Machine Learning Classifier Approach with Gaussian Process, Ensemble boosted Trees, SVM, and Linear Regression for 5G Signal Coverage Mapping

Akansha Gupta*, Kamal Ghanshala, R. C. Joshi

Graphic Era Deemed to be University, Dehradun (India)

Abstract

This article offers a thorough analysis of the machine learning classifiers approaches for the collected Received Signal Strength Indicator (RSSI) samples which can be applied in predicting propagation loss, used for network planning to achieve maximum coverage. We estimated the RMSE of a machine learning classifier on multivariate RSSI data collected from the cluster of 6 Base Transceiver Stations (BTS) across a hilly terrain of Uttarakhand-India. Variable attributes comprise topology, environment, and forest canopy. Four machine learning classifiers have been investigated to identify the classifier with the least RMSE: Gaussian Process, Ensemble Boosted Tree, SVM, and Linear Regression. Gaussian Process showed the lowest RMSE, R- Squared, MSE, and MAE of 1.96, 0.98, 3.8774, and 1.3202 respectively as compared to other classifiers.

Keywords

I. Introduction

Wireless Networks have developed rapidly during the last 2-3 decades.1G arrived in the late 1980s which worked on analog signals and supported voice calls only. 2G arrived during the 1990s and was used for voice calls and data transmission having a bandwidth of 64kps.In 2000, 3G was launched with a bandwidth of 1Mbps to 2Mbps and supported not only voice calls but also video calls and conferencing.4G came into existence in 2009 with its data transmission speed of 100Mbps - 1Gbps.4G network was expanded world-wide and found its industrial applications also. Advance wireless technology is evolving very rapidly with the implementation of the 5G-NR network until 2020 [1]-[3]. Features of a High-speed next-generation 5G-NR network are high density (1 million nodes per Km2), high capacity (10Tbps per Km2), high data rates (Multi Gbps peak rates), low latency (1 ms), high reliability (1 out of 100 million packets lost), low energy (10 + years of battery life), low complexity, and high speed of 1 Gbps. With the establishment of this new technology, it may bring some challenges and difficulties like security, privacy, etc. Millimeter (MM) waves required for 5G propagation have a limitation of their effects on human cells and tissues, getting absorbed during transmission, require a small size antenna and cause unpredictable loss of signal during propagation [4]-[6].

In advanced wireless networks remarkable enhancement in information is observed [7]-[9]. Manual extraction of relevant information from an enormous amount of data is not possible, and if done, it will be prone to inevitable flaws. For capturing such big data companies no longer limit themselves to surveys and questionnaires rather big data capturing devices are deployed which include smart phone's, cameras, online browsing, etc. Machine learning seems to hold a promising solution for the analysis of big data [10]-[13]. Generally, data patterns are learned using information hidden in the big data, and then effective predictions can be proposed depending upon the final analysis. There are many Machine Learning algorithms available out of which appropriate selection of machine learning algorithms can be done using the hit and trial method [14]-[16]. We briefly describe the content in the following sections of the paper. In Section II and III a literature review and an outline of Machine learning algorithms is given respectively. In Section IV we describe measurement setup with data collection methodology. Finally, in Section V and VI, we report experimental results, discussions, conclusions, and future scope which clearly show the usefulness of the machine learning approach in predicting signal coverage.

II. Literature Review

Hajar El Hammouti et al. [17] proposed a signal mapping model based on field measured data which is applied for predicting signal coverage in an outdoor network by utilizing an S-shaped sigmoid function. Effectively modeled neural networks provide a better approximation of coverage mapping. Amir Ghasemi [18] presented a crowd-sourced analysis of the Long-Term Evaluation (LTE) network to build a wireless coverage predictive model of the radio access network (RAN). Janne Riihijarvi et al. [19] explored signal coverage mapping with machine learning algorithms using a data set derived from an extensive drive test and analyzed that Random Forest, Exponential smoothing of time series, and Gaussian Process are machine learning methods that produce better results for signal coverage. It improves the Quality of user experience and concurrently reduces

* Corresponding author.

