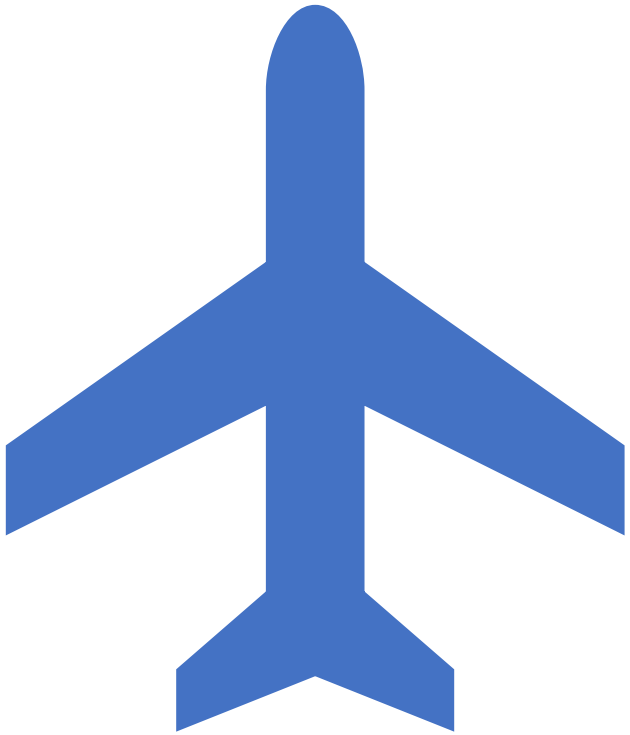


Predicting Aviation Accident Severity

C.J. Argue, Jake Caldwell, Inkee Jung, Jinting Liu

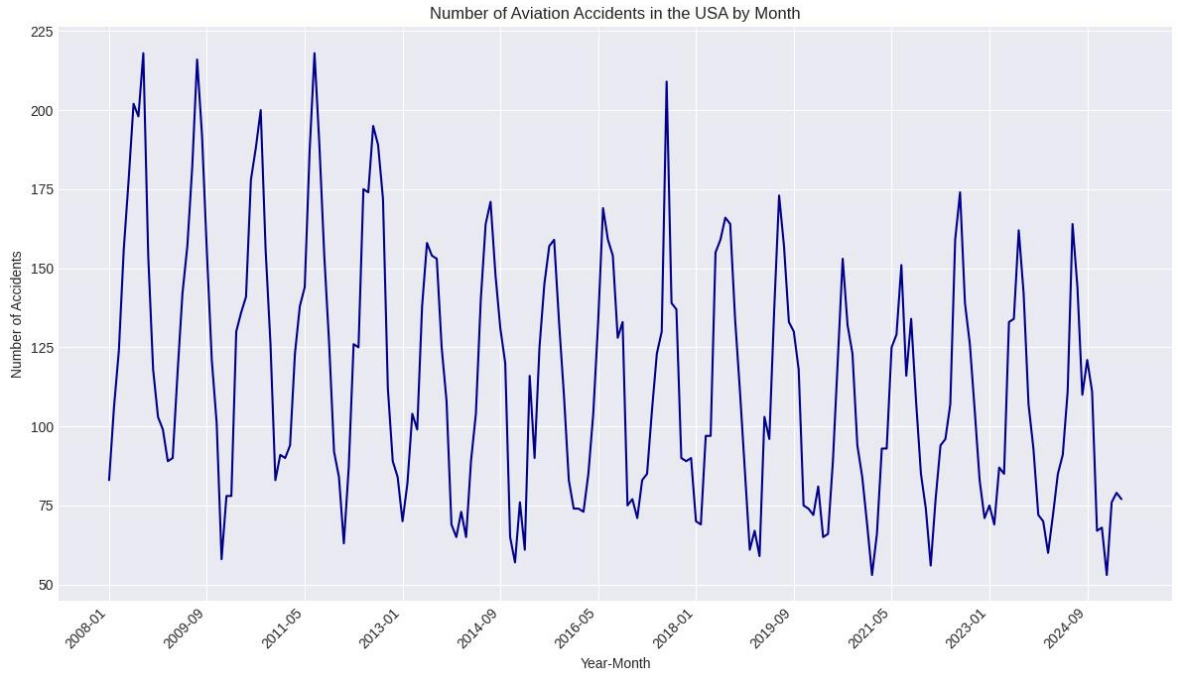
Problem Definition

Motivation

- Increasing air travel demands call for **ongoing improvements** in aviation safety
- Data-driven insights can support **lawmakers** and **airlines** in shaping **preventive policies** and **risk mitigation strategies**

Key Questions

- Can we **predict** whether a given accident will result in serious **aircraft damage** or **casualties**?
- What are the most influential factors that lead to **“severe” accident outcomes?**
- What trends can be observed in **the number of accidents** over time?



National Travel Safety Board (NTSB) investigation database

Dataset of investigations into airplane accidents

Fact gathering includes flight logs, maintenance records

On-flight injuries
Severity
Aircraft damage.

