Exploring Pediatric Appendicitis

Muhammad al-Akhdar muhammad3lakder@gmail.com

2025-02-17

Abstract

In this analysis, we explore the challenges of diagnosing appendicitis, specifically in pediatrics, by analyzing clinical and ultrasound data. We used a concise data science workflow that included data cleaning, exploratory visualizations, hypothesis testing, and logistic regression modeling. This approach not only revealed key differences in clinical characteristics, but also demonstrated how our analysis can improve healthcare decision making. We also discuss the limitations of the current dataset and methodology, and invite constructive feedback and collaboration for further refinement.

Introdution

In 2015, approximately 11.6 million cases of appendicitis were reported, resulting in approximately 50,100 deaths worldwide, deaths could be attributed to many reasons, such as lack of access to medicine, quality of care, and medical diagnostic error.

Helping healthcare professionals make more informed decisions is one way to reduce diagnostic errors.

There are many criteria to do this, such as history taking, physical examination, risk scores (e.g., Alvarado score), and imaging techniques such as ultrasound and CT.

The accuracy of the diagnosis could be improved, for example, by using machine learning. A recent study built a model to predict appendicitis in pediatrics using an interpretable unsupervised machine learning method.

Since many models have been built using only history and physical exam as predictors, we would use the same dataset to first explore the disease and then build models focusing on ultrasound as a diagnostic tool.

Methodology

The dataset was acquired in a retrospective study from a cohort of pediatric patients admitted with abdominal pain to Children's Hospital St. Hedwig in Regensburg, Germany.

- Taking a look at the dataset

Table 1: First Ten Rows of the Pediatric Patients Dataset

Sex	US_Performed	Severity	Management	Diagnosis
female	yes	complicated	primary surgical	appendicitis

Sex	US_Performed	Severity	Management	Diagnosis
male	yes	uncomplicated	conservative	appendicitis
female	yes	uncomplicated	conservative	no appendicitis
male	yes	uncomplicated	conservative	appendicitis
female	yes	uncomplicated	conservative	no appendicitis
male	yes	uncomplicated	conservative	no appendicitis
female	yes	complicated	primary surgical	appendicitis
male	yes	uncomplicated	primary surgical	appendicitis
male	yes	uncomplicated	conservative	appendicitis
female	yes	uncomplicated	conservative	no appendicitis

```
## Number of missing values is 14008
```

