

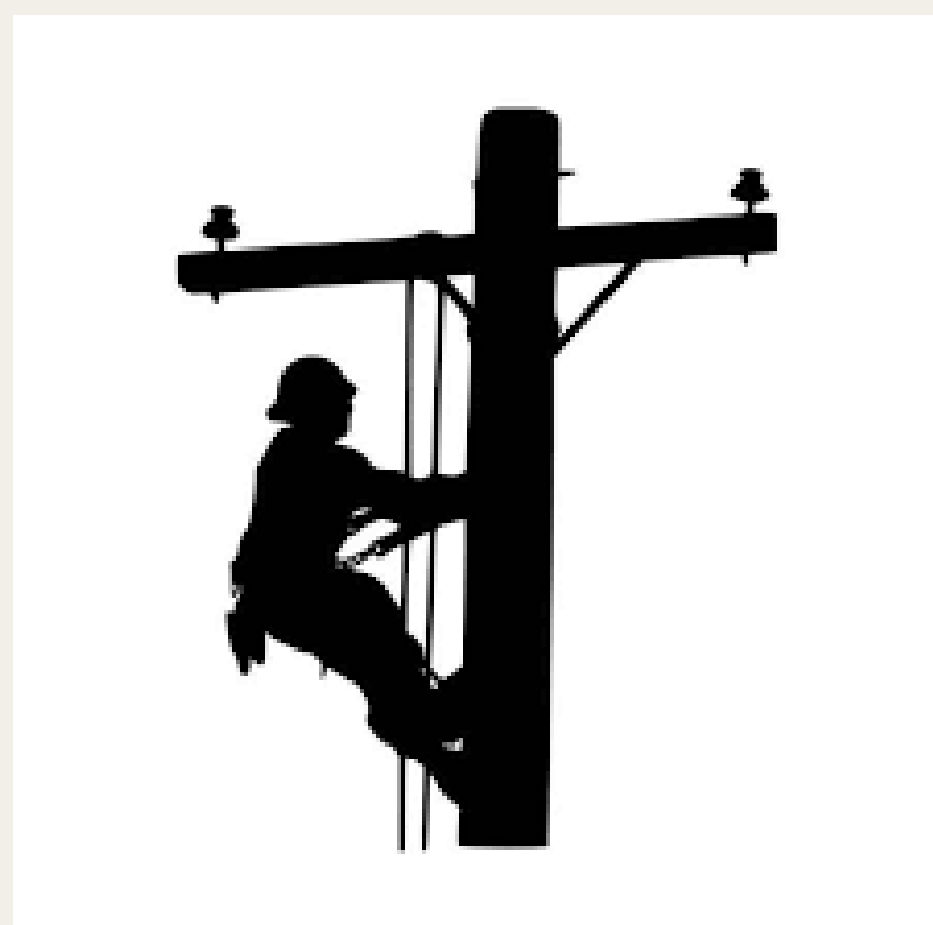
Classification of Distribution Power System Outages

