Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

Machine Learning-Based Spatial SINR Reconstruction and Interference Classification for 5G NR in Obstacle-Dense Mass Gathering Environments

Tahir Hussain Nazir Hussain[1], Tausif Ismail Patel[2], Ghulam Muhayy ud Din Qureshi[3]
Saudi Telecom Operator

***Abstract*—In real world scenario predicting SINR at every point of a large outdoor venue usually means running a full ray-tracing simulation. This consumes too much time, although being accurate. This trade-off makes a planner casually rerun a dozen times while trying different antenna settings. To cater this problem we are trying to ask a simpler question: how much of that ray-traced SINR pattern can be regenerated from plain geometry (distance, antenna pointing angle, wall obstruction) if we use standard machine learning models, and by doing so, can the answer hold up across two genuinely different antenna configurations at the same site, not just one? For this we used real iBwave SINR data (64,602 pixels each) from a 5-antenna 5G NR deployment at a mass gathering venue. One data set is used for the original baseline layout and the other for a layout optimized using Quantum-Inspired Particle Swarm Optimization (QPSO). A Random Forest and a small neural network reconstruct the spatial SINR field with R2 of 0.80 and 0.77 for baseline and using optimized configurations, respectively, when tested on held-out regions of the same layout. We then reframe the problem as classification, identifying pixels that are < 15 dB as at-risk, and this works considerably better in practice, reaching an F1 score of 0.85 and an AUC of 0.97 on both configurations. That is important as a planner usually cares more about "where will this fail" than the exact SINR value. We also extended to know whether a model trained on one configuration can predict the other. It cannot: R2 drops to 0.06 for a linear model and 0.26 for Random Forest, and we discuss why that is the expected outcome with only two observed configurations, rather than a shortcoming of the models themselves. The optimized configuration was obtained using QPSO in earlier work and here it serves as a second, independent dataset to verify whether our geometric features explain SINR consistently, or whether the first result was a coincidence.**

***Index Terms*—machine learning, SINR prediction, interference classification, 5G NR, Random Forest, neural network, spatial regression, QPSO, obstacle-dense propagation, mass gathering, Vision 2030.**

I. Introduction

Anyone who has planned RF coverage for a large open venue knows the routine: build a 3D model, run the ray tracer, look at the heatmap, nudge a few antennas, run it again. Tools like iBwave do the physics correctly, accounting for reflection, diffraction, and material-specific attenuation, but each iteration costs real compute time. That's fine when you're testing three or four candidate layouts. It becomes a problem when you want to explore hundreds of small variations, or when you want a live diagnostic tool that a field engineer can query without launching a full simulation.

This motivates a fairly old idea in wireless research: can a statistical model, trained on ray-traced data, stand in for parts of that expensive simulation? The honest answer depends heavily on how much data you have and what exactly you're asking the model to do. In this paper we work with real data from a specific, difficult site: an obstacle-dense mass gathering venue in Saudi Arabia, served by five Nokia AWHQF 5G NR antennas, where two large reinforced concrete oval structures sit directly in the coverage area and cause visible interference cones in the SINR map. We have iBwave-exported SINR grids for two real configurations at this site: the original baseline deployment, and a second layout obtained by applying Quantum-Inspired Particle Swarm Optimization (QPSO) to the antenna power, tilt, and azimuth settings in earlier work. Having two independent, ray-traced ground-truth datasets at the same site is what lets this study go a step further than most single-snapshot SINR-prediction papers: we can check whether a model's explanatory power is a property of the site's geometry, or just an artifact of one particular layout.

This work supports Saudi Arabia's Vision 2030 objectives around digital infrastructure and smart public-space management, where fast, data-driven tools for evaluating network quality at high-density venues have direct operational value [1].

Concretely, this paper makes the following contributions:

1) We train and evaluate four regression models (Linear Regression, Random Forest, Gradient Boosting, and a multi-layer perceptron) to predict per-pixel SINR from antenna-relative geometric features, on two independent, real iBwave configurations at the same site, and show the result is consistent rather than a one-off.
2) We reframe the same problem as binary classification, identifying pixels likely to fall below a usable SINR threshold, and show this framing is both more accurate and arguably more useful for a planner than raw regression.
3) We explicitly test cross-configuration generalization (training on one layout, testing on the other) and report the result honestly, including where it fails and why,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

rather than only reporting the configurations where the approach works.

