

Reliability-Aware SDR Beamforming for Underground Environments: A Laboratory Study

Marta Sieradzka, Adam Michalak, Piotr Woźniak, Bartosz Berezowski, Krzysztof Cichon
 Science Club Spektrum, Poznan University of Technology
 Polanka 3, 60-965 Poznan, Poland
 Email: krzysztof.cichon@put.poznan.pl

Abstract—This paper presents a laboratory feasibility study of a reliability-assessment layer intended for direction-aware mobile MIMO beamforming in future multipath-rich underground deployments. A centrally controlled software-defined radio system performs repeated phase-commanded sweeps and records angular power profiles from three receive channels. Eight interpretable descriptors quantify angular selectivity, receiver agreement, temporal stability, peak separation, main-lobe width, and profile asymmetry. A Random Forest classifier maps these descriptors to an operational profile label. In this paper, reliability denotes profile-level consistency rather than verified beam direction or communication-link quality. The classification dataset contains 16 comparable physical three-receiver cases, including 7 reliable and 9 ambiguous profiles; eight additional single-receiver analyzer measurements follow a different acquisition protocol and are not used for classification. No synthetic cases are used in the reported results. Leave-one-session-out evaluation yields 68.8% accuracy, 65.9% balanced accuracy, a macro F1-score of 65.4%, and an ROC-AUC of 63.5% for Random Forest. These exploratory results support the feasibility of the processing chain but do not establish generalization to actual tunnels, mines, or unseen hardware configurations.

Index Terms—mobile MIMO, beamforming, software-defined radio, reliability assessment, machine learning, multipath propagation, laboratory feasibility study, underground communications

I. INTRODUCTION

Underground radio environments are challenging for communication, localization, and sensing because tunnels, technical galleries, and mines can exhibit strong reflections, diffraction, shadowing, and waveguide-like propagation [1]. These effects may produce angular power profiles with several local maxima, broad responses, or receiver-dependent dominant directions. Tunnel geometry and repeated reflections can therefore make a maximum-power steering rule difficult to interpret: the strongest measured state need not be the most stable or mutually consistent state across receivers.

Mobile MIMO and distributed software-defined radio (SDR) systems provide a flexible platform for adapting spatial transmission and reception characteristics [6]. Beamforming concentrates transmitted or received energy by applying controlled phase relations across array branches. In the present system, direction awareness is represented by a finite grid of commanded steering states. The controller therefore needs not only a maximum-power state, but also an indication of whether the measured angular response is sufficiently distinct

and repeatable to be used without an immediate verification sweep.

Reliable underground connectivity supports personnel safety, operational coordination, monitoring, and positioning in mines and transportation tunnels. In contrast to open environments, radio propagation in confined underground structures is strongly influenced by tunnel cross-section, wall properties, carrier frequency, antenna polarization, and transmitter–receiver geometry. Surveys and measurement campaigns report a combination of multipath and waveguide-like behavior, with attenuation and field structure depending on the particular underground environment [2]–[4]. These characteristics make underground communication a distinct propagation problem rather than a direct extension of conventional open-space wireless links.

For MIMO and direction-aware transmission, the spatial structure of this field is especially relevant. Experimental investigations in arched tunnels have shown that field statistics and spatial correlation vary with propagation distance and can materially affect MIMO channel behavior [5]. Consequently, a pronounced maximum in an angular sweep does not by itself establish that a steering state is sufficiently distinct, repeatable, or mutually consistent across spatially separated observations. This motivates the profile-level assessment considered here, which combines angular shape, temporal repeatability, and cross-receiver consistency instead of relying only on the largest measured power value.

The target application is reliability-aware beamforming in multipath-rich underground environments. The measurements reported here were nevertheless collected entirely under laboratory conditions. This paper is therefore a feasibility study of the measurement-to-decision processing chain, not validation in a tunnel or mine.

