Project 2

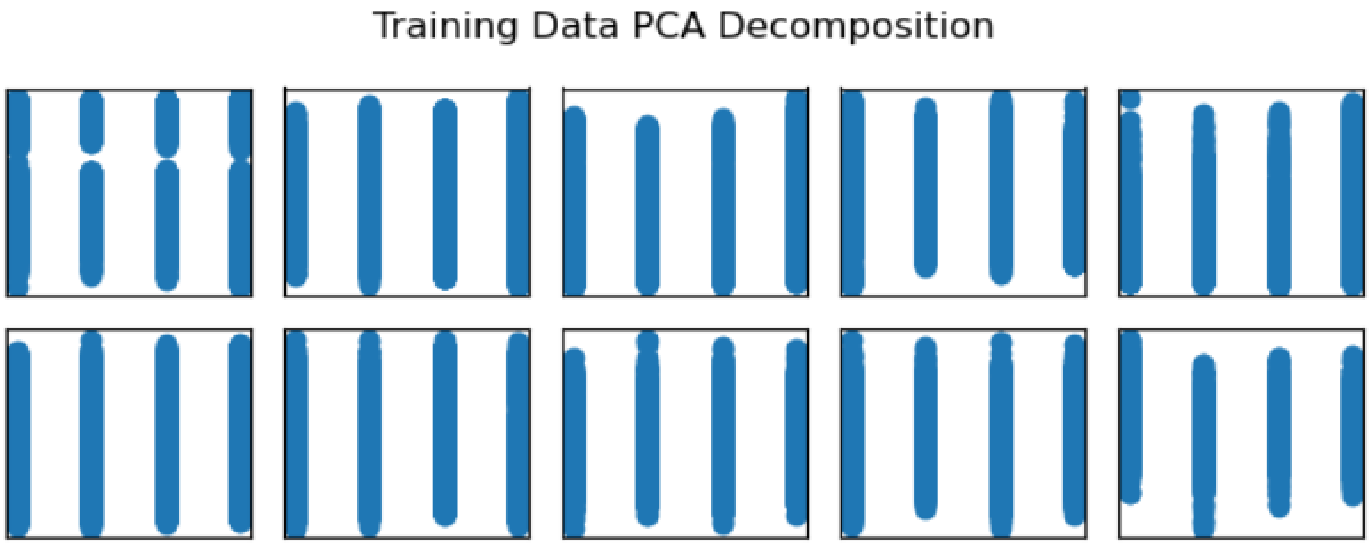
Jack Hanling

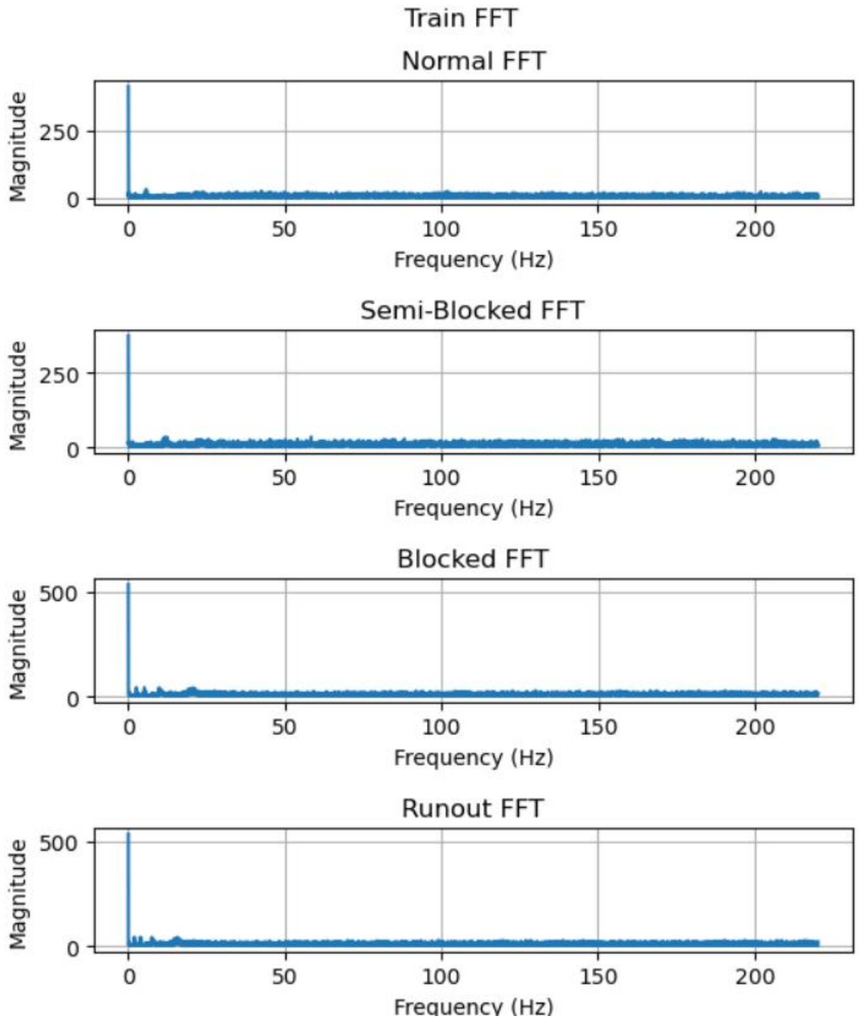
Saihari Kota

Introduction

- Goal: Build an ML model to accurately predict 3D Printer health status
- 4 states of 3D printer health:
  - Normal
  - Semi-blocked
  - Blocked
  - Run Out of Material
- 20 Features: data collected by Acoustic Emission (AE) Sensor
- Training Set Size: 10592, 12261, 11079, 14544
- Test Set Size: 1998, 2000, 1999, 2000

Methodology: PCA & FFT





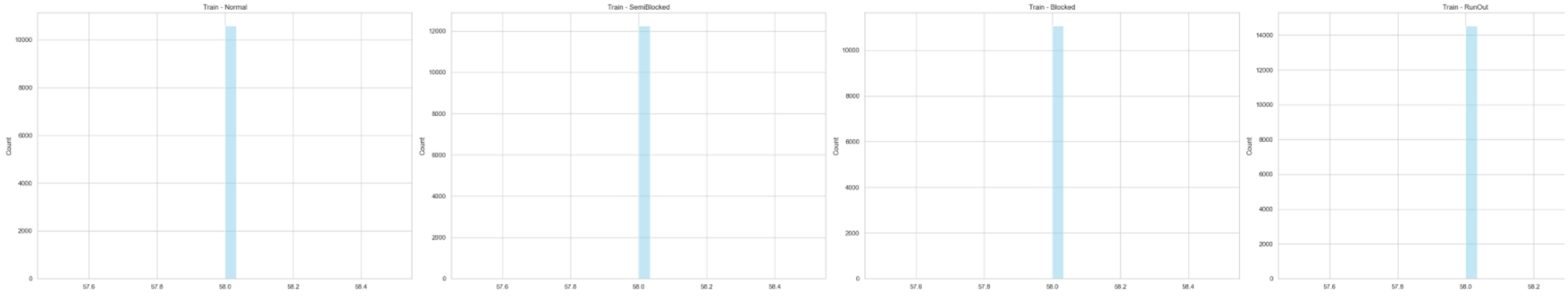
Methodology: KNN and MLP

- Preprocessing
  - Normalized [0,1]
  - FFT for Temporal Features (100 point batches)
  - PCA for Spatial (Characteristic) Features (Averaged over 100 batches)
- K Nearest Neighbors (250 Neighbors)
- Multilayer Perceptron Classifier
  - Activation Function: 'tanh'
  - Solver = 'adam'

Methodology: Data Processing

- Plotted histograms for each feature to determine:
    - Distribution shape
    - Mean and standard deviation for each state
    - Statistical significance of each feature
- Eliminated insignificant feature: Threshold
- Overall Goal: Identify critical features using ANOVA test
    - AE count, amplitude, RMS, ASL, ABS Energy, FREQPP1, FREQPP2, FREQPP3

