

# Spatio-Temporal Regression of LTE/5G Signal Metrics for ITU/3GPP Quality Classification on Drive-Test Data from Manaus

Myke Valadão, Ana Corrêa, Alexandre Senna, Celso Carvalho and Waldir Sabino

**Abstract**—This paper investigates the feasibility of predicting LTE/5G signal-quality metrics (RSRP, RSRQ, SINR, and RSSI) from purely spatio-temporal inputs (latitude, longitude, hour, weekday) and using the predicted values to recover the ITU/3GPP-inspired composite quality label. A drive-test dataset of 17,646 measurements collected in Manaus, Brazil (April 13–16, 2026) is processed through a four-stage pipeline: (i) anomaly detection and removal, (ii) per-metric ITU/3GPP threshold-based labeling, (iii) multi-output regression of the four signal metrics, and (iv) re-classification of the predicted metrics for label validation. Seven regression models are compared on the same stratified 80/20 split. The best performer, a Decision Tree, achieves  $R^2 = 0.884$  averaged across the four metrics and recovers the original ITU/3GPP-inspired label with 86.51% accuracy. The results indicate that spatio-temporal features alone capture a substantial portion of the signal-quality structure in the studied area, supporting the feasibility of prospective coverage estimation without on-device measurement. The Android application used in the experiments is available in GitHub.

**Keywords**—LTE, 5G NR, signal quality, ITU-T G.1028, 3GPP TS 36.214, regression, drive test, spatio-temporal modelling, Manaus.

## I. INTRODUCTION

Mobile network coverage characterization traditionally relies on direct measurement campaigns, in which user equipment (UE) reports physical-layer metrics such as reference signal received power (RSRP), reference signal received quality (RSRQ), signal-to-interference-plus-noise ratio (SINR), and received signal strength indicator (RSSI). These measurements are associated with standardized LTE physical-layer measurement procedures, such as 3GPP TS 36.214 [1], and can be related to Quality of Service (QoS) definitions and service-level assessment frameworks, such as ITU-T E.800 and ITU-T G.1028 [2], [3]. Such indicators support applications ranging from network optimization to QoS-aware service design.

While effective, this measurement-driven approach is reactive and costly: a quality estimate at a given location and time usually requires a sample to be collected there [4], [21]. A natural question is whether the dependency between space, time, and signal quality is strong enough to enable prospective estimation, predicting the metrics themselves at locations and times without prior measurements. If feasible, such prediction

could support route planning, proactive network management, and coverage interpolation in regions with sparse measurement campaigns.

Recent studies have increasingly explored data-driven alternatives to dense measurement campaigns, including mobile-network quality prediction from GPS information, drive-test measurements, and machine-learning-based coverage estimation [4], [12], [13]. These works indicate that spatial information can be a strong proxy for network behavior, especially when the propagation environment exhibits recurrent geographic patterns. In parallel, radio map estimation has emerged as an important research direction for wireless network planning, digital twins, and coverage prediction, since it aims to infer spatially continuous signal-quality information from sparse measurements or environmental descriptors [14]–[16].

This paper addresses that question through a controlled experiment on a drive-test dataset collected in Manaus, Amazonas, Brazil. Rather than predicting the quality label directly, which would mask the underlying physical relationships, we frame the task as a multi-output regression problem: given only  $\langle \text{latitude, longitude, hour, weekday} \rangle$  as input, the model must predict the four signal metrics simultaneously. The predicted metrics are then re-classified using the same ITU/3GPP-inspired thresholds applied to the ground-truth measurements, allowing us to assess both the regression quality and the resulting label agreement.

The main contributions of this work are threefold. First, we propose a reproducible four-stage pipeline that separates data cleaning, threshold-based labeling, regression, and re-classification, making the prediction process transparent and interpretable. Second, we provide empirical evidence that location and time alone can explain a substantial portion of the signal-quality variance in an urban Brazilian scenario, with the best models achieving  $R^2$  values between 0.86 and 0.90 across the four evaluated metrics. Third, we present a comparative evaluation of seven regression models using identical train/test splits, covering tree-based ensembles, instance-based learning through k-NN, and a linear baseline.

