Haneda, Katsuyuki; Desai, Rachana

Predicting the Line-of-Sight Existence Using Radio Channel Properties and Relative Antenna Locations

Published in:
17th European Conference on Antennas and Propagation, EuCAP 2023

DOI:
10.23919/EuCAP57121.2023.10133661

Published: 01/01/2023

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Predicting the Line-of-Sight Existence Using Radio Channel Properties and Relative Antenna Locations

Katsuyuki Haneda[1] and Rachana Desai[2]
[1] Aalto University School of Electrical Engineering, Espoo, Finland E-mail: katsuyuki.haneda@aalto.fi
[2] Rapidminer, Dortmund, Germany

***Abstract*—Existence of the line-of-sight (LoS) in a radio link is an important feature determining radio link and localization performance. This study explores relationship between multi-dimensional nature of radio channel properties such as delay and angular spreads and link pathloss with the LoS existence along with the relative antenna geometry of the two communicating devices in an attempt to predict the LoS existence. A non-linear classification method of the decision tree is trained by extensive wideband double-directional channel sounding data in a microcellular environment. It was found that the BS-MS link distance and link pathloss were most influential features in predicting the LOS existence. While the use of less influential features of delay and angular spreads led to 75 % accuracy due to not being able to find the exact pattern across different MS routes. We thereby demonstrate the usefulness of incorporating geometrical knowledge of the link to predict the LoS state, despite not using a map of the cellular site, in addition to using channel parameters.**

I. Introduction

Radio channel modeling is an essential technical element in developing modern wireless communications. Complexity of channel modeling increases as the cellular generation evolves because knowledge of new dimensions of radio channels is required for improved link performance. Time/Doppler, delay/frequency, angle/space and the wave polarizations are most notable dimensions to consider in channel models [1]. Characteristics of channels in these dimensions are possibly related to service, environmental and geometrical conditions of the radio link such as cell types, link distances and base station (BS) antenna heights. Possible simplification of channel modeling is therefore relevant by finding relations and dependencies between different dimensions of radio channel properties and with environmental and geometrical conditions of links. It is known that a unique relation exists between for example the Doppler frequency, a velocity of mobile station (MS) and the angle-of-arrival power profile of a downlink channel, while existence of such relation between *condensed channel parameters* such as the pathloss, delay and angular spreads and that with the environmental and geometrical conditions of links would lead to too complex mathematical formulas. This paper therefore resorts to a machine learning (ML) algorithm to analyze statistical relation between condensed parameters of multi-dimensional mobile radio channels and geometrical link condition. As an example, we study microcellular street canyon environment and relate the *existence of line-of-sight (LoS)* to other geometrical link conditions and condensed parameters using the decision tree [2]–[5]. Figure 1 shows the parameters of multi-dimensional channel model and geometrical conditions we aim at relating: pathloss, delay spread, angular spread at the BS and MS and finally three-dimensional BS-MS link distance.

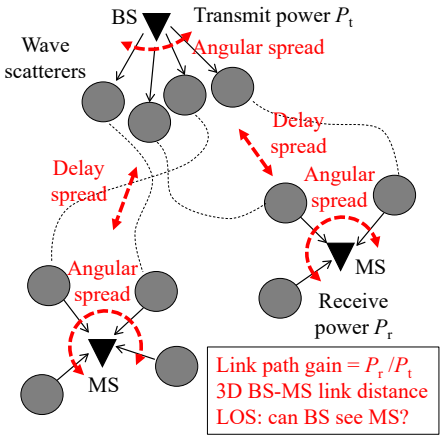

Fig. 1. Condensed parameters, highlighted in red color, of cellular downlink wave propagation channels that we concern in this paper.

