Peak-Aware LSTM-CNN for Milan Internet Traffic Prediction

-This document is not used for poster, only demonstrating progress

Model Description: Peak-Aware LSTM-CNN

The proposed model is a **Peak-Aware LSTM-CNN** designed to predict urban internet traffic in Milan with fine-grained temporal and spatial resolution. It integrates convolutional, recurrent, and attention mechanisms with a custom loss to emphasize high-traffic periods.

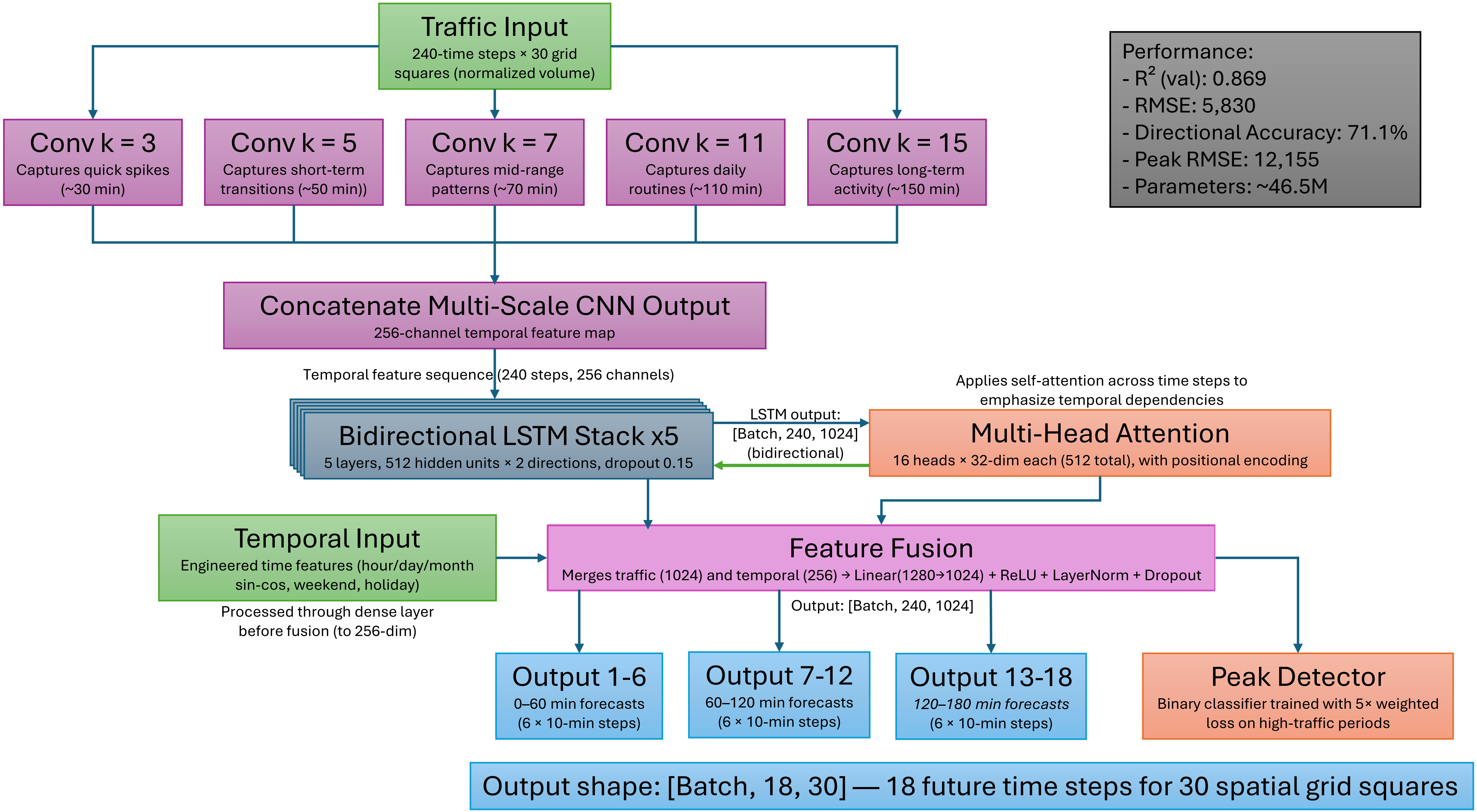
The architecture processes two input streams: a **traffic volume sequence** ([Batch, 240, 30]) capturing 40 hours of historical data across 30 city grid squares, and a **temporal feature sequence** ([Batch, 240, 10]) encoding time-of-day, day-of-week, and calendar context.

