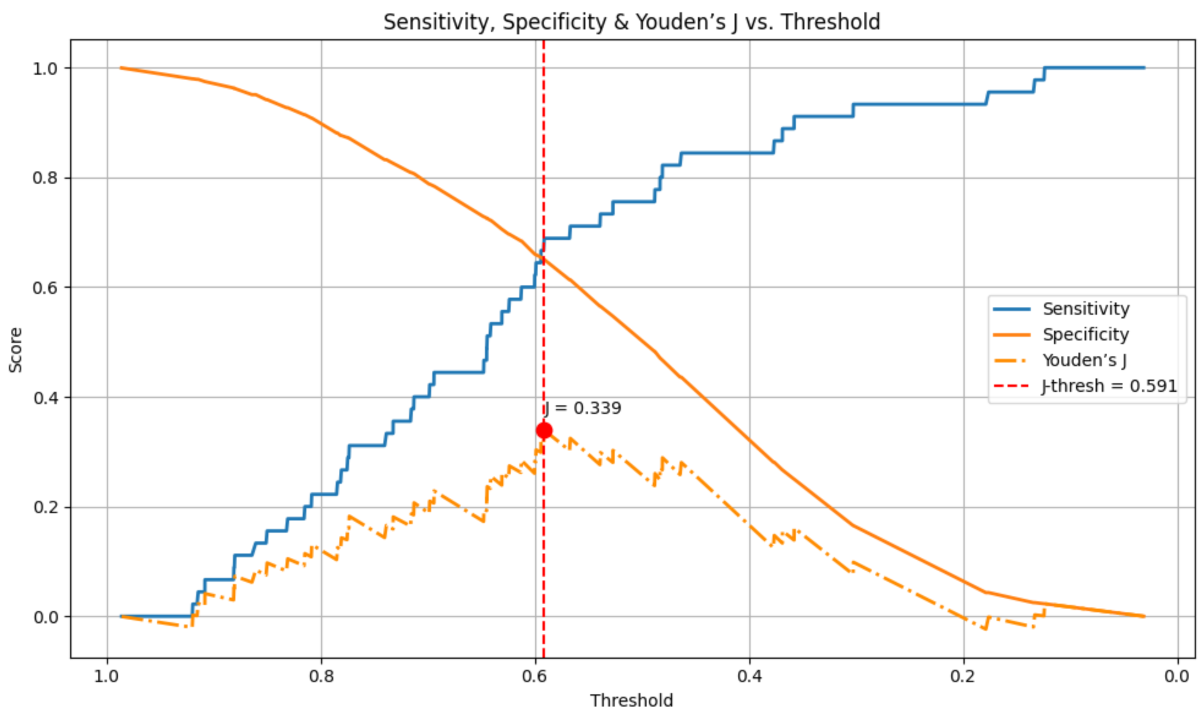


Figure 14: Sensitivity, Specificity, and Youden's J vs Threshold Augmented Model

The following ROC curves illustrate these differences. The augmented fusion model achieved a slightly higher AUC (0.669 vs 0.631), a lower false positive rate (0.35 vs 0.45), and a greater vertical separation from the no-skill classifier (0.339 vs 0.281), all indicative of better classification performance.

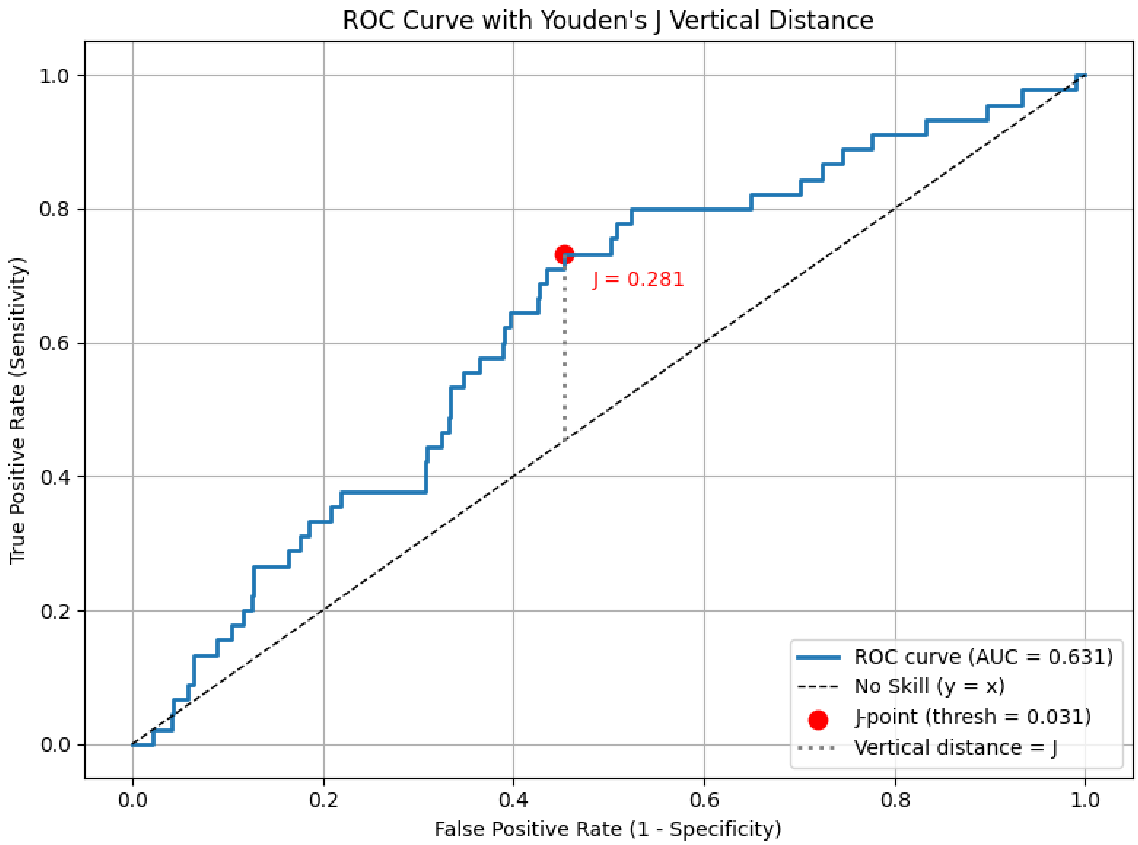


Figure 15: ROC-AUC with Youden J-index for Fusion model without data augmentation

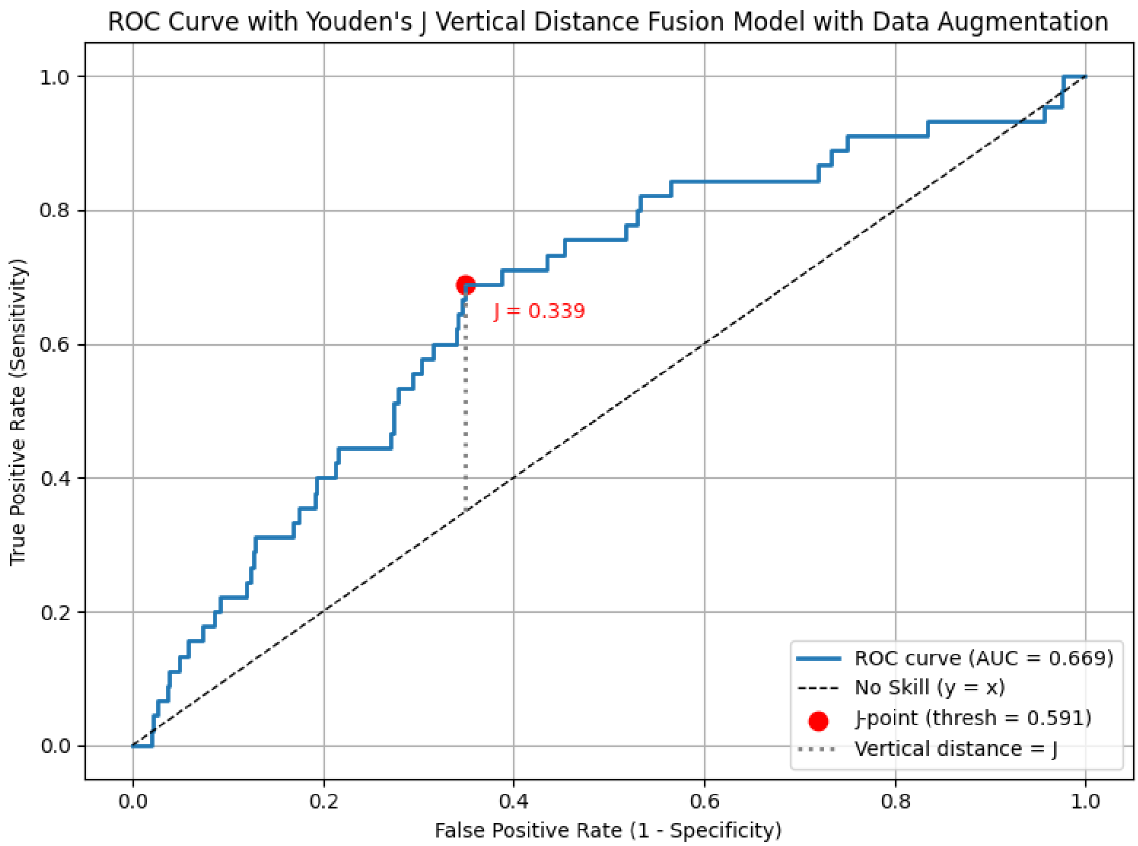


Figure 16: ROC-AUC with Youden J-index for Fusion model with data augmentation

Confusion matrices reinforced this finding, revealing a notably higher false-positive rate in the non-augmented model.

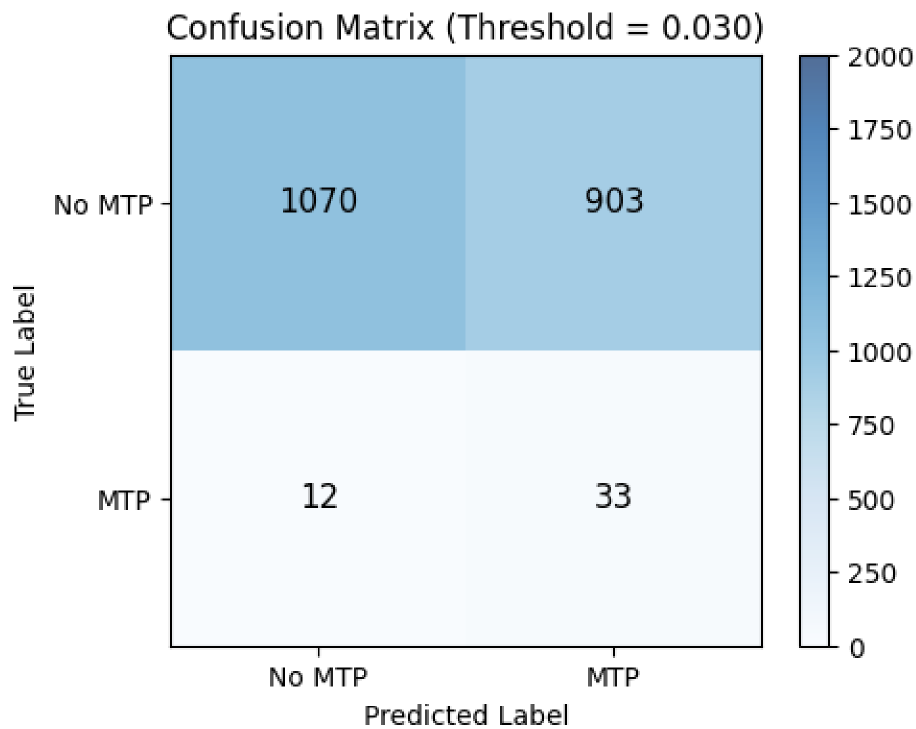


Figure 17: Confusion matrix for the non-augmented model

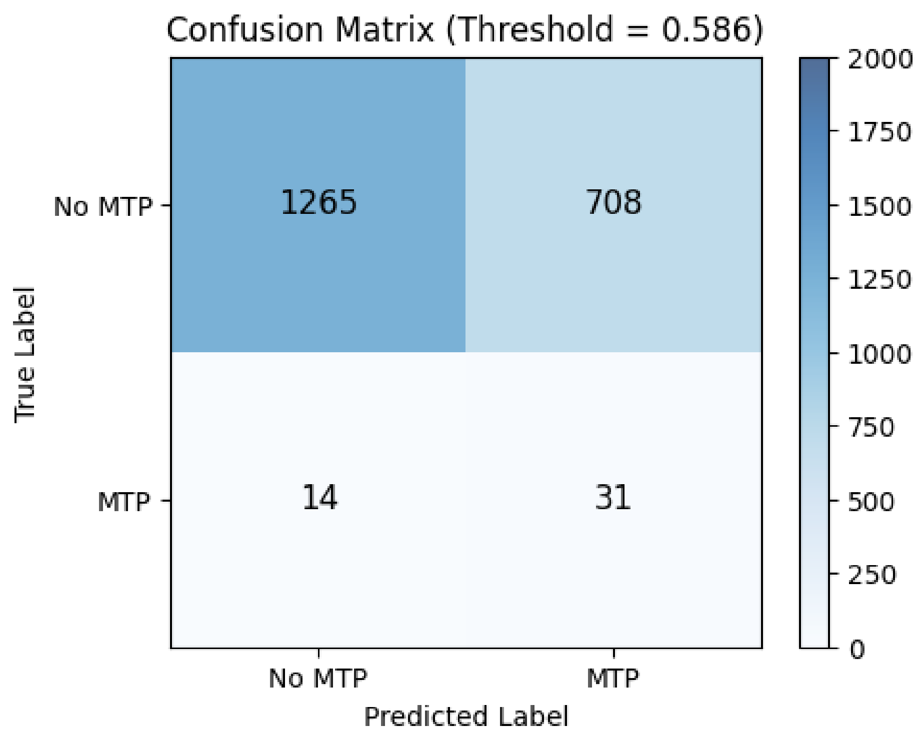


Figure 18: Confusion matrix for the augmented model

Benchmarking Against ABC Score from the PROPPR Trial.

We evaluated six final models: MLP (with and without SMOTE), CNN (with and without X-ray augmentation), and fusion models (with and without both SMOTE and augmentation). The fusion model with both SMOTE and augmentation performed best overall. To contextualize its performance, we benchmarked it against the ABC score results reported in the Pragmatic Randomized Optimal Platelet and Plasma Ratios (PROPPR) trial.

Notably, the PROPPR trial dataset had a class distribution of approximately 57% non-MT and 43% MT, whereas our dataset was heavily imbalanced (98:2), which likely contributed to lower PPV and higher NPV in our models.

	MLP	MLP	CNN	CNN	Fusion	Fusion	ABC Score
	w/o SMOTE	with SMOTE	w/o Augment	with Augment	w/o Augment	with Augment	PROPPR Trial
Sensitivity	0.616	0.256	0.595	0.571	0.719	0.605	0.66
Specificity	0.846	0.966	0.629	0.607	0.705	0.837	0.37-0.45
PPV	0.056	0.1	0.042	0.039	0.101	0.137	0.88-0.90
NPV	0.993	0.989	0.983	0.98111	0.986	0.98	0.13-0.15

Evaluation of results using Grad-CAM

While evaluating the chest x-rays and heatmaps, that were produced by the models, we have noticed that machine may be able to pick-up very subtle ques, that are not picked up by a radiologist on plan chest x-ray. The cases below illustrate this. This is a true negative case of a person, who was assaulted with a metal bar to the face and head, there was no chest injury and no transfusion needed. The model correctly classified him as a true negative, with no notable activation on the Grad-CAM heatmap:

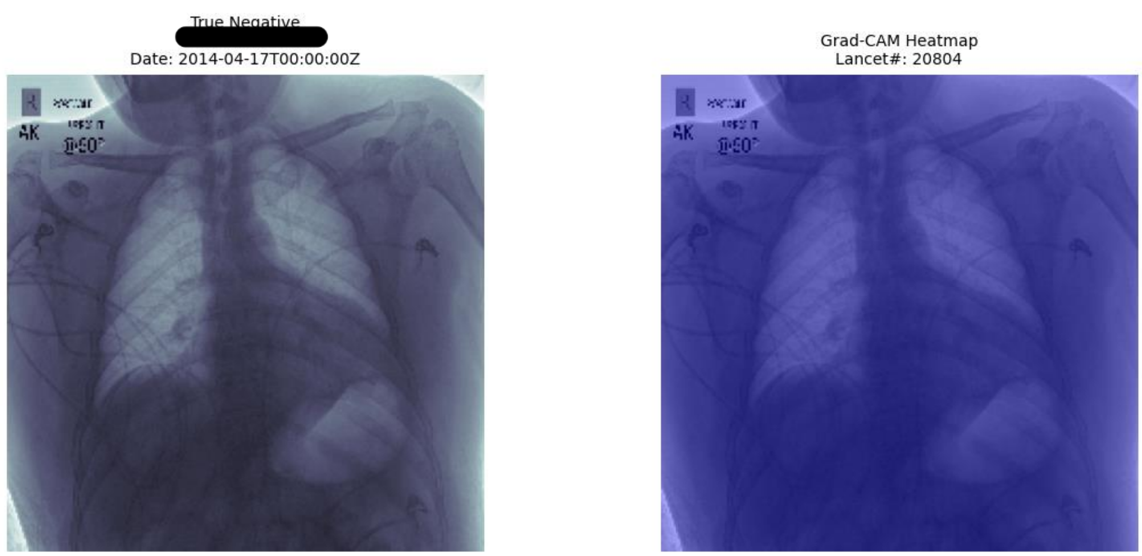


Figure 19: True Negative Case

This is a true positive case: a 58 y.o. gentleman who was pinned under a crane at a construction site, he came in hypotensive, required 2 units of whole blood while in the trauma bay and was taken to the OR for emergency exploratory laparotomy with small bowel resection. As you can see, the machine correctly identifies him as someone having abdominal injuries.

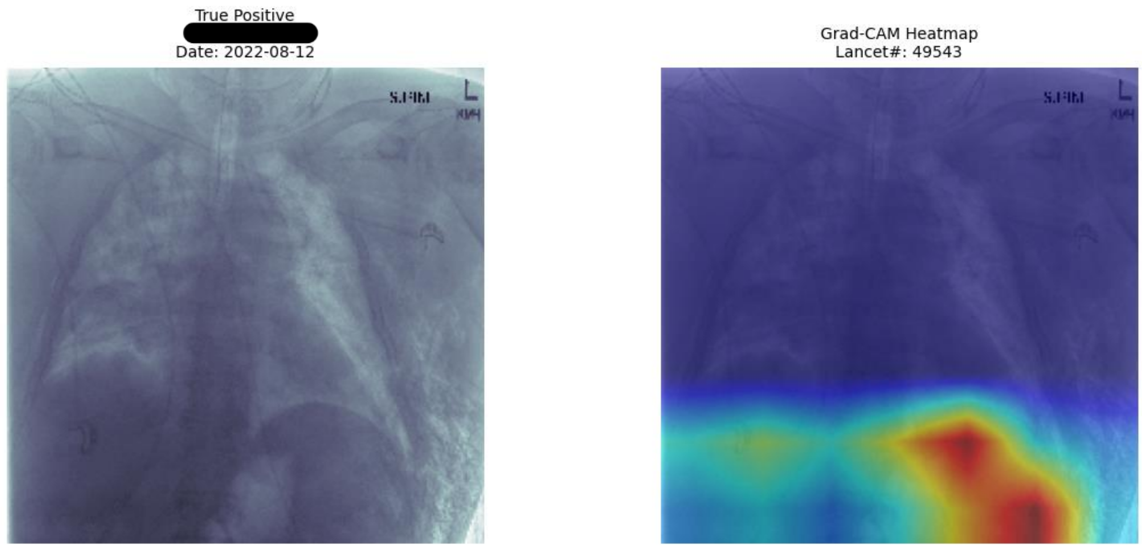


Figure 20: True Positive Case

Next, we have an 84 y.o. male transferred as a full trauma activation after being stomped on the chest by a cow. This patient came in intubated, hypotensive with BP 50/30 and massive transfusion protocol was initiated in the trauma bay. His injuries included multiple bilateral rib fractures, sternal fracture, blunt cardiac injury, and retrosternal hematoma. His Chest X-ray's heatmap correctly identifies most injuries being in the chest and lower third of sternum.