Some Plots

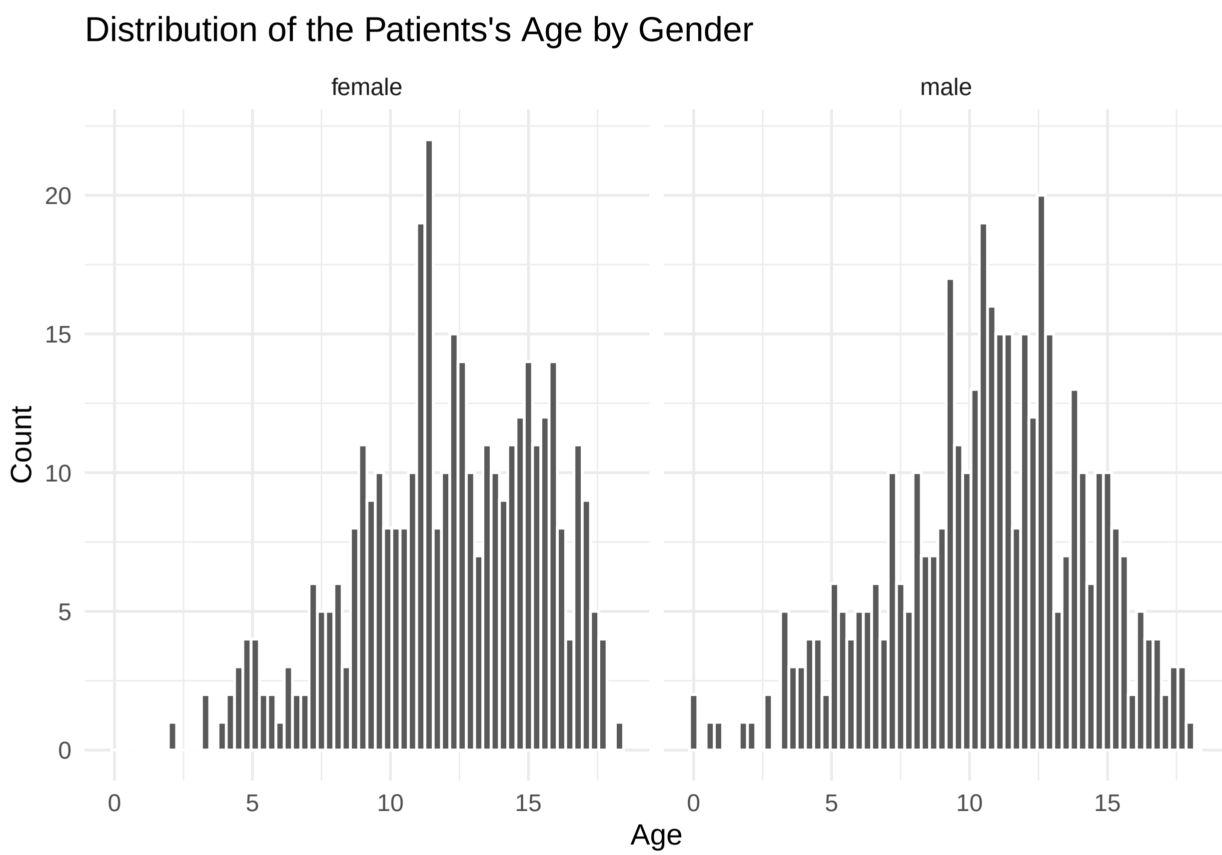


Figure 1: Distribution of the Patients's Age by Gender

Table 2: The Mean Age of Patients By Gender

Sex	Mean
female	12.06
male	10.68

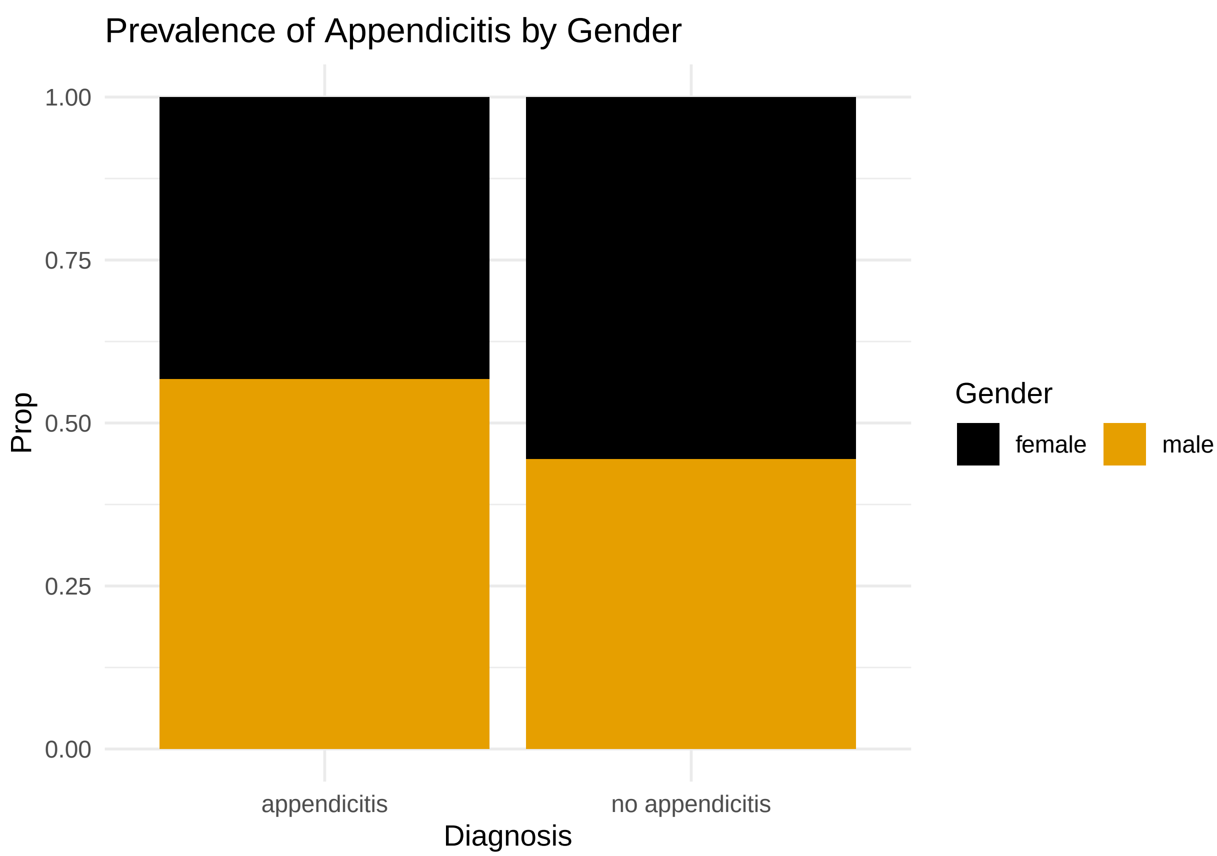


Figure 2: Prevalence of Appendicitis by Gender

Table 3: Prevalence of Appendicitis by Gender

Sex	Diagnosis	n	p
female	appendicitis	200	0.53
female	no appendicitis	176	0.47
male	appendicitis	262	0.65
male	no appendicitis	141	0.35

- Figure 2 shows that the prevalence in appendicitis is more males than females, which is consistent with existing findings, but it is not that substantial.