## II. METHODOLOGY

### A. Overview of the Proposed Pipeline

The proposed methodology follows a four-stage pipeline designed to evaluate whether LTE/5G signal-quality indicators can be inferred from compact spatio-temporal information,

Myke Valadão, Celso Carvalho, and Waldir Sabino are with the Federal University of Amazonas and the Center for R&D in Electronics and Information Technology (UFAM/CETELI), Amazonas, Brazil, e-mails: mykedouglas@ufam.edu.br; julia.correa@ufam.edu.br; alexandre.antonaccio@ufam.edu.br; ccarvalho\_@ufam.edu.br; waldirjr@ufam.edu.br.

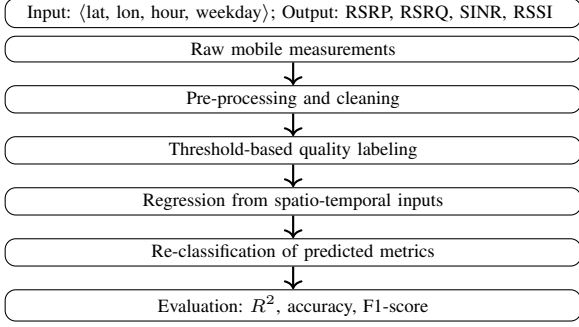


Fig. 1: Proposed pipeline for predicting and re-labeling LTE/5G signal-quality measurements from spatio-temporal inputs.

following the broader premise of spatio-temporal network-quality prediction and radio-map learning [14], [20]. Instead of directly training a classifier from location and time to quality labels, the pipeline first reconstructs the continuous radio measurements and only then maps the predicted values back to network quality classes. This design allows the model to preserve the physical meaning of the original signal indicators, while also enabling a direct comparison between measured and estimated values.

In the first stage, the raw mobile measurements are post-processed to remove invalid records, missing values, anomalous samples, and systematic artefacts, such as the anomalous values observed in some NR SINR measurements. After this cleaning step, each valid sample is represented by its geographic position, temporal attributes, and measured signal-quality indicators, namely RSRP, RSRQ, SINR, and RSSI. In the second stage, a threshold-based labeling procedure is applied using ITU/3GPP-inspired signal-quality criteria. This produces a composite network quality label for each sample, assigning it to one of four classes: Excellent, Good, Moderate, or Poor.

In the third stage, different regression models are trained to estimate the continuous signal-quality metrics from the spatio-temporal input vector ⟨latitude, longitude, hour, weekday⟩. The goal is to assess how much of the signal behavior can be predicted from location and time alone, without using direct radio measurements as input. Finally, in the fourth stage, the predicted signal metrics are converted again into quality labels using the same threshold-based rule adopted for the original measurements. The resulting predicted labels are then compared with the reference labels, allowing the evaluation of both regression performance and re-labeling accuracy. This pipeline, as shown in Fig. 1 provides a structured way to analyze the predictability of mobile network quality while maintaining interpretability at each processing stage.

### B. Dataset Generation and Preprocessing

To generate the dataset, we developed an Android application to collect mobile network measurements directly from user devices. The use of field measurements collected from user equipment follows previous drive-test and spatio-temporal network-quality analysis studies [4], [20]. The collected attributes include latitude, longitude, RSRP, RSRQ, SINR, RSSI, timestamp, and day of the week. The measurement campaign was conducted in Manaus, Amazonas, Brazil, between April

TABLE I: Per-metric ITU/3GPP-inspired thresholds used to assign individual quality scores before computing the composite QoS label.

| Class     | RSRP<br>(dBm)  | RSRQ<br>(dB) | SINR<br>(dB) | RSSI<br>(dBm) |
|-----------|----------------|--------------|--------------|---------------|
| Excellent | $\geq -80$     | $\geq -10$   | $\geq 20$    | $\geq -70$    |
| Good      | $[-90, -80)$   | $[-13, -10)$ | $[13, 20)$   | $[-85, -70)$  |
| Moderate  | $[-100, -90)$  | $[-15, -13)$ | $[0, 13)$    | $[-95, -85)$  |
| Weak      | $[-110, -100)$ | $[-17, -15)$ | $[-3, 0)$    | $[-110, -95)$ |
| Poor      | $< -110$       | $< -17$      | $< -3$       | $< -110$      |

13 and April 16, 2026, from 07:00 to 20:00. Three participants used the application in both static and mobile conditions, allowing the dataset to capture different mobility patterns and propagation scenarios. The application recorded network measurements at one-second intervals.