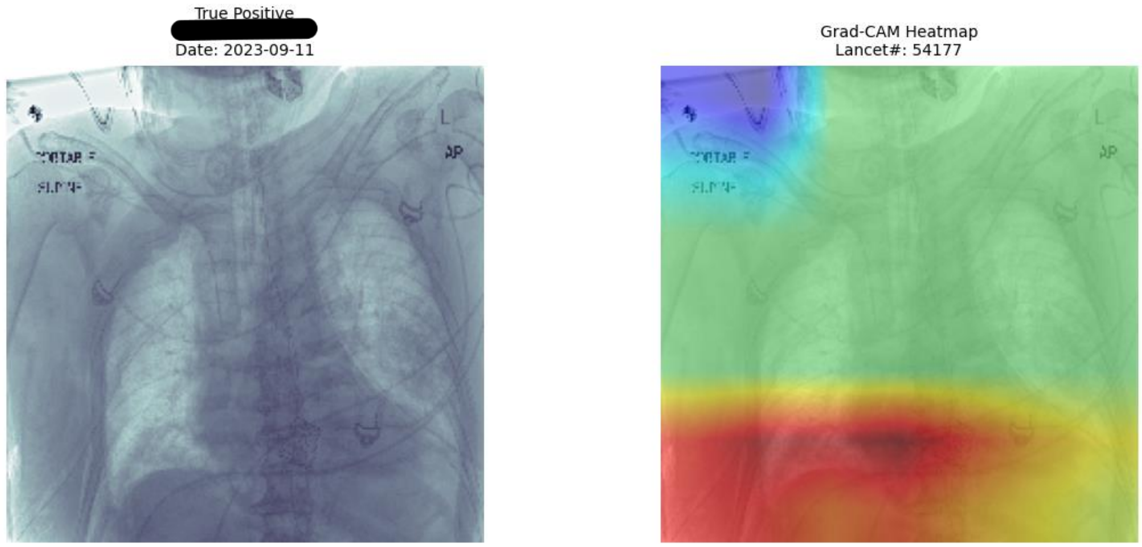


Figure 21: True Positive Case

However, CNN's are famous for being sensitive to artifacts and the next two false positive cases illustrate that phenomenon. The first one shows that machine has learned text used in trauma bay portable x-ray machine might be an indicator of transfusion:

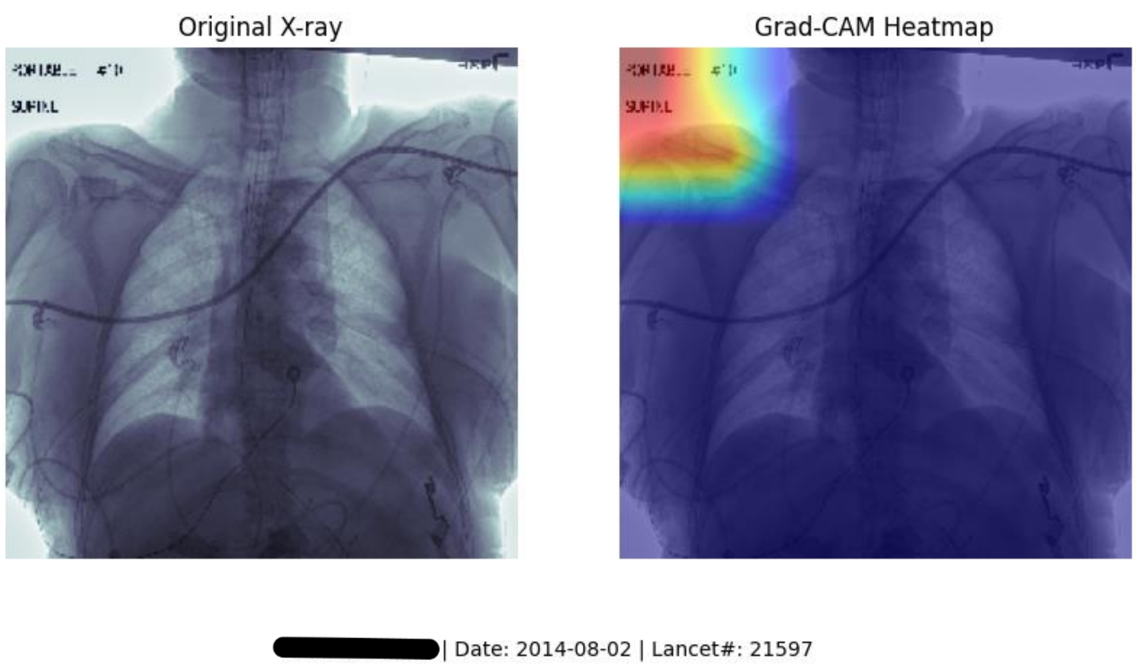


Figure 22: False Positive Case

And the second one shows an artifact in patient's dressing, which happened to be over right upper quadrant the machine considered to be a sign of possible transfusion:

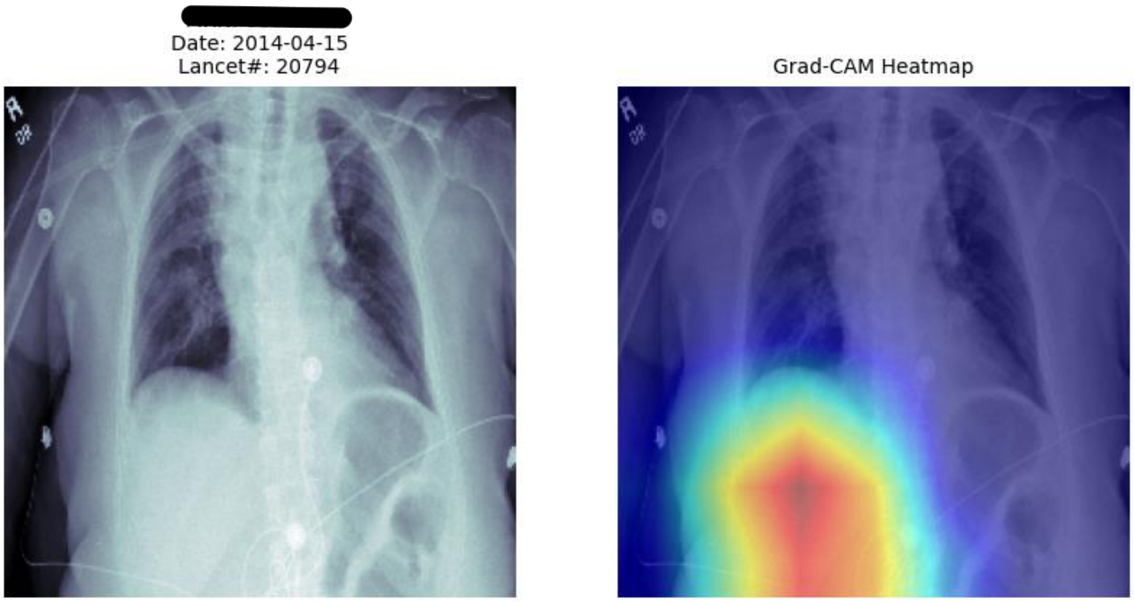


Figure 23: False Positive Case

However, the CT abdomen gives a little better view of an object in the dressing over skin defect:

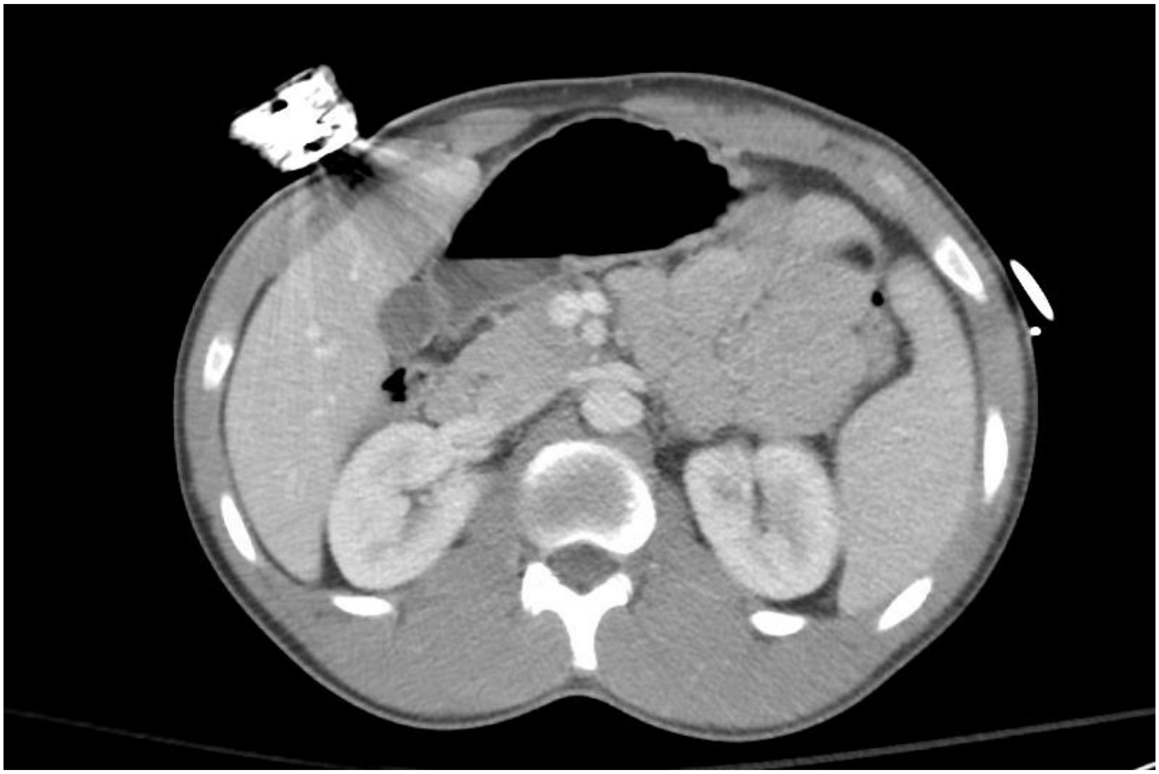


Figure 24: An Artifact in Dressing

Challenges and Future Directions

Despite the promise of ML in predicting MT needs, several challenges persist. Although sensitivity, specificity, and negative predictive values appear to be apar or better with traditional tests, the overall predictive power of the model leaves room for improvement. External validation was not performed; however, model performance was benchmarked against the ABC score results from the PROPPR trial to provide a frame of reference. Additionally, ML models must be transparent and interpretable to gain clinician trust and ensure ethical use in medical practice. One of the directions the future research should focus on improving model's explainability by reviewing heatmaps of the chest x-rays with trained radiologists to determine why the machine classified them into one or another group. Another direction is looking into using pelvic x-rays, as they are also taken in the trauma bay and may increase the accuracy of prediction tremendously.

Summary of Findings

In this study, we developed and tested a joint fusion model combining structured data and chest X-ray imaging to predict massive transfusion need upon trauma admission. Our best-performing model, DenseNet-121 plus MLP with SMOTE and X-ray augmentation, outperformed standalone models and was benchmarked against the ABC score from the PROPPR trial. Despite class imbalance challenges, the fusion model demonstrated better calibration, improved ROC-AUC, and greater interpretability via Grad-CAM heatmaps, making it a promising tool for clinical decision support.

Conclusion

The integration of machine learning into healthcare represents a paradigm shift towards more accurate and timely medical interventions. Predicting the need for massive blood transfusion using ML models can significantly improve patient outcomes and optimize resource utilization. As technological advancements continue and data availability increases, the potential for ML in transforming healthcare grows exponentially. Our research shows that fusion model utilizing data augmentation and real patient data is a feasible project and may help clinicians to identify patients in need of transfusion faster, especially when the seconds count.

Chapter 5: Predicting Admission Volume and Blood Demand Using Machine Learning

Use of ML for Clinical Operations Support

Introduction

Since 2008, a sustained decline in blood collection has been observed across the United States, with confirmation of this trend continuing through 2015 and beyond[38,117]. This reduction in the national blood supply poses a growing challenge for trauma centers, where rapid hemorrhage control and balanced transfusion remain central to the treatment of hypovolemic shock[73]. While maintaining adequate blood supply is critical to prevent mortality, overstocking may contribute to expiration-related waste, increased healthcare costs, and inefficient resource allocation.

In response to these challenges, we aim to predict blood demand at a major trauma center, using historical data. Specifically, we seek to model not only blood utilization, but also trauma volume and mechanism trends in relation to environmental and temporal variables.

Recent review of 1,814 articles suggested overall highlighted an increasing global demand for blood transfusions, driven by an aging population, increasing number of surgical procedures, especially oncological, and greater circulatory system care demands, while simultaneous decrease in blood donations by younger population and growth of older population, which is unable to donate blood[118]. These factors predict simultaneous increase of demand and decline in the supply of blood components.

Previous studies have investigated trauma seasonality and prediction using various machine learning techniques, including artificial neural networks (ANN)[119-121] and correlations to duration of the daylight[122]. Some have extended this approach to include weather-based features such as daylight hours, precipitation, and temperature. Additionally, hidden Markov models and recurrent neural networks (RNNs) have shown promise for modeling asynchronous healthcare events[123]. Building on this literature, we integrate trauma registry, EHR-derived transfusion data, and granular weather information to construct and evaluate predictive models for trauma-related blood demand.

Methods

In this study, we used two primary sources of data. Patient-level information was extracted from the trauma registry and supplemented with electronic health record (EHR) data on daily admissions and blood product orders. Together, these sources were merged to create a comprehensive dataset containing the number of patients admitted per day, their mechanisms of injury, and the blood products utilized.

Daily weather data, generously provided by CustomWeather, Inc., included the following variables: maximum, minimum, and average temperatures; heating and cooling degree days; precipitation; average relative humidity; average wind speed; average dew point; average visibility; and average sea level pressure.

We used several R packages for data processing and analysis. For data import and cleaning, we employed readr[49], janitor[45], dplyr[69], and reshape2[50]. Exploratory data analysis and summary statistics were conducted using broom[47], psych[51], moderndive[59], and gtsummary[54]. To engineer time-based features, we used lubridate[52], chron[62], and tis[124] to extract weekdays, holidays, and weekends. The lunar[125] package was used to encode moon phases, and StreamMetabolism[126] provided geographic-specific calculations of daily solar duration.
Once the dataset was constructed, we explored correlations, assessed seasonality, and evaluated time-series stationarity. We first applied naïve forecasting methods to establish a baseline. Next, we implemented several classical and machine learning modeling approaches, including ARIMA, Holt–Winters exponential smoothing, multilayer perceptron (MLP), XGBoost, random forest (RF), recurrent neural networks (RNN), convolutional neural networks (CNN), linear regression, support vector regression (SVR), generalized autoregressive conditional heteroskedasticity (GARCH), and Facebook Prophet.

Time Series Characteristics of Trauma Admissions

Daily trauma admissions at our center exhibit clear temporal structure and non-random variation, making them suitable for time series analysis. Over the past ten years, we observed a steady increase in admissions volume, with a pronounced surge during the

COVID-19 pandemic period, which persisted for several years thereafter. This long-term growth trend is visualized in **Figure 25**.

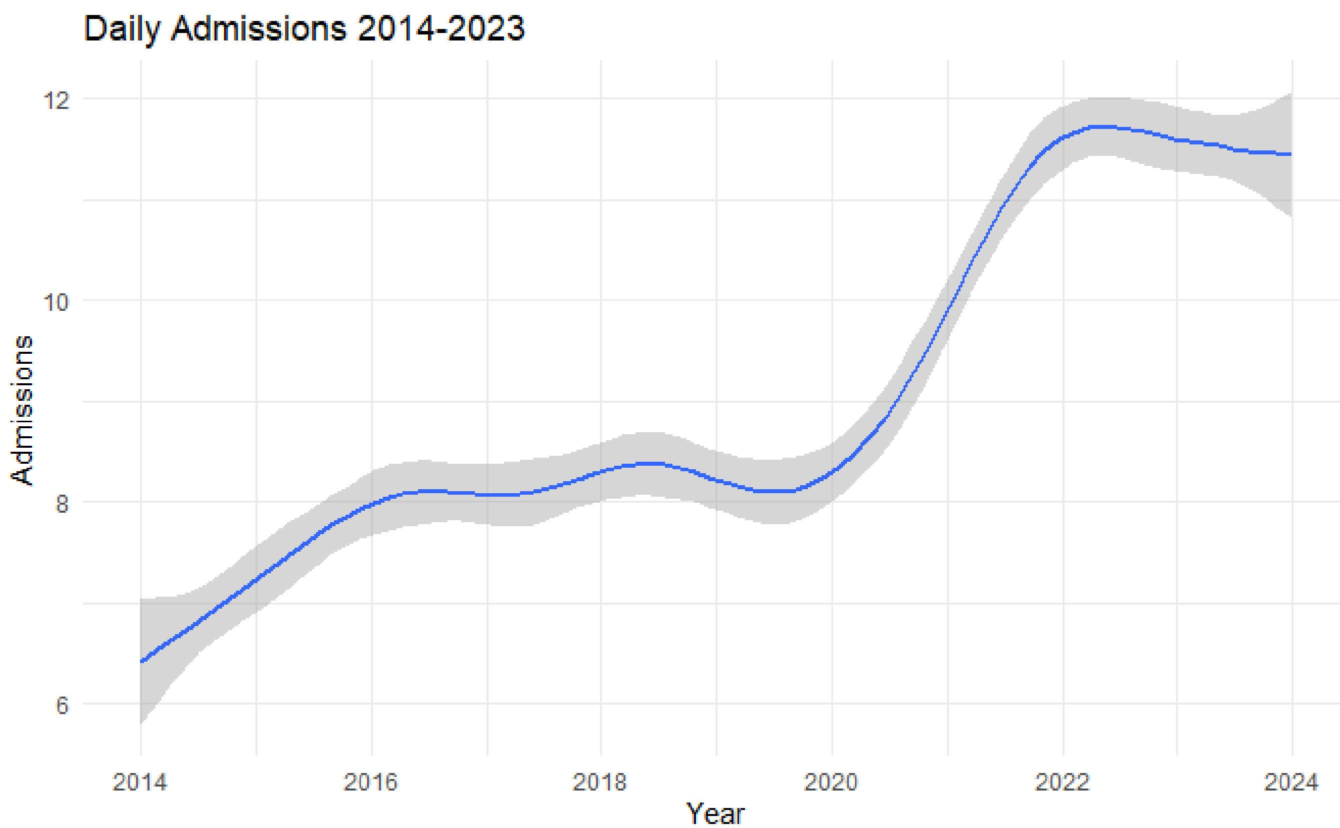


Figure 25: Daily Admissions 2014-2024

Annual and Weekly Seasonality

Admission volumes show strong seasonal patterns on both annual and weekly scales. The annual seasonality reflects increased trauma admissions during warmer months, particularly summer, which coincides with greater outdoor activity, travel, and alcohol-related incidents. This is supported by environmental correlations (e.g., higher admissions on hot days and during longer daylight hours).

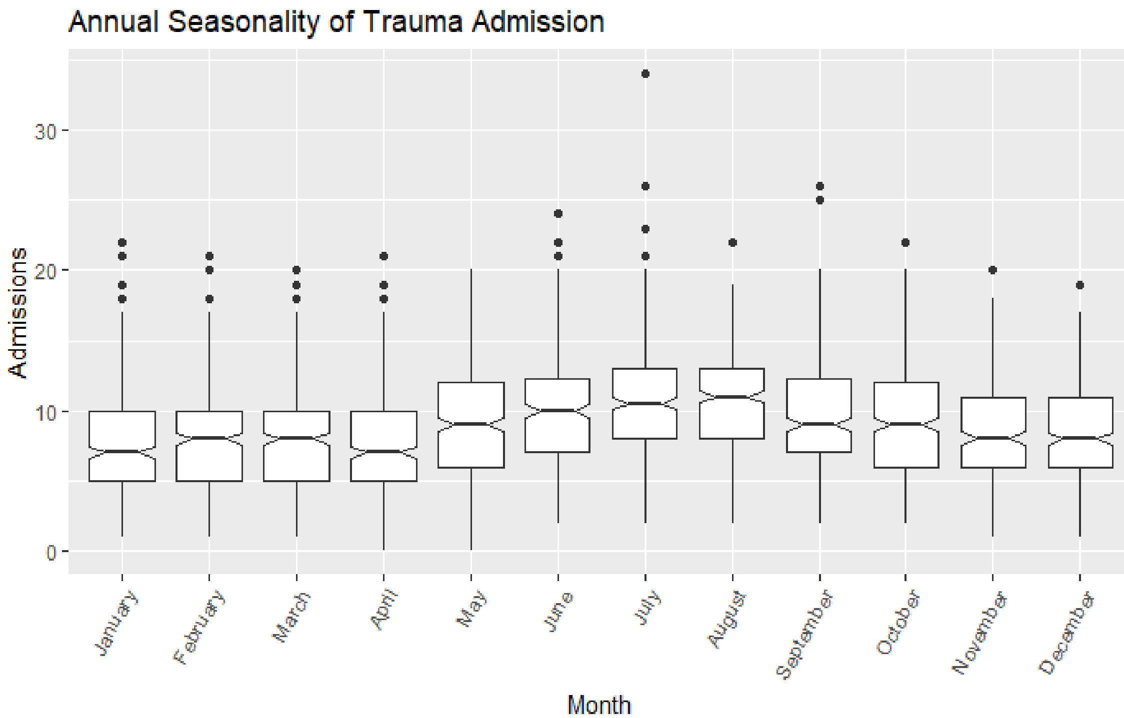


Figure 26: Annual Seasonality of Trauma Admission

Daily Daylight Duration as a Proxy for Annual Seasonality

To quantify annual seasonality in a continuous and geographically specific way, we incorporated daily daylight duration into our predictive modeling framework. Unlike categorical representations such as "summer vs. winter" or monthly dummies, daylight duration offers a continuous variable that naturally encodes the progression of the seasons across the calendar year.

This variable was derived using the StreamMetabolism R package, which calculates daylight hours based on the center's geographic latitude and date. The resulting variable was included in both exploratory analysis and as a feature in machine learning models. As shown in **Figure L**, trauma admissions increase in tandem with longer daylight hours, peaking during mid-summer when daylight reaches its annual maximum. This pattern reflects heightened outdoor activity, increased vehicular traffic, and seasonal risk behaviors that contribute to trauma burden.

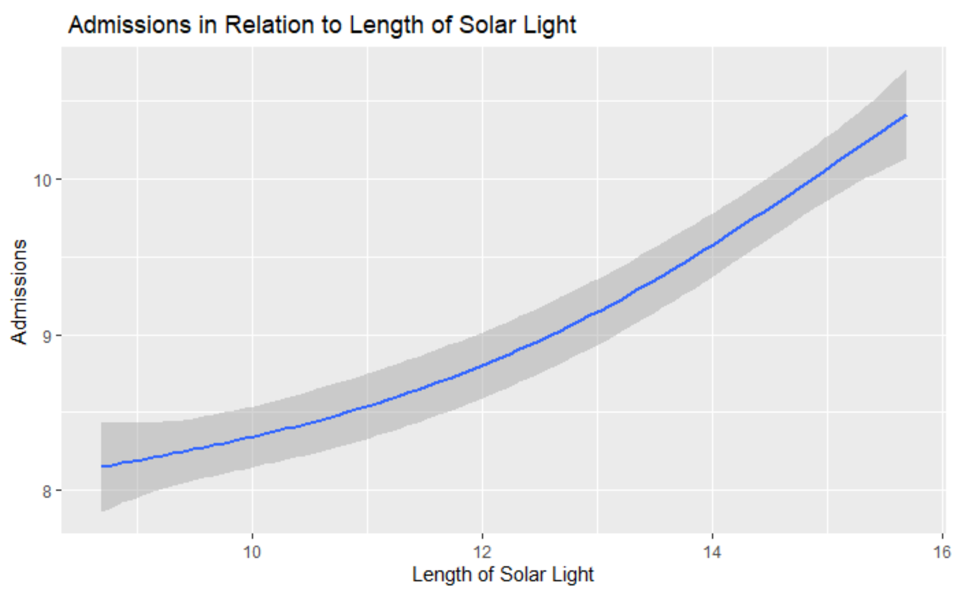


Figure 27: Admission in Relation to Length of Solar Light

Daily trauma admissions plotted against duration of daylight (hours) across the calendar year. A positive nonlinear relationship is observed, with higher admission volumes corresponding to longer days, supporting the inclusion of daylight duration as a continuous marker of seasonality.

In addition, a clear weekly seasonality is present. Admissions tend to rise on weekends and public holidays, with notable peaks on Fridays, Saturdays, and Sundays (see **Figure**

Y). These peaks may reflect increased recreational activity, vehicular traffic, and alcohol consumption.

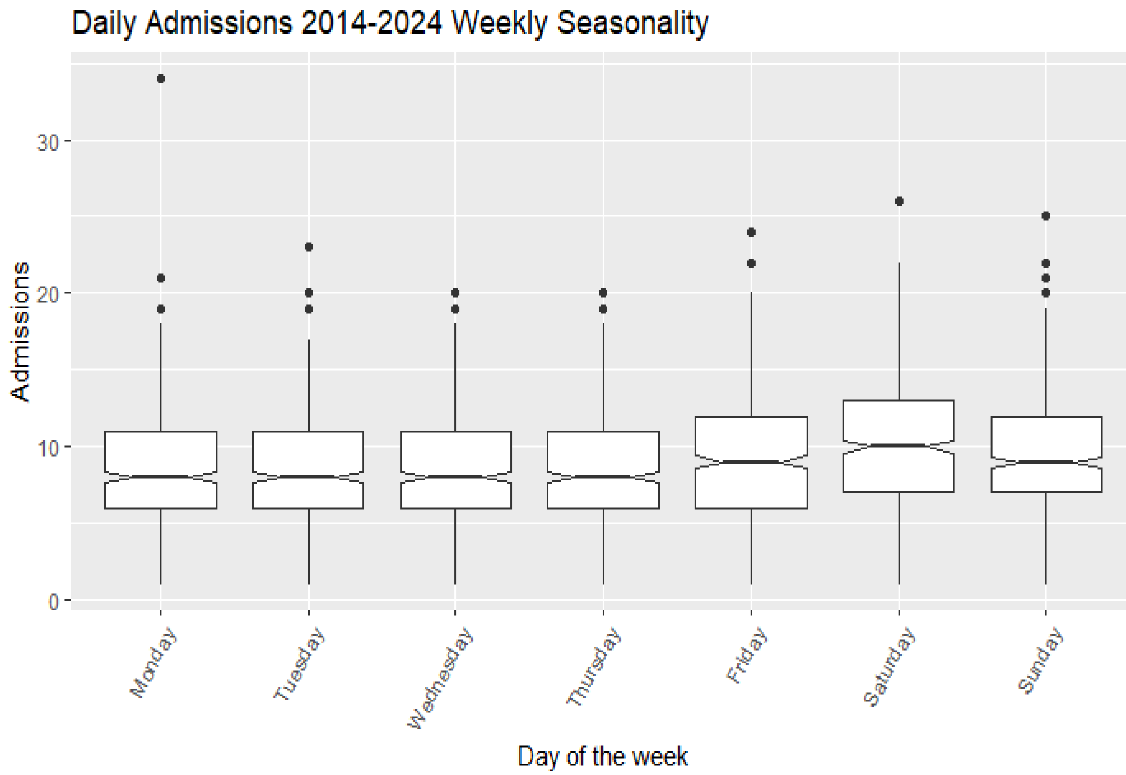


Figure 28: Daily Admissions 2014-2024 Weekly Seasonality

There were multiple variables affecting the overall trauma admission rate. Interestingly, despite longstanding myths in clinical folklore, we found no significant association between lunar phase and trauma admissions (p = 0.60), as shown in **Figure 29**.

