Adjudication of Adverse Events Using AI

Saman Parvaneh, PhD
Director, Data Science and AI, Edwards Lifesciences

November 15, 2023

Present

Edwards Lifesciences 2020-Present

Structured and unstructured data

2020

2015

Philips Research 2015-2020

Electronic Health Record and Biosignals

Postdoc 2012-2014

Wearable Sensors (Frailty and Outcome Prediction)

Ph.D. 2005-2011

ECG (Atrial Fibrillation Prediction)

2010

M.Sc. 2003-2005

Medical Imaging (MRI) and Speech (Infant Cry)

2005

B.Sc. 1998-2003

Bioinstrumentation

2000

Disclaimer

The views expressed in this presentation are my own and do not reflect the views of my employer. I am presenting my personal opinions and analysis, and any errors or omissions are solely my responsibility.

Outline

Endpoint Adjudication and Clinical Event Committee (CEC)

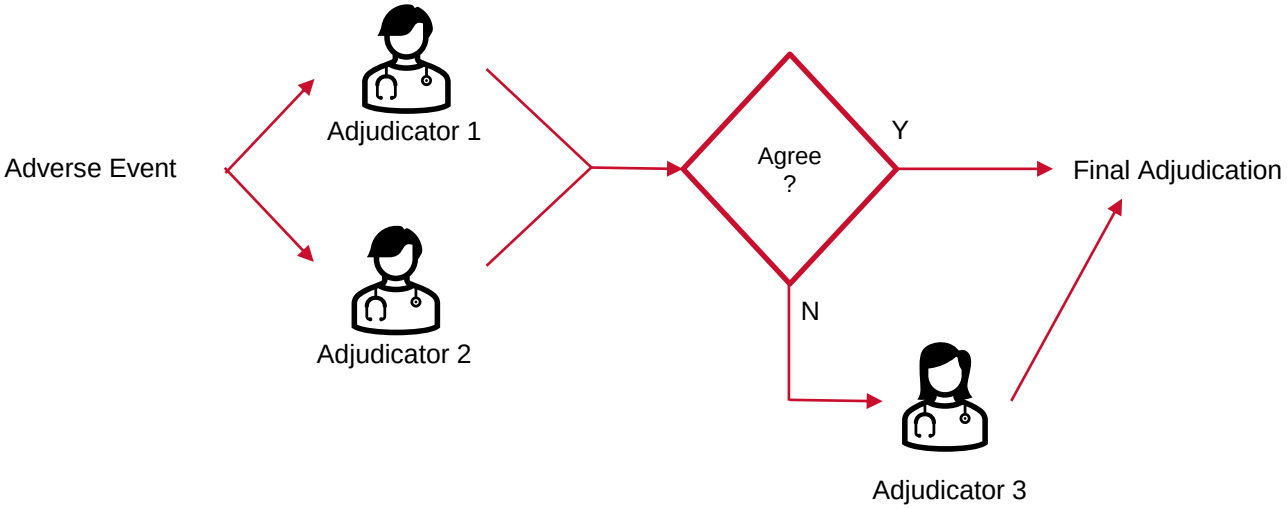
Artificial Intelligence for endpoint adjudication (eCEC)

Results and lesson learned

Endpoint Adjudication & CEC

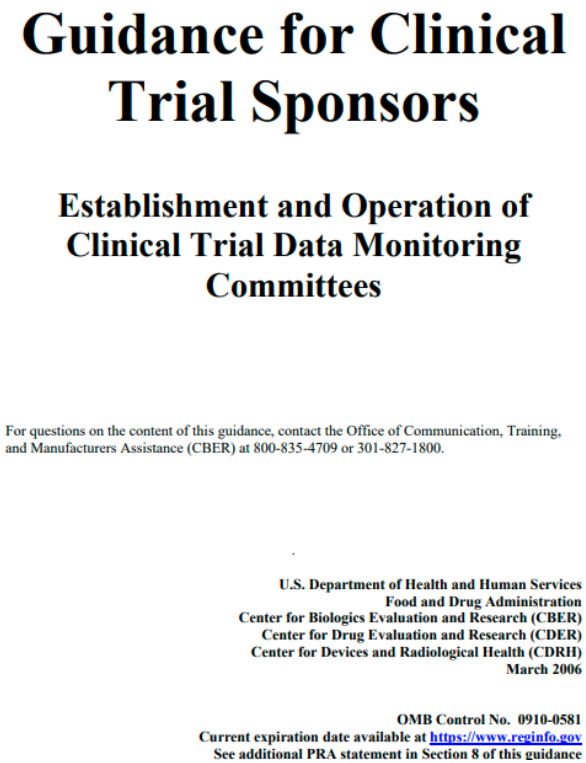
- Endpoint adjudication in clinical trials is a process to ensure that consistent criteria were applied throughout the trial in the assessment of study endpoints.
- Often, clinical outcomes are adjudicated by Clinical Event Committees (CEC) to ensure data quality and patient safety.
- Clinical Event Committee (CEC) is a group of medical experts that review pre-determined adverse events throughout a trial.
- The CEC provides an independent, unbiased, and objective adjudication of the protocol-specified events for a trial.

Adjudication by CEC



Site Reported AEs + Source Documents + EW Narratives → CEC Adjudication → BioStats Analysis

CEC Adjudication and FDA



Guidance for Clinical Trial Sponsors

Establishment and Operation of Clinical Trial Data Monitoring Committees

For questions on the content of this guidance, contact the Office of Communication, Training, and Manufacturers Assistance (CBER) at 800-835-4709 or 301-827-1800.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Biologics Evaluation and Research (CBER)
Center for Drug Evaluation and Research (CDER)
Center for Devices and Radiological Health (CDRH)
March 2006

OMB Control No. 0910-0581
Current expiration date available at https://www.reginfo.gov
See additional PRA statement in Section 8 of this guidance

