VSB Power Line Fault Detection

Myunggun Kim, Sehee Moon, Minsu Lee, Ingon Chu

2026S IMEN574 Term Project

Introduction

Competition Overview

- **Competition** : VSB Power Line Fault Detection
  - **Host**: ENET Centre(Energy Research Center), VŠB – Technical University of Ostrava(Czech)
  - **Period**: Dec 2018 – Mar 2019
  - **Participants**: 7,600+ entrants / 1,400+ teams

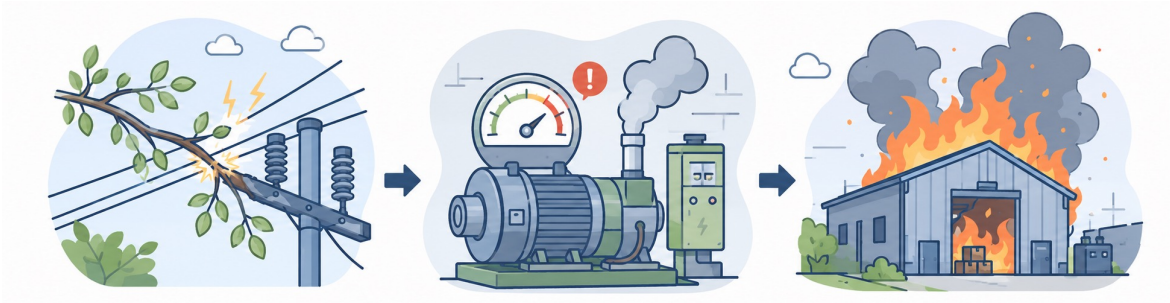
- **Background of the Competition**
  - Partial discharge (PD) is an early indicator of insulation failure in overhead power lines
  - PD often does not cause an immediate outage, so faults can progress undetected
  - Detecting PD early helps prevent line failures before they escalate
  - However, traditional inspection methods are costly and inefficient

➢ **Need for automated, signal-based fault detection through PD pattern analysis**

Motivation

- **Potential for application in various fields – military domain**
  - Power infrastructure of military bases is directly linked to operational sustainability
  - Applicable to monitoring of electric wires installed in extensive and difficult-to-access areas
  - Suitable for military domain -> significant cost and operational constraints
- **Research experience on unique 3-phase data structures**
  - The 3-phases share the same electric wires, making interdependent-channel time-series classification insufficient for capturing cross-phase information
  - This competition provided research experience in analyzing interdependent multichannel signals and physically coupled time-series data.

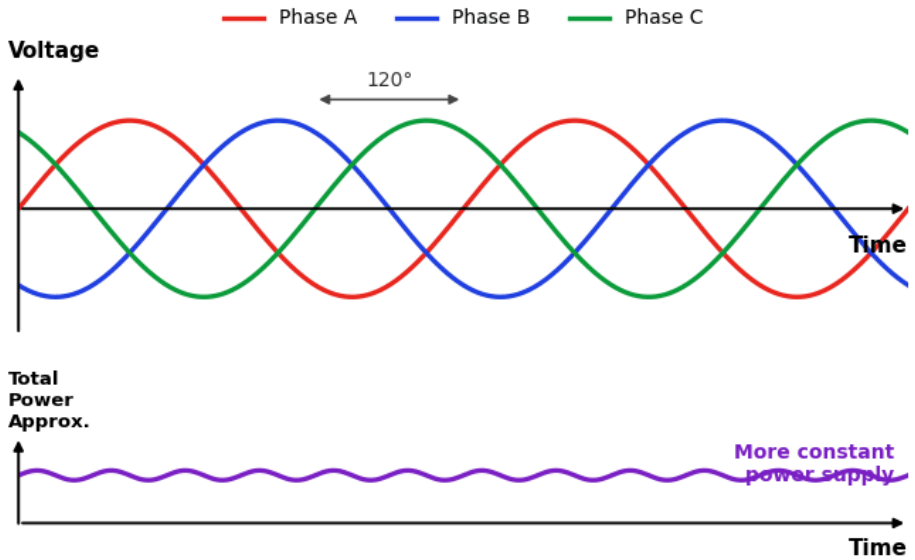




Data

Dataset Description

- The VSB dataset contains **high-frequency voltage signals** collected from a **three-phase** power system.
- Each measurement event consists of three simultaneously recorded phase signals, and each signal contains 800,000 measurements over 20 ms, corresponding to one full **50 Hz cycle**.
- The task is to classify each signal as **normal or faulty** based on **partial discharge patterns**.