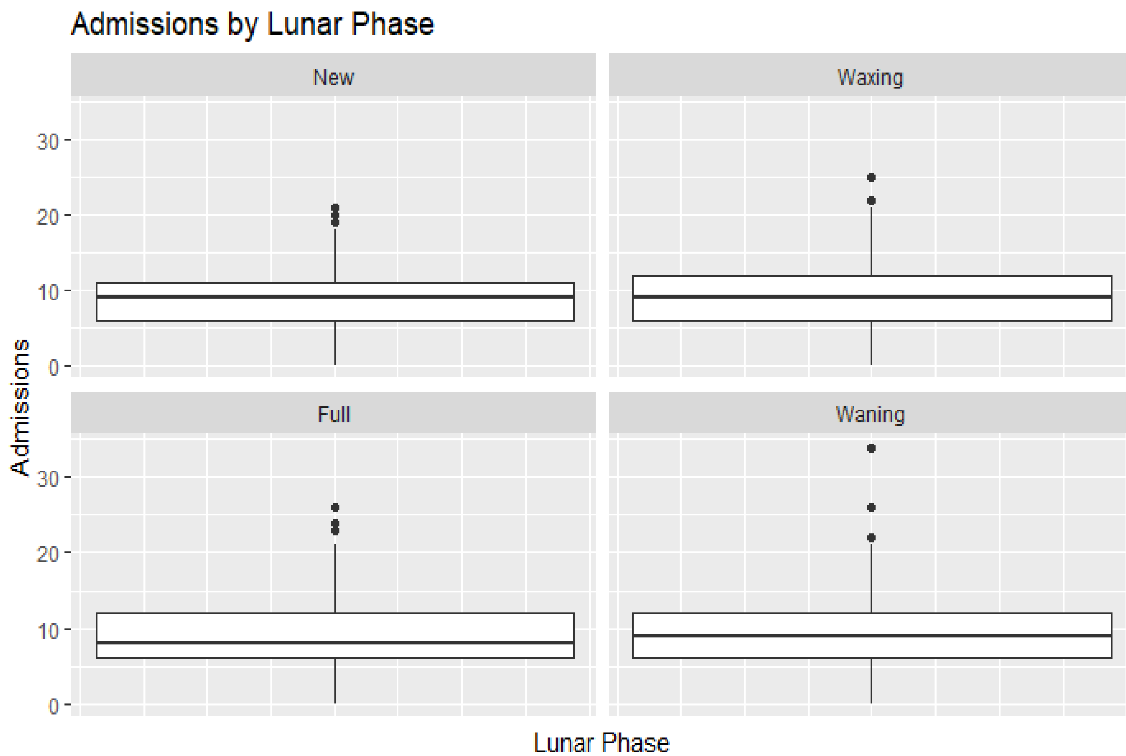


Figure 29: Admission by Lunar Phase

Environmental and Calendar Characteristics Affecting Admission Rate

Characteristic	N	Beta	95% CI1	p-value
Max Temperature (°F)	3,782	0.07	0.07, 0.08	<0.001
Min Temperature (°F)	3,782	0.08	0.07, 0.10	<0.001
Average Temperature (°F)	3,782	0.09	0.08, 0.10	<0.001
Heating Degree Days	3,782	-0.10	-0.11, -0.08	<0.001
Cooling Degree Days	3,782	0.25	0.22, 0.28	<0.001
Precipitation (in)	3,782	-0.02	-0.02, -0.01	<0.001
Avg Relative Humidity (%)	3,782	-0.07	-0.08, -0.06	<0.001
Avg Wind Speed (mph)	3,782	-0.07	-0.11, -0.03	<0.001

Environmental and Calendar Characteristics Affecting Admission Rate

Characteristic	**N**	**Beta**	**95% CI**1	**p-value**
Avg Dew Point (°F)	3,782	0.08	0.06, 0.09	<0.001
Avg Visibility (miles)	3,782	0.27	0.18, 0.37	<0.001
Avg Sea Level Pressure	3,782	-0.28	-0.94, 0.39	0.4
Weekend	3,782	1.4	1.1, 1.7	<0.001
Holiday	3,782	0.61	-0.15, 1.4	0.12
Days Off	3,782	1.4	1.1, 1.7	<0.001
Solar Day Length (hr)	3,782	0.30	0.25, 0.35	<0.001
Rain Indicator	3,782	-1.5	-1.7, -1.2	<0.001
Snow Indicator	3,782	-1.3	-2.1, -0.61	<0.001
Ice Indicator	3,782	-1.4	-2.8, 0.02	0.054
Lunar Phase	3,782	0.02	-0.05, 0.09	0.6
Monday	3,782	-0.43	-0.79, -0.07	0.019
Tuesday	3,782	-0.84	-1.2, -0.48	<0.001
Wednesday	3,782	-0.70	-1.1, -0.34	<0.001
Thursday	3,782	-0.61	-0.96, -0.25	<0.001
Friday	3,782	0.23	-0.13, 0.59	0.2
Saturday	3,782	1.6	1.2, 1.9	<0.001
Sunday	3,782	0.77	0.41, 1.1	<0.001

1CI = Confidence Interval

Some of the strong admissions predictors included days-off:

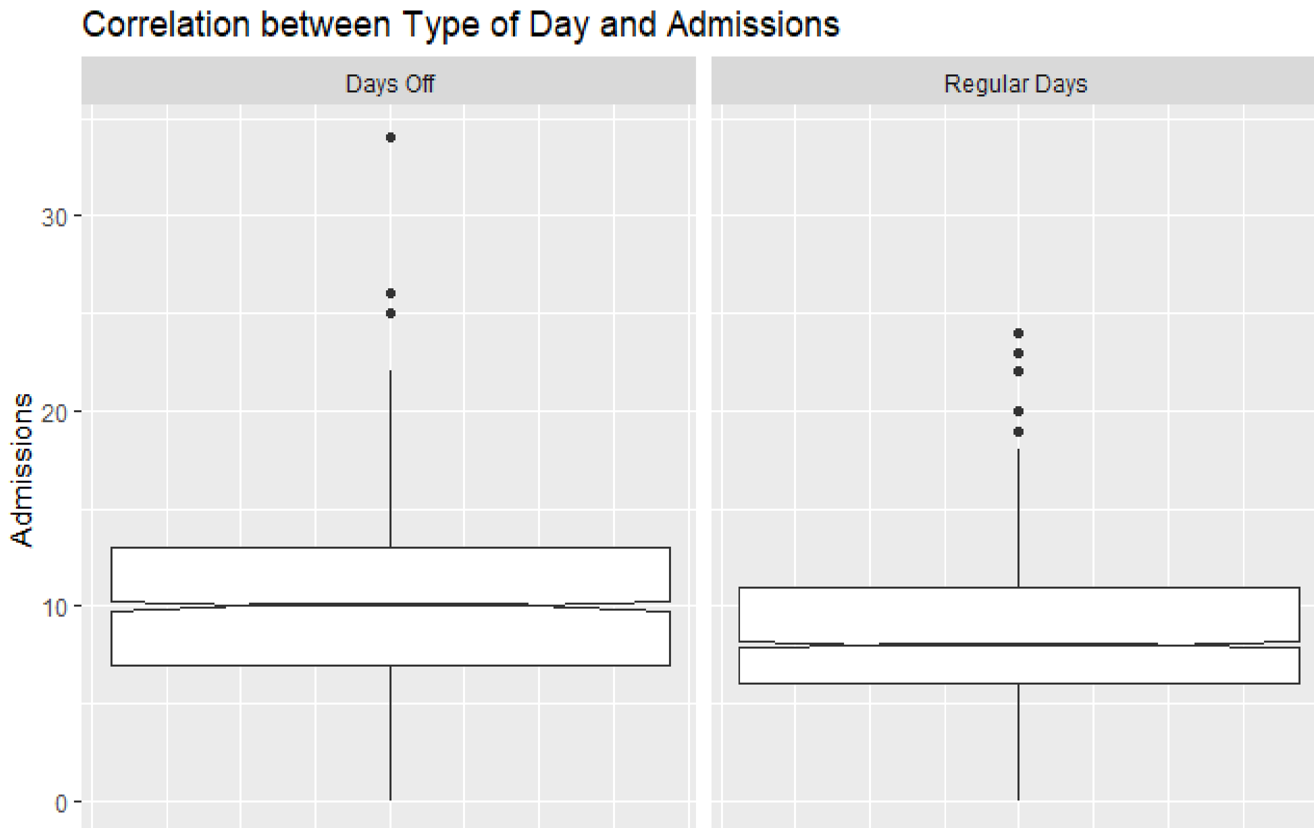


Figure 30: Admission Days-Off vs. Regular Days

Environmental variables, such as rain, snow, and ice correlated negatively with admission rate:

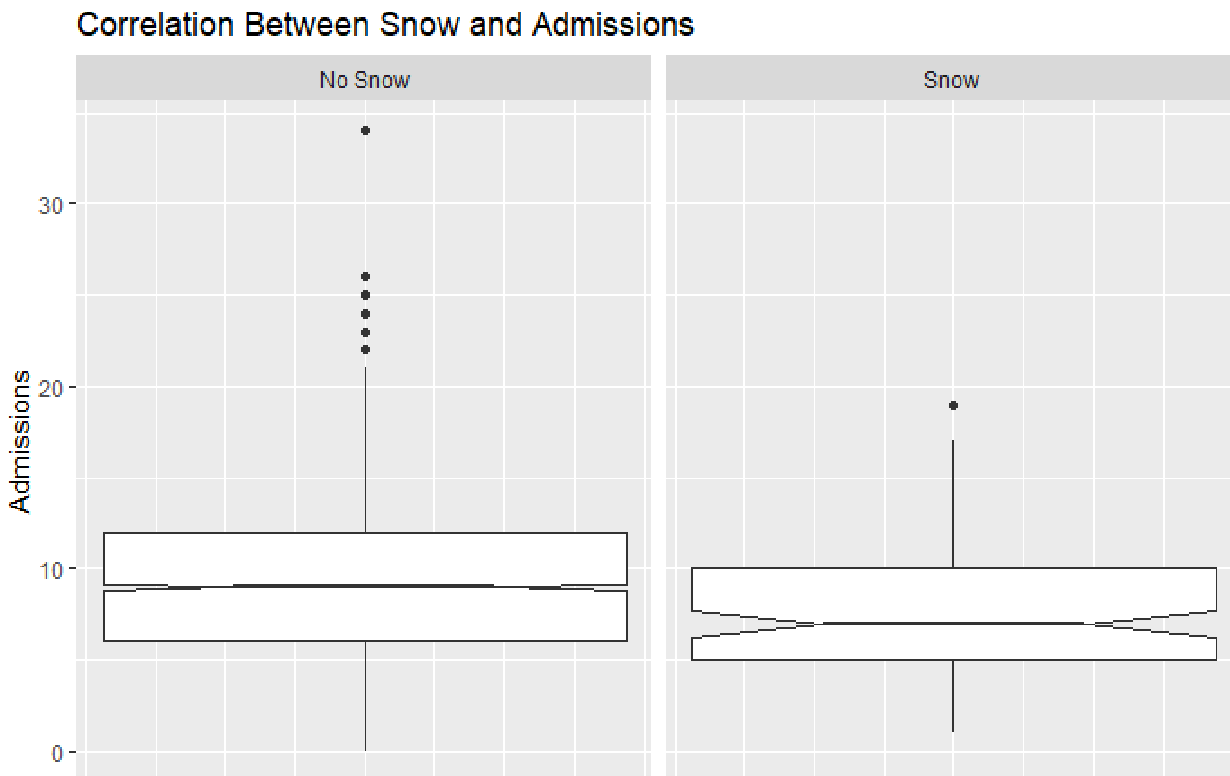


Figure 31: Admissions Affected by Snow

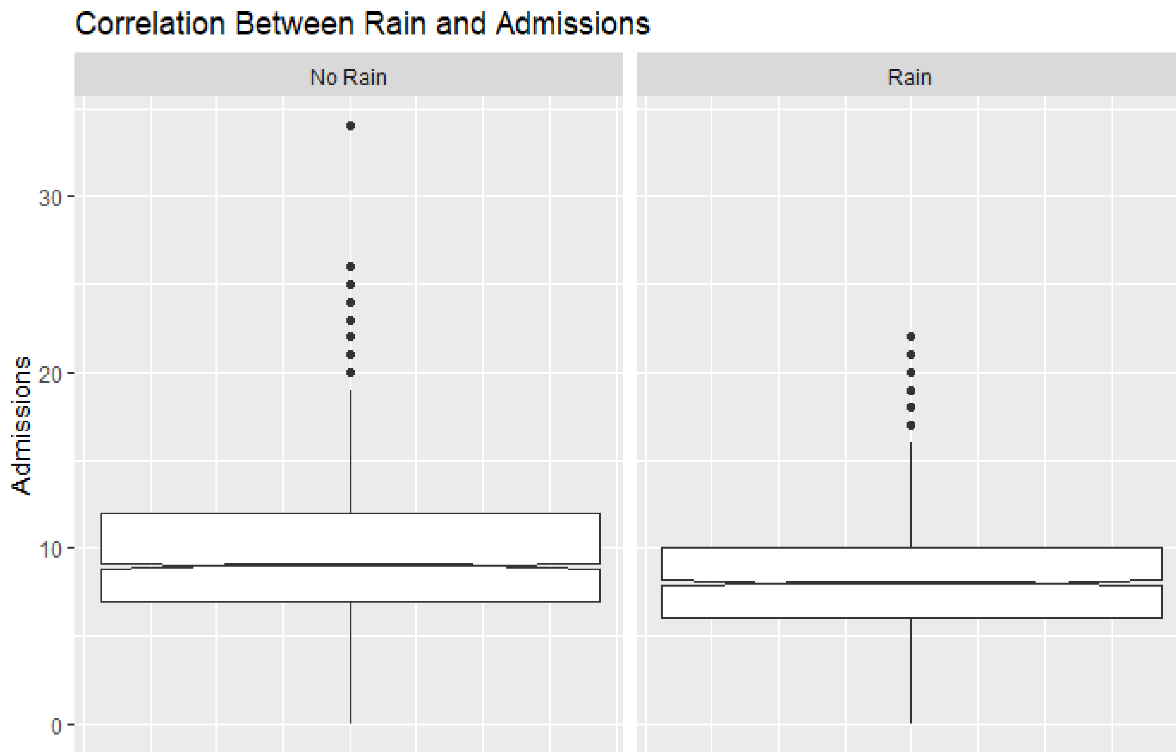


Figure 32: Admissions Affected by Rain

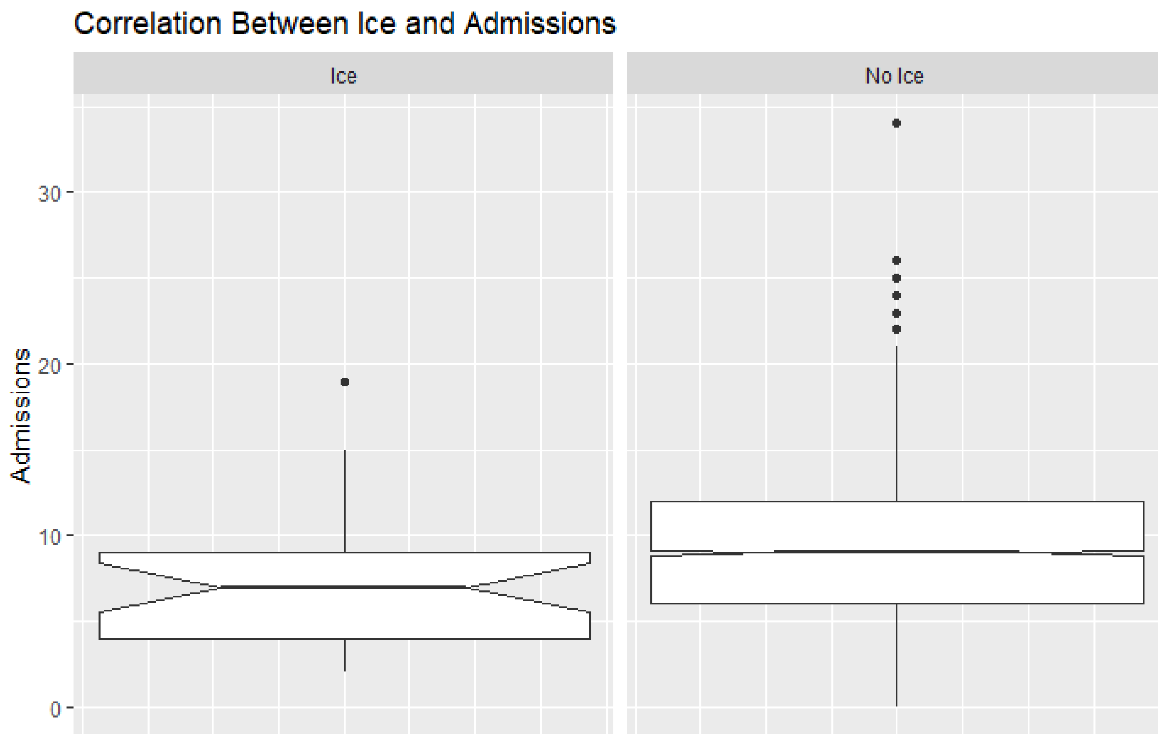


Figure 33: Admissions Affected by Ice

Distribution and Variance

The distribution of daily trauma admissions approximates a normal distribution but exhibits moderate skewness and kurtosis, reflecting the real-world heterogeneity in trauma burden. Over time, we also observed non-constant variance, or heteroscedasticity, suggesting that the data violates assumptions of constant error variance required for some linear models.

To illustrate this, we plotted the rolling variance of daily admissions, revealing clear fluctuations over the course of the year (**Figure 34**). This pattern suggests that both the mean and variability of admissions change over time, reinforcing the need for models that can adapt to seasonality and nonstationary patterns.

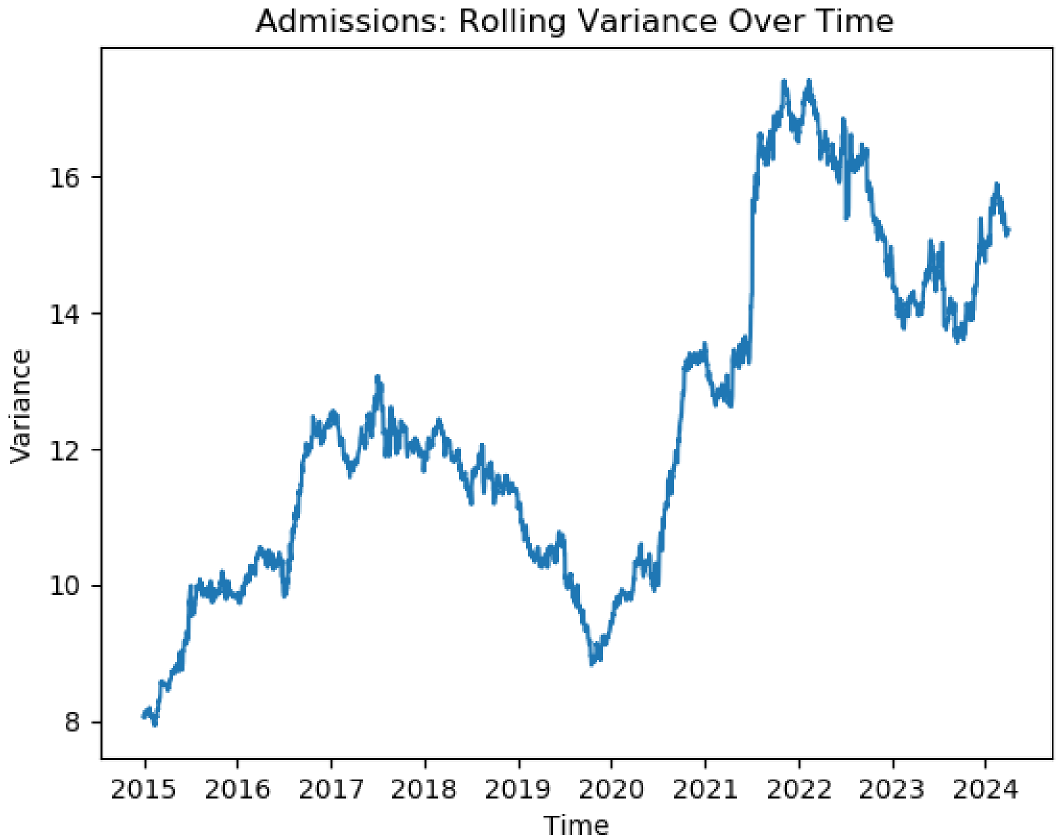


Figure 34: Admissions Rolling Variance Over Time

Stationarity Assessment

To formally evaluate whether trauma admissions could be modeled using time-series methods that assume stationarity, we applied the Augmented Dickey–Fuller (ADF) test. Results indicated the presence of a unit root, meaning the series is non-stationary in its raw form (ADF p-value > 0.05). Non-stationary data limits the performance of traditional models and necessitates transformation before modeling.

Box–Cox Transformation

To address this issue, we applied the **Box–Cox transformation**, a power transform introduced by Box and Cox in 1964[127]. This method stabilizes variance and improves normality of the series without altering the relative order of values—an essential property for preserving temporal structure. It has been successfully used for data normalization, which allowed subsequently to use traditional statistical methods for normally distributed data[128,129], and to prepare data for use in ML[130].

The Box–Cox transformation is defined as:

$$y(\lambda) = \begin{cases} \frac{y^{\lambda}-1}{\lambda}, & \lambda \neq 0 \\ \ln(y), & \lambda = 0 \end{cases}$$

Figure 35: Box-Cox Conversion Formula

In our dataset, applying this transformation to daily trauma admissions improved Gaussianity and reduced variance, enabling better performance of forecasting models downstream. The effect of this transformation is illustrated in **Figure 36**, showing a more stable variance and a more symmetric distribution after transformation.