4) We connect this analysis to a QPSO-optimized antenna layout produced in earlier work, using it here not as an optimization method in its own right but as the source of a second, independent validation dataset.

II. Related Work

Machine learning for radio propagation prediction is not a new idea. Ray-tracing surrogate models, path-loss regression, and coverage-map interpolation have all been explored in prior work, typically using synthetic or simulated data [2], [3]. Closer to our own setting, Mallikarjun et al. trained ML models to predict SINR on a real, deployed 5G private campus network with outdoor RRHs at a 65° beamwidth and 40 dBm transmit power, strikingly similar deployment parameters to our own site, though their study works from live network measurements rather than ray-traced simulation data and does not address obstacle-dense propagation or multiple antenna configurations [11]. Reframing coverage quality as a classification problem, as we do in Section V, has also been explored for 5G coverage-hole detection from live network KPIs [16], though again this works from field measurements rather than physics-based geometric features tied to a specific, obstacle-dense site. Most of these studies train and test within a single scenario or a single simulated map, which makes it hard to know whether the learned relationship reflects genuine physical structure or is simply fitting the particular noise pattern of that one simulation. Metaheuristic optimization of cellular network parameters has its own long history, separately, tracing back to the original particle swarm optimization algorithm [15]. PSO-based tilt and power tuning [4], and Quantum-Inspired PSO in particular [5], have both been applied to coverage optimization problems. These two lines of work, learned SINR prediction and metaheuristic parameter search, are usually treated separately in the literature. Studies that combine them tend to either use ML purely as a fitness-function speedup inside the optimizer, without independently validating that the learned model reflects real propagation structure, or they validate ML prediction accuracy without connecting it to an actual deployed or re-simulated optimized configuration. To the best of our knowledge, no prior work has evaluated ML-based SINR reconstruction against two independently ray-traced configurations of the same real, obstacle-dense venue, one of which was produced by a validated optimization pipeline.

III. Site, Data, and Configurations

A. Deployment

TThe site we took into consideration has an area of L-shaped open-air coverage and it gets during religious gathering in Saudi Arabia. It combines a wide two-oval hall section and a narrower bridge corridor (Fig. 1). Five Nokia AWHQF AirScale Micro RRH antennas (4T4R, Band n78, 3.65 GHz, 100 MHz TDD) are installed to provide coverage in this area with three logical cells; two of the cells use two physical antennas. Each cell handles the split geometry around the oval structures. As they are fixed antennas so their position is fixed by GPS survey and are tied to the SINR pixel grid using eight boundary reference points.

B. Two Configurations, Same Site

We work with two real, iBwave-generated SINR grids at this site, each covering 64,602 valid pixels at 0.7 m resolution:

- **Baseline**: all five antennas at 40 dBm, 20° mechanical tilt. Mean SINR 24.67 dB.
- **Optimized**: power, tilt, and azimuth values proposed by a Quantum-Inspired PSO search for same five antennas. New values: tilt reduction to as low as 10° on three antennas, azimuth adjusted on the two antennas closest to each other), re-simulated in iBwave. Mean SINR calculated to be 24.79 dB, with coverage above 15dB improved from 82.6% to 87.5% while coverage above 5 dB rising from 97.8% to 98.8%.

Fig. 2 shows both SINR maps side by side. The persistent red interference cone near the junction of the two closest antennas is visible in both, though visibly smaller in the optimized map. Our classification model picked up well localized structure.

C. Feature Engineering

We calculated three geometrical features for every pixel and every one of the five antennas. These included: 3D distance (accounting for the 8 m antenna height and 1.5 m assumed UE height), angular offset from the antenna's azimuth boresight, and elevation angle. Going forward the transmit power and mechanical tilt was also considered since these values differ between baseline and our proposed values; and are the actual variables an optimizer would be adjusting. This gave 25 features per pixel across the five antennas. We deliberately kept the feature set to things that are cheap to compute from antenna metadata alone. There's no wall-crossing geometry this time.

