Model Comparison Report: PTB-XL Dataset Analysis

Analysis Team

May 2, 2025

1 Executive Summary

This report comprehensively compares different models trained on the PTB-XL dataset, analyzing their performance across three classification types: Binary, Superclasses, and Subclasses. The analysis covers both the original PTB-XL dataset and a modified version with simulated signal gaps, along with performance comparisons between standard and optimized models using pruning and quantization techniques.

2 Dataset Overview

- **Original PTB-XL**: Standard dataset without modifications
- **Modified PTB-XL**: Same dataset with simulated signal gaps:
  - 0.5-second zero-value gaps periodically
  - In average 10 gaps per 10-second signal (5% missing data)
  - Simulates real-world signal loss scenarios

3 Model Overview

- **Standard Model**: Baseline CNN model
- **Optimized Model**: Same architecture with:
  - Model pruning (64.3% size reduction)
  - Weight quantization (76% size reduction)
  - Final size: 600KB (8.6% of original 7MB)

4 Overall Performance Comparison

Table 1 presents the comparison of key performance metrics across all models and classification types.

Table 1: Overall Model Performance Comparison

Dataset	Classification	Accuracy	Precision	Recall	F1 Score	AUC
Original PTB-XL (Standard)	Binary	0.8708	0.8783	0.8712	0.8717	0.9460
	Superclasses	0.8914	0.7969	0.7686	0.7824	0.9258
	Subclasses	0.9650	0.7555	0.6170	0.6794	0.9164
Original PTB-XL (Optimized)	Binary	0.8826	0.8905	0.8854	0.8846	0.9482
	Superclasses	0.8734	0.7672	0.7185	0.7327	0.9186
	Subclasses	0.9636	0.6836	0.5372	0.5622	0.9012
Modified PTB-XL (Standard)	Binary	0.8508	0.8565	0.8567	0.8550	0.9322
	Superclasses	0.8821	0.7914	0.7293	0.7589	0.9211
	Subclasses	0.9645	0.7662	0.6170	0.6794	0.9164
Modified PTB-XL (Optimized)	Binary	0.8512	0.8572	0.8535	0.8534	0.9329
	Superclasses	0.8795	0.7815	0.7299	0.7548	0.9203
	Subclasses	0.9637	0.8001	0.5290	0.6368	0.9164

5 Performance Analysis

5.1 Original PTB-XL Dataset

- **Binary Classification**:
  - Improvements: Accuracy (+1.18%), Precision (+1.22%), Recall (+1.42%), F1 (+1.29%), AUC (+0.22%)
  - Suggests effective noise removal through optimization
- **Superclasses Classification**:
  - Degradation: Accuracy (-1.80%), Precision (-2.97%), Recall (-5.01%), F1 (-4.97%), AUC (-0.72%)
  - Indicates potential feature loss during optimization
- **Subclasses Classification**:
  - Degradation: Accuracy (-0.14%), Precision (-7.19%), Recall (-7.98%), F1 (-11.72%), AUC (-1.52%)
  - Complex patterns affected by optimization

5.2 Modified PTB-XL Dataset

- **Binary Classification**:
  - Minimal changes: Accuracy (+0.04%), Precision (+0.07%), Recall (-0.32%), F1 (-0.16%), AUC (+0.07%)
  - Robust to both optimization and signal gaps
- **Superclasses Classification**:
  - Slight changes: Accuracy (-0.26%), Precision (-0.99%), Recall (+0.06%), F1 (-0.41%), AUC (-0.08%)

- Better tolerance to optimization with modified data

- **Subclasses Classification**:
  - Mixed results: Accuracy (+0.02%), Precision (+3.39%), Recall (-8.80%), F1 (-4.26%), AUC (no change)
  - Trade-off between precision and recall

6 Detailed Class-wise Analysis

6.1 Binary Classification

Table 2: Binary Classification Class-wise Performance

Class	Metric	Original	Modified	Change	Status
Class 0	Precision	0.9303	0.8979	-0.0324	↓
	Recall	0.8324	0.8332	+0.0008	↑
	F1	0.8786	0.8643	-0.0143	↓
	AUC	0.9460	0.9322	-0.0138	↓
Class 1	Precision	0.8115	0.8034	-0.0081	↓
	Recall	0.9211	0.8868	-0.0343	↓
	F1	0.8628	0.8430	-0.0198	↓
	AUC	0.9460	0.9329	-0.0131	↓