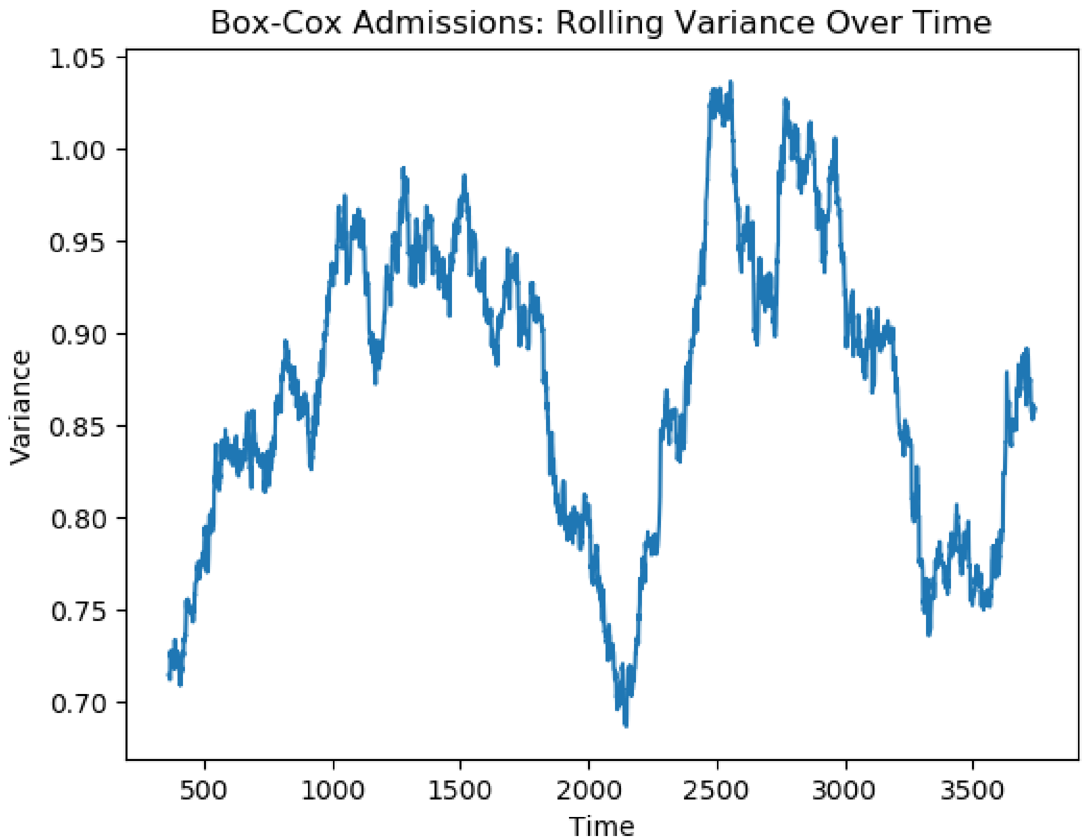


Figure 36: Admissions Rolling Variance After Box-Cox Transformation

Naïve and Holt–Winters Forecasting Models

Naïve Forecasting as Baseline

To establish a minimum threshold for predictive performance, we first employed naïve forecasting, a simple yet commonly used baseline in time-series modeling. Under this approach, the forecast for any future time point is assumed to be equal to the most recent observed value:

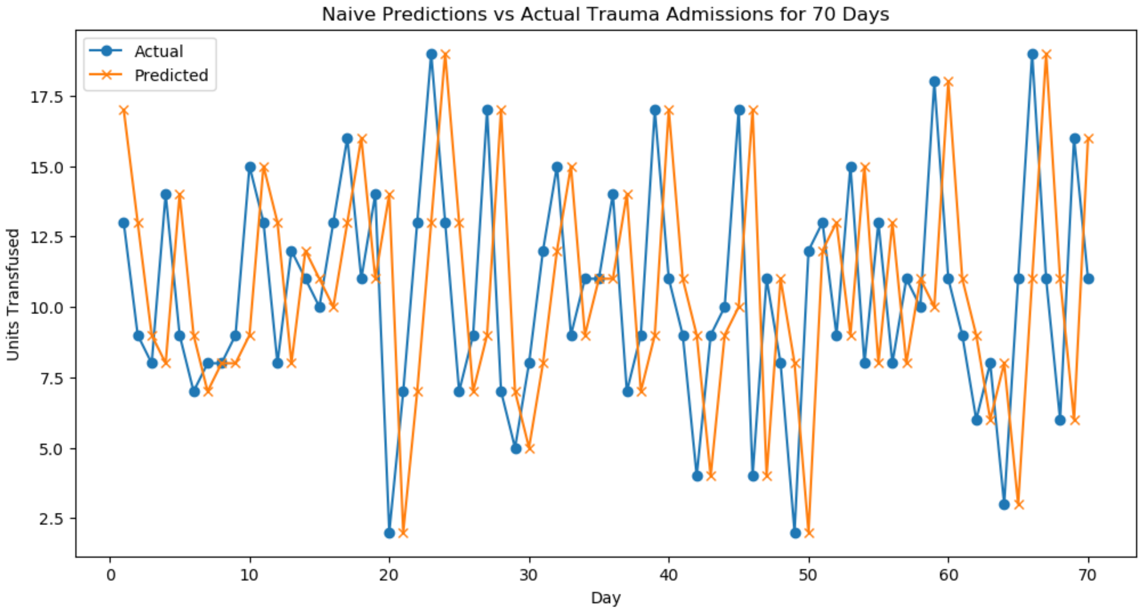
$$\hat{y}_{t+1} = y_t$$

Figure 37: Naive Forecasting Formula

This method does not account for trends, seasonality, or external influences, but it provides a reference point for evaluating the value added by more complex models. Applied to our trauma admissions dataset, the naïve model yielded the following performance metrics:

- Mean Absolute Error (MAE): 20.5
- Root Mean Squared Error (RMSE): 4.5
- R^2: –0.38

These results, particularly the negative R^2, confirm poor model fit and illustrate the need for models capable of capturing seasonality and temporal dynamics (**Figure A**).



Holt–Winters Exponential Smoothing

A more sophisticated alternative is the Holt–Winters method, which explicitly models both trend and seasonal components using exponential smoothing. We applied the additive version of the model, suitable when the seasonal variations are roughly constant over time:

$$\hat{y}_{t+h} = (L_t + h \cdot T_t) + S_{t+h-m(k+1)}$$

L_t = level estimate
T_t = trend estimate
S_t = seasonal component
m = seasonal period
h = forecast horizon

Figure 38: Holt-Winters Formula

This model demonstrated substantial improvement over naïve prediction:

- Training MAE: 4.78
- Training RMSE: 5.02
- Training R^2: –4.86
- Test MAE: 6.93
- Test RMSE: 7.04
- Test R^2: –17.99

While the MAE and RMSE values improved considerably, the negative R^2 values—especially on the test set—highlight that while the model able to follow general trend and seasonality, it struggled with generalization and are not useful for predicting actual volume of admissions. (Figure 39).

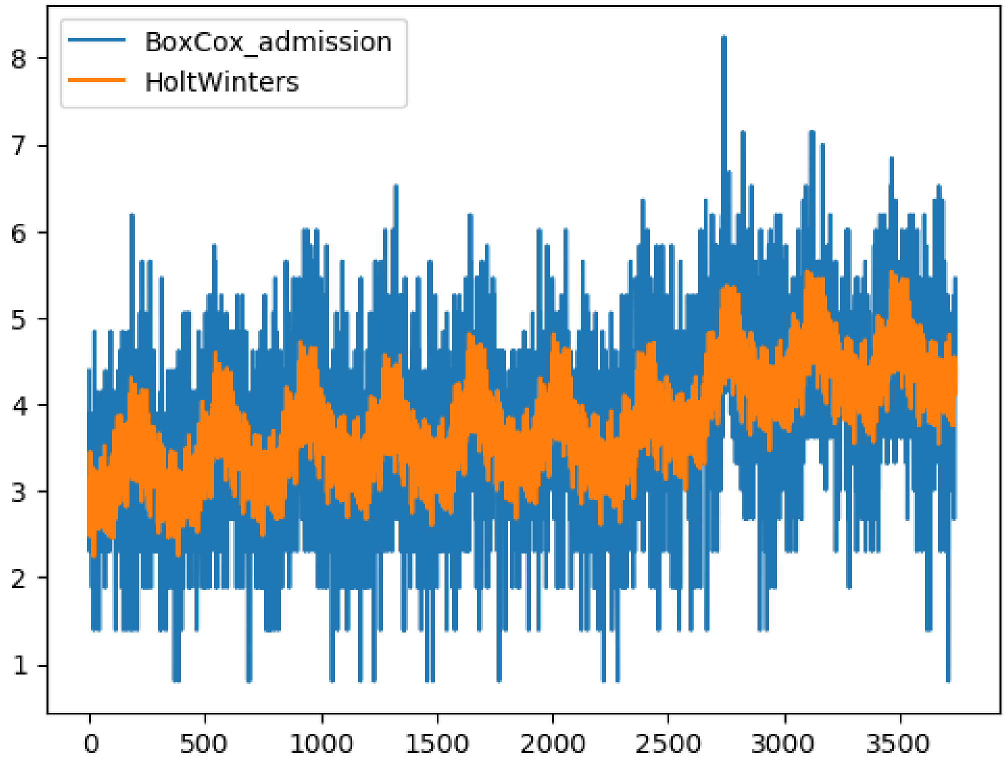


Figure 39: Holt-Winters Predictions

On visual inspection, the model effectively learned the average seasonal pattern and general level of admissions, but it failed to capture short-term fluctuations or sudden changes in trauma volume (**Figure 40**).

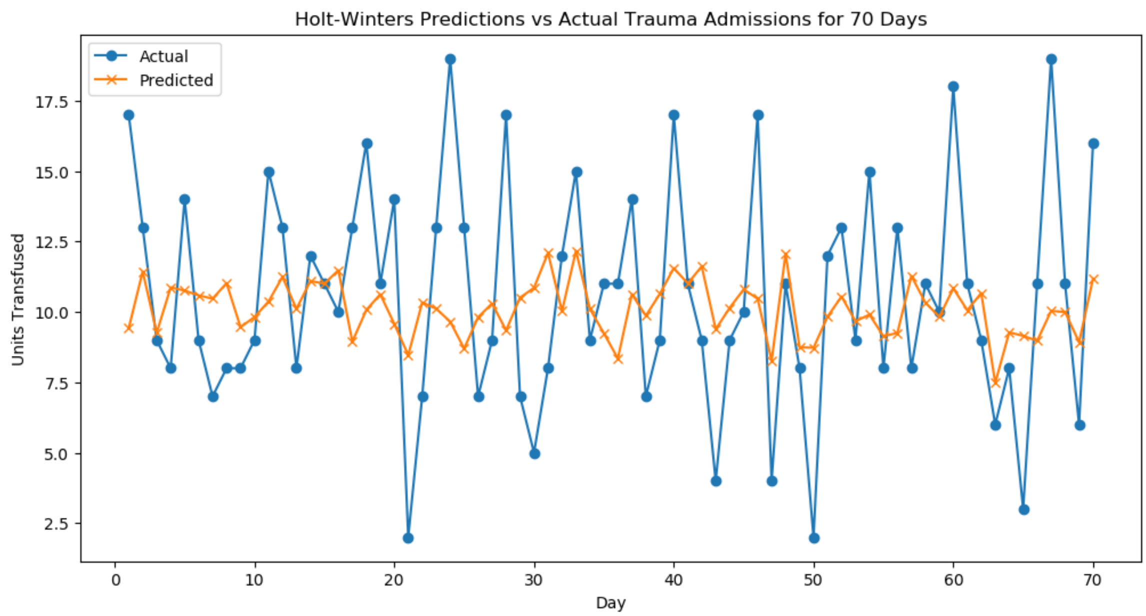


Figure 40: Holt-Winters Predictions for Horizon of 70 Days

Summary and Implications

Despite its simplicity, the Holt–Winters method provided a meaningful baseline for trend and seasonality-aware forecasting. However, its inability to adapt to local or abrupt changes in trauma admission rates underscores the need for more flexible, data-driven models, such as machine learning approaches. These classical models helped quantify the underlying signal structure and set the stage for more accurate forecasts from neural and ensemble models.

Multilayer Perceptron (MLP) for Admissions Forecasting

To capture nonlinear relationships between trauma admissions and environmental, temporal, and historical features, we trained a Multilayer Perceptron (MLP) model using engineered features from the combined trauma registry, EHR, and weather dataset.

The MLP was implemented using the MLPRegressor from the scikit-learn package. The architecture consisted of a single hidden layer with 20 neurons, using the ReLU activation function and early stopping to prevent overfitting.

Model Training and Forecasting

The model was trained for up to 200 epochs, with early stopping triggered after approximately 180 iterations, indicating convergence. Visual inspection of the training loss curve showed a smooth descent and no signs of instability or overfitting (**Figure 41**).

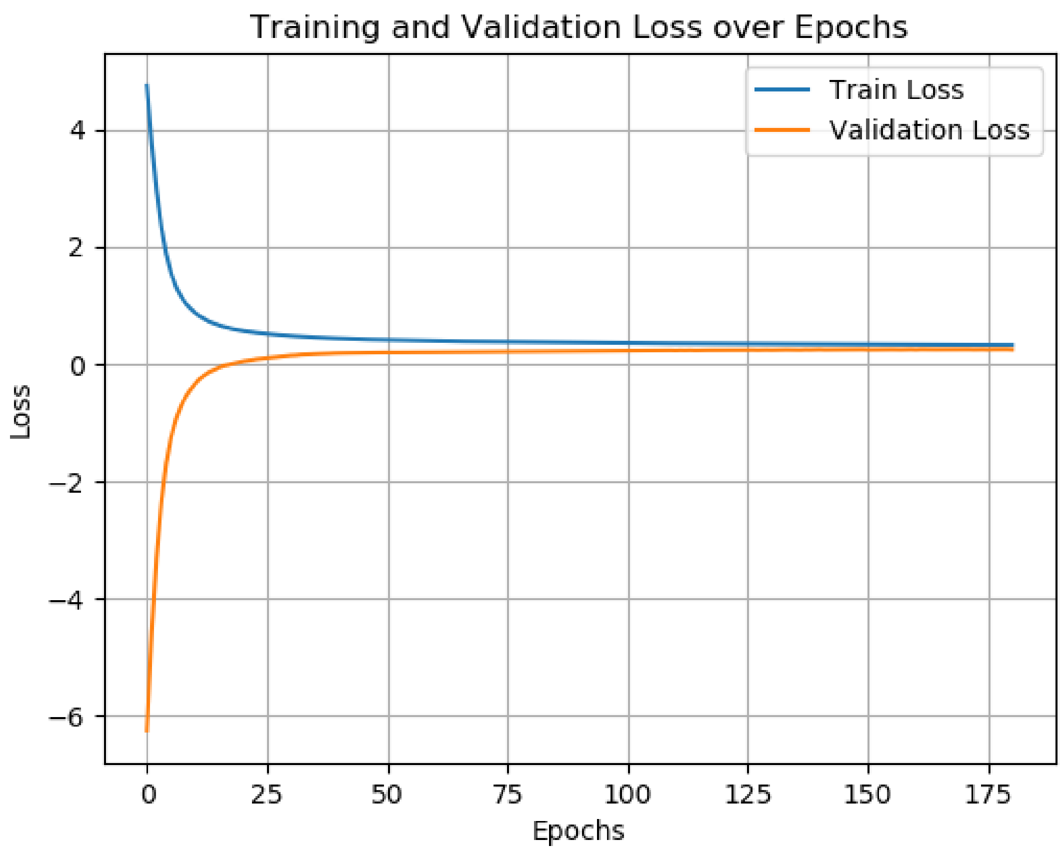


Figure 41: Training and Validation Loss for MLP

On the training data, the model learned the underlying structure of trauma admissions well, while maintaining generalizability on the testing set:

- Training MAE: 2.43
- Training RMSE: 3.08
- Training R^2: 0.38
- Test MAE: 2.69
- Test RMSE: 3.42
- Test R^2: 0.17

Compared to Holt–Winters and naïve baselines, the MLP exhibited substantially improved accuracy and positive R^2 values, demonstrating that the model successfully captured meaningful, generalizable relationships (**Figure 42: MLP 60-day forecast**).

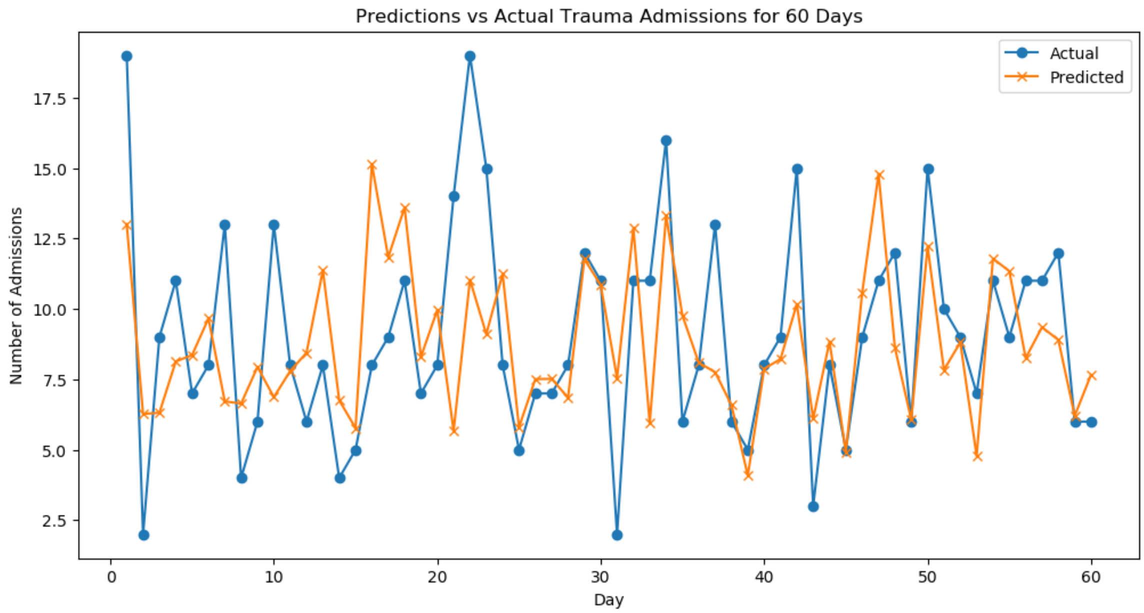


Figure 42: MLP 60 Days Forecast

Residual Distribution Analysis

To assess the calibration and error structure of the MLP model, we plotted the distribution of residuals (i.e., the difference between observed and predicted trauma admissions) for both training and testing sets.

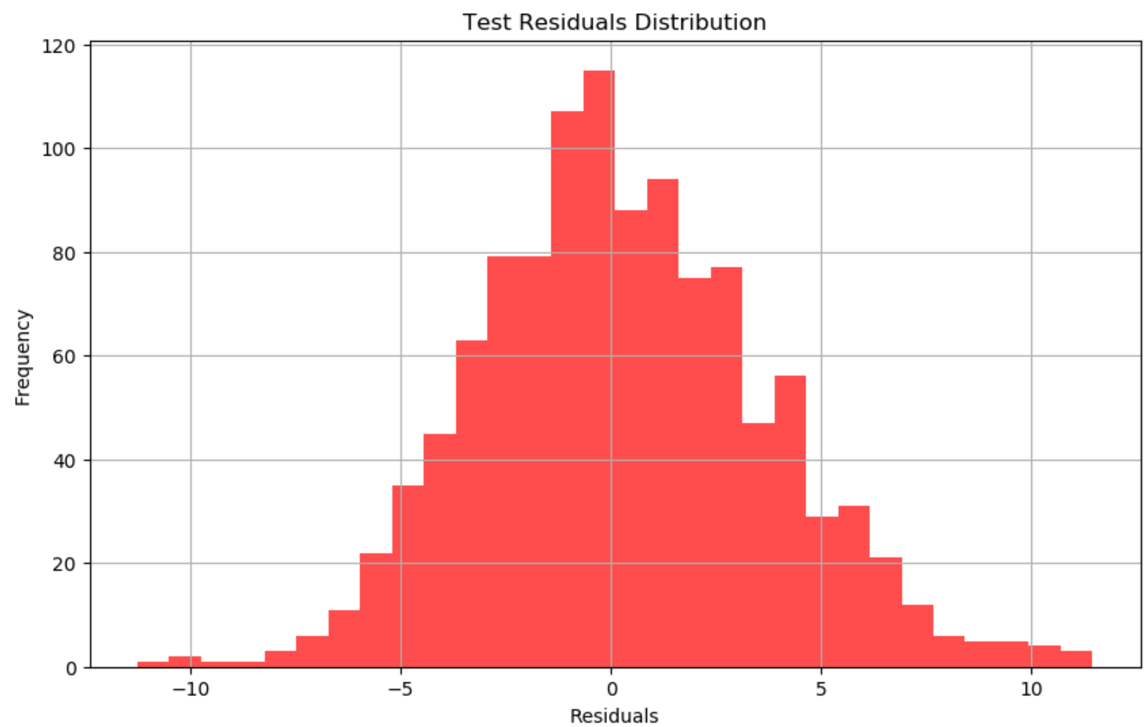


Figure 43: Test Residual Distribution

As shown in **Figure 43**, the residuals are centered around zero, indicating that the model does not consistently over- or under-predict, and approximately normally distributed, with mild deviations at the tails, which is expected given the real-world variability in trauma admission volumes.

These findings support the use of this model in forecasting trauma admissions, as normally distributed residuals are a desirable property in predictive modeling—they imply that the majority of predictions are unbiased and fall within a reasonable error margin.

To assess prediction error distribution, we analyzed residuals on both training and test sets. Key findings include:

Training residuals:

Skewness = 0.40

Kurtosis = 0.53

Shapiro–Wilk p-value = 1.50×10^{-11}

Testing residuals:

Skewness = 0.24

Kurtosis = 0.17

Shapiro–Wilk p-value = 0.0022

While the Shapiro–Wilk test rejected perfect normality, Q–Q plots showed that residuals closely approximated a normal distribution within ±2 standard deviations, with deviations occurring mostly in the tails—consistent with real-world trauma data characterized by occasional outliers and heteroscedasticity (**Figure 44**).