E-mail address: akanksha3000@gmail.com

the operational cost. H. Braham et al. [20] analyzed that accurate evaluation of coverage gives better coverage optimization by utilizing the Fixed Rank Kringing (FRK) algorithm which further provides an accurate prediction of signal coverage of the locations where field measurements are not easily accessible. FRK frames a coverage map from geo located measurement by interpolating them spatially. Carlos Oroza et al. [21] estimated the performance of a machine learning based signal loss model for different terrain and vegetation environments. Four major machine learning algorithms were explored with minimum error value: K-Nearest-Neighbor, Adaboost, Random Forest, and Neural Networks. Random Forest outperforms among them with the least error. Machine learning model accomplishes a 37% reduction in average prediction error. Many researchers [22]-[23] executed exhaustive field signal measurement on Received Signal Strength Indicator (RSSI) transmission multiband channel to get the maximum signal coverage and angular power arrival. A hybrid approach adopting sub 6-G and Millimeter (mm) wave bands was found to be promising. In [24]-[25] authors proposed Backtracking Spiral Algorithm (BSA) information detection and recognition of cellular coverage using the big-data method. The distribution potential of the individual cell was diagnosed in a small granular geographical grid. The high efficiency and detection capability of the proposed algorithm validate it over other existing algorithms. Aldebaro Klautau et al. [26] represented 5G scenarios by creating channel realizations using ML applied to the PHY layer. Tadilo Endeshaw et al. [27] studied the deployment of AI by combining machine learning, NLP, and data analytics techniques for increasing the competence of wireless networks. Deussom Djomadji et al. [28] tuned the propagation models using Particle Swarm Optimization. Data is collected using the network navigation tool IX EVDO rev B. Comparison of RMSE has been done for the optimized model and the Okumura Hata model and it was concluded that an optimized model using PSO performs better than Okumura Hata model. Chao-Kai Wen et al. [29] estimated the wireless channel based on Sparse Bayesian Learning techniques.

III. Outdoor Path Loss Prediction

5G Network planning consists of outdoor propagation modeling required to predict the outdoor signal loss. Outdoor channel models consider the effect of diffraction, reflection, scattering, and refraction of EM waves, when traveling in free space between transmitter and receiver [38], [47]-[49]. Different channel models have been designed to consider the interference effect due to terrain, the height of a building, vegetation, rain, terrain, etc. Ideal path loss is predicted using the free-space path loss equation. Path loss models have been classified into canonical, empirical, deterministic, and stochastic propagation models [30]-[32].

A. Free Space Path Loss Model

It is based on the Friis transmission equation and one of its simplest kinds of first-order approximation connectivity models. EM waves travel without any interference in Line of Sight (LOS) in free space. Receive Signal Strength (RSS) decreases as the square of the distance between Tx-Rx increases with a single path after neglecting the ground effect [33]-[35].

The free space path loss model is expressed in equation (1):

$$\frac{P_r}{P_t} = \frac{G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \tag{1}$$

where,

P_t = transmitted power (*dB*)

P_r = power received by receiver (*dB*)

G_r = gain of receiving antenna

G_t = gain of transmitting antenna

λ = signal wave length (*m*)

d = distance between Tx-Rx

L = system loss factor (L=1 for FSL)

Channel loss (dB) can also be expressed using equation (2).

$$L(dB) = 32.44 + \log(d) + 20\log(f) \tag{2}$$

B. Machine Learning Algorithms

Machine learning algorithms are classified as [36], [39]-[42]:

Supervised Learning: This algorithm learns from the specimen data and related results to predict the correct result when given new data similar to specimen data.

Unsupervised Learning: This algorithm learns from the specimen data without any related results where the algorithm itself has to determine underlying data patterns and groups or cluster data based on some kind of similarity in their features.

Reinforcement Learning: This algorithm learns from specimen data that lack labels where the result or outcome is rewarded or penalized. It is like learning by trial & error.

The below described algorithms are supervised learning algorithms.

