Decision-making in pediatric blunt solid organ injury: a deep learning approach to predict massive transfusion, need for operative management, and mortality risk

Niti Shahi, Ashwani K. Shahi, Ryan Phillips, Gabrielle Shirek, Denis Bensard, Steven L. Moulton

PII: S0022-3468(20)30770-3

DOI: https://doi.org/10.1016/j.jpedsurg.2020.10.021

Reference: YJPSU 59982

To appear in: *Journal of Pediatric Surgery*

Received date: 12 October 2020

Accepted date: 20 October 2020

Please cite this article as: N. Shahi, A.K. Shahi, R. Phillips, et al., Decision-making in pediatric blunt solid organ injury: a deep learning approach to predict massive transfusion, need for operative management, and mortality risk, *Journal of Pediatric Surgery* (2020), https://doi.org/10.1016/j.jpedsurg.2020.10.021

October 10th, 2020

Dear Dr. Holcomb,

Thank you for your consideration of our manuscript entitled Decision-Making in Pediatric Blunt Solid Organ Injury: A Deep Learning Approach to Predict Massive Transfusion, Need for Operative Management, and Mortality Risk for publication in *The Journal of Pediatric Surgery* for the WPTC meeting special edition. This abstract was accepted for oral presentation at WPTC. We have no affiliations or financial conflicts of interest to disclose. This manuscript has not been published previously, nor is it being submitted elsewhere.

Thank you again for your consideration of our manuscript.

Sincerely,

Niti Shahi, MD

Decision-Making in Pediatric Blunt Solid Organ Injury: A Deep Learning Approach to Predict Massive Transfusion, Need for Operative Management, and Mortality Risk

Niti Shahi*[1,2,3], Ashwani K. Shahi*, Ryan Phillips[1,2], Gabrielle Shirek[1], Denis Bensard[1,2,4], Steven L. Moulton[1,2]

[1]Division of Pediatric Surgery, Children's Hospital Colorado, Aurora, CO, USA; [2]Department of Surgery, University of Colorado School of Medicine, Aurora, CO, USA; [3]Department of Surgery, University of Massachusetts, Worcester, MA, USA; [4]Department of Surgery, Denver Health Medical Center, Denver, CO, USA

*Co-first authors

Corresponding Author:

Niti Shahi, MD
Department of Surgery
University of Massachusetts School of Medicine
55 Lake Avenue North,
Worcester, MA 01655
niti.shahi@umassmemorial.org
Fax: (844)-804-3061

Presentations: Accepted for oral presentation at Western Pediatric Trauma Conference (cancelled due to COVID19)

Funding Sources/Disclosures: The authors have no conflicts of interest, financial or otherwise, to disclose. No grant funding was used to support this research.

IRB Approval: This study was conducted under a protocol reviewed and approved by The Colorado Multiple Institutional Review Board (COMIRB) at the University of Colorado. A waiver of consent was reviewed and approved by the IRB.

Abstract

Background:

The principal triggers for intervention in the setting of pediatric blunt solid organ injury (BSOI) are declining hemoglobin values and hemodynamic instability. The clinical management of BSOI is, however, complex. We therefore hypothesized that state-of-art machine learning (computer-based) algorithms could be leveraged to discover new combinations of clinical variables that might herald the need for an escalation in care. We developed algorithms to predict the need for massive transfusion (MT), failure of non-operative management (NOM), mortality, and successful non-operative management without intervention, all within four hours of emergency department (ED) presentation.

Methods:

Children (≤18 years) who sustained a BSOI (liver, spleen, and/or kidney) between 2009-2018 were identified in the trauma registry at a pediatric level 1 trauma center. Deep learning models were developed using clinical values [vital signs, shock index-pediatric adjusted (SIPA), organ injured, and blood products received], laboratory results [hemoglobin, base deficit, INR, lactate, thromboelastography (TEG)], and imaging findings [focused assessment with sonography in trauma (FAST) and grade of injury on computed tomography scan] from pre-hospital to ED settings for prediction of MT, failure of NOM, mortality, and successful NOM without intervention. Sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve (AUC) were used to evaluate each model's performance.

Results:

A total of 477 patients were included, of which 5.7% required MT (27/477), 7.2% failed NOM (34/477), 4.4% died (21/477), and 89.1% had successful NOM (425/477). The accuracy of the models in the validation set were as follows: MT (90.5%), failure of NOM (83.8%), mortality (91.9%), and successful NOM without intervention (90.3%). Serial vital signs, the grade of organ injury, hemoglobin, and positive FAST had low correlations with outcomes.

Conclusion:

Deep learning-based models using a combination of clinical, laboratory and radiographic features can predict the need for emergent intervention (MT, angioembolization, or operative management) and mortality with high accuracy and sensitivity using data available in the first four hours of admission. Further research is needed to externally validate and determine the feasibility of prospectively applying this framework to improve care and outcomes.