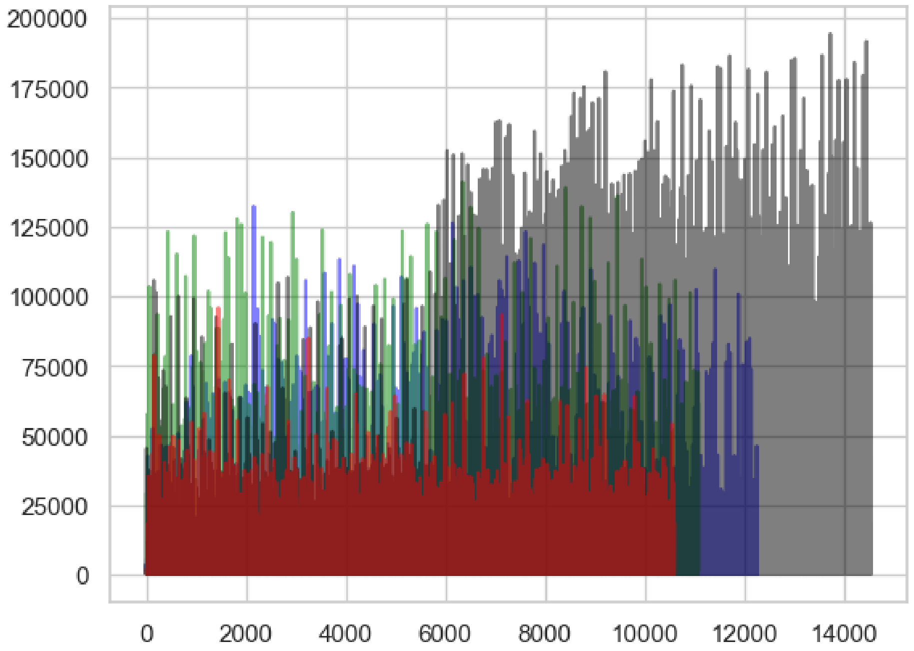
	Feature	Normal	Semi-blocked	Blocked	Run-out	ANOVA_p
6	Feature_6	0.03 ± 0.00	0.03 ± 0.00	0.03 ± 0.00	0.03 ± 0.00	0.000000e+00
7	Feature_7	51.54 ± 0.63	51.71 ± 0.94	52.59 ± 0.65	51.42 ± 1.10	0.000000e+00
4	Feature_4	58.96 ± 1.10	59.02 ± 1.16	59.57 ± 1.48	58.95 ± 1.36	0.000000e+00

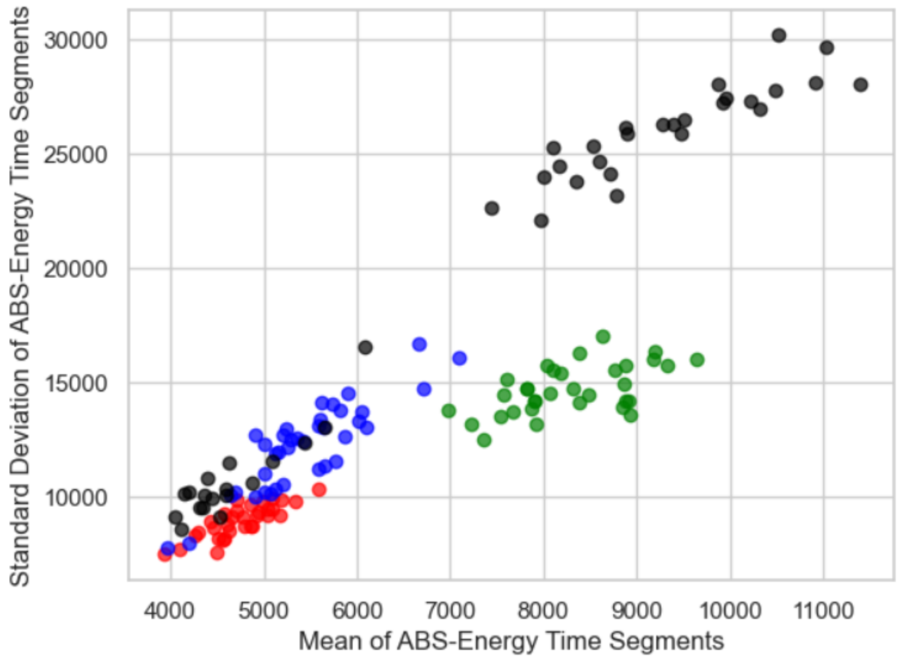


Distribution of Feature 10

Methodology: Time Segmenting

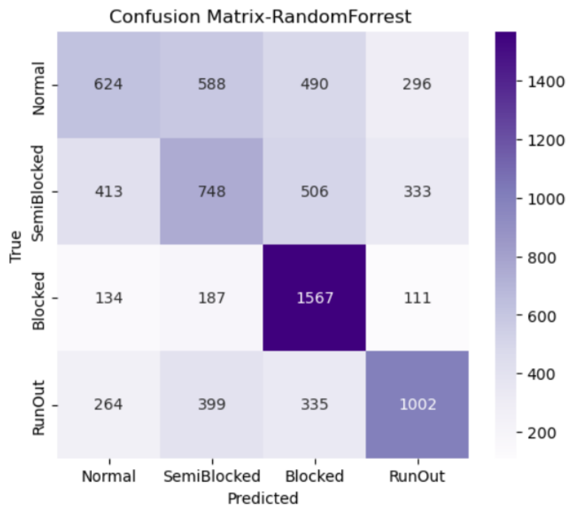
- Segmented training and test datasets in windows of time
  - Reduces effect of noise
  - Captures machine behavior and patterns over time
  - Differentiates states much better
- Windows of 500ms were chosen