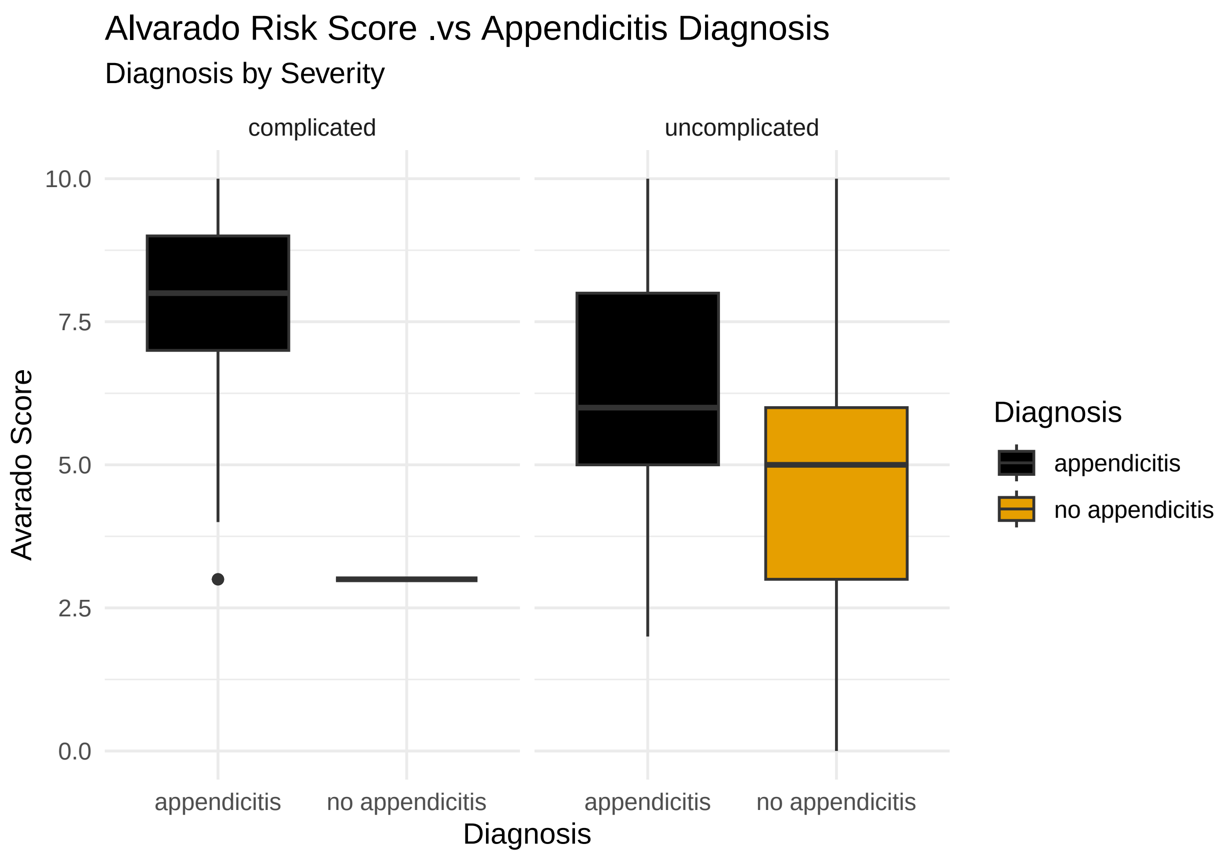


Figure 3: Alvarado Risk Score .vs Appendicitis Diagnosis

Table 4: Alvarado Risk Score .vs Appendicitis Diagnosis

Diagnosis	mean	median
appendicitis	6.67	7
no appendicitis	4.83	5

- Alvarado score is a system that have been developed to identify people who are likely to have appendicitis, as a score below 5 suggests against a diagnosis of appendicitis, while a score of 7 or more is predictive of acute appendicitis, but it is performance varies. Here, the severity of the diagnosis was added to see if the score also differed.