National Travel Safety Board (NTSB) investigation database

- ~30000 aviation investigations since 2008
- ~250 variables
- .mdb file parsed using ‘mdb-tools’

- Extracted data
  - Selected ~50 variables
  - Merged columns split across files

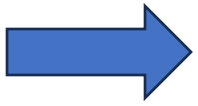
events	
PK	ev_id
	ntsb_no
	ev_type
	ev_date
	ev_dow
	ev_time
	ev_tmzn
	ev_city
	ev_state

aircraft	
PK,FK1	ev_id
PK	Aircraft_Key
	regis_no
	ntsb_no
	acft_missing
	far_part
	flt_plan_filed
	flight_plan_activated
	damage

Merged datasets to get aircraft-level data

Event-level Aircraft-level Sub-aircraft level

Light condition	Make	Fatalities
Light	Cessna	Crew: 0
		Pass: 0
Dark	Boeing	Crew: 1
		Pass: 4
Dark	Cessna	Crew: 2
		Pass: 0
	Piper	Crew: 0
		Pass: 0



Light condition	Make	Fatalities
Light	Cessna	0
Dark	Boeing	5
Dark	Cessna	2
Dark	Piper	0

Propagated to aircraft

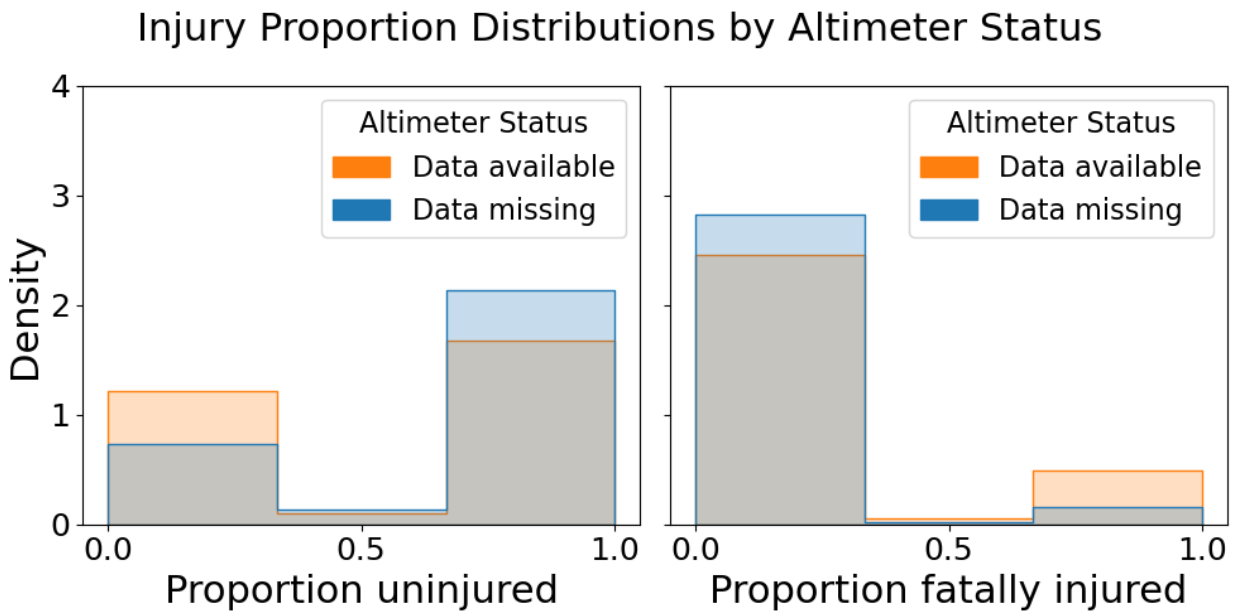
Aggregated by aircraft

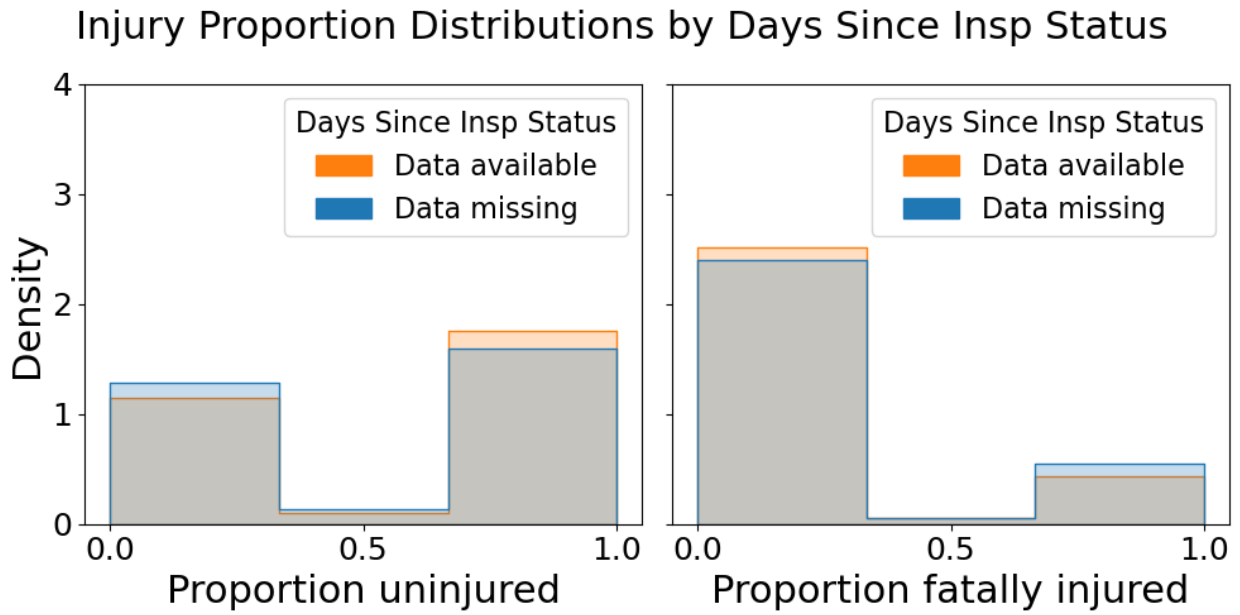
Removed rows, reduced categories, and imputed to get complete data with manageable feature set

