Supervised Machine Learning in Digital Power Line Communications

- **Authors:** *Kushal Thapa, Dr. Stan McClellan, Dr. Damian Valles*
- **Presenter:**

Kushal Thapa

Ingram School of Engineering

Texas State University

San Marcos, TX, USA

Email: k_t260@txstate.edu

TEXAS STATE UNIVERSITY

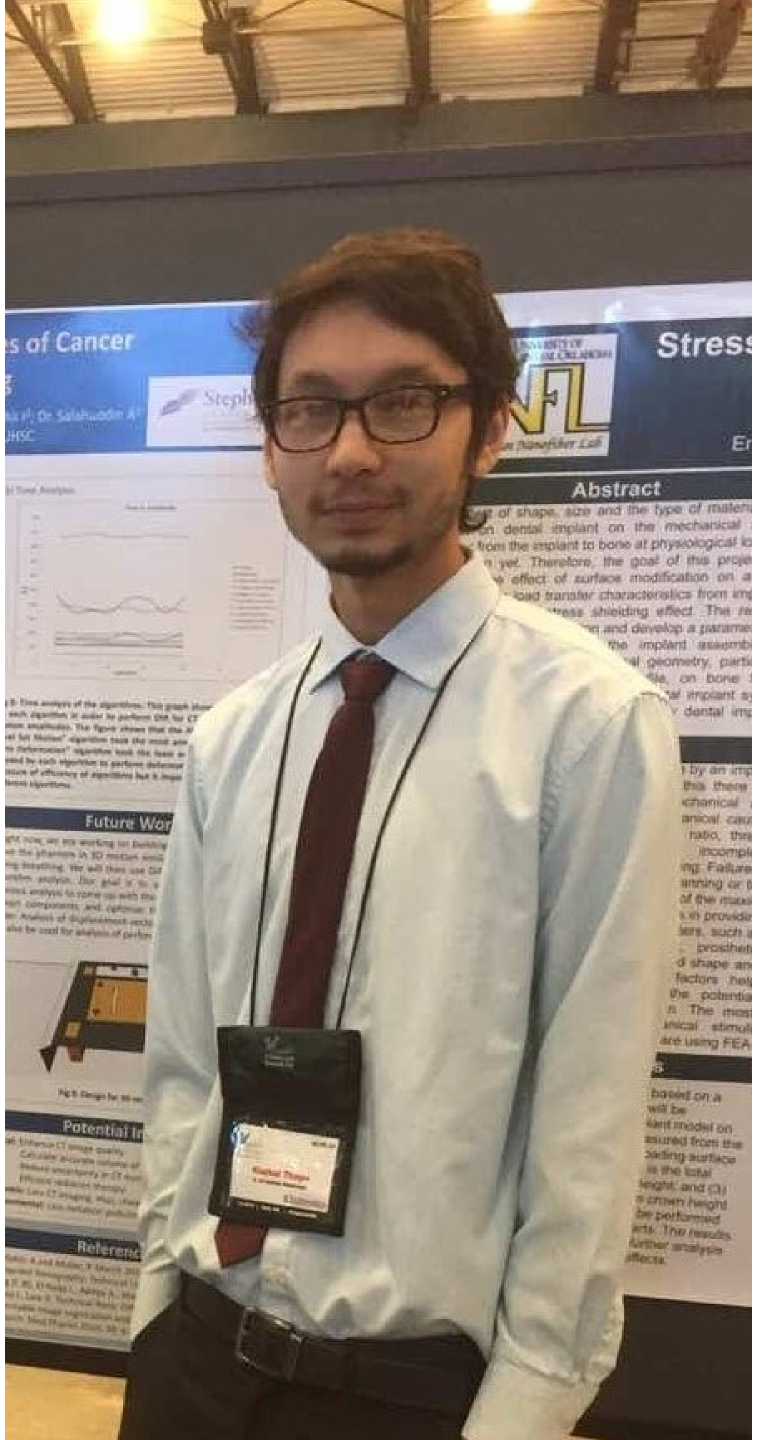
The rising STAR of Texas

IARIA

Presenter’s Bio

- B.S. in Biomedical Engineering from *University of Central Oklahoma*
- Currently a M.S. in Engineering- Electrical Engineering student at *Texas State University*
- Graduate Research Assistant for Dr. Stan McClellan in Ingram School of Engineering
- Website: kushalthapa.wordpress.com

  https://github.com/kushal-thapa



Research Interest

- Power line communications
- Machine Learning in communications
- Distributed network architecture
- Remote filesystem monitoring
- Characterization of heat trace cable

Introduction

- **Power line communications (PLC)**
  - A technology of sending communication signal over the power lines.

