Predicting Cervical Spine Injuries in Children

Final Project: Statistics 215A, Fall 2021

Andy Shen, Licong Lin, Seunghoon Paik

1 Introduction and Domain Problem

Cervical spine injuries (CSI) include injuries sustained to the spinal cord connecting the head and neck to the rest of the body. These injuries, albeit rare, are extremely severe for children whose bodies are still developing. CSI detection strategies include the use of a **computed tomography (CT) scan**, among others. While these interventions are effective at pinpointing injuries, they can inadvertently expose children to ionizing radiation and cause unnecessary pain (Leonard et al. (2011)).

The rarity of a CSI, combined with the potentially harmful effects of CSI detection methods, force clinicians and specialists to assess the risk of performing CT scans against the potential of missing a critical injury diagnosis. The decision of whether to perform a CT scan or not is supported through a **clinical decision rule (CDR)** which uses prior data of potential CSI injuries to predict whether a patient has sustained a true CSI or not. Doctors use the results of the CDR and their own assessment of the situation to decide if the risk of a CT scan outweighs its potential benefits.

Current predictive capabilities in this regime have high sensitivity but low specificity values. This means that, while more children are undergoing CT scans, these scans do not detect injury, resulting in the aforementioned consequences. Leonard et al. (2011) performed multiple logistic regression analyses, resulting in 94% sensitivity with 39% specificity. **Therefore, our objective is to improve on current CDR methods by increasing specificity without sacrificing sensitivity, all while maintaining interpretability throughout the process.**

In this project, we develop a CDR for predicting CSI in children, based on data provided by the Pediatric Emergency Care Applied Research Network (PECARN). The data are described in Section 2. We ensure all components of our CDR are interpretable and realistically accessible to physicians when speculation of CSI is called into question. Throughout the modeling process, we ensure that our predictions are accurate, stable, and capture the breadth of possible injuries for each patient.

The structure of this report is as follows: In Section 2, we explain our data and the features used to predict CSI. In Section 3, we conduct an exploratory data analysis. Sections 4, 5, and 6 concern the modeling process and stability checks, followed by concluding remarks in Section 7.

2 Data Collection

Our data contains patient CSI information from a previous CDR study conducted by Leonard et al. (2011). The data consist of 12 datasets, each containing various categories of information about the patients, such as demographics, medical history, mechanism of injury, their appearance at the time of response and arrival at the hospital, among others.

The previous study contained children both with and without CSI. The children without CSI were placed into three control groups: mechanism of injury (MOI), emergency medical services (EMS) and a random control group.

The MOI group is created by matching non-injured patients to injured patients with a similar injury mechanism (1012 patients), The EMS group is created by matching non-injured patients who required EMS out-of-hospital care to injured patients who also required this service (702 patients). The random control group (1060 patients) is simply a random matching of patients were who were not injured to patients who were injured. There were 540 patients in the case group (the truly injured patients)

In our study, we are only interested in distinguishing true injury from simple trauma. We group all of the sub-control groups into a single control group for more refined predictions.

Some datasets have the same name, but are denoted with an `field`, `out` or `site` at the end of the file name (e.g `clinicalpresentationfield` vs `clinicalpresentationsite`). The suffix "field" denotes measurements of the patient at the location where the injury was sustained, such as the site of the car crash or the patient's home. The suffix "out" refers to information reported by an outside hospital referring the patient to the PECARN system. Note that this information is not available for patients who were not referred through an outside hospital. The suffix "site" refers to information collected at the PECARN hospital site. **The site information is always collected last**, with the outside hospital information collected between field and site if pertinent.

It is also important to recognize that this data is extremely prone to human and/or machine error. We describe potential lapses in the data collection process in Section 2.1. The data cleaning process is discussed in Section 3.2.

2.1 Meaning

In order to better understand the variables in our data, we separate them into **three distinct categories: demographic information, injury mechanism, and trauma presentation.** Demographic information simply refers to the patient's basic information, such as age, gender, and ethnicity. While these features may have very little predictive power, they can be useful in grouping injured or non-injured patients post-hoc.

Injury mechanism refers to how the patient sustained their trauma. Examples include a vehicle accident, child abuse, assault, or falling. Each category of injury is then further elucidated into more specific modes of injury (such as a rear-end collision vs a side impact collision). Most mechanisms of injury in the given data are injuries sustained from various youth sports or from a motor vehicle accident.

Finally, trauma presentation refers to the condition of the patient when they arrive to the hospital or at the site of trauma. These features are spanned across multiple datasets. Features in this category include whether the patient is conscious or not, whether they report neck or facial pain, whether their mental status appears to be normal, among others.

Note here that some features across these three categories preclude accurate reporting of others. For example, if a patient is unconscious or cannot communicate properly, they will be unable to properly describe whether they are in pain or not.

For our study, we prioritize using the "site" data since it is most recent in terms of a patient's trauma presentation. Features in the other two categories (demographics and injury mechanism) do not differ across location.