Three-Phase Voltages and Composite Power

Why three-phase signals?

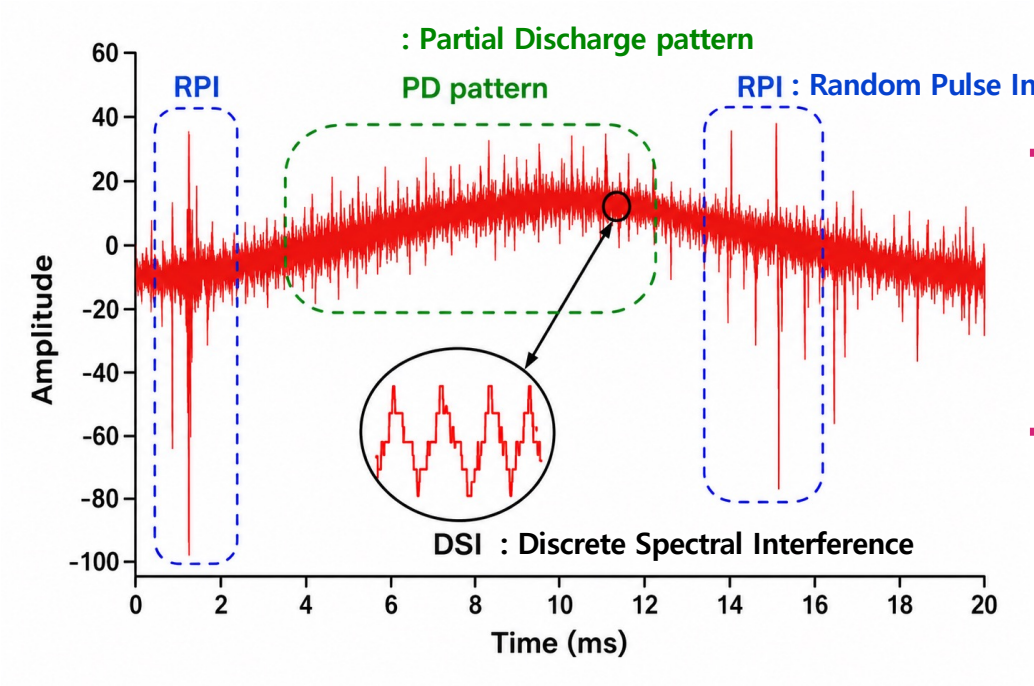
Balanced power delivery
Phase-shifted signals produce smoother total power.

More reliable power delivery
The three phases complement each other, reducing power fluctuations.

Real power-line structure
Transmission systems commonly operate with three-phase AC power.

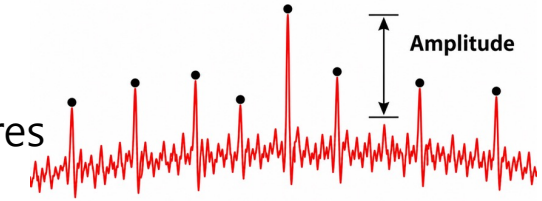
Data Preprocessing

- Raw voltage signals contain various types of **noise** and **electrical interference**.
- To detect line faults, we need to identify **PD-related abnormal impulse patterns** hidden in the signal.
- To highlight these patterns, we applied **data preprocessing** approaches.



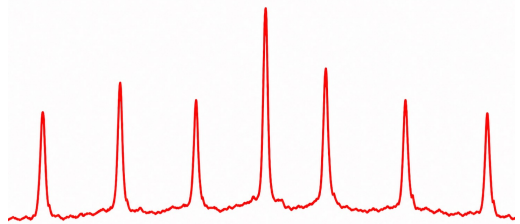
Feature Extraction

- Extract peak / statistical features
- Represent abnormal patterns



Noise Reduction

- Suppress noise & false peaks
- Preserve PD-related impulses



Methodology

Approach 1: Methods & Results

❑ Insights from Related Work and Feature Extraction

- Related work: A complex classification approach of partial discharges from covered conductors in real environment (2017, IEEE)