The use of ML algorithms to predict channel parameters has been a popular topic for research, e.g., [6], Section II-A as exhaustive summary and hence references therein. Most closely-related work to this paper is [7] where the channel properties from vehicular scenarios are related to each other through various ML algorithms. However, the studies were mostly based on channel sounding campaigns with the use of a handful of the mentioned condensed parameters and/or link geometries. Sensitivity of the ML algorithms on inherent patterns of multi-dimensional data is therefore not fully explored. In contrast, this paper studies data-driven ML modeling of channels based on *34000 fully polarimetric double-directional wideband outdoor microcellular channel sounding* at 5.3 GHz, obtained by former colleagues of the authors of the present paper in the same research group. This is the largest dataset of measured mobile channels that have been used for training an ML algorithm, to the best of the authors' knowledge. We analyze statistical inter-dependence between LoS existence and mentioned multi-dimensional channel and geometrical link features using an ML algorithm, decision tree. It will be shown that the BS-MS link distance and link pathloss are the most influential features in predicting the LOS existence

in radio links. When predicting the LoS existence using the condensed parameters without the link distance and pathloss, the decision tree was unable to learn the exact pattern of datasets coming from different measurements of radio links, leading to limited prediction accuracy.

The rest of the paper is organized as follows: Section II denotes an analytical relationship between channel model parameters of interests that we would like to analyze the inter-(in)dependence. Section III describes a decision tree and its application to LoS state prediction based on multipath channel parameters. Section IV introduces our extensive wideband double-directional channel sounding in an urban microcellular site 5.3 GHz. Section V explains the LoS state prediction model established from the extensive measurement data, and their possible applicability to different urban microcellular sites. Finally, Section VI summarizes our work.

II. Rigorous Relations

This section introduces the rigorous relation between the parameters in our study, i.e., 1) a link pathloss on a decibel scale $PL_{\rm dB}$, 2) azimuth and polar angular spreads S_ϕ and S_θ, 3) delay spread S_τ, 4) relative coordinates of the MS in reference to BS $\boldsymbol{p}_{\rm MS} = [x\ y\ z]$ and 5) LoS existence in the channel. A channel impulse response h of a link l is defined by [1]

$$h_l(\boldsymbol{p}_{\rm MS}) = \sum_{n=0}^{N} \boldsymbol{g}_{\rm MS}^{\rm H}(\boldsymbol{\Omega}_n^{\rm MS})\boldsymbol{\alpha}_n\boldsymbol{g}_{\rm BS}(\boldsymbol{\Omega}_n^{\rm BS}) \times \delta(\boldsymbol{\Omega}^{\rm BS} - \boldsymbol{\Omega}_n^{\rm BS})\delta(\boldsymbol{\Omega}^{\rm MS} - \boldsymbol{\Omega}_n^{\rm MS})\delta(\tau - \tau_n) \quad (1)$$

for a static MS and hence the Doppler frequency is zero. In (1), $0 \le n \le N$ denotes indices of the number of plane wave propagation paths, where $n = 0$ is assigned to a LoS path while $1 \le n \le N$ are allocated for any other propagation paths undergoing reflection, diffraction, transmission and scattering. Furthermore, $\delta(\cdot)$ is a dirac delta function, $\boldsymbol{\Omega} = [\phi\ \theta]$ is a vector of azimuth and polar angles of a path observed at link ends, τ is a propagation delay time of the path from the BS to MS antennas; $\boldsymbol{g} \in \mathbb{C}^2$ is a row vector of complex polarimetric antenna transfer function at link ends; $\boldsymbol{\alpha} \in \mathbb{C}^{2\times 2}$ is a complex polarimetric path amplitude matrix, and finally $.^{\rm H}$ denotes the Hermitian transpose. The formula (1) represents an LoS channel when $\boldsymbol{\alpha}_0 \neq \boldsymbol{0}$ while it is an NLoS channel if $\boldsymbol{\alpha}_0 = \boldsymbol{0}$. The following defines the five parameters of our main interests in multi-dimensional channel models that can be determined using symbols introduced in (1).