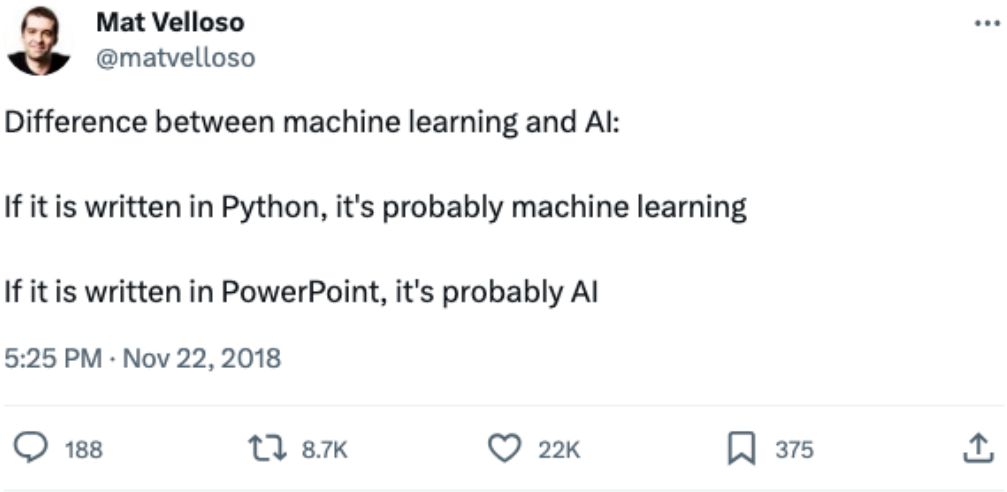
Sponsors may also choose to establish an endpoint assessment/adjudication committee (these may also be known as clinical events committees) in certain trials to review important endpoints reported by trial investigators to determine whether the endpoints meet protocol-specified criteria [1].

[1] Guidance for Clinical Trial Sponsors
Establishment and Operation of Clinical Trial Data Monitoring Committees - https://www.fda.gov/media/75398/download

Automation of Endpoint Adjudication

- Automation of endpoint adjudication represents an opportunity to enhance adjudication consistency, improve turnaround time, and reduce costs when CEC adjudication is no longer performed.
- The availability of clinical data adjudicated by CEC allows the use of Artificial Intelligence (AI) for endpoint adjudication.

Artificial Intelligence and Machine Learning



Mat Velloso
@matvelloso

Difference between machine learning and AI:

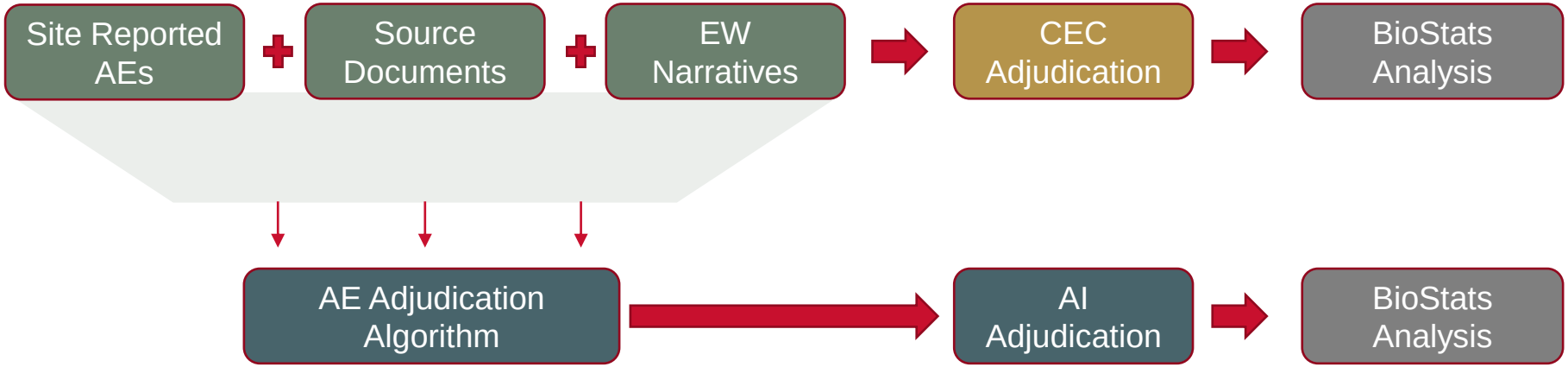
If it is written in Python, it's probably machine learning

If it is written in PowerPoint, it's probably AI

5:25 PM · Nov 22, 2018

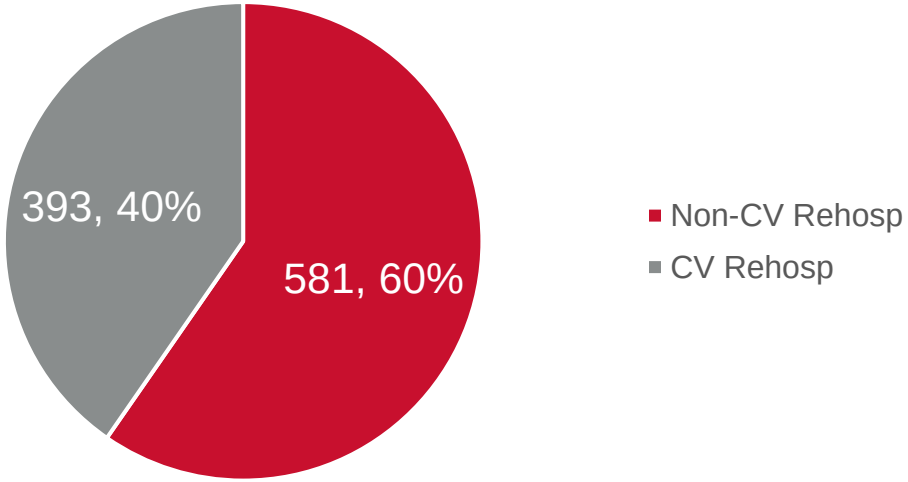
188 8.7K 22K 375

AI-based Adverse Events Adjudication



CEC vs. AI-based Adjudication (EW: Edwards Lifesciences, AE: Adverse Event)

Rehospitalization Adjudication using AI



Variable Group	Number of Variables	Examples
Numeric	1	Number of days from procedure to event
Categorical	2	MedDRA System Organ Classes (SOC) and MedDRA Preferred Term (PT)

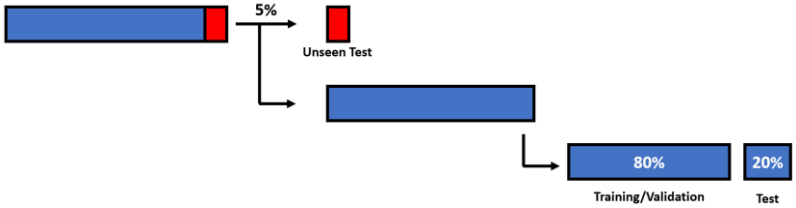
MedDRA Coding

- MedDRA stands for Medical Dictionary for Regulatory Activities.
- It is a standardized medical terminology to facilitate the sharing of regulatory information internationally for medical products used by humans.
- MedDRA is used for registration, documentation and safety monitoring of medical products before and after being authorized for sale.

