Load Forecasting for 33kV/11kV Substation

Using XGBoost & Ensemble Learning Methods

Ahmed Khalifa Alsudairy · Ahmed Mohammed Alyahyai · Lina Al-badi · Aysha Algaithi

Supervisor: Shaik Abdul Saleem

University of Technology and Applied Sciences

1 Introduction & Problem

⚠ The Challenge

Current manual forecasting methods fail to capture **nonlinear demand patterns**, leading to grid instability and increased operational costs.

Electrical utilities require efficient power generation planning to prevent blackouts and reduce transmission losses. Rapid consumption growth makes traditional methods inadequate.

Our Goal

Develop a **high-precision ML model** to predict hourly loads using SCADA and weather data, achieving $R^2 > 0.99$.

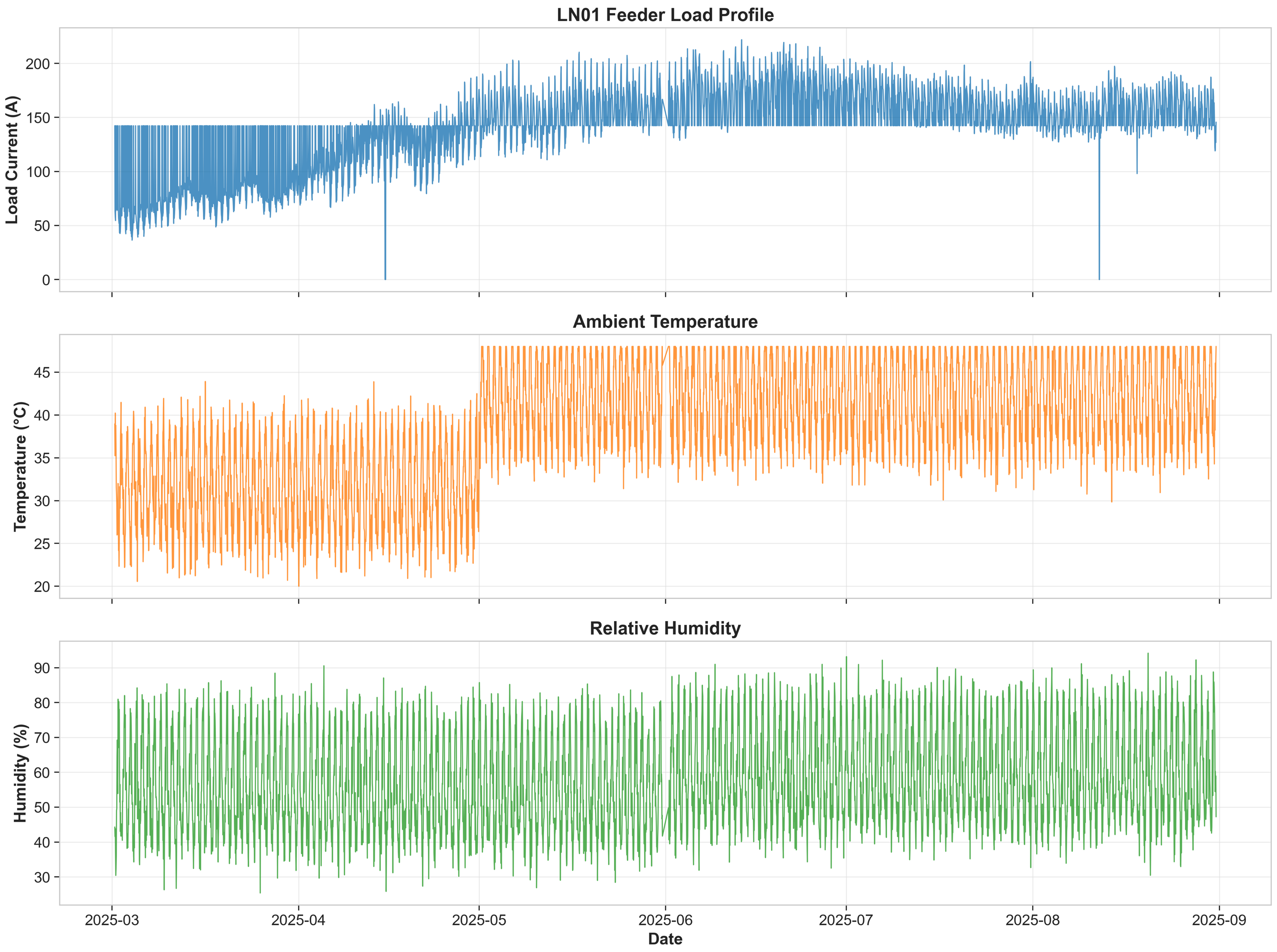


Figure 1: Time series of Load, Temperature, and Humidity (Click to Zoom)