- Pros:
  - Pre-built infrastructure
- Cons:
  - Heterogenous medium
  - Lots of transitions

- Applications:
  - Power line home network
  - Broad-band over power line (BPL)
  - Smart utility meters

Research goal

- Problem
  - **Signal degradation** caused by transformers and **dynamic noise** of the power-grid makes demodulating PLC signal at the receiver difficult

- Proposed solution
  - Use **Machine Learning (ML)** to extract information carried by the communication signal because of its high sensitivity, ease of use, and scalability

- Goals
  - Build a test architecture to send and collect PLC signal
  - Use various ML algorithms to classify the received PLC signal

Signal Flow

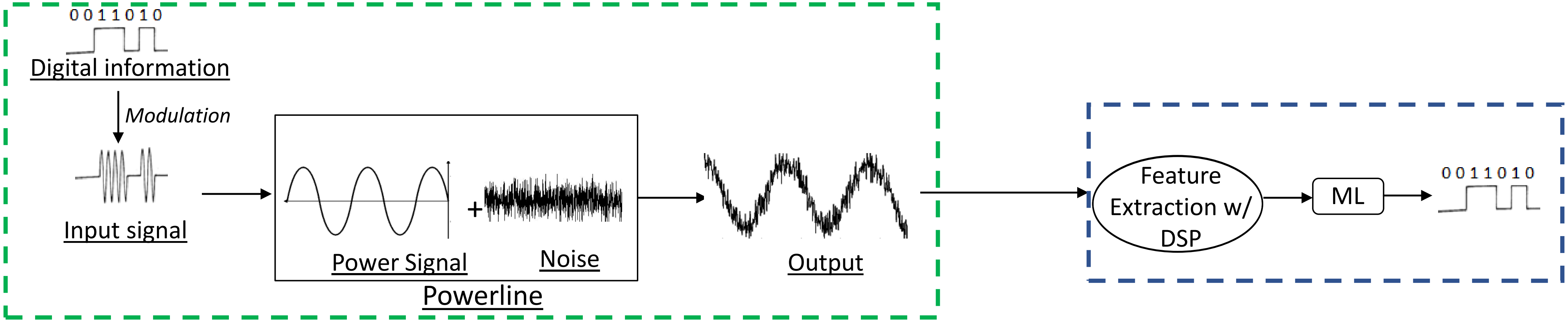


Figure 1. Signal flow from digital input to ML output. The diagram within the green dotted box shows the transmitted and received PLC signal. This raw output is then processed and fed into ML models to extract the original digital information as shown in the blue dotted box.

PLC Network Architecture

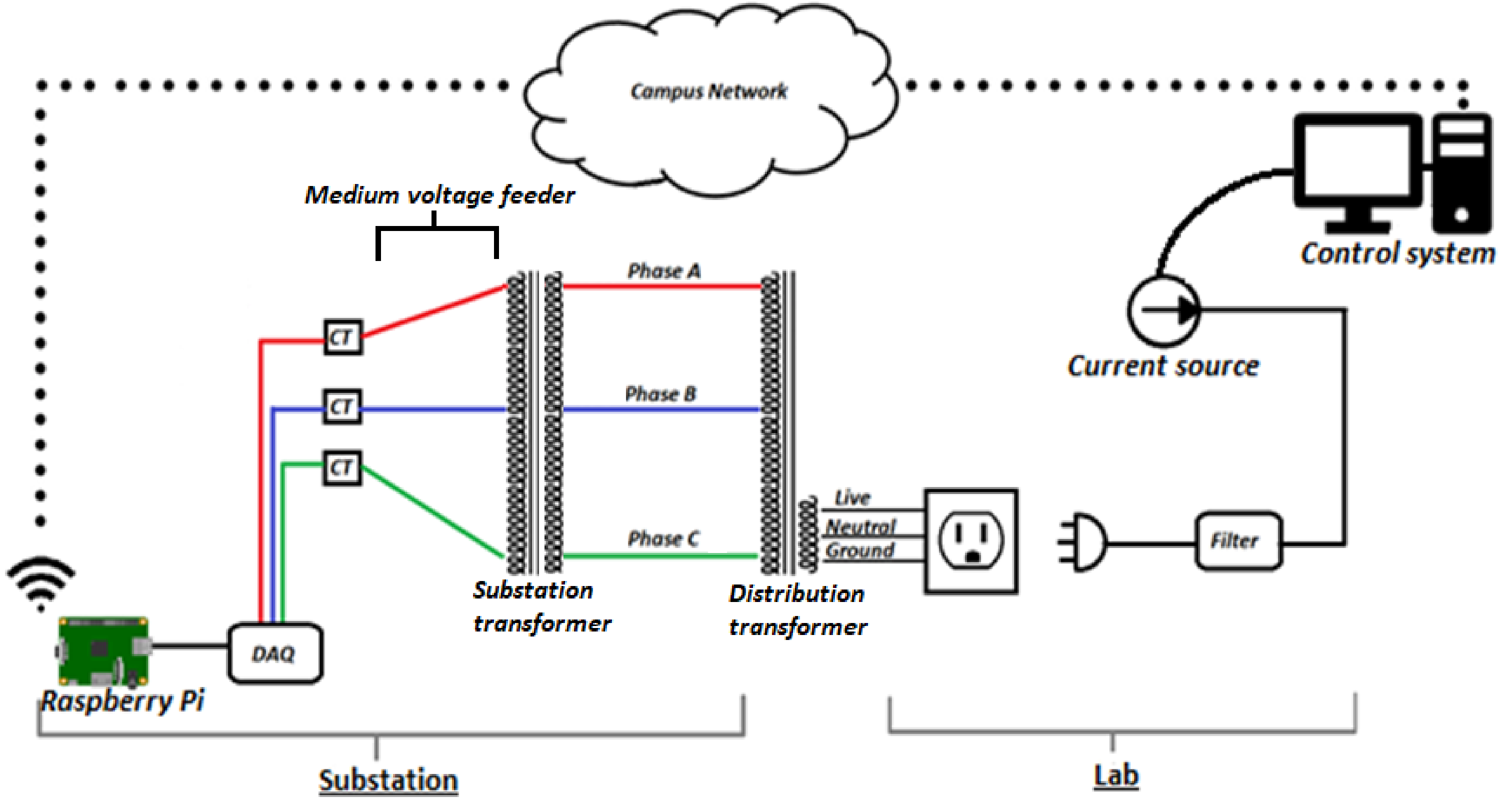


Figure 2. Experimental setup for sending and receiving PLC signals in a distribution power grid. The signal originates at the current source in the lab which is then injected into the power grid and is collected at the substation using Data Acquisition Device (DAQ).