1. Linear Regression

Linear regression is the most suitable machine learning algorithm for prediction[18].In linear regression, a set of independent input parameters(x) are considered to determine the output parameter(y) and there exists an association between the input and output parameter which is expressed in the form of the linear equation as:

$$y = wx + \epsilon$$

Where ϵ = intercept on y-axis

w = slope of the line

The linear regression algorithm tries to find the best fit line by minimizing the root mean square error (RMSE) between true and predicted value.

2. Support Vector Machine (SVM)

Support Vector Machine (SVM) can perform both classification and regression of the data. The data is classified into one of the groups by finding a hyperplane that divides the input instances into two classes. The input vectors located on the hyperplane are the support vectors. In cases where the input data is not linearly separable, suitable kernel functions are applied which map data into higher dimensions where data can be easily classified.

3. Gaussian Process

Gaussian regression process is a non-parametric Bayesian approach that computes the probability distribution over permissible functions in the data [29]. The posterior probability is deduced from the prior distribution and the data. For a linear function

$$y = wx + \epsilon$$

The posterior probability is obtained using Baye's rule, equation (3):

$$p\left(\frac{w}{y}, X\right) = \frac{p\left(\frac{y}{X}, w\right) p(w)}{p\left(\frac{y}{X}\right)}$$

$$\text{Posterior Probability} = \frac{\text{likelihood x prior probability}}{\text{marginal likelihood}} \tag{3}$$

To make predictions at some random point x* one needs to calculate the predictive probability where a weighted measure of all the posterior probability distribution is evaluated using equation (4).

$$p\left(f^*/_{x^*}, y, X\right) = \int p\left(f^*/_{x^*}, w\right) p\left(w/_{y, X}\right) dw \tag{4}$$

Predictive probability is computed from posterior probability so that uncertainty measurements on the predictions can be provided by calculating their mean and variance.

4. Ensemble Learning

Sometimes a single machine learning algorithm is not able to provide the desired results, the expected result can be obtained by combining available algorithms [43]. The final result can be calculated by voting or averaging the result of individual algorithms. Major ensemble algorithms used are Bagging and Boosting.

a) Ensemble Bagging Tree with Random Forests

In Bagging firstly initial data set is utilized to reproduce a replica of the training set by using the Bootstrap sampling method. The bootstrap sampling method is used to create random samples from the initial data set where the sample size used as a training set is the same as the initial data set [44]. The random sample is generated from initial data by duplicating some sets of data multiple times and some records are not even considered once. The test data set is the initial and random sample sets that are used as the training data set. Secondly, multiple models are built from the training sets when the same algorithm is applied to them. Random Forests is a very good example to represent bagging benefits. In Random Forests, the best feature is selected for classification that converges the algorithm faster to a unique result. When the same data set is used a similar tree structure with associated prediction is obtained. Whereas the random forest after every split, bagging provides a random set of features for classification that probably result in the negligible association among classifications from sub-models.

b) Ensemble Boosting Tree with AdaBoost

AdaBoost implies Adaptive Boosting. Bagging works on 'simple voting' where each model is developed independently to provide an outcome [45]. The final result is obtained after analyzing the majority outcomes of the parallel ensemble. Boosting works on 'weighted voting' where each model provides an outcome that is based on majority selection. The final result is obtained by generating a sequential ensemble where greater weights are designated to the instances of preceding models that are misclassified. In every iteration, a model is built by rectifying the misclassification of the preceding model until no further corrections are required [46].

