AI-Driven Geospatial Analysis and Prediction of Cellular Coverage Using Multi-Carrier Measurements

Aliaa Salim, Southern University and A&M College, Louisiana
Fred Lacy, Southern University and A&M College, Louisiana
Yasser Ismail, Southern University and A&M College, Louisiana

BACKGROUND

- This study investigates whether collected LTE and 5G signal metrics can be used to identify and predict areas with weak or unstable cellular coverage.
- Two key metrics were analyzed:
  - **RSRP** – signal strength
  - **RSRQ** – signal quality
- A **Random Forest machine learning model** was developed using geographic and network-context features.
- Measurements were collected from **AT&T, T-Mobile, and Verizon** across suburban, urban, and campus environments.
- The goal is to generate predictive coverage maps that support proactive wireless network optimization.

RESEARCH QUESTIONS

This study focuses on one research question selected from a broader set of ten research questions related to wireless network performance analysis:

Can collected signal metrics be used to predict areas of weak or unstable coverage?

The objective is to determine whether field-collected LTE and 5G signal measurements can be used to develop predictive models that identify geographic areas likely to experience degraded cellular coverage.

METHODS AND MATERIALS

Methodology:

• Dataset Collection
SigCap measurement data were collected from AT&T, T-Mobile, and Verizon between October and November 2025 across suburban, urban, and campus environments.

• Data Cleaning and Validation
Raw files were cleaned by normalizing text fields, validating required parameters, and removing incomplete or invalid records.

• Serving-Cell Extraction
Only rows labeled as “primary” were retained to analyze active serving-cell measurements, while neighboring-cell records were excluded.

• Spatial Aggregation
Measurements were grouped into 100 m × 100 m geographic zones, and adjacent zones were combined into 500 m × 500 m super-zones for spatially independent analysis.

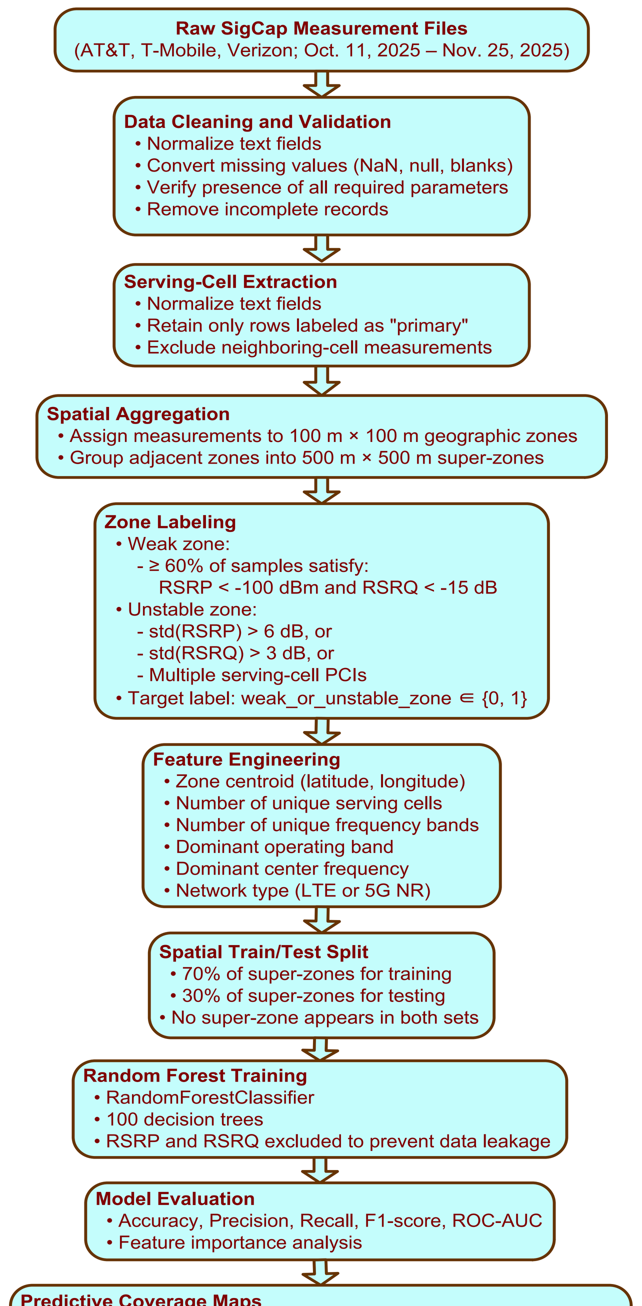
• Zone Labeling
Zones were classified as weak or unstable using threshold-based criteria for RSRP, RSRQ, signal variability, and serving-cell changes. A binary target variable, weak_or_unstable_zone, was then created.