L1	System Organ Class (SOC) (27)
L2	High Level Group Term (HLGT) (337)
L3	High Level Term (HLT) (1,737)
L4	Preferred Term (PT) (26,180)
L5	Lowest Level Term (LLT) (87,590)

SOC	Cardiac Disorders
HLGT	Cardiac Valve Disorders
HLT	Cardiac valve disorders NEC
PT	Cardiac valve disease
LLT	Valvular heart disease NOS

Rehospitalization Adjudication using AI

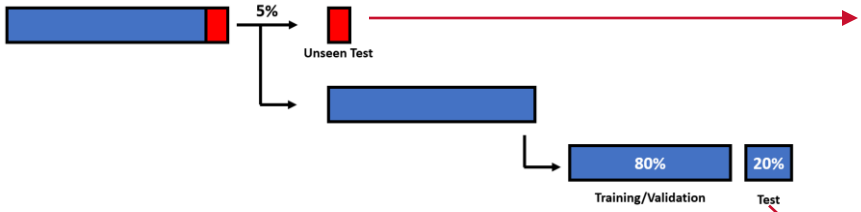


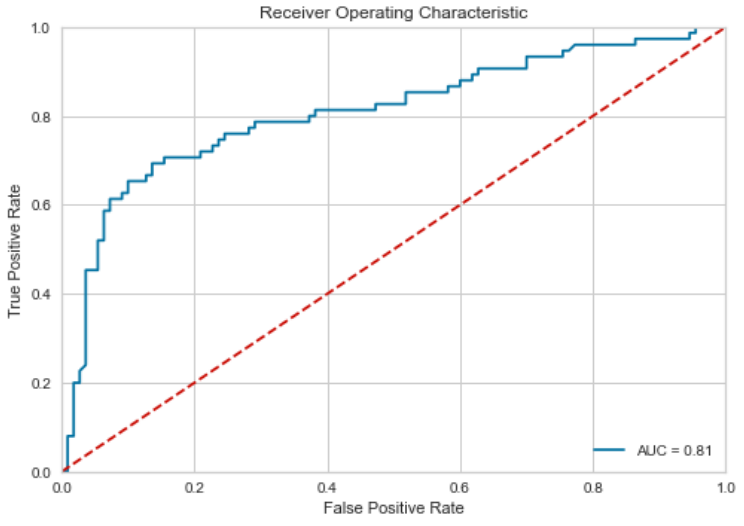
	Non-CV Rehospitalization	CV Rehospitalization
Training/validation	442 (59.73%)	298 (40.27%)
Test	110 (59.46)	75 (40.54)
Unseen test	29 (59.18)	20 (40.82)

Rehospitalization Adjudication using AI

- One-hot encoding was used for categorical variables.
- Seventeen different AI algorithms were tested.
- Data is imbalanced. Therefore, AI model with the highest F1 score (Catboost Classifier) was selected for further improvement.
- Catboost is an advanced boosting algorithm.
- To find feature importance and provide interpretability for AI-based adjudication, SHAP (SHapley Additive explanation) method is employed.

Rehospitalization Adjudication using AI





Non-Cardiovascular

Cardiovascular

True Class	Non-Cardiovascular	Cardiovascular
Non-Cardiovascular	26	3 False Positive
Cardiovascular	3 False Negative	17

