Path Loss Prediction Modelling for Next-generation Internet-of-Things Applications Using Different Boosting Machine Learning Methods

Vasileios P. Rekkas
ELEDIA@AUTH, School of Physics
Aristotle University of Thessaloniki
541 24 Thessaloniki, Greece
vrekkas@physics.auth.gr

Sotirios P. Sotiroudis
ELEDIA@AUTH, School of Physics
Aristotle University of Thessaloniki
541 24 Thessaloniki, Greece
ssoti@physics.auth.gr

Georgia Athanasiadou
Department of Informatics
and *Telecommunications*
University of Peloponnese
Tripolis, Greece
gathanas@uop.gr

Panagiotis Sarigiannidis
Department of Informatics
and Telecommunications Engineering
University of Western Macedonia
Kozani, Greece
psarigiannidis@uowm.gr

George V. Tsoulos
Department of Informatics
and *Telecommunications*
University of Peloponnese
Tripolis, Greece
gtsoulos@uop.gr

Sotirios K. Goudos
ELEDIA@AUTH, School of Physics
Aristotle University of Thessaloniki
541 24 Thessaloniki, Greece
sgoudo@physics.auth.gr

***Abstract*—Wireless channel propagation characteristics are crucial for wireless network systems. The accuracy of the path loss (PL) prediction determines the quality of the received signal and optimization of wireless communication networks.In this paper, we apply and compare various machine learning (ML) boosting methods for the prediction of path loss in cellular communications using a flying base station (FBS). We use a ray tracing technique to obtain the dataset for the training process of the models. The work at hand generates prediction models, based on five different boosting ML learners to accurately predict the path loss of an unmanned aerial vehicle (UAV). The proposed approach exhibits great accuracy and efficiency in predicting the path loss.**

***Index Terms*—Ray tracing, cellular communications, machine learning, boosting**

I. Introduction

Accurate wireless propagation models can play a crucial role in planning, designing and evaluating the performance of wireless communication networks [1]. PL can be affected by a variety of factors, including absorption of electromagnetic waves, reflection, weather conditions and refraction [2]. PL prediction needs to be accurately modeled, in order to achieve optimal frequency and interference analysis, management of dynamic coverage, reliable system design and link budget analysis and optimal network performance [2], [3]. The advent of NGIoT next-generation Internet-of-Things (NG-IoT) brings different research challenges and priorities. The identified priorities encompass several components of the IoT stack and, as a result, relate to 6G, Distributed Ledgers, Big Data, Artificial Intelligence, Cyber Security, and Cloud Computing. Especially, the emerging sixth-generation (6G) mobile networks require features like massive connectivity, increased network capacity, and extremely low-latency. UAVs represent a crucial part of the NGIoT ecosystem. Conventional empiric and deterministic methods have been utilized for path loss modeling [4]. Empirical modeling is based on frequency and attenuation data measurements and use mathematical formulation for the statistical description of the propagation factors between transmitter and receiver antennas. The study of the wave propagation and network planning of fifth generation (5G) cellular networks using low-altitude UAVs in urban environment seems to be a necessity [5]. ML algorithms can offer various services and become a vital aspect of future wireless networks. PL prediction can be modelled as a supervised regression problem. Conventional ML algorithms, can achieve great accuracy and efficiency, reduced computing time and resources and excellent applicability [6]. In contrast to the conventional ML approach, in this paper we aim to evaluate the efficiency of various boosting ML methods in urban environments for cellular communications. The rest of this paper is organized as follows. Section II provides the formulation procedure and modeling details. Section III describes the boosting ML methods we employ. In section IV simulation metrics and results are presented, while conclusions are included in Section V.

II. FORMULATION

A 3D in-house ray tracing software is used to compute the path loss in an urban cellular network in the city center of Tripolis, Greece. A FBS is placed in an outdoor urban environment, at a point A (x_{BS},y_{BS}) and flies in three different heights (2, 4, and 6m) obtaining the path loss values for different horizontal coordinates x_i,y_i for the UAV heights.