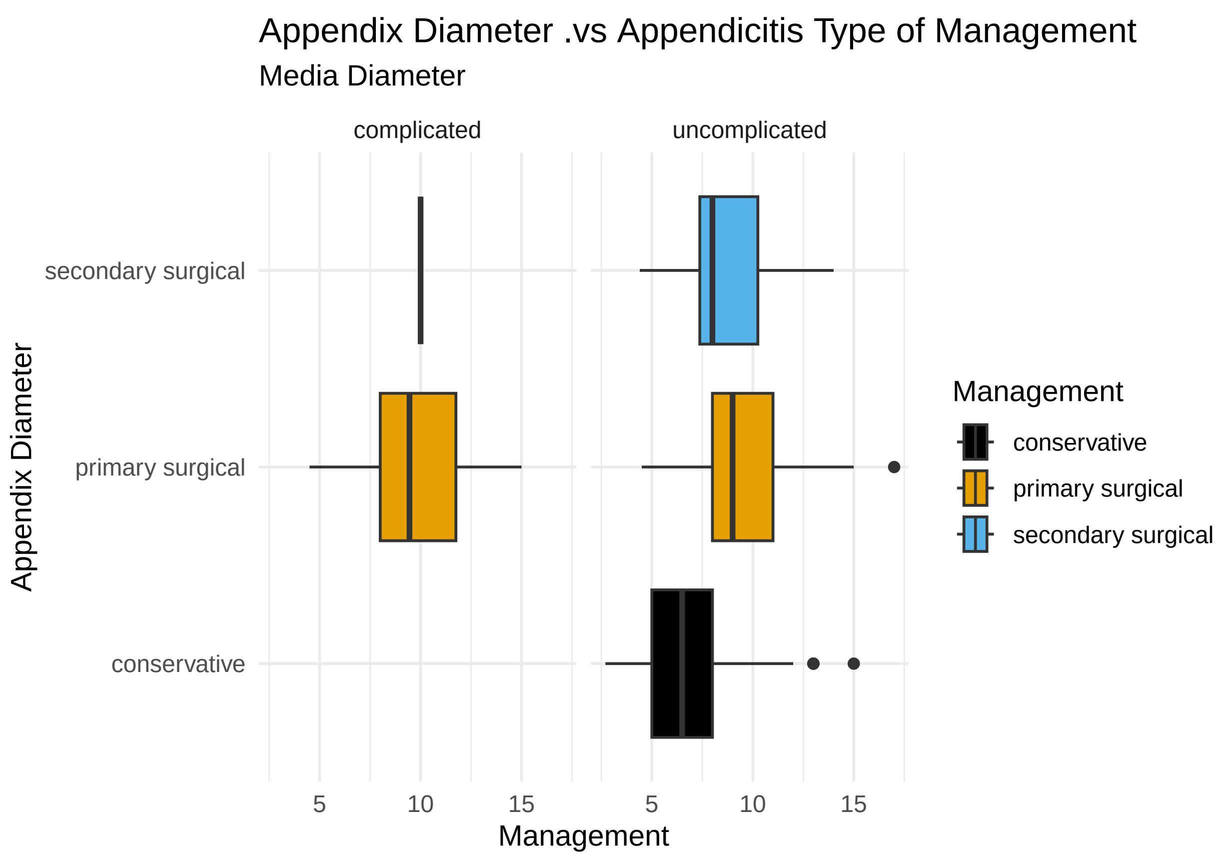


Figure 4: Appendix Diameter .vs Appendicitis Type of Management

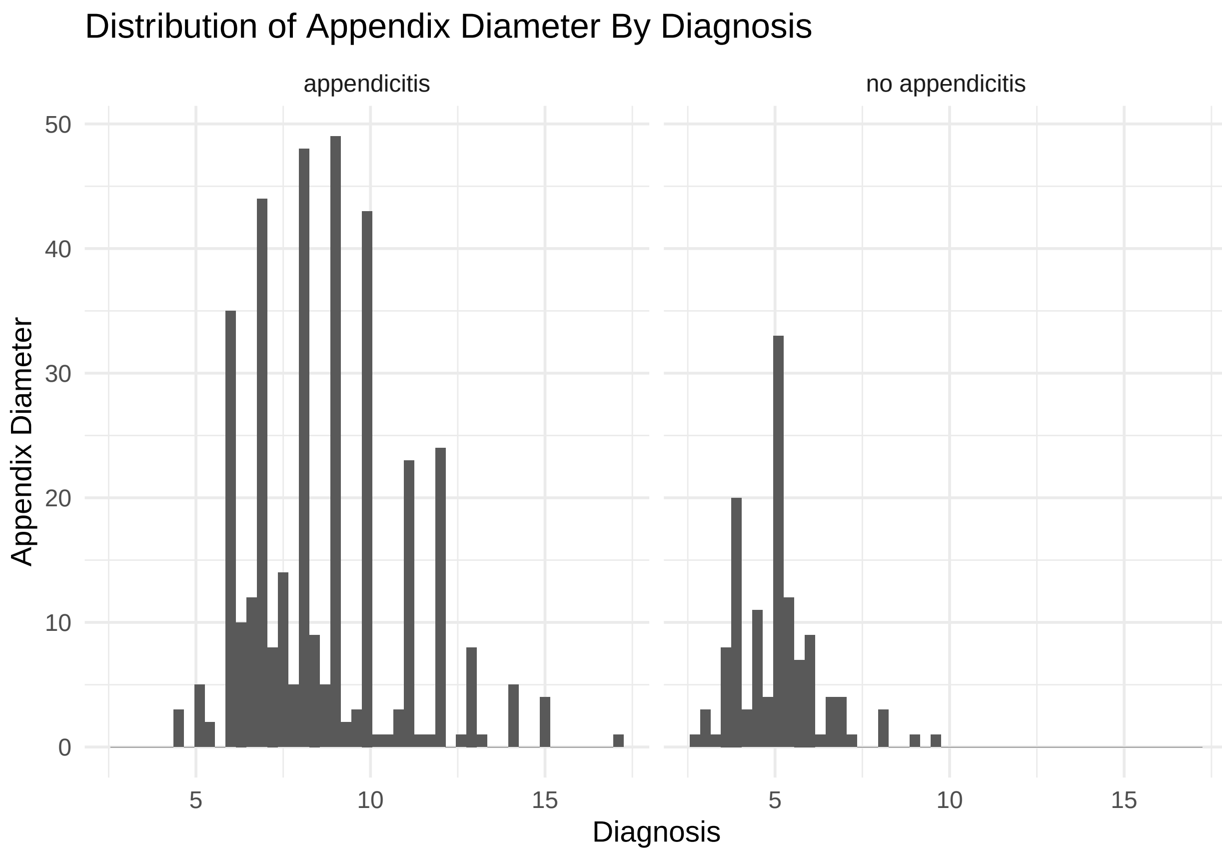
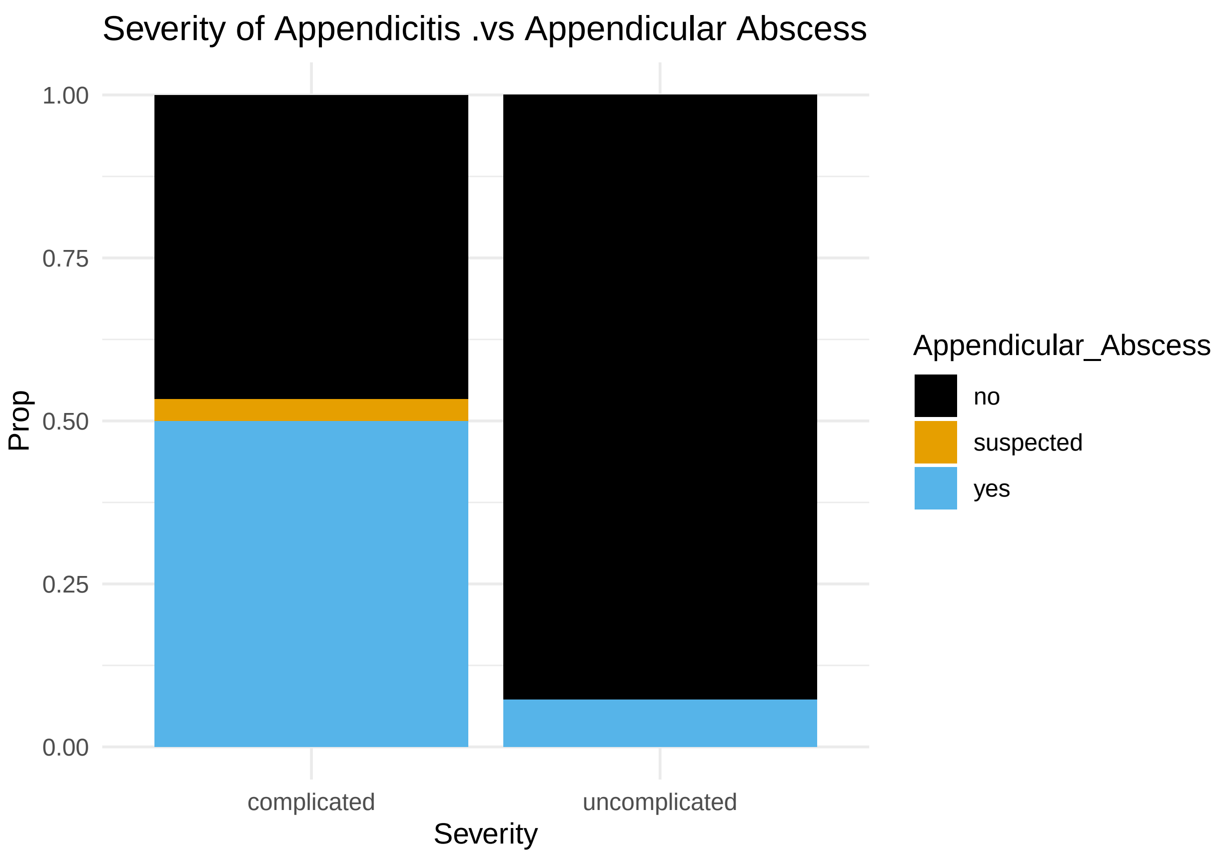


Figure 5: Appendix Diameter .vs Appendicitis Type of Management

Table 5: Mean of Appendix Diameter By Appendicitis Management

Diagnosis	mean	median
appendicitis	8.70	8.2
no appendicitis	5.04	5.0

- 

- In severe cases, abscess can be seen and this Figure shows that the proportion is higher for complicated cases.