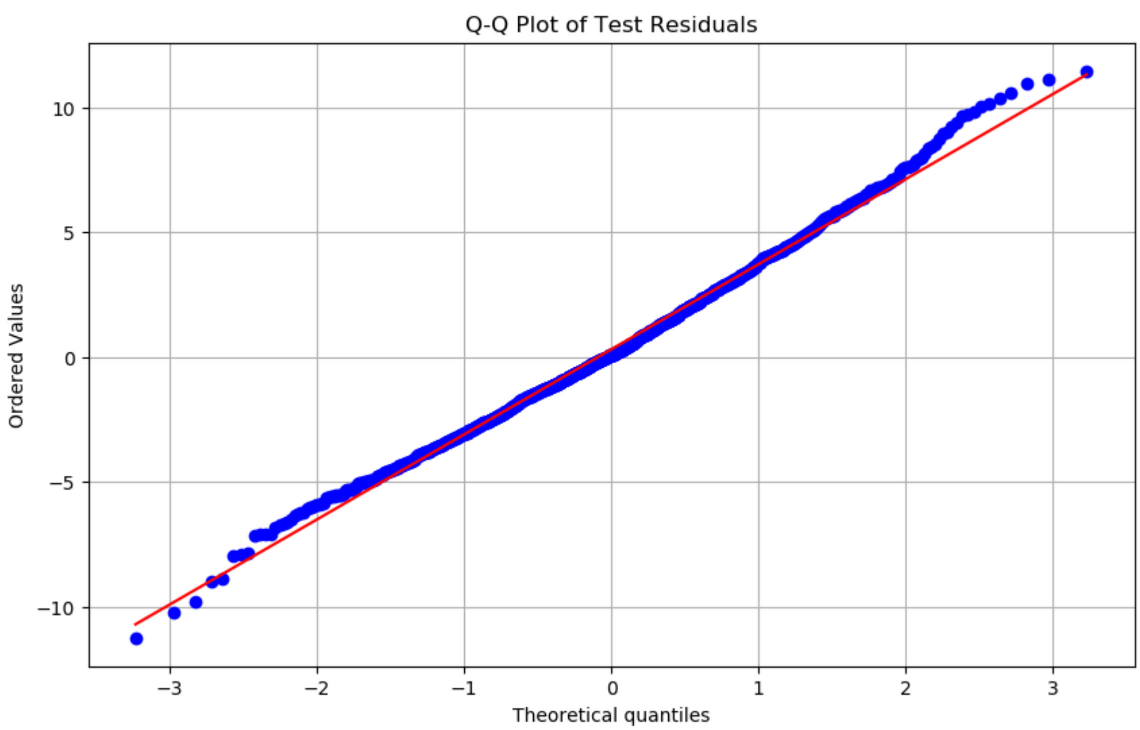


Figure 44: Test Residual Q-Q Plot

Feature Importance

To interpret the MLP model's behavior, we evaluated feature importance using permutation importance scores. The most influential predictors included:

- **Environmental features**:
    - Minimum and average temperature
    - Visibility
    - Precipitation
- **Temporal features**:
    - Days of the week (notably Sunday and Friday)
    - Days off and holidays
    - Solar day length
- **Recent trauma trends**:
    - 30-day rolling averages for mechanisms such as **MVCs**, **GSWs**, **pedestrian injuries**, and **falls**

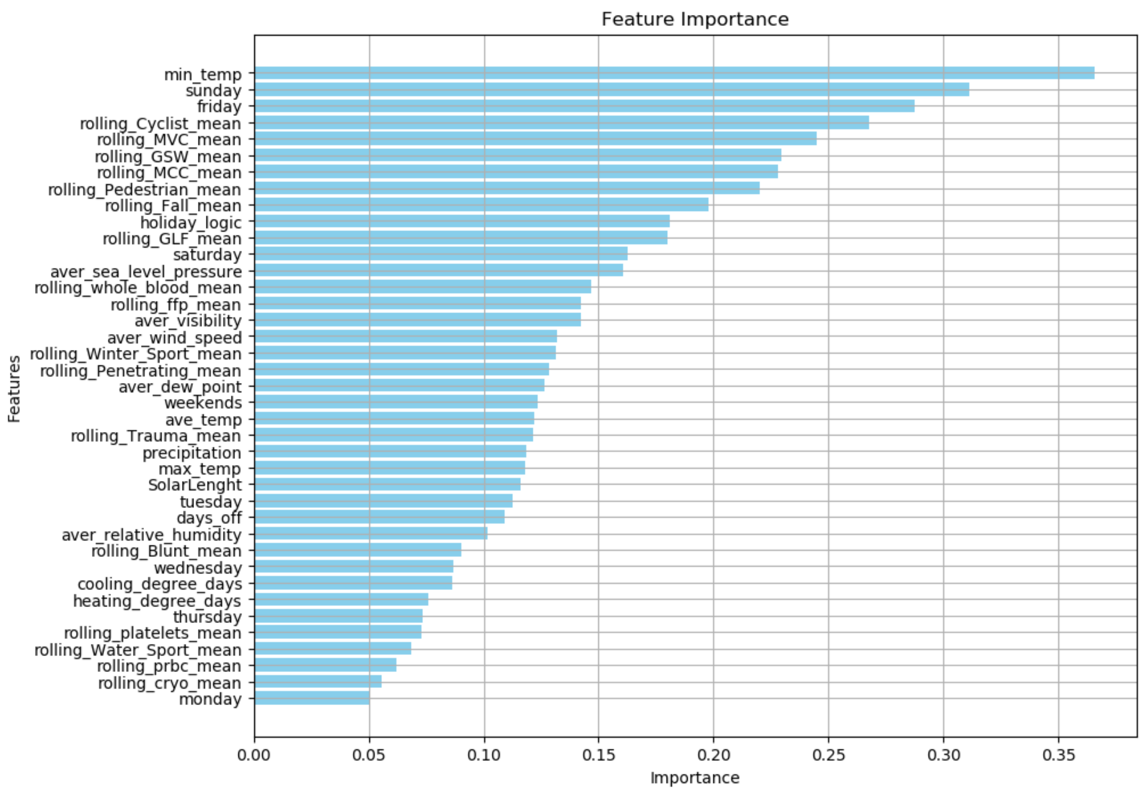


Figure 45: Ranked Feature Importance Plot

These findings suggest that the MLP model effectively leveraged both external context (weather and calendar variables) and internal clinical signals (historical injury trends) to make accurate forecasts.

Summary

The MLP model demonstrated strong performance relative to traditional time-series methods. It balanced complexity with generalizability, producing accurate forecasts with interpretable patterns in residuals and feature contributions. While more complex architectures may offer incremental improvements, the MLP proved to be a computationally efficient and clinically relevant model for forecasting trauma admissions.

Extreme Gradient Boosting (XGBoost) for Trauma Admissions

To complement the MLP model and compare performance across model families, we implemented an Extreme Gradient Boosting (XGBoost) regressor. XGBoost is a tree-based ensemble learning algorithm known for its high accuracy, resistance to overfitting, and interpretability via feature importance. It is particularly well-suited for tabular data with mixed data types and complex, nonlinear relationships.

The model was trained using the same set of input variables as the MLP model, including temporal, environmental, and rolling average mechanism-specific features. Hyperparameters were tuned to balance bias and variance, with early stopping based on test set performance.

Model Performance

The XGBoost model achieved strong predictive performance, closely mirroring the MLP model in both accuracy and error structure:

- Training MAE: 1.90

- Training RMSE: 2.47
- Training R^2: 0.60
- Test MAE: 2.62
- Test RMSE: 3.30
- Test R^2: 0.23

These results demonstrate a slightly better fit and generalization compared to the MLP model. Forecast visualization over a 60-day horizon confirmed that XGBoost effectively captured the general direction and shape of trauma volume trends, although it, too, struggled to capture extreme high and low admission days (**Figure 46**).

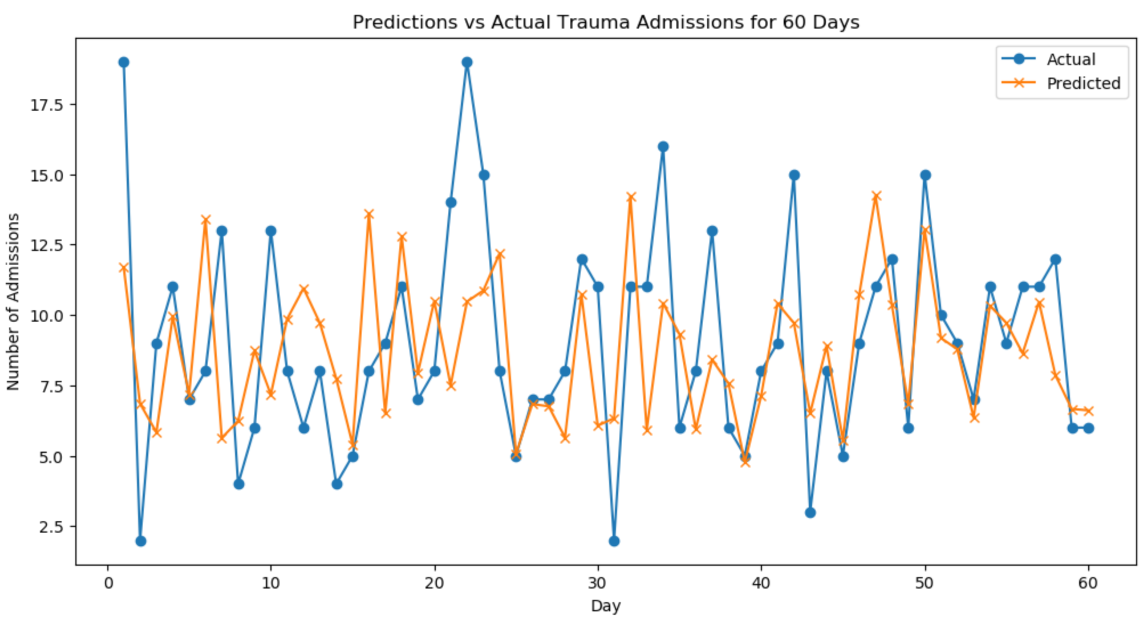


Figure 46: XGBoost Predictions 60 Days Horizon

Residual Distribution Analysis

Similarly to MLP, XGBoost's residual analysis showed distributions that were approximately normally distributed and centered around zero:

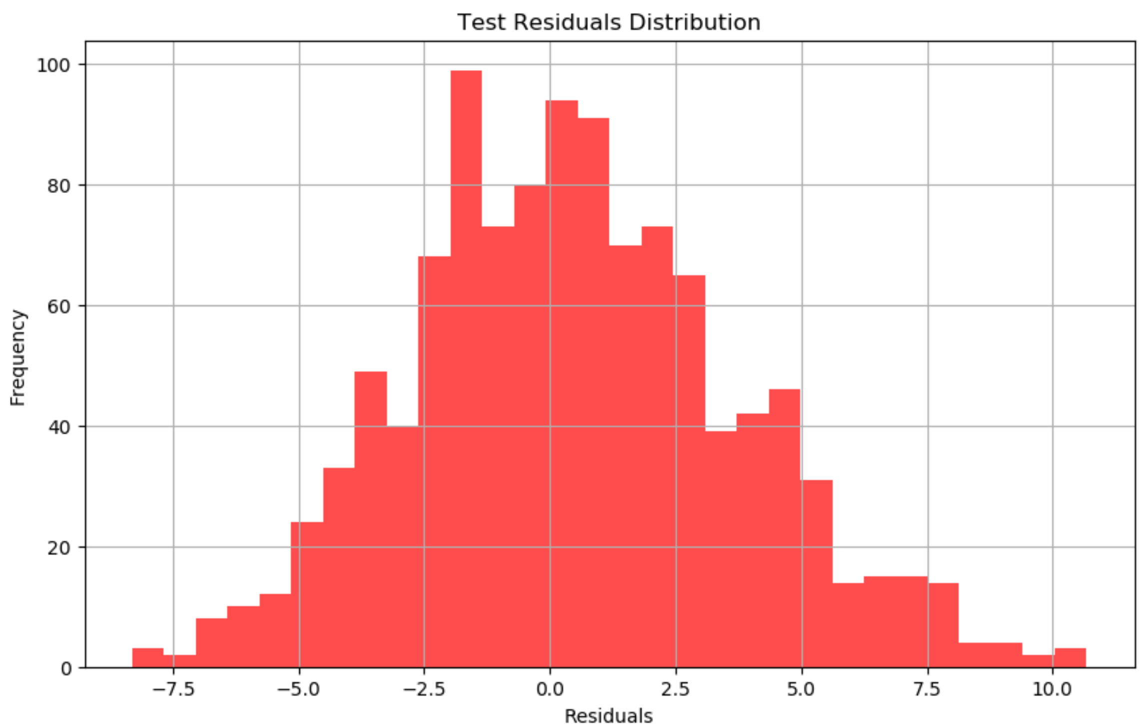


Figure 47: XGBoost Test Residual Distribution

Residuals were evaluated for distribution and normality:

- **Training residuals:**
  - Skewness: 0.70
  - Kurtosis: 1.16
  - Shapiro–Wilk p-value: 5.01×10^{-21}
- **Testing residuals:**
  - Skewness: 0.25
  - Kurtosis: –0.11
  - Shapiro–Wilk p-value: 0.00056

Q–Q plots of residuals again showed close alignment with the theoretical normal distribution within ±2 standard deviations, with only mild tail deviations (**Figure 48**).

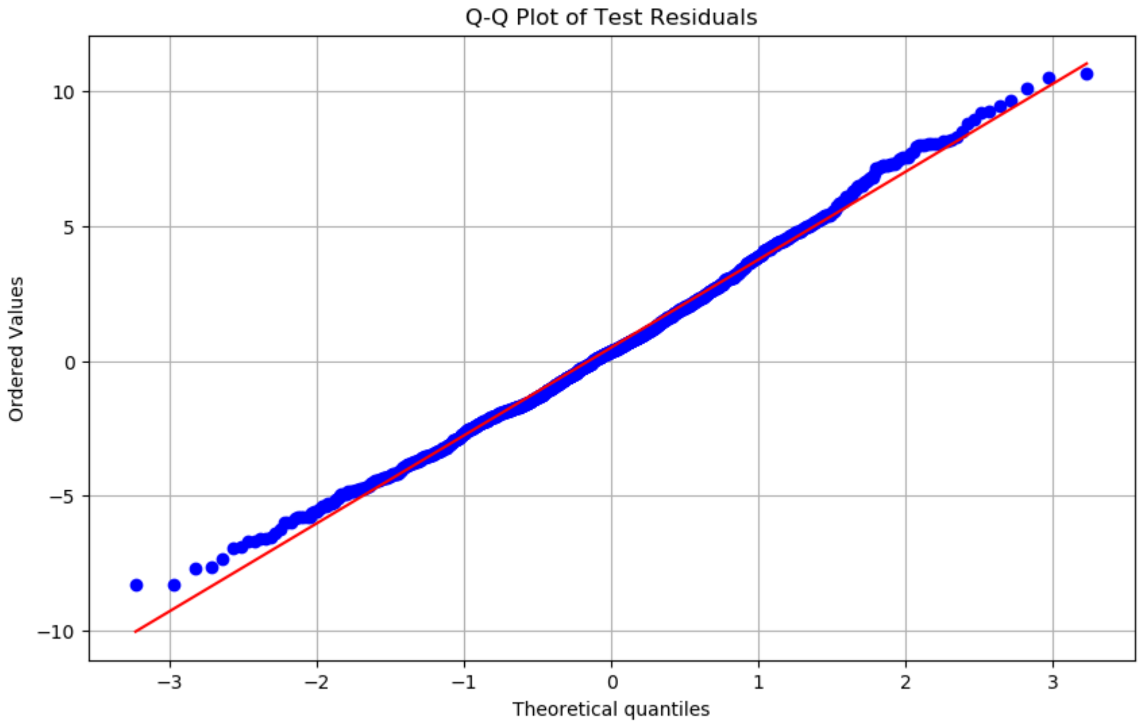


Figure 48: XGBoost Q-Q Test Residuals Plot

These findings mirror the residual characteristics of the MLP model and further validate the robustness of the predictions.

Feature Importance

XGBoost provides native feature importance scores, which we used to rank the relative contribution of each predictor. The results closely aligned with those from the MLP model, with the most important features including environmental predictors, such as average temperature, visibility, dew point, temporal patterns, including days off, holidays, weekend indicators and historical trauma indicators, such as 30-day rolling averages of GSWs, MCCs, MVCs, and falls

These results confirm that recent mechanism-specific trends and temporal signals are key predictors of daily trauma admissions (**Figure 49** shows the feature importance plot).

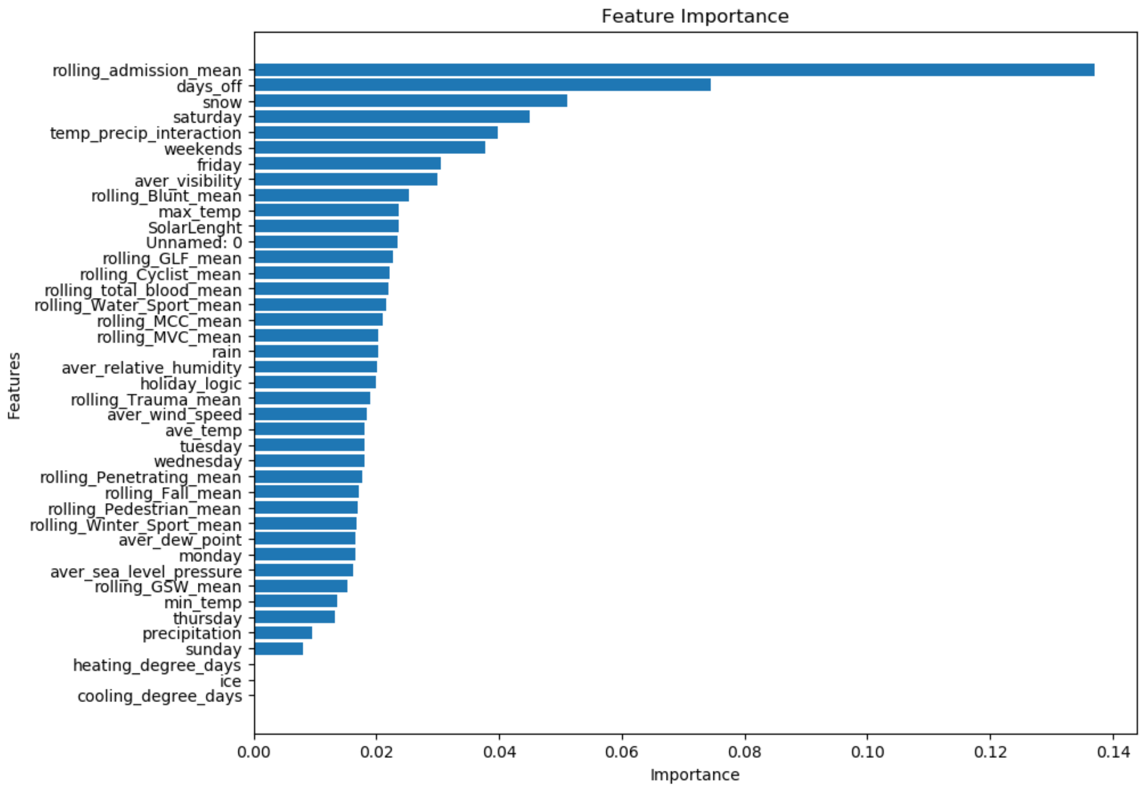


Figure 49: XGBoost Feature Importance

Summary

XGBoost slightly outperformed the MLP model across all metrics, particularly in terms of training fit and generalization to unseen data. Its strength lies in its ability to model complex nonlinear relationships and interactions between features without requiring data normalization or transformation. The residual patterns, feature priorities, and accuracy metrics suggest that XGBoost is a highly effective model for forecasting trauma admissions.

Model Performance Comparison

To synthesize the results across modeling approaches, Table X compares the predictive performance of the naïve baseline, Holt–Winters exponential smoothing, MLP, and XGBoost models. While classical statistical models captured overall trends and seasonal structure, they performed poorly on unseen data, particularly in terms of generalization (negative R^2 values on the test set).

In contrast, both MLP and XGBoost models demonstrated substantial improvements, with positive R^2 values on test data and lower error metrics across the board. XGBoost outperformed MLP slightly, achieving the lowest MAE and RMSE values and the highest explained variance (R^2) in both training and testing sets.

Table X. Performance Metrics of Forecasting Models

Model	Training MAE	Training RMSE	Training R^2	Test MAE	Test RMSE	Test R^2
Naïve Prediction	—	—	—	20.5	4.5	–0.38
Holt–Winters	4.78	5.02	–4.86	6.93	7.04	–17.99
MLP	2.43	3.08	0.38	2.69	3.42	0.17
XGBoost	1.90	2.47	0.60	2.62	3.30	0.23

Trauma-Related Blood Consumption Patterns

Between 2014 and early 2024, daily blood product usage in trauma patients at our center increased steadily, reflecting broader trends in trauma burden and resource intensity. This upward trajectory is visualized in **Figure 50**, which shows a consistent rise in blood consumption over the past decade.

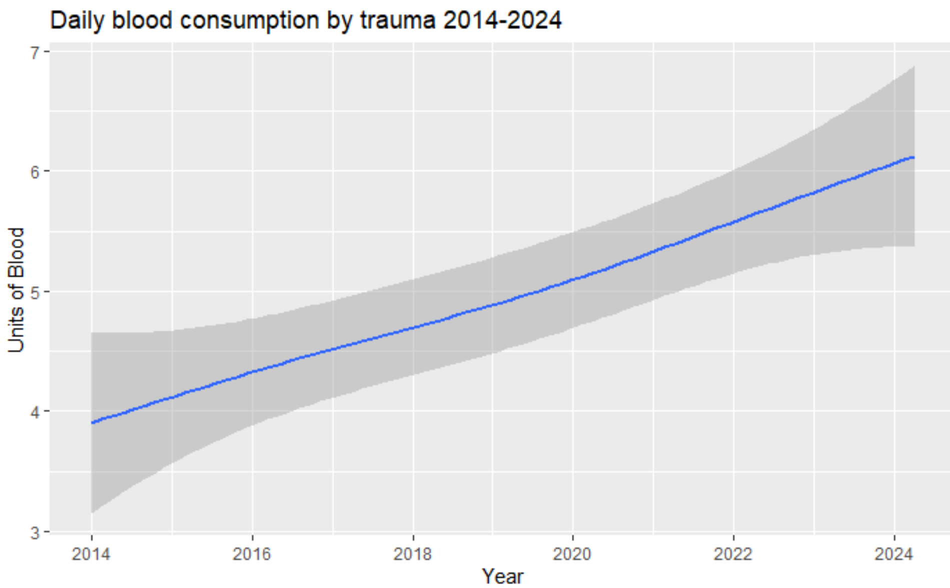


Figure 50: Consistent Rise in Blood Consumption

Annual Seasonality: Present but Subtle

Although transfusion volume follows a similar annual pattern to trauma admissions with increased usage during summer months, the seasonal amplitude is far less pronounced. As shown in **Figure 51** (monthly boxplot), the variation in transfusion volume across months is muted compared to admission volume, suggesting that severity of cases, rather than volume alone, drives transfusion trends.