The FBS transmissions are used for the calculation of the path loss, with the transmission frequency at 2100 MHz (3G/4G network). A data-set is constructed,using the results of the software and is used to train the ML models. The four input parameters are the geometrical height h of the FBS, the horizontal coordinates (x_i,y_i) of the point of interest and the distance d of the point from the FBS, measured in meters, while the single output parameter is the path loss value at the point of interest, in dBm. The data-set is consisted of 18.898 samples and, for validation purpose, is randomly split into a set for the training procedure (80%) and a test dataset (20%). Each boosting ML prediction model uses this dataset as input feature and suitable performance metric indicators are used to validate the results.

III. MACHINE LEARNING MODELS

ML techniques have been widely investigated for ray tracing and path loss prediction problems. In this work various boosting ML methods are applied to study the performance and accuracy approximation in path loss problems. Boosting is a ensemble learning approach, that trains and combines base classifies/regressors to produce an accurate classification/regression model [7].

A. Adaptive Boosting(Adaboost)

Adaboost is a strong ensemble method that combines various weak classifiers with poor accuracy, in order to obtain a strong classifier. AdaBoost increases the weight of false classified data points, or decreases the weight of the correct-classified data points. In each iteration, Adaboost trains the data sample and ensures accurate predictions, by highlighting the false classified data and obtaining new distribution samples every time,based on the results. Adaboost can reduce the weight of the unimportant data and prioritize the key data points, on the top of the training procedure. This way, Adaboost, while being simple, can improve the accuracy and convergence speed, despite any increase in diversity of the samples [8].

B. Gradient Boosting Decision Tree (GBDT)

GBDT is an ensemble algorithm that combines Gradient Descent Method, ensemble Learning and Decision Trees. It uses weak decision trees as the base learners and uses a gradient boosting technique to sum the predictions and conclusion of the series of trees, as the final model. GBDT adjusts the errors, trains a new decision tree from the previously trained one and uses, at each iteration, the loss function to establish the direction of the gradient descent to fit the residual between actual value and the prediction of the method. GBDT is accurate, efficient and has great interpretability, becoming rather popular for different wireless network challenges [9].

C. Extreme gradient boosting (XGBoost)

XGBoost is an ensemble based on trees or linear classifiers. It combines several weak classifiers to form a model with better efficiency. XGBoost optimizes the structured loss function by using a second-order Taylor expansion to optimize the objective function and improve accuracy. XGBoost changes the weight of training samples in the training process and uses the weights of the leaf nodes and the tree depth, to reduce and adjust the complexity of the final approach. XGBoost is widely used due to its high accuracy, high speed and strong anti-noise ability [10].

D. Light Gradient Boosting Machine (LGBM)

LGBM is a gradient boosting ensemble framework that builds a strong regression tree model by combining weak tree learners. It uses a leaf-wise splitting algorithm to develop vertically and a histogram-based method to define the best split parameters and reduce standard deviations. LGBM chooses the leaf with the highest growth loss, to grow the final tree model. LGBM has increased accuracy, higher computing speed and lower system memory usage [11].

E. Categorical boosting (Catboost)

CatBoost is a high-performance novel learning algorithm based on GBDT and uses binary decision trees as weak base learners to generate an accurate model. One primary contrast among CatBoost and other boosting algorithms is that the CatBoost uses symmetric tree and a new method of calculating the values of the leaf nodes to generate the tree, which helps improve the robustness of the model.In Catboost, the set of the feature points are randomly arranged to generate various random permutations, which helps maintain the diversity of the coupled input points and prevent over fitting. CatBoost models can process categorigal variables, as well as numerical. It can address the challenge of prediction bias, thus improving accuracy, has less predicting time and is rather efficient for low latency environments [12].