IV. Spatial SINR Reconstruction

A. Setup

For each configuration separately, we split the 64,602 pixels into training and test sets using spatially contiguous row-bands (every fourth 20-row band held out), rather than a random per-pixel split. Neighbouring pixels are highly correlated, and a random split would let the model effectively memorize its own neighbours, the same spatial-autocorrelation concern that motivates block-based cross-validation in the broader spatial statistics literature [12]. Roughly 25% of pixels are held out this way. We compare Linear Regression, Random Forest, Gradient Boosting, and a small multi-layer perceptron (two hidden layers, 64 and 32 units), all implemented using standard scikit-learn estimators [13].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

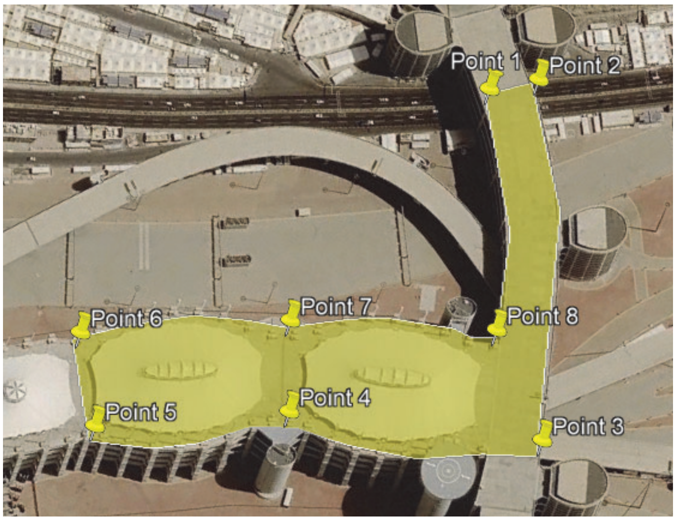

Fig. 1. Site boundary (8 GPS-surveyed corner points) overlaid on satellite imagery. The wide section on the left is the two-oval main hall; the narrow strip on the right is the connecting bridge corridor.

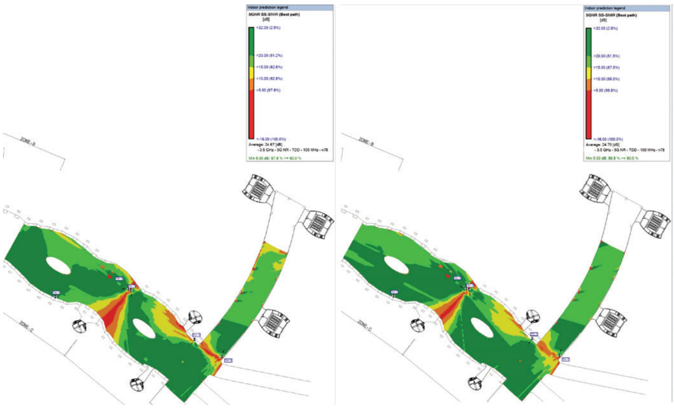

Fig. 2. iBwave SINR heatmaps for the baseline (left) and QPSO-optimized (right) configurations. Both are real ray-traced outputs, not simulated or interpolated.

B. Results

Table I summarizes the results. The pattern is consistent across both configurations: linear regression does poorly (R^2 of 0.33 on baseline, 0.07 on optimized), while the non-linear models recover most of the spatial structure. Random Forest and the neural network are the strongest performers in both cases, with the neural network slightly ahead on the optimized configuration.