The non-demographic variables all measure some aspect of the patient's trauma and allow us to assess how serious it is. Each piece of data was recorded from multiple different perspectives. The measurements were collected differently across 2-3 different locations and at 2-3 different time periods – the actual values could differ across time and location. Moreover, the instruments used to take each measurement could be calibrated differently or the individual taking the measurement could have made a mistake. It is very important to keep these sources of error in mind in the modeling process and when questioning unusual observations.

By assuming our data is correct, we are assuming that no errors were made in the data recording process. We also make other assumptions stated in more detail in Section 2.2.

2.2 Relevance and Randomness

The data provided are relevant to our study. The features described in our model are all accessible to a physician when they make their clinical decision. If any additional data is necessary, it could include other potentially useful features that are typically made available to a physician at the time of making the clinical decision. We do not collect more data since we are provided with all of the data recorded by PECARN, and it is highly unlikely that additional data that is pertinent to this study even exists.

Moreover, this problem is an observational study. Patients are assigned to case or control group based on whether they have a CSI, and each patient in the case group has several matching controls. Thus, the patients in our data cannot be treated as totally independent observations from a population. However, the dependence among the patients only exists within each matching pair, and different matching pairs can be treated as independent. By using random splitting (discussed in Section 4.1), we ensure that our training, validation and test sets have roughly the same distribution conditioned on the given data, and therefore we can hope that the models fitted using training data generalize well to the test set.

3 Exploratory Data Analysis

3.1 Reducing the feature space

Our initial feature space included over 300 features and an interpretable model only requires a small fraction of these features. We trim down our initial subset of features through the consultation of domain experts in cervical spine injury. Utilizing the assistance of Dr. Michael Boyle, a physician in the UC San Francisco School of Medicine, we reduced our feature space to roughly 51 variables to perform variable selection on.

The omitted features were mostly redundant information or information that would not be provided to a physician when making a clinical decision. These omissions were largely determined through domain-expert judgment calls. When in doubt, we retained a questionable feature to avoid potentially losing crucial predictive information. As discussed in Section 4.2, we use a more rigorous variable selection process after this step to determine our best features for modeling.

3.2 Data Cleaning

In order to create an interpretable decision rule, we map our categorical features into binary 0-1 features, **with 1 corresponding to "yes" and 0 corresponding to "no."** However, not all features can be uniformly binarized with this mapping, so we implement several layers of judgment calls described below:

Ambiguous values: Certain features convey a degree of ambiguity in the recorded value. This could indicate a variety of things such as whether the data was unavailable to the nurse or EMT, or if the patient is in a state where a reliable response cannot be elicited (such as being inebriated or unconscious). These values are imputed using *defensive imputation* which assumes the worst-case scenario for the patient. For example, a value of `S` which stands for "suspected, but unknown" would map to "yes." Fortunately, the data contains documentation which provides meaning to the possible values for each feature. Note that this does not default to imputing all missing values with "yes." We must consider what each feature is asking for and determine the worst-case scenario accordingly.

Ordinal features: Our feature space contains an ordinal feature, `AVPU`, which stands for "Alert, Verbal, Pain, Unresponsive." This is a widely accepted health care protocol to determine a patient's overall condition. Each letter in "AVPU" corresponds to a more urgent patient status. Alertness and unresponsiveness are simply whether the patient is alert or unresponsive. "Verbal" and "Pain" correspond to whether the patient is receptive to verbal or painful stimulation. In our case, we treat anything that is not alert as the worst-case scenario and impute accordingly. Failing to respond to any stimuli whatsoever should be heavily scrutinized and our imputation takes this into account.

Missing values: Certain columns in the data contain blank or missing values. Features with a proportion of missing values above 15% are removed from this analysis. The remaining missing values are imputed using the *median* value of the binary features. In other words, we impute missing values with whatever response (0 or 1) has the highest frequency for the non-missing values. For instance, if we do not have data on whether a patient has tenderness in their neck, we impute based on which value occurs most frequently.

All judgment calls that directly involved medical subject-matter expertise were verified by Dr. Boyle.

Our final cleaned data contains 2774 control patients and 540 patients who were truly injured with CSI.

3.3 Correlation between features

Figure 1 shows a correlation heatmap of the reduced feature space of size 44 (excluding the outcome). While most features are not strongly correlated with each other, there are some features that exhibit high correlations. These features are typically those of the same type, such as whether the patient reported pain in their neck and whether tenderness was observed in the patient's neck. It is expected for features like these to be influenced by one another.

Moreover, most features have absolute correlation $|\rho| < 0.5$, indicating weak association throughout the feature space in general. Because of this, we feel confident that we will not encounter any issues surrounding variance inflation and multicollinearity in the modeling stage.