Feature Selection

Using the 7 features with high MI presented in the paper

- Max/Min/Mean width
- Max/Min/Mean height
- Number of peaks

3-phases Structure

80% of defects are defective in all three phases

If one phase is defective, the probability that the other phases are **also defective is 92.8%**

Added a feature

- Group each of the 7 features by 3 phases and add **Max / Min / Mean / Std** as new features
- Compute the **relative value** of each original feature against those statistics, and add as new features

→ **42 features in total**

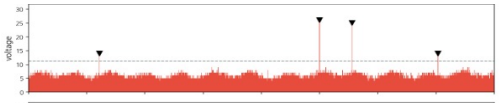
Before (7 feature)

0.5806

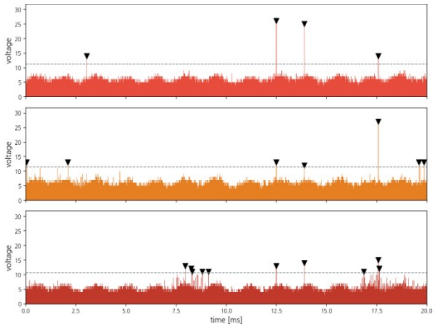
After (42 feature)

0.6545

Key insight
Adds three-phase statistics and relative values to a model that previously viewed only one phase independently



Number of peaks	4
Max height	0.61
Min height	0.30
Mean height	0.41
.....	.....





Max

min

mean

Std

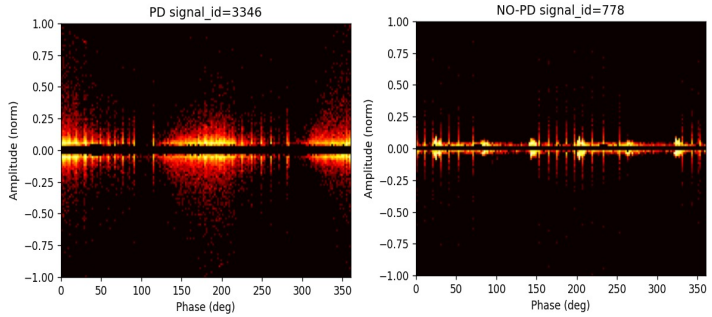
Relative

Number of peaks	Max0.31 height	.....
13	0.61	.....
4	0.31	.....
7	0.46	.....
5.2	0.15	.....
0.74	0.32	.....

Approach 2: Methods & Results

❑ Deep Learning with PD pattern image

- **PRPD : 2D plot that maps PD events onto the phase angle**
  - x-axis : phase angle / y-axis : amplitude / color : number of events
- **PD pattern recognition via CNN model**



1. Data preprocessing

Denoising : Filter cutoff & Pulse threshold

- Filter cutoff → 1kHz high-pass
- Pulse threshold : 1.0σ / 2.0σ / 3.0σ

2. Model training

Training with PRPD image

- Backbone : ResNet18 + imageNet
- Optimizer : AdamW
- Scheduler : Cosine-Annealing

3. Measuring

GroupKfold validation (K = 5)

OOF prediction : Validate each fold with val. Datasets (20%)

4. Experiment

Adapting different variables

- Changing aggregations (**mean**/max/median)
- Threshold tuning **best t : 0.39**

❑ Submission to Kaggle

- About 47th ..? **Not enough!**

Final submission score

Score: 0.66492

Public score: 0.60440

46	▲ 949	OlegPy		0.66523
47	▲ 929	One Old Dog		0.66462
48	▼ 47	MA		0.66422

Agentic Research Workflow

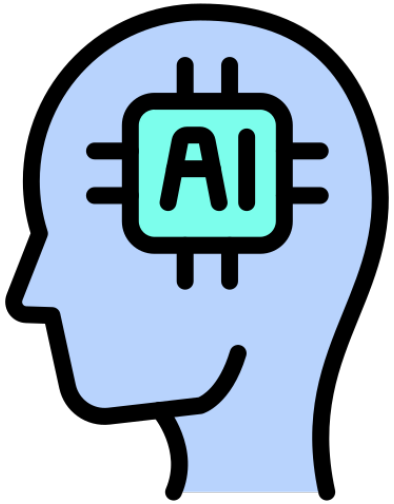
Toward an Agentic Research Workflow

- Conventional Kaggle Workflow
  1. Exploratory Data Analysis
  2. Paper & Literature Review
  3. Data Preprocessing
  4. Modeling
  5. Performance Evaluation & Insight Generation
  6. Hypothesis Refinement