Raw Data

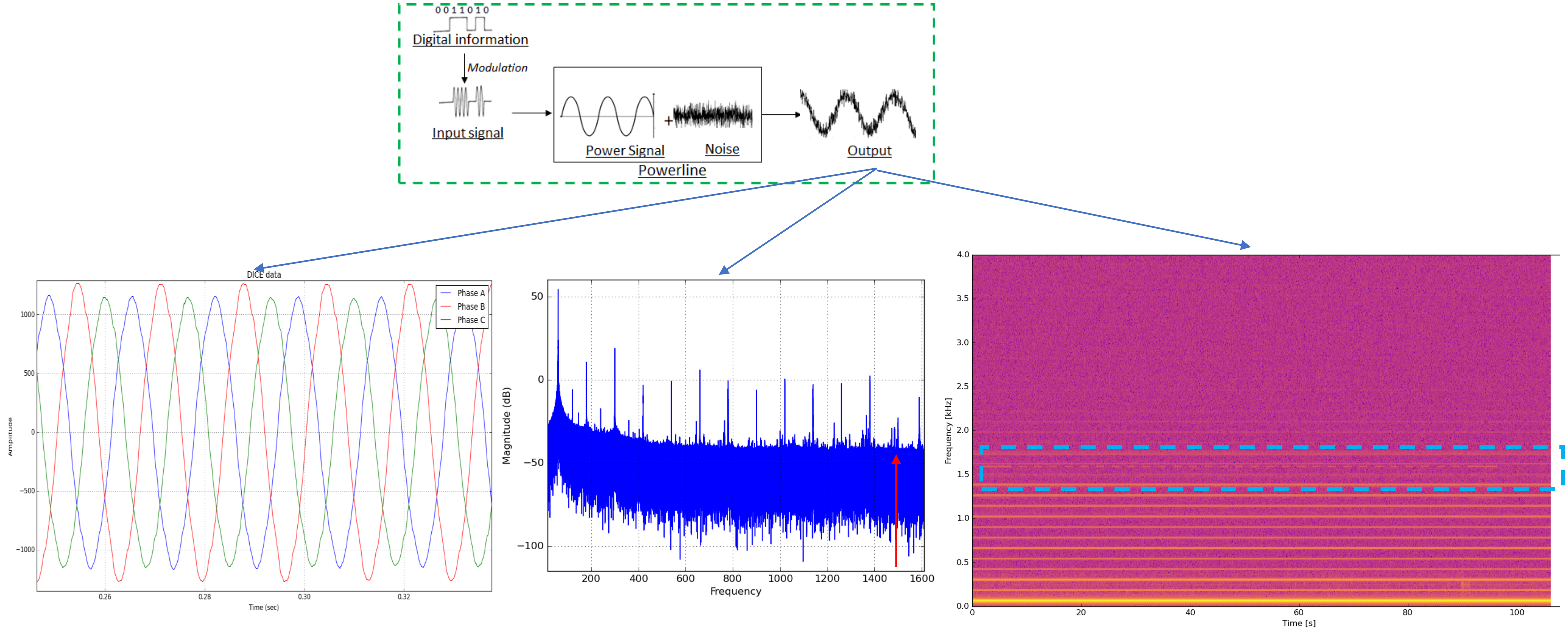
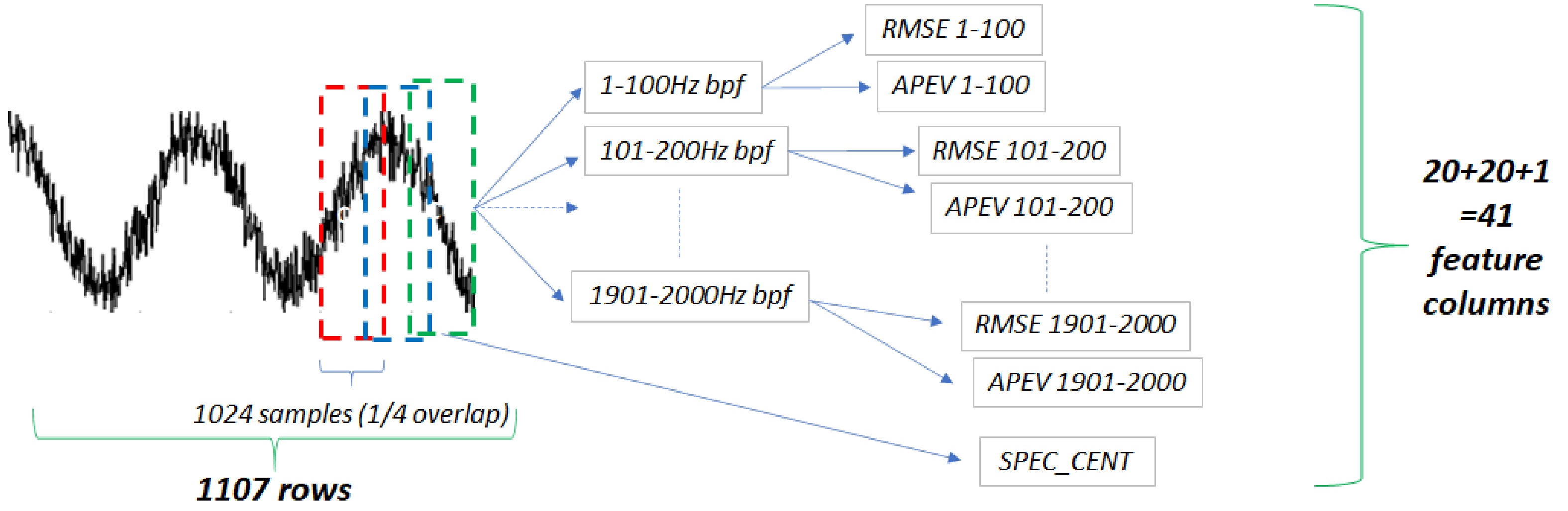