IV. PERFORMANCE MEASURES AND RESULTS

A. Metrics

A quantitative examination of the ML algorithms that are used for the path loss estimation, can be made with suitable evaluation statistic metrics. These error metrics are calculated by comparing the predicted target values of the model with the actual values of the measured test data. The error measurement metrics that are used in our study, are the Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE) and the Mean Absolute Percent Error (MAPE). Their definitions are given in the following equations (1)-(3) [13]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

$$RMSE = \sqrt{(\frac{1}{n})\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{y_i - \hat{y}_i}{y_i}|x100\% \tag{3}$$

where n denotes the test set number of input records, y are the real measured data and $\hat{y}$ the predicted ones of the i-th data record.

B. Numerical Results

In order to obtain the best suitable model for this case, we have evaluated different boosting learners, namely AdaBoost, GTBD, XGB, LGBM and Catboost. For the evaluation of our models we use the Scikit-Learn open source ML library implemented in Python language , by using suitable conventional error measurement metrics as shown in Table I.

CatBoost performs better than the other algorithms in terms of path loss prediction errors, as it has outperformed all other methods in all the previously mentioned performance indicators. The MAE value is 1.76 dB, the RMSE value is 2.789 dB and the MAPE value is 1.464%. AdaBoost and LGBM methods are close to the optimal, with MAPE values of 1.499% and 1.491% respectively, which suggests them as alternative approaches to the path loss prediction problem. Even though XGB obtains the worst results, with a value of MAPE 1.758%, the percentage error is low, which indicates that XGB can efficiently predict the path loss. In general, all methods showed satisfactory accuracy in predicting the path loss values, as MAPE values were less than 2%.

TABLE I
ERROR MEASUREMENT METRICS FORPATH LOSS PREDICTION

Algorithm	MAE	MSE	RMSE	MAPE %
Adaboost	1.865	9.903	3.147	1.499
GTB	1.927	9.487	3.080	1.612
XGB	2.076	11.270	3.357	1.758
LGBM	1.80	8.238	2.870	1.491
CatBoost	1.760	7.777	2.789	1.464

Figs. 1-2 are scatter plots that present the correlation between the actual measured values (black line) and the predicted values obtained by the ML methods (coloured dots). The prediction model is more accurate, when the prediction dots are closer to the line, that represents the actual measured test values. The correlation shows a diminutive difference between real and predicted values, due to the small MAPE values for estimating the path loss values. Figs. 3-4 are histograms that present the statistical difference distribution of the true test and the model predicted values. In Figs. 1-4 the results of Catboost and LGBM are depicted, as these approaches acquired the best accuracy in predicting the path loss.

Fig. 2b is an histogram that shows the correlation between estimated and ground truth values.

Figs. 5-7 show the comparative results of all methods for path loss prediction. It is evident that Catboost outperforms the other methods, due to the use of symmetric trees and the robustness of the method, while the worst performance is measured for the XGB learner. It should be pointed that the computational power required for ML model computation is quite lower than the one required for RT computation. The proposed ML approaches have low complexity, high accuracy, robustness and computing speed.

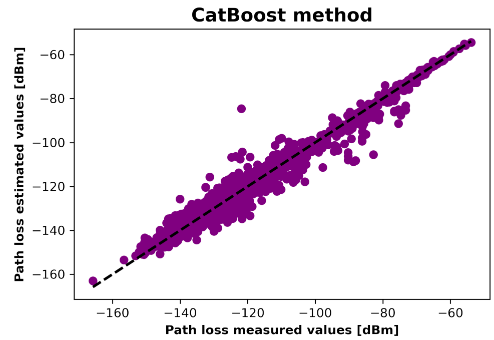

Fig. 1. Estimated versus real measurement values.