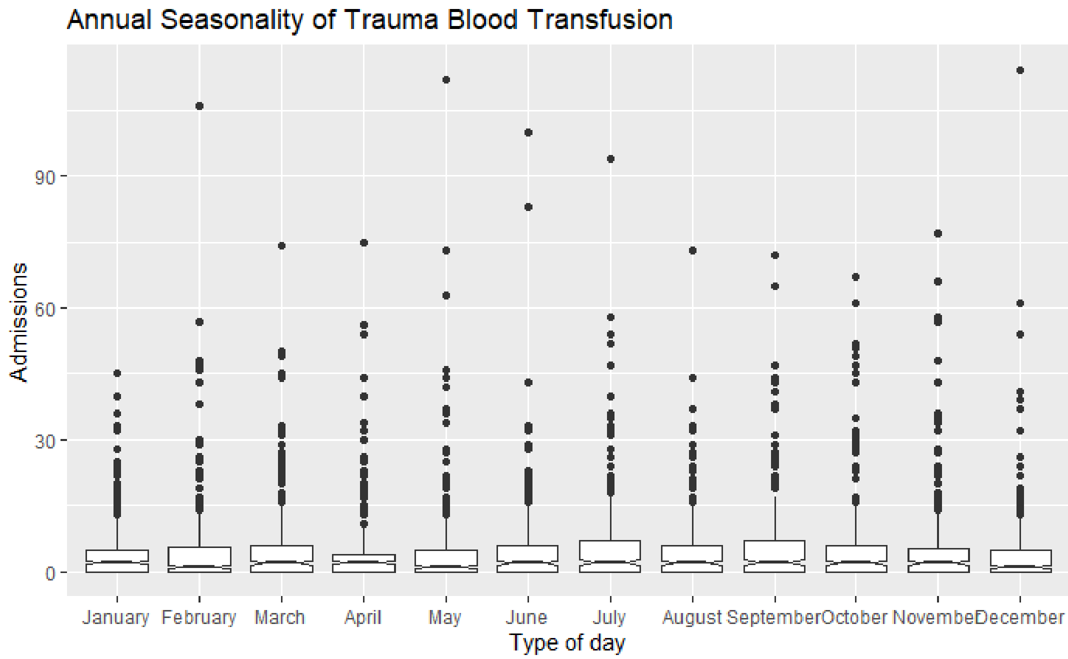


Figure 51: Annual Blood Consumption Seasonality Boxplot

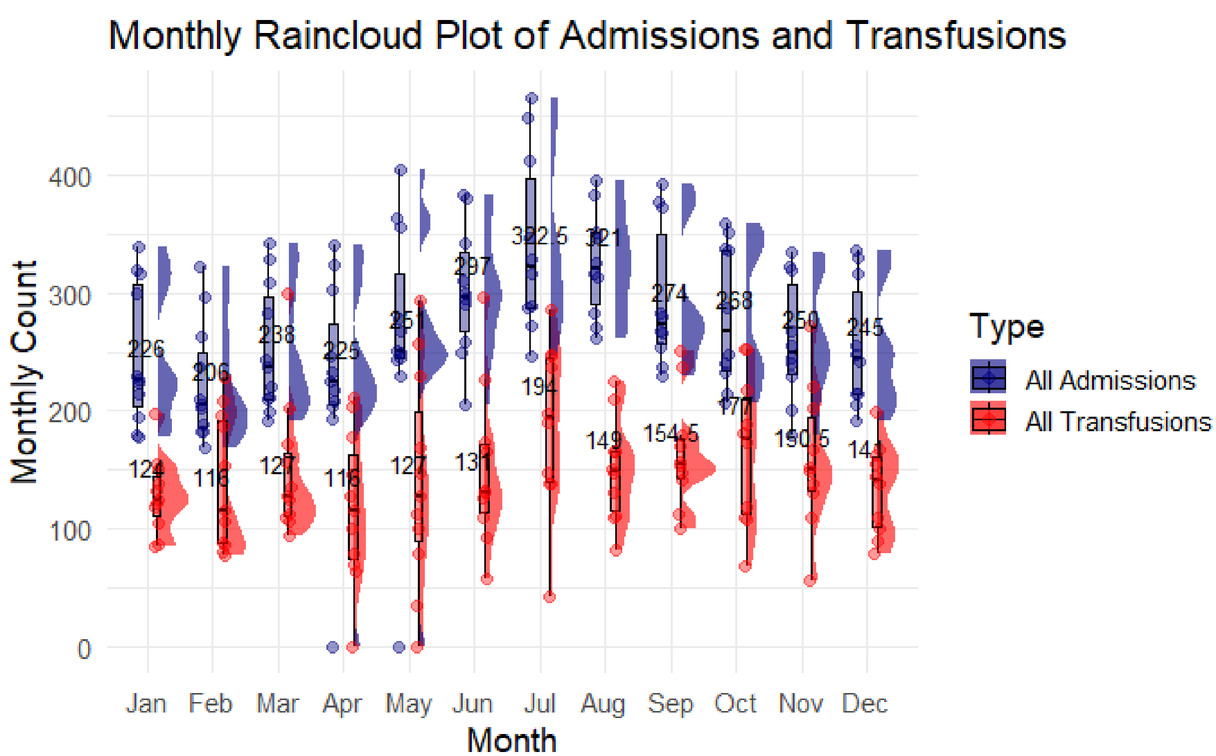


Figure 52: Raincloud Annual Admission and Transfusion Volume Trends

A raincloud plot (**Figure 52**) further illustrates this relationship, revealing sporadic bursts of high transfusion activity superimposed on a mild seasonal background. These bursts correspond to isolated clusters of severe trauma events rather than predictable patterns tied to calendar features and can be also illustrated by this gif plot:

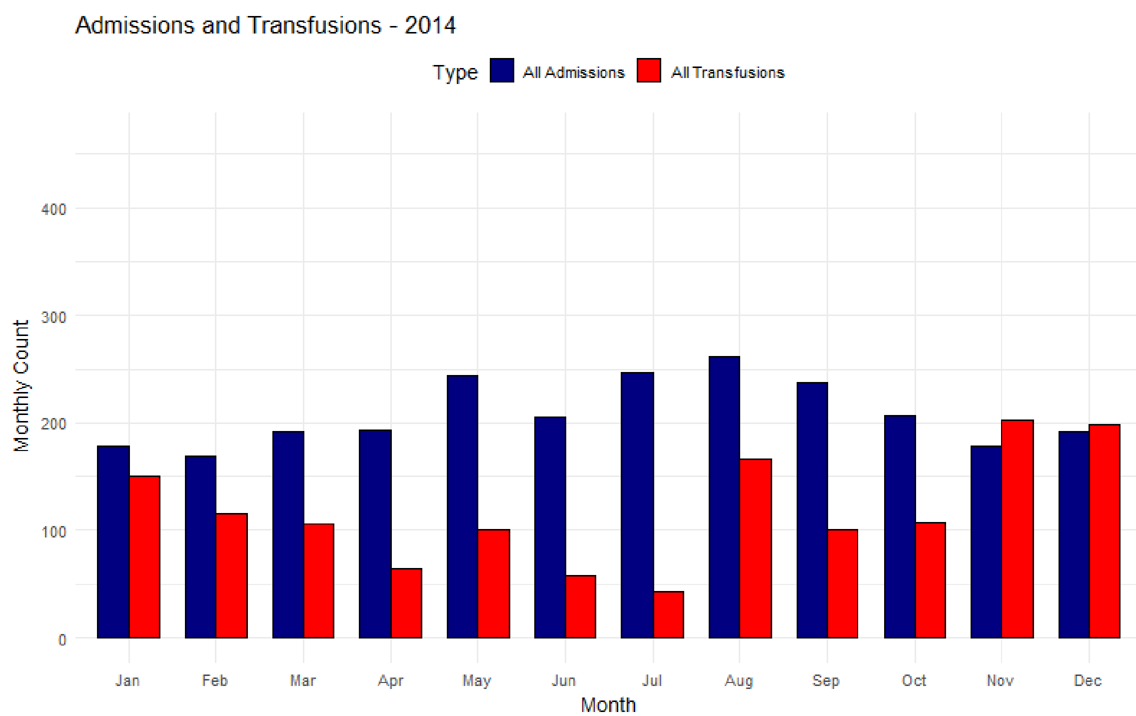


Figure 53: Admission and Transfusion GIF Plot

Weekly and Holiday Effects: Minimal

Unlike trauma admissions, which show a clear weekend effect, blood product utilization exhibits no significant weekly pattern. As shown in **Figure 54**, boxplots grouped by day of the week demonstrate relatively flat distributions. Similarly, there is no statistically meaningful difference in blood consumption between regular weekdays and designated days off or holidays. This reflects the heterogeneity in transfusion needs, which depend more on injury mechanism and acuity than on temporal patterns.

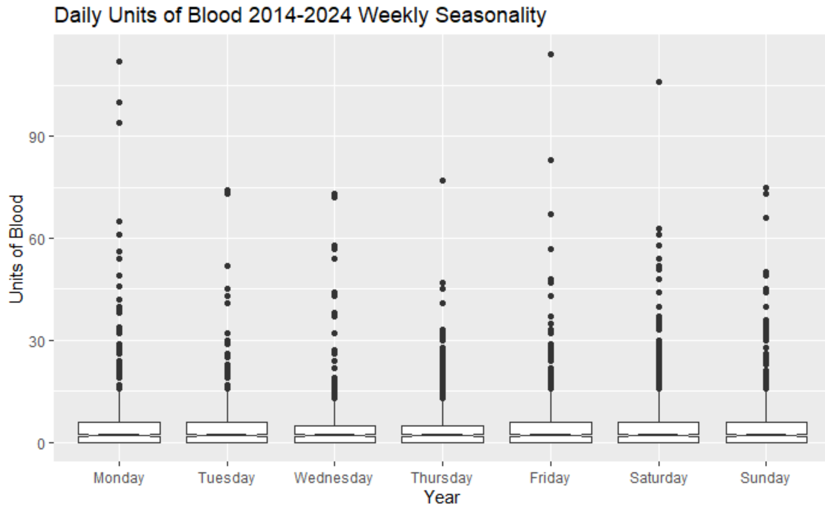


Figure 54: Weekly Transfusion Seasonality

Correlation with Admissions: Present but Weak for Forecasting

While a positive correlation exists between daily trauma admissions and blood transfusion volume (**Figure 55**), this relationship is highly variable.

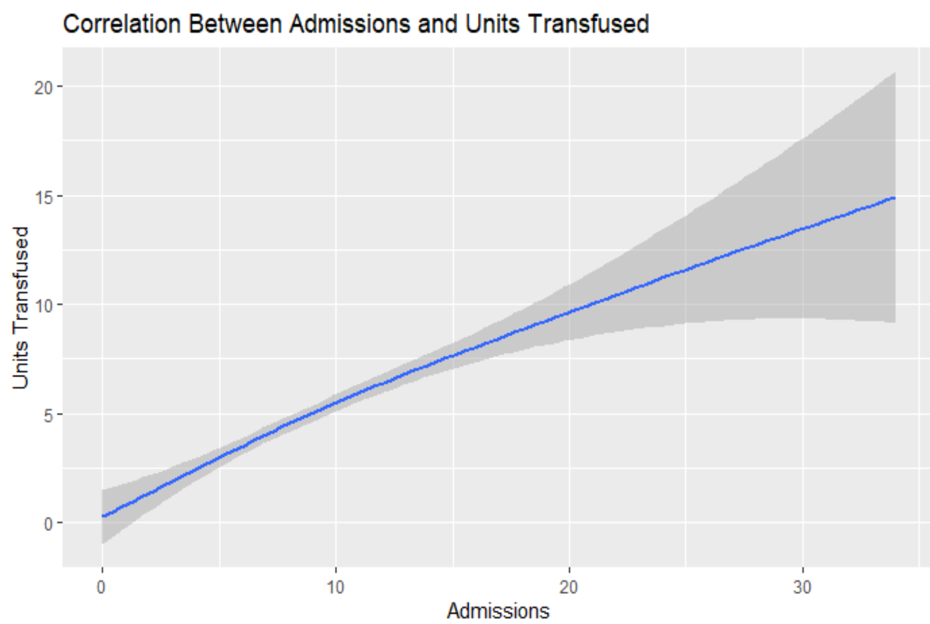


Figure 55: Admissions vs Units Transfused

A scatterplot with linear fit reveals many high-leverage outliers, representing days with disproportionate transfusion use relative to admissions volume—often due to rare but extreme injuries (e.g., mass casualty events, high-velocity penetrating trauma).

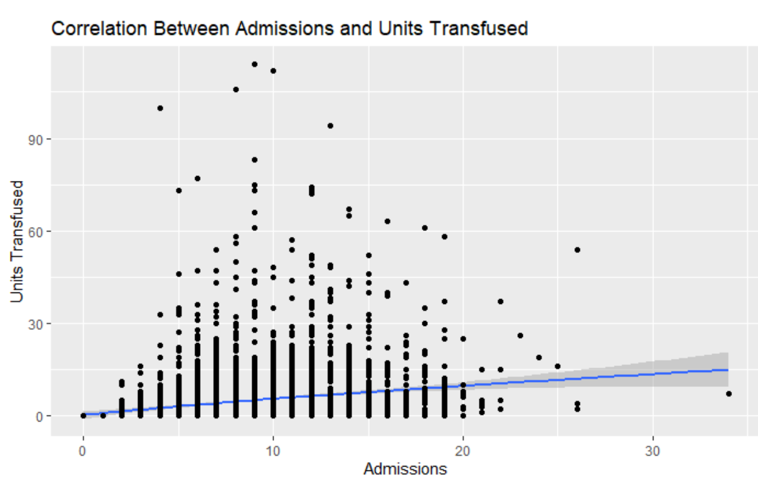


Figure 56: Admissions vs Transfusion with scatterplot

This high variance **(Figure 57)** severely limits predictive modeling of blood demand, even when using advanced methods such as XGBoost.

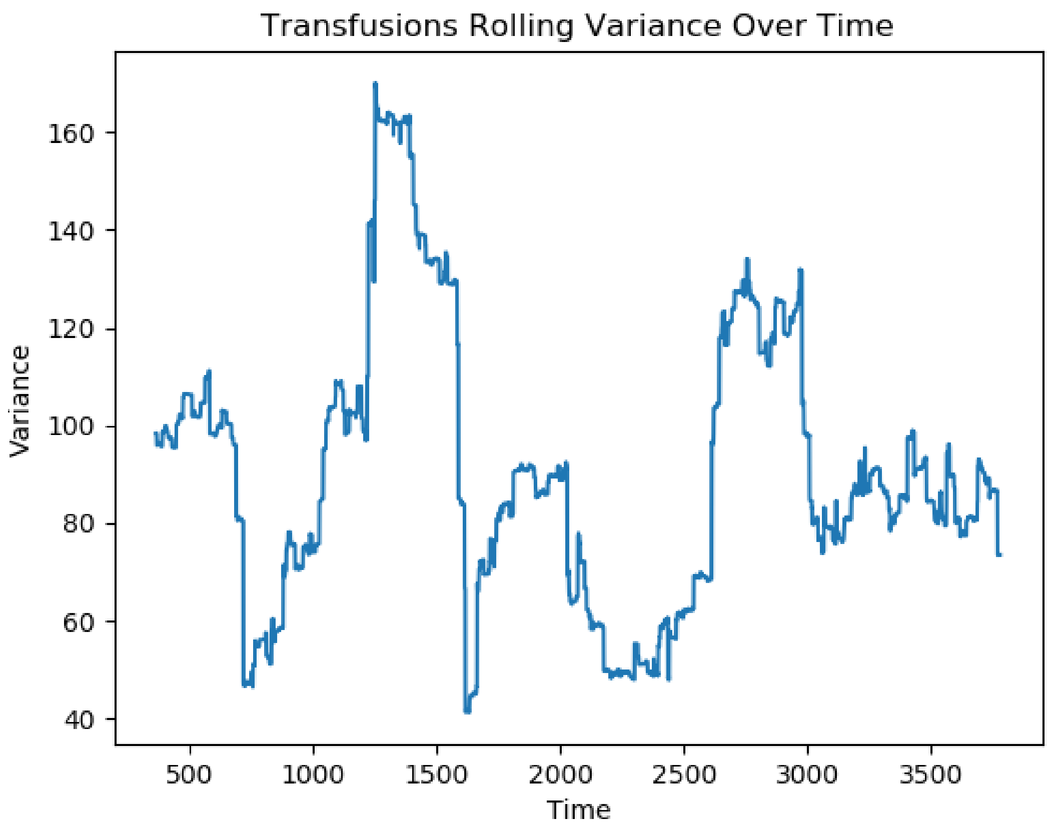


Figure 57: Transfusion Volume Rolling Variance

Despite incorporating environmental, temporal, and clinical features, the model performed poorly:

- Training MAE: 4.32
- Training RMSE: 9.85
- Training R^2: –0.08
- Test MAE: 4.54
- Test RMSE: 9.81
- Test R^2: –0.11

These results indicate that the model failed to capture meaningful signal and performed worse than a simple mean-based predictor. As illustrated in **Figure 58**, the 60-day forecast produced by the model fluctuates narrowly around the median transfusion level, failing to capture the magnitude of actual peaks and troughs. This underreaction to real-world variability further highlights the limitations of using global, aggregate-level modeling for predicting trauma-related blood demand.

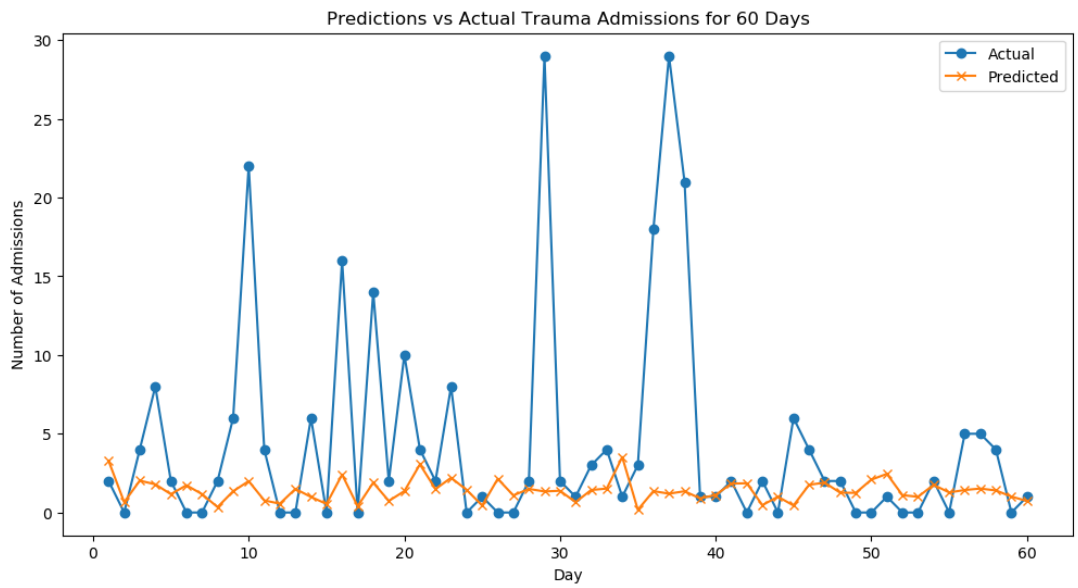


Figure 58: XGBoost 60-day Forecast

Residuals: Non-Normal and Highly Skewed

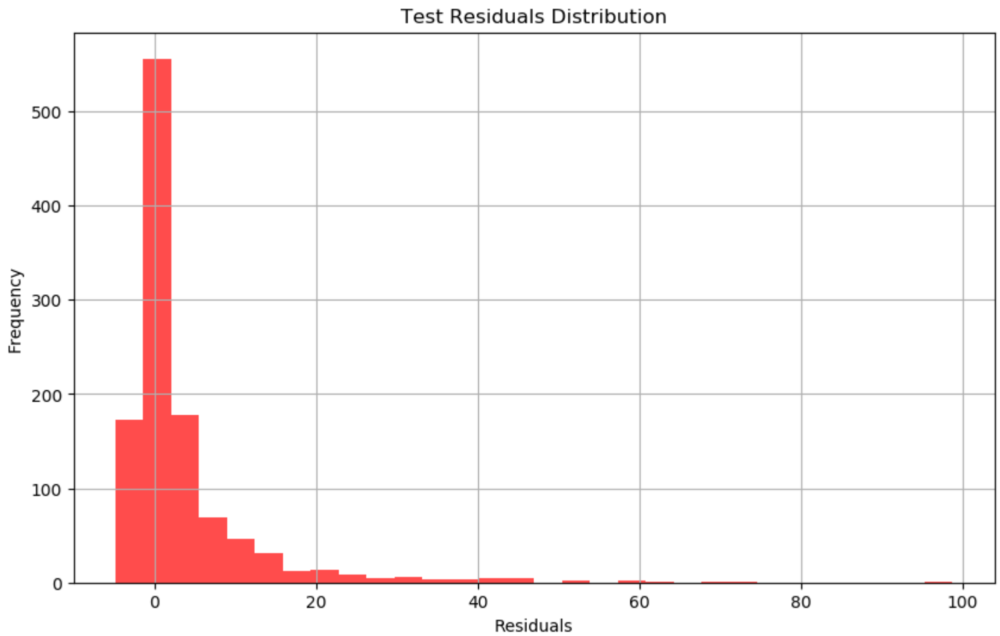


Figure 59: Test Residuals Distribution

Model diagnostics further confirm the unreliability of transfusion forecasting with global models. Residuals were highly right-skewed, with extreme kurtosis values:

- Train Skewness: 4.95
- Train Kurtosis: 35.99
- Test Skewness: 4.13
- Test Kurtosis: 23.62

Both training and test residuals failed the Shapiro-Wilk test for normality ($p < 0.001$), and the corresponding Q–Q plot was U-shaped, indicating severe deviation from normal distribution (**Figure 60**).

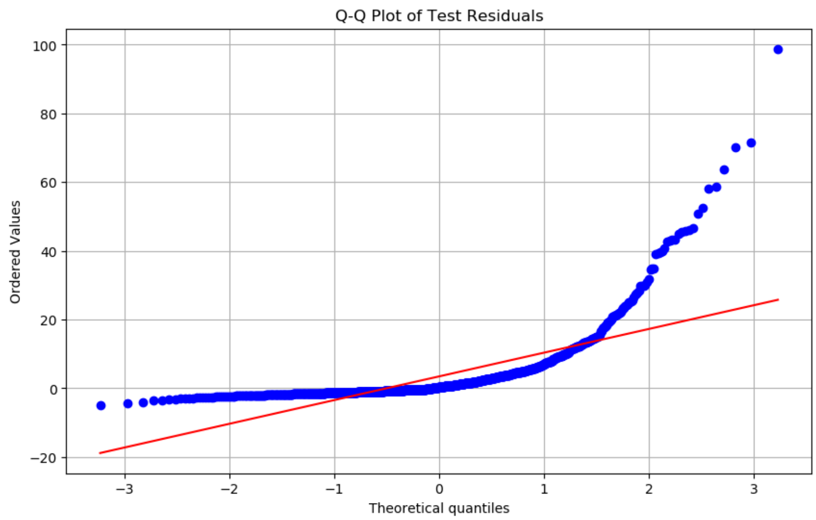


Figure 60: Test Residuals Q-Q Plot

Conclusion

While trauma admissions and blood product usage are directionally related, blood consumption exhibits more irregularity, outlier sensitivity, and mechanism-driven variability. These characteristics significantly limit the effectiveness of aggregate-level forecasting models. In the next section, we explore whether mechanism-specific modeling can improve performance by addressing this heterogeneity directly.

Trauma Heterogeneity and the Case for Mechanism-Specific Blood Demand Modeling

A key limitation in forecasting trauma-related blood demand lies in the underappreciated heterogeneity of trauma mechanisms. Trauma encompasses a wide spectrum of injuries, each with distinct physiological, temporal, and resource-use profiles. Attempts to predict transfusion needs using aggregate data obscure these differences and lead to poor model generalizability.

To better understand this heterogeneity, we analyzed the annual seasonality of transfusion needs across common mechanisms of injury.

Mechanism-Specific Seasonality Patterns

Gunshot Wounds (GSWs): As shown in Figure W, GSW-related transfusions show minimal annual seasonality but according to our statistical analysis exhibit a strong weekly effect, peaking on weekends and holidays. These injuries are typically associated with high transfusion volumes, despite low frequency.