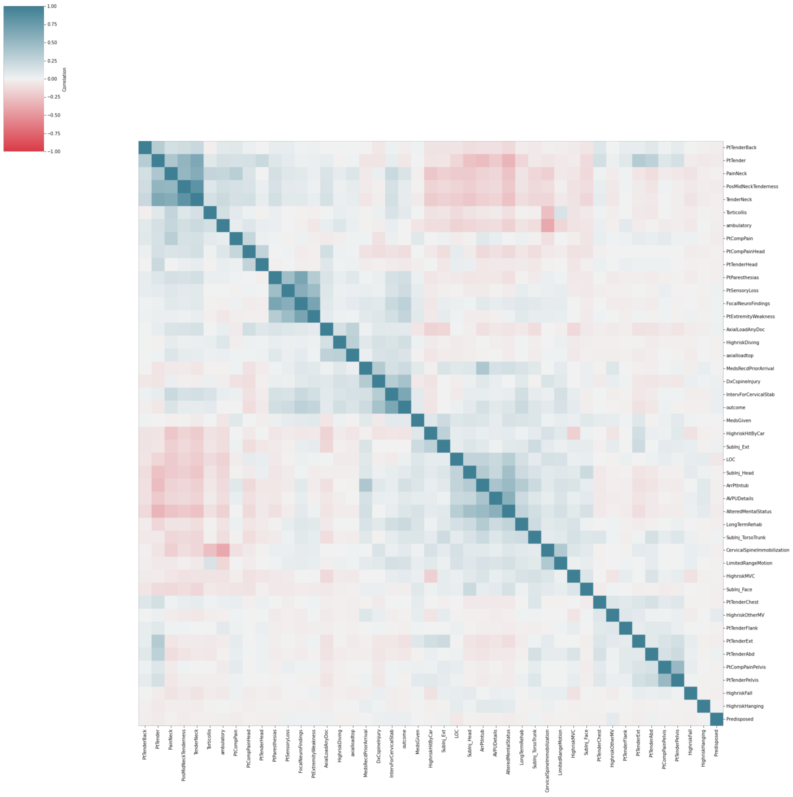


Figure 1: Correlation heatmap of reduced feature space of 44 features.

3.4 Correlation with outcome

After seeing which features are associated with each other, we then examine the correlation of each feature with respect to the outcome variable (whether the patient truly has a CSI). We find that most features are not highly correlated with the outcome in either direction, with some exceptions: the `IntervForCervicalStab` feature, which tells whether the patient underwent any cervical stabilization measures at the site, has the highest correlation (above 0.60) with the outcome. This should not be surprising since a cervical stabilization measure is only imposed if a medical official on-site thinks it is necessary, so we expect most true injuries to be highly associated with this intervention.

3.5 Frequency of feature values

We further explore our data by examining the frequency of “yes” and “no” values for the case and controls group. We plot the five features with the greatest and smallest *absolute differences* in relative frequency for the case group in Figure 2. We repeat this for the controls group in Figure 3.

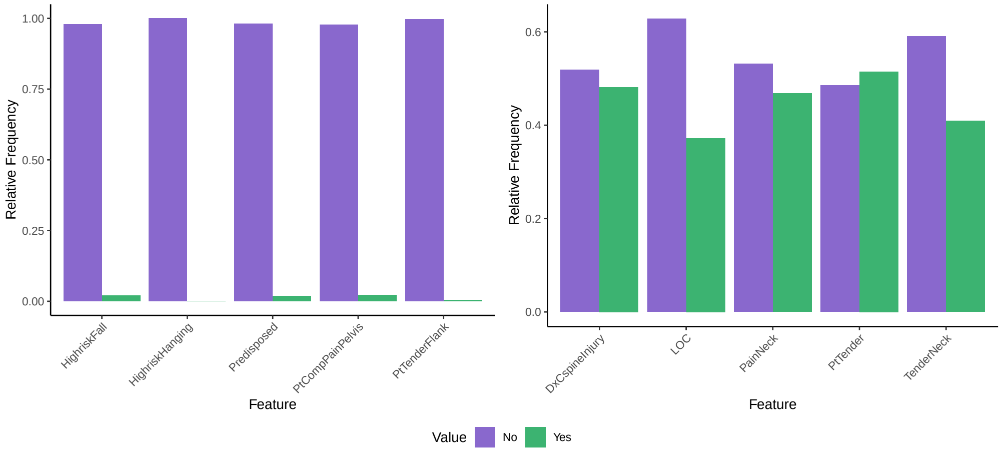


Figure 2: Top five features with the greatest and smallest absolute differences in relative frequency for the CASE group.

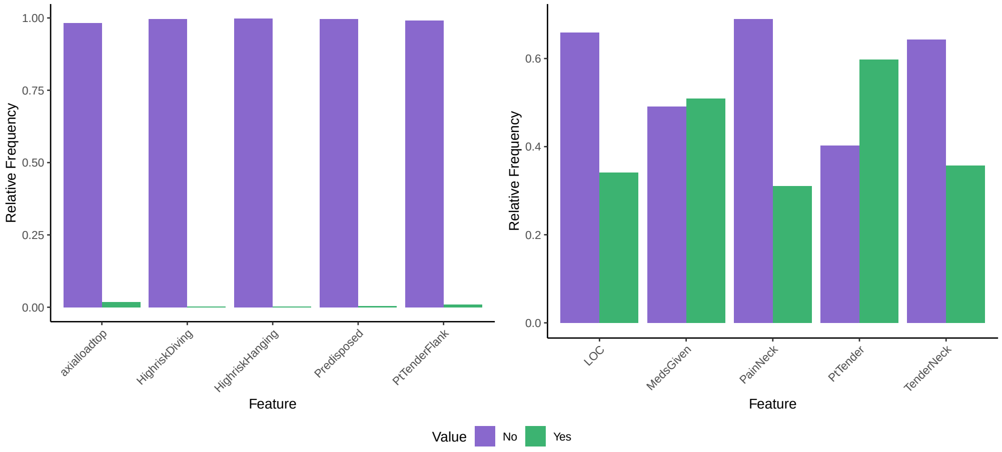


Figure 3: Top five features with the greatest and smallest absolute differences in relative frequency for the CONTROLS group.