A **multi-scale CNN module** with five parallel 1D convolution layers (kernel sizes 3 to 15) captures patterns ranging from quick spikes to extended daily cycles. These outputs are concatenated and fed into a **5-layer bidirectional LSTM stack** (512 hidden units per direction) to learn long-range temporal dependencies.

To focus on the most relevant time steps, the LSTM output is passed through a **multi-head self-attention mechanism** (16 heads × 32 dimensions). Simultaneously, the temporal input is projected into a 256-dimensional feature space via a dense layer. The attention and temporal features are **fused** through a linear layer, normalization, and dropout to form a unified representation.

A parallel **peak detector** branch classifies whether each time step represents a high-traffic period, with a peak-aware loss function (5× weighting on peaks) used during training to improve critical predictions.

Finally, the fused features are processed by **18 output heads**, grouped into three 1-hour segments, each predicting traffic volume for all 30 grid squares in 10-minute increments. The model outputs [Batch, 18, 30] shaped predictions with strong performance on metrics like **R^2 = 0.869** and **71.1% directional accuracy**.


Traffic Input
240-time steps × 30 grid squares (normalized volume)
Conv k = 3
Captures quick spikes (~30 min)
Conv k = 5
Captures short-term transitions (~50 min))
Conv k = 7
Captures mid-range patterns (~70 min)
Conv k = 11
Captures daily routines (~110 min)
Conv k = 15
Captures long-term activity (~150 min)
Performance:
- R^2 (val): 0.869
- RMSE: 5,830
- Directional Accuracy: 71.1%
- Peak RMSE: 12,155
- Parameters: ~46.5M
Concatenate Multi-Scale CNN Output
256-channel temporal feature map
Temporal feature sequence (240 steps, 256 channels)
Applies self-attention across time steps to emphasize temporal dependencies
Bidirectional LSTM Stack x5
5 layers, 512 hidden units × 2 directions, dropout 0.15
LSTM output: [Batch, 240, 1024] (bidirectional)
Multi-Head Attention
16 heads × 32-dim each (512 total), with positional encoding
Temporal Input
Engineered time features (hour/day/month sin-cos, weekend, holiday)
Processed through dense layer before fusion (to 256-dim)
Feature Fusion
Merges traffic (1024) and temporal (256) → Linear(1280→1024) + ReLU + LayerNorm + Dropout
Output: [Batch, 240, 1024]
Output 1-6
0–60 min forecasts (6 × 10-min steps)
Output 7-12
60–120 min forecasts (6 × 10-min steps)
Output 13-18
120–180 min forecasts (6 × 10-min steps)
Peak Detector
Binary classifier trained with 5× weighted loss on high-traffic periods
Output shape: [Batch, 18, 30] — 18 future time steps for 30 spatial grid squares

Parameters:

Core Architecture		
Parameter	Value	Description
MODEL_TYPE	enhanced_lstm	Enhanced LSTM-based architecture
HIDDEN_DIM	512	Hidden size of the LSTM (bidirectional = 1024)
NUM_LAYERS	5	Number of stacked LSTM layers
DROPOUT	0.15	Dropout rate between layers
BIDIRECTIONAL	TRUE	Uses bidirectional LSTM
PREDICTION_HORIZON	18	Predicts 3 hours ahead (18 × 10-minute steps)
SEQUENCE_LENGTH	240	Input sequence length (40 hours)

Attention Config		
Parameter	Value	Description
USE_ATTENTION	TRUE	Enables attention mechanism
NUM_HEADS	16	Number of attention heads
ATTENTION_DIM	512	Attention output dimensionality
ATTENTION_LAYERS	2	Number of attention layers

CNN Config		
Parameter	Value	Description
USE_MULTI_SCALE	TRUE	Enables multi-scale convolution
KERNEL_SIZES	[3,5,7,11,15]	Kernel sizes for different time scales

Regularization		
Parameter	Value	Description
USE_BATCH_NORM	TRUE	Applies batch normalization
USE_LAYER_NORM	TRUE	Applies layer normalization
USE_RESIDUAL_CONNECTIONS	TRUE	Enables residual connections

Peak Loss Config		
Parameter	Value	Description
USE_PEAK_LOSS	TRUE	Applies extra weighting to peak-period errors
PEAK_WEIGHT	5	Weight multiplier for peak errors
PEAK_THRESHOLD_PERCENTILE	70	Defines peaks above the 70th percentile