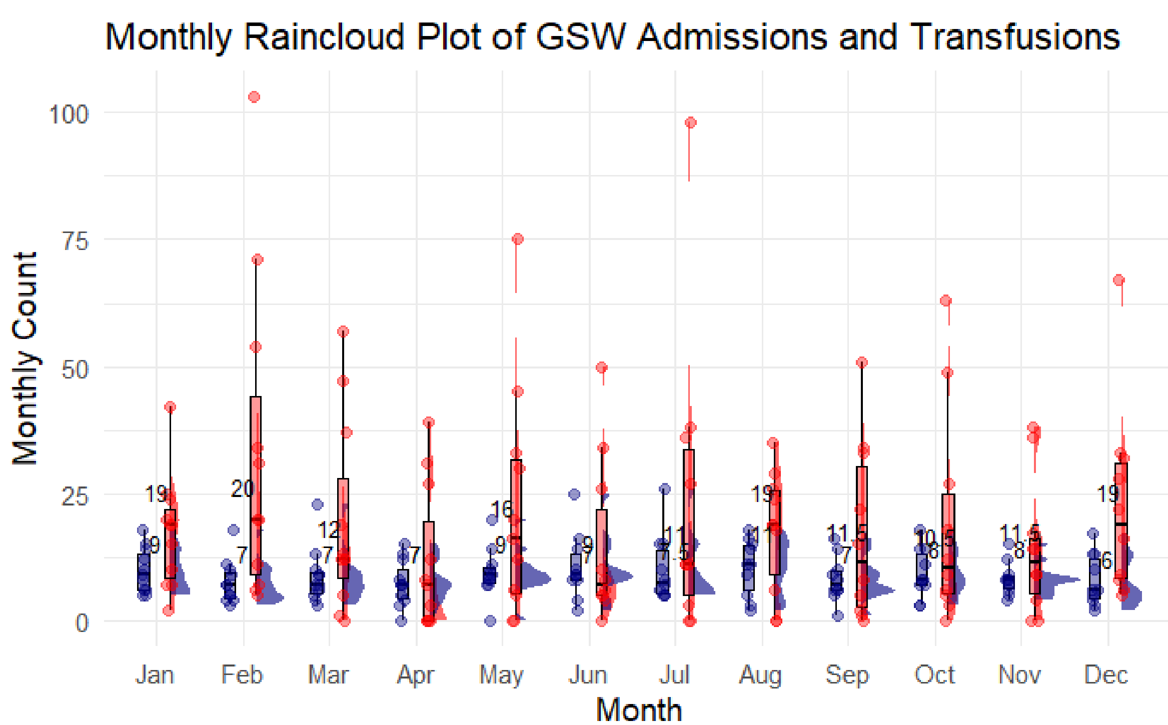


Figure 61: GSW Incidence vs Transfusion Requirements

Motorcycle Collisions (MCCs): Figure X illustrates a pronounced annual peak in summer months, consistent with seasonal increases in motorcycle use. These events are often associated with polytrauma and moderate to high transfusion needs.

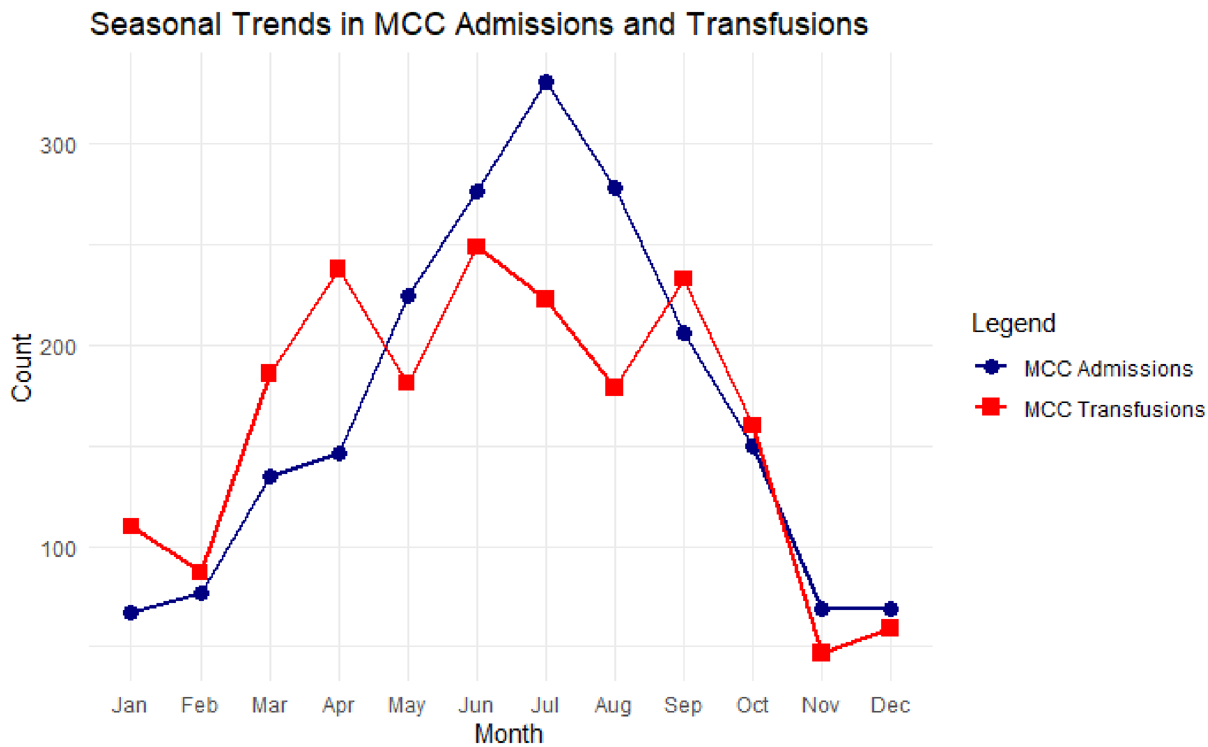


Figure 62: MCC Incidence vs Transfusion Requirements

Motor Vehicle Collisions (MVCs): As seen in Figure 63, MVCs follow a broad annual trend, peaking in summer and tapering in winter, though to a lesser degree than MCCs. They contribute a substantial portion of moderate transfusion volume due to their high frequency.

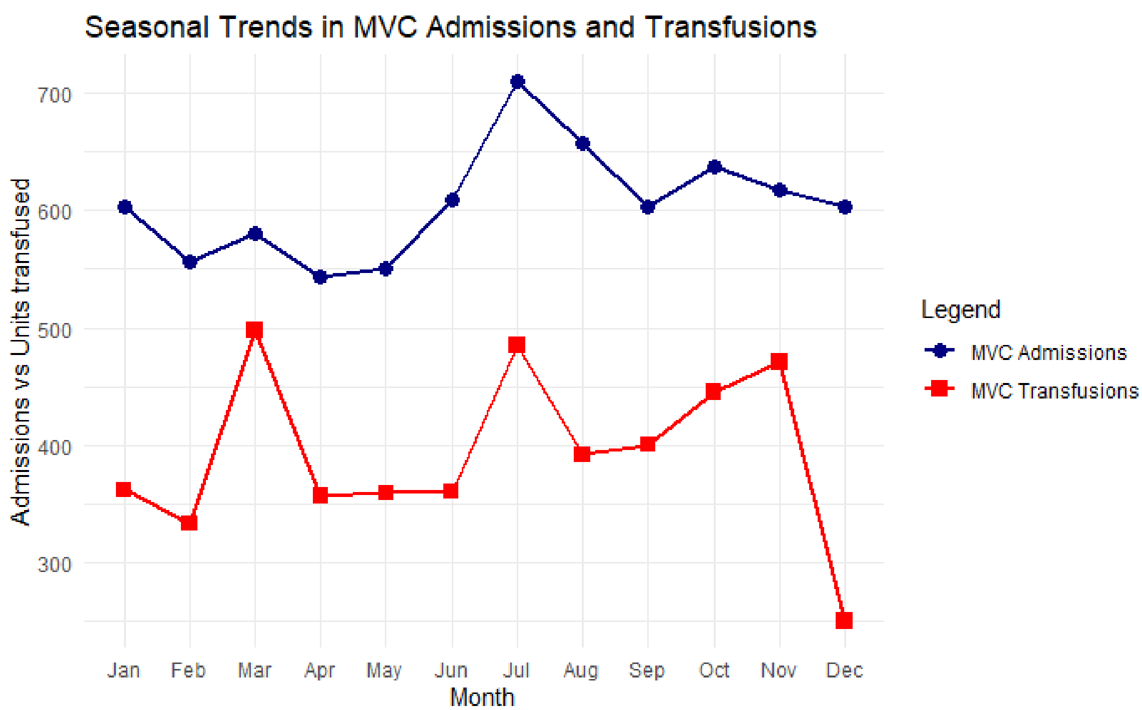


Figure 63: MVC Incidence vs Transfusion Requirements

Ground-Level Falls (GLFs): Figure 64 shows that GLFs are evenly distributed across the year with low transfusion requirements. These are often geriatric injuries with minimal hemorrhagic risk and represent volume without acuity.

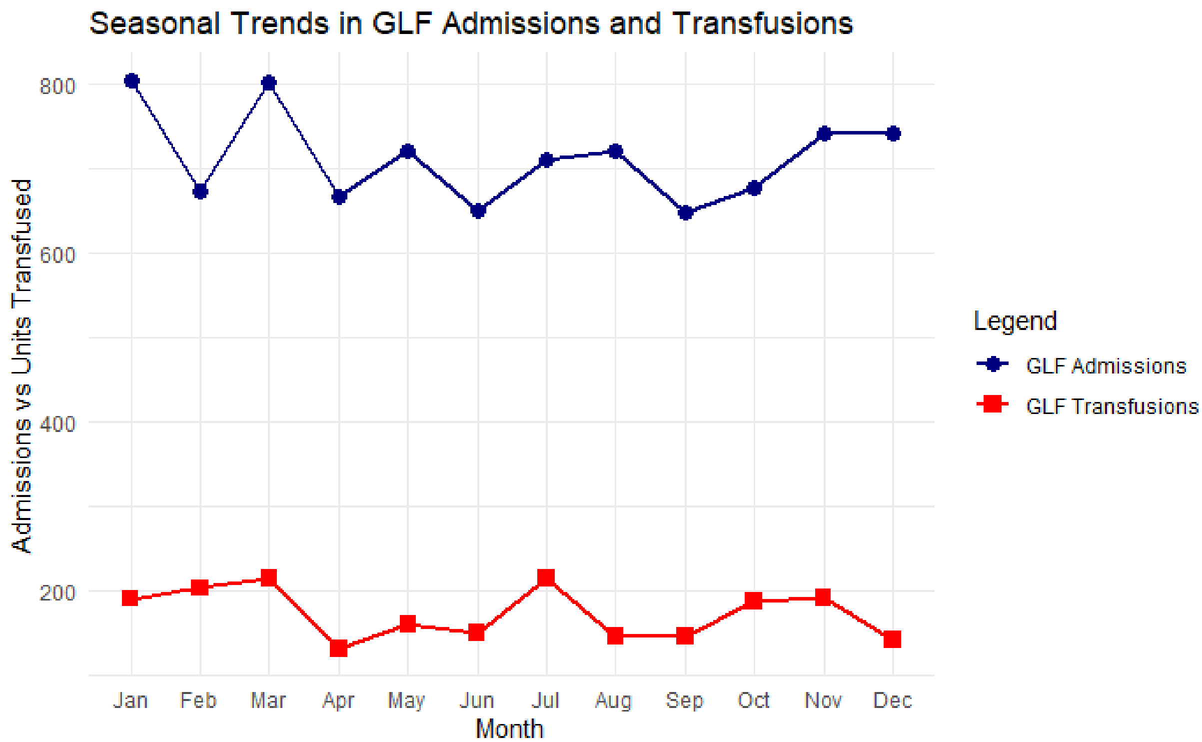


Figure 64: GLF Incidence vs Transfusion Requirements

These patterns highlight that transfusion seasonality is not uniform, but mechanism-dependent. GSWs defy annual seasonality, GLFs require few transfusions, and vehicular collisions demonstrate varying degrees of predictable seasonal trends.

Conclusion

The complexity and variability of trauma care, particularly in transfusion needs, underscore the limitations of aggregate, one-size-fits-all forecasting models. Our findings reveal that while trauma admissions and blood product usage share broad seasonal patterns, the predictive relationship between them is highly inconsistent due to significant heterogeneity in injury mechanisms and clinical severity.

Given this variability, future modeling efforts should shift toward mechanism-specific predictive approaches. These models would be trained using inputs such as mechanism-specific admission rates, rolling historical trends (e.g., 7-day or 30-day averages), environmental and temporal features (weather, daylight, day-of-week), and clinical indicators such as predicted average injury severity score and or penetrating vs blunt trauma, as has been previously predicted by Dennis and Stonko[119-121].

By aligning model structure with the distinct profiles of different trauma types, might be able to improve forecast accuracy for high-risk, high-transfusion mechanisms (e.g., GSWs, MCCs), reduce the impact of noise from low-acuity, low-transfusion cases (e.g., GLFs), enable the generation of mechanism-targeted alerts, supporting more efficient blood inventory planning and trauma team preparedness.

This tailored approach will better reflect the real-world clinical landscape, providing actionable insights that improve resource utilization and patient care in high-stakes trauma environments.

Chapter 6: Discussion

In this dissertation, we investigated the problem of blood transfusion prediction across multiple dimensions—from individual clinical decision-making in the trauma bay to system-level forecasting of trauma admissions and blood product demand. We approached the issue using a mixed-methods framework that combined qualitative inquiry and quantitative analysis with advanced machine learning and deep learning methodologies. This multidimensional exploration reflects the inherent complexity of trauma care and the critical importance of timely transfusion decisions in saving lives and preserving limited blood resources.

Key Findings

Three overarching insights emerged from this work:

1. **Decision-making in the trauma bay can be augmented**: We found that transfusion decisions often begin as intuitive, fast-acting judgments—characteristic of *System 1* thinking—driven by the trauma surgeon's perception, experience, and limited early data. While this heuristic process is essential in the acute setting, our qualitative and quantitative findings suggest that it can be effectively augmented through real-time access to structured, visualized, and analytically derived data. Tools like visual dashboards, computerized check-lists, and AI-powered clinical decision support can enhance *System 1* decisions with the precision of *System 2* analysis—bringing data-driven rigor to high-stakes moments.
2. **Clinical and operational data infrastructure is sufficient for AI-based modeling**: We demonstrated that well-managed EHR data, integrated with trauma registry information, offers a powerful substrate for predictive modeling. Our machine learning models—whether aimed at patient-level transfusion needs or forecasting hospital-wide trauma admissions—show that institutional data assets are mature enough to support the development of accurate, scalable, and clinically relevant AI solutions. These models offer not just improvements in individual patient care, but a path toward institutional readiness and operational efficiency.

3. **Fusion models**: The development of a joint fusion model combining structured clinical data with chest X-ray imaging outperformed individual modality models, highlighting the value of multimodal integration in transfusion prediction.
4. **Trauma heterogeneity necessitates a new modeling paradigm:** Our findings reveal that trauma is not a monolith, but an assortment of mechanisms of injury that differ in their patterns, transfusion demands, and seasonal variation. This heterogeneity demands a shift away from monolithic, aggregate-level prediction models and toward mechanism-specific modeling, which can better reflect the underlying dynamics of blood consumption and trauma care.

Alignment with Study Aims

These findings directly support the achievement of all four research aims:

Aim 1: Understanding the decision-making process

Through surveys and in-depth semi-structured interviews with trauma attendings, we characterized the clinical logic and variables that influence transfusion decisions. Our thematic analysis revealed key data elements (e.g., GCS, systolic BP, shock index, patient appearance) that clinicians rely on and validated their statistical significance via regression analysis.

Aim 2: Building ML models to identify transfusion needs

We developed and tested multiple ML models—including MLPs, CNNs, and a novel joint fusion model. The fusion model demonstrated strong sensitivity and interpretability, especially when paired with data augmentation and SMOTE, outperforming standalone approaches and approximating benchmark performance reported in the PROPPR trial.

Aim 3: Predicting admissions and blood demand at a system level

Using 10 years of trauma and weather data, we constructed predictive models of trauma volume and blood consumption. While trauma admissions could be forecasted with reasonable accuracy using MLP and XGBoost models, the prediction of blood usage proved more challenging due to the erratic and mechanism-specific nature of high-consumption events.

Aim 4: Evaluating model performance

We assessed models using a variety of metrics (AUC, Youden's J, MAE, RMSE, residual distribution) and explainability tools (e.g., Grad-CAM for CNNs). The results underscore the importance of calibration, class balancing, and multimodal integration—while also revealing limitations and directions for further optimization.

Implications for Clinical and Operational Practice

This work suggests that AI can play a transformative role in trauma transfusion decision support if designed with clinical reality in mind. Predictive tools must align with existing workflows, communicate information visually and intuitively, and adapt to the variability of trauma presentations. Moreover, they must move beyond "one-size-fits-all" approaches and tailor their logic to the mechanism, timing, and context of injury.

At the operational level, our findings support the feasibility of using ML models to support staffing and maybe institutional blood bank planning, if blood usage is more monotonous and predictable than in our center. Mechanism-specific seasonally informed forecasting, and adaptive inventory management systems could help trauma centers better prepare for surges in trauma admissions, while minimizing wasteful practices.

Chapter 7: Summary and Conclusion

Recent advances in scientific computing, from deep learning algorithms to GPU-accelerated infrastructure and institutional computing centers, have allowed artificial intelligence (AI) to expand into multiple areas of medicine, including ophthalmology, endocrinology, cardiology, and surgery. However, trauma care, despite being one of the most time-sensitive and resource-intensive domains, remains underexplored by AI research. Hemorrhagic shock is a leading cause of preventable death in trauma, especially among younger populations, and the inherent chaos, urgency, and cognitive overload in trauma bays make it an ideal setting for clinical decision support.

This dissertation explored blood transfusion prediction from both the patient-level and system-level perspectives. We began by examining how decisions are made in the trauma bay, showing that the initial, fast-paced clinical reasoning (System 1) can be augmented by timely, analytically driven tools. By pairing qualitative findings with statistical correlations, we identified specific variables, such as shock index, mechanism of injury, lactate levels, and mental status that strongly influence transfusion decisions, and highlighted opportunities to improve decision-making through structured data visualization and checklists.

We then demonstrated the feasibility of applying machine learning to create a fusion model that integrates structured clinical data and chest X-rays. This approach improved predictive sensitivity while maintaining interpretability using Grad-CAM heatmaps, offering a promising method for early identification of patients likely to require massive transfusion. Additionally, we developed a forecasting system for trauma admissions, incorporating weather, temporal, and mechanism-based predictors. These models successfully captured annual and weekly trends and offer potential utility in operational planning and blood inventory management.

However, attempts to forecast daily blood product usage on an institutional level showed limited success. Despite robust models for admissions trends, the irregularity of transfusion events driven heterogeneity of trauma population introduced high variance and poor predictive performance. Upon closer analysis, we identified trauma heterogeneity, particularly the variation by mechanism of injury, as a major barrier to reliable system-level forecasts. Our analysis suggests that disaggregating modeling efforts by mechanism may offer a more accurate and clinically actionable path forward.

In conclusion, we present three major contributions:

1. A new method to augment heuristic decision-making in trauma care with structured, analytical data;
2. A joint fusion model combining imaging and clinical variables to identify transfusion needs at the point of care; and
3. A system-level forecasting framework for trauma admissions using environmental and temporal signals.

While we did not achieve reliable institution-wide prediction of blood demand, we uncovered a key challenge: the heterogeneity of trauma and offered a pathway for future research using mechanism-specific modeling. Taken together, these findings point toward a future where machine learning can enhance not only clinical accuracy at the bedside but also operational preparedness across the trauma care continuum.

Appendix 1: Qualitative questionnaire

Questionnaire:

1. Are you an Attending Trauma Surgeon?
    a. Yes – will proceed with questionnaire.
    b. No – will skip the questionnaire and proceed to the "thank you" screen.

2. Hello dear Attending Trauma Surgeon!

    This research is related to discovering and analyzing the clinical decision-making process, related to blood transfusion in trauma victims, and will be conducted as an on-line questionnaire. The questionnaire will take about 5-10 minutes of your time.

    Your answers will be coded and analyzed in order to establish the decision-making process related to initiation or discontinuation of blood transfusion. If you are interested to participate, please mark "yes" below. If you are disagree with these terms and conditions, please reply "no".

    Also if you are willing to participate in a follow-up interview, please answer the corresponding question in the questionnaire and I will contact those, who are willing to share their expertise to schedule an interview.

Thank you, Mike Kolesnikov,

Trauma NP, PhD-Candidate,

Principal investigator IRB # STUDY00024780

a. Yes – will proceed with questionnaire
b. No – will skip the questionnaire and proceed to the "thank you" screen

3. Are you able to identify the need for blood transfusion in a trauma patient before they arrive to the resuscitation bay?
    a. Yes -> proceed to questions 4 & 5
    b. No -> skip questions 4 & 5
    c. Sometimes -> proceed to questions 4 & 5

4. What information was most useful to you?
    a. Text answer

5. When you were wrong, what information led you to not transfuse?
    a. Text answer

6. When deciding about a blood transfusion:
    a. I prefer to use a heuristic model, such as use of a combination of vitals, history, and physical signs. These are the things I consider when making this decision:

i. Mechanism of Injury

ii. Systolic Blood Pressure (SBP)

iii. Heart Rate (HR)

iv. Glasgow Coma Scale Score (GCSS)

v. Presence of long bone fracture(s)

vi. Reported Estimated Blood Loss (EBL)

vii. Other: (Text answer)

viii. None of the above

b. I prefer to use a formal analytical model with formalized scoring system or indexes, such as (mark all scoring systems you are using)

i. Assessment of Blood Consumption (ABC),

ii. Shock Index (SI),

iii. Trauma Associated Severe Hemorrhage Score (TASHS),

iv. Emergency Transfusion Score (ETS)

v. McLaughlin MT Scoring System

vi. Prince of Wales Hospital (PWH) Scoring system

vii. Trauma Bleeding Severity Score (TBSS)

viii. Other: (text answer)

ix. None of the above

c. I use a combination of formal and heuristic models:

i. Mechanism of Injury

ii. Systolic Blood Pressure (SBP)

iii. Heart Rate (HR)

iv. Glasgow Coma Scale Score (GCSS)

v. Presence of long bone fracture(s)

vi. Reported Estimated Blood Loss (EBL)

vii. Other: (Text answer)

viii. Assessment of Blood Consumption (ABC),

ix. Shock Index (SI),

x. Trauma Associated Severe Hemorrhage Score (TASHS),

xi. Emergency Transfusion Score (ETS)

xii. McLaughlin MT Scoring System

xiii. Prince of Wales Hospital (PWH) Scoring system

xiv. Trauma Bleeding Severity Score (TBSS)

xv. Other: (text answer)

xvi. None of the above

d.

7. Have you ever had a patient who came into the resuscitation bay and needed a blood transfusion unexpectedly?
   a. Yes -> goes to question 8
   b. No -> goes to question 9

8. Was any information missing from the report that would have helped you to identify the need for blood transfusion sooner?
    a. Text answer

9. When you transfuse less blood than was initially ordered, what information helps you to decide that you need to discontinue the blood transfusion?
    a. Any specific events, lab tests, vital signs, gestalt? (text answer)

10. What information do you wish you had from the field or predictive analytics tools, to determine the need for blood transfusion? (text answer)

11. Would you be able to refer me to another trauma attending, who may be interested in participating in this research?
    a. Yes
    b. No -> skips name, contact information, and best time to contact

    c. Name
    d. Contact information
    e. Best time to contact.