Key Words: machine learning, big data, deep learning, trauma, massive transfusion, pediatric, artificial intelligence

Level of Evidence: III

Study Type: Retrospective comparative study (Prognosis/Care Management)

Introduction

Most children who sustain blunt solid organ injuries (BSOI) do not require significant intervention (1). Some, however, will require aggressive interventions, such as massive transfusion (MT) and/or surgery as life-saving measures. Early BSOI clinical practice guidelines used the grade of solid organ injury to guide management (2). The management of BSOI has evolved, however, such that most interventions are now guided by hemodynamic status (2). Yet, even with close monitoring of hemodynamic status, it can be challenging to identify which patients are more likely to require MT or fail non-operative management (NOM). There are limitations to using vital signs to evaluate hemodynamic status and predict impending shock in children. First, hypotension in children with traumatic injuries and BSOI may be due to severe traumatic brain injuries as opposed to hemorrhagic shock (3). Second, as Partrick et al. highlighted, "children are recognized as having an increased physiologic reserve and therefore may have nearly normal vital signs even in the presence of severe shock (4)."

Several recent studies have demonstrated promise using the shock index- pediatric adjusted (SIPA) score, grading systems, and thromboelastography (TEG) to predict MT and/or failure of NOM in pediatric trauma patients (2, 5-9). For MT, several different grading systems have been developed with variable performance. The most widely recognized is the ABC score, which has been validated in adult trauma patients. It is comprised of four components with one point for each of the following: penetrating mechanism, positive focused abdominal sonography for trauma (FAST), systolic blood pressure (SBP) <90, and heart rate (HR) $\geq$120. A score higher than two supports the decision for triggering MT (5). While it was initially found to be 75% sensitive and 86% specific, later studies have shown that "the ABC score overestimates the need

for transfusion, with a positive predictive value of 50 percent to 55 percent (6)." Additionally, the ABC score relies on the FAST exam, which has been found to have poor sensitivity in children, and it utilizes vital sign values based on abnormal adult ranges (2). Phillips et al. therefore developed the ABCD score to more accurately assess children who sustain blunt or penetrating injuries. It is comprised of the ABC score with age-specific SIPA values (abnormal versus normal) replacing heart rate and systolic blood pressure, together with lactate and base deficit. An ABCD score >3 had a sensitivity of 77.4%, specificity of 78.8%, and a 77.6% accuracy in identifying the need for massive transfusion (7). More recent work has shown that specific rapid TEG findings are also associated with the need for massive transfusion in blunt and penetrating trauma, including: ACT ≥ 128 sec, angle (α) ≤ 65, maximum amplitude (MA) ≤ 55 mm, and LY30 ≥ 5% %(8). Linnaus et al. similarly found that a high percentage of children who sustained BSOI injuries and required transfusion or failed NOM had elevated SIPA values in the trauma bay (9).

Machine learning (ML) has the potential to build upon the above findings by identifying individual and combinations of features associated with outcomes. Deep learning (DL) is a subset of machine learning, which does not require extensive feature engineering based on domain knowledge to extract features from raw data (10). Instead, DL has the potential to automatically determine features and combinations of features from raw data through linear and non-linear models (10). There has been limited application of deep learning thus far in pediatric trauma outcomes research. This study aimed to develop DL models to help in decision making for pediatric BSOI by predicting which patients: 1) may need massive transfusion; 2) may fail NOM; 3) are at risk for mortality; or 4) can be successfully managed with NOM without intervention.

Methods

Setting:

Children's Hospital Colorado (CHCO) is a 444-bed, free-standing, regional referral pediatric hospital. It is the only American College of Surgeons (ACS) verified Level 1 Pediatric Trauma Center in Colorado and the adjacent seven states of North Dakota, South Dakota, Nebraska, Kansas, New Mexico, Wyoming, and Montana.

Data Collection and Inclusion Criteria:

This study was approved by the Colorado Multi-Institutional Review Board (COMIRB) with a waiver of informed consent. The institution's trauma registry was queried for all patients < 18 years old with a BSOI (liver, spleen, or kidney) from 2009-2018. Data collection included demographics (age, gender, race, ethnicity, and insurance type), emergency department (ED) vital signs [heart rate (HR) and blood pressure (BP)], ED SIPA, clinical characteristics [Glasgow Coma Scale (GCS), intubation status, weight, blood products received, and injury severity score (ISS)], imaging findings [Focused Assessment with Sonography in Trauma (FAST) findings, as well as organ(s) injured and grade of injury on computed tomography (CT)], and laboratory findings [serial hemoglobin values, base deficit, INR, lactate, and TEG] (11). MT was defined as receiving >40 cc/kg within 6 hours of presentation (7). All data was de-identified before the development of the four models.

Development of the Models:

Many researchers are applying Deep Neural Networks (DNNs) with small datasets across various domains. Regression and classification problems formerly treated by traditional machine learning methods (like Support Vector Machines, Random Forest, etc.) with a small dataset are being solved by DNNs with higher accuracy and better generalization performance. For example, in domains like materials science, DNNs with small datasets are being used to predict material defects (12). Though DNNs with big datasets is the optimal solution, DNN with small datasets can be a reasonable choice when big datasets are unavailable. Various strategies for applying deep learning tools to small datasets include carefully selecting loss function (i.e. hinge or cosine loss for optimization), transfer learning, regularization techniques like stochastic drop-out training to reduce overfitting, and better optimization tools (i.e. batch normalization and learning rate) for preventing underfitting (13).