Loss Weights		
Loss Type	Weight	Description
MSE	1	Mean Squared Error (primary loss)
MAE	0.2	Mean Absolute Error (auxiliary loss)
Directional Loss	0.1	Loss for directional trend accuracy

Training Optimization		
Parameter	Value	Description
BATCH_SIZE	16	Mini-batch size for training
LEARNING_RATE	0.0001	Small learning rate for stability
WEIGHT_DECAY	0.00001	L2 regularization strength
SCHEDULER_TYPE	cosine_warm_restart	Cosine annealing learning rate
PATIENCE	50	Early stopping patience
GRADIENT_CLIP_VALUE	0.3	Threshold for gradient clipping

Model Architecture: Peak-Aware LSTM-CNN for Milan Internet Traffic Prediction

1. Traffic Input
Function: Feeds the main traffic data into the network.
Details: 240 time steps representing 40 hours of past internet activity across 30 of Milan's busiest grid squares. All values are normalized.
Shape: [Batch, 240, 30]

2. Temporal Input
Function: Provides time-based context to the model.
Details: Includes engineered features such as hour/day/month (encoded as sin-cos), and binary indicators for weekends and holidays.
Shape: [Batch, 240, 10]

3. Multi-Scale CNN Feature Extractor
Function: Captures temporal patterns at multiple scales.
Layers:
- **Conv k=3**: Detects quick spikes (~30 min)
- **Conv k=5**: Captures short-term transitions (~50 min)
- **Conv k=7**: Focuses on mid-range patterns (~70 min)
- **Conv k=11**: Models daily routines (~110 min)
- **Conv k=15**: Extracts long-term activities (~150 min)

Each convolution operates in parallel and processes the same traffic input. The outputs are concatenated.
Output Shape: [Batch, 240, 256]

4. Concatenation Layer
Function: Combines the outputs of all convolution branches into a single temporal feature sequence.
Output Shape: [Batch, 240, 256]

5. Bidirectional LSTM Stack ×5
Function: Learns long-range temporal dependencies from the concatenated features.
Details: 5 stacked Bi-LSTM layers with 512 hidden units each and dropout of 0.15. Bidirectional to consider both past and future context.
Output Shape: [Batch, 240, 1024]

6. Multi-Head Self-Attention
Function: Allows the model to weigh different time steps differently based on their relevance.
Details: 16 attention heads, each 32 dimensions, with sinusoidal positional encoding.
Output Shape: [Batch, 240, 1024]

7. Temporal Feature Processor
Function: Reduces and transforms the temporal input for fusion.
Details: A dense layer maps [Batch, 240, 10] to [Batch, 240, 256].