Before applying the labeling procedure, the collected data were preprocessed to remove invalid and inconsistent samples. This cleaning step included the removal of records containing missing values, such as NaN entries, as well as anomalous measurements outside physically meaningful or expected ranges for the considered radio indicators. By filtering these samples prior to labeling, we ensured that the composite QoS classes were computed only from valid and reliable measurements.

After data collection and preprocessing, each sample was labeled according to a threshold-based signal quality procedure inspired by ITU and 3GPP recommendations, particularly 3GPP TS 36.214 [1], ITU-T G.1028 [2], and the general QoS terminology defined in ITU-T E.800 [3]. The labeling process was applied only to samples containing valid values for the four selected signal quality indicators: RSRP, RSRQ, SINR, and RSSI.

The proposed labeling method consists of three steps. First, each raw signal quality value is independently evaluated according to the threshold intervals shown in Table I. These intervals assign each metric to one of five quality levels: Excellent, Good, Moderate, Weak, or Poor. Second, each categorical level is converted into a numerical score  $s_k \in [0, 4]$ , where Poor corresponds to 0, Weak to 1, Moderate to 2, Good to 3, and Excellent to 4. This scoring step normalizes heterogeneous radio measurements, originally expressed in different units and value ranges, into a common quality scale, which is consistent with the need to combine physical-layer indicators into service-oriented quality descriptors [2], [3]. Third, the four per-metric scores are averaged to obtain a composite score  $\bar{s}$ , which is then mapped into a final QoS label  $y \in \{\text{Excellent, Good, Moderate, Poor}\}$ .

The composite score and the final QoS label are defined as follows:

$$\bar{s} = \frac{1}{4} \sum_{k=1}^4 s_k, \quad y = \begin{cases} \text{Excellent} & \bar{s} \geq 3.0, \\ \text{Good} & 2.0 \leq \bar{s} < 3.0, \\ \text{Moderate} & 1.0 \leq \bar{s} < 2.0, \\ \text{Poor} & \bar{s} < 1.0. \end{cases} \quad (1)$$

Thus, Table I is used to compute the individual contribution of each radio metric, while Equation 1 combines these contributions into a single end-to-end QoS label. Although each metric is initially classified into five levels, the final label is represented using four classes to provide a more compact

and interpretable quality indicator for network monitoring and semantic analysis tasks. This procedure ensures that the labeling process is transparent, reproducible, and directly derived from standardized measurement definitions and QoS-oriented signal-quality criteria [1], [2].

### C. Multi-Output Regression

Let  $\mathbf{x}_i \in \mathbb{R}^4$  denote the input vector  $\mathbf{x}_i = (\text{lat}_i, \text{lon}_i, h_i, w_i)$ , where  $h_i$  is the hour expressed as a decimal value,  $h_i = \text{hour} + \text{minute}/60$ , and  $w_i$  is the day-of-week index. The target vector is  $\mathbf{y}_i = (\text{RSRP}_i, \text{RSRQ}_i, \text{SINR}_i, \text{RSSI}_i)$ . Thus, the regression task consists of learning a multi-output function  $f : \mathbb{R}^4 \rightarrow \mathbb{R}^4$  that minimizes the average squared prediction error over all samples and all four output metrics,  $\min_f \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^4 (f_k(\mathbf{x}_i) - y_{i,k})^2$ , following a standard multi-output regression formulation [11].

Seven regression families are compared: Decision Tree (DT), Random Forest (RF), Extra Trees (ET), Gradient Boosting (GB), Histogram-based Gradient Boosting (HGB), k-NN ( $k = 5$ , distance-weighted), and Ridge Regression with standardised inputs (linear baseline). The model set covers recursive partitioning through CART decision trees [5], bagging-based tree ensembles through Random Forests [6], highly randomized tree ensembles through Extra Trees [7], additive boosting models through Gradient Boosting [8], instance-based learning through nearest neighbors [9], and regularized linear regression through Ridge Regression [10]. All models are trained on a stratified 80/20 split (stratification on the composite label). Hyper-parameters were chosen from default values commonly used in the scikit-learn library [11]; the focus of this study is on the comparative behaviour of model families rather than on hyper-parameter tuning.