Figure 3. Visualization of raw data: 3-phase time-series signal (bottom-left), FFT diagram of the phase A signal (bottom-mid), and spectrogram of the phase A signal (bottom-right). The frequency of the input signal was 1595Hz and the arrow in the FFT diagram and the blue dotted box in the spectrogram shows this frequency where the trace of the input signal can be seen.

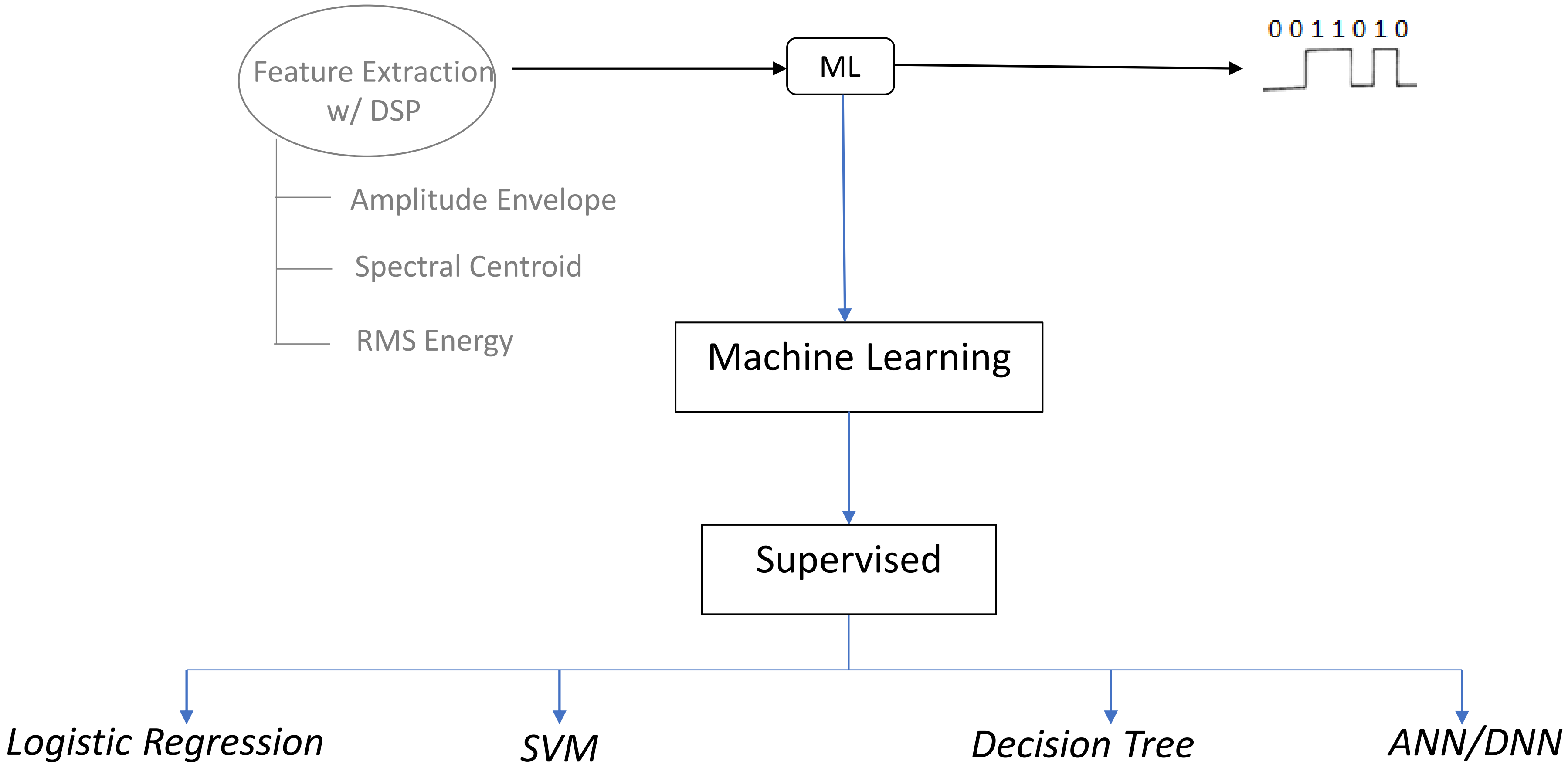
Feature Extraction



Index	MSE 1-10	MSE 101-2	MSE 201-3	MSE 301-4	MSE 401-5	MSE 501-6	MSE 601-7	MSE 701-8	......	APEV 1501-1600	APEV 1601-1700	APEV 1701-1800	APEV 1801-1900	APEV 1901-2000	SPEC_CENT	Target
0	727.161	137.573	44.485	30.0338	22.347	17.6945	15.0706	13.7141		1070.88	1079.46	1079.46	1073.32	1074.5	145.691	0
1	764.007	103.652	17.6669	14.9728	2.83444	3.31799	5.84989	2.75421		1077.25	1079.75	1075.48	1074.53	1074.89	164.091	0
2	753.868	101.585	17.6219	14.9859	2.93606	3.25943	5.8733	2.83213		1075.8	1082.92	1079.82	1078.72	1078.72	159.372	0
3	764.028	103.56	17.6966	14.9785	3.05845	3.3071	5.84552	2.78956		1077.97	1077.13	1074.38	1076.32	1072.56	158.026	0
4	751.976	101.553	17.2255	14.5067	3.03499	3.37509	5.85301	2.79199	......	1069.66	1070	1075.24	1072	1075.67	163.087	0
5	765.779	103.414	17.6677	14.9232	2.90059	3.18236	6.01005	2.77238		1067.74	1070.73	1075	1072.03	1074.66	161.504	0
6	753.836	102.001	17.8004	15.0667	2.98042	3.22529	5.91431	2.8863		1071.92	1075.41	1075.41	1072.52	1074.95	159.707	0
7	765.672	103.179	17.9968	15.3215	3.01923	3.25779	5.93764	2.94785		1068.44	1073.94	1068.17	1067.15	1066	164.873	0