• Feature Engineering
Geographic and network-context features, including location, serving-cell diversity, frequency information, and network type, were extracted for each zone.

• Random Forest Training
A Random Forest classifier with 100 decision trees was trained to predict weak or unstable coverage while excluding RSRP and RSRQ from the input features to prevent data leakage.

• Model Evaluation
Model performance was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics on spatially independent training and testing datasets.

• Predictive Coverage Maps
The trained model generated predictive coverage maps showing the probability of weak or unstable coverage for each geographic zone.



Spatial Partitioning Strategy:

• 100 m × 100 m Geographic Zones
Measurements were aggregated into 100 m × 100 m geographic zones to reduce short-term signal fluctuations and represent local coverage conditions more accurately. Measurements within the same zone were combined to generate zone-level statistics used for labeling and machine learning.

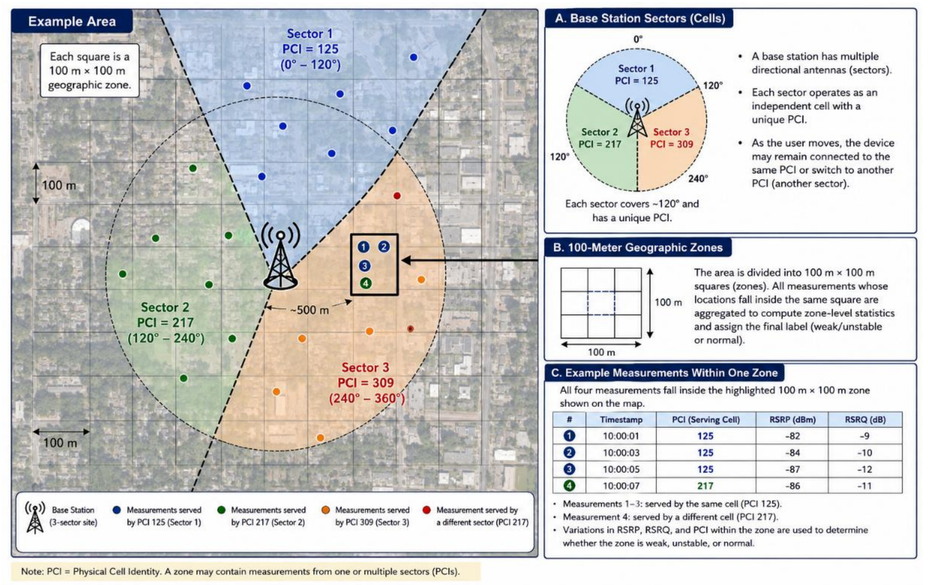


Figure 1. Relationship between base station sectors (cells), serving-cell measurements, and the 100 m × 100 m geographic zones used in this study.

• 500 m × 500 m Super-Zones
Adjacent 100 m × 100 m zones were grouped into 500 m × 500 m super-zones to reduce spatial data leakage during training and testing. All zones within the same super-zone were assigned entirely to either the training or testing dataset to ensure spatial independence.