The main contributions are:

- a centralized SDR measurement and decision architecture for repeated multi-receiver angular sweeps;
- a precise set of eight interpretable profile descriptors;
- a leakage-resistant evaluation using only comparable physical three-receiver cases and leave-one-session-out validation;
- quantitative comparison with a majority baseline, single-feature threshold rules, Logistic Regression, and a linear Support Vector Machine (SVM);

- an exploratory robustness analysis covering model randomness, feature ablation, descriptor correlation, and statistical comparison with the strongest rule-based baseline.

The remainder of this paper is organized as follows. Section II reviews related work on underground propagation and machine-learning-assisted beamforming. Section III describes the SDR system model and phase-commanded measurement procedure. Section IV presents the laboratory dataset and operational labelling procedure, while Section V defines the angular-profile descriptors. Section VI describes the evaluation protocol, and Section VII presents the classification, uncertainty, and feature-sensitivity results. Section VIII summarizes the principal limitations, and Section IX concludes the paper.

II. RELATED WORK

Underground propagation has been investigated through analytical channel modelling for mines and road tunnels [1], broad surveys of underground-mine wireless communication [2], extensive RF measurements in mines and tunnels [3], experimental MIMO propagation studies in arched tunnels [5], [6], and distributed-antenna measurements in metro tunnels [7]. A recent review also summarizes tunnel communication technologies and their deployment constraints [4]. These studies motivate the intended underground application, but the present paper does not attempt to reproduce their propagation environments; its measurements remain laboratory based.

Machine-learning approaches have also been applied to beam selection and coordinated beamforming [8], [10]–[12]. Related work has additionally considered the security of AI-driven beam-selection mechanisms [9]. In contrast, the present prototype does not predict an optimal beam index or a communication-level metric. It screens measured angular profiles for operational consistency before any downstream controller action.

III. SYSTEM MODEL

A. Phase-Commanded Transmit Sweep

The transmitting subsystem comprises four SDR-based branches, denoted TX0–TX3. For steering state k , transmitter m generates

$$x_{m,k}[n] = a \exp \left(j \left(2\pi f_{\text{bb}} \frac{n}{f_s} + \phi_{m,k} \right) \right), \quad (1)$$

where a is the normalized amplitude, f_{bb} is the baseband tone frequency, and f_s is the sampling frequency. The phase command is

$$\phi_{m,k} = m\Delta\phi_k, \quad (2)$$

with

$$\Delta\phi_k = -\frac{2\pi d}{\lambda} \sin(\theta_k). \quad (3)$$

The laboratory prototype used $d = \lambda/2$ and 25 nominal steering states

$$\Theta = \{-60^\circ, -55^\circ, \dots, 55^\circ, 60^\circ\}. \quad (4)$$

The nominal steering angle is a software command used to generate phase states; it is not a direct measurement of angle of arrival or departure.

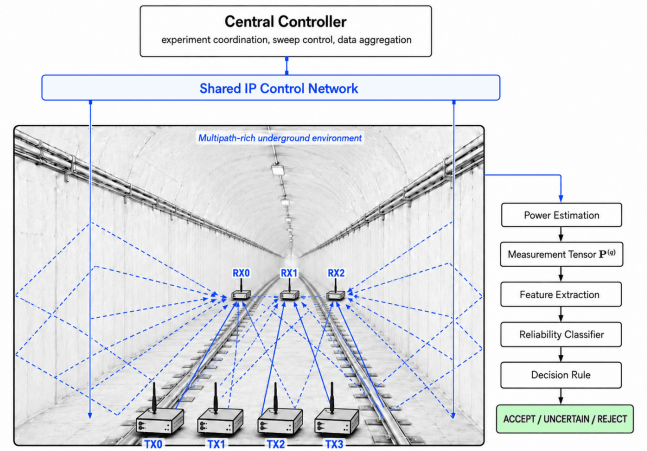


Fig. 1. Conceptual target deployment and reliability-processing architecture. The tunnel scene illustrates the intended application and is not the environment used for the reported measurements.