Given the trained regressor  $\hat{f}$ , the predicted signal vector for a test sample  $i$  is  $\hat{\mathbf{y}}_i = \hat{f}(\mathbf{x}_i)$ . We then apply the same composite-labeling rule, Eq. (1), to  $\hat{\mathbf{y}}_i$ , yielding  $\hat{y}_i$ . The agreement  $\mathbb{1}[\hat{y}_i = y_i]$  between the two label assignments is the primary classification metric reported alongside the standard regression scores ( $R^2$ , MAE, RMSE [22]).

## III. EXPERIMENTS AND RESULTS

Although the RSSI column retains 4,698 valid samples after cleaning, the multi-output regression task requires all four metrics to be present simultaneously in each row, since they are the four outputs of the regressor. The 367 NR samples for which only RSSI is available, RSRP, RSRQ, and SINR being missing in those rows, are therefore excluded from the regression dataset by row-wise listwise deletion, not by any form of metric-level downsampling. After this filtering and the additional GPS-validity constraint, 4,331 fully-observed records remain. These are split into 3,464 training and 867 test samples (stratified by the composite label). Composite-label distribution in the test set is approximately 7.5% Excellent, 34.1% Good, 34.1% Moderate, and 24.2% Poor.

Figure 2 presents the spatial distribution of the data collected in the city of Manaus, Amazonas, after the post-processing stage, following the radio-map perspective in which

TABLE II: Model comparison on the held-out test set ( $n = 867$ ). Models are sorted by re-labeling accuracy.

| Model | Acc.         | Avg.<br>$R^2$ | MAE   | RSRP<br>$R^2$ | RSRQ<br>$R^2$ | SINR<br>$R^2$ | RSSI<br>$R^2$ | Train<br>(s) |
|-------|--------------|---------------|-------|---------------|---------------|---------------|---------------|--------------|
| DT    | <b>86.51</b> | 0.884         | 1.610 | 0.897         | 0.861         | 0.880         | 0.897         | 0.003        |
| RF    | <b>86.16</b> | 0.883         | 1.630 | 0.896         | 0.861         | 0.880         | 0.896         | 0.678        |
| k-NN  | 84.54        | 0.861         | 1.608 | 0.878         | 0.829         | 0.858         | 0.878         | 0.002        |
| ET    | 83.39        | 0.884         | 1.608 | 0.897         | 0.862         | 0.881         | 0.897         | 0.499        |
| GB    | 80.28        | 0.880         | 1.785 | 0.895         | 0.855         | 0.875         | 0.895         | 1.296        |
| HGB   | 79.12        | 0.881         | 1.725 | 0.895         | 0.856         | 0.877         | 0.895         | 9.310        |
| Ridge | 36.56        | 0.223         | 6.098 | 0.222         | 0.368         | 0.081         | 0.222         | 0.003        |

signal measurements are interpreted as spatially varying quantities over a geographic region [14], [16]. Each point on the map represents a geographic location where mobile network measurements were recorded, considering the RSRP, RSRQ, RSSI, and SINR indicators. The point colors follow the scale shown in the legend of each subfigure, where colors closer to red indicate lower signal levels, while colors closer to blue represent better signal conditions. It is important to highlight that these visualizations show only the measured values after post-processing, before the samples are labeled into network quality classes.

### A. Model Comparison

Table II ranks the seven models by re-labeling accuracy. The DT leads with 86.51% accuracy and an average  $R^2$  of 0.884, narrowly ahead of the RF (86.16%,  $R^2 = 0.883$ ). The two are statistically close—both are based on axis-aligned partitions of the spatio-temporal feature space, a property inherited from tree-based partitioning methods [5], [6]. This appears well suited to the near-discrete sectorial structure of the urban coverage. GB variants are competitive in regression quality but slightly lower in re-labelling accuracy (79–80%), consistent with their additive boosting formulation [8]. Ridge regression collapses ( $R^2 = 0.22$ , 36.6%), confirming that the signal-to-location relationship is strongly non-linear in the studied area and not adequately represented by a single regularized linear mapping [10].