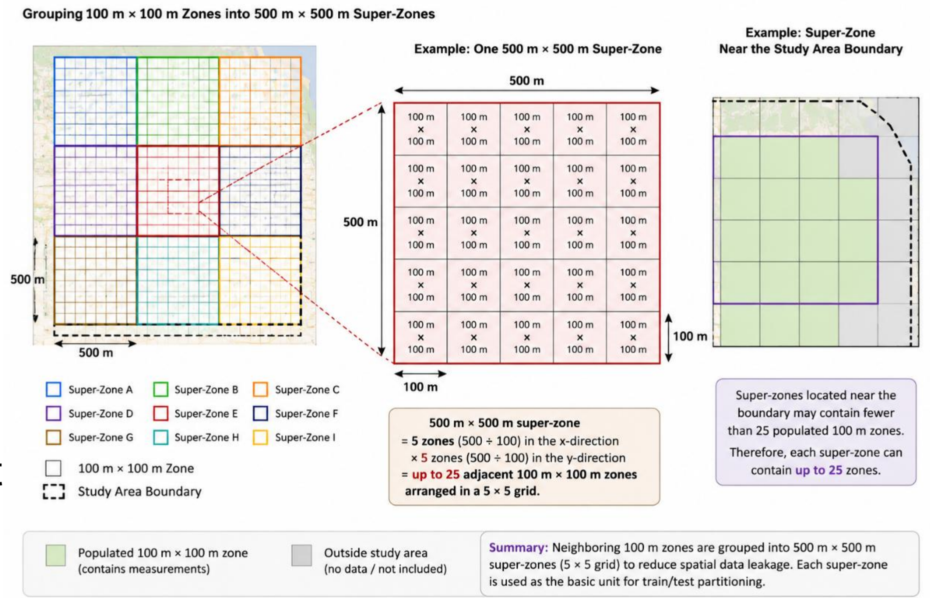


Figure 2. Hierarchical grouping of 100 m × 100 m geographic zones into 500 m × 500 m super-zones for spatially independent machine learning evaluation.

RESULTS

1. Cross-Carrier Performance Comparison

As shown in Table 1, the Random Forest classifier achieved strong predictive performance across the AT&T, Verizon, and T-Mobile datasets. AT&T achieved the highest overall accuracy (95.41%) and ROC-AUC (0.9893), while Verizon achieved perfect precision (100%). These results demonstrate that geographic and network-context features can effectively predict weak or unstable cellular coverage across different wireless environments.

Table 1. Cross-Carrier Comparison of RQ8 Results.

Metric	AT&T	Verizon	T-Mobile
Measurement Environment	Suburban	University Campus	Urban
Serving-Cell Measurements	200,331	91,955	8,167
100 m Geographic Zones	371	219	35
500 m Super-Zones	115	41	10
Weak Zones	5	9	3
Unstable Zones	42	86	19
Weak-or-Unstable Zones	44	89	20
Positive-Class Percentage	11.86%	40.64%	57.14%
Training Zones	262	142	24
Testing Zones	109	77	11
Accuracy	**95.41%**	**89.61%**	**81.82%**
Precision	76.92%	**100.00%**	**100.00%**
Recall	**83.33%**	75.76%	66.67%
F1-Score	0.8000	**0.8621**	0.8000
ROC-AUC	**0.9893**	0.9291	0.8000
False Positives	3	0	0
False Negatives	2	8	2

2. Cross-Carrier Performance Comparison

As shown in Figure 3, Feature importance analysis showed that serving-cell diversity, PCI diversity, geographic location, and frequency-band diversity were the strongest predictors of degraded coverage. These findings indicate that weak or unstable coverage follows repeatable spatial and network-dependent patterns.

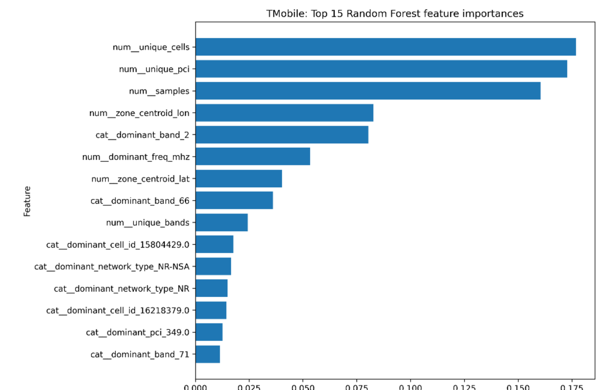


Figure 3. Cross-Carrier Comparison of RQ8 Results.

3. Predictive Coverage Maps

AT&T: Figures 4, 5, and 6 show strong agreement between the predicted and observed weak or unstable coverage zones in the AT&T testing regions. The Random Forest model successfully identified geographic hotspots of degraded coverage in areas not used during training, demonstrating that instability and degraded coverage follow predictable spatial patterns. Results also showed that most positive zones in the AT&T dataset were associated with instability-driven degradation rather than with persistent weak-signal conditions.