Predicted Class

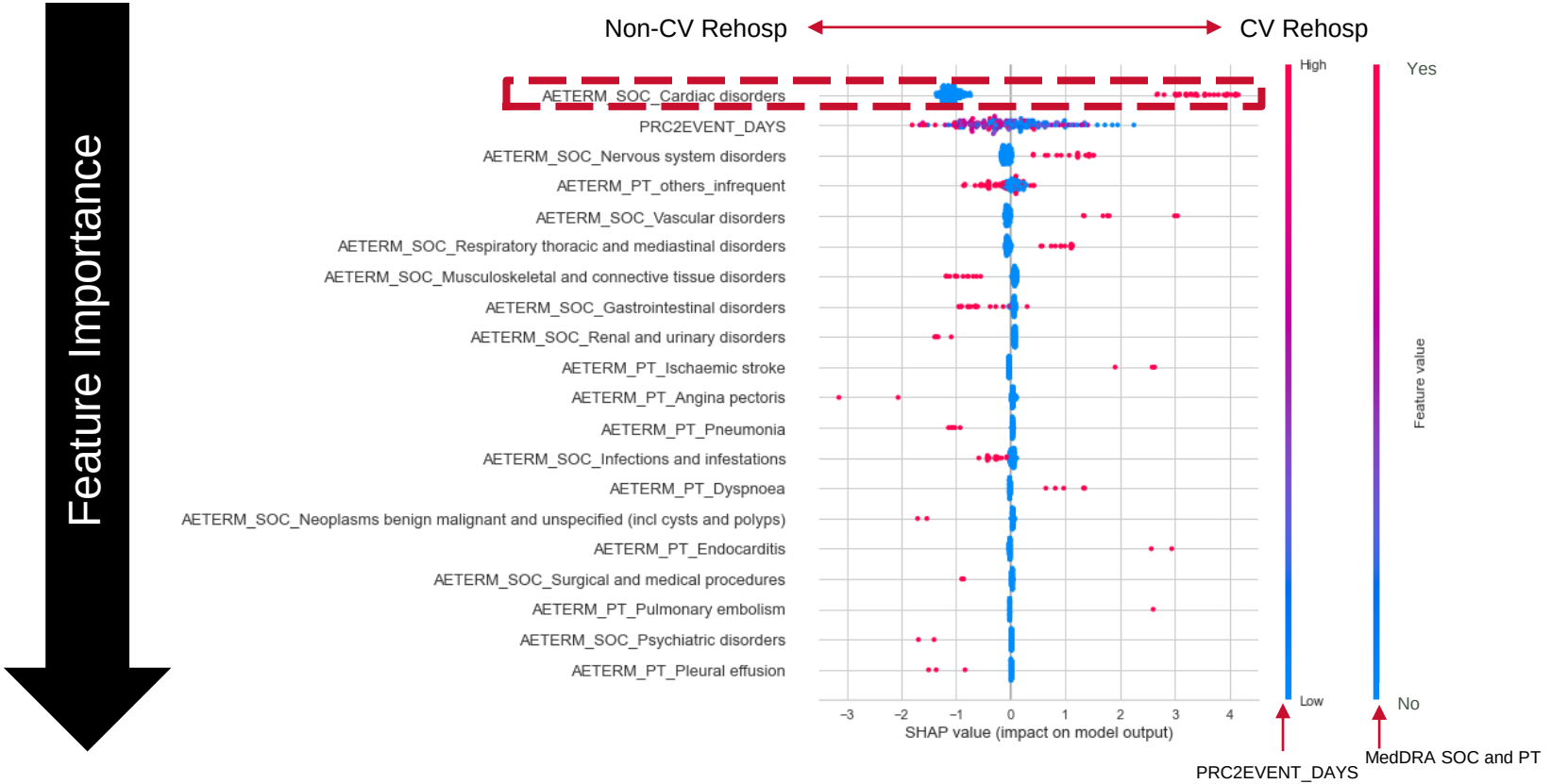
Non-Cardiovascular

Cardiovascular

True Class	Non-Cardiovascular	Cardiovascular
Non-Cardiovascular	95	15 False Positive
Cardiovascular	25 False Negative	50

Predicted Class

Interpreting Model



Example Case (1)

CV/Non-CV Rehospitalization

Site: CV Rehosp
CEC: CV Rehosp
eCEC: CV Rehosp
(Confidence Score: 98.3%)

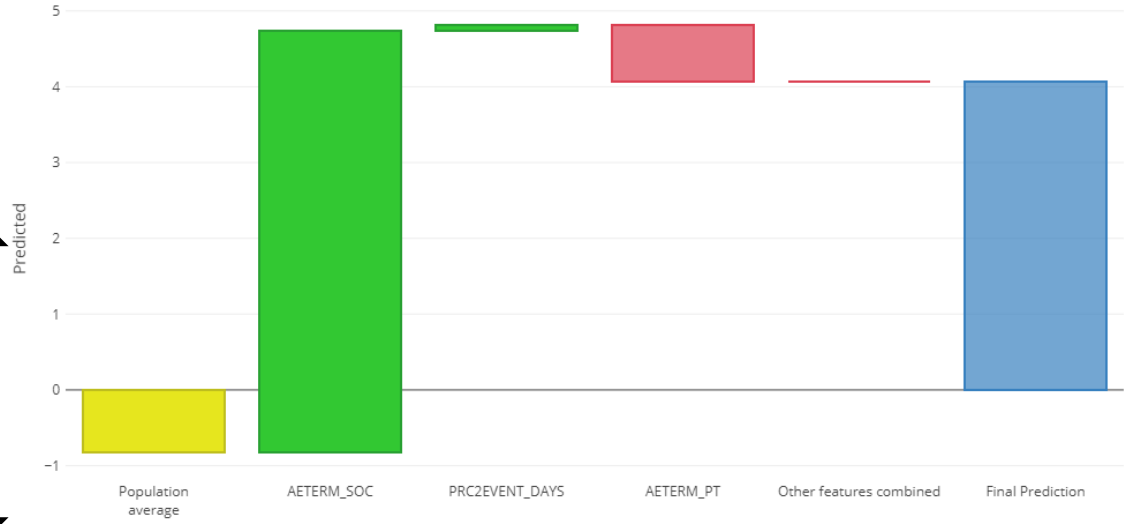
AETERM_SOC = Cardiac disorders

PRC2EVENT_DAYS = 993.0

AETERM_PT = others_infrequent

CV

Non-CV



Example Case (2)

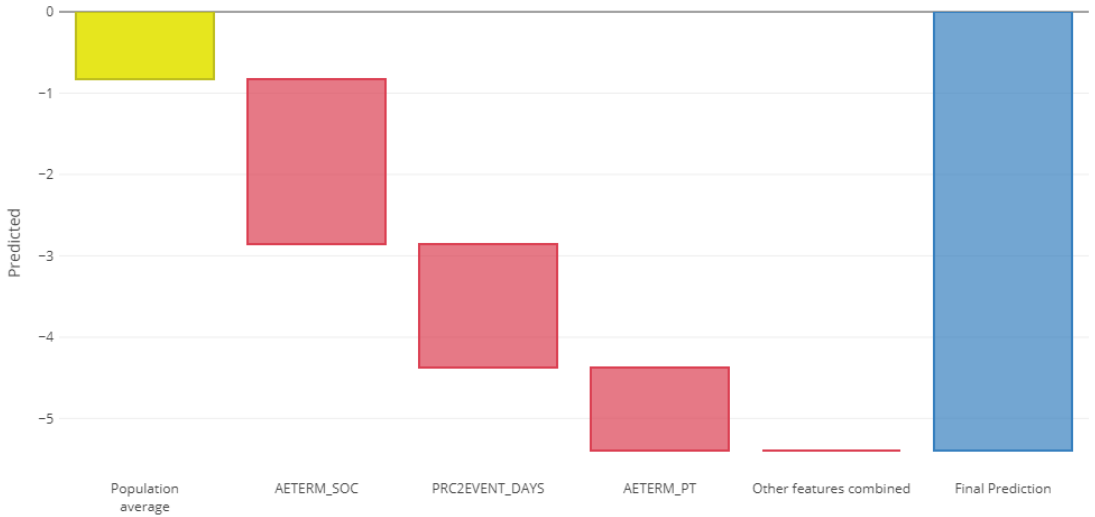
CV/Non-CV Rehospitalization

Site: Non-CV Rehosp
CEC: Non-CV Rehosp
eCEC: Non-CV Rehosp
(Confidence Score: 99.5%)

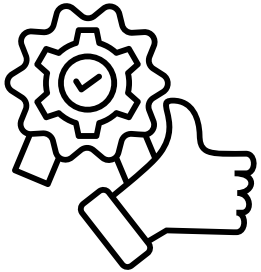
AETERM_SOC = Infections and infestations