Figure 3 presents the  $R^2$  values obtained for each signal quality metric, namely RSRP, RSRQ, SINR, and RSSI, across the evaluated models. The tree-based and instance-based models achieved consistently high scores, generally ranging from 0.86 to 0.90, indicating a strong ability to estimate the original signal indicators after post-processing. Among these models, GB, ET, RF, and DT showed very similar behavior, with small variations depending on the evaluated metric. On the other hand, Ridge regression presented substantially lower  $R^2$  values, around 0.21–0.24, confirming that the signal behavior is not well represented by a simple linear model. These results suggest that the dataset contains non-linear spatial and signal propagation patterns that are better captured by decision-tree-based approaches and other non-parametric models [5], [7], [9].

Figure 4 compares the per-class F1-score of the five best-performing models considering the four network quality classes: Excellent, Good, Moderate, and Poor. The results show that the extreme classes, Excellent and Poor, are classi-

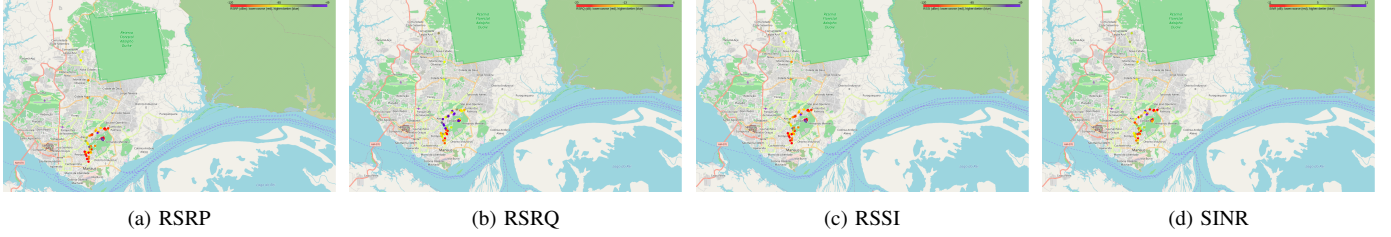


Fig. 2: Spatial distribution of the measured LTE/5G signal indicators in Manaus.

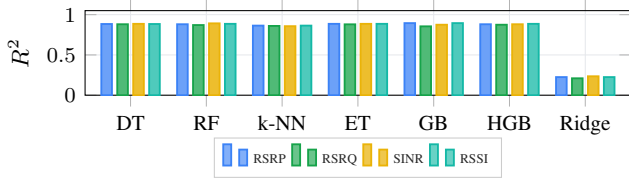
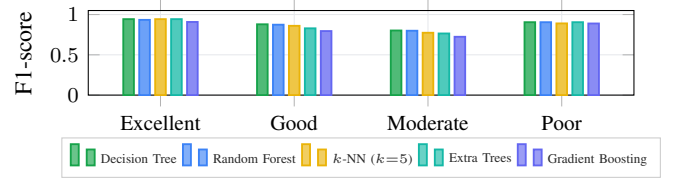

 Fig. 3:  $R^2$  per metric for each model.


Fig. 4: Per-class F1 across the top five models.

fied with higher F1-scores across all models, reaching values close to or above 0.90. This indicates that samples with very good or very poor signal conditions are easier to distinguish. In contrast, the intermediate classes, Good and Moderate, present lower F1-scores, especially the Moderate class, which is more difficult to separate due to its proximity to neighboring quality levels. Overall, Decision Tree and Random Forest achieved the most balanced performance among the evaluated classes, while the lower scores for the intermediate classes highlight the inherent ambiguity of signal quality transitions in real mobile network measurements.

### B. Error Analysis

Table III presents the confusion matrix for the DT model, which achieved the best overall re-labeling accuracy among the evaluated models. The results show that most predictions are concentrated along the main diagonal, indicating a strong agreement between the true and predicted network quality labels. The Excellent class was classified with high precision, with no samples incorrectly assigned to it from other classes. Similarly, the Poor class presented a high number of correct predictions, with most errors occurring only at the boundary with the Moderate class. The main source of confusion appears between the Good and Moderate classes, where 53 Moderate samples were predicted as Good and 17 Good samples were predicted as Moderate. This behavior is expected because these two classes represent intermediate signal quality levels, where the transition between categories is less clearly separated than in the extreme classes.