ev_type	country	phase	acft_make	num_eng	rwy_len
ACC	USA	Maneuvering	Bell	1	
ACC	USA	Takeoff	Cessna	1	1717
ACC	USA	Initial Climb	Piper	1	
ACC	USA	Takeoff	RANS	1	
ACC	USA	Enroute			
ACC	USA	Landing	Cessna	1	6000
ACC	USA	Landing	Cessna	1	2221
ACC	USA	Takeoff	CESSNA	2	6505
ACC	USA	Enroute	PIPER	1	
ACC	USA	Approach	Cessna	1	3499
ACC	USA	Approach	CESSNA	1	
ACC	USA	Approach	HARMS R.	1	
ACC	USA	Maneuvering	Van Heeswyk		
INC	USA		AEROSPATIAL	2	
INC	USA				
ACC	GE				
ACC	PE				

ev_type	country	phase	acft_make	num_eng
ACC	USA	Enroute	bell	1
ACC	USA	Takeoff	cessna	1
ACC	USA	Takeoff	piper	1
ACC	USA	Takeoff	other/unknown	1
ACC	USA	Enroute	other/unknown	1
ACC	USA	Landing	cessna	1
ACC	USA	Landing	cessna	1
ACC	USA	Takeoff	cessna	2
ACC	USA	Enroute	piper	1
ACC	USA	Landing	cessna	1
ACC	USA	Landing	cessna	1
ACC	USA	Landing	other/unknown	1
ACC	USA	Enroute	other/unknown	2