B. Multi-Receiver Measurement Tensor

For receiver $j \in \{0, 1, 2\}$, steering state k , and sweep repetition s , the average linear power is estimated from a block of complex baseband samples as

$$\bar{p}_{j,k,s} = \frac{1}{N_{j,k,s}} \sum_{n=0}^{N_{j,k,s}-1} |y_{j,k,s}[n]|^2. \quad (5)$$

A labelled measurement case q contains the complete tensor

$$\mathcal{P}^{(q)} = \left[P_{j,k,s}^{(q)} \right] \in \mathbb{R}^{3 \times 25 \times S_q}, \quad (6)$$

where $P_{j,k,s}^{(q)}$ is the logarithmic representation of the measured power. The complete angular response, its repetition across slots, and its consistency across receivers are used in the profile assessment. In the available three-receiver dataset, S_q ranges from three to five complete sweeps per case; every complete sweep contains all 25 steering states for RX0–RX2.

Fig. 1 shows the target deployment concept and the processing chain. The tunnel graphic is illustrative; the evaluation in this paper used laboratory measurements.

C. Decision-Support Flow

The central controller waits until the transmit and receive components report readiness, executes the repeated sweep, assembles $\mathcal{P}^{(q)}$, and passes it to the reliability layer. Fig. 2 summarizes this procedure.

The classifier returns a predicted class $\hat{y}^{(q)} \in \{\text{reliable}, \text{ambiguous}\}$ and a model score $c^{(q)}$. A deterministic downstream mapping may convert the output to

$$\mathcal{D} = \{\text{ACCEPT}, \text{UNCERTAIN}, \text{REJECT}\}. \quad (7)$$

The score is not treated as a calibrated probability, and the three-way mapping is shown only as a controller interface. Decision thresholds are deployment parameters; they are neither optimized nor operationally validated in the present dataset.

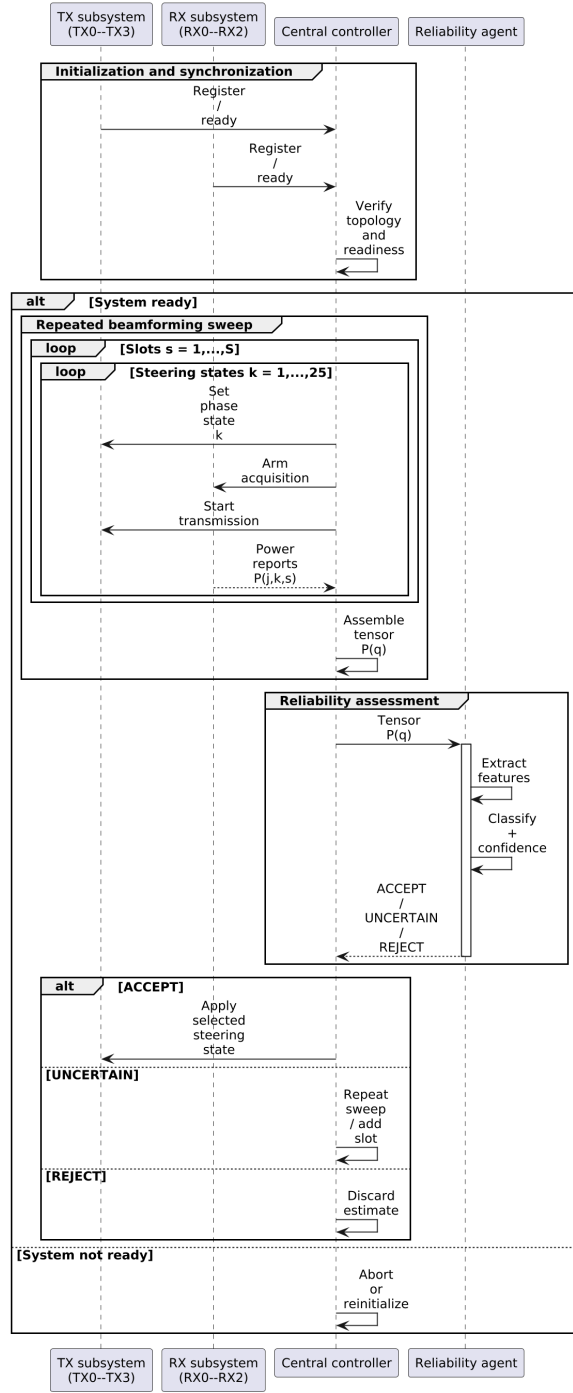


Fig. 2. Coordinated measurement cycle and separation between acquisition, profile inference, and deterministic radio control.