IV. Experimental Setup and Data Collection

This research has been carried out in Dehradun, Uttarakhand-India which lies in the Himalayan ranges. Its geographical coordinates are within latitude 78.0322° E, longitude 30.3165° N. Uttarakhand is often regarded as a terrain full of the tree canopy and suburban environment with a combination of mountain, forest, residential building, commercial complex(2 to 6 storied), and free space. Since 2018 exhaustive field measurement has been carried out at fringe areas of Uttarakhand to measure the effect of the tree, forest, mountain, snow, and buildings on propagation loss.4 clusters (6 BTS each) were identified, covering a 36 Km^2 area shown in Fig. 1. These BTS are strategically chosen to cover RSSI to the tree canopy and mountain canopy. At each test point, exhaustive measurements were carried out repeatedly to calculate changing average RSSI signal variation due to environmental conditions. Hours of driving and route tracking were carried out around identified BTS to collect RSSI samples using CATIA and TEMS navigation tools as shown in Fig. 3. When the dense forest area started, RSSI samples became inaccessible.
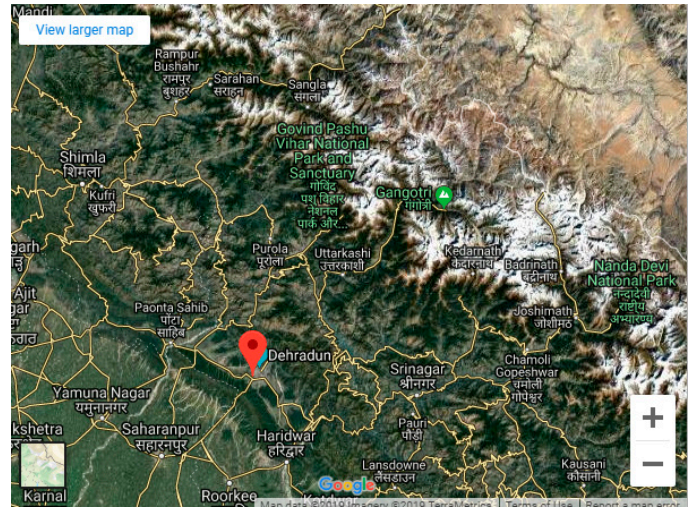

Fig. 1. The 4 clusters (6 BTS each) in the Dehradun,Uttarakhand-India.

A drive test was conducted to measure the data. As shown in Fig. 2 and Fig. 3 drive test tools consist of drive vehicle, laptop, Garmin (GPS) global position system, sockets, test cables, Sony mobile handset equipped with TEMS navigation software, MapInfo or Deskcat, and drive test route [35],[28].

Battery and charger   GPS (Ext Antenna and Data Cable)
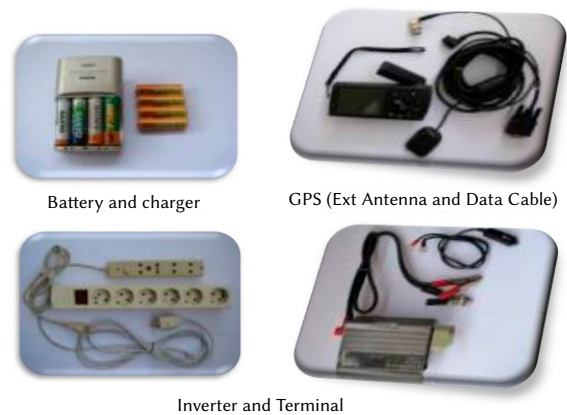
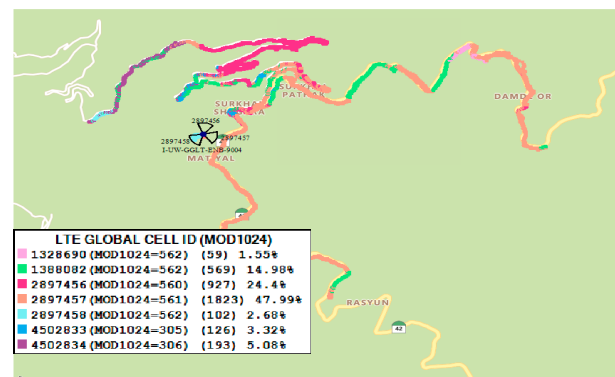Inverter and Terminal

Fig. 2. 3D Drive test tool.

Fig. 3. Drive test route. Each BTS site is identified by a 3-letter codename.

GPS was mounted on a vehicle and a Sony Ericsson mobile handset was used.RSS measurements were recorded at each test point around the selected base stations (BS) covering all roads, forests, mountains, and populated areas.

A. Dataset

42,500 RSSI samples of field measurement dataset are utilized for applying the machine learning approach on signal coverage prediction.