These figures show that, while there is some overlap across the two sub-groups, we cannot say that there is complete uniformity in which features stand out between the case and control groups. For instance, the `DxCspineInury` feature, which tells us whether the patient is arriving with a diagnosis or suspicion of a CSI, has the second-lowest difference in yes/no rate for the case group (48% vs 52%), whereas it is somewhere in the middle for the control group (where roughly 92% were not suspected). This feature proves to be extremely crucial in the modeling stage and it is discussed further in Section 4.4.

4 Modeling

4.1 Translation and Comparability

Our statistical goal is to accurately diagnose a cervical spine injury (minimize Type II error) while also keeping the Type I error rate as low as possible.[1] In the context of CSI, a Type II error occurs when our CDR determines no CSI is present when when the patient has actually sustained a CSI. This statistical flaw bears serious and life-threatening consequences on children and we must pay special attention to this rate. A Type I error occurs when our CDR determines CSI is present even though the patient is actually not injured.

Reducing the Type I error rate can reduce the number of unnecessary CT scans and hence improve the efficiency. Our goal is then translated into **maximizing sensitivity while maintaining a sufficiently high specificity**.

Note that if Type II errors are our only area of concern, there is no need for any decision rule since a clinician would diagnose a CSI and perform a CT scan for all situations. As mentioned in the Introduction (Section 1), CT scans could result in unnecessary exposure to ionizing radiation, which is also harmful. Therefore, we must strike a balance between taking the conservative route of always diagnosing someone with CSI and the worst-case scenario of not diagnosing a serious injury.

Regarding comparability, the training and test data come from the PECARN sites and are independent. We have no reason to assume that the data comes from another distribution. Moreover, the injury status of one patient is independent of another patient's injury status. Even if two children were injured from one another (colliding in a soccer game, for example), whether one child sustains a CSI has no bearing on whether the other child does. This is an extreme example and most data points are indeed completely independent.

Prior to modeling, we split our data into *training* (60% of data), *tuning* (20% of data), and *testing* (20% of data) sets. The training set will be used to construct our candidate models. The tuning set will be used to select an optimal model, and the testing set will be used to corroborate the performance of our models.

These splits were determined randomly as there is no underlying temporal or spatial phenomena that makes the outcome of one patient dependent on another.

4.2 Feature and hyperparameter selection

We use **logistic regression** followed by **backwards selection** to select our model features. We first fit the outcome against all features with logistic regression and then omit the feature with the smallest absolute fitted coefficient value. There is no need to scale our features since they are all binary. This process repeats until m features remain, where m is a pre-specified number. This variable selection process is called recursive feature elimination (RFE) in Python and can be done automatically.

To select the number of features m, we run logistic regression model with $m = 1, 2, ..., 10$ features on the training data and select the one with the lowest misclassification rate on the tuning data, resulting in an optimal value of $m = 9$. However, the stability check performed in Section 5 reveals that 8 features results in a more optimal CDR. Seeing that the data science life cycle is an iterative process, we select 8 features in our models. These 8 features correspond to:

- Whether the patient was intubated or not (`ArrPtIntub`),

[1] Here we slightly abuse the concept of Type I and II error since there is no distributional assumption in our problem.

- Whether the patient is suspected of having a CSI (`DxCspineInjury`) (see Section 4.4),
- Whether the patient has focal neurological deficits such as a spinal cord issue (`FocalNeuroFindings`),
- If the sustained trauma was the result of diving (`HighriskDiving`),
- Whether the patient required a cervical stability intervention such as a collar or brace (`IntervForCervicalStab`),
- Whether the patient has extremity weakness (`PtExtremityWeakness`),
- Whether the patient has sensory loss (`PtSensoryLoss`), and
- If the trauma was sustained to the patient's torso or trunk (`SubInj_TorsoTrunk`).

These features were discussed with Dr. Boyle, who verified their legitimacy in a real-life clinical context. Moreover, common sense knowledge tells us that these features are meaningful and serious with regards to CSI's.

4.3 Candidate Models

We create three candidate CDR's using three different learning techniques: AdaBoost, Decision Trees, and Logistic Regression. Each model has their own strengths and weaknesses in terms of interpretability and predictability. For completeness, we also include the baseline model, Bayesian Rule List, Greedy Rule List (GRL), and RuleFit models provided in the code skeleton.

The sensitivity and specificity rates for all three models are shown in Figure 4. These values are generated from the tune dataset. The corresponding Receiver Operating Characteristic (ROC) curves for all three models are shown in Figure 5. In general, we achieve over three times the specificity of Leonard et al. (2011) while maintaining the same sensitivity. The results from Leonard et al. (2011) are shown in the black dot. Statistically speaking, we increase the power over threefold without sacrificing the potential for a Type I error.