Haoyuan Sun

Advisors: Michela Taufer and Mike Wyatt

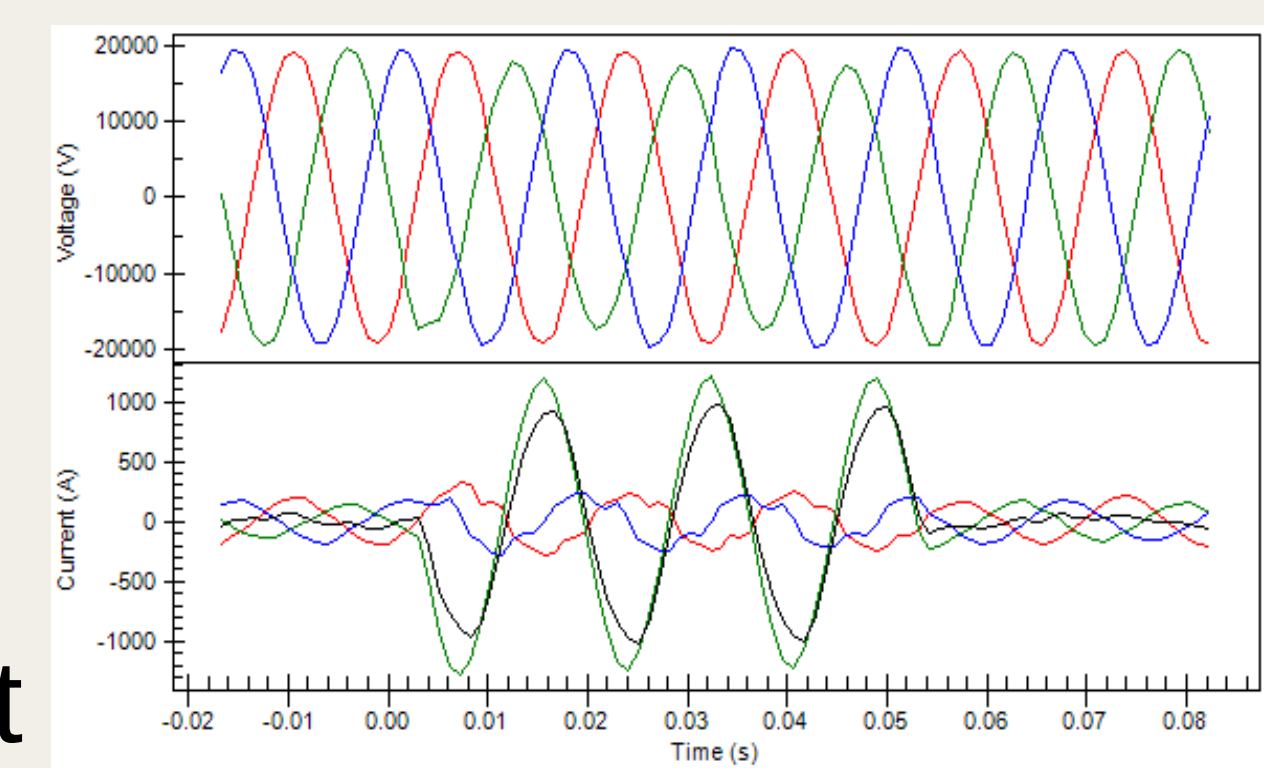
Motivation

- Replace manual classification and thus save manpower
- Help figure out cause of the outage and provide hints on how to clear it
- Reduce system recovery time



Power System Outage Data Set

- An EPRI / DOE dataset
- 295 outage records
- Data is imperfect
 - Some entries are missing
 - Some categories are redundant