8	752.564	102.089	17.8643	15.1625	3.11703	3.36963	6.08173	2.95874		1073.37	1070.99	1072.09	1075.13	1075.16	165.582	0

Figure 4. The time-series output from the data-capture was divided into multiple overlapping frames. Spectral centroid (SPEC_CENT) was calculated for each frame. Then, each of these frames were passed through 100Hz bandwidth bandpass filters from 1Hz to 2000Hz. RMS Energy (RMSE) and amplitude envelope (APEV) were calculated for the filtered signals. The resulting dataset had 41 columns and 1107 rows.

Machine Learning Algorithms



Results

Grid Search

Phase A only:

Classifiers	Best parameters	Training acc	Testing acc
Logistic Regression	C=1.0, solver=lbfgs	94.06	93.99
SVM	C=1000, gamma=0.001	94.45	95.19
Decision Tree	Max_depth=1, Min_samples_split=1.0	94.19	95.19

Full dataset:

Classifiers	Best parameters	Training acc	Testing acc
Logistic Regression	C=1.0, solver=lbfgs	77.27	76.73
SVM	C=10, gamma=0.1	77.15	75.52
Decision Tree	Max_depth=5, Min_samples_split=7	73.84	72.22

Results

Feature selection

Phase A only:

Classifiers	Two best features	Training acc	Testing acc
Logistic Regression	RMSE 201-300 and **RMSE 1501-1600**	93.29	94.58
SVM	RMSE 1301-1400 and **RMSE 1501-1600**	95.53	96.78
Decision Tree	RMSE 1-100 and **RMSE 1501-1600**	95.70	95.19

Full dataset:

Classifiers	Two best features	Training acc	Testing acc
Logistic Regression	RMSE 501-600 and **RMSE 1501-1600**	73.43	72.21
SVM	RMSE 701-800 and **RMSE 1501-1600**	76.60	74.92
Decision Tree	RMSE 501-600 and **RMSE 1501-1600**	78.86	74.51