IV. LABORATORY DATASET

A. Case Construction and Labels

The classification dataset consists of 16 physical three-receiver measurement cases: 7 labelled *reliable* and 9 labelled *ambiguous*. Each case contains three to five repeated sweeps from -60° to 60° with a 5° step. An additional set of eight single-receiver analyzer measurements was collected under a

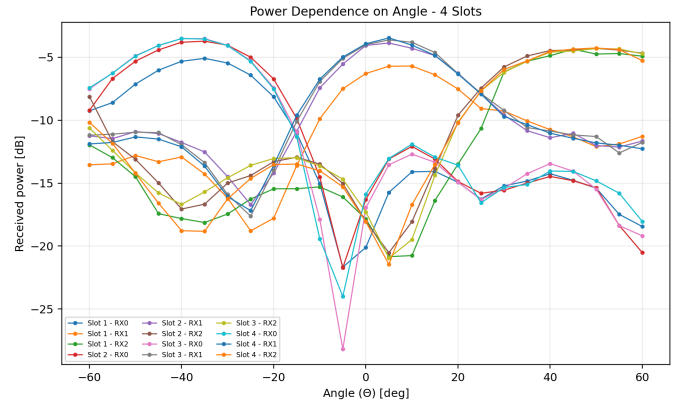


Fig. 3. Example laboratory angular response supplied with the label *reliable*; four complete slots are present in this case.

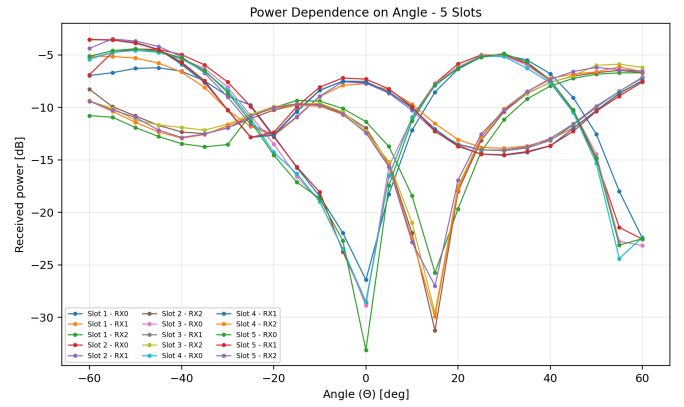


Fig. 4. Example laboratory angular response supplied with the label *ambiguous*; five complete slots are present in this case.

different acquisition protocol. All eight carry a reliable label, but they are not part of the classification dataset because the cross-receiver descriptor is undefined and acquisition mode would otherwise be confounded with class.

The labels are operational rather than geometrical. According to the supplied labelling procedure, a profile was labelled *reliable* when visual inspection showed a distinct, sharp response that remained stable across repeated slots. Flat, multi-peak, or chaotic profiles were labelled *ambiguous*; the broader system description also treats disagreement between receiver indications as an ambiguity indicator. The archive provides one supplied label per case. No independent geometric reference, blinded relabelling, multiple annotators, or inter-rater agreement measurement is available. Consequently, the classifier should be interpreted as reproducing this operational profile distinction, not as establishing the true propagation path or communication-link reliability.

Figs. 3 and 4 show examples from the supplied labelled data. These figures illustrate the operational distinction; they do not constitute geometrical proof of a direct or reflected propagation path.