Table IV summarizes the per-class precision, recall, and F1-score obtained by the DT model on the test set. The model achieved a weighted F1-score of 0.864, confirming its strong overall performance in the re-labeling task. The Excellent class obtained the highest precision, equal to 1.000, indicating that all samples predicted as Excellent were correctly classified. The Poor class also showed a balanced performance, with precision, recall, and F1-score all equal to 0.905. In contrast, the Moderate class presented the lowest recall and F1-score, with values of 0.753 and 0.802, respectively, mainly due to its overlap with the Good class. These results reinforce that the

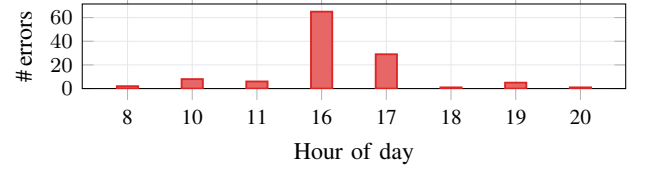


Fig. 5: Distribution of misclassification errors across hours of day for the Decision Tree model.

Decision Tree model is effective for identifying clear signal quality conditions, while most classification errors occur in the central quality ranges, where the signal measurements are naturally more ambiguous.

### C. Discussion

Figure 5 shows the distribution of misclassification errors across the hours of the day for the DT model, which achieved the best overall performance among the evaluated models. The results indicate that the errors are not uniformly distributed over time, with a clear concentration at 16:00, where 65 misclassifications were observed. A secondary peak also appears at 17:00, with 29 errors, while the remaining hours present substantially lower error counts. This temporal concentration suggests that the model had greater difficulty re-labeling samples collected during the late afternoon period. One possible explanation is that this interval may correspond to a peak-usage or high-mobility period in the urban environment, when increased user movement, traffic density, and network load can produce more unstable radio conditions, a factor also considered in spatio-temporal mobile-network quality analysis [12], [20]. Such variability may make the boundaries between adjacent quality classes less distinct, especially between intermediate classes such as Good and Moderate. Therefore, the peak in errors around 16:00–17:00 does not necessarily indicate a limitation of the DT model alone, but may also reflect more complex propagation and usage conditions during this time window.

The proposed pipeline can also be interpreted as a compact radio-map learning strategy, since it estimates signal-quality indicators from location and time and then converts them into interpretable network-quality labels. This connects the present

TABLE III: Confusion matrix — Decision Tree, test set ( $n = 867$ ).

| True label | Predicted label |      |          |      | $\Sigma$ |
|------------|-----------------|------|----------|------|----------|
|            | Excellent       | Good | Moderate | Poor |          |
| Excellent  | 58              | 7    | 0        | 0    | 65       |
| Good       | 0               | 279  | 17       | 0    | 296      |
| Moderate   | 0               | 53   | 223      | 20   | 296      |
| Poor       | 0               | 0    | 20       | 190  | 210      |

study to recent works on radio map construction, pathloss map prediction, and RF coverage estimation using empirical measurements, deep learning, and generative models [15], [17]–[19]. These connections suggest that future versions of the pipeline may benefit from richer environmental descriptors, such as building density, road topology, terrain elevation, and base-station proximity, while still preserving the interpretability of the threshold-based re-labeling stage.

#### IV. CONCLUSION

This study investigated the predictability of LTE/5G signal-quality metrics from purely spatio-temporal inputs in an urban Brazilian scenario. Using a four-stage pipeline that separates data cleaning, threshold-based labeling, regression, and re-classification, we showed that a compact input representation composed only of ⟨latitude, longitude, hour, weekday⟩ can capture a substantial portion of the variability in RSRP, RSRQ, SINR, and RSSI, reaching an average  $R^2$  close to 0.88 for the best-performing models. After mapping the predicted signal metrics back to the ITU/3GPP-inspired composite quality label, the best model achieved a re-labeling accuracy of 86.5%.

The results also revealed relevant characteristics of the dataset and the prediction task. The systematic anomalous artefact observed in NR SINR samples was identified and removed during post-processing, avoiding label distortion without compromising valid measurements. Moreover, tree-based models consistently outperformed both the linear baseline and smoother regression approaches, while Ridge regression presented a significant performance drop. This behavior confirms the strongly non-linear nature of the location-to-signal mapping in the considered urban environment. The per-class analysis further showed that the extreme quality classes, Excellent and Poor, are recovered with higher fidelity, whereas most errors occur near the Good/Moderate boundary, where signal-quality transitions are naturally more ambiguous.

