# Individual Work Summary – Battery RUL Prediction Project

**Name: Tianhao Yang**

**Project Title:** Machine Learning for Lithium-Ion Battery Remaining Useful Life (RUL) Prediction

**Institution:** Lakehead University, Electrical and Computer Engineering(Course-based)

## 1. Personal Efforts and Contributions

Over the course of the project, I played a central role in building the full end-to-end machine learning pipeline for lithium-ion battery Remaining Useful Life (RUL) prediction. My work covered data preprocessing, model development, visualization, and technical documentation.  
  
I began by setting up the overall Python project architecture, including a clear modular structure (src/load\_data.py, src/visualize.py, src/train\_model.py, and main.py). I defined the workflow logic from raw data ingestion to final result visualization and ensured full reproducibility.  
  
In data handling, I developed functions to load and standardize .mat files from the NASA battery dataset (B0005, B0006, B0007, B0018). I implemented consistency checks, outlier filtering, and moving-average smoothing to reduce measurement noise. The final standardized dataset includes key fields such as cycle, capacity, SOH, and RUL, which serve as the foundation for model training.  
  
For visualization, I created several analytical plots—capacity vs. cycle, SOH vs. cycle, and end-of-life (EOL) cycle histograms—to help identify degradation patterns. These visual tools were essential for debugging data issues and for clearly demonstrating model outputs during evaluation.  
  
In the modeling phase, I implemented two baseline predictors: Linear Regression and Random Forest Regression. I designed the model training function with GroupKFold validation to prevent data leakage between different battery cells. Model evaluation metrics (MAE, RMSE, R²) were integrated for automatic reporting, and trained models were serialized with joblib for later use. I also explored early-life forecasting—training models using only the first N cycles—to test predictive capability at limited observation horizons.  
  
Beyond coding, I focused heavily on project documentation. I authored and formatted the Bilingual Translation Report and maintained detailed in-code comments and structure explanations to support future reproducibility.

## 2. Technical Skills Applied

• Python, NumPy, pandas, SciPy, scikit-learn, matplotlib  
• Data cleaning, transformation, and normalization  
• Machine learning regression, cross-validation, and metrics  
• Visualization and automation of output reporting  
• Scientific writing and bilingual documentation

## 3. Collaboration and Teamwork

Within the group, I contributed to integrating my data processing and visualization modules with teammates’ model experiments. I helped align code conventions, debug common runtime errors, and ensure consistency in dataset labeling and directory structures. My communication and coordination supported the smooth merging of code and the overall reproducibility of results.

## 4. Reflection and Next Steps

This project deepened my understanding of both battery degradation mechanisms and machine learning for prognostics. I learned to design modular, interpretable ML pipelines and to maintain scientific rigor in data-driven experiments.  
  
In future stages, I plan to extend this work by implementing advanced models (e.g., Gradient Boosting, CNN-LSTM hybrid architectures) and incorporating uncertainty quantification for more reliable RUL estimation.  
  
Overall, my contribution ensured the project’s technical foundation, code reliability, and documentation quality, enabling the team to produce a functional and academically sound RUL prediction framework.