Statistical Analysis

Hypothesis testing

- Since the use of ultrasound is less expensive and less harmful than CT, we will use ultrasound data here to see the impact on the diagnostic process.

Here we will test if the addition of **appendiceal diameter** will have a discernible difference on the outcome of the diagnosis, the method we will use is hypothesis testing with **randomization**, and set the discernability level to be 0.05 (i.e the level of rejection).

The two populations of interest in this study are pediatric patients who do or do not have appendicitis.

Let p *= the true mean of appendix diameter in pediatric patients.

- So our hypotheses are

$H_0 : p_{Appendicitis} = p_{no\ Appendicitis}$

$H_A : p_{Appendicitis} \neq p_{no\ Appendicitis}$

- With a p-value of 0, which is smaller than the discernability level of 0.05, we reject the null hypothesis. The data provide convincing evidence that there is a difference between the mean appendix diameters of pediatric patients with and without appendicitis.

Table 6: 95% confidence interval for the difference in mean of appendiceal diameter between patients diagnosed with appendicitis or no appendicitis.

lower_ci	upper_ci
3.37	3.94

- *We are 95% confident that the mean diameter of the appendix in pediatric patients with "appendicitis" is 3.37 to 3.94 greater than in pediatric patients without appendicitis..*

Modeling

- To help the health workers make more informed decisions (i.e, Accurately diagnosing the appendicitis) we would use a supervised machine learning.

We will build Supervised explainable models using logistic regression then test and validate the model.

For evaluation metrics will use **Cross Validation** as way to build the model then we would use **ROC** to check the models precision and accuracy.

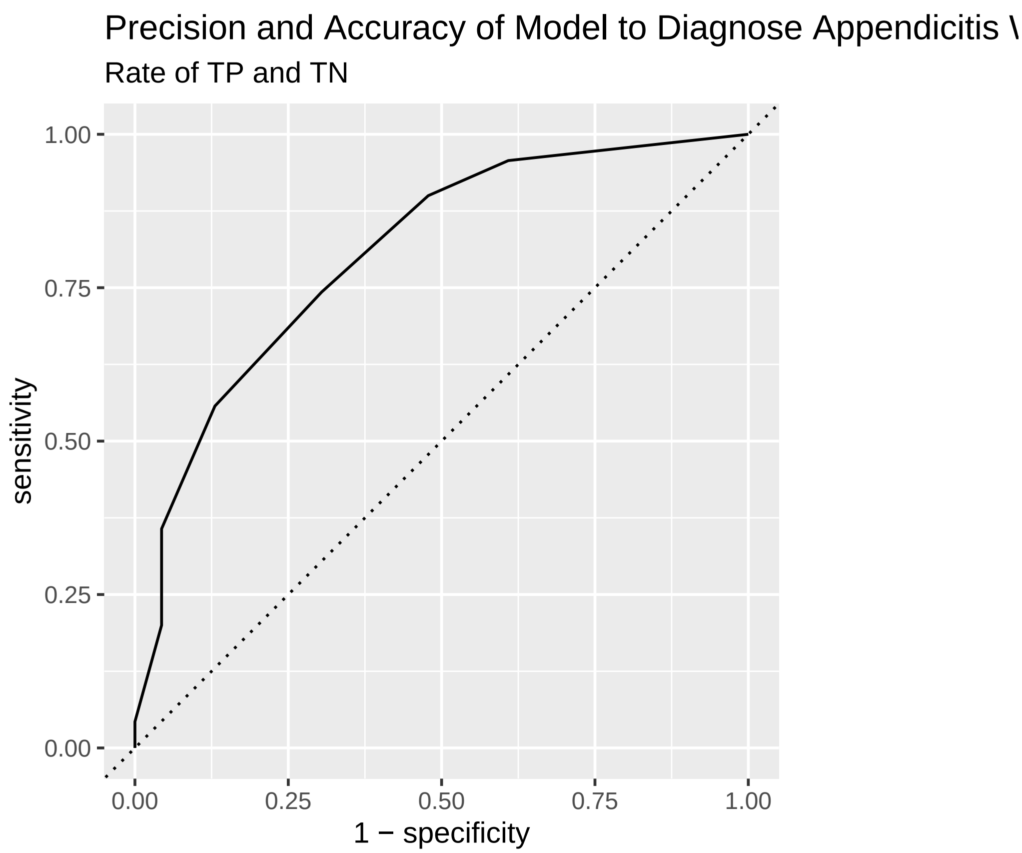
- Model 1: Logistic Regression with Alvarado Score