A key question is how to best fit machine learning models to relatively small "training" data sets, so that accurate predictions can be made on new data. In machine learning jargon, this is the question of *generalization*. Per conventional wisdom in machine learning, a model that is too simple will *underfit* the true patterns in the training data, and thus, it will poorly predict on new data. A model that is too complicated will *overfit* spurious patterns in the training data; such a model will also poorly predict on new data.

Recent deep learning practice appears to eschew this conventional wisdom that was applicable to traditional statistical machine learning models. Bornschein et al, in a recent paper from the International Conference on Machine Learning (ICML) showed that one can train on a smaller subset of the training data while maintaining generalizable results, even for large overparameterized models (14). Highly overparameterized neural networks (where the number of model parameters exceeds the number of training data) can display strong generalization

performance, even on small datasets. In our study, we observed the same generalization behavior. In each of our models, the number of model parameters exceeded training data size. Due to an imbalanced data set, models were built by under-sampling the majority class. But the models performed well on the validation data from the majority class and did not suffer from overfitting.

Four models were developed: MT, failure of NOM, mortality, and successful NOM without intervention. Deep Learning models were developed on Google Cloud Platform using Google Colaboratory (Colab) using TensorFlow/Keras 2.0. For each model, two experiments were conducted – a model with a 4-hour data set and a model with 24-hour data set. Due to the small sample size and unbalanced dataset (i.e. 21 deaths vs. 456 survivors), the majority class was under-sampled to create a balanced set for model training. For the MT model, a training set of 37 was used and a validation set of 440 was used. For the failure of NOM model, a training set of 47 and a validation set of 430 was used. For the mortality model, a training set of 30 and a validation set of 447 was used. Lastly, for the successful NOM without intervention model, a training set of 66 and validation set of 411 was used. Each deep learning model consisted of input layers, hidden layers, and output layers. Leaky rectified linear unit (Leaky ReLU) activation functions were used with hinge loss functions for training and optimization of the classifiers. Dropout layers were used to regularize and reduce overfitting.

Features:

The following features were used in the development of the three models: demographics (gender, age, weight), GCS scores, vital signs (HR and BP; for pre-hospital, ED arrival, as well as 2 hours and 4 hours after ED arrival), SIPA scores (calculated as heart rate divided by blood

pressure; pre-hospital, ED arrival, as well as 2 hours and 4 hours after ED arrival), ED TEG values [R-time (R), alpha angle, maximum amplitude (MA), and lysis at 30 minutes (LY30)], lab values [hemoglobin (ED arrival, as well as 2 hours and 4 hours after ED arrival), INR, base deficit, and lactate], resuscitation metrics [fluid administered in pre-hospital and hospital settings (cc/kg) and blood transfusion in pre-hospital and hospital settings (up to 4 hours after presentation)], clinical events (intubation in pre-hospital setting or ED, in addition to cardiopulmonary resuscitation (CPR) in the pre-hospital setting or ED), presence of a head injury, multiple solid organ injuries, and imaging findings (FAST and CT grade of injury). Vital signs and laboratory values were used as both continuous variables and categorical variables. Categorical inputs were further converted into numerical variables (as required by deep learning models) as flags with one or zero value for the following: abnormal, normal, or unknown. Normality of lab values was determined based on institutional ranges. An additional set of models were run with the same clinical information available at 24 hours after presentation.

Statistical Analysis:

Demographic and outcomes data are presented as medians with interquartile ranges for continuous variables and as frequencies with percentages for categorical variables. Accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC) were used to assess performance. Extensive exploratory data analysis including calculation of descriptive statistics (mean, median, IQR, minimum, maximum, and missing data counts) was conducted to study and compare statistics between populations that received an intervention and the population that was successfully managed non- operatively. Statistical techniques such as t-tests were used to calculate p-values using scikit-learn python libraries like scipy.stats and

stats.ttest_ind. For each input feature, counts of missing data by outcome for each population group were calculated.

Various approaches like Pearson correlations, chi-square, and recursive feature elimination with cross-validation (RFECV) were used for determining feature importance for each model. This helped to inform which features could be excluded from the model when a large number of observations were missing a data input. Deleting data can result in reduced statistical power, biased estimators, reduced representativeness of the sample, or incorrect inferences and conclusions. For handling missing data, we imputed missing values.