Fig. 4 shows the architecture of 3D channel modeling and the complete procedure of real-time data collection.
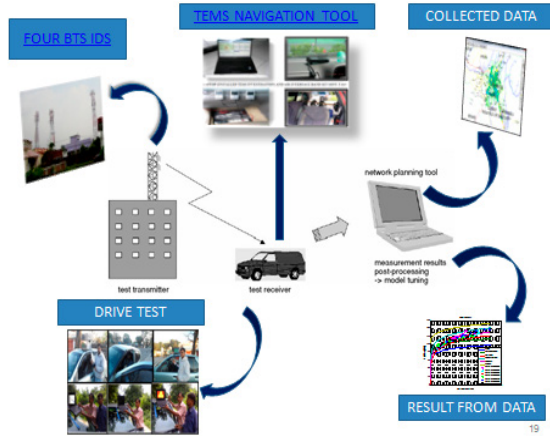

Fig. 4. Illustration of 3D Channel Modeling Architecture.

TABLE I. Features of Wireless Network

Features	Value
Coverage Objective RSRP (dBm)	-106.6
Cluster Sectors	I-UW-GGLT-ENB-9004-0 (A)
Coverage Overshooting Radius (m)	4095
Band	850,2300,1800
Antenna Longitude	10
Antenna Latitude	50
Antenna Height (m)	37
Antenna Azimuth	10
Antenna Tilt Electrical	8
Antenna Tilt Mechanical	3

Table I illustrates features of wireless network which affect network planning and required for optimum signal coverage. The received signal is a combination of signals coming from different directions due to reflection, diffraction, and scattering.

RSSI signal strength is measured 360 degrees around each BTS in 3 sectors alpha, beta, and gamma to analyze the maximum coverage of signal within a cell. The coverage threshold values of the network in 3 sectors are summarizes in Table II.

TABLE II. Coverage Objective Threshold Value in Alpha, Beta, Gamma Sectors

Coverage Objective Threshold	Sector Alpha, Beta, Gamma
Coverage Objective RSRP (dBm)	-106.6
Coverage Objective Percentile (%)	80
Coverage Overshooting Radius (m)	4095
Coverage Overshooting RSRP (dBm)	-91.6
Coverage Overshooting Percentile (%)	10
Coverage Swap Percentile (%)	50
Coverage SideLobe Percentile (%)	30
Coverage Radius Inner Percentile (%)	10

V. Experimental Results and Discussions

In this section, the performance of Machine learning classifiers is evaluated using data collected from the experimental setup. The validation scheme has been chosen before tuning to estimate the performance of the model on new data. Validation also helps to examine the predictive accuracy of the fitted models and avoids over fitting. 3 types of validation schemes were available:

a) Cross-Validation is used for small data sets and uses a full portion of the data set.

b) Hold out Validation is used for large data sets and uses some portion of the data set.

c) No Validation signifies no protection against overfitting.

5 fold cross-validation was used to divide the original data set into 5 disjoint sets as by using 5 fold cross-validation, the predictive accuracy of trained models was well estimated on the entire data set where each fold:

a) Trains a model

b) Evaluate the performance of model

c) Calculates average test error

A. Predicted Vs Response Plot

The Predicted Vs Response plot analyzed the performance of classifiers by evaluating the efficiency of the regression model by investigating the prediction for varying response values. The predicted response of models was laid against the true response. An efficient regression model had a predicted response nearly identical to the true response, therefore response values lay close to the diagonal line. The perpendicular separation between the diagonal line to each point was the deviation of the prediction for the point under consideration. An efficient classifier has minimum errors and points distributed roughly identical about the diagonal line.
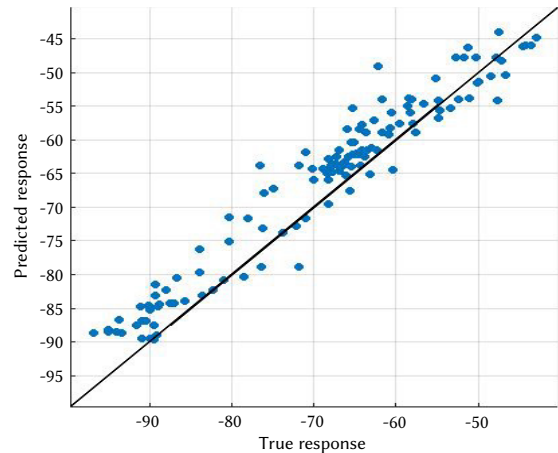

Fig. 5. Predicted Vs True response of Ensemble Boosted Tree.