Table 7: A Model to Diagnose Appendicitis With Avarado Score

pred_class	pred_appendicitis	pred_no_appendicitis	alvarado_score
appendicitis	0.78	0.22	6
appendicitis	0.95	0.05	9
appendicitis	0.68	0.32	5
appendicitis	0.86	0.14	7
appendicitis	0.56	0.44	4
appendicitis	0.56	0.44	4
appendicitis	0.78	0.22	6
no appendicitis	0.30	0.70	2
appendicitis	0.86	0.14	7
appendicitis	0.86	0.14	7

Table 8: Precision and Accuracy of Model to Diagnose Appendicitis With Avarado Score

.pred_class	Diagnosis	n	p	decision
appendicitis	appendicitis	67	0.96	True positive
no appendicitis	appendicitis	3	0.04	False negative
appendicitis	no appendicitis	14	0.61	False positive
no appendicitis	no appendicitis	9	0.39	True negative



- Model 2: Multivariate Logistic Regression

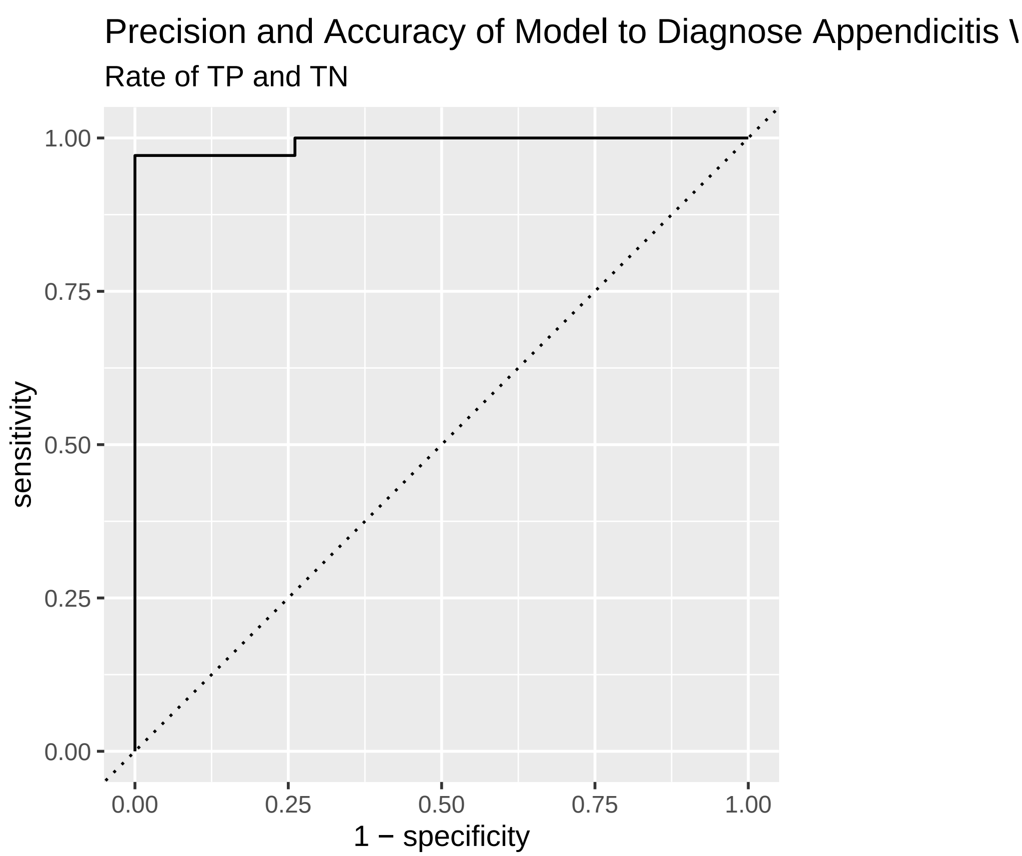
Table 9: A Model to Diagnose Appendicitis With Avarado Score, Appendix_Diameter, Weight and BMI

pred_class	pred_appendicitis	pred_no_appendicitis
no appendicitis	0.01	0.99
no appendicitis	0.10	0.90
no appendicitis	0.29	0.71
appendicitis	0.80	0.20
appendicitis	0.96	0.04
no appendicitis	0.46	0.54
appendicitis	1.00	0.00
no appendicitis	0.03	0.97

pred_class	pred_appendicitis	pred_no_appendicitis
appendicitis	0.85	0.15
appendicitis	1.00	0.00

Table 10: Precision and Accuracy of Model to Diagnose Appendicitis With Avarado Score, Appendix_Diameter, Weight and BMI

.pred_class	Diagnosis	n	p	decision
appendicitis	appendicitis	67	0.96	True positive
no appendicitis	appendicitis	3	0.04	False negative
no appendicitis	no appendicitis	23	1.00	True negative



- The second model that used **Ultra-sonography** results showed lower **False positives and negatives**. per table above.