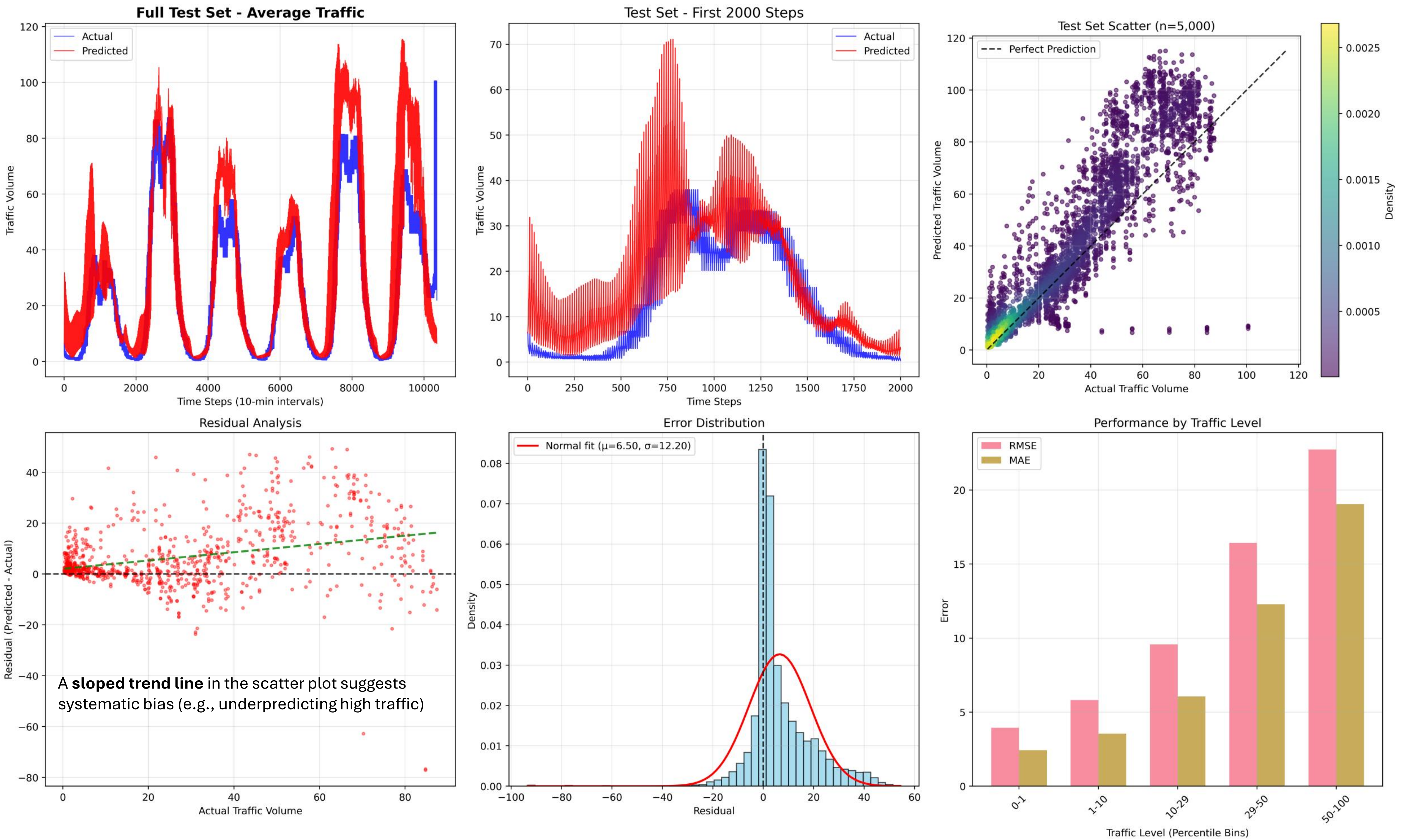
8. Feature Fusion
Function: Combines the outputs of the attention mechanism and the processed temporal features.
Details: Concatenates [1024 + 256] features, passes through a linear layer (1280 → 1024), ReLU activation, LayerNorm, and Dropout (0.15).
Output Shape: [Batch, 240, 1024]

9. Peak Detector (Auxiliary Branch)
Function: Identifies high-traffic periods to enhance learning.
Details: A binary classifier trained with a 5× weighted loss for samples in the 70th percentile or above. It operates only during training and influences the loss function.
Architecture:
Linear(1024→512)→ReLU→Linear(256)→ReLU→Linear(1)→Sigmoid

