

PHASE TWO IMPLEMENTATION

Four datasets with anomalies were obtained from the GRU, LSTM, RNN and ANN models.

70% of each dataset was training data, and 30% was testing data.

Eight classification models were used: Extreme Gradient Boosting (XGBoost), Logistic Regression, Adaptive Boosting (AdaBoost), Gradient Boosting, Decision Trees, Random Forests, SVM and K Nearest Neighbors (KNN).

Each model was run once without and one with hyperparameter tuning.

In the following, the parts in **yellow** are done, but the parts in **green** are still in progress.

unknown fault detection:

4 variants of DAE should be trained using **only healthy data** under different noise levels, i.e., GRU-DAE, CNN-DAE, CLSTM-DAE and ANN-DAE.

1. **Training process:** model performance during the training process under different hyperparameters (LR, epoch, layers, latent size..) should be documented. In particular, the reconstruction error of the model should be calculated under different training trials until the best model with optimal hyperparameters is found.
2. **Testing process1:** Fault detection performance of GRU-DAE (reconstruction error, recall, precision and F1 score) should be calculated and compared with other variants, e.g., CNN-DAE, CLSTM-DAE and ANN-DAE **under different thresholds.**
3. **Testing process2:** Fault detection performance of GRU-DAE (reconstruction error, recall, precision and F1 score) should be calculated and compared with other variants, e.g., CNN-DAE, CLSTM-DAE and ANN-DAE **under different level of noise.**
4. **Testing process3:** Fault detection performance of GRU-DAE (reconstruction error, recall, precision and F1 score) using testing data **from different driving scenarios** should be calculated.

Question: What are the different driving scenarios ? are they the drift, delay, etc data

Phase 2: fault classification:

1. **Training process:** model performance during the training process against different hyperparameters (LR, epoch, layers) should be documents. Specifically, the training and validation accuracy of the model should be calculated under different training trials until the best model with optimal hyperparameters is found.
2. **Testing process1:** Fault classification performance of trained ensemble model (recall, precision and F1 score) should be calculated and compared with other

traditional classifier, e.g., CNN, SVM and LSTM under single and concurrent faults !

3. **Testing process2:** Fault classification performance of trained ensemble model (recall, precision and F1 score) should be calculated and compared with other traditional classifier, e.g., CNN, SVM and LSTM under different level of fault combinations (2, 3,4).
4. **Testing process3:** Fault classification performance of ensemble model (recall, precision and F1 score) should be analysed under different data size (number of input variables, number of training samples)

Question: How to label the single and concurrent faults?