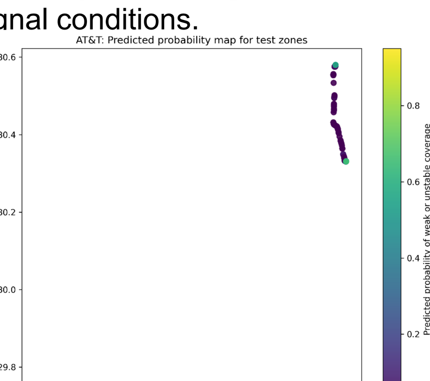


Figure 4. Predicted Probability Map.

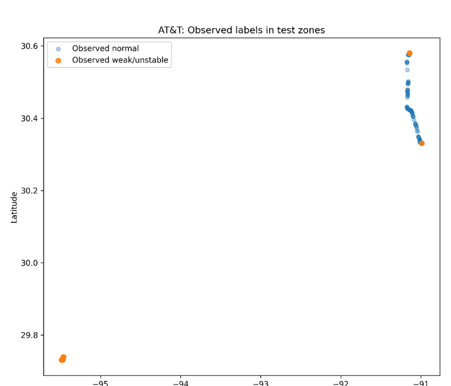


Figure 5. Observed Weak/Unstable Zone Labels.

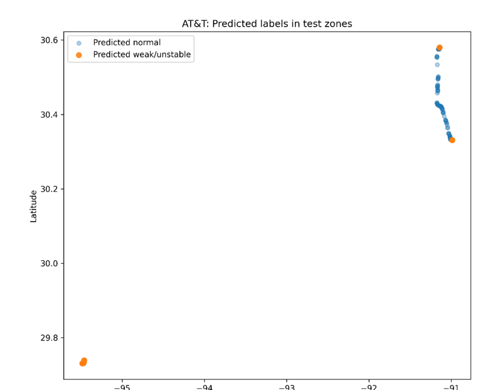


Figure 6. Predicted Weak/Unstable Zone Labels.

Verizon: Figures 7, 8, and 9 show strong agreement between the observed and predicted weak or unstable coverage zones across Verizon's testing regions. The Random Forest model successfully identified the major degradation hotspots, including the northern cluster and the isolated southern hotspot, while assigning high probabilities to the most degraded areas. These results demonstrate that the proposed framework can accurately localize degraded coverage patterns in geographically unseen regions using geographic and network-context features.

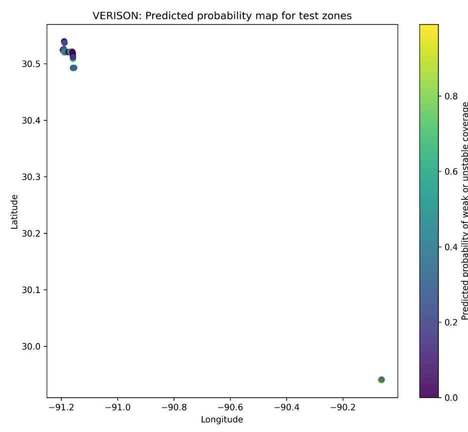


Figure 7. Predicted Probability Map.

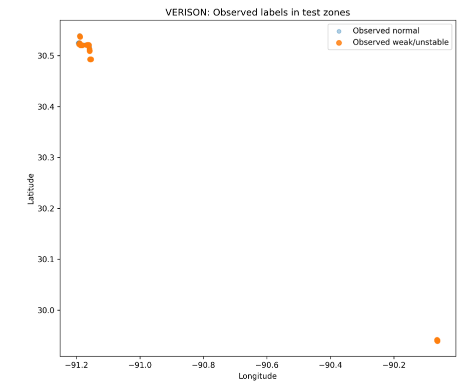


Figure 8. Observed Weak/Unstable Zone Labels.

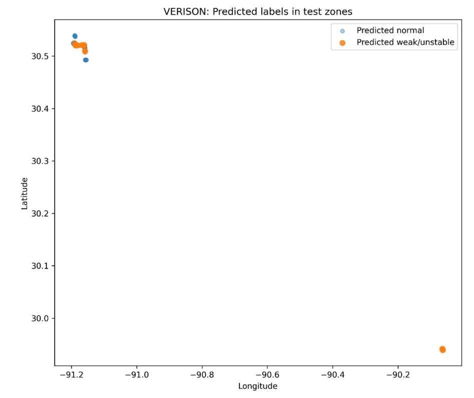


Figure 9. Predicted Weak/Unstable Zone Labels.