Results

Demographics and Clinical Characteristics

A total of 477 pediatric trauma patients sustained BSOI during the study period. The median age at the time of injury was 10.0 (IQR 6.0, 14.0) years old. Sixty-five percent of injured children were males (311/477). Two-hundred sixty-one patients (54.7%) had liver injuries, 250 (52.4%) had spleen injuries, and 35 (7.4%) had kidney injuries; a total of 65 patients had multiple BSOI injuries (13.6%). Twenty-seven patients (5.7%) required MT. Four patients (0.8%) underwent angioembolization, and 34 patients failed non-operative management. Overall, 21 (4.4%) of the patients died. There were 425 (89.1%) patients who were successively managed nonoperatively and survived. The remainder of the demographic and clinical characteristics are summarized in Table 1, divided into cohorts by patients who underwent successful NOM without intervention and survived versus those who did not. Correlations between clinical characteristics and outcomes are demonstrated in Table 2.

Performance of the 4-hour Models:

For MT, the model achieved 90.5% accuracy, 88.9% sensitivity, and 90.5% specificity with an AUC of 0.90 for the validation set. For failure of NOM, the model had 83.8% accuracy, 91.7% sensitivity, and 83.5% specificity with an AUC of 0.88 for the validation set. For the outcome of mortality, the model achieved 91.9% accuracy, 100.0% sensitivity, and 91.8% specificity with an AUC of 0.96 for the validation set. Lastly, for successful NOM without intervention, the model had a 90.3% accuracy, 90.4% sensitivity, and 88.2% specificity with AUC of 0.89.

Massive Transfusion

The clinical characteristics with the highest absolute correlation with MT was if the patient received any blood products within four hours (r= 0.68), intubation status (r=0.48), abnormal LY30 (r=0.53), and GCS (r= 0.47). We identified 17 patients (63.0%; 17/27) who met ABCD criteria who received MT (7). Another 17 patients (3.8%; 17/450) who met ABCD criteria did not receive MT (7).

Failure of NOM

A majority of the clinical characteristics had a low correlation with failure of NOM. Factors with the highest absolute correlation with failure of NOM were LY30 (r=0.43), R (r=0.40), and MA (r=0.38). FAST had a weak correlation with failure of NOM (r= 0.15). Grade of organ injury (liver, spleen, and/or kidney) had weak correlation with failure of NOM (all r's <0.2).

Mortality

The clinical factors that had the highest absolute correlations with mortality were history of CPR in the ED (r= 0.68), history of CPR in the field (r= 0.66), ED base deficit values (r=0.65), and ED INR value (r= 0.61).

Successful NOM with No Intervention

The demographics and outcomes of the patients who were successfully treated with NOM with no intervention and survived versus those who underwent an intervention (MT, angioembolization, and/or surgical management) are demonstrated in Table 1. The clinical factors that had the highest absolute correlation in this model were GCS (r= 0.53), presence/absence of CPR in the ED (r= 0.52), and intubation in the ED (r= 0.50).

Review of False Positives

In a review of the false positives for all four models, common themes were identified where model prediction incorrectly identified an outcome or condition. For the MT and failure of NOM models, there was a cohort of patients with severe traumatic brain injuries that affected their hemodynamic status. Additionally, many patients had concomitant orthopedic injuries such as pelvic fractures that contributed to pre-hospital hemodynamic instability, but did not require MT. For the mortality model, traumatic brain injury, orthopedic polytrauma, or significant cardiac and aortic injuries were common in the false positive patients and likely contributed to their initial hemodynamic instability. Lastly, for the successful NOM without intervention model, patients classified in the false positive category were incorrectly classified as successful NOM. These patients were hemodynamically stable on presentation, and they had worsening

physical exam findings over time or signs of bowel or retroperitoneal organ injury requiring operative management.

Review of False Negatives

False negatives were reviewed in all four models, where the model failed to predict an outcome or condition. The MT model had one patient who was incorrectly classified as not needing MT, who actually required MT. This patient presented in hemorrhagic shock and required MT for stabilization. This patient was missing base deficit, lactate, and INR values. He responded to early implementation of MT, and his ED vital signs and SIPA were within normal limits for his age. The failure of NOM model had several false negatives, and common themes in these patients included initial hemodynamic stability, followed by worsening physical exam findings or CT findings warranting operative management. The mortality model had no false negatives. In the successful NOM model, most false negatives were patients who were initially hemodynamically unstable. Several of these patients also had traumatic brain injuries and orthopedic injuries, which affected their initial hemodynamic status. Over time, they were managed with blood transfusion or intravenous fluid administration and ultimately stabilized.

Comparison of 4 hour and 24-hour models

The DL models were run with data available at 24 hours after ED presentation for comparison with the models described above (4-hour models). The sensitivity of the MT, failure of NOM, mortality models in addition to the successful NOM without intervention model are presented in Table 3. The four-hour models outperformed the 24 models for all outcomes.

Discussion

The present study demonstrates the potential utility of using deep learning to identify children with BSOI at risk for poor outcomes, within four hours of presentation. To date, there are limited studies using machine and deep learning techniques in the pediatric trauma literature. The present study demonstrates the feasibility and efficacy of its use with high accuracy, sensitivity, and specificity for all four outcomes in a small dataset. Future work to build upon this model using a larger data set could lay the foundation for prospective validation of the deep learning-based approach.