- Omni-directional *pathloss* of a link on decibel-scale $PL_{\rm dB}$ is given by

$$PL_{\rm dB} = -10\log_{10}\sum_{n=0}^{N}|\boldsymbol{\alpha}_n|^2. \quad (2)$$

- *Azimuth spread* S_ϕ of a link is given by [1]

$$S_\phi = \sqrt{\frac{\sum_{n=0}^{N}(e^{{\rm j}\phi_n} - \mu_\phi)^2|\boldsymbol{\alpha}_n|^2}{\sum_{n=0}^{N}|\boldsymbol{\alpha}_n|^2}}, \quad (3)$$

where

$$\mu_\phi = \sqrt{\frac{\sum_{n=0}^{N}e^{{\rm j}\phi}|\boldsymbol{\alpha}_n|^2}{\sum_{n=0}^{N}|\boldsymbol{\alpha}_n|^2}}, \quad (4)$$

is the mean azimuth angle. The *polar angular spread* S_θ can also be derived in the same manner by replacing ϕ by θ.

- *Delay spread* S_τ of a link is given by [1]

$$S_\tau = \sqrt{\frac{\sum_{n=0}^{N}\tau_n^2|\boldsymbol{\alpha}_n|^2}{\sum_{n=0}^{N}|\boldsymbol{\alpha}_n|^2} - T_{\rm m}}, \quad (5)$$

where μ_θ is the mean delay defined as

$$T_{\rm m} = \frac{\sum_{n=0}^{N}\tau_l|\boldsymbol{\alpha}_n|^2}{\sum_{n=0}^{N}|\boldsymbol{\alpha}_n|^2}. \quad (6)$$

Other than the pathloss, the above parameters are referred to as *condensed parameters*. They are essential inputs to multi-dimensional channel models such as ITU-R IMT-2020 [8]. It is clear that rigorous relationships between pathloss and condensed parameters do not reveal unique mathematical relationship and hence inter-(in)dependence between them. A data-driven approach to analyze the relationships between them would therefore be the more feasible analysis method.

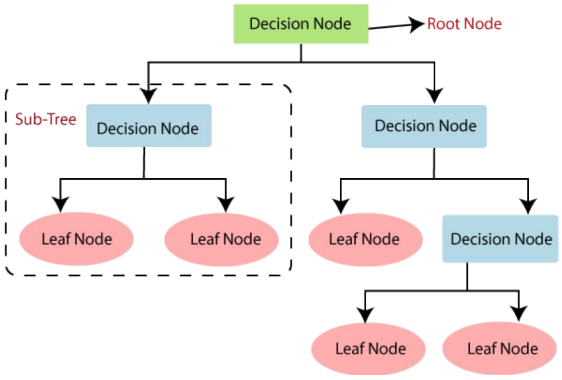

Fig. 2. Decision Tree Layout [9].

III. Decision Tree

This section introduces the principle of decision tree and the reasoning for using it for the LoS state identification. Classification And Regression Tree (CART) is a collective name for two separate machine learning approaches; numerical target labels such as condensed parameters in this study are represented by a regression tree, whereas categorical or binary target labels such as the LoS existence for each BS-MS link is represented by a classification tree. A decision tree is a flowchart-like structure consisting of the following parts.

- *Root Node*: The top-level node symbolizes the overall goal or major decision to make.
- *Decision Node*: A node below the root node representing a test for classification and regression based on an *attribute* of the target label.

- *Branches*: The options, regression and classification outcomes, or courses of action that are available in relation to the test in the decision node are represented by the branches that grow from the nodes. They are usually denoted by an arrow line and frequently include associated costs as well as the chance of occurrence.
- *Leaf Node*: Each action's alternative outcomes are represented by the leaf nodes or terminal nodes, which are attached at the ends of the branches. Leaf nodes are divided into two types: decision nodes, which signal the need for further decision, and leaf nodes, which represent a chance occurrence or undetermined consequence. The leaf nodes indicate different *classes*.