Analyzed missingness to avoid data leakage





Status strongly predictive of fatality → Drop feature

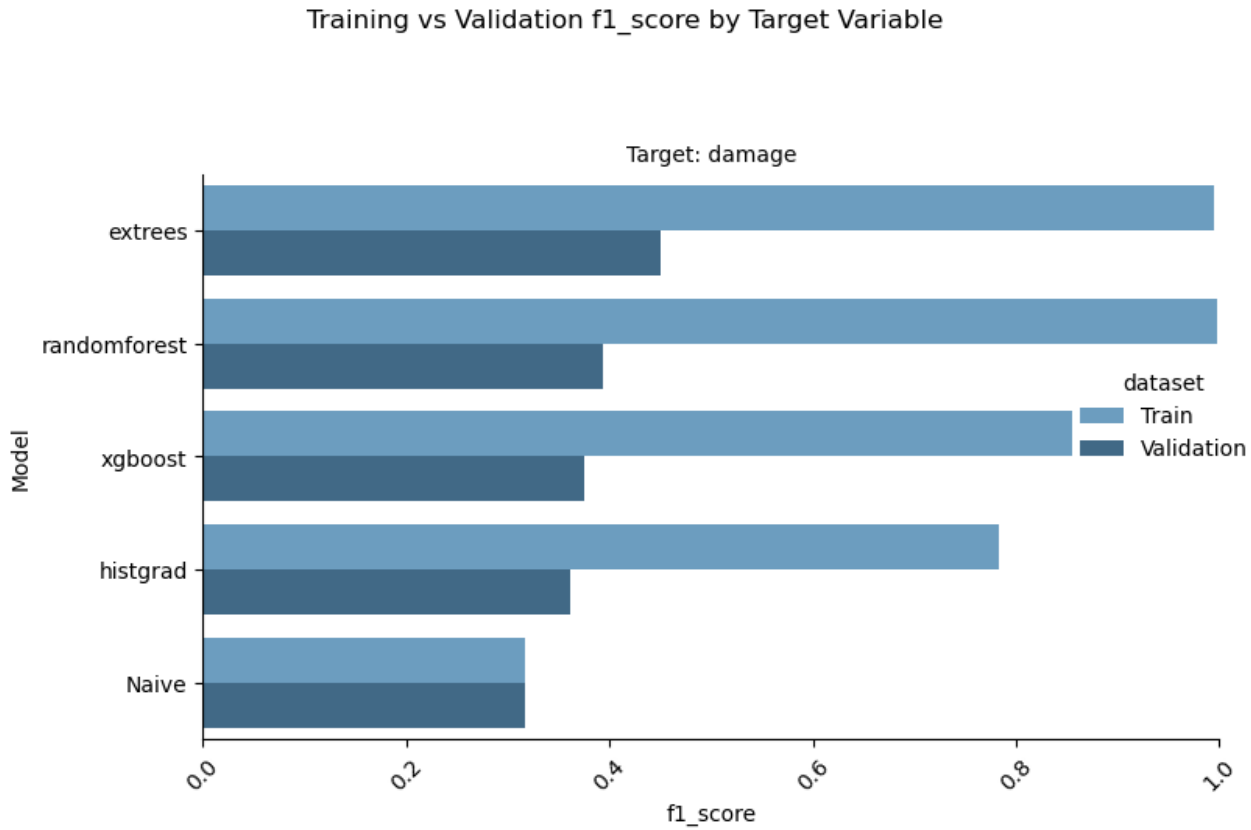
Status weakly predictive → Keep feature

Modeling approach

- Series of Learners
  - Random Forest (Classifier/Regressor)
  - Extra Trees (Classifier/Regressor)
  - Histogram Gradient Boost (Classifier/Regressor)
  - XGBoost (Classifier/Regressor)
  - Bagging (Classifier/Regressor)

- Hyperparameter tuning for each target
  - 5-fold cross-validation

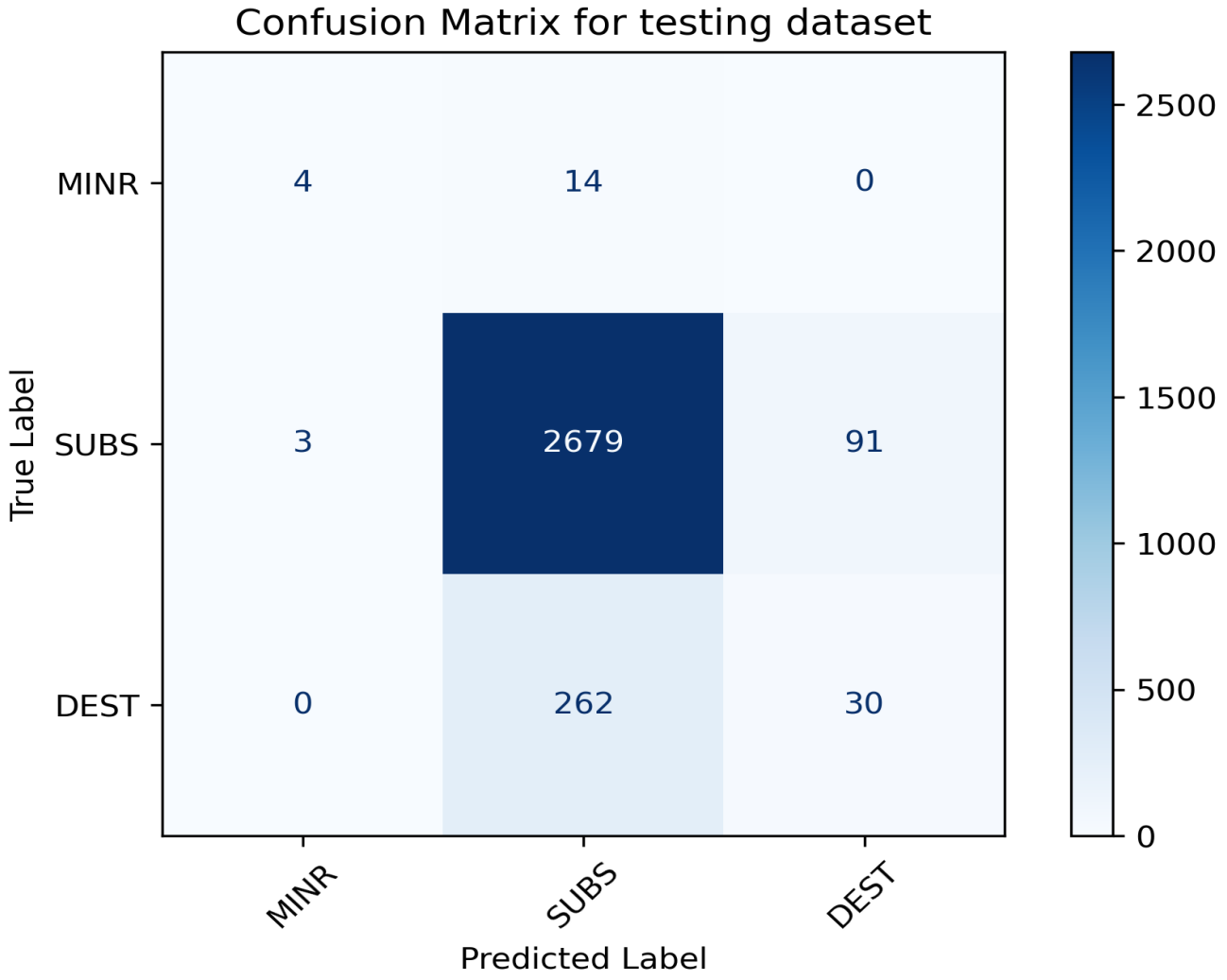
Classification Model Performances: Training and Validation Sets