V. RELIABILITY FEATURES

For each case, receiver powers are converted to the linear domain and averaged across the three receivers and all available slots. The resulting aggregate profile is converted back to decibels and denoted $\bar{P}(\theta_k)$. The feature vector is

$$\mathbf{x} = [P_{\text{peak}}, S_{\text{peak}}, \sigma_P, A_{\text{RX}}, S_{\text{slot}}, R_{\text{PSL}}, W_{\text{ang}}, B_{\text{asym}}]. \quad (8)$$

The features are defined as follows:

- P_{peak} : maximum value of $\bar{P}(\theta_k)$;
- S_{peak} : difference between the maximum and the mean of $\bar{P}(\theta_k)$;
- σ_P : standard deviation of $\bar{P}(\theta_k)$ across steering states;
- A_{RX} : mean, over slots, of the standard deviation of the three receiver peak angles;
- S_{slot} : standard deviation of the aggregate-profile peak power across slots;
- R_{PSL} : difference between the main peak and the maximum outside the contiguous -3 dB main-lobe region;
- W_{ang} : angular width of the contiguous -3 dB region containing the main peak;
- B_{asym} : normalized linear-power imbalance on the left and right sides of the main peak,

$$B_{\text{asym}} = \frac{|P_{\text{left}} - P_{\text{right}}|}{P_{\text{left}} + P_{\text{right}}}. \quad (9)$$

These definitions correct two ambiguities in the prototype implementation: asymmetry is calculated in the linear-power domain, and the sidelobe is selected outside the contiguous main-lobe region rather than as the adjacent angular sample.

VI. EVALUATION PROTOCOL

The primary classifier is a Random Forest with 300 trees and class weighting [13]. The tree count and remaining classifier settings are treated as fixed implementation choices; no hyperparameter search is performed on the 16 cases. Feature-level Gaussian augmentation is not used because independent perturbation does not preserve measured correlations and may generate physically inconsistent feature combinations.

To reduce leakage between measurements collected under the same conditions, evaluation is performed with leave-one-session-out cross-validation. Each laboratory session is held out once, and all cases from that session are predicted by a model trained on the remaining sessions. Thus, each physical case receives exactly one out-of-fold prediction.

Random Forest is compared with: (i) a majority-class predictor, (ii) a one-level decision tree using only A_{RX} , (iii) a one-level decision tree using only R_{PSL} , (iv) class-weighted Logistic Regression, and (v) a class-weighted linear SVM. Thresholds for the one-feature rules are learned only from the training sessions in each fold. Balanced accuracy is the mean recall across the two classes, macro F1 is the unweighted mean of the two class-wise F1-scores, and ROC-AUC summarizes ranking performance over score thresholds.

Because the dataset is small, three additional exploratory checks are reported rather than used for model selection: (i) the 300-tree Random Forest is repeated over 50 random

TABLE I
LEAVE-ONE-SESSION-OUT PERFORMANCE ON 16 PHYSICAL
THREE-RECEIVER CASES.

Method	Accuracy	Balanced acc.	Macro F1	ROC-AUC
Majority class	0.563	0.500	0.360	—
RX-agreement rule	0.375	0.397	0.365	0.492
PSLR rule	0.563	0.563	0.561	0.540
Logistic Regression	0.438	0.452	0.435	0.508
Linear SVM	0.438	0.452	0.435	0.492
Random Forest	0.688	0.659	0.654	0.635

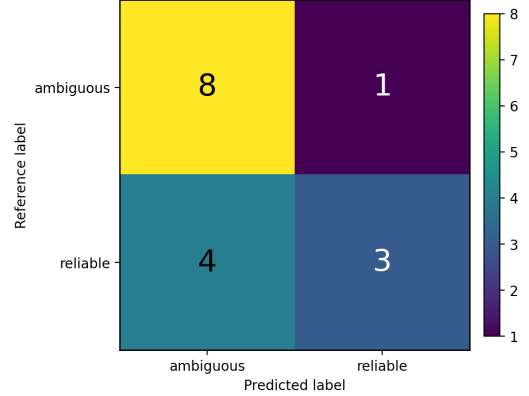


Fig. 5. Out-of-fold Random Forest confusion matrix for leave-one-session-out validation.

seeds to quantify sensitivity to ensemble randomness; (ii) each descriptor is removed once in a leave-one-feature-out ablation; and (iii) the Random Forest and the strongest single-feature baseline are compared with an exact McNemar test on their paired out-of-fold predictions. Pairwise Pearson correlations are also inspected to identify descriptor redundancy.