Each decision node indicates a splitting rule for a value/category of the target label based on its attribute. For example, a radio link may be classified into LoS or NLoS condition at an decision node having a splitting rule related to the delay spread of the link. The link state is the target label, while its state of LoS or NLoS is is an binary attribute of the target label in this case. The rule separates the target value/state into different classes for classification, and for regression, each resulting class has a reduced error after split by the rule. New decision nodes are created until user-set figure-of-merits are fulfilled for regression or classification based on the test data for training. The figure-of-merits to define the optimum regression or classification can be of different types such as gain ratio, information gain, accuracy, Gini index and least square errors. A numerical value estimate of the target label in a class is obtained by averaging the classified values in the corresponding leaf node, while a forecast of the target label is given as one of the classes and corresponding numerical value estimate. An attribute of a class is determined based on the majority of examples that reached this leaf during training. The most widely utilized method for predicting and classifying future events is decision trees [9]–[11].Building and trimming are the two primary processes in developing such trees. During the first phase, the training data set is recursively partitioned until the majority of the entries in each partition have the same value. After that, the second phase removes some branches that contain erroneous classification or those with the largest estimated error rate in the regression to keep the tree compact. Decision trees are frequently criticized for failing to capture complicated and non-linear relationships between target labels. Despite this, research reveals that decision trees have a high level of accuracy given the training data requirements [9]–[11].

Because decision trees are non-linear, they are more flexible to explore, prepare, and predict multiple alternative consequences for decisions to make. Furthermore, by displaying a simplified view of cause-and-effect links, decision trees efficiently express complicated processes. Finally, they allow thorough validation of intuition about relationship between a target label and attributes before detailed analysis of the same through, e.g., physical and analytical models.

IV. Wideband MIMO Channel Sounding

As an alternative to the mathematical model based approach to analyze inter-dependence of pathloss, LoS existence and condensed parameters, a data based modeling approach of the same is tested. This section introduces the data and channel sounding campaign from which the data originate. Since the data have never been presented in papers, we describe them here briefly for complete understanding of the present work.

TABLE I
Geometrical settings of the MIMO channel sounding at 5.3 GHz in downtown Helsinki

Route	MS travel length [m]	# of BS-MS links L	LoS/ NLoS	BS height z_{BS} [m]
s03	42.3	1766	NLoS	10
s04	61.6	4712	Both	10
s05	28.3	3285	NLoS	10
s06	89.7	5997	both	10
s07	41.1	3129	both	10
s08	65.1	4574	LoS	10
s09	46.5	4263	LoS	10
s10	0	1031	NLoS	16-40
s11	70.8	4508	NLoS	40
s13	0	1181	NLoS	16-40

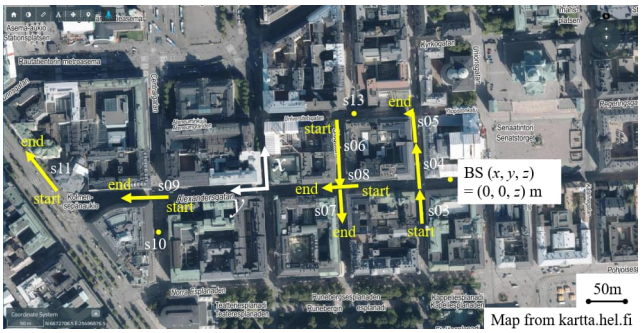

Fig. 3. Dynamic MIMO channel sounding at 5.3 GHz in a downtown Helsinki. A location of the BS along with MS routes s03–s11 are overlaid.

A large campaign of dynamic multiple-input multiple-output (MIMO) channel sounding was performed in downtown Helsinki at 5.3 GHz by former colleagues of one of the authors in the same research group using the channel sounder introduced in [12]. The BS is fixed at one of streets with elevated height using a crane. Several routes of a mobile station are illustrated in Fig. 3. The routes have varying length, the number of MIMO channel transfer functions L and the BS heights as summarized in Table I. Existence of the optical LoS is also indicated in the same table. Routes s10 and s13 were performed while the MS antenna is static. Instead, the BS antenna height changes between 16 and 40 m continuously. For route s13, about 60 links at the lowest BS antenna height were outage where no meaningful signal was observed during channel sounding. Routes s04, s06 and s07 include both LOS and NLOS conditions depending on the location of the MS.