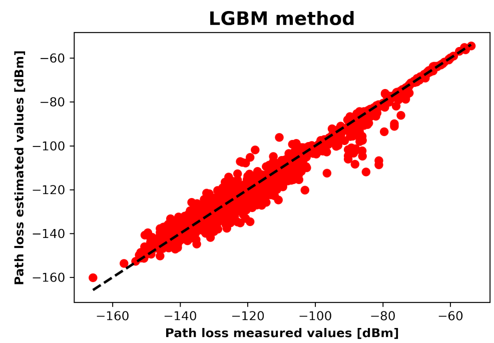

Fig. 2. Estimated versus real measurement values.

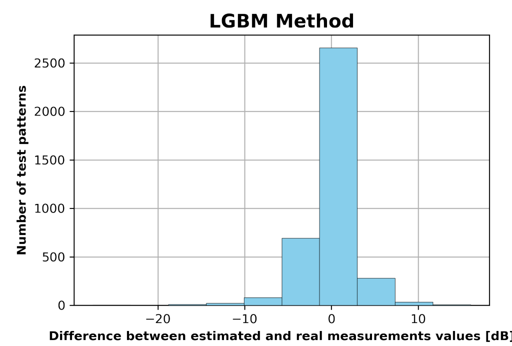

Fig. 3. Statistical distribution of the difference between estimated and real values.

V. Conclusion

In this paper, we have proposed a boosting modeling procedure for the modeling of path loss derived from FBS in an urban environment for cellular communications. We combined ray tracing data with five different boosting ML methods, namely AdaBoost, GTBD, XGB, LGBM and Catboost. A comparison regarding the efficiency in path loss prediction of the methods was conducted, with satisfactory results. Catboost outperformed the other algorithms, with all approaches scoring MAPE values below 2%, showing that boosting ML approach can offer a solution to the path loss prediction and planning. Future challenges include expanding and testing this

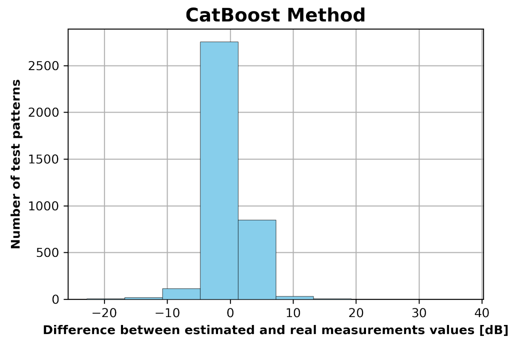

Fig. 4. Statistical distribution of the difference between estimated and real values.

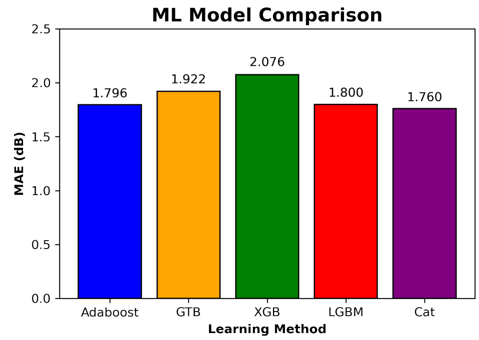

Fig. 7. MAE (dB).

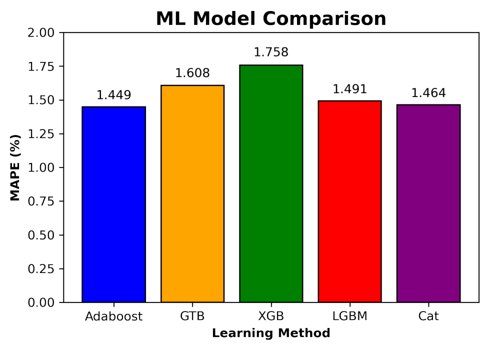

Fig. 5. MAPE (dB).