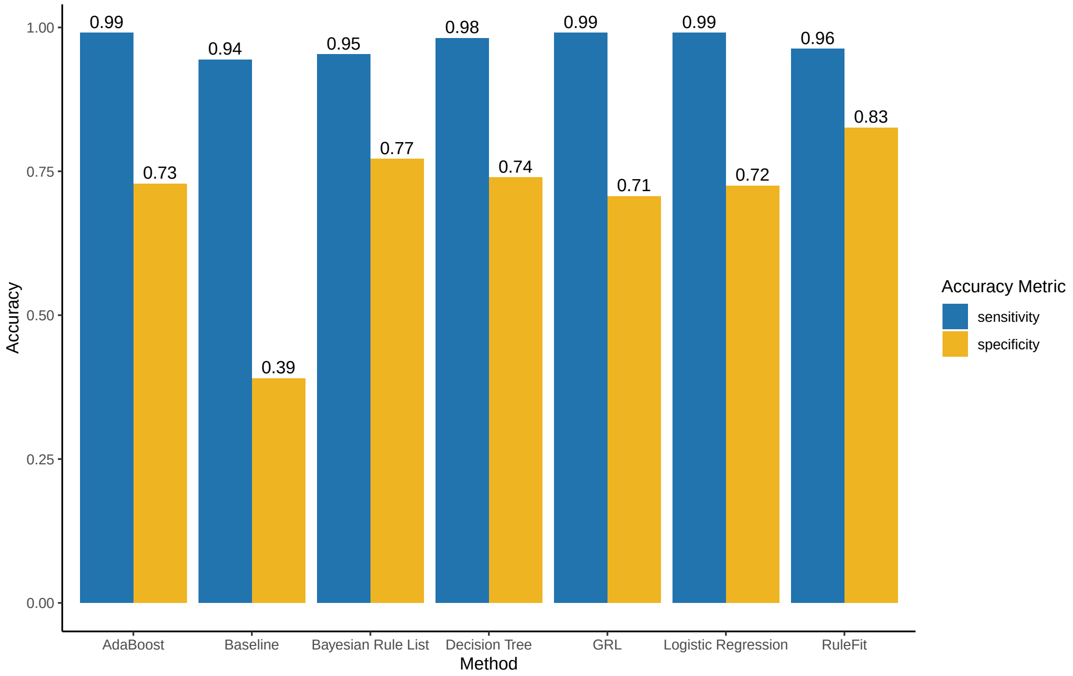


Figure 4: Sensitivity and specificity rates for candidate models and baseline model (second from left). Accuracy rates are assessed at each model's optimal hyperparameter value.