Results

Learning Curves

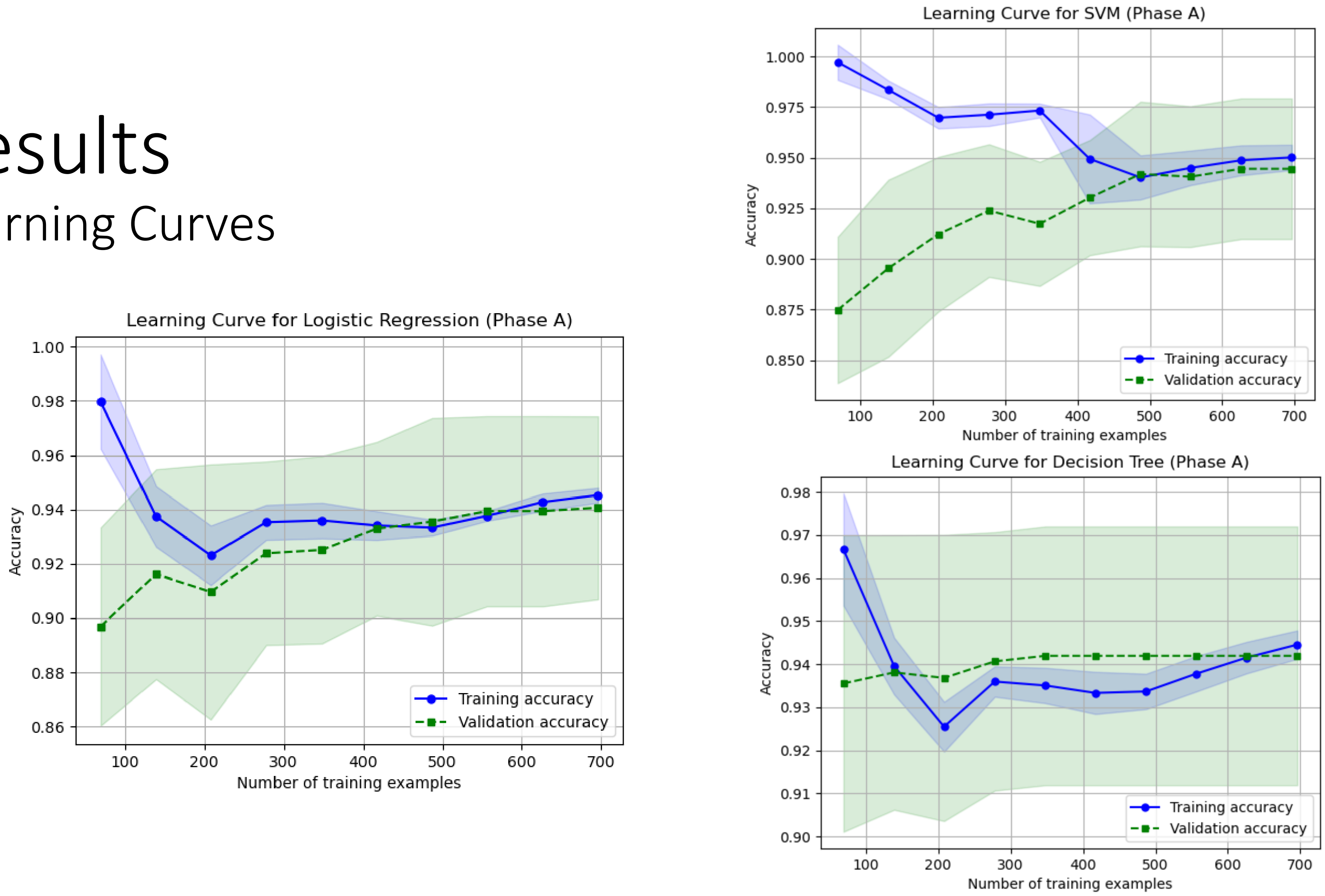


Figure 5. Learning curves of Logistic Regression (left), SVM (top right) and Decision Tree (bottom right) models for our Phase A PLC dataset. The convergence of training and validation accuracy curves in these plots show that none of these models were overfitted or underfitted.

Results

Ensemble learning

Classifiers	Phase A ROC AUC	Phase A Accuracy	Fullset ROC AUC	Fullset Accuracy
Logistic Regression	0.97 (+/-0.02)	0.94 (+/-0.02)	0.86 (+/-0.02)	0.77 (+/-0.02)
SVM	0.97 (+/-0.02)	0.93 (+/-0.02)	0.85 (+/-0.02)	0.76 (+/-0.02)
Decision Tree	0.91 (+/-0.02)	0.92 (+/-0.02)	0.82 (+/-0.02)	0.74 (+/-0.02)
Majority Voting	0.97 (+/-0.02)	0.93 (+/-0.02)	0.87 (+/-0.02)	0.76 (+/-0.02)