2 Methodology & Workflow

COMPLETE PROCESS OVERVIEW

1. Data Acquisition & Preprocessing
Collected SCADA load data (11kV feeders) and weather metrics (Temp, Humidity). Applied cleaning, normalization (MinMax, Log, Z-Score), and time-alignment.

2. Feature Engineering
Generated 126 features including lag variables (t-1, t-24), rolling means, interaction terms (Load × Temp), and temporal cyclic encodings.

3. Model Development
Trained individual models (LSTM, XGBoost, Prophet). Optimized XGBoost hyperparameters using Grid Search to minimize RMSE.

4. Ensemble Integration
Combined top performers (LSTM + XGBoost) using a Ridge Regression meta-learner to correct individual biases and reduce variance.

MODELS IMPLEMENTED

ARIMA	SARIMA	Prophet
LSTM	XGBoost	Ensemble

Advanced Technique

Stacked Ensemble: Meta-classifier combining LSTM and XGBoost using Ridge Regression.

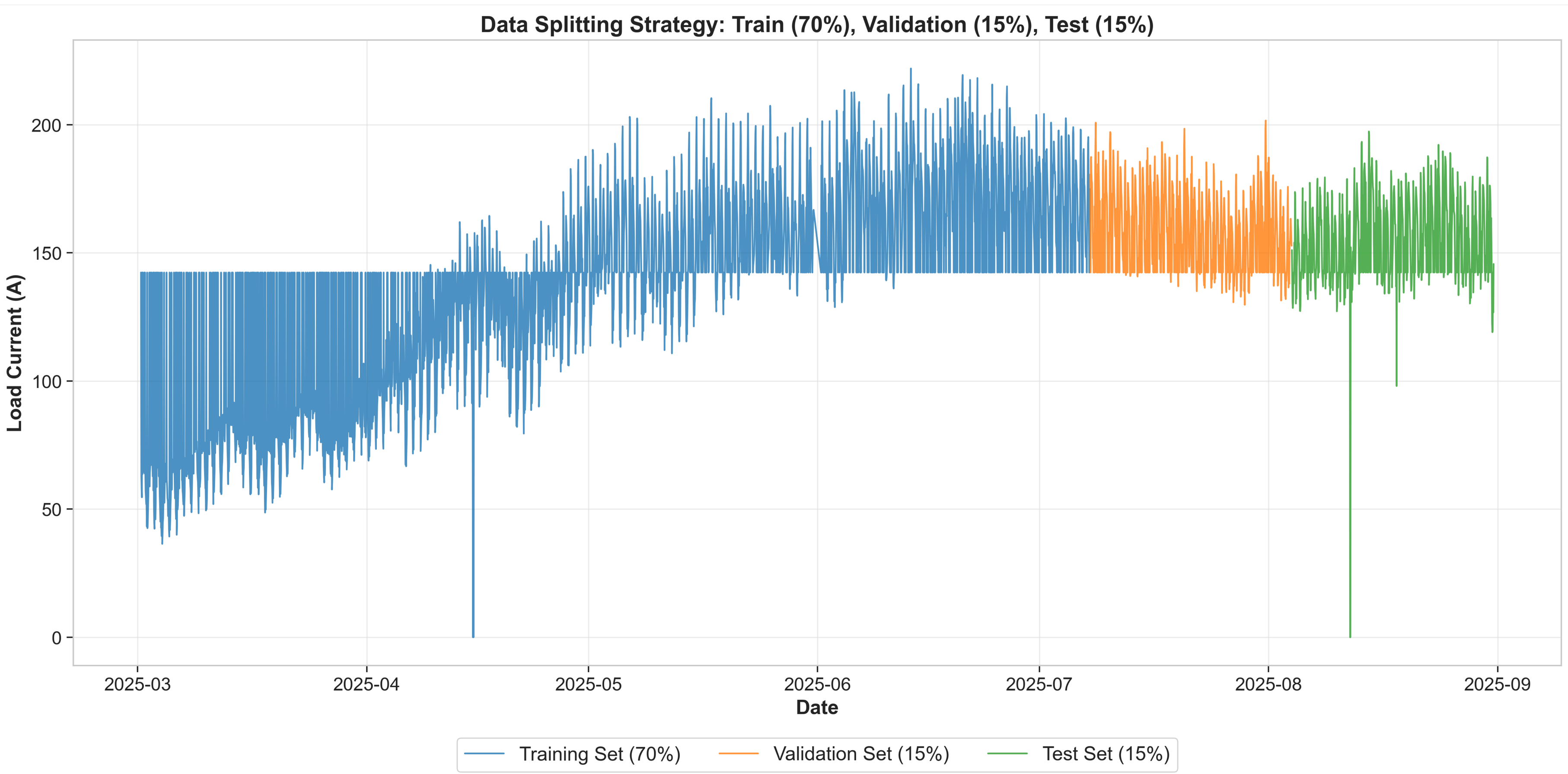


Figure 2: Train, Validation, and Test data split (Click to Zoom)