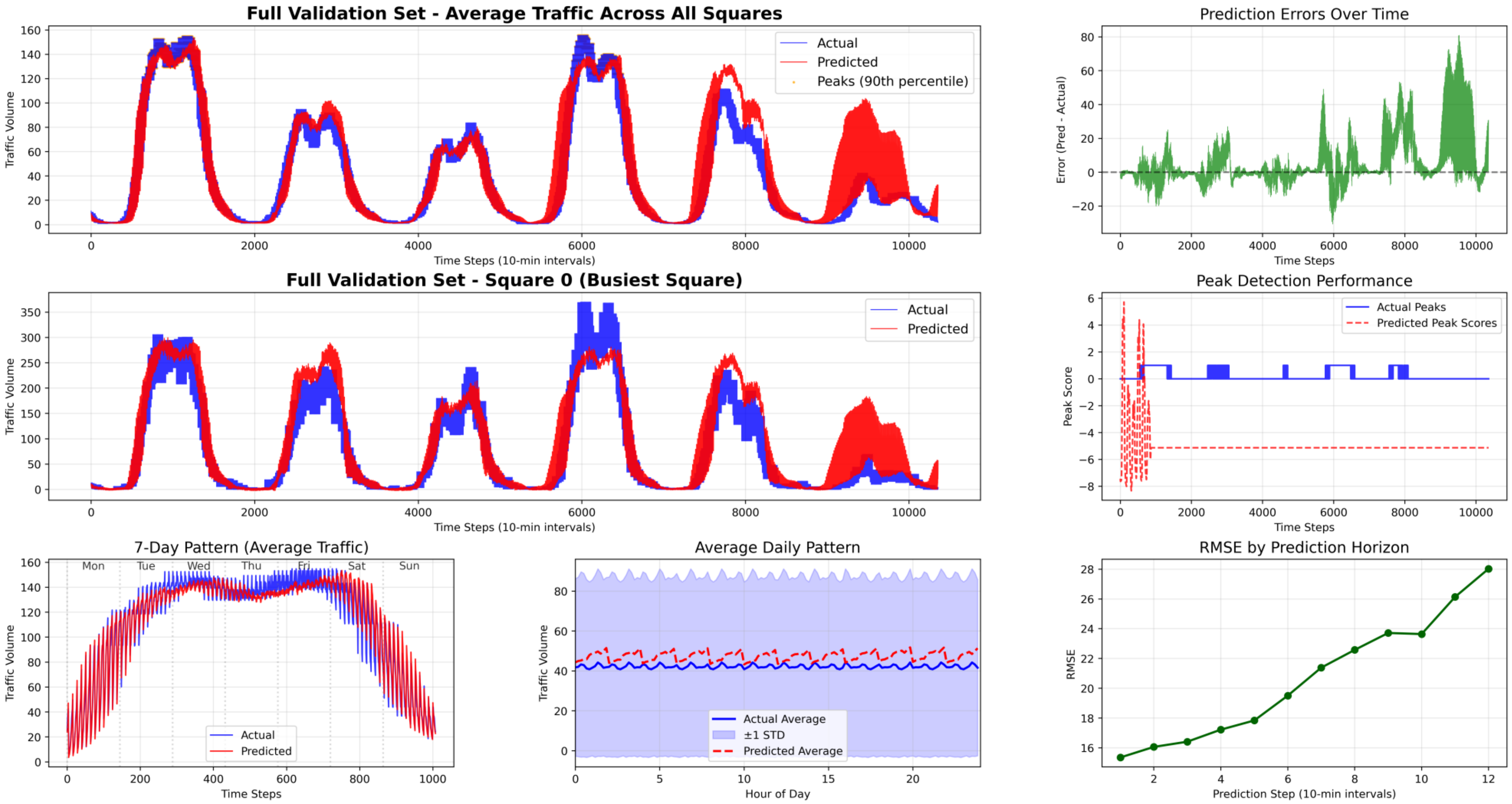
10. Multi-Horizon Output Heads
Function: Generate traffic volume predictions across 18 future 10-minute intervals.
Structure: Three grouped outputs:
- **Output 1–6**: 0–60 min forecast
- **Output 7–12**: 60–120 min forecast
- **Output 13–18**: 120–180 min forecast

Each head:
Linear(1024→512)→ReLU→Dropout(0.15)→Linear(256)→ReLU→Linear(30)
Output Shape: [Batch, 18, 30]