6.2 Superclasses Classification

Table 3: Superclasses Classification Class-wise Performance

Class	Metric	Original	Modified	Change	Status
Class 0	Precision	0.8409	0.7553	-0.0856	↓
	Recall	0.6714	0.6976	+0.0262	↑
	F1	0.7466	0.7254	-0.0212	↓
	AUC	0.9218	0.9011	-0.0207	↓
Class 1	Precision	0.6157	0.7609	+0.1452	↑
	Recall	0.5687	0.4008	-0.1679	↓
	F1	0.5913	0.5250	-0.0663	↓
	AUC	0.8987	0.8768	-0.0219	↓
Class 2	Precision	0.7996	0.8129	+0.0133	↑
	Recall	0.6964	0.6164	-0.0800	↓
	F1	0.7444	0.7011	-0.0433	↓
	AUC	0.9227	0.9113	-0.0114	↓
Class 3	Precision	0.8150	0.8200	+0.0050	↑

Table 3 – continued from previous page

Class	Metric	Original	Modified	Change	Status
	Recall	0.9242	0.8795	-0.0447	↓
	F1	0.8662	0.8487	-0.0175	↓
	AUC	0.9490	0.9383	-0.0107	↓
Class 4	Precision	0.8079	0.7571	-0.0508	↓
	Recall	0.7505	0.7658	+0.0153	↑
	F1	0.7781	0.7614	-0.0167	↓
	AUC	0.9325	0.9264	-0.0061	↓

6.3 Subclasses Classification

Table 4: Subclasses Classification Class-wise Performance

Class	Metric	Original	Modified	Change	Status
Class 1	Precision	0.8188	0.7723	-0.0465	↓
	Recall	0.7386	0.7647	+0.0261	↑
	F1	0.7766	0.7685	-0.0081	↓
	AUC	0.9675	0.9591	-0.0084	↓
Class 2	Precision	0.7937	0.8542	+0.0605	↑
	Recall	0.9259	0.7593	-0.1666	↓
	F1	0.8547	0.8039	-0.0508	↓
	AUC	0.9968	0.9944	-0.0024	↓
Class 3	Precision	0.7619	0.8462	+0.0843	↑
	Recall	0.8889	0.8148	-0.0741	↓
	F1	0.8205	0.8302	+0.0097	↑
	AUC	0.9973	0.9978	+0.0005	↑
Class 5	Precision	0.6571	0.7191	+0.0620	↑
	Recall	0.7034	0.6575	-0.0459	↓
	F1	0.6795	0.6869	+0.0074	↑
	AUC	0.9285	0.9120	-0.0165	↓
Class 6	Precision	0.6031	0.7284	+0.1253	↑
	Recall	0.7054	0.5268	-0.1786	↓
	F1	0.6502	0.6114	-0.0388	↓
	AUC	0.9744	0.9741	-0.0003	↓
Class 10	Precision	0.7238	0.7119	-0.0119	↓
	Recall	0.7318	0.7039	-0.0279	↓
	F1	0.7278	0.7079	-0.0199	↓
	AUC	0.9718	0.9702	-0.0016	↓
Class 14	Precision	0.7444	0.6848	-0.0596	↓
	Recall	0.4626	0.5888	+0.1262	↑
	F1	0.5706	0.6332	+0.0626	↑

Table 4 – continued from previous page

Class	Metric	Original	Modified	Change	Status
	AUC	0.9307	0.9201	-0.0106	↓
Class 15	Precision	0.8101	0.8579	+0.0478	↑
	Recall	0.9304	0.8214	-0.1090	↓
	F1	0.8661	0.8393	-0.0268	↓
	AUC	0.9497	0.9358	-0.0139	↓
Class 20	Precision	0.5846	0.5508	-0.0338	↓
	Recall	0.1712	0.2928	+0.1216	↑
	F1	0.2648	0.3824	+0.1176	↑
	AUC	0.8797	0.8729	-0.0068	↓

7 Key Findings

- **Model Size**:
  - 91.4% total reduction (7MB → 600KB)
  - Two-phase optimization: pruning (64.3%) and quantization (76%)
- **Performance Patterns**:
  - Binary Classification: Consistent high performance ($AUC > 0.93$)
  - Superclasses Classification: Moderate impact ($AUC > 0.91$)
  - Subclasses Classification: Significant trade-offs ($AUC > 0.90$)
- **Dataset Impact**:
  - Modified dataset shows more consistent performance
  - 5% missing data has minimal impact on AUC
  - Optimization effects vary by classification type

8 Recommendations

1. Use an optimized model for binary classification
2. Evaluate trade-offs for superclasses and subclasses
3. Consider the standard model for superclasses with original data
4. Future work should focus on:
   - Improving recall in complex classifications
   - Maintaining binary classification gains
   - Reducing performance gaps between datasets
   - Exploring adaptive optimization strategies

9 Conclusion

The optimization process achieves significant model size reduction (91.4%) while maintaining strong performance in binary classification. The modified dataset shows consistent performance across models, with minimal impact on AUC values. The choice between standard and optimized models should consider the specific classification requirements and the relative importance of precision versus recall. The high AUC scores (> 0.93) across all configurations indicate robust discrimination ability, making both models suitable for ECG classification tasks.