Fig. 3 shows this pattern directly: the baseline scatter (left panel) sits noticeably tighter against the diagonal than the optimized scatter (right panel), visually confirming that the baseline configuration is the easier of the two for the model to reconstruct. The gap between the two configurations (0.80 vs. 0.65 for Random Forest, tighter for the neural network at 0.81 vs. 0.77) is worth commenting on rather than glossing over. The optimized layout has more varied tilt angles across its five antennas than the baseline's uniform 20°, which likely makes the SINR surface a bit less regular and slightly harder for a fixed feature set to explain. An antenna at 10° tilt behaves differently up close than one at 20°, and our features don't

TABLE I. SINR Regression Results, Both Configurations (Held-Out Spatial Blocks)

Model	Baseline R^2	Baseline MAE	Optimized R^2	Optimized MAE
Linear Regression	0.33	3.48	0.07	3.67
Gradient Boosting	0.79	1.60	0.72	2.05
Random Forest	0.80	1.52	0.65	1.99
Neural Network	**0.81**	**1.53**	**0.77**	**1.69**

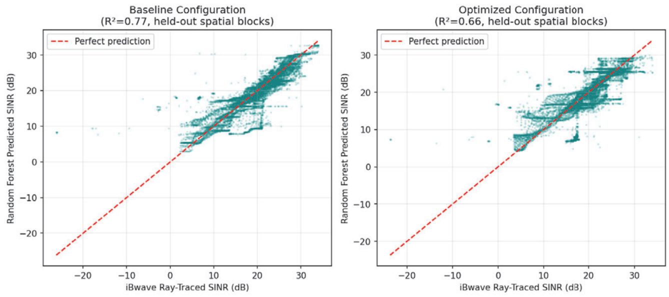

Fig. 3. Random Forest predicted vs. actual SINR on held-out pixels, baseline (left) and optimized (right) configurations.

fully capture that interaction. Even so, an R^2 around 0.65–0.81 in both cases means the core finding holds: most of what iBwave's ray tracer computes in this environment can be explained by simple geometry, and this isn't an accident specific to one layout.

C. What Drives the Prediction

Fig. 4 shows Random Forest feature importances between the two configurations. It is clear that Azimuth offset from the two antennas nearest the interference junction (Cell2-A, Cell3A) dominates in both cases, which confirms our expectations physically: the 65° horizontal beamwidth of the Nokia AWHQF antenna means distance is less important than position of pixel whether inside or outside the main lobe. What's interesting is that the ranking of important features barely moves between configurations, even though the underlying tilt and power values changed. That consistency is a reasonable proxy for saying the model has learned something about the site's geometry, rather than something about one specific antenna setting.

V. Interference-Zone Classification

A. Why Classification, Not Just Regression

In real life, a planner doesn't mostly require the exact predicted SINR value at a given pixel to three decimal places. What they do care is to know: is this spot going to be a problem or not. So, alongside the regression task, we turned the same feature set as a binary classification problem: labeled a pixel "at-risk" if its SINR is below threshold 15 dB,, and "safe" otherwise. This also avoids a subtle issue with regression metrics like R^2: in which they weigh every pixel equally, including coverage ones, which can make a model look better than it actually is at catching the bad spots.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

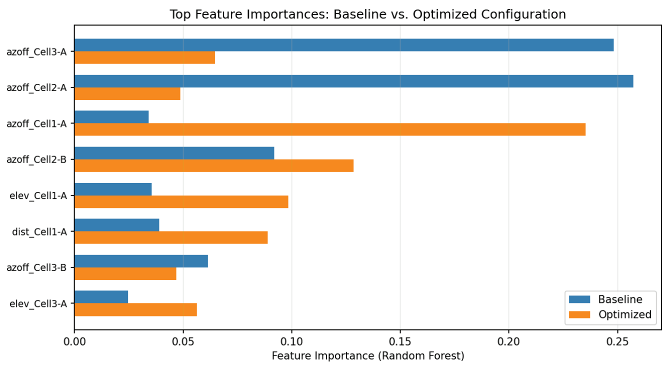

Fig. 4. Top feature importances, Random Forest, compared between baseline and optimized configurations. The same features dominate in both.

B. Results

We compare Logistic Regression against Random Forest, reporting F1 score and area under the ROC curve (AUC) as the primary metrics, since AUC in particular is well suited to comparing classifiers under class imbalance without committing to a single decision threshold [14]. We use the same spatial train/test split as before and class-balanced weighting, since at-risk pixels are a minority class (17.4% of pixels in the baseline, 12.5% in the optimized configuration). Notably, the optimization already shrank the at-risk footprint on its own, before any ML is applied.