Diagnosing

Predicting a case with these values Alvarado Score = 6, Appendix Diameter = 10, Weight = 30, BMI = 20, using out second model we built.

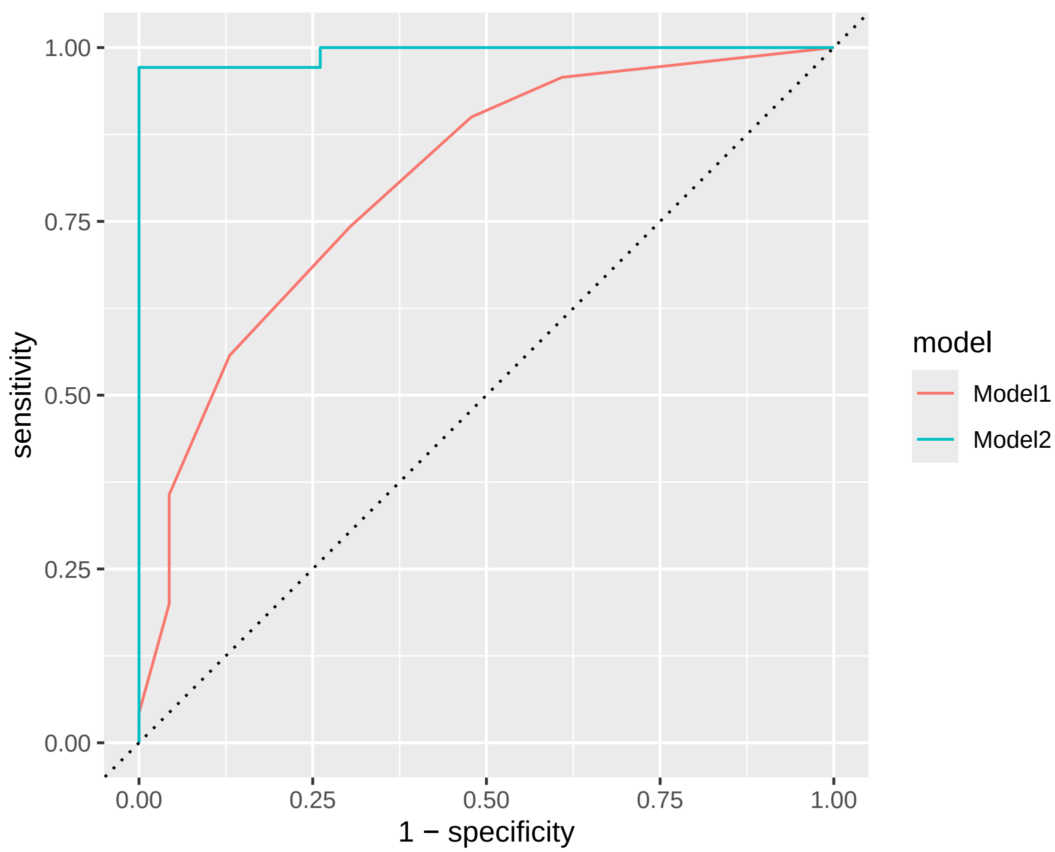


Figure 6: Comparing the Accuracy and Precision of both Models

Table 11: Prediction Probabilities for a Sample Case

.pred_appendicitis	.pred_no appendicitis
0.999642	0.000358

Limitations

While our analysis provides valuable insights, there are limitations that we should note:

1- Generalizability: The data set is derived from a single hospital cohort, which may limit the generalizability of our findings.

2- Model assumptions: Logistic regression assumes linearity between predictors and log odds of the outcome, which we did not test.

3- Potential biases: We do not address the possibility of selection bias and measurement error in clinical assessments.

4- Data quality: Missing values and potential data entry errors could affect model performance. Future work could explore validation of the results with external datasets.

Conclusion

This analysis demonstrates an approach to diagnosing pediatric appendicitis by combining statistical analysis with machine learning. Our findings explore the potential of combining clinical scores with imaging results to improve diagnostic accuracy. We welcome any constructive feedback and collaboration to further refine this analysis. **If you have any suggestions or would like to collaborate, please contact us via the discussion forum or email**.