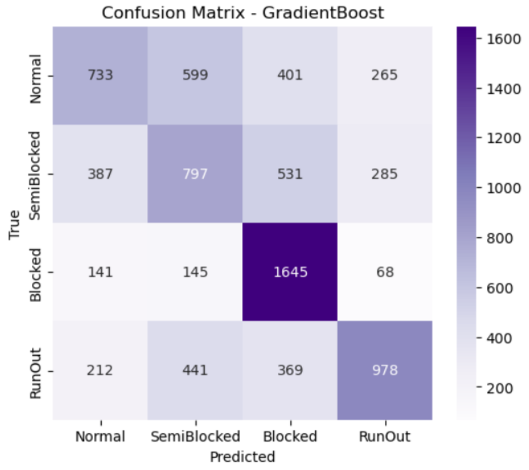


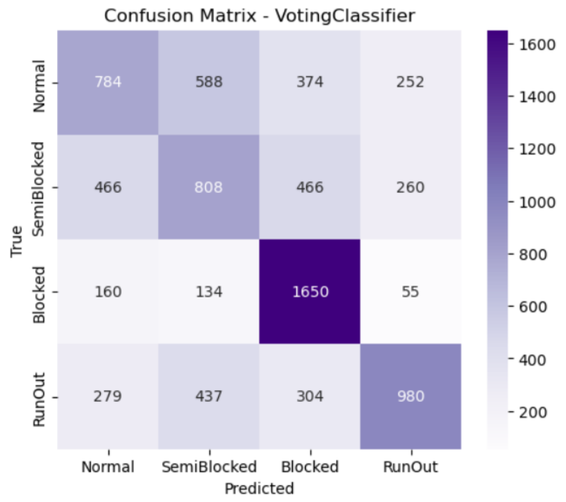


Results: Basic Models

- Training several models on the base training dataset resulting in poor accuracy
- RandomForest: 49.28%
- ExtraTrees: 51.54%
- DecisionTree: 49.46%
- MLPClassifier: 30.04%
- GradientBoosting: 51.93%
- VotingClassifier (RandomForest, GradientBoost, Logistic Regression) : 52.80%







Results: FFT & PCA Preprocessing

- FFT was used on batches of 100 samples to find periodic trends
- Used with an MLP Classifier and K Nearest Neighbors Classifier
- KNN = 25% Accuracy
- MLP = 57% Accuracy (1000,100,100)

- PCA was conducted on 20 independent variables, and averaged over 100 time intervals to be consistent with FFT
- Used with an MLP Classifier and K Nearest Neighbors Classifier
- K Nearest Neighbors: 25% Accuracy
- MLP Classifier: 25% Accuracy (100,100)

- Combination of PCA and FFT used with MLP and KNN
- K Nearest Neighbors: 25%
- MLP Classifier: 55%

	precision	recall	f1-score	support
1	1.00	0.45	0.62	22
2	0.90	1.00	0.95	9
3	0.40	0.44	0.42	9
4	0.00	0.00	0.00	0
accuracy			0.57	40
macro avg	0.57	0.47	0.50	40
weighted avg	0.84	0.57	0.65	40