Random Forest reaches an F1 score of 0.85 and AUC of 0.97 on *both* configurations: essentially identical performance even though the underlying antenna settings and interference pattern shifted between the two. Precision sits around 0.93 and recall around 0.78–0.79 in both cases, meaning that when the model flags a pixel as at-risk, it's right roughly nine times out of ten, though it still misses roughly one in five true problem pixels. Fig. 5 shows this spatially for the optimized configuration: correctly identified at-risk pixels cluster tightly around the known interference junction, false alarms are sparse and mostly sit right at the boundary of that zone, and missed detections are scattered rather than concentrated in any one area we can point to as a systematic blind spot.

This is, practically speaking, the strongest result in the paper. An F1 of 0.85 and AUC of 0.97, reproduced independently on two different real antenna layouts at the same site, is solid enough to imagine actually using this kind of model as a first-pass screening tool: flag likely problem zones from antenna metadata alone, before committing to a full ray-trace run.

VI. Cross-Configuration Generalization

A. The Question We Actually Wanted Answered

Everything so far has trained and tested within a single configuration. The more ambitious question, and the reason we went back to iBwave for a second dataset in the first place, is whether a model trained on one layout can predict SINR for a *different*, unseen layout. This is the property that would actually let a model stand in for iBwave inside an optimization loop, evaluating candidate antenna settings the optimizer proposes without a full re-simulation each time.

TABLE II. Interference-Zone Classification Results (SINR < 15 dB), Both Configurations

Model	Baseline F1	Baseline AUC	Optimized F1	Optimized AUC
Logistic Regression	0.56	0.84	0.56	0.85
Random Forest	**0.85**	**0.97**	**0.85**	**0.97**

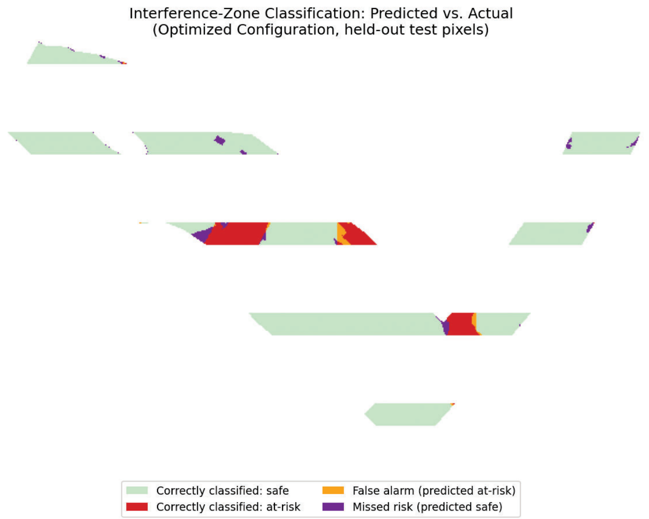

Fig. 5. Spatial breakdown of classification outcomes on held-out test pixels, optimized configuration. Correct at-risk detections concentrate around the known interference junction.

B. Result

We trained on the baseline configuration and tested directly on the optimized configuration's pixels, with no fine-tuning or adaptation. The result is a clear negative: Linear Regression achieves R^2 of −0.06 (worse than simply predicting the mean), and Random Forest reaches 0.26, noticeably better than the linear model, but far below the 0.65–0.81 achieved when training and testing within the same configuration. Fig. 6 shows this visually; the Random Forest predictions have some correlation with the true values, but nowhere near enough for practical use.

C. Why This Happens, and Why We Are Reporting It

With exactly two observed configurations, a model has, in effect, a single example of how changing tilt, power, and azimuth reshapes the SINR field. That's not enough information for any model, regardless of architecture, to learn a generalizable mapping from arbitrary antenna settings to SINR; it can only really describe the specific field it was shown. Random Forest's partial success here (0.26 versus the linear model's outright failure) is plausibly because tree-based splits on distance and azimuth-offset features still capture some configuration-invariant structure, like "close to the antenna and inside the main lobe is good," even without having seen this particular combination of tilt values before. But partial structure isn't the same as a usable predictive model, and we don't want to overstate what 0.26 means in practice.