#### REFERENCES

- [1] 3GPP, “Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer; Measurements,” 3GPP TS 36.214, 3rd Generation Partnership Project, Technical Specification.
- [2] ITU-T, “End-to-end quality of service for video telephony over 4G mobile networks,” Recommendation ITU-T G.1028.1, International Telecommunication Union, 2019.
- [3] ITU-T, “Definitions of terms related to quality of service,” Recommendation ITU-T E.800, International Telecommunication Union, 2008.
- [4] S. Farthofer, M. Herlich, C. Maier, S. Pochaba, J. Lackner, and P. Dorfinger, “An open mobile communications drive test data set and its use for machine learning,” *IEEE Open Journal of the Communications Society*, vol. 3, pp. 1688–1701, 2022.

TABLE IV: Per-class evaluation — DT.

| Class               | Precision    | Recall       | F1           | Support |
|---------------------|--------------|--------------|--------------|---------|
| Excellent           | 1.000        | 0.892        | 0.943        | 65      |
| Good                | 0.823        | 0.943        | 0.879        | 296     |
| Moderate            | 0.858        | 0.753        | 0.802        | 296     |
| Poor                | 0.905        | 0.905        | 0.905        | 210     |
| <b>Weighted avg</b> | <b>0.868</b> | <b>0.865</b> | <b>0.864</b> | 867     |

- [5] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA, USA: Wadsworth, 1984.
- [6] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [7] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees,” *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.
- [8] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [9] T. M. Cover and P. E. Hart, “Nearest neighbor pattern classification,” *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [10] A. E. Hoerl and R. W. Kennard, “Ridge regression: Biased estimation for nonorthogonal problems,” *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
- [11] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [12] O. Sahin and V. Sathya, “Network quality prediction in a designated area using GPS data in LTE and 5G environments,” *Journal of Network and Computer Applications*, vol. 231, p. 104002, 2024.
- [13] S. Zeng *et al.*, “Mobile network coverage prediction using multi-modal model based on deep neural networks and semantic segmentation,” *Sensors*, vol. 24, no. 16, p. 5178, 2024.
- [14] B. Feng, M. Zheng, W. Liang, and L. Zhang, “A recent survey on radio map estimation methods for wireless networks,” *Electronics*, vol. 14, no. 8, p. 1564, 2025.
- [15] C. Yapar, F. Jaensch, R. Levie, G. Kutyniok, and G. Caire, “Overview of the first pathloss radio map prediction challenge,” *IEEE Open Journal of Signal Processing*, vol. 5, pp. 947–956, 2024.
- [16] R. Shrestha, T. N. Ha, P. Q. Viet, and D. Romero, “Radio map estimation: Empirical validation and analysis,” arXiv preprint arXiv:2310.11036, 2023.
- [17] Y. Li, C. Zhang, W. Wang, and Y. Huang, “RMTransformer: Accurate radio map construction and coverage prediction,” arXiv preprint arXiv:2501.05190, 2025.
- [18] S. Sarkar, M. H. Manshaei, and M. Krunz, “RADI-ANCE: Radio-frequency adversarial deep-learning inference for automated network coverage estimation,” arXiv preprint arXiv:2308.10584, 2023.
- [19] K. S. Gill, S. Nguyen, M. M. Thein, and A. M. Wyglinski, “Three-way deep neural network for radio frequency map generation and source localization,” arXiv preprint arXiv:2111.12175, 2021.
- [20] J. Toimela, “Spatio-temporal prediction for mobile network quality analysis,” Master’s thesis, Tampere University, 2020.
- [21] M. J. Taheri, A. Diyanat, M. A. Ahmadi, and A. Nazari, “Predicting drive test results in mobile networks using optimization techniques,” arXiv preprint arXiv:2502.09305, 2025.
- [22] Valadao, Myke, et al. “Predicting noise and user distances from spectrum sensing signals using transformer and regression models.” *Applied Sciences* 15.8 (2025): 4296.