

## ABSTRACT

This research proposes a lightweight misbehavior detection system for communication-based train control using ensemble learning models. The study evaluates Bagging-based methods, including Random Forest and k-Nearest Neighbors with Bagging, alongside Boosting-based approaches such as AdaBoost, XGBoost, and LightGBM. The models were tested on the CBTCSet dataset, addressing data imbalance and assessing performance based on accuracy, precision, recall, F1-score, testing time, and fit stability to meet real-time CBTC requirements.

The results indicate that Random Forest with the Weighted Imbalance method provides the best balance between detection performance and computational efficiency, achieving 92% accuracy, 92% precision, 92% recall, and a 92% F1-score. The total testing time for 15% of the dataset, consisting of 173,843 data entries, was 12.21 seconds, resulting in an average processing time of 70.23  $\mu$ s per entry. While other models demonstrated specific advantages, some suffered from overfitting, underfitting, or excessive processing time, limiting their feasibility for real-time deployment.

These findings confirm that Bagging-based models, particularly Random Forest, offer the most effective trade-off between detection accuracy and computational feasibility, making them the most viable choice for real-time CBTC operations to enhance safety and system resilience.

**Keywords:** Intelligent Transportation Systems, Urban Railway Systems, Communication-Based Train Control, Ensemble Learning, Supervised Learning, Misbehavior Detection System, Cybersecurity, Processing Time Optimization.