The BS was equipped with a planar array of 32 dual-polarized patch antennas, while the mobile station was with a spherical antenna array consisting of 32-element dual-polarized patch antennas. The mobile was on a trolley at the lowest height of 1.6 m above the ground to measure MIMO radio channel responses on 10 MS routes stretching

over 445 m. Since the channel response was acquired every quarter- wavelength, we have obtained a total of about 34,000 MIMO channel responses. From the measured MIMO channel responses, a set of parameters was estimated for a propagation path, i.e., 1) angle-of-departure (azimuth/polar) at the BS $\mathbf{\Omega}^{\mathrm{BS}}$, 2) angle-of-arrival (azimuth/polar) at the MS $\mathbf{\Omega}^{\mathrm{MS}}$, 3) normalized propagation delays to the shortest delayed path in the link $\hat{\tau}_{ln} = \tau_{ln} - \min_l(\tau_{ln})$ and 4) polarimetric complex amplitude $\boldsymbol{\alpha}$, using a high-resolution algorithm [12]. The high-resolution algorithm compensates for the radiation characteristics of the arrays at link ends using the knowledge of complex multi-polarized antenna patterns. Condensed parameters defined in Section II were derived for each link using the set of propagation path parameters.

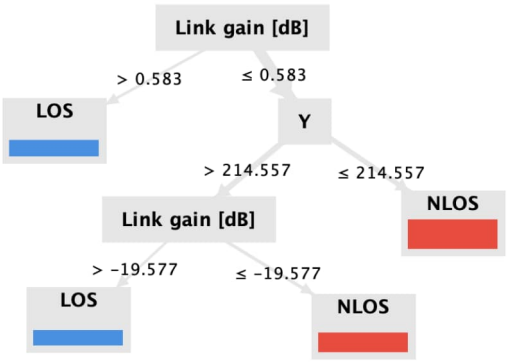

Fig. 4. The decision tree defined by influential features, i.e., pathloss [dB] and BS-MS link distance [m], to predict LoS existence of MIMO microcellular mobile links.

TABLE II
CONFUSION MATRIX FOR PERFORMANCE EVALUATION

Prediction Actual	**LoS**	**NLoS**	**Precision**
LoS	3072	55	90.2%
NLoS	5673	14076	70.9%
Recall	34.7%	99.6%	

V. APPLICATION OF DECISION TREE TO LOS STATE PREDICTION

A. Setup

The condensed parameters of MIMO links derived from the extensive channel sounding were split into training and testing subsets. Feature selection and parameter tuning for the training set was performed first. Among the CART, the classification tree is used in this study to predict existence of LoS of a link, given its condensed parameters. In general, ML algorithms help extract significant features from the data for insightful predictions. The decision tree was validated using test dataset to empower prediction strengths of the LoS/NLoS identification without knowing the specificity of data.

B. Results and Discussions

Figure 4 illustrates the decision tree with most influential features, i.e., the link pathloss and BS-MS link distance, to predict the LoS existence in cellular mobile links with 100 % accuracy. When correlation is analyzed between the LOS existence and other target labels, the link gain has a higher score of 0.6 than the others, which have a normal range from 0.0 to 0.5.

If the influential features are different for multiple MS routes, the decision tree will have to be retrained for a new route. Furthermore, it is possible that the attribute distribution is not identical for different MS routes, implying that using multiple datasets from different MS routes may not yield improved predictions of the target label. Since every training of the decision tree based on datasets from different MS routes may not lead to the exact pattern prediction of the target label, some degrees of inaccuracies are expected. For example, Figure 5 illustrates a decision tree without the link pathloss or BS-MS link distance as target labels. Each decision node has two branches depending on the attribute of delay and angular spreads. The leaf node is existence of the LoS or NLoS. Table II depicts a confusion matrix that compares classification outcomes with actual observations. There are four potential outcomes here:

- *True Positive (TP)* means that the decision tree predicts that the link has a LoS, which is actually true; 3072 links fell into this case.
- *False Positive (FP)* is LoS prediction but actual link condition is NLoS; 55 links were classified to this case.
- *False Negative (FN)* refers to NLoS prediction while actually the link was LoS; 5673 cases observed.
- *True Negative (TN)* cases are correct prediction of the NLoS condition; 14076 links fell into this case.