PRC2EVENT_DAYS = 1133.0

AETERM_PT = Cellulitis



Benefits of Using Interpretable Approach



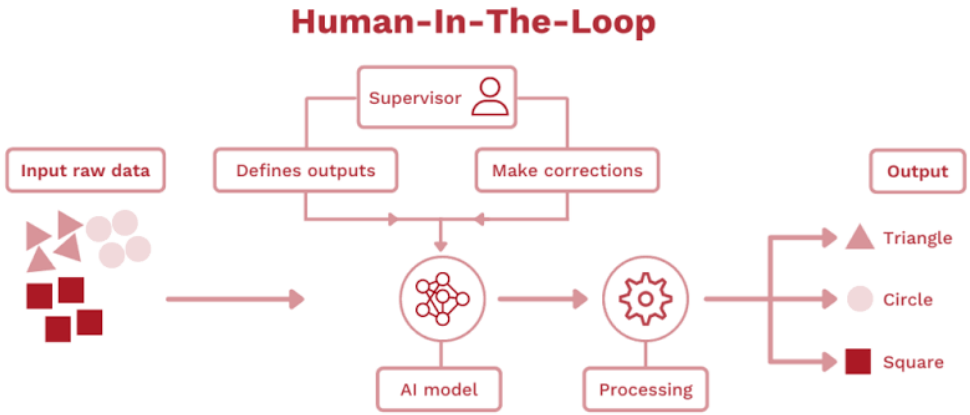
Building Trust

Better Adoption

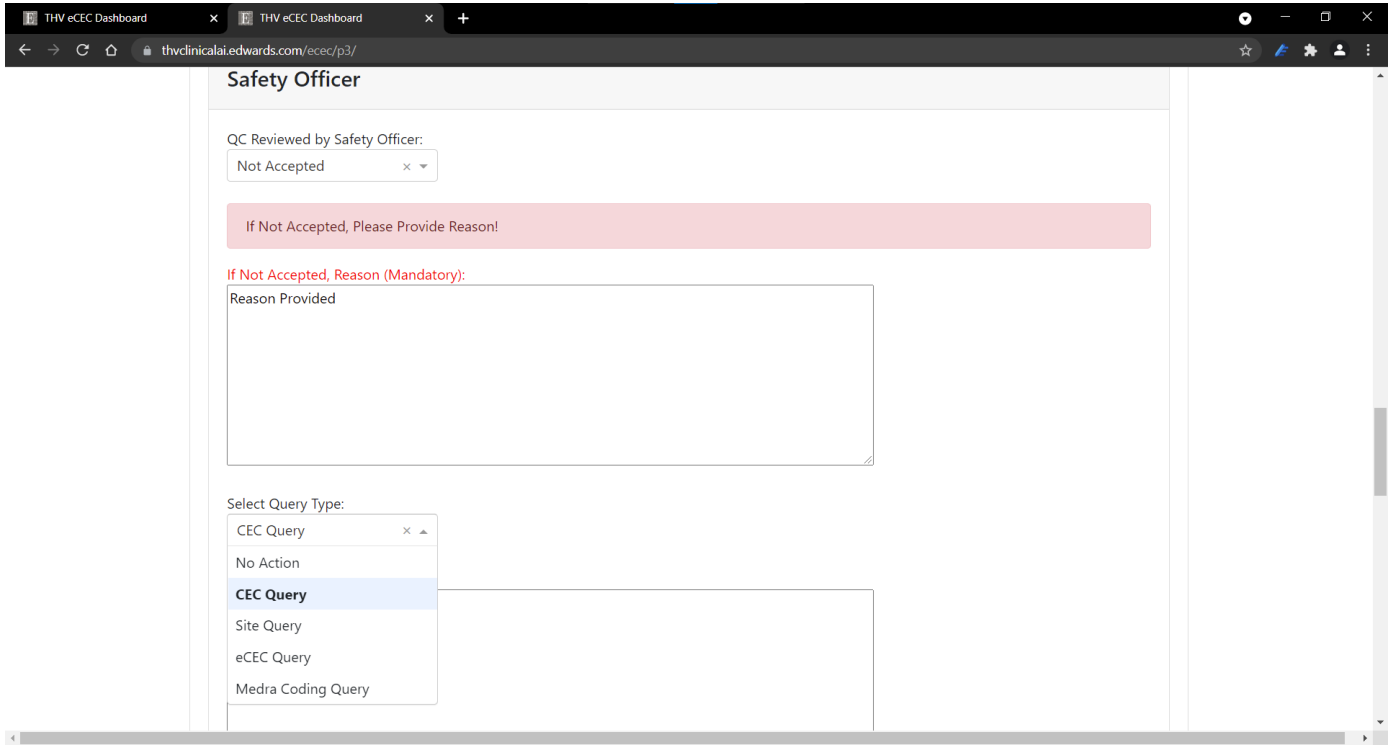
Feedback for Improvement

Human-In-The-Loop

- It is a design paradigm that involves integrating humans into the decision-making process of artificial intelligence (AI) systems.
- The term "human-in-the-loop" refers to the idea that **AI systems** should be designed to **work with human operators** rather than replace them entirely.