- Mixed numerical and categorical data



Data Pre-Processing

Delete and Merge Redundant Categories

Planned
Customer request
Other
Undetermined
missing
unknown

Not Outages

Unknown

What is One-hot Encoding and why

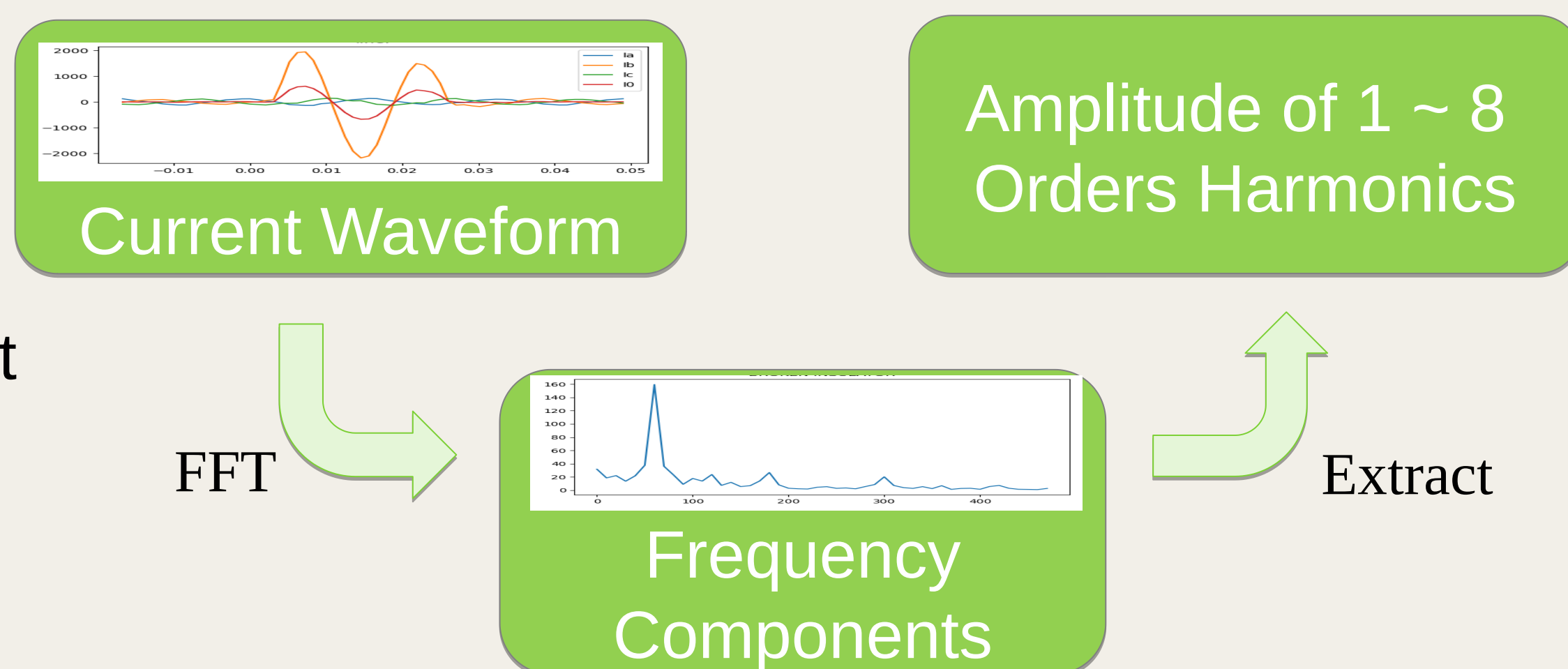
- Convert categorical data into numerical form
- Suitable for data without ordinal relationship
- Avoid predictions halfway between categories