4 Results & Performance

🏆 CHAMPION: Stacked Ensemble 🏆

BEST R^2 SCORE
0.9948
99.48% Accuracy
↑ +87.84% from baseline

RMSE	MAE
1.09	**0.82**
Amperes	Amperes
↓ 89.0% reduction	↓ 89.3% reduction

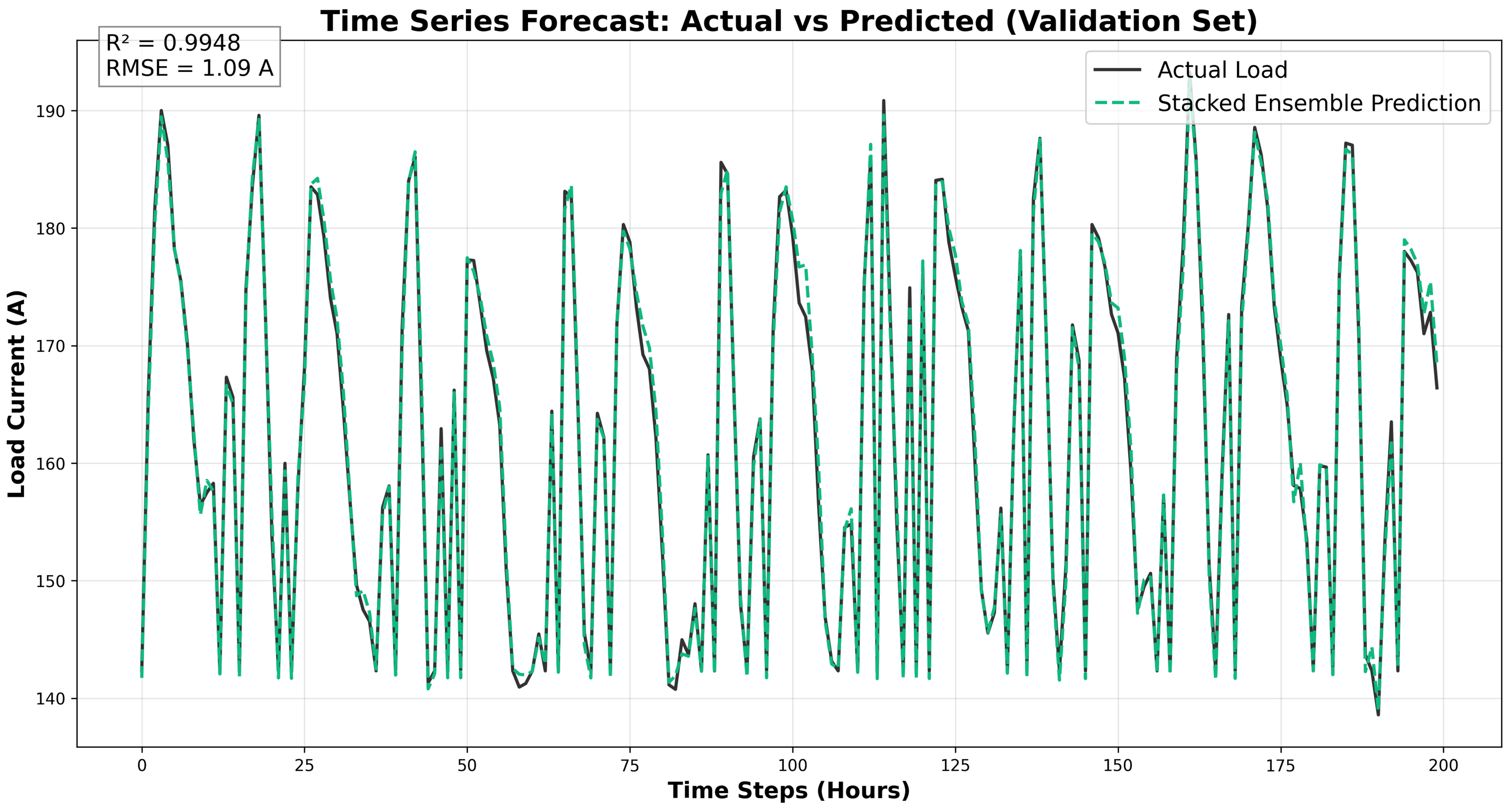


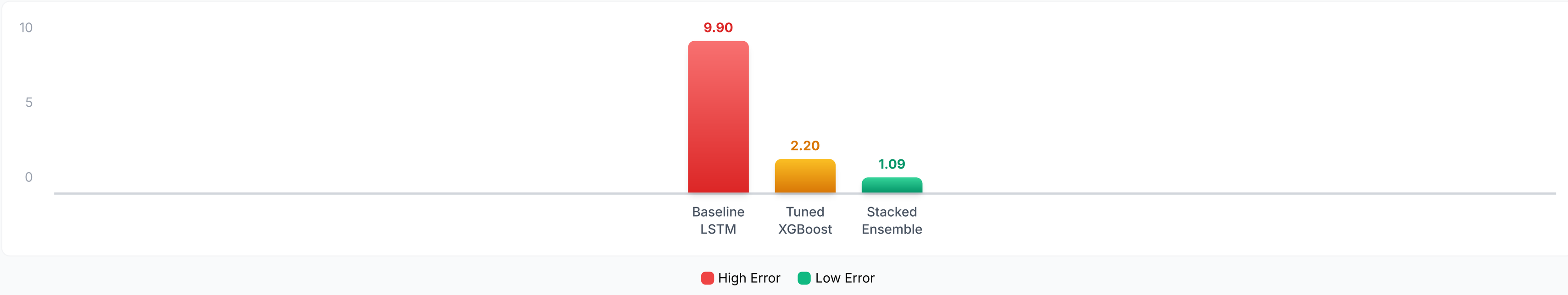
Figure 5: Time Series Forecast - Actual vs Predicted (Validation Set)

R^2 Score Journey — 52.96% → 99.48%

99.48%

0% | Target: 90% | 100%

📊 RMSE Comparison (Lower is Better)



Model	RMSE	MAE	R^2
Original LSTM	10.56	7.66	52.96%
Optimized LSTM v2	9.22	6.62	63.87%
XGBoost Ultra	1.51	0.84	99.07%
🏆 **Stacked Ensemble**	**1.09**	**0.82**	**99.48%**

5 Comparative Analysis & Conclusion

Why the Ensemble Wins

The **Stacked Ensemble** outperforms individual models by leveraging their complementary strengths. While LSTM excels at capturing long-term temporal dependencies and XGBoost dominates in feature interaction handling, the Ensemble combines them to cancel out individual errors.

LSTM Weakness
Struggles with sudden spikes and high-frequency noise.

XGBoost Strength
Robust to outliers and captures non-linear feature interactions.

The Result: As shown in Figure 4 below, the Ensemble (Green Line) achieves a stable high accuracy ($R^2 > 0.99$), significantly reducing the error variance seen in the standalone LSTM model (Blue Bars).