Fig. 5 depicted RMSE value for Ensemble was 4.692 with R Squared value of 0.89. MSE was 22.015, MAE 3.88209, prediction speed 1800 obs/Sec, and training time of 32.441 Sec.

RMSE for Support Vector Machine was 3.9823 with R Squared value of 0.92, MSE 15.858, MAE 3.1148, prediction speed 11000 obs/Sec, and training time of 38.888Sec as shown in Fig. 6.

RMSE value for Linear Regression was 4.2946 with R Squared value of 0.91, MSE 18.444, MAE 3.5477, prediction speed 4000 obs/Sec, and training time of 25.623 Sec as shown in Fig. 7.

Fig. 8 shows RMSE value for Gaussian Process was 1.9691 with R Squared value of 0.98, MSE 3.8774, MAE 1.3202, prediction speed 8600 obs/Sec, and training time of 38.888 Sec.
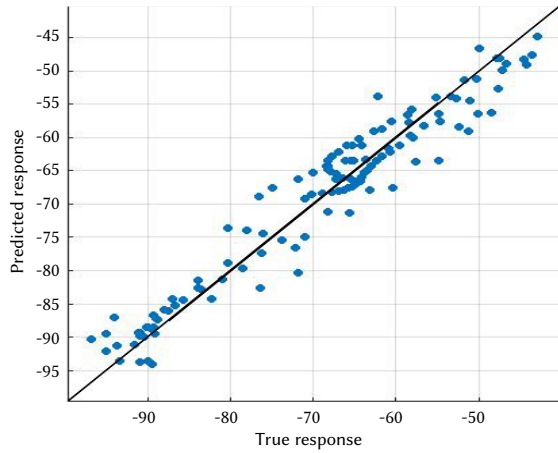
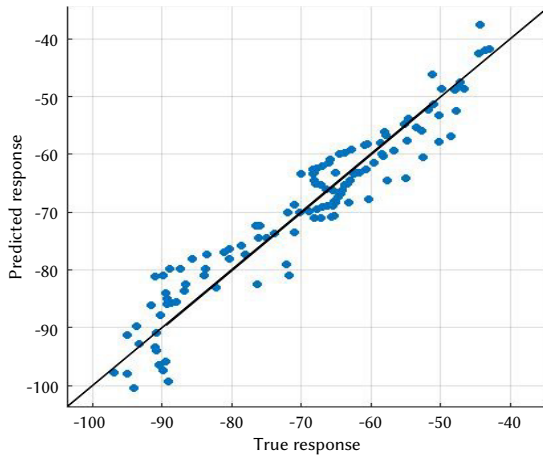Fig. 6. Predicted Vs True response of Support Vector Machine.

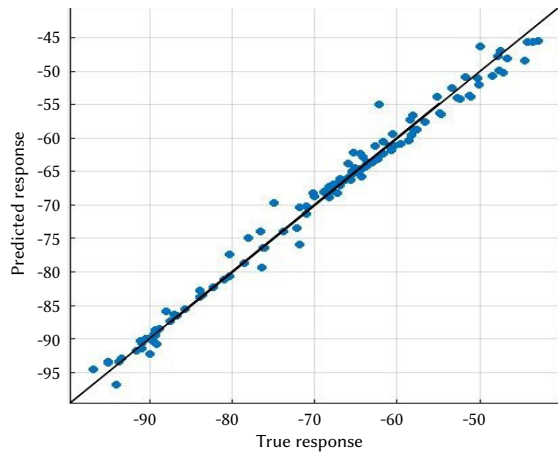Fig. 7. Predicted Vs True response of Linear Regression.