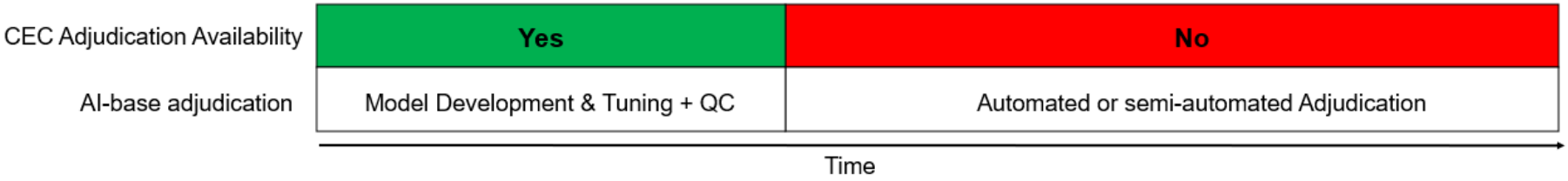


Human-In-The-Loop and AI-based Adjudication

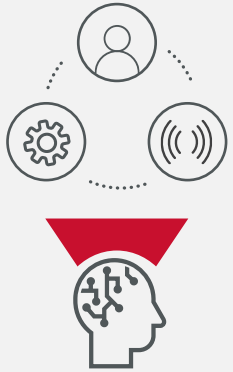


Integration of AI-based adjudication into workflow over time

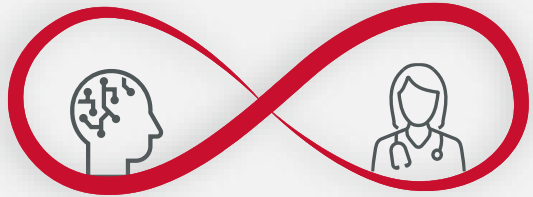
- Out of 91 cases with a mismatch between CEC and AI-based algorithm:
  - 17 cases require re-review by CEC
  - 6 cases required a MedDRA coding update
- Results suggest the potential for using AI-based adjudication for QC when CEC adjudication is required.



Is AI **replacing** CEC?



AI could be **helpful in resource-limited setting**



Human-in-the-loop is the key for a successful AI for healthcare

Key Takeaways

- Working with a multidisciplinary team is the key to success.
- Endpoint adjudication using AI techniques is possible.
- By increasing data size, further improvement of the AI algorithm is possible.
- AI-based adjudication can reduce the cost and increase efficiency to get consistent endpoint adjudication in place of CEC adjudication where CEC adjudication is not required or feasible.
- AI confidence score and/or disagreement between eCEC and other human judgments could be used for prioritizing QC.
- The robustness of AI-based adjudication across different trials needs to be explored, as endpoint definitions may change from trial to trial.
- Collaboration across different companies and FDA provide an opportunity to have more accurate and robust AI-based adjudication

Team

Zaniar Ardalan

Sr Data Scientist
Data Science & AI

Matt Song

Sr Data Engineer
Data Science & AI

Ihsan Hasan

Director
Medical Safety

Elena Khury

Sr Analyst
Clinical Safety Systems

Carlie Gaunty

Director
Medical Safety

Michael Lu

Director
Biostatistics

Wei Liu

VP
Biostatistics

Terri Johnson

Sr Director
Biostatistics

Thank you

saman_parvaneh@edwards.com

Edwards Lifesciences • One Edwards Way, Irvine CA 92614 USA • edwards.com

Edwards

AI is a broad term

Artificial intelligence (AI)

Any technique that enables a machine to mimic human intelligence

Machine learning (ML)

Mathematical and statistical methods that enable machines to learn tasks from data without explicitly programming

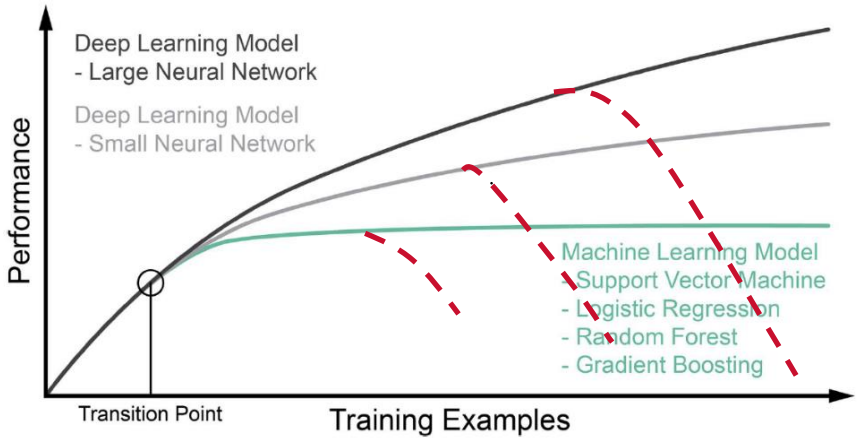
Deep learning

Neural networks with many layers that learn representations and patterns directly from data

Our goal is to create AI that is

- Simple
- Reliable
- **Generalizable**
- Interpretable
- Fair

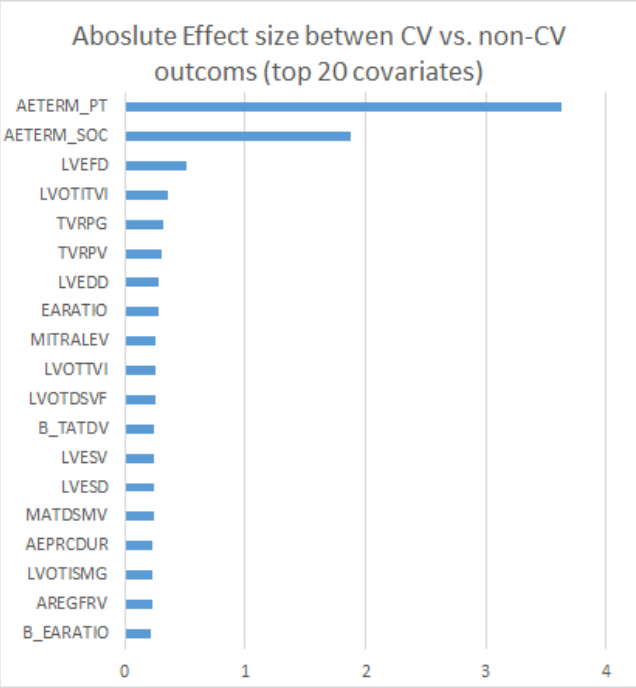
The myth of big data in medicine



“This promise turns out to be a myth. The more and more data you have, the less you are able to curate the quality of that data.” Professor G. Clifford

Statistical Approach for initial data screening

- Univariable Analysis



- Correlation/Collinearity among covariates
  - Some echo parameters have high collinearity.
  - Variables with >0.8 absolute correlation will be eliminated
  - Keep the one variable with higher absolute effect size if highly collinear among multiple variables.

- After removing columns with much missing and high collinearity, we kept 107 variables into AI modeling.