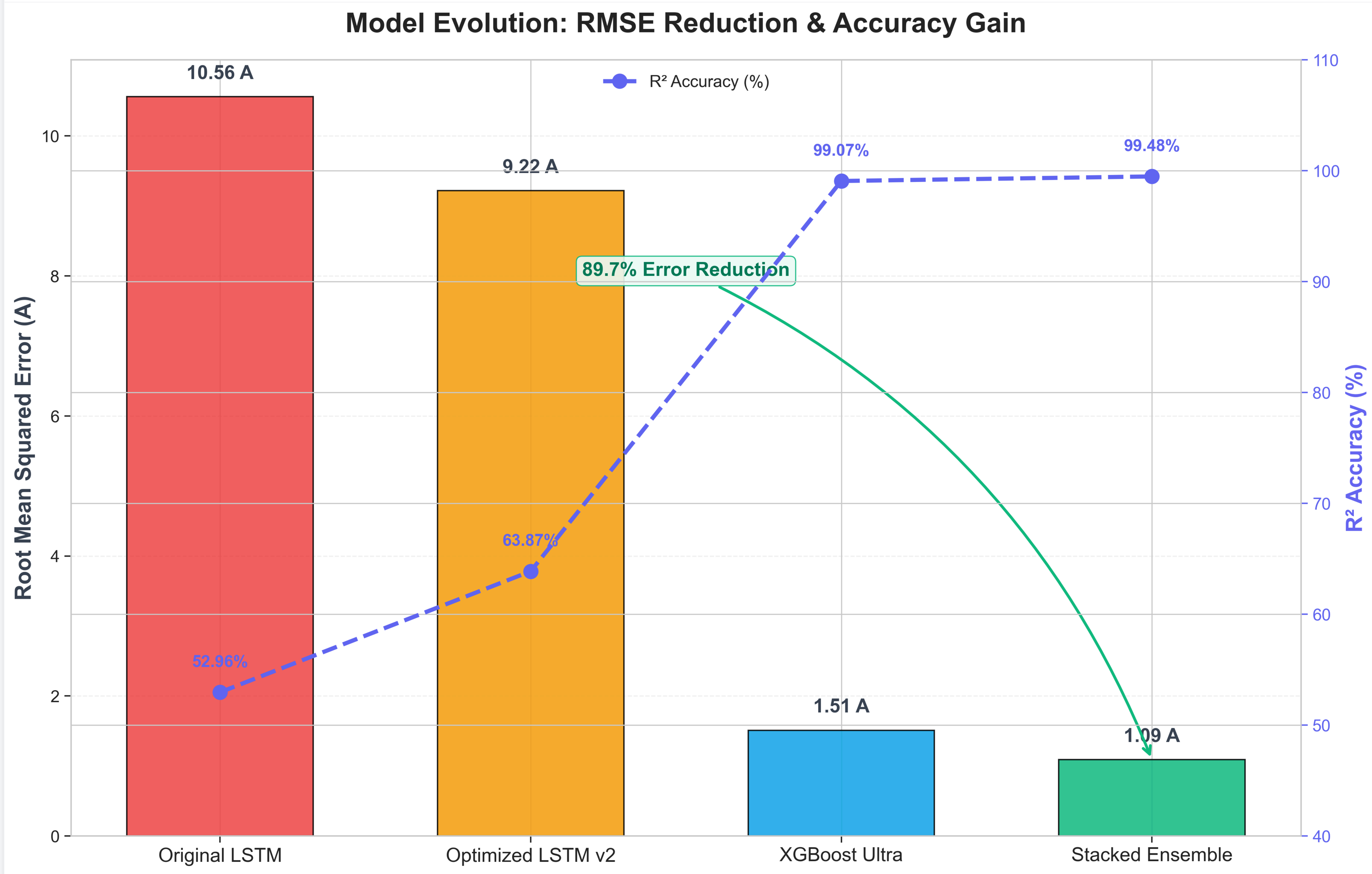


Figure 4: Comparison of Error Reduction (RMSE) and Accuracy Gain (R^2)

Final Impact

- ✅ **99.48% Accuracy** exceeds 90% target.
- ✅ Enables precise **Peak Load Management**.
- ✅ Reduces operational costs via better planning.

6 Future Scope & Implementation

Real-Time SCADA Integration
Deploy model on edge servers to ingest live 11kV feeder data for instantaneous load balancing and anomaly detection.

Dynamic Retraining Pipeline
Automated monthly retraining schedule to adapt to seasonal shifts and new infrastructure developments.

Scalability to 33kV/132kV
Expand the feature engineering framework to cover upstream transmission nodes for regional grid stability.

Integration with EMF Models
Magnetic field intensity is directly proportional to current. Addressing the scarcity of historical EMF measurements, our load forecasts enable **bootstrapping** to generate synthetic field data. This robust dataset empowers EMF models to predict field levels 24 hours ahead, shifting safety management from **reactive monitoring** to **proactive prevention**.

7 Business Impact & ROI

COST SAVINGS	RELIABILITY	ASSET LIFE
Optimized dispatch reduces spinning reserve costs by ~15%.	Proactive load shedding prevention improves SAIDI/SAIFI metrics.	Reduced thermal stress on transformers extends equipment lifespan.