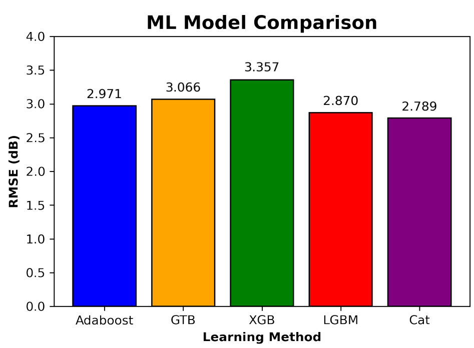

Fig. 6. RMSE (dB).

framework for 5G wireless networks, study of different Deep-Learning (DL) approaches and evalueation of the framework in a more complex environment (e.g. additional FBS heights, different point of interest, sub-urban environment etc.)

ACKNOWLEDGMENT

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 957406 (TERMINET).

REFERENCES

[1] T. T. Nguyen, R. Caromi, K. Kallas, and M. R. Souryal, "Deep learning for path loss prediction in the 3.5 GHz CBRS spectrum band," in *2022 IEEE wireless communications and networking conference (WCNC)*, 2022, pp. 1665–1670.

[2] B. Alfaresi, Z. Nawawi, and B. Y. Suprapto, "Prediction of path loss propagation in case of light rail transit (LRT) using machine learning approaches," in *2021 international conference on converging technology in electrical and information engineering (ICCTEIE)*, 2021, pp. 106–111.

[3] A. Gupta, J. Du, D. Chizhik, R. A. Valenzuela, and M. Sellathurai, "Machine learning-based urban canyon path loss prediction using 28 GHz manhattan measurements," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 6, pp. 4096–4111, 2022.

[4] Y. Sun, J. Zhang, Y. Zhang, L. Yu, Z. Yuan, G. Liu, and Q. Wang, "Environment features-based model for path loss prediction," *IEEE Wireless Communications Letters*, pp. 1–1, 2022.

[5] S. K. Goudos, G. Athanasiadou, G. V. Tsoulos, and V. Rekkas, "Modelling ray tracing propagation data using different machine learning algorithms," in *2020 14th European Conference on Antennas and Propagation (EuCAP)*, 2020, pp. 1–4.

[6] N. Moraitis, L. Tsipi, and D. Vouyioukas, "Machine learning-based methods for path loss prediction in urban environment for LTE networks," in *2020 16th international conference on wireless and mobile computing, networking and communications (WiMob)*, 2020, pp. 1–6.

[7] C. Chelmis and W. Qi, "Hierarchical MultiClass AdaBoost," in *2021 IEEE international conference on big data (big data)*, 2021, pp. 5063–5070.

[8] S. Baloch and M. S. Muhammad, "An intelligent data mining-based fault detection and classification strategy for microgrid," *IEEE access : practical innovations, open solutions*, vol. 9, pp. 22 470–22 479, 2021.

[9] H. Zhang, B. Hu, S. Xu, B. Chen, M. Li, and B. Jiang, "Feature fusion using stacked denoising auto-encoder and GBDT for wi-fi fingerprint-based indoor positioning," *IEEE access : practical innovations, open solutions*, vol. 8, pp. 114 741–114 751, 2020.

[10] Y. Feng, L. Liu, and J. Shu, "A link quality prediction method for wireless sensor networks based on XGBoost," *IEEE access : practical innovations, open solutions*, vol. 7, pp. 155 229–155 241, 2019.

[11] H. R. Seireg, Y. M. K. Omar, F. E. A. El-Samie, A. S. El-Fishawy, and A. Elmahalawy, "Ensemble machine learning techniques using computer simulation data for wild blueberry yield prediction," *IEEE access : practical innovations, open solutions*, vol. 10, pp. 64 671–64 687, 2022.

[12] M. Massaoudi, S. S. Refaat, H. Abu-Rub, I. Chihi, and F. S. Wesleti, "A hybrid bayesian ridge regression-CWT-catboost model for PV power forecasting," in *2020 IEEE kansas power and energy conference (KPEC)*, 2020, pp. 1–5.

[13] N. Ahmad, Y. Ghadi, M. Adnan, and M. Ali, "Load forecasting techniques for power system: Research challenges and survey," *IEEE Access*, vol. 10, pp. 71 054–71 090, 2022.