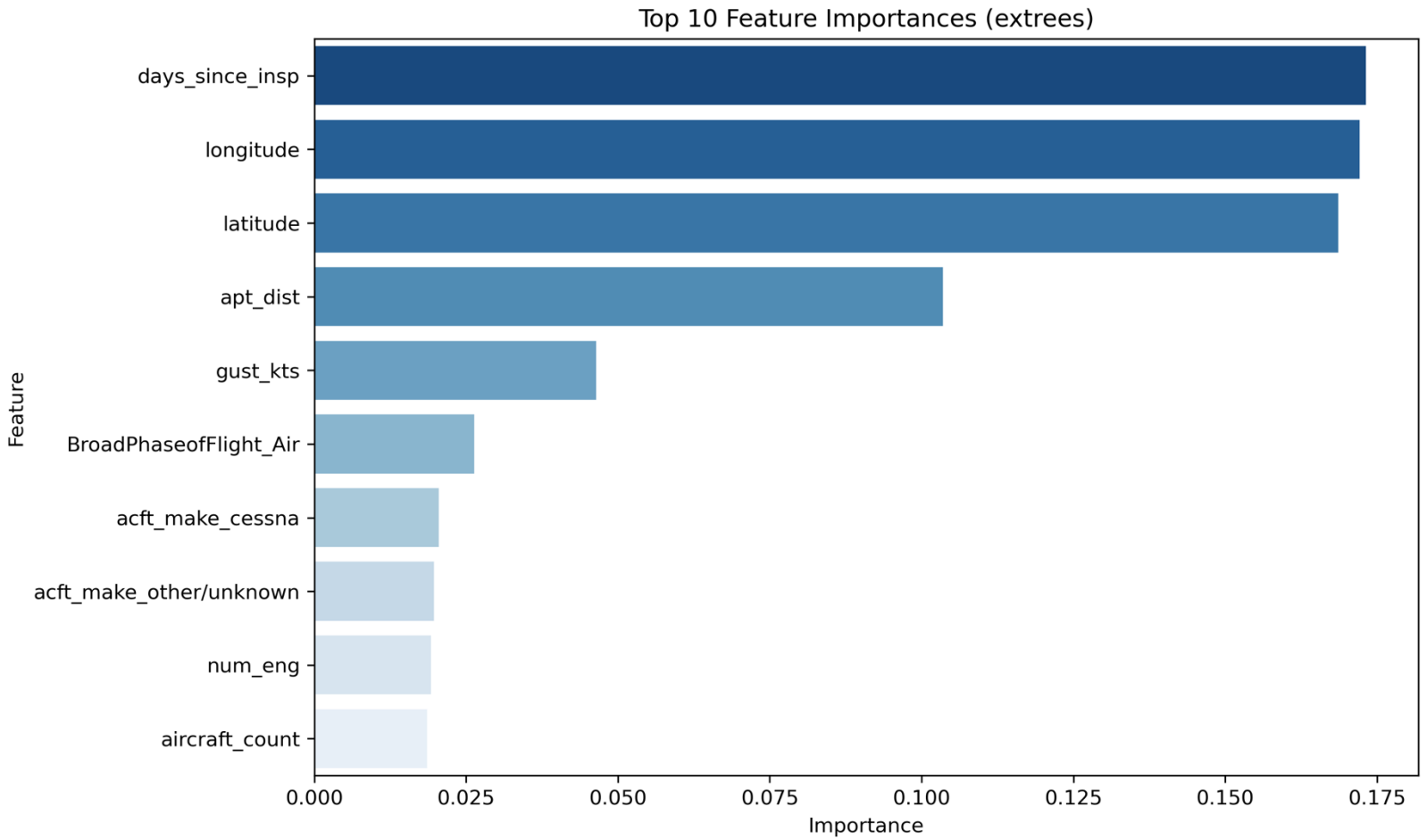
- For each, Extra Trees Classifier has highest validation F1(macro) score.
- Slight edge over “Naïve” estimator
- For all, every model overfits training data