12. May I contact you with a short follow-up interview?

    a. Yes
    b. No -> skips name, contact information, and best time to contact

c. Name

d. Contact information

e. Best time to contact

Appendix 2: Interview Guide

Hi, my name is Mike, I am doing research on blood transfusion in context of decision making, so I will ask you a few questions regarding blood transfusions.

1. In what situations do you use blood transfusion in trauma patient population?

2. Were you ever able to identify a patient who will need a blood transfusion before they were brought by an EMS?

    a. If #2 is a yes: What information did you have at that time and what was the meaning of it? Did you ever change your mind about blood transfusion after the patient arrived?

        i. (If mind changed) Can you tell more about that?

    b. If # 2 is a no: What was the earliest, you have ever identified a patient in need of blood transfusion? Can you please describe how you have made this decision?

3. Did you ever have a decision about starting a blood transfusion coming to you as a surprise?

    a. If #2 is a yes: please describe what happened, why were you surprised?

b. If # 2 is a no: What was the latest, you have identified a patient in a need of blood transfusion? In hindsight, could it be earlier? What information did you need to make this decision sooner?

4. Now, let's talk about stopping a blood transfusion. Could you please tell me how do you make the decision of discontinuing a blood transfusion?
    a. If the answer is too brief: What specific information helps you to make this decision?
        i. Follow-up: Are you waiting for specific lab values, or do you follow physical signs, or completion of damage control surgery?

5. If you had a magic wand, what information would help you to predict a need of blood transfusion at any point either before the patient arrives or before the need for transfusion is obvious?

    a. If interviewee says "I don't know" or answer is too brief:
        i. Do you feel knowing the dynamics of changes in vital signs during transportation would be helpful?
        ii. Do you think being able to predict vital signs upon arrival could help?
        iii. Do you think patient's age and size matters?
        iv. Are there any imaging studies you wish the transfer patient had before coming, that could help you identify a patient at risk?

6. What else do you think is important in make the decision of initiation or stopping a blood transfusion?

7. Can you refer me to another trauma surgeon, who routinely makes decisions about blood transfusion and might be interested to participate in this research?

References:

1. Robertson OH, Bock AV. Blood Volume in Wounded Soldiers : I. Blood Volume and Related Blood Changes after Hemorrhage. *J Exp Med.* 1919;29(2):139-153.
2. Landsteiner K. Individual Differences in Human Blood. *Science.* 1931;73(1894):403-409.
3. Landsteiner K. [Agglutination phenomena of normal human blood]. *Wien Klin Wochenschr.* 2001;113(20-21):768-769.
4. Czerwinski M, Kaczmarek R, Glensk U. Ludwik Hirszfeld: A pioneer of transfusion and immunology during the world wars and beyond. *Vox Sang.* 2022;117(4):467-475.
5. Benjamin AJ, Young AJ, Holcomb JB, et al. Early Prediction of Massive Transfusion for Patients With Traumatic Hemorrhage: Development of a Multivariable Machine Learning Model. *Ann Surg Open.* 2023;4(3):e314.
6. Abuzeid AM, O'Keeffe T. Review of massive transfusion protocols in the injured, bleeding patient. *Curr Opin Crit Care.* 2019;25(6):661-667.
7. Alser O, Dorken-Gallastegi A, Proano-Zamudio JA, et al. Using the Field Artificial Intelligence Triage (FAIT) tool to predict hospital critical care resource utilization in patients with truncal gunshot wounds. *Am J Surg.* 2023;226(2):245-250.
8. Strickland M, Nguyen A, Wu S, et al. Assessment of Machine Learning Methods to Predict Massive Blood Transfusion in Trauma. *World J Surg.* 2023;47(10):2340-2346.
9. Avital G, Gelikas S, Radomislensky I, et al. A prehospital scoring system for predicting the need for emergent blood product transfusion. *Transfusion.* 2021;61 Suppl 1:S195-S205.
10. Jadwin DF, Fenderson PG, Friedman MT, et al. Determination of Unnecessary Blood Transfusion by Comprehensive 15-Hospital Record Review. *Jt Comm J Qual Patient Saf.* 2023;49(1):42-52.

11. El-Menyar A, Mekkodathil A, Abdelrahman H, et al. Review of Existing Scoring Systems for Massive Blood Transfusion in Trauma Patients: Where Do We Stand? *Shock.* 2019;52(3):288-299.
12. Euripides. *Medea.* Penguin Classics; 1963.
13. Shakespeare W. *King Lear.* JB Lippincott & Co; 1880.
14. Hirschfeld L, Hirschfeld H. SEROLOGICAL DIFFERENCES BETWEEN THE BLOOD OF DIFFERENT RACES.: THE RESULT OF RESEARCHES ON THE MACEDONIAN FRONT. *The Lancet.* 1919;194(5016):675-679.
15. Cusack L, De Buck E, Compernolle V, Vandekerckhove P. Blood type diets lack supporting evidence: a systematic review. *Am J Clin Nutr.* 2013;98(1):99-104.
16. Wang J, Jamnik J, García-Bailo B, Nielsen DE, Jenkins DJA, El-Sohemy A. ABO Genotype Does Not Modify the Association between the "Blood-Type" Diet and Biomarkers of Cardiometabolic Disease in Overweight Adults. *The Journal of Nutrition.* 2018;148(4):518-525.
17. Giangrande PL. The history of blood transfusion. *Br J Haematol.* 2000;110(4):758-767.
18. Roussel J. *Transfusion of human blood by the method of J. Roussel.* Digital copy: Ahttps://wellcomecollection.org/works/nmp52xs9 ed. London: J.&A. Churchill; 1877.
19. Lewisohn R. Blood transfusion by the citrate method. *Am J Med.* 1952;13(5):550-555.
20. Hogman CF, Meryman HT. Red blood cells intended for transfusion: quality criteria revisited. *Transfusion.* 2006;46(1):137-142.
21. Hess JR. An update on solutions for red cell storage. *Vox Sang.* 2006;91(1):13-19.
22. Massive Transfusion Scoring Systems: Appendices to Clinical Practice Guideline. *J Emerg Nurs.* 2019;45(6):e30-e42.
23. Barton CA, Oetken HJ, Hall NL, et al. Whole blood versus balanced resuscitation in massive hemorrhage: six of one or half dozen of the other? *J Trauma Acute Care Surg.* 2024.
24. Hooper N, Armstrong TJ. Hemorrhagic Shock. In: *StatPearls.* Treasure Island (FL) ineligible companies. Disclosure: Tyler Armstrong declares no relevant financial relationships with ineligible companies.2023.
25. Burman S, Cotton BA. Trauma patients at risk for massive transfusion: the role of scoring systems and the impact of early identification on patient outcomes. *Expert Rev Hematol.* 2012;5(2):211-218.
26. Kalkwarf KJ, Cotton BA. Resuscitation for Hypovolemic Shock. *Surg Clin North Am.* 2017;97(6):1307-1321.
27. Baird EW, Lammers DT, Abraham P, et al. Diagnostic performance of the ABC score in the PROPPR trial. *Injury.* 2024;55(8):111656.
28. Mitterecker A, Hofmann A, Trentino KM, et al. Machine learning-based prediction of transfusion. *Transfusion.* 2020;60(9):1977-1986.
29. Russell Stuart NP. *Artificial Intelligence: A Modern approach.* 4th Edition ed. Available from: VitalSource Bookshelf: Pearson Education (US).
30. Lin YK, Chou YY. A hybrid genetic algorithm for operating room scheduling. *Health Care Manag Sci.* 2020;23(2):249-263.
31. Xiang RF, Quinn JG, Watson S, Kumar-Misir A, Cheng C. Application of unsupervised machine learning to identify areas of blood product wastage in transfusion medicine. *Vox Sanguinis.* 2021;116(9):955-964.

32. Schilling M, Rickmann L, Hutschenreuter G, Spreckelsen C. Reduction of Platelet Outdating and Shortage by Forecasting Demand With Statistical Learning and Deep Neural Networks: Modeling Study. *JMIR Med Inform.* 2022;10(2):e29978.
33. Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ Digit Med.* 2020;3:136.
34. Hashmi MF, Katiyar S, Keskar AG, Bokde ND, Geem ZW. Efficient Pneumonia Detection in Chest Xray Images Using Deep Transfer Learning. *Diagnostics (Basel).* 2020;10(6).
35. Schwartz JM, George M, Rossetti SC, et al. Factors Influencing Clinician Trust in Predictive Clinical Decision Support Systems for In-Hospital Deterioration: Qualitative Descriptive Study. *JMIR Hum Factors.* 2022;9(2):e33960.
36. Capizzi A, Woo J, Verduzco-Gutierrez M. Traumatic Brain Injury: An Overview of Epidemiology, Pathophysiology, and Medical Management. *Med Clin North Am.* 2020;104(2):213-238.
37. Refaai MA, Blumberg N. The transfusion dilemma--weighing the known and newly proposed risks of blood transfusions against the uncertain benefits. *Best Pract Res Clin Anaesthesiol.* 2013;27(1):17-35.
38. Jones JM, Sapiano MRP, Savinkina AA, et al. Slowing decline in blood collection and transfusion in the United States - 2017. *Transfusion.* 2020;60 Suppl 2(Suppl 2):S1-S9.
39. Team RS. RStudio: Integrated Development Environment for R. 2024.
40. Team RC. R: A Language and Environment for Statistical Computing. 2024.
41. Rinker TW, Kurkiewicz D. {pacman}: {P}ackage Management for {R}. 2018.
42. McLean MW. RefManageR: Import and Manage BibTeX and BibLaTeX References in R. *The Journal of Open Source Software.* 2017.
43. Boshnakov GN, Putman C. rbibutils: Read ’Bibtex’ Files and Convert Between Bibliography Formats. 2024.
44. Dragulescu A, Arendt C. xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files. 2020.
45. Firke S. janitor: Simple Tools for Examining and Cleaning Dirty Data. 2023.
46. Wickham H, Henry L. purrr: Functional Programming Tools. 2023.
47. Robinson D, Hayes A, Couch S. broom: Convert Statistical Objects into Tidy Tibbles. 2024.
48. Wickham H, Bryan J. readxl: Read Excel Files. 2023.
49. Wickham H, Hester J, Bryan J. readr: Read Rectangular Text Data. 2024.
50. Wickham H. Reshaping Data with the {reshape} Package. *Journal of Statistical Software.* 2007;21(12):1-20.
51. William R. psych: Procedures for Psychological, Psychometric, and Personality Research. 2024.
52. Grolemund G, Wickham H. Dates and Times Made Easy with {lubridate}. *Journal of Statistical Software.* 2011;40(3):1-25.
53. Wickham H, Averick M, Bryan J, et al. Welcome to the {tidyverse}. *Journal of Open Source Software.* 2019;4(43):1686.
54. Sjoberg DD, Whiting K, Curry M, Lavery JA, Larmarange J. Reproducible Summary Tables with the gtsummary Package. *{The R Journal}.* 2021;13(1):570-580.
55. Gohel D, Skintzos P. flextable: Functions for Tabular Reporting. 2024.

56. Pedro A, Garnier, Simon, et al. {viridis(Lite)} - Colorblind-Friendly Color Maps for R. 2024.
57. Wickham H. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York; 2016.
58. Sjoberg DD, Fei T. tidycmprsk: Competing Risks Estimation. 2024.
59. Kim AY, Ismay C, Kuhn M. Take a moderndive into introductory linear regression with R. *The Journal of Open Source Education.* 2021;4(41).
60. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using ’ggplot2’. 2024.
61. Lesnoff, M, Lancelot, R. aod: Analysis of Overdispersed Data. 2012.
62. James D, Hornik K. chron: Chronological Objects which Can Handle Dates and Times. 2023.
63. Lüdecke D. sjPlot: Data Visualization for Statistics in Social Science. 2024.
64. Long JA. jtools: Analysis and Presentation of Social Scientific Data. 2022.
65. Auguie B. gridExtra: Miscellaneous Functions for ”Grid” Graphics. 2017.
66. Schauberger P, Walker A. openxlsx: Read, Write and Edit xlsx Files. 2024.
67. Barrett T, Dowle M, Srinivasan A, et al. data.table: Extension of ‘data.frame‘. 2024.
68. Sjoberg D, Yogasekaram A, de la Rua E. cardx: Extra Analysis Results Data Utilities. 2024.
69. Wickham H, François R, Henry L, Müller K, Vaughan D. dplyr: A Grammar of Data Manipulation. 2023.
70. Iannone R, Cheng J, Schloerke B, et al. gt: Easily Create Presentation-Ready Display Tables. 2024.
71. Chang W. webshot: Take Screenshots of Web Pages. 2023.
72. Gohel D, Moog S, Heckmann M. officer: Manipulation of Microsoft Word and PowerPoint Documents. 2024.
73. Taghavi S, Nassar, A.K., Askari, R. Hypovolemic Shock. *StatPearls [Internet].* 2024.
74. Nishtala MV, Reed M, Schumacher J, et al. Catastrophic health expenditure in nonneurological injury due to motor vehicle crash. *J Trauma Acute Care Surg.* 2023;95(2):172-180.
75. Ambady NR, R. Thin Slices of Expressive Behavior as Predictors of Interpersonal Consequences. *Psychological Bulletin.* 1992;111(2):256-274.
76. Kahneman D. *Thinking, Fast and Slow.* 2011.
77. Norman G. Dual processing and diagnostic errors. *Adv Health Sci Educ Theory Pract.* 2009;14 Suppl 1:37-49.
78. Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. *Acad Med.* 2011;86(3):307-313.
79. Akl HE, Mostafa YA. Digitization and validation of the open bite checklist manifesto: a step toward artificial intelligence. *Angle Orthod.* 2024;94(1):51-58.
80. Jama T, Lefering R, Lauronen J, Handolin L. Factors affecting physicians' decision to start prehospital blood product transfusion in blunt trauma patients: A cohort study of Helsinki Trauma Registry. *Transfusion.* 2024;64 Suppl 2:S167-s173.
81. Feng YN, Xu ZH, Liu JT, Sun XL, Wang DQ, Yu Y. Intelligent prediction of RBC demand in trauma patients using decision tree methods. *Mil Med Res.* 2021;8(1):33.

82. Christie SA, Conroy AS, Callcut RA, Hubbard AE, Cohen MJ. Dynamic multi-outcome prediction after injury: Applying adaptive machine learning for precision medicine in trauma. *PLoS One.* 2019;14(4):e0213836.
83. Ngufor C, Murphree D, Upadhyaya S, et al. Predicting Prolonged Stay in the ICU Attributable to Bleeding in Patients Offered Plasma Transfusion. *AMIA Annu Symp Proc.* 2016;2016:954-963.
84. Ngufor C, Caraballo PJ, O'Byrne TJ, et al. Development and Validation of a Risk Stratification Model Using Disease Severity Hierarchy for Mortality or Major Cardiovascular Event. *JAMA Netw Open.* 2020;3(7):e208270.
85. Trentino KM, Schwarzbauer K, Mitterecker A, et al. Machine Learning-Based Mortality Prediction of Patients at Risk During Hospital Admission. *J Patient Saf.* 2022;18(5):494-498.
86. El Moheb M, Gebran A, Maurer LR, et al. Artificial intelligence versus surgeon gestalt in predicting risk of emergency general surgery. *J Trauma Acute Care Surg.* 2023;95(4):565-572.
87. Park HJ, Jung DY, Ji W, Choi CM. Detection of Bacteremia in Surgical In-Patients Using Recurrent Neural Network Based on Time Series Records: Development and Validation Study. *J Med Internet Res.* 2020;22(8):e19512.
88. Murphree D, Ngufor C, Upadhyaya S, et al. Ensemble learning approaches to predicting complications of blood transfusion. *Conf Proc IEEE Eng Med Biol Soc.* 2015;2015:7222-7225.
89. Breiman L. Random forests. *Machine learning.* 2001;45:5-32.
90. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. Paper presented at: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining2016.
91. Cortes C, Vapnik V. Support-Vector Networks. *Mach Learn.* 1995;20(3):273–297.
92. Freund Y, Schapire RE. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences.* 1997;55(1):119-139.
93. Nederpelt CJ, Mokhtari AK, Alser O, et al. Development of a field artificial intelligence triage tool: Confidence in the prediction of shock, transfusion, and definitive surgical therapy in patients with truncal gunshot wounds. *J Trauma Acute Care Surg.* 2021;90(6):1054-1060.
94. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE.* 1998;86(11):2278-2324.
95. Huang G, Liu Z, Maaten LVD, Weinberger KQ. Densely Connected Convolutional Networks. Paper presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 21-26 July 2017, 2017.
96. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. Paper presented at: 2017 IEEE International Conference on Computer Vision (ICCV); 22-29 Oct. 2017, 2017.
97. Ahamed MKU, Islam MM, Uddin MA, et al. DTLCx: An Improved ResNet Architecture to Classify Normal and Conventional Pneumonia Cases from COVID-19 Instances with Grad-CAM-Based Superimposed Visualization Utilizing Chest X-ray Images. *Diagnostics (Basel).* 2023;13(3).

98. Hamza A, Attique Khan M, Wang SH, et al. COVID-19 classification using chest X-ray images based on fusion-assisted deep Bayesian optimization and Grad-CAM visualization. *Front Public Health.* 2022;10:1046296.
99. Sobahi N, Atila O, Deniz E, Sengur A, Acharya UR. Explainable COVID-19 detection using fractal dimension and vision transformer with Grad-CAM on cough sounds. *Biocybern Biomed Eng.* 2022;42(3):1066-1080.
100. Jahmunah V, Ng EYK, Tan RS, Oh SL, Acharya UR. Explainable detection of myocardial infarction using deep learning models with Grad-CAM technique on ECG signals. *Comput Biol Med.* 2022;146:105550.
101. Kolekar S, Gite S, Pradhan B, Alamri A. Explainable AI in Scene Understanding for Autonomous Vehicles in Unstructured Traffic Environments on Indian Roads Using the Inception U-Net Model with Grad-CAM Visualization. *Sensors (Basel).* 2022;22(24).
102. Kumaran SY, Jeya JJ, R MT, Khan SB, Alzahrani S, Alojail M. Explainable lung cancer classification with ensemble transfer learning of VGG16, Resnet50 and InceptionV3 using grad-cam. *BMC Med Imaging.* 2024;24(1):176.
103. Livieris IE, Pintelas E, Kiriakidou N, Pintelas P. Explainable Image Similarity: Integrating Siamese Networks and Grad-CAM. *J Imaging.* 2023;9(10).
104. Hammad A, Khalil S, Moretti S. Search for mono-Higgs signals in $b\overline{b}$ final states using deep neural networks. *Physical Review D.* 2023;107(7):075027.
105. Luo J, Xiong C. Youden index and Associated Cut-points for Three Ordinal Diagnostic Groups. *Commun Stat Simul Comput.* 2013;42(6):1213-1234.
106. Yin J, Tian L. Joint confidence region estimation for area under ROC curve and Youden index. *Stat Med.* 2014;33(6):985-1000.
107. Yin J, Samawi H, Tian L. Joint inference about the AUC and Youden index for paired biomarkers. *Stat Med.* 2022;41(1):37-64.
108. Yin J, Mutiso F, Tian L. Joint hypothesis testing of the area under the receiver operating characteristic curve and the Youden index. *Pharm Stat.* 2021;20(3):657-674.
109. Wen J, Tang X, Lu J. An imbalanced learning method based on graph tran-smote for fraud detection. *Sci Rep.* 2024;14(1):16560.
110. Swana EF, Doorsamy W, Bokoro P. Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset. *Sensors (Basel).* 2022;22(9).
111. Kim SY, Jin JJ, Ha A, et al. SMOTE-Enhanced Explainable AI Model for Predicting Visual Field Progression in Myopic Normal-Tension Glaucoma. *J Glaucoma.* 2025.
112. Li Y, Du C, Ge S, et al. Hematoma expansion prediction based on SMOTE and XGBoost algorithm. *BMC Med Inform Decis Mak.* 2024;24(1):172.
113. Schaudt D, Spate C, von Schwerin R, et al. A Critical Assessment of Generative Models for Synthetic Data Augmentation on Limited Pneumonia X-ray Data. *Bioengineering (Basel).* 2023;10(12).
114. Nishio M, Noguchi S, Matsuo H, Murakami T. Automatic classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy on chest X-ray image: combination of data augmentation methods. *Sci Rep.* 2020;10(1):17532.

115. Muhammad U, Hoque MZ, Oussalah M, Keskinarkaus A, Seppanen T, Sarder P. SAM: Self-augmentation mechanism for COVID-19 detection using chest X-ray images. *Knowl Based Syst.* 2022;241:108207.
116. Dong H, Zhu B, Zhang X, Kong X. Use data augmentation for a deep learning classification model with chest X-ray clinical imaging featuring coal workers' pneumoconiosis. *BMC Pulm Med.* 2022;22(1):271.
117. Ellingson KD, Sapiano MRP, Haass KA, et al. Continued decline in blood collection and transfusion in the United States-2015. *Transfusion.* 2017;57 Suppl 2(Suppl 2):1588-1598.
118. Oliveira EM, Reis IA. What are the perspectives for blood donations and blood component transfusion worldwide? A systematic review of time series studies. *Sao Paulo Med J.* 2020;138(1):54-59.
119. Stonko DP, Guillamondegui OD, Fischer PE, Dennis BM. Artificial intelligence in trauma systems. *Surgery.* 2021;169(6):1295-1299.
120. Dennis BM, Stonko DP, Callcut RA, et al. Artificial neural networks can predict trauma volume and acuity regardless of center size and geography: A multicenter study. *J Trauma Acute Care Surg.* 2019;87(1):181-187.
121. Stonko DP, Dennis BM, Betzold RD, Peetz AB, Gunter OL, Guillamondegui OD. Artificial intelligence can predict daily trauma volume and average acuity. *J Trauma Acute Care Surg.* 2018;85(2):393-397.
122. Roislien J, Sovik S, Eken T. Seasonality in trauma admissions - Are daylight and weather variables better predictors than general cyclic effects? *PLoS One.* 2018;13(2):e0192568.
123. Wu S, Liu S, Sohn S, et al. Modeling asynchronous event sequences with RNNs. *J Biomed Inform.* 2018;83:167-177.
124. Hallman J. tis: Time Indexes and Time Indexed Series. 2021.
125. Lazaridis E. lunar: Lunar Phase & Distance, Seasons and Other Environmental Factors. 2022.
126. Sefick S. Stream Metabolism-A package for calculating single station metabolism from diurnal Oxygen curves. 2023.
127. Box GEP, Cox DR. An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological).* 2018;26(2):211-243.
128. Sibono L, Grosso M, Tejedor-Calvo E, et al. A critical analysis of Adaptive Box-Cox transformation for skewed distributed data management: Metabolomics of Spanish and Argentinian truffles as a case study. *Anal Chim Acta.* 2025;1345:343704.
129. Kim S, Chen MH, Ibrahim JG, Shah AK, Lin J. Bayesian inference for multivariate meta-analysis Box-Cox transformation models for individual patient data with applications to evaluation of cholesterol-lowering drugs. *Stat Med.* 2013;32(23):3972-3990.
130. Blum L, Elgendi M, Menon C. Impact of Box-Cox Transformation on Machine-Learning Algorithms. *Front Artif Intell.* 2022;5:877569.