Over the past decade there has been an evolution in the management children with BSOIs. In the past, grade of injury based on CT findings guided management. Nowadays, hemodynamic status is recognized as a more individual-specific means to tailor therapy to the degree of solid organ injury and the needs of the child. For example, many institutions previously performed serial hemoglobin/hematocrit studies to assess for ongoing hemorrhage, in addition to vital signs monitoring. Recent studies have shown, however, that repeat or serial hemoglobin levels following BSOI are of limited utility (15, 16). Additionally, a prior study by Acker et al demonstrated that pediatric BSOI patients who failed non-operative management did so at a median of 4 hours from the time of injury (1). Our findings corroborate the lack of utility of trending hemoglobin lab values beyond four hours, as the serial hemoglobin values poorly correlated with failure of NOM, mortality, and successful NOM without intervention. Moreover, our models demonstrate that the clinical history and laboratory values available at four hours outperformed the models that utilized clinical history and laboratory values available at 24 hours.

Thus, the information available within four hours of presentation is often adequate for decision-making in this critically ill patient population.

Our study sheds additional light on the benefits of using TEG to identify severely injured pediatric trauma patients. The use of TEG and its association with MT and mortality has primarily been explored in the adult literature. Specifically, Coleman et al. found that adults with blunt solid organ injuries were hypercoagulable upon admission, as demonstrated in their ED TEGs (17). Over half of the study patients had evidence of fibrinolysis shutdown on admission. TEG may help with the early identification of patients with severe blunt solid organ injuries, who will require an intervention and are at risk for poor outcomes. Specifically, in this study, we found that LY30 had a positive correlation with the need for MT and failure of NOM. Future work with an increased number of patients and universal TEG measurement would be logical next steps to building upon these initial findings.

As a result of this study, we recommend routine laboratory evaluation of BSOI patients with ED hemoglobin, base deficit, lactate, INR, and TEG. While many of these features (i.e. laboratory studies) showed moderate correlation with outcomes, only a few had strong linear correlations. The deep learning models further demonstrated that the best steps in the clinical management of select patients who need MT, angioembolization, and/or surgical management is complex. The Pearson correlations provided some insight into the linear correlations between clinical characteristics and outcomes; however, in real life, all outcomes are not linearly correlated with inputs. The advantage of deep learning is that it transforms raw inputs into meaningful outcomes by learning the complex relationships between combinations of inputs and outcomes. Thus, improvements in the accuracy of the models will arise from larger volumes of empirically validated data across a variety of clinical domains (18).

There are various options to deploy machine learning models into production at scale, to use them in real-world clinical applications. One common approach is to save the models and build an application with Representational State Transfer (REST) endpoints to deploy the models using a cloud provider. TensorFlow Serving is another option. It is an efficient model server that can sustain a high load and has a model repository to automatically deploy the latest versions. Integration of the models in either the electronic medical record or a cloud-based platform could allow for easy access and rapid application in the pre-hospital or trauma bay setting.

There are multiple limitations to our study. First, this was a retrospective single center study. Second, deep learning models typically require large datasets, and our dataset only had 477 patients. Third, there was missing data in our cohort. For example, several patients did not have available pre-hospital vital sign data and only 23 patients had available TEG data.

.

Conclusion

Deep learning models show promise in the early identification of pediatric blunt trauma patients at risk for adverse outcomes. One advantage of deep learning models is that they do not require specific components used by traditional scoring systems to predict need for MT or mortality. In this preliminary, single-center study of children with BSOI, applying a DL based algorithm helped correctly identify patients who were successfully treated without intervention. The MT model identified patients needing emergent intervention with higher sensitivity and specificity compared to existing approaches like ABCD (7). There is no widely used predictive model for the failure of NOM. The failure of NOM model had high a sensitivity of 91.7% and specificity of 83.5%. The mortality model provided high sensitivity, specificity, and accuracy. As such, it could be envisioned as an early warning system to alert clinicians of impending

deterioration. Considering nearly 90% of patients were successfully managed non-operatively, the successful NOM without intervention DL model might be a useful tool for identifying the majority of patients who can be successfully managed with fewer resources in less intensive clinical settings, thus lowering the cost of care.

Our models demonstrated that clinical findings within four hours of presentation could be used for critical clinical decision making for pediatric BSOI patients as model performance did not improve with trended data from the first 24 hours following admission. This suggests that serial blood draws beyond four hours may not be needed. Similarly, a positive FAST exam did not make a significant difference in prediction performance. Conversely, considering the relatively high correlation between TEG and outcomes, we hypothesize that inclusion of TEG may help to identify those patients at greatest need for an emergent intervention.

Further research with a larger population, with less emphasis on the FAST exam and universal application of TEG, is needed to further validate the feasibility of applying a DL framework to the management of pediatric trauma patients with BSOIs. Denser patient data, including continuous physiological data and natural language processing of semantic data, are next steps to improving the models. In our false positive and false negative analysis, the presence of a severe traumatic brain injury or pelvic/femur fracture were important factors that affected initial hemodynamic status. Future models with input signals for comorbid injuries may further improve the performance and utility of DL models.