Results Summary

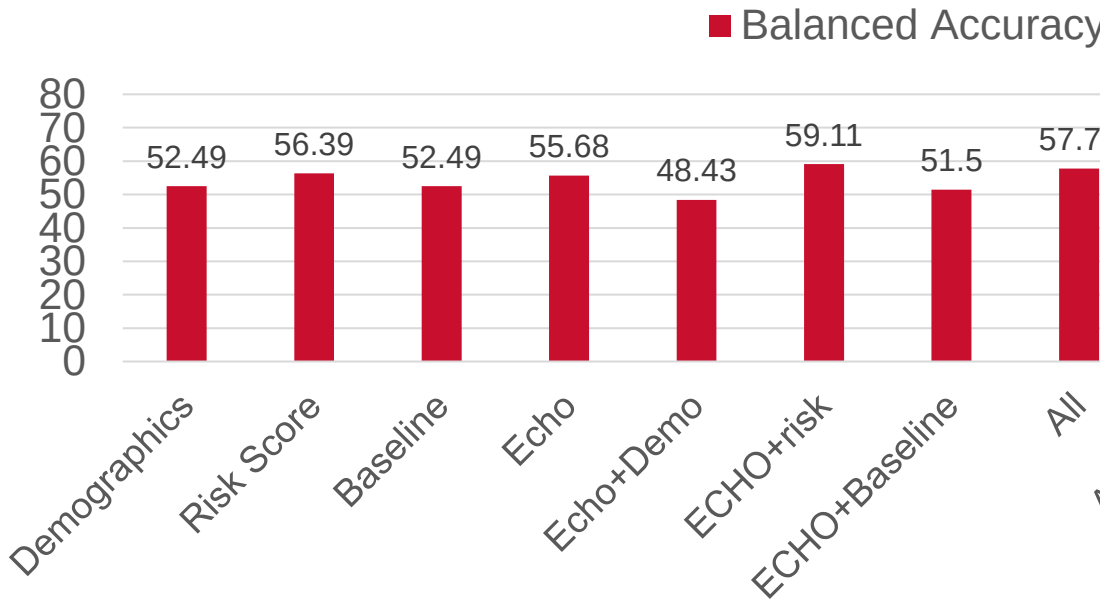
Measure	Test	Unseen Test	Derivations
Sensitivity	0.6667	0.8500	TPR = TP / (TP + FN)
Specificity	0.8636	0.8966	SPC = TN / (FP + TN)
Precision	0.7692	0.8500	PPV = TP / (TP + FP)
Negative Predictive Value	0.7917	0.8966	NPV = TN / (TN + FN)
False Positive Rate	0.1364	0.1034	FPR = FP / (FP + TN)
False Discovery Rate	0.2308	0.1500	FDR = FP / (FP + TP)
False Negative Rate	0.3333	0.1500	FNR = FN / (FN + TP)
Accuracy	0.7838	0.8776	ACC = (TP + TN) / (P + N)
F1 Score	0.7143	0.8500	F1 = 2TP / (2TP + FP + FN)

Rehospitalization Adjudication using AI/ML

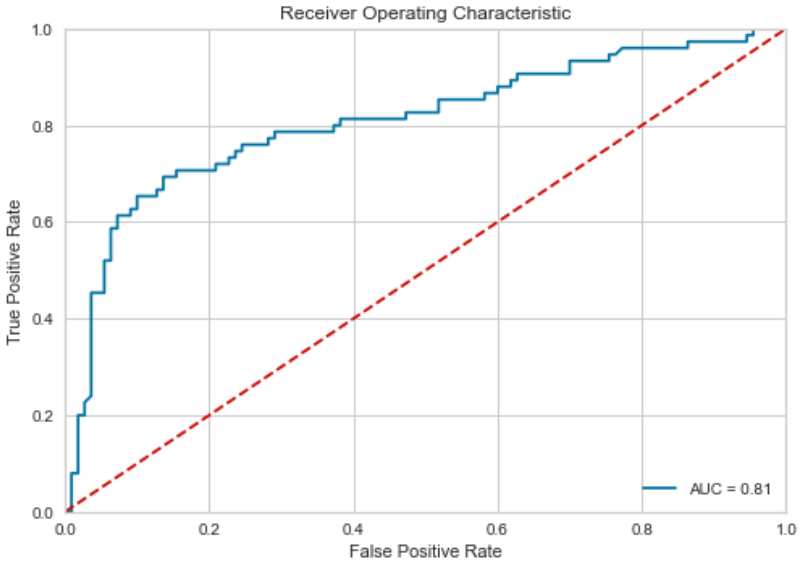
Variable Group	Number of Variables	Examples
Demographics	4	RACE, AGE, SEX, BMI
Risk Scores	8	STSRSCR, FRIDX07, GRASPAVG, ALBUM
Baseline Measures	39	HEARTRATE, RHYTHM, LVEDD, LVEF, AOVMG, NYHA
Echocardiography	53	HEARTRATE, RHYTHM, LVEDD, LVEF, AOVMG, NYHA
Text/Categories	3	AE TERM, AE MEDDRA PREFERED TERM (PT), AE DESCRIPTION

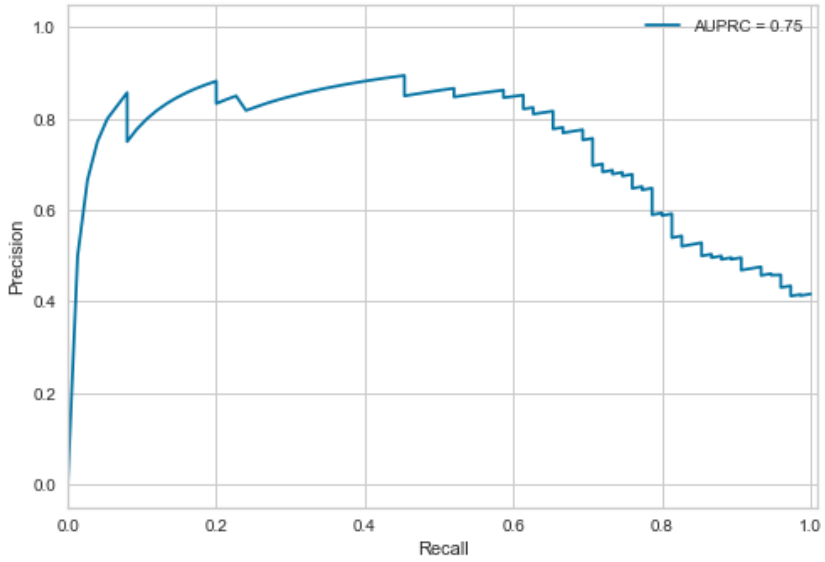
Results

- Removed 5 subjects where time between event and adjudication was negative
- Handled missing values
- Handled correlated variables