Cause	Code	Weather	Code
1.Equipment	1000000	1.Clear Weather	10000000
2.Tree	0100000	2.Lightning	01000000
3.Vehicle	0010000	3.Major Storm	00100000
4.Lighting	0001000	4.Wind	00010000
5.Animal	0000100	5.Thunderstorm	00001000
6.Weather	0000010	6.Snow/Ice	00000100
7.Customer Caused	0000001	7.Windy	00000010
8.Unknown	0000000	8.Heat	00000001
		9.Unknown	00000000

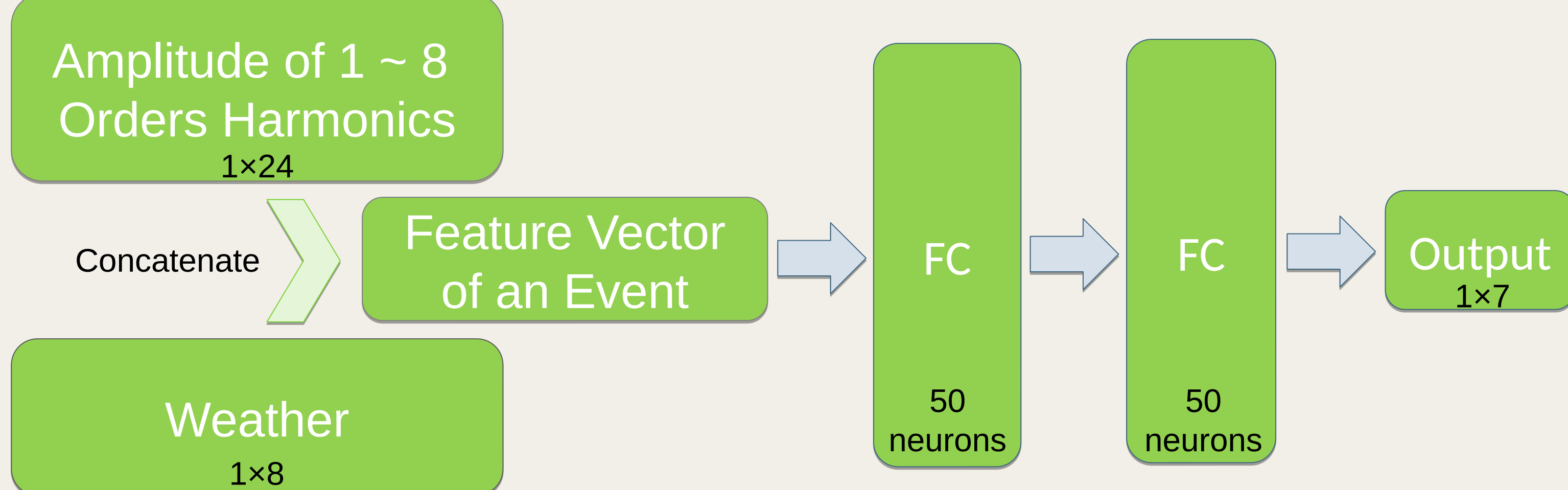
Feature Extraction

Features

- ✓ 1 ~ 8 orders harmonic amplitude
- ✓ Weather condition at time of the outage

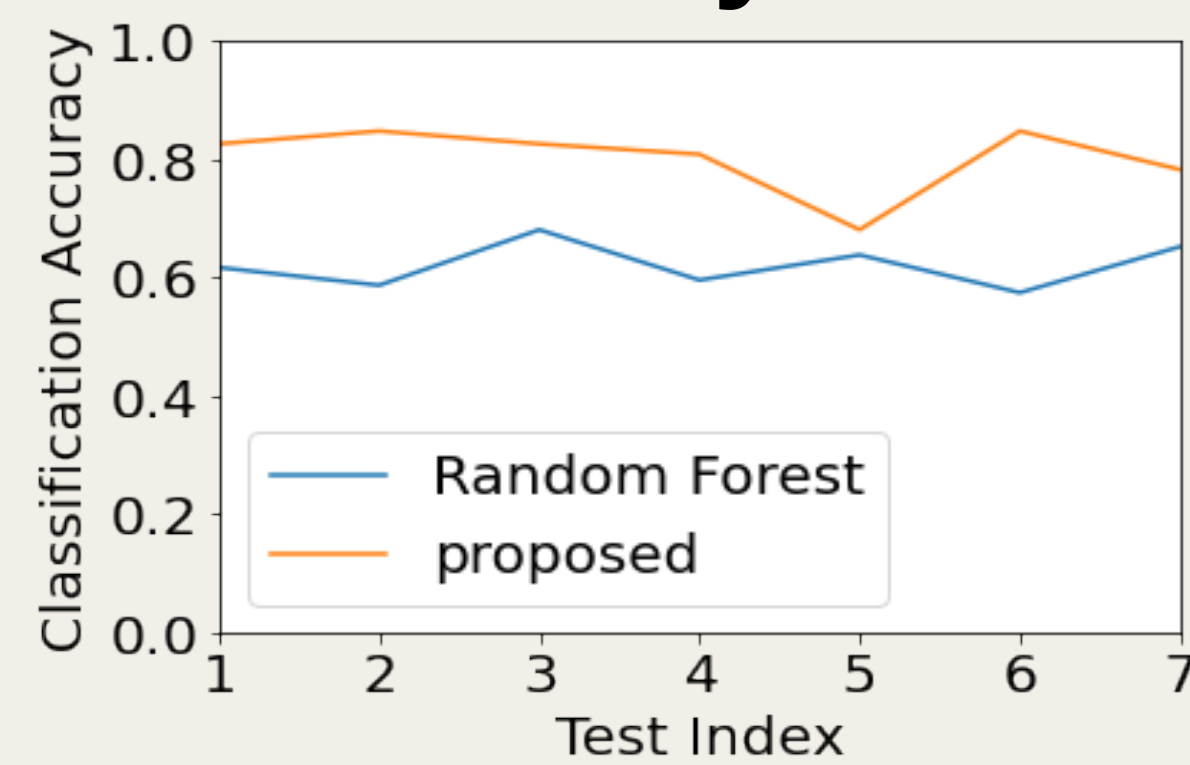


Proposed Classifier



Method Evaluation

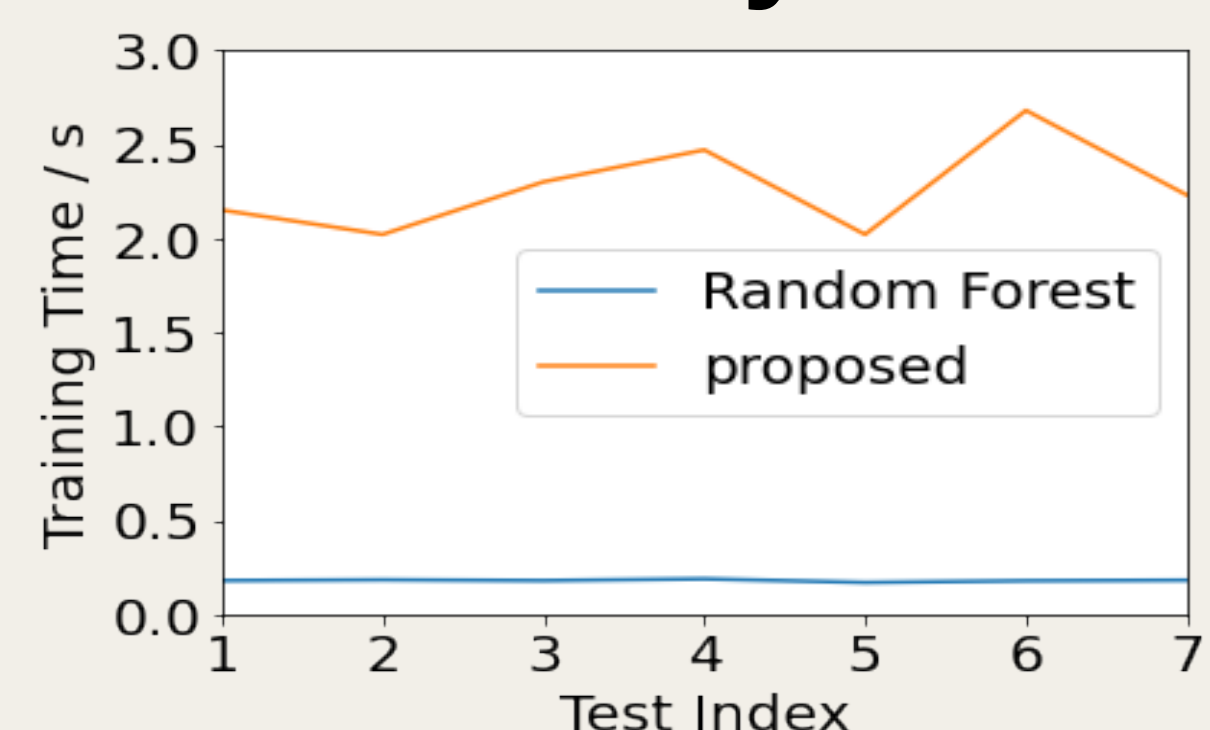
Accuracy



Confusion matrix

Baseline Method	Proposed Method
[[21 0 0 0 0 0 0]]	[[20 0 1 0 0 0 0]]
[[4 5 0 0 0 0 0]]	[[2 3 0 4 0 0 0]]
[[5 0 1 0 0 0 0]]	[[1 0 5 0 0 0 0]]
[[2 0 0 3 0 0 0]]	[[0 0 0 5 0 0 0]]
[[3 0 0 0 0 0 0]]	[[0 0 0 0 3 0 0]]
[[2 0 0 0 0 0 0]]	[[0 0 1 1 0 0 0]]
[[1 0 0 0 0 0 0]]	[[1 0 0 0 0 0 0]]

Efficiency



[[18 0 2 0 0 0 0]]	[[20 0 0 0 0 0 0]]
[[5 5 0 0 0 0 0]]	[[1 4 0 3 2 0 0]]
[[4 0 2 0 0 0 0]]	[[1 0 5 0 0 0 0]]
[[3 0 0 2 0 0 0]]	[[0 0 0 5 0 0 0]]
[[3 0 0 0 0 0 0]]	[[2 0 0 0 1 0 0]]
[[1 0 1 0 0 0 0]]	[[0 1 0 0 0 0 1]]

Conclusion

The proposed method ...

- Achieves consistently higher performance than baseline
- Works well on first four categories
- Does not work well on last three categories. Very likely due to lack of data for these categories
- Takes more time to train but is acceptable

Project Summary

In this project, I ...

- Get more familiar with data frame operation
- Learn stratified split of train and test set
- Know more about Matplotlib
- Make my first poster ever

Future Works

- Try more powerful feature extracting methods to further distinguish between the categories that are now mixed
- Deal with categories that only have a few cases

Reference

Wilson, Aaron, "A hierarchical approach to automated identification of anomalous electrical waveforms" (2019). *Masters Theses and Doctoral Dissertations*.