Classification Models: Results

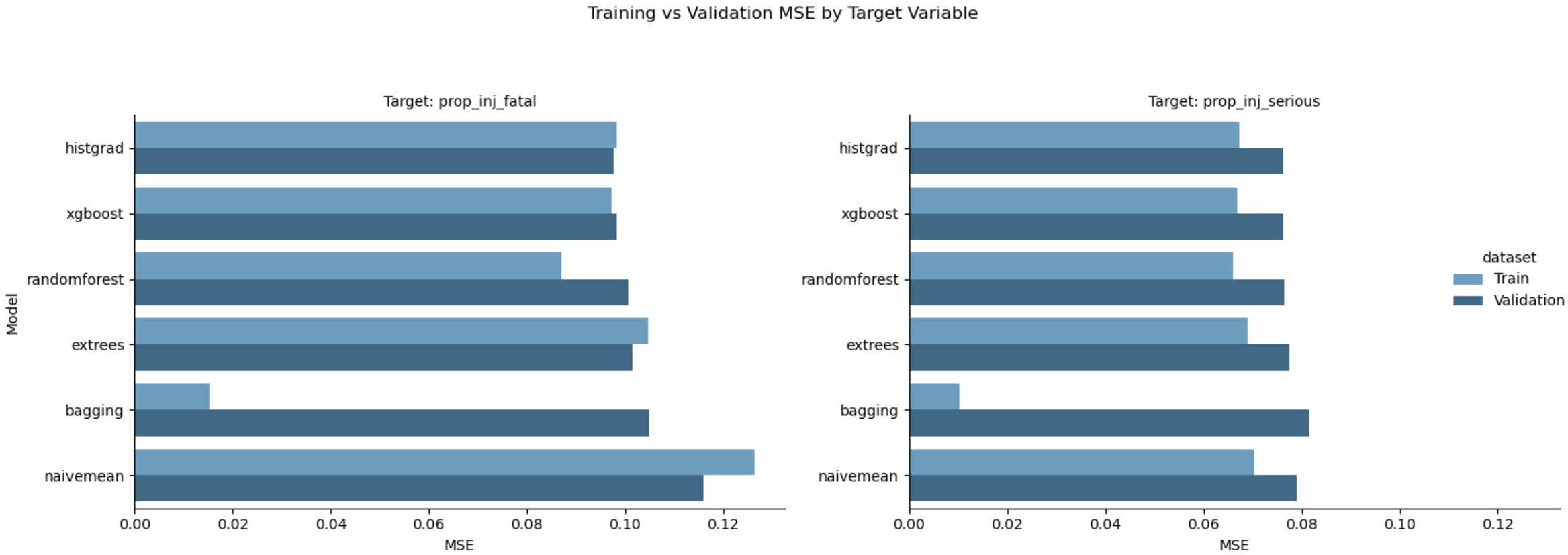
Target: Damage	
Model	**F1 score (Macro average)**
Extra Trees Classifier	0.459
Naïve predictor (Baseline)	0.316



Damage classification:
Extra Trees Feature Importances



Regression Model Performances: Training and Validation Sets



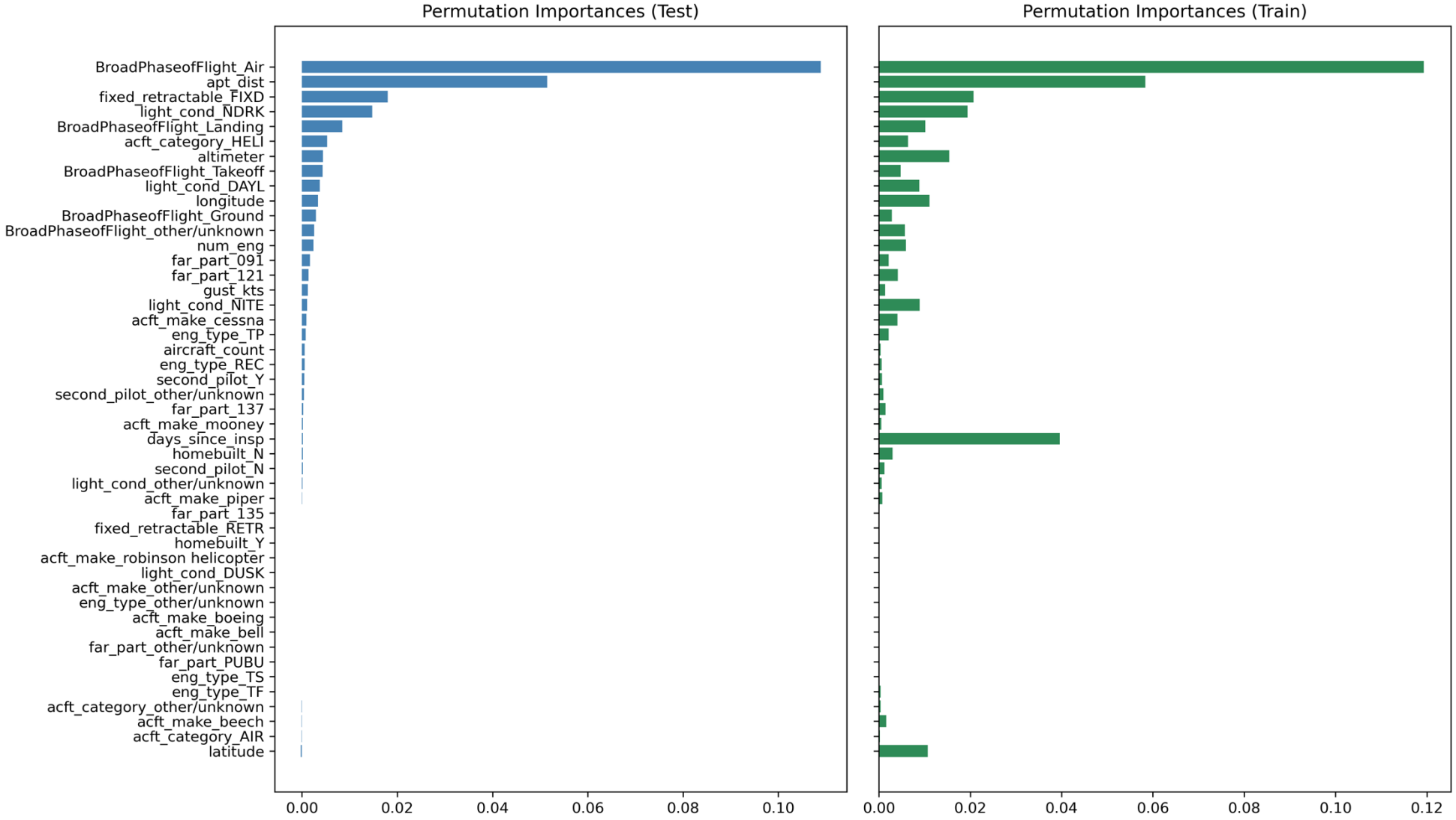
- For each, Histogram Gradient Boosting Regressor has lowest validation MSE.
- Slight edge over “Naïve Mean” estimator