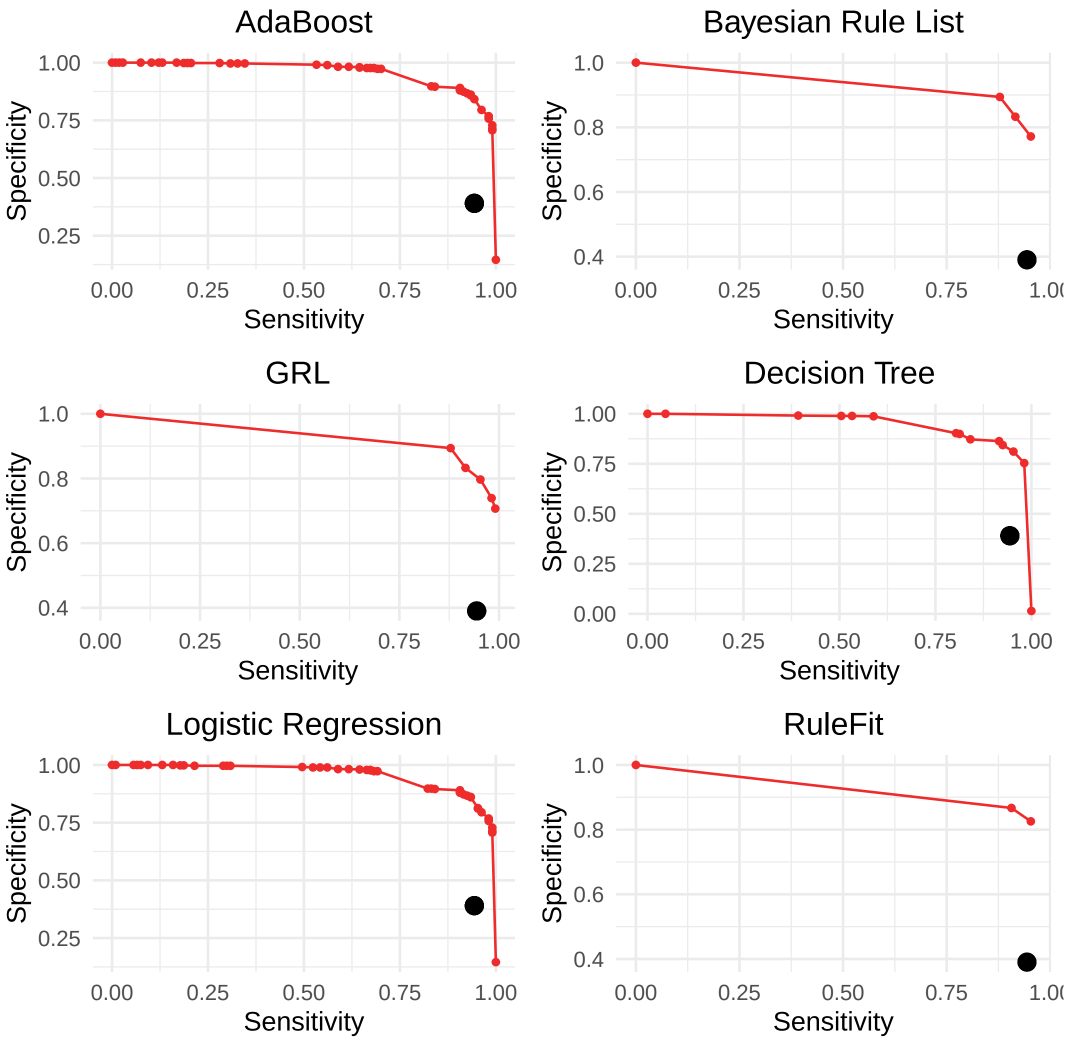
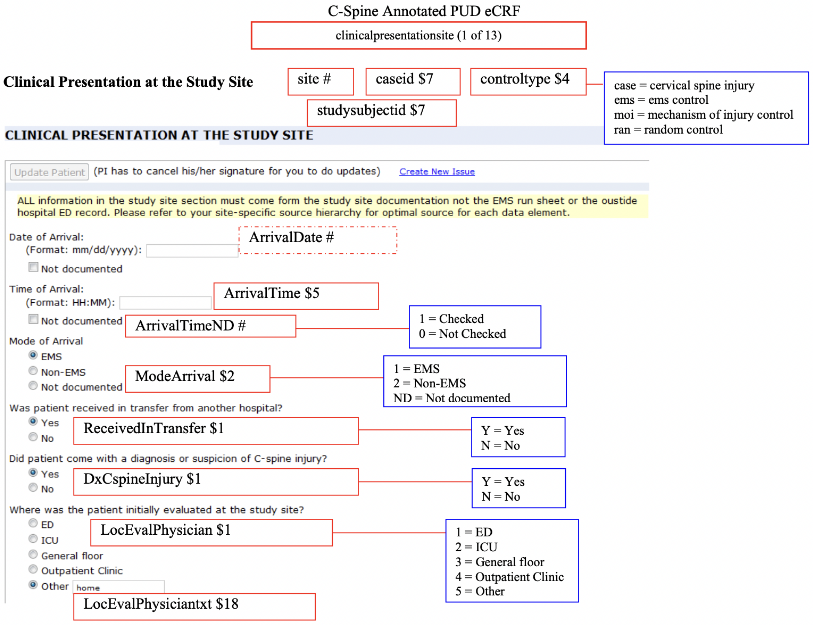


Figure 5: ROC curves for the three candidate models. The black dot represents the CDR from Leonard et al.

4.4 Subjective Features

The `DxCspineInjury` feature (explained in Section 3.4) has a unique context. This feature is literally a guess as to whether this person has a CSI or if they arrived with a diagnosis. Such information could be revealing of the true outcome even though it is a feature made available to a physician prior to making the CDR. In order to use a feature like this, we must provide a thorough investigation of the feature and ensure that it is indeed reasonable.

A rigorous inspection of the raw data reveals that this feature has **zero** missing or indeterminate values. Thus, we can rely on this information to be provided in general. We then examined how this feature was recorded. Fortunately, our data documentation contains a copy of the form that a nurse or EMT fills out when examining the patient. The question of CSI suspicion appears in the survey along with the other questions that are part of our feature space. If we envision ourselves as the nurse or physician, we can see them record their personal judgment of the patient just like they will record other measurements, such as AVPU or neck tenderness, for example. A copy of the questionnaire with the `DxCspineInjury` feature is shown in Figure 6.



C-Spine Annotated PUD eCRF

clinicalpresentationsite (1 of 13)

Clinical Presentation at the Study Site site # caseid $7 controltype $4 studysubjectid $7

case = cervical spine injury
ems = ems control
moi = mechanism of injury control
ran = random control

CLINICAL PRESENTATION AT THE STUDY SITE

(PI has to cancel his/her signature for you to do updates) Create New Issue

ALL information in the study site section must come form the study site documentation not the EMS run sheet or the oustide hospital ED record. Please refer to your site-specific source hierarchy for optimal source for each data element.

Date of Arrival:
(Format: mm/dd/yyyy): ArrivalDate #
□ Not documented

Time of Arrival:
(Format: HH:MM): ArrivalTime $5
□ Not documented ArrivalTimeND #

1 = Checked
0 = Not Checked

Mode of Arrival
● EMS
○ Non-EMS
○ Not documented
ModeArrival $2

1 = EMS
2 = Non-EMS
ND = Not documented

Was patient received in transfer from another hospital?
● Yes
○ No
ReceivedInTransfer $1

Y = Yes
N = No

Did patient come with a diagnosis or suspicion of C-spine injury?
● Yes
○ No
DxCspineInjury $1

Y = Yes
N = No

Where was the patient initially evaluated at the study site?
○ ED
○ ICU
○ General floor
○ Outpatient Clinic
● Other home
LocEvalPhysician $1
LocEvalPhysiciantxt $18

1 = ED
2 = ICU
3 = General floor
4 = Outpatient Clinic
5 = Other

Figure 6: Copy of data recording mechanism for CSI suspicion.