Author Contribution Form:

Study conception and design: NS, AKS

Data acquisition: NS, RP, GS

Analysis and data interpretation: NS, AKS

Drafting of the manuscript: NS, RP, GS, DB, SLM, AKS

Critical revision: NS, RP, GS, DB, SLM, AKS

Tables Legend:

Table 1: Characteristics of Children with Successful Non-operative Management (with no MT) vs Other Outcomes [Required intervention (MT or surgery) or Mortality]

Table 2: Top 25 Features associated with Outcomes Based on Absolute Pearson's Correlation Coefficient Values

Table 3: Demonstration of the 4 and 24-hour Deep Learning Models (Validation set)

References

1. Acker SN, Petrun B, Partrick DA, Roosevelt GE, Bensard DD. Lack of utility of repeat monitoring of hemoglobin and hematocrit following blunt solid organ injury in children. Journal of Trauma and Acute Care Surgery. 2015;79(6):991-4.
2. Notrica DM. Pediatric blunt abdominal trauma: current management. Curr Opin Crit Care. 2015;21(6):531-7.
3. Gardner AR, Diz DI, Teeze JA, Miller CD, Petty J. Injury patterns associated with hypotension in pediatric trauma patients: A national trauma database review. Journal of trauma and acute care surgery. 2015;78(6):1143-8.
4. Partrick DA, Bensard DD, Janik JS, Karrer FM. Is hypotension a reliable indicator of blood loss from traumatic injury in children? The American journal of surgery. 2002;184(6):555-9.
5. Nunez TC, Voskresensky IV, Dossett LA, Shinall R, Dutton WD, Cotton BA. Early prediction of massive transfusion in trauma: simple as ABC (assessment of blood consumption)? Journal of Trauma and Acute Care Surgery. 2009;66(2):346-52.
6. ACS Committee on Trauma. ACS TQIP Massive Transfusion in Trauma Guidelines. 2014.
7. Phillips R, Acker SN, Shahi N, Meier M, Leopold D, Recicar J, et al. The ABC-D score improves the sensitivity in predicting need for massive transfusion in pediatric trauma patients. J Pediatr Surg. 2020;55(2):331-4.
8. Einersen PM, Moore EE, Chapman MP, Moore HB, Gonzalez E, Silliman CC, et al. Rapid thrombelastography thresholds for goal-directed resuscitation of patients at risk for massive transfusion. J Trauma Acute Care Surg. 2017 Jan;82(1):114-9.

9. Linnaus ME, Notrica DM, Langlais CS, Peter SDS, Leys CM, Ostlie DJ, et al. Prospective validation of the shock index pediatric-adjusted (SIPA) in blunt liver and spleen trauma: An ATOMAC study. J Pediatr Surg. 2017;52(2):340-4.
10. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-44.
11. Acker SN, Ross JT, Partrick DA, Tong S, Bensard DD. Pediatric specific shock index accurately identifies severely injured children. J Pediatr Surg. 2015;50(2):331-4.
12. Feng S, Zhou H, Dong H. Using deep neural network with small dataset to predict material defects. Mater Des. 2019;162:300-10.
13. Barz B, Denzler J. Deep learning on small datasets without pre-training using cosine loss. The IEEE Winter Conference on Applications of Computer Vision; 2020.
14. Bornschein J, Visin F, Osindero S. Small Data, Big Decisions: Model Selection in the Small-Data Regime. arXiv preprint arXiv:2009.12583. 2020.
15. Acker SN, Hill LR, Bensard DD, Moulton S, Partrick DA. The benefits of limiting scheduled blood draws in children with a blunt liver or spleen injury. J Pediatr Surg. 2020;55(7):1219-23.
16. Madbak F, Price D, Skarupa D, Yorkgitis B, Ebler D, Hsu A, et al. Serial hemoglobin monitoring in adult patients with blunt solid organ injury: less is more. Trauma Surgery & Acute Care Open. 2020;5(1):e000446.
17. Coleman JR, Kay AB, Moore EE, Moore HB, Gonzalez E, Majercik S, et al. It's sooner than you think: Blunt solid organ injury patients are already hypercoagulable upon hospital admission-Results of a bi-institutional, prospective study. The American Journal of Surgery. 2019;218(6):1065-73.
18. Hestness J, Narang S, Ardalani N, Diamos G, Jun H, Kianinejad H, et al. Deep learning scaling is predictable, empirically. arXiv preprint arXiv:1712.00409. 2017.