In our balanced dataset with the similar number of LoS and NLoS link states in the measurement, the accuracy of 75 % was achieved with the decision tree. The prediction performance shows some rooms for improvement using optimizing parameters to extract and learn more patterns for the model. The current studies were based on datasets of condensed parameters from extensive MIMO channel sounding with 2 LoS, 5 NLoS and 3 mixed LoS/NLoS MS routes. However, the number of datasets were still limited to extract meaningful patterns using the decision tree. More BSs and cellular sites, leading to extended MS measurement routes, may help find more meaningful patterns of the condensed parameters. The inter-relationships between condensed parameters are as such complex as elaborated in Section II, making it hard to find commonalities or dissimilarities in the present data sets from extensive MIMO channel sounding. Relating the condensed parameters with the geometry-related target labels like the user's location, heights, position and density, among others may lead to finding more meaningful patterns.

VI. CONCLUDING REMARKS

This paper presented the use of the decision tree to find inherent patterns of various condensed parameters of radio channels and geometrical radio link setup to relate them with the LoS existence of the same link. Fully data-driven modeling approach is studied because of complicated and non-unique

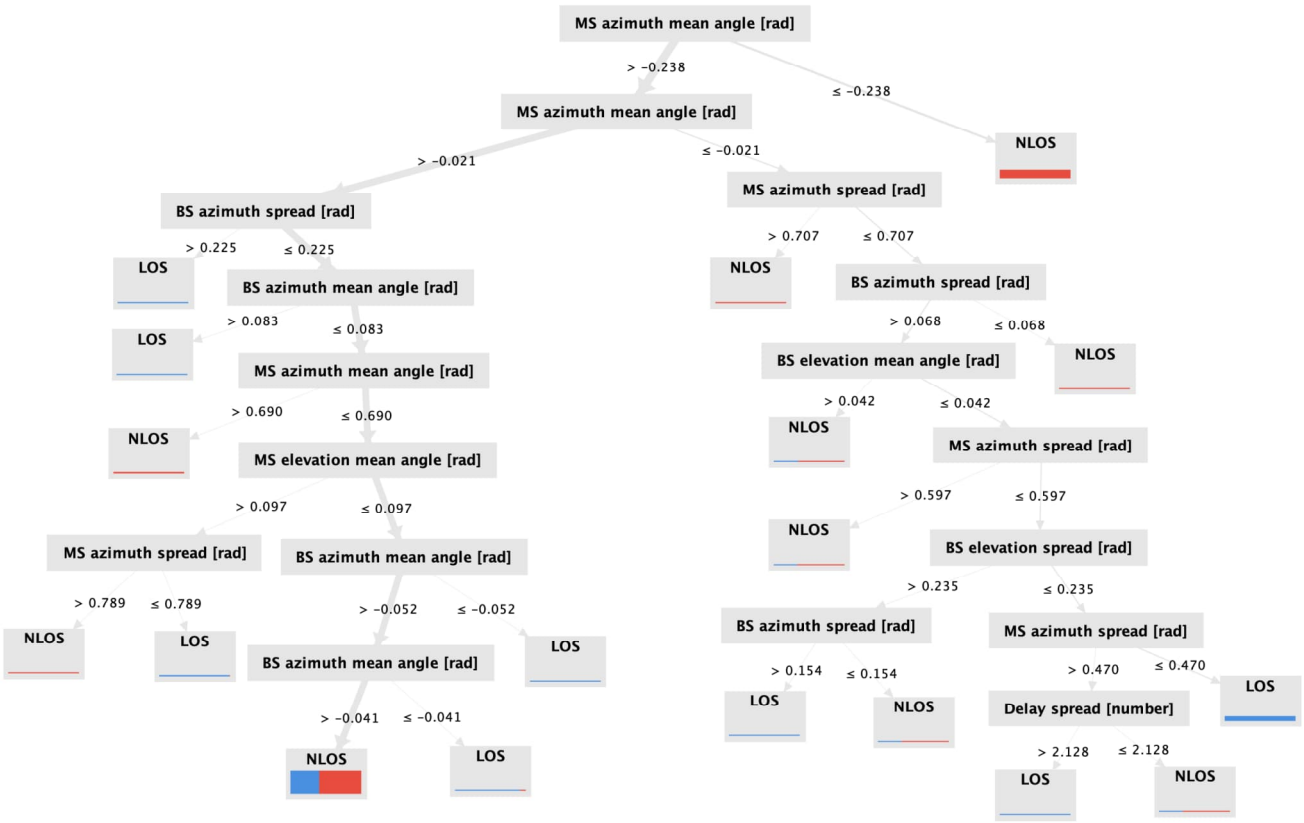

Fig. 5. The decision tree explained by all the condensed channel parameters but the pathloss and BS-MS link distance. The prediction accuracy of the LoS existence in MIMO microcellular mobile links was 75 % by this tree. In the figure, the delay spread is defined by a number of taps, where each tap corresponds to a delay width of 16.6 ns equivalent to 5 m in a path length.

mathematical relationship between the mentioned parameters and setups. The radio channel and geometrical link data were obtained from extensive wideband polarimetric double-directional channel sounding in a city center of Helsinki at 5.3 GHz, featuring about 34000 links. It was found that the BS-MS link distance and link pathloss were most influential features in predicting the LOS existence. While the use of less influential features of condensed parameters led to 75 % accuracy in predicting the target label due to not being able to find the exact pattern across different MS routes. The use of geometric link setups, with only relative positions of the BS and MS, in addition to condensed channel parameters seems promising way forward when predicting one/some of them.