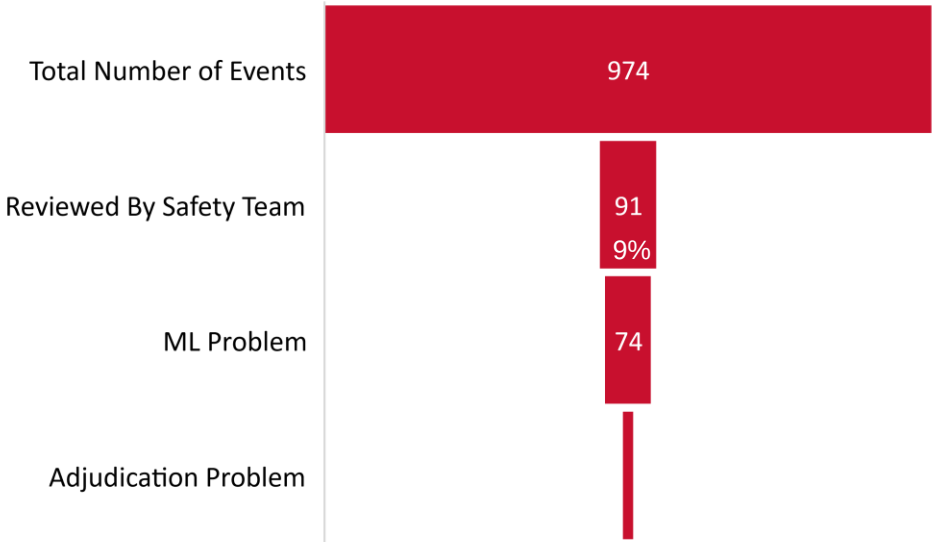


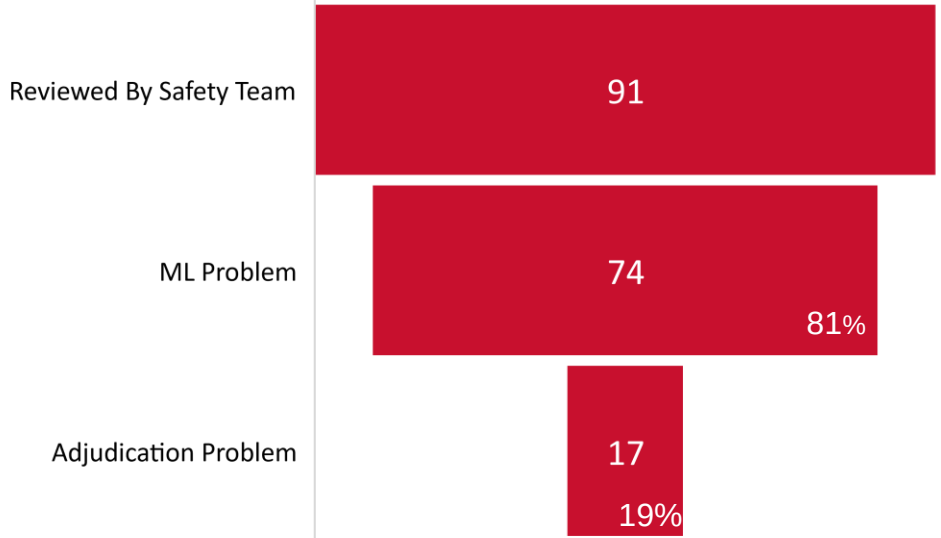
Performance on Test Data



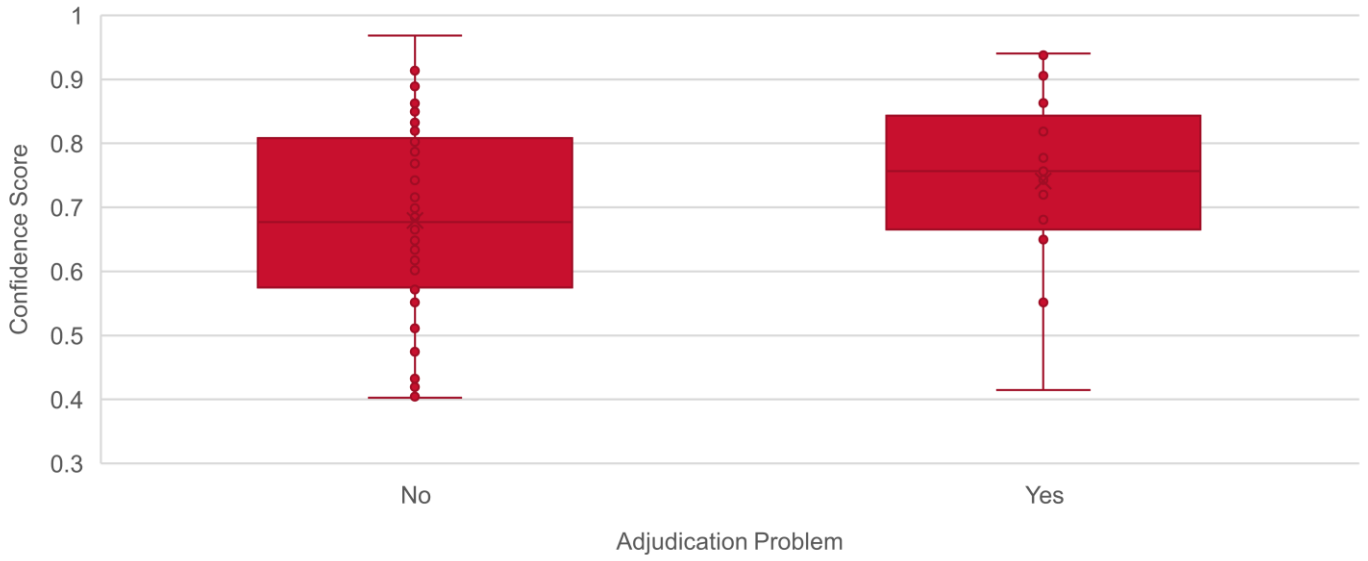


QC





6 out of 74 cases in ML Problem groups had coding problem.


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
Confidence Score
No
Yes
Adjudication Problem