Deep research
perplexity
CURSOR
CLAUDE CODE
ChatGPT
Claude

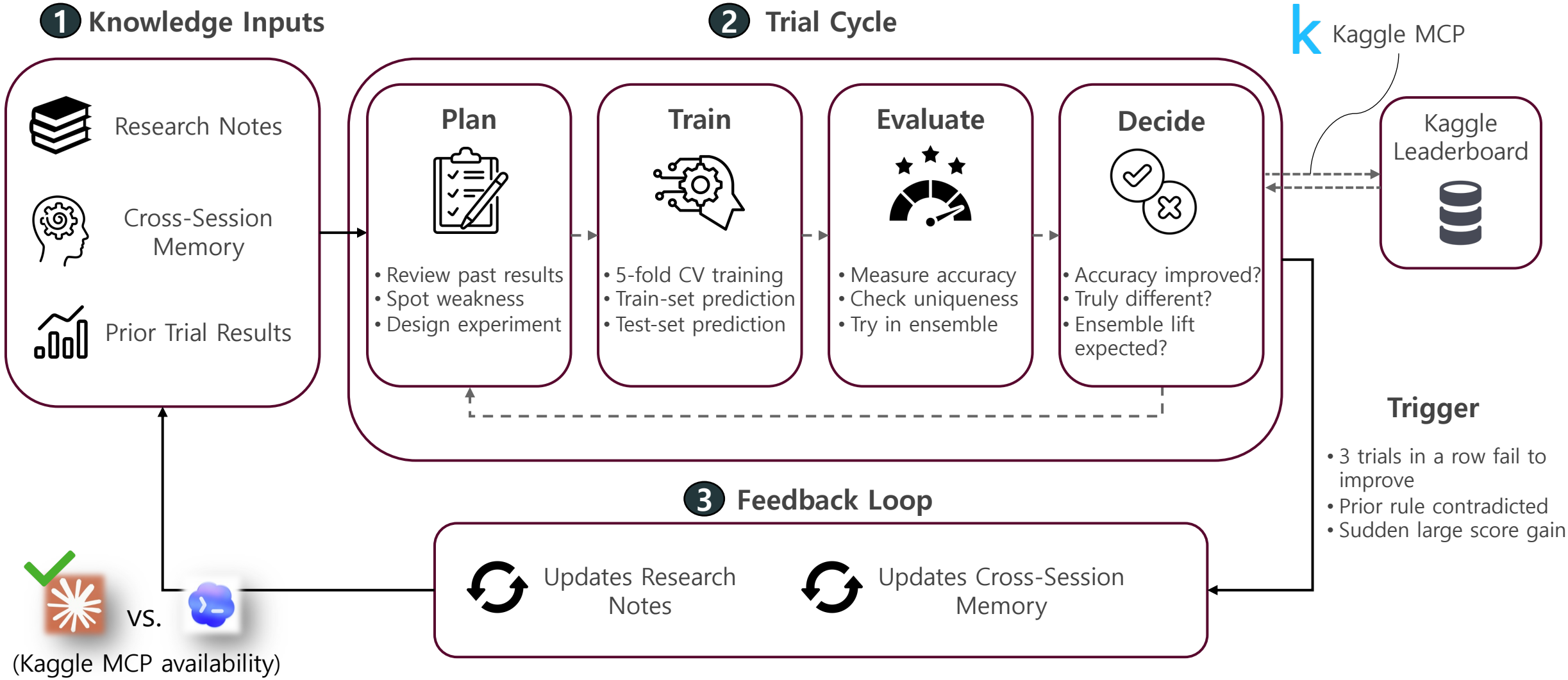


Human-Driven (AI-powered)