T-Mobile: Figures 10, 11, and 12 demonstrate strong agreement between the predicted and observed weak- or unstable-coverage zones in the T-Mobile test regions. The probability map successfully identified the primary high-risk hotspot, while the predicted labels closely matched the observed degradation patterns. Only a small number of boundary zones were misclassified, primarily in regions where the degradation patterns were less distinct. Overall, the results confirm that the model effectively captured the geographic structure of degraded wireless coverage in the T-Mobile dataset.

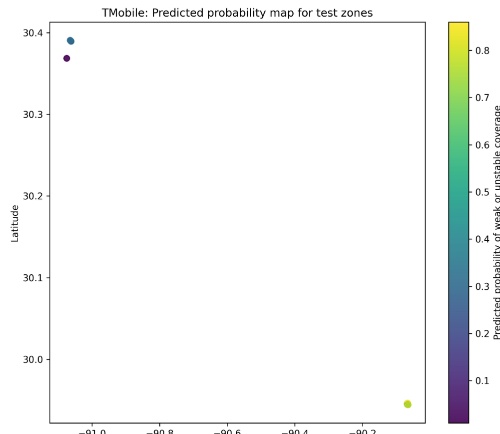


Figure 10. Predicted Probability Map.

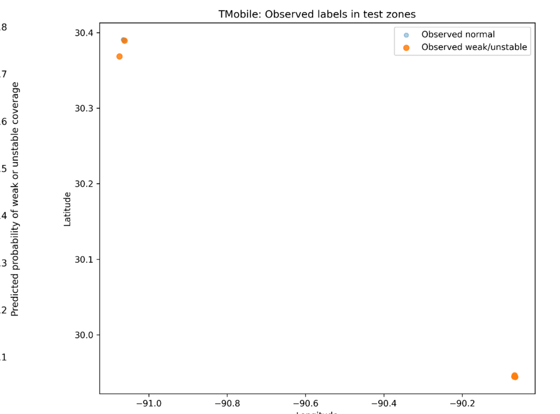


Figure 11. Observed Weak/Unstable Zone Labels.

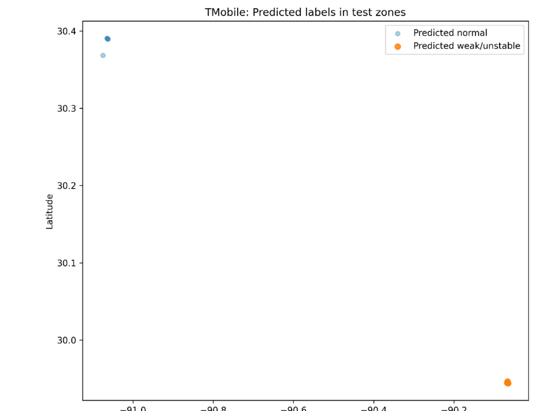


Figure 12. Predicted Weak/Unstable Zone Labels.

CONCLUSION

This study demonstrated that weak or unstable LTE/5G coverage can be accurately predicted using geographic and network-context features, without directly using signal-strength measurements as model inputs. The proposed Random Forest framework successfully identified degraded coverage hotspots across AT&T, Verizon, and T-Mobile datasets while maintaining strong performance on geographically unseen testing regions. Results showed that degraded coverage exhibits stable, predictable spatial patterns strongly associated with serving-cell instability, geographic location, and radio-environment complexity. The predictive coverage maps further confirmed the practical feasibility of machine-learning-based wireless coverage prediction for supporting future network planning and optimization.

ACKNOWLEDGEMENTS

This work has been supported in part through SpectrumX, the U.S. National Science Foundation (NSF) Spectrum Innovation Center, funded via Award AST 21-32700 and operated under Cooperative Agreement with NSF by the University of Notre Dame.

REFERENCES

[1] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
[2] Ubiquiti Inc., “Understanding Cellular Signal Strength and Quality,” Ubiquiti Help Center. Available: https://help.ui.com/hc/en-us/articles/31173907212567-Understanding-Cellular-Signal-Strength-and-Quality. Accessed: May 8, 2026.