Results

ROC AUC Curves

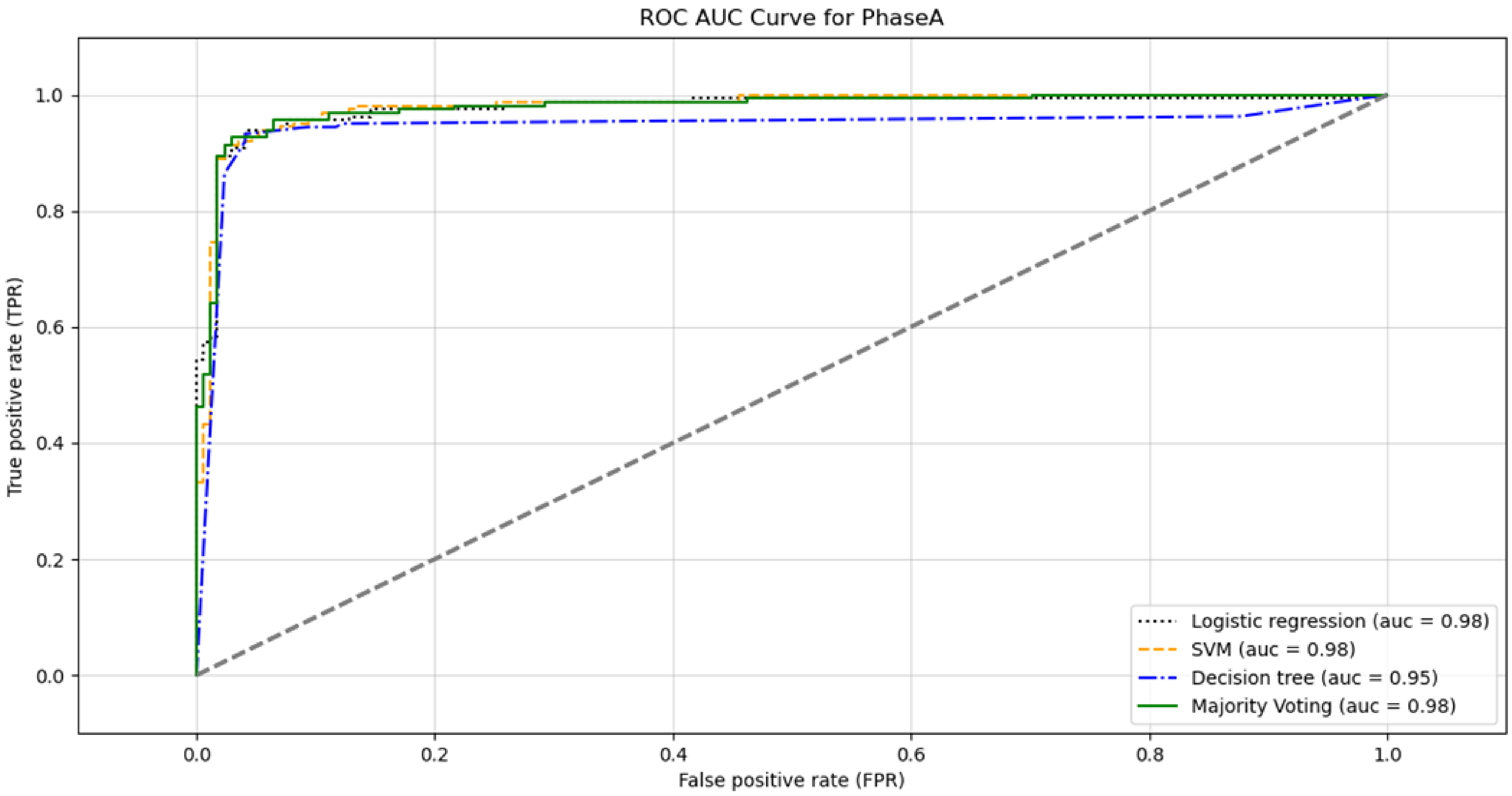
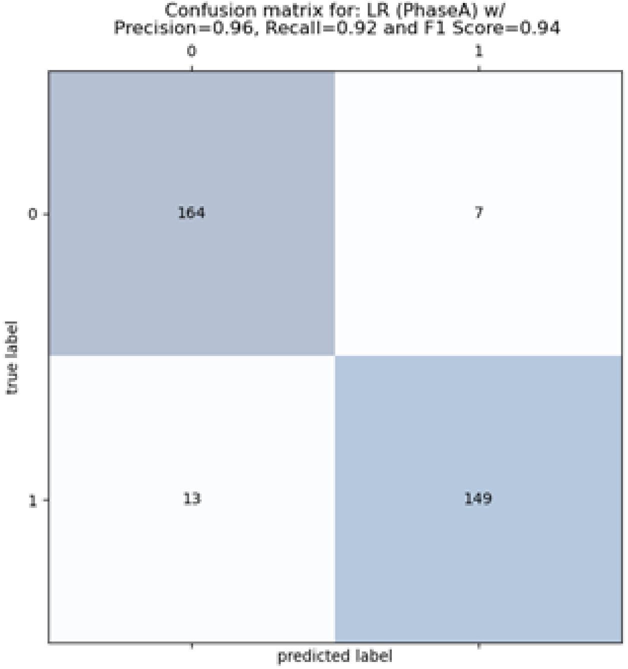
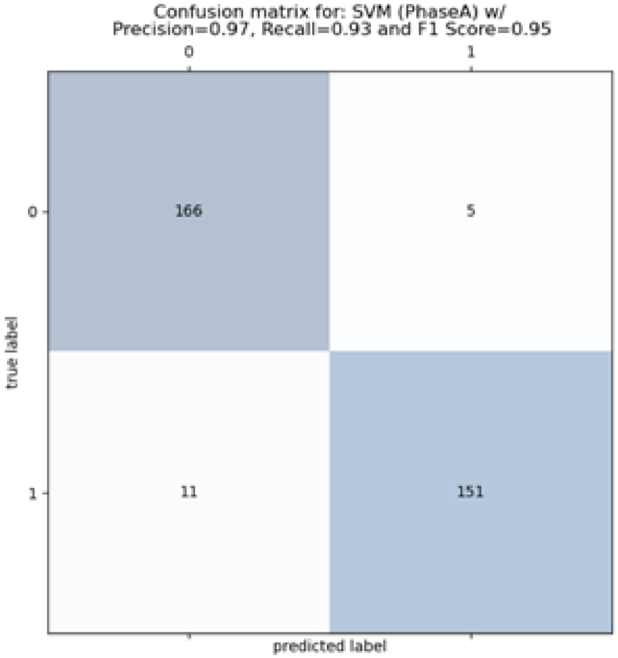


Figure 6. ROC Curve of LR, SVM, TREE and Majority Voting models with AUC score. The curves and AUC score show that all these models performed similarly with Decision Tree being slightly worse than the rest.