Regression Models: Results

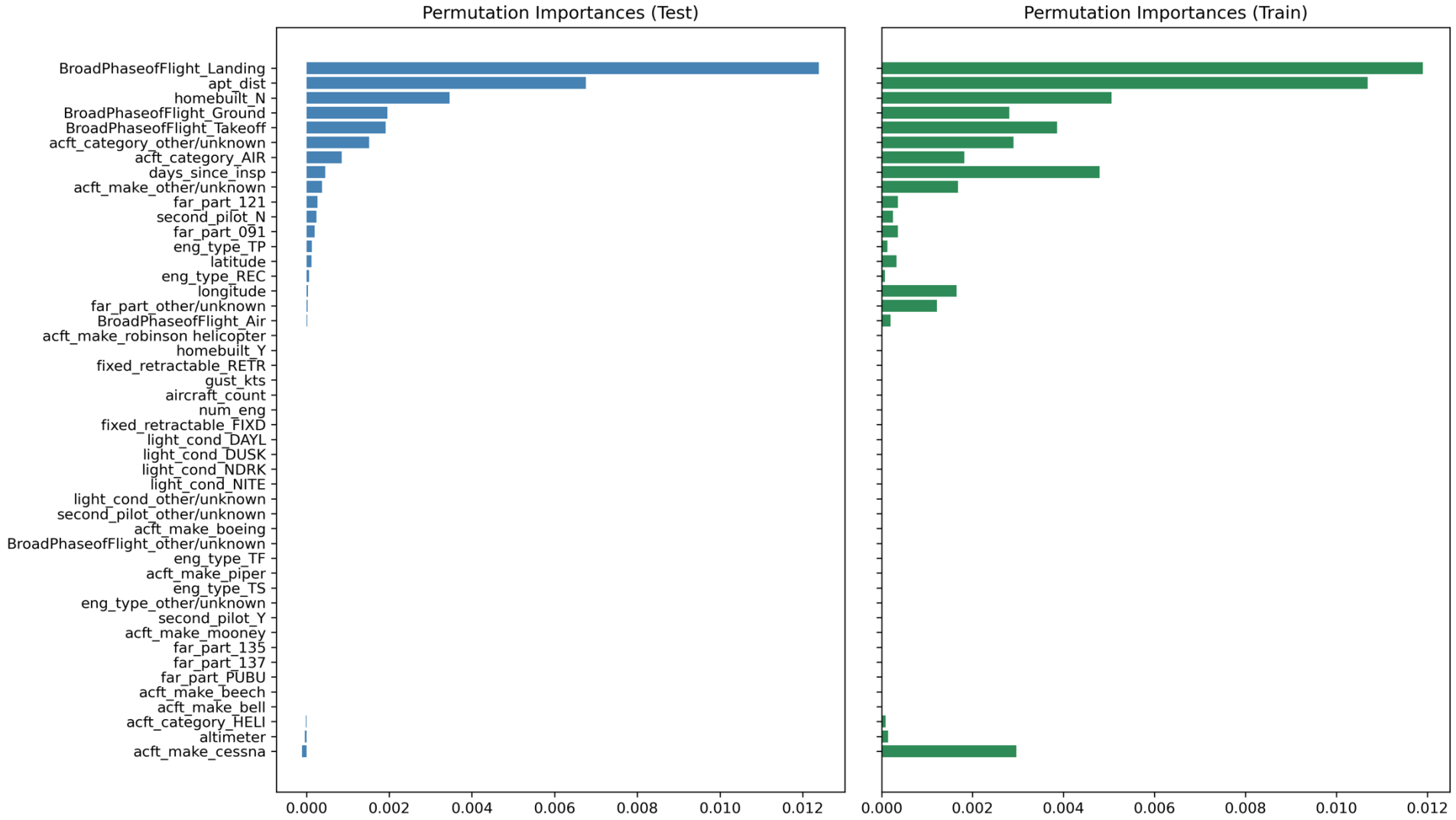
Target: Proportion of **Fatal** Injuries		
Model	**MSE**	**MAE**
Histogram Gradient Boost	0.114	0.232
Naïve Mean (Baseline)	0.133	0.267