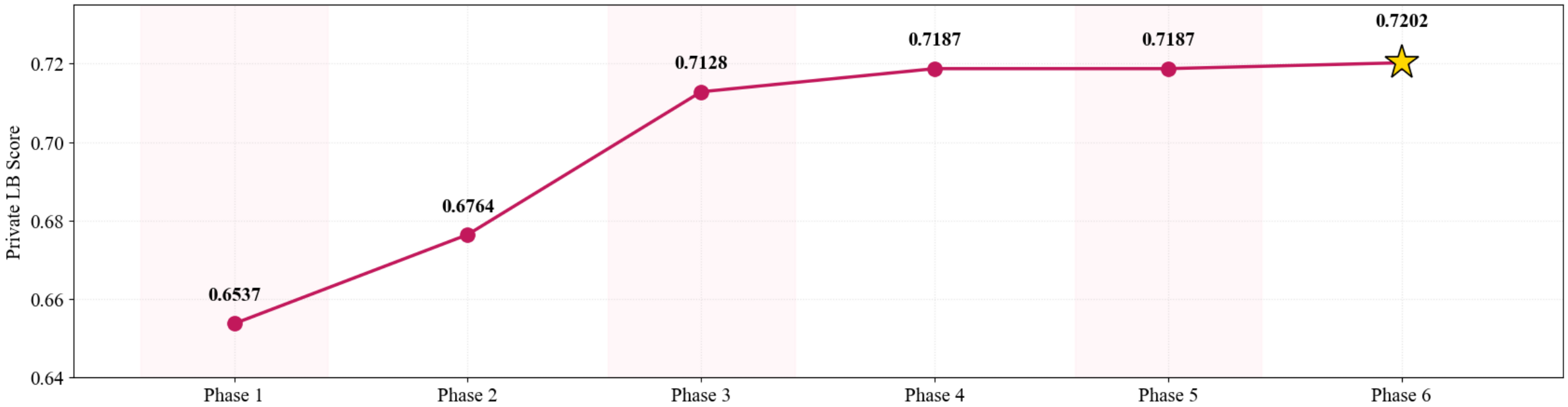
AI Agent-Driven

Autonomous Kaggle Research Agent



Trial Journey on Private Score Progression

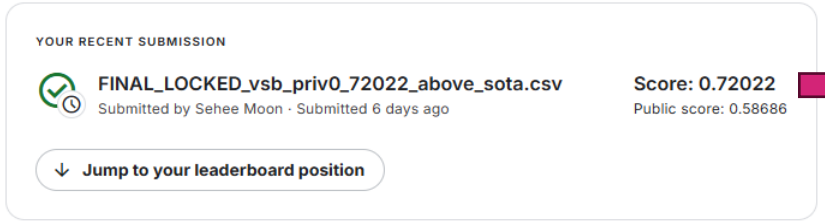
- 6 phases of trials climbed the Private Score from 0.6537 to 0.7202
- Each phase resolved different limitations
  - similar models → same input → same approach → missing domain insight



	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6
Approach:	**Single model baseline (BiLSTM)**	**Diverse architectures (Transformer variants)**	**Add new model type (TimesNet)**	**New data representations (PulseTransformer)**	**Refine existing approaches (Joint fusion)**	**3-phase post-process rule (no new model)**
Key Insight:	*Retraining the same model produces nearly identical answers*	*Many architectures, same data: all plateau around 0.68*	*Need different kinds of input,not just more model variations*	*Tweaks to new models stay too close to existing ones*	*A single combined model overfits the training data*	*In 90% of cases, faults appear in all 3 phases together*

Final Result

- Private Leaderboard Score
  - Late Submission, 1446 Teams



Public Private

The private leaderboard is calculated with approximately 43% of the test data.
This competition has completed. This leaderboard reflects the final standings.

Prize Winners

#	△	Team	Members	Score	Entries	Last	Solution
1	▲ 15	mark4h		0.71899	31	7y	
2	▲ 36	HELLO TANG		0.71501	16	7y	
3	▲ 657	g6gg6g		0.70488	32	7y	

- Beyond the Leaderboard

AI Agent	Manual Approach
✓ Highest Score (0.72022)	✓ Interpretable
✓ 1st on leaderboard	✓ Single lightweight model
✗ 7 models, black-box	✓ Domain-grounded
✗ Hard to interpret/deploy	✗ Lower score

Thank You

Myunggun Kim, Sehee Moon, Minsu Lee, Ingon Chu

2026S IMEN574 Term Project