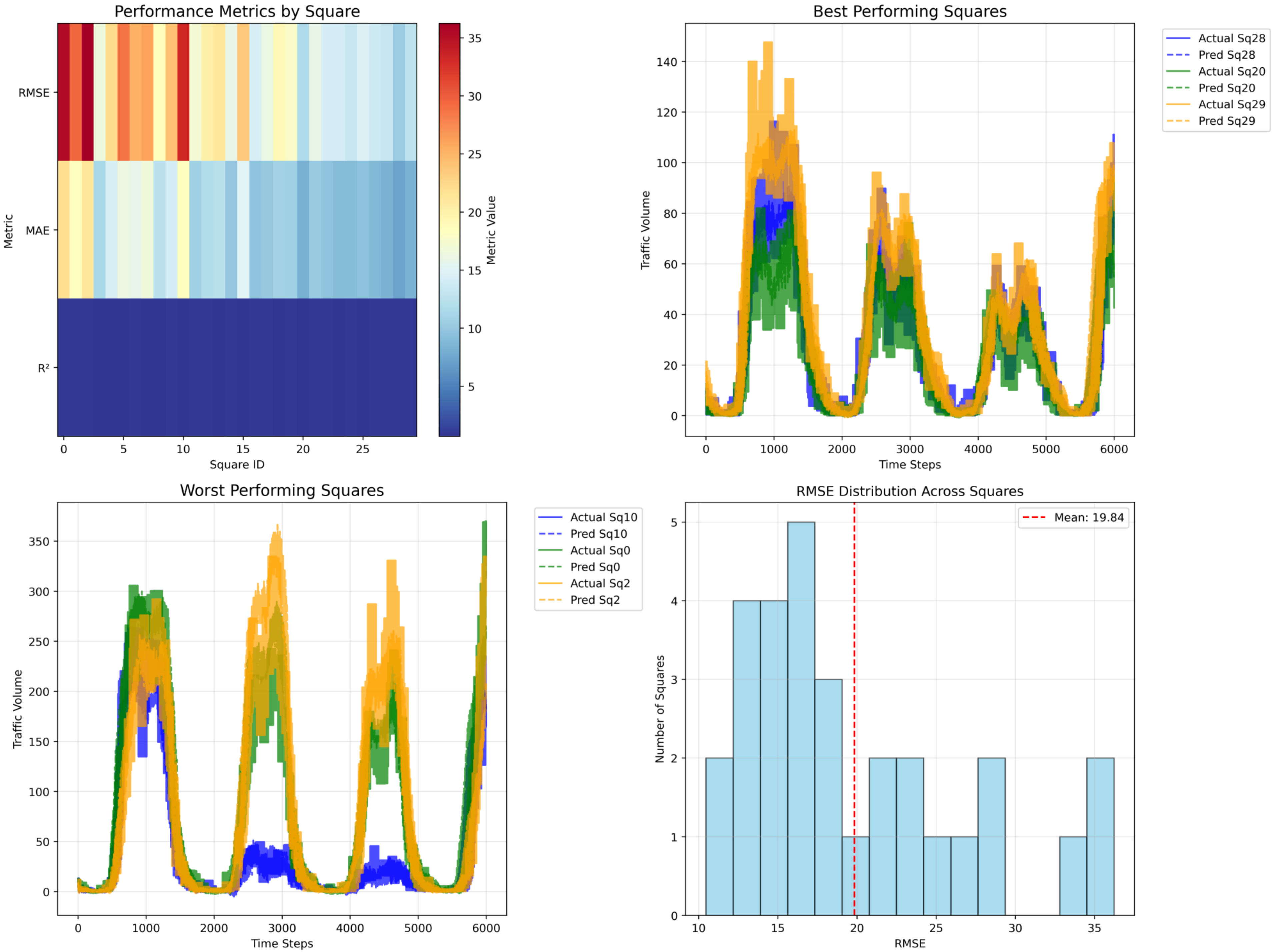
11. Output Summary
- **Final Output**: 18 future time steps (10-min each) for 30 city grid squares
- **Output Shape**: [Batch, 18, 30]
- **Used For**: Real-time traffic forecasting, capacity planning, and peak anticipation.