Results
Confusion Matrix





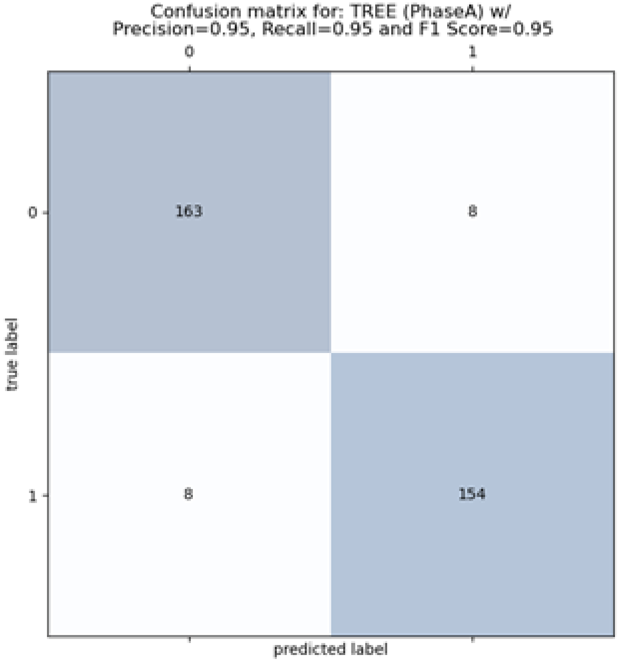


Figure 7. Confusion matrix plots of the LR (left), SVM (mid) and TREE (right) for Phase A PLC dataset. Precision, Recall and F1 scores of the models are also shown on the plot titles. Based on these scores and the confusion matrix, SVM performed slightly better than the rest.

Results
ANN

- **Optimal parameters:**
  - 2 hidden layers
- 50 nodes each
- Activation functions: tanh of both hidden layers and sigmoid of output layer
- Adam optimizer with learning rate of 0.01 and beta decay of 1e-5

- **Best accuracy:**
- Training – 98.19%
- Testing – 94.29%

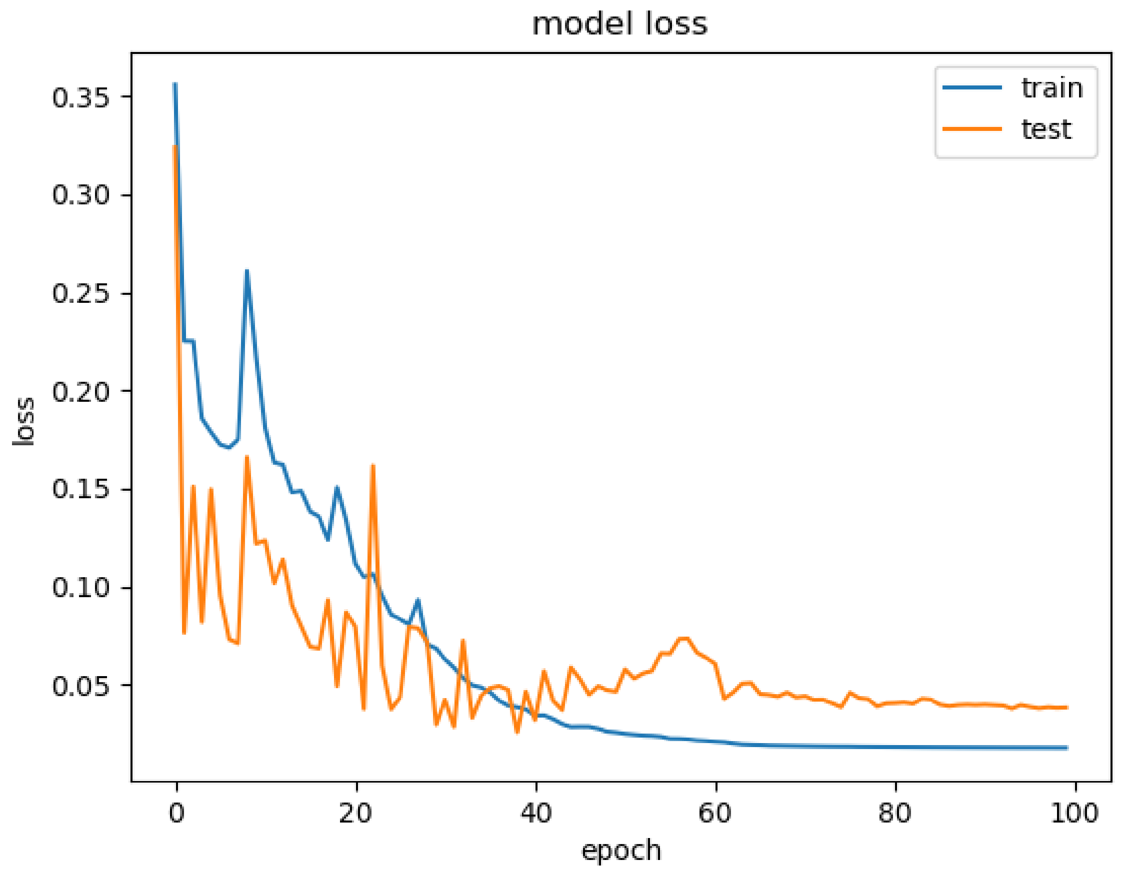


Figure 8. Loss curve of the ANN model showing stabilization of the training and validation (test) loss within 100 epochs.

Conclusion

- SVM algorithm performed slightly better than Logistic Regression and Decision Tree because of its non-linear classification for our PLC data.
- ANN model outperformed the one-neuron models (LR, SVM and TREE) in terms of accuracy.
- Phase A dataset had higher accuracies across all ML models compared to the full dataset (3 phase combined).
  - This is because this was the primary phase where the PLC signal was first injected. Phases B and C had some image of the signal, but it was not as prominent in these phases.
- We showed the applicability of ML in PLC