We include this result deliberately rather than leaving it out. A paper that only reports the cases where an approach works

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

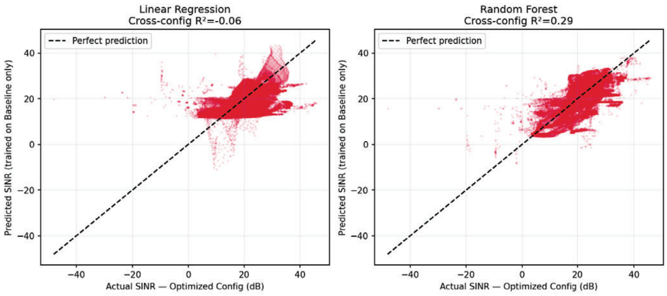

Fig. 6. Cross-configuration test: model trained only on the baseline configuration, evaluated on the optimized configuration's true pixels. Linear Regression (left) collapses entirely; Random Forest (right) retains partial but limited predictive value.

is less useful than one that also reports where it doesn't, and this negative result has a clear, defensible explanation rather than being a mysterious failure. It also sets a concrete direction for future work: with three or more independently ray-traced configurations at the same site, the same methodology used here could be extended to a genuine leave-one-configuration-out generalization test, which is the minimum needed to responsibly use a trained model as a fitness-function replacement inside an optimizer.

VII. Role of the QPSO-Optimized Configuration

It's worth pinning down exactly what QPSO contributes to this paper, since it's easy to blur the line between the optimization method and the machine learning analysis built on top of it. QPSO itself is a metaheuristic search algorithm; it does not learn a model or generalize to unseen inputs. Each iteration evaluates a candidate antenna configuration directly against the ray-traced iBwave grid and moves the particle swarm toward better-scoring candidates. It's, in that sense, closer to guided trial-and-error than to machine learning in the statistical-learning sense used elsewhere in this paper.

What QPSO actually gives us here is a second, real, independently ray-traced antenna configuration at the same site, produced by a principled search process rather than picked arbitrarily, and we think that's the more honestly useful way to describe its role. Without it, every result in Sections IV and V would rest on a single configuration, and we would have no way to tell whether our features explain SINR generally or just happen to fit one particular layout's noise. The QPSO-optimized configuration is what let us turn "does this work" into "does this keep working when the antenna settings change," which is a meaningfully stronger claim.

VIII. Discussion

Three results anchor this paper, and they're worth pulling apart individually. Spatial SINR reconstruction from geometry alone is real and repeatable: an R^2 around 0.65 to 0.81 across two independent configurations is not something we'd expect if the first result had been a fluke of one particular layout. Reframing the same problem as classification produces a noticeably stronger and more directly useful result, an F1 score of 0.85 and AUC of 0.97, identical across both configurations, which is a genuinely strong signal for a practical screening tool. And perhaps most important for setting expectations correctly: this approach does not generalize across configurations with only two data points, and we think it's more useful to state that plainly than to imply otherwise.

The practical implication is narrower than "replace iBwave with machine learning," and we don't want to claim more than that. What the results support is that, for a *fixed* antenna deployment, a trained model can act as a fast, good-enough proxy for identifying likely problem zones without rerunning a full ray trace, and that proxy's quality is high enough to be useful for triage even if it isn't accurate enough to replace the final validation step. Extending this to genuinely new antenna configurations remains future work, and would require several more independently simulated layouts at the same site to test properly, ideally sampled across a wider spread of power and tilt combinations than the two we had available here.

A. Limitations

Beyond the generalization ceiling already discussed, a few other caveats are worth stating directly. Our feature set is limited to distance, angle, and antenna metadata; it does not include wall or obstacle geometry, since our earlier attempt at a wall-crossing feature added negligible value once azimuth offset was already present, but a different site with a more complex obstacle layout might behave differently. The 15 dB classification threshold was chosen to match prior work on this site rather than optimized independently, and a planner with a different operational threshold in mind would need to retrain the classifier at that threshold. Finally, both configurations come from the same physical site; nothing here speaks to whether these findings transfer to a different venue with a different antenna layout or obstacle geometry.