Moreover, subject-matter expertise tells us that there does indeed exist precedents for utilizing judgment calls in decision rules. Dr. Boyle verified that many current decision instruments utilize physician suspicion, such as the Wells' Criteria for Pulmonary Embolism[2] and the HEART Score[3]. Seeing that such features are utilized in everyday medical practice, there is no reason to exclude it from a decision rule.

Statistically speaking, as depicted in Figures 2 and 3, this guess is only correct about 50% of the time in the case group compared to 92% of the time in the controls group. We can infer that most of these guesses are not in favor of a CSI. This is indeed true, as roughly 85% of all patients are not predicted to have a CSI.

All of the arguments above were discussed in detail with Dr. Boyle. Therefore, we are confident that this feature can be utilized in a real-world context. Not only is its distribution statistically reasonable, but we have no reason to believe that this feature is fundamentally different from other features in terms of how it is documented. The only difference is that it relies on a human judgment call, but such judgment calls are commonly utilized in medical practices today.

5 Stability

We verify the stability of our models using three different perturbations. **The perturbations are implemented independently from each other.**

We created 10 different train/tune/test set combinations using our existing data. For all 10 data splits, we checked the following:

- The stability of the feature selection process described in Section 4.2,
- The stability of the sensitivity and the specificity reported in Section 4.3, and
- How the model responds when we change the control group.

[2] https://www.mdcalc.com/wells-criteria-pulmonary-embolism
[3] https://www.mdcalc.com/heart-score-major-cardiac-events

5.1 Feature Selection

Table 1 below shows the count of the features we chose in Section 4.2 for the 10 re-generated train/tune/test sets.

Naturally, we observe slight shifts in the selected features for the perturbed data. These extra features include `HighriskFall` (3 occurrences), `Predisposed` (2 occurrences), `LongTermRehab`, `HighriskMVC`, `SubInj_Face`, `PtTenderBack`, and `PtTenderChest` which all occur one time. The remaining selected features were all included in our original model's. Our variable selection process is quite stable and we are confident in our choice of features.

Table 1: Feature (re-)selection counts on the 10 perturbed datasets.

Feature	Occurrences
ArrPtIntub	10
DxCspineInjury	10
HighriskDiving	10
IntervForCervicalStab	10
FocalNeuroFindings	8
PtExtremityWeakness	8
PtSensoryLoss	8
SubInj_TorsoTrunk	6

We mention in Section 4.2 that our original feature selection process yielded 9 variables. The ninth feature was `PtTenderExt` which does not show up in any of the selected features for the 10 perturbed datasets. This stability check allowed us to re-update our feature space by omitting `PtTenderExt` from our candidate models.

5.2 Specificity and Sensitivity

Using the same 8 variables delineated in Section 4.2, we compute the average sensitivity and specificity for the perturbed tune data, shown in Table 2. The predictability is stable and Adaboost and Logistic regression consistently demonstrate similar sensitivity and specificity rates compared to our original model. Thus, we are confident in the stability of our accuracy rates.

Table 2: Average accuracy rates across the 10 perturbed datasets.

	Sensitivity	Specificity
AdaBoost	0.981	0.728
Decision Tree	0.971	0.752
Logistic Regression	0.981	0.732

5.3 Different control groups

We then check our models' stability by using only one type of control (Random EMS, MOI) at a time instead of grouping all control groups into a singular homogeneous control group. The results are shown in Figure 7.