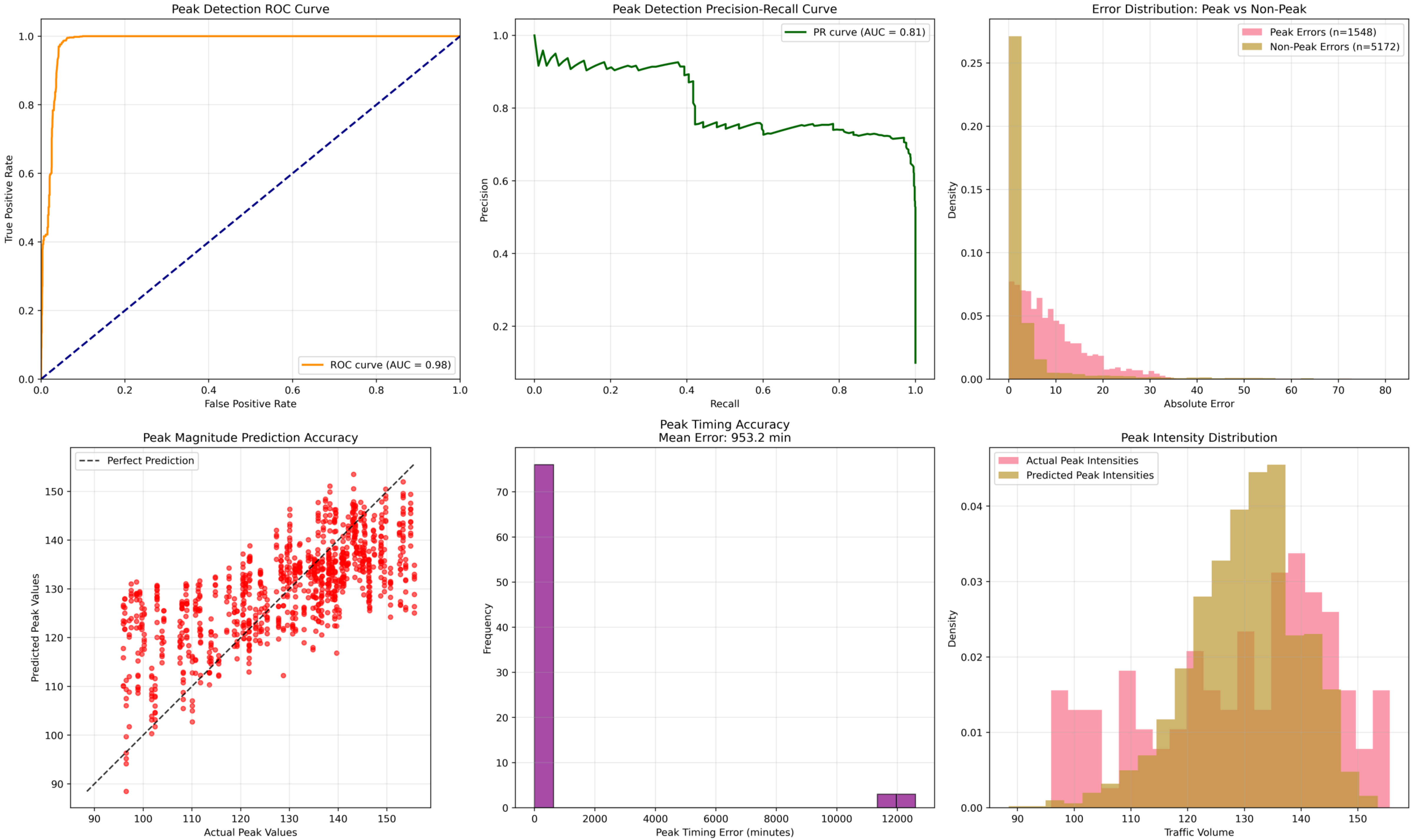

Full Test Set - Average Traffic
Actual
Predicted
Traffic Volume
Time Steps (10-min intervals)
Test Set - First 2000 Steps
Actual
Predicted
Traffic Volume
Time Steps
Test Set Scatter (n=5,000)
Perfect Prediction
Predicted Traffic Volume
Actual Traffic Volume
Density
Residual Analysis
Residual (Predicted - Actual)
Actual Traffic Volume
A **sloped trend line** in the scatter plot suggests systematic bias (e.g., underpredicting high traffic)
Error Distribution
Normal fit (μ=6.50, σ=12.20)
Density
Residual
Performance by Traffic Level
RMSE
MAE
Error
0-1
1-10
10-29
29-50
50-100
Traffic Level (Percentile Bins)