Table 1: Characteristics of Children with Successful Non-operative Management (with no MT) vs Other Outcomes [Required intervention (MT or surgery) or Mortality]

	Count	Successful NOM without Intervention	Count	Other Outcomes (Required intervention or Mortality)	P-value
Demographics					
Age (years), median(IQR)	425	11 (6.0,14.0)	52	9.5 (5.5,1.0)	0.2516
Gender, Male/Female	425	273 M/152 F	52	38 M/14 F	0.2072
Clinical Characteristics					
GCS, median (IQR)	425	15.0 (15.0,15.0)	52	3.0 (3.0,15.0)	< 0.0001
Injury severity score, median (IQR)	424	12.0 (9.0,17.0)	52	34.0 (25.0,45.0)	
Liver grade, median (IQR)	229	3.0 (2.0,3.0)	32	3.0 (2.0,5.0)	0.2745
Spleen grade, median (IQR)	219	3.0 (2.0,3.0)	31	3.0 (2.0,4.0)	0.2552
Kidney grade, median (IQR)	31	3.0 (2.0,3.5)	4	3.5 (2.8,4.3)	0.9102
Multiple organ injuries, n(%)	423	49 (11.6%)	52	14 (26.9%)	0.5117
Head injury, n(%)	425	42 (9.9%)	52	34 (65.4%)	< 0.0001
Isolated BSOI, n(%)	422	222 (52.5%)	52	8 (15.4%)	
Pulmonary contusion, n(%)	416	120 (28.8%)	51	30 (58.8%)	
Major orthopedic injury, n(%)	423	140 (33.1%)	52	26 (50.0%)	
Pancreatic injury, n(%)	423	8 (1.9%)	52	7 (13.5%)	
Intestinal injury, n(%)	423	2 (0.5%)	52	12 (23.1%)	
Intubated field, n(%)	425	30 (7.1%)	52	31 (59.6%)	
Intubated ED, n(%)	425	34 (8.0%)	52	33 (63.5%)	
CPR in field, n(%)	425	2 (0.5%)	52	15 (28.8%)	
CPR in ED, n(%)	425	0 (0.0%)	52	15 (28.8%)	
Any blood transfused, n(%)	425	49 (11.53%)	52	46 (88.46%)	
Received blood transfusion pre-hospital, n(%)	425	22 (5.18%)	51	15 (28.8%)	
Labs & Work-up					

INR, median(IQR)	191	1.2 (1.1,1.3)	47	1.5 (1.3,2)	< 0.0001
Lactate, median(IQR)	59	2.5 (1.3,3.7)	29	4.4 (2.8,8.2)	< 0.0001
Base deficit, median(IQR)	124	-5.0 (-7.0,-3.0)	45	-9.0 (-14.0,-6.0)	< 0.0001
Pre-hospital SIPA, median(IQR)	393	1.0 (0.9,1.2)	50	1.4 (1.2,1.9)	0.2731
ED SIPA, median(IQR)	425	0.9 (0.7,1.1)	52	1.3 (1.0,1.8)	< 0.0001
R-time, median(IQR)	9	4.5 (4.2,5.0)	14	5.8 (4.5,7.3)	< 0.0001
Angle, median(IQR)	9	67.5 (61.1,69.3)	14	57.5 (45.5,67)	< 0.0001
MA, median(IQR)	9	59.9 (56.6,61.9)	14	55.0 (50.1,62.1)	< 0.0001
LY30, median(IQR)	4	0.0% (0.0, 3.0%)	13	0.0 (0,2.5%)	< 0.0001
Abnormal Values					
Abnormal INR value, n(%)	425	74 (17.4%)	52	43 (82.7%)	< 0.0001
Abnormal Base Deficit, n(%)	425	84 (19.8%)	52	41 (78.8%)	< 0.0001
Abnormal Lactate, n(%)	425	34 (8.0%)	52	28 (53.8%)	< 0.0001
Abnormal Pre-hospital SIPA, n(%)	425	204 (48.0%)	52	42 (80.8%)	0.2731
Abnormal ED SIPA, n(%)	425	109 (25.6%)	52	40 (76.9%)	< 0.0001
Abnormal R time, n(%)	425	3 (0.7%)	52	6 (11.5%)	< 0.0001
Abnormal MA, n(%)	425	2 (0.5%)	52	7 (13.5%)	< 0.0001
Abnormal LY30, n(%)	425	3 (0.7%)	52	12 (23.1%)	< 0.0001
Abnormal Angle, n(%)	425	1 (0.2%)	52	5 (9.6%)	0.1541
Abnormal Hemoglobin in ED, n(%)	425	117 (27.5%)	52	30 (57.7%)	< 0.0001
Abnormal Hemoglobin at 2 hours, n(%)	425	387 (91.1%)	52	42 (80.8%)	0.0199
Abnormal Hemoglobin at 4 hours, n(%)	425	374 (88.0%)	52	43 (82.7%)	0.2769

Abnormal ED HR, n(%)	425	243 (57.2%)	52	45 (86.5%)	< 0.0001
Abnormal ED BP, n(%)	425	183 (43.1%)	52	36 (69.2%)	< 0.0001
FAST positive, n(%)	425	83 (19.5%)	52	19 (32.7%)	0.0404
Outcomes					
Hospital length of stay days, median(IQR)	425	3.0 (2.0,5.0)	52	8.0 (2.0,16.0)	
Ventilation days, median(IQR)	425	0.0 (0.0,0.0)	52	2.0 (0.8,5.0)	
ICU length of stay, median(IQR)	425	0.0 (0.0,1.0)	52	3.0 (1.0,6.3)	
Required orthopedic surgery, n(%)	425	46 (10.8%)	52	4 (7.7%)	