VII. RESULTS

Table I presents the leave-one-session-out results. Random Forest obtains the highest numerical balanced accuracy and macro F1-score among the evaluated methods. However, the absolute performance remains modest and the sample count is too small for a strong generalization claim.

The Random Forest confusion matrix in Fig. 5 contains eight correctly identified ambiguous cases, one ambiguous case classified as reliable, three correctly identified reliable cases, and four reliable cases classified as ambiguous. This corresponds to an ambiguous-class precision of 66.7%, recall of 88.9%, and F1-score of 76.2%. For the reliable class, precision is 75.0%, recall is 42.9%, and F1-score is 54.5%.

The small case count produces wide uncertainty. For 11 correct predictions out of 16, the 95% Wilson interval for accuracy is 44.4–85.8%. The corresponding intervals for ambiguous-class recall (8/9) and reliable-class recall (3/7) are 56.5–98.0% and 15.8–75.0%, respectively. The Random Forest is correct on five cases where the PSLR rule is wrong, whereas the PSLR rule is correct on three cases where Random Forest is wrong. An exact McNemar test gives $p = 0.727$, so

TABLE II
EXPLORATORY LEAVE-ONE-FEATURE-OUT RANDOM FOREST ABLATION.

Omitted descriptor	Balanced acc.	Macro F1
None	0.659	0.654
P_{peak}	0.659	0.654
S_{peak}	0.730	0.733
σ_P	0.492	0.492
A_{RX}	0.659	0.654
S_{slot}	0.659	0.654
R_{PSL}	0.603	0.600
W_{ang}	0.675	0.676
B_{asym}	0.730	0.733

the observed numerical advantage of Random Forest is not statistically established on this dataset.

At fixed 300 trees, repeating only the Random Forest seed 50 times gives mean accuracy 0.701 ± 0.042 (standard deviation across seeds; range 0.563–0.750) and mean balanced accuracy 0.677 ± 0.044 (range 0.548–0.730). This sensitivity confirms that individual percentage-point differences should not be over-interpreted at the current sample size.

The ablation results in Table II are exploratory rather than a feature-selection procedure. Removing σ_P causes the largest decrease in balanced accuracy, while removing S_{peak} or B_{asym} increases the score in this particular small sample. The descriptors are also not independent: the strongest pairwise Pearson correlations are $r = 0.773$ for S_{peak} and σ_P , $r = 0.619$ for S_{peak} and R_{PSL} , and $r = -0.583$ for S_{peak} and W_{ang} . These observations reinforce the need for a larger dataset before assigning stable physical importance to individual descriptors.

The high recall for ambiguous profiles and low recall for reliable profiles indicates a false-rejection tendency: the model rejects many profiles carrying a reliable reference label. The present measurements contain no application-level cost model, so this trade-off cannot be claimed as beneficial. The model score is also not probability-calibrated, and the ACCEPT/UNCERTAIN/REJECT interface remains conceptual rather than a validated operating policy.

VIII. LIMITATIONS

The study has several principal limitations. First, the classification evaluation contains only 16 comparable physical cases, so the reported metrics and feature effects are statistically unstable. Second, labels are supplied operational judgments rather than independent geometrical ground truth, and no multi-annotator agreement data are available. Third, all evaluated profiles were collected in a laboratory and do not validate operation in a tunnel or mine. Fourth, the supplied dataset contains angular received-power sweeps rather than communication-level outcome records; therefore BER, SINR, throughput, packet-delivery performance, and closed-loop beam-switching success are not evaluated. Fifth, the auxiliary analyzer measurements use a single receiver and are not interchangeable with the three-receiver classification cases. Finally, model scores and downstream controller thresholds are not calibrated against application-level error costs.