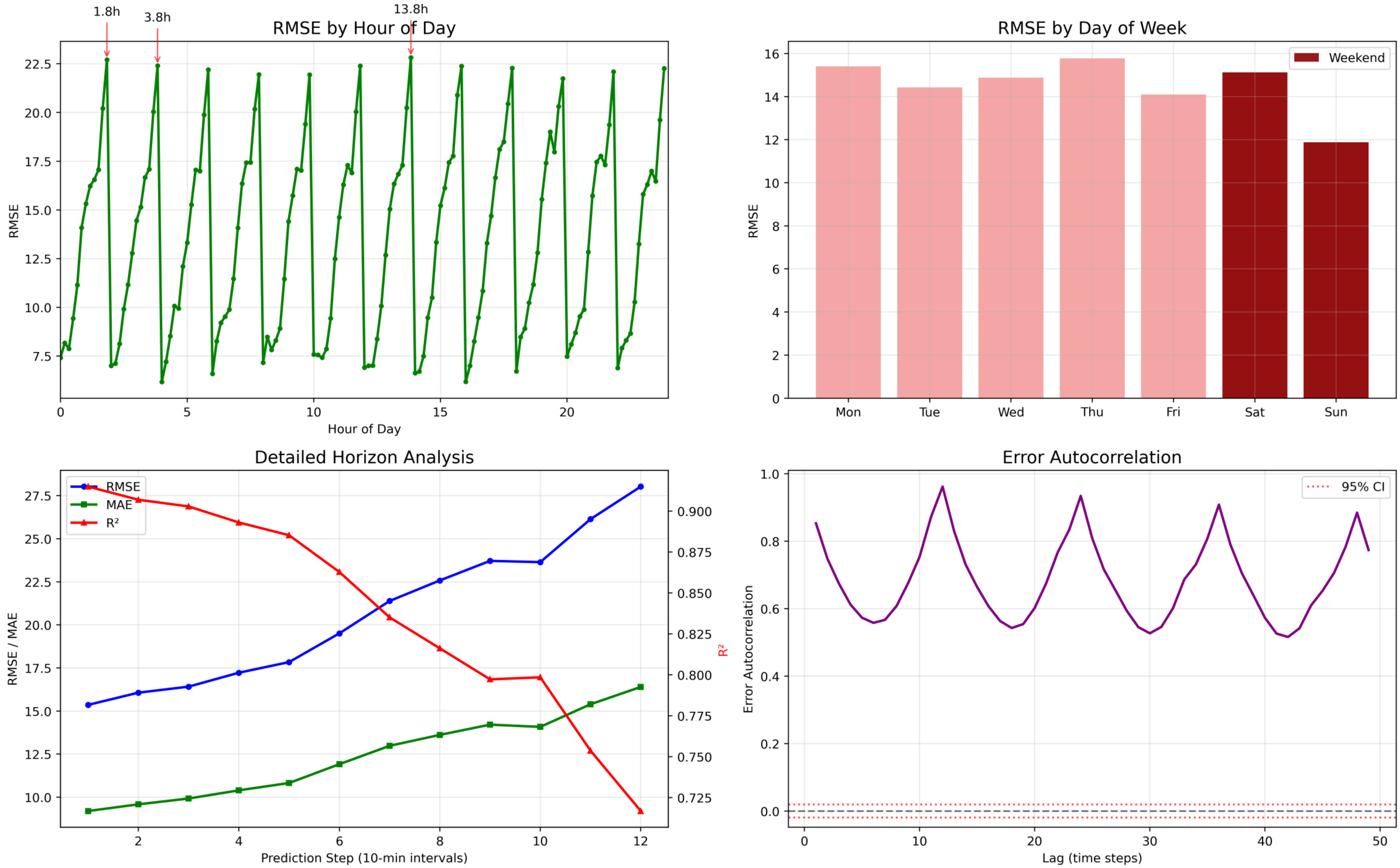
Enhanced Validation Set Analysis




```
Overall Performance:
  RMSE: 14.42  |  MAE: 8.09  |  R²: 0.898

Peak Traffic Performance (>90th percentile):
  RMSE: 9.20  |  Samples: 1,032

Low Traffic Performance (<10th percentile):
  RMSE: 1.67  |  Samples: 1,032

Model Configuration:
  Architecture: Enhanced LSTM  |  Hidden Dim: 512  |  Layers: 5
  Peak Loss: True  |  Multi-Scale: True  |  Attention: True

Data Info:
  Validation samples: 10,356  |  Prediction horizon: 12  |  Squares: 30
```


Performance Metrics by Square
RMSE
MAE
R²
Metric
Square ID
0 5 10 15 20 25
Metric Value
35 30 25 20 15 10 5
Best Performing Squares
Actual Sq28
Pred Sq28
Actual Sq20
Pred Sq20
Actual Sq29
Pred Sq29
Traffic Volume
Time Steps
0 1000 2000 3000 4000 5000 6000
0 20 40 60 80 100 120 140
Worst Performing Squares
Actual Sq10
Pred Sq10
Actual Sq0
Pred Sq0
Actual Sq2
Pred Sq2
Traffic Volume
Time Steps
0 1000 2000 3000 4000 5000 6000
0 50 100 150 200 250 300 350
RMSE Distribution Across Squares
Mean: 19.84
Number of Squares
RMSE
10 15 20 25 30 35
0 1 2 3 4 5


Peak Detection ROC Curve
ROC curve (AUC = 0.98)
True Positive Rate
False Positive Rate
Peak Detection Precision-Recall Curve
PR curve (AUC = 0.81)
Precision
Recall
Error Distribution: Peak vs Non-Peak
Peak Errors (n=1548)
Non-Peak Errors (n=5172)
Density
Absolute Error
Peak Magnitude Prediction Accuracy
Perfect Prediction
Predicted Peak Values
Actual Peak Values
Peak Timing Accuracy
Mean Error: 953.2 min
Frequency
Peak Timing Error (minutes)
Peak Intensity Distribution
Actual Peak Intensities
Predicted Peak Intensities
Density
Traffic Volume



RMSE by Hour of Day
1.8h
3.8h
13.8h
RMSE
Hour of Day
RMSE by Day of Week
Weekend
RMSE
Mon
Tue
Wed
Thu
Fri
Sat
Sun
Detailed Horizon Analysis
RMSE
MAE
R²
RMSE / MAE
R²
Prediction Step (10-min intervals)
Error Autocorrelation
95% CI
Error Autocorrelation
Lag (time steps)