Abbreviations: Glasgow coma score (GCS), International Normalized Ratio (INR), Interquartile range (IQR), Cardiopulmonary resuscitation (CPR), Emergency Department (ED), Shock index-pediatric adjusted (SIPA), R-time (R), MA (Maximum amplitude), thromboelastography lysis at 30 minutes (LY30), Heart rate (HR), Blood pressure (BP), Intensive care unit (ICU)

Table 2: Top 25 Features associated with Outcomes Based on Absolute Pearson's Correlation Coefficient Values

Massive Transfusion Features	R	Failed NOM (Required Surgery) Features	R	Mortality Features	R	Non-operative Management without Intervention Features	R
Required pRBCs at 4 hours	0.68	LY30	0.43	CPR in the field	0.68	GCS	0.53
Required FFP	0.65	Abnormal	0.42	CPR in the ED	0.66	CPR in ED	0.52

at 4 hours		LY30					
Required Platelets at 4 hours	0.58	R-time	0.40	Base Deficit	0.65	Intubated in the ED	0.50
Abnormal LY30	0.53	MA	0.38	INR	0.61	Intubated in the Field	0.49
Intubated in the ED	0.48	Angle	0.38	Lactate	0.58	ED SIPA	0.48
GCS	0.47	Abnormal INR	0.35	GCS	0.54	CPR in the Field	0.48
Intubated in the field	0.45	ED SIPA	0.34	Intubated in the ED	0.53	Abnormal INR	0.47
Abnormal MA	0.43	CPR in the ED	0.32	Intubated in the field	0.53	Abnormal Lactate	0.42
INR	0.43	Abnormal MA	0.32	Required FFP at 4 hours	0.50	Abnormal Base Deficit	0.42
ED SIPA	0.42	Abnormal Base Deficit	0.32	Required pRBCs at 4 hours	0.49	LY30	0.41
Base Deficit	0.42	Base Deficit	0.29	Head injury	0.46	Abnormal LY30	0.40
Head injury	0.41	ED Blood Pressure	0.29	ED SIPA	0.42	ED Blood Pressure	0.38
Received pRBCs pre-hospital	0.40	Abnormal Lactate	0.28	ED Blood Pressure	0.38	R time	0.36
Angle	0.39	GCS	0.28	Required Platelets at 4 hours	0.36	MA	0.35
Abnormal Angle	0.38	Abnormal ED SIPA	0.27	Pre-hospital Blood Pressure	0.35	Angle	0.35
CPR in the ED	0.37	Intubated in the ED	0.26	Pre-hospital SIPA	0.31	Lactate	0.35
Abnormal Base Deficit	0.37	Abnormal Angle	0.26	R-time	0.31	Base deficit	0.35
ED Blood Pressure	0.37	Abnormal R-time	0.26	Abnormal INR	0.31	Abnormal ED SIPA	0.34
Abnormal INR	0.37	Intubated in the field	0.26	Abnormal Base Deficit	0.29	Abnormal MA	0.30
Abnormal R-time	0.37	Lactate	0.23	Abnormal Lactate	0.28	INR	0.29
Lactate	0.35	Blood pressure at 4 hours	0.23	Received blood pre-hospital	0.27	Blood pressure 4 hours after presentation	0.28
Received blood pre-	0.34	INR	0.21	Presence of Intestinal	0.27	Abnormal Angle	0.26

hospital				Injury			
Received FFP pre-hospital	0.34	CPR in the field	0.21	Hemoglobin in ED	0.26	ED heart rate	0.25
MA	0.34	Heart rate at 4 hours	0.19	Failure of Non-operative management	0.26	Abnormal R time	0.25
Pre-hospital SIPA	0.32	Abnormal Pre-hospital SIPA	0.17	Hemoglobin at 2 hours	0.24	Heart rate at 4 hours	0.21

Abbreviations: packed red blood cells (pRBCs), frozen fresh plasma (FFP), thromboelastography lysis at 30 minutes (LY30), International Normalized ratio (INR), Shock index- pediatric adjusted (SIPA), cardiopulmonary resuscitation (CPR), Maxiumum amplitude (MA), Glasgow Coma score (GCS), Emergency department (ED)

Table 3: Demonstration of the 4 and 24 hour Deep Learning Models (Validation set)

Outcome	Model	Accuracy	Sensitivity	Specificity	AUC
Massive Transfusion	4 Hours	90.5%	88.9%	90.5%	0.90
	24 Hours	90.0%	88.9%	90.0%	0.90
Failure of NOM/Need Surgery	4 Hours	83.8%	91.7%	83.5%	0.88
	24 Hours	82.4%	91.7%	82.1%	0.87
Mortality	4 Hours	91.9%	100.0%	91.8%	0.96
	24 Hours	91.9%	100.0%	91.8%	0.96
Successful NOM without Intervention	4 Hours	90.3%	90.4%	88.2%	0.89
	24 Hours	86.9%	86.8%	88.2%	0.88