Target: Proportion of **Serious** Injuries		
Model	**MSE**	**MAE**
Histogram Gradient Boost	0.075	0.163
Naïve Mean (Baseline)	0.077	0.166

Fatal Injury Proportions: Permutation Importances



Serious Injury Proportions: Permutation Importances



Results

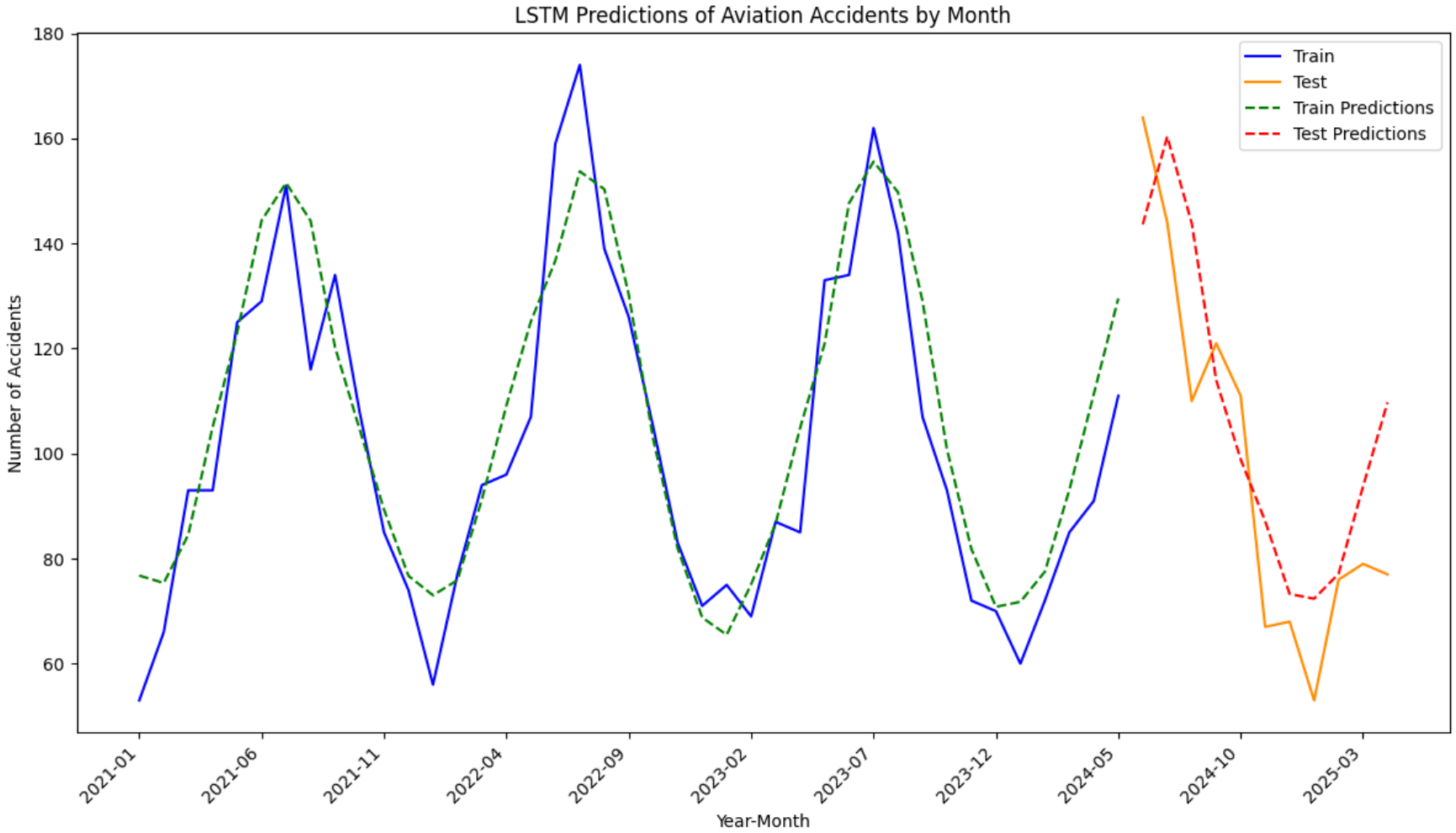
- Predicting damage:
  - Decent (45%) improvement over baseline
    - Indicates location and inspections important
- Proportion of injuries:
  - Modest (14%) improvement for predicting fatal injuries
  - No (2.5%) improvement for predicting serious injuries

Insights

- Across all our models, features from the various NTSB datasets show **limited predictive power**
- They offer only **marginal improvement** over naive baseline predictors
- Given these limitations, we shift focus to a different task:

Using time series models to predict the number of aircraft accidents per month

Time Series Prediction on Monthly Accidents



Conclusion & Future Work

- Accident severity may be linked to the time since last inspection.
- Regulators should assess if more frequent inspections would reduce severity.
- Use more advanced techniques to uncover complex patterns in the data.
- Build a web-based tool to visualize trends and explore "what-if" scenarios.

Thank you!