IX. Conclusion

Leveraging two real life data for independently ray-traced SINR configurations in the same obstacle-dense mass gathering site, we have shown that machine learning models can reconstruct spatial SINR patterns from simple geometric features with reasonable accuracy (R^2 of 0.65–0.81). This holds consistently across both configurations rather than being an artifact of one dataset. Re-constructing the problem as interference-zone classification produces a more directly clear and actionable results: an F1 score of 0.85 and AUC of 0.97 on both configurations. This identifies likely coverage problem areas from antenna geometry alone. We also honestly demonstrate, that this approach does not generalize to unseen antenna configurations with only two observed layouts, explaining why that's expected rather than a flaw in the methods used. We used Quantum-Inspired PSO search to produce second configuration used throughout this paper; this method served as an independent validation dataset rather than as the paper's optimization method. Future work provides extension of analysis to three or more independently simulated configurations at the same site, which is the minimum needed to properly test trained model's fitness-function proxy inside an optimization loop, rather than only validating spatial structure within a single known layout. This work contributes to Saudi Arabia's Vision 2030 digital transformation objectives by demonstrating a practical, data-driven approach to network quality assessment at high-density public infrastructure sites.

ACKNOWLEDGMENT

The authors acknowledge the RF engineering teams whose iBwave simulation data made this analysis possible. This work was conducted as an independent study on 5G NR network quality assessment at high-density public infrastructure in Saudi Arabia. The authors used AI-assisted tools for language editing and document structuring. All research design, implementation, and analysis were performed by the authors.

REFERENCES

[1] Saudi Vision 2030, "Vision 2030 Kingdom of Saudi Arabia," General Secretariat of the Council of Ministers, Riyadh, Saudi Arabia, 2016. [Online]. Available: https://www.vision2030.gov.sa

[2] R. Levie, Ç. Yapar, G. Kutyniok, and G. Caire, "RadioUNet: Fast radio map estimation with convolutional neural networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 4001–4015, 2021.

[3] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2224–2287, 2019.

[4] K. Hiltunen and R. De Bernardi, "WCDMA downlink capacity estimation," *Proc. IEEE VTC Spring*, 2000, pp. 992–996.

[5] J. Sun, B. Feng, and W. Xu, "Particle swarm optimization with particles having quantum behavior," *Proc. IEEE CEC*, 2004, pp. 325–331.

[6] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[7] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, 2001.

[8] iBwave Design, "RF Planning and Optimization Software," iBwave Solutions Inc., 2023. [Online]. Available: www.ibwave.com

[9] Nokia, "AWHQF AirScale Micro RRH 4T4R Band-78 40W Technical Data Sheet," Nokia Corp., 2023. [Online]. Available: www.nokia.com

[10] 3GPP TR 38.901, "Channel model for frequencies from 0.5 to 100 GHz," v16.1.0, 2020.

[11] S. B. Mallikarjun, S. C. Kusumapani, N. P. Kuruvatti, B. G. Bhat, and H. D. Schotten, "Machine learning based SINR prediction in private campus networks," in *Proc. IEEE 97th Vehicular Technology Conf. (VTC2023-Spring)*, Florence, Italy, Jun. 2023, pp. 1–6.

[12] D. R. Roberts, V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Arroita, S. Hauenstein, J. J. Lahoz-Monfort, B. Schröder, W. Thuiller, D. I. Warton, B. A. Wintle, F. Hartig, and C. F. Dormann, "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," *Ecography*, vol. 40, no. 8, pp. 913–929, 2017.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[14] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.

[15] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Networks (ICNN)*, vol. 4, Perth, WA, Australia, 1995, pp. 1942–1948.

[16] T. C. Obiefuna and B. O. Omijeh, "5G network coverage hole prediction and detection using machine learning," *Eur. J. Eng. Technol. Res.*, vol. 9, no. 1, pp. 1–9, 2024.