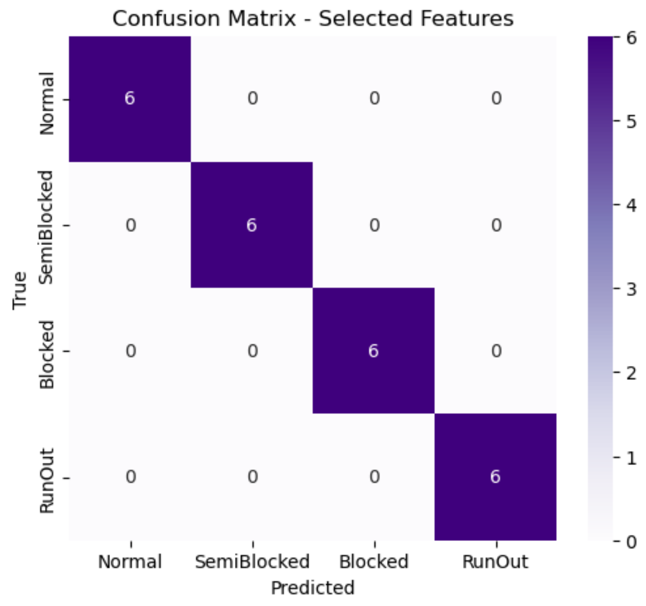
precision recall f1-score support

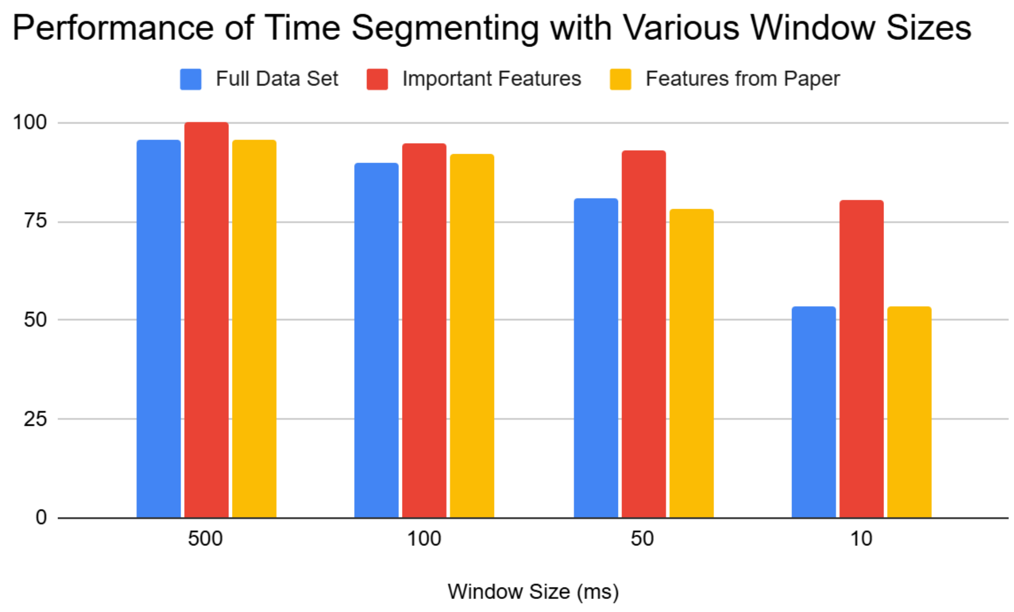
	precision	recall	f1-score	support
1	0.00	0.00	0.00	0
2	0.00	0.00	0.00	0
3	1.00	0.25	0.40	40
4	0.00	0.00	0.00	0
accuracy			0.25	40
macro avg	0.25	0.06	0.10	40
weighted avg	1.00	0.25	0.40	40

	precision	recall	f1-score	support
1	0.90	0.43	0.58	21
2	0.90	0.90	0.90	10
3	0.10	0.33	0.15	3
4	0.30	0.50	0.38	6
accuracy			0.55	40
macro avg	0.55	0.54	0.50	40
weighted avg	0.75	0.55	0.60	40

Results: Time Segmenting

- Used RandomForest Classification model with time segmented data
- Full Data Set: n_estimators: 1
  - 95.83% accuracy
- Using features from research paper (Count, RMS, ABS Energy): n_estimators: 4
  - 95.83% Accuracy
- Using important features determined by statistical significance (AE count, amplitude, RMS, ASL, ABS Energy, FREQPP1, FREQPP2, FREQPP3): n_estimators:25
  - 100% accuracy
- 100ms window
  - Full Data Set: 89.66%
  - Important Features: 94.83%
  - Research Paper Features: 92.24%
- 50ms window
  - Full Data Set: 80.77%
  - Important Features: 92.79%
  - Research Paper Features: 78.37%
- 10ms Window
  - Full Data Set: 53.64%
  - Important Features: 80.44%
  - Research Paper Features: 53.53%





Summary

- Chosen Model: Random Forest with n_estimators: 25, data segmented in 500 ms windows and features: AE count, amplitude, RMS, ASL, ABS Energy, FREQPP1, FREQPP2, FREQPP3
  - Accuracy: 100%
  - 102.92% increase in accuracy over RandomForest model without time segmenting

Future Work:

- Advanced time series transformer
  - Can capture long range dependencies and relationships
  - Can weigh different time steps unlike fixed window segmenting
- Larger training and test data would further validate model
  - Currently there are only 24 500ms windows in the test set