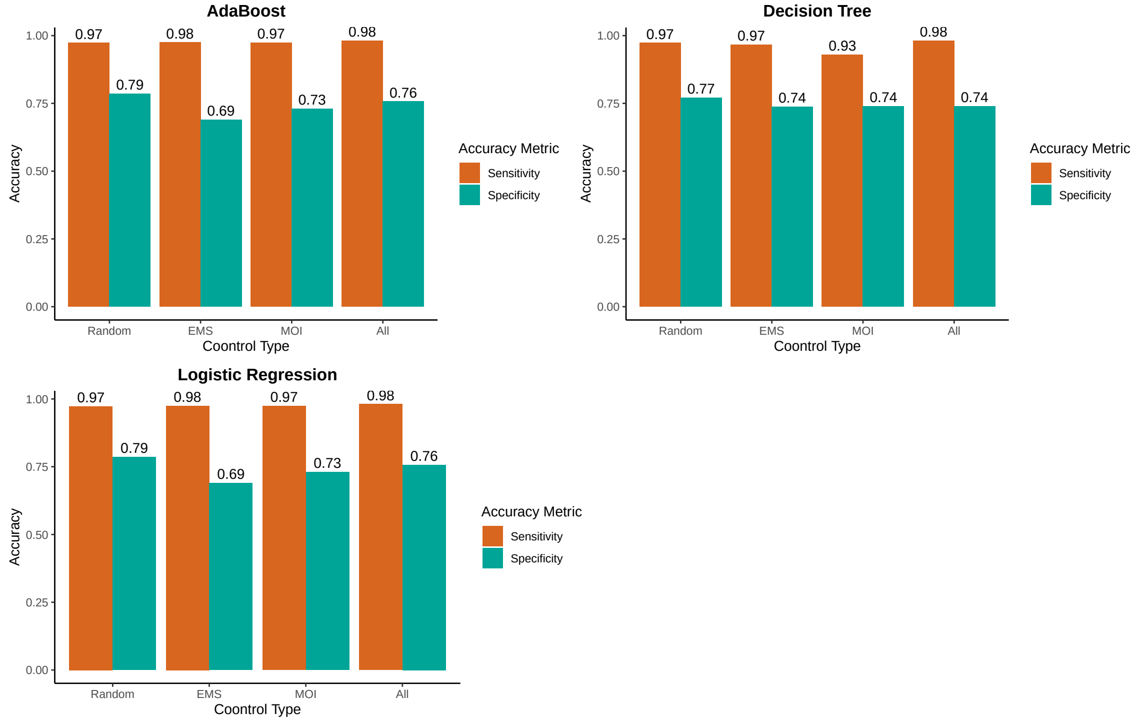


Figure 7: Sensitivity and specificity rates for three models with varying control types.

We see that, for all three models, the sensitivity is higher (above 98%) if we use group all controls uniformly. This makes sense since the prediction accuracy is likely to increase if we have a larger training set. When comparing one control at a time, all three models can achieve sensitivity above 97% with about 70% specificity except for the decision tree with MOI control.

In general, this stability check demonstrates that our models are robust to the choice of control group in almost every case. The only counterexample here is probably due to the discrete nature of decision tree, in which the confidence output (the confidence of the model that a patient has a cervical injury) takes its value in a finite set whose cardinality is the number of leaves. The performance of Adaboost and Logistic Regression are identical in all cases. This suggests that Adaboost and Logistic Regression can converge to the same model after training in our specific problem.

6 Test set performance

We evaluate our three models on the test set, the results of which are shown in Table 3. The thresholds used for these algorithms are obtained from the tuning set.

We see that all three models achieve sensitivities above 98% on the test set with over 74% specificity. These results are very similar to their counterparts on the tuning set. This all indicates that our models do not suffer from overfitting and can generalize well to new data from the same population. Note that the both the sensitivity and specificity of the baseline model are strikingly lower than the other models.

Table 3: Performance of models on test data.

	Sensitivity	Specificity
AdaBoost	0.983	0.763
Tree	0.983	0.742

	Sensitivity	Specificity
Logistic	0.983	0.763
Baseline	0.944	0.388

Ultimately, we believe AdaBoost is the best model for a clinical decision rule. In our study, Adaboost is implemented by averaging the outcomes of 100 adaptively generated one-step decision trees that classify the patient using the value of *a single feature*. The weak learner in Adaboost is the one-step decision tree. Therefore, is interpretable and easy-to-follow without sacrificing specificity. A clinician at a patient's bedside can quickly reach a decision simply by answering a series of yes or no questions, which is effectively what AdaBoost is.

7 Conclusion

Using the PECARN dataset, we discovered the important features to diagnose CSI in children and created a CDR that is powerful and interpretable. Our candidate models show higher sensitivity with much higher specificity than the current literature. Among the selected features, `FocalNeuroFindings`, `HighriskDiving`, and `SubInj_TorsoTrunk` are present for all three control groups in the original paper. Additionally, they are the only features that both the original paper and our study have in common.

There are similarities and differences between this study and the previous work in terms of variable selection. Interestingly, both studies selected 8 features for modeling. The choices, however, are not the same. Our feature space includes features utilized in Leonard et al. (2011) among others. However, Leonard et al. (2011) restricted themselves to a single dataset out of the 12 provided, whereas we utilized information from additional datasets. We believe this is a key reason for the stronger performance of our model, specifically with regards to specificity.

There is a lot of power and credibility attached to a statistician. One can always speculate the performance of stronger models to diagnose CSI. However, interpretability is a bedrock for others to understand the value of statistics and data science. Projects like these demonstrate the importance of creating a powerful product that others can understand.

8 Acknowledgments

The authors would like to thank Michael Boyle from the UCSF School of Medicine for his advice and expertise in the EDA and modeling process. We would not have been able to assess the real-world components of our model without this subject-matter expertise.

Finally, the authors would like to thank the STAT 215A teaching team, Bin Yu and Omer Ronen, for providing invaluable support and teaching us how to be active learners. This project, along with the many lessons we learned throughout the semester, will serve us well in our future careers as statisticians.

References

Leonard, Julie C, Nathan Kuppermann, Cody Olsen, Lynn Babcock-Cimpello, Kathleen Brown, Prashant Mahajan, Kathleen M Adelgais, et al. 2011. "Factors Associated with Cervical Spine Injury in Children After Blunt Trauma." *Annals of Emergency Medicine* 58 (2): 145–55.