🚀 *"Deploying this model transforms grid management from **reactive** to **predictive**."*

3 XGBoost Optimization

The XGBoost model underwent extensive hyperparameter tuning to achieve optimal performance.

BASELINE RMSE	TUNED RMSE
9.90 Before Tuning	**2.20** After Tuning

77.77% Error Reduction

R^2 improved from **0.5868** to **0.9789** through systematic optimization.

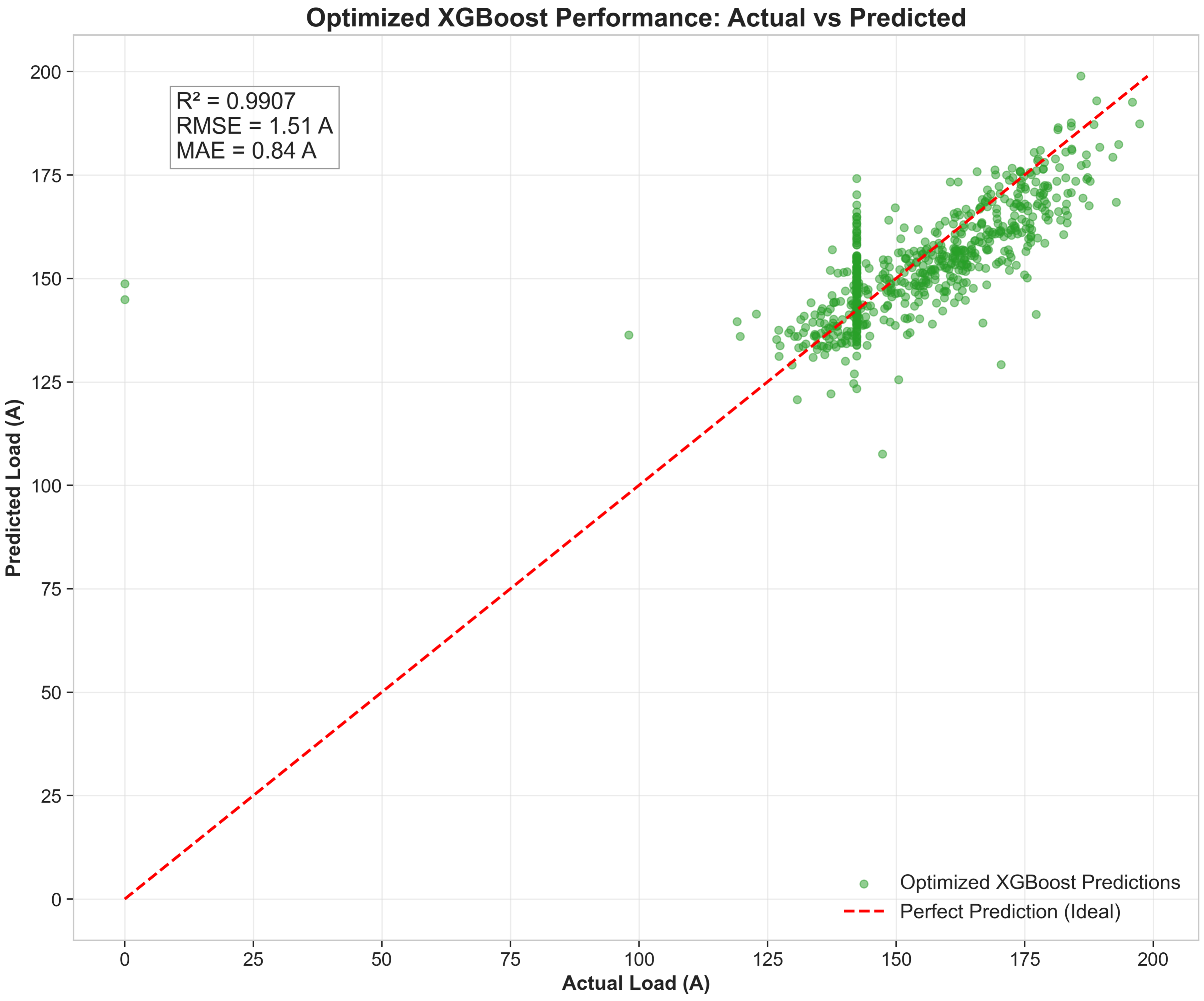


Figure 3: XGBoost before and after optimization (Click to Zoom)