ACKNOWLEDGEMENT

This work has been in part funded by the European Commission through the H2020 project ARIADNE (Grant Agreement no. 871464) and by the Academy of Finland – NSF joint call pilot "Artificial intelligence and wireless communication technologies", decision #345178.

REFERENCES

[1] A. F. Molisch, *Wireless Communications*. John Wiley and Sons, Ltd., 2011.

[2] A. Myles, R. Feudale, Y. Liu, N. Woody, and S. Brown, "An introduction to decision tree modeling," *J. Chemometrics*, vol. 18, pp. 275–285, 2004.

[3] J. Fürnkranz, *Decision Tree*. In: Sammut, C., Webb, G.I. (eds.) Encyclopedia of Machine Learning, Springer, Boston, MA, 2011.

[4] Y.-Y. Song and Y. Lu, "Decision tree methods: applications for classification and prediction," *Shanghai archives of psychiatry*, vol. 27, no. 2, pp. 130–135, 2015.

[5] J. Grus, *Data Science from Scratch*, 2nd ed. O'Reilly Media, Inc., Apr. 2019.

[6] C. Huang, R. He, B. Ai, A. F. Molisch, B. K. Lau, K. Haneda, B. Liu, C.-X. Wang, M. Yang, C. Oestges, and Z. Zhong, "Artificial intelligence enabled radio propagation for communications – Part I: Channel characterization and antenna-channel optimization," *IEEE Trans. Ant. Prop.*, vol. 70, no. 6, pp. 3939–3954, 2022.

[7] C. Huang, A. F. Molisch, R. He, R. Wang, P. Tang, B. Ai, and Z. Zhong, "Machine learning-enabled LOS/NLOS identification for MIMO systems in dynamic environments," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 3643–3657, 2020.

[8] Report ITU-R M.2412-0, "Guidelines for evaluation of radio interface technologies for IMT-2020," Oct. 2017. [Online]. Available: https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-M.2412-2017-PDF-E.pdf

[9] JavaTpoint, "Decision tree classification algorithm." [Online]. Available: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm

[10] G. Seif, "A guide to decision trees for machine learning and data science," Nov. 2018. [Online]. Available: https://towardsdatascience.com/a-guide-to-decision-trees-for-machine-learning-and-data-science-fe260\\7241956

[11] V. Kanade, "What is a decision tree? Algorithms, template, examples, and best practices," May 2022. [Online]. Available: https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-decision-tree/

[12] V.-M. Kolmonen, J. Kivinen, L. Vuokko, and P. Vainikainen, "5.3-GHz MIMO radio channel sounder," *IEEE Trans. Instr. Meas.*, vol. 55, no. 4, pp. 1263–1269, 2006.