Future work should therefore use a larger and consistently collected multi-receiver dataset, independent or blinded labelling, representative underground measurements, communication-level ground truth, and a separate final test environment before operational controller thresholds are calibrated.

IX. CONCLUSION

This paper presented a laboratory feasibility study of a reliability-assessment layer for a centrally coordinated SDR beamforming system. The contribution is the architecture linking repeated multi-receiver sweeps, interpretable feature extraction, classification, and a possible downstream controller interface. The evaluation uses only comparable physical three-receiver cases, no synthetic augmentation, and session-level separation between training and testing.

On 16 cases, Random Forest achieved 68.8% accuracy, 65.9% balanced accuracy, and a macro F1-score of 65.4%, detecting eight of nine ambiguous cases but only three of seven reliable cases. Additional analysis shows wide statistical uncertainty, sensitivity to Random Forest randomness, correlated descriptors, and no statistically established advantage over the strongest single-feature rule on the available sample. The results therefore support only a preliminary profile-screening feasibility claim, not validated underground or communication-link performance. A substantially larger and independently evaluated measurement campaign is required before operational deployment claims can be made.

REFERENCES

- [1] Z. Sun and I. F. Akyildiz, "Channel modeling and analysis for wireless networks in underground mines and road tunnels," *IEEE Transactions on Communications*, vol. 58, no. 6, pp. 1758–1768, 2010.
- [2] A. E. Forooshani, S. Bashir, D. G. Michelson, and S. Noghanian, "A survey of wireless communications and propagation modeling in underground mines," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1524–1545, 2013.
- [3] C. Zhou, T. Plass, R. Jacksha, and J. A. Waynert, "RF propagation in mines and tunnels: Extensive measurements for vertically, horizontally, and cross-polarized signals in mines and tunnels," *IEEE Antennas and Propagation Magazine*, vol. 57, no. 4, pp. 88–102, 2015.
- [4] J. Zhu, Y. Dong, J. Zhang, F. Guo, Q. Lu, B. Lv, and J. Wu, "Review on tunnel communication technology," *Sustainability*, vol. 14, no. 18, Art. no. 11451, 2022.
- [5] J.-M. Molina-Garcia-Pardo, M. Lienard, and P. Degauque, "Propagation in tunnels: Experimental investigations and channel modeling in a wide frequency band for MIMO applications," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Art. no. 560571, 2009.
- [6] J.-M. Molina-Garcia-Pardo, M. Lienard, P. Degauque, C. Garcia-Pardo, and L. Juan-Llaser, "MIMO channel capacity with polarization diversity in arched tunnels," *IEEE Antennas and Wireless Propagation Letters*, vol. 8, pp. 1186–1189, 2009.
- [7] A. Abdykadyrov *et al.*, "Investigation of underground communication quality using distributed antenna systems considering radio-frequency signal propagation characteristics in Almaty metro tunnels," *Network*, vol. 6, no. 1, p. 15, 2026.
- [8] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," *IEEE Access*, vol. 6, pp. 37328–37348, 2018.
- [9] O. F. Tuna and F. E. Kadan, "Security of AI-driven beam selection for distributed MIMO in an adversarial setting," *IEEE Access*, vol. 12, pp. 42028–42041, 2024.

- [10] J. Zhang, S. Xu, C. Li, Y. Huang, and L. Yang, "Efficient beam selection for ISAC in cell-free massive MIMO via digital twin-assisted deep reinforcement learning," *IEEE Transactions on Wireless Communications*, vol. 25, pp. 9875–9890, 2026.
- [11] A. Klautau, P. Batista, N. Gonzalez-Prelcic, Y. Wang, and R. W. Heath, "5G MIMO data for machine learning: Application to beam-selection using deep learning," in *Proc. Information Theory and Applications Workshop*, 2018, pp. 1–9.
- [12] R. Boutaba *et al.*, "A comprehensive survey on machine learning for networking: Evolution, applications and research opportunities," *Journal of Internet Services and Applications*, vol. 9, no. 1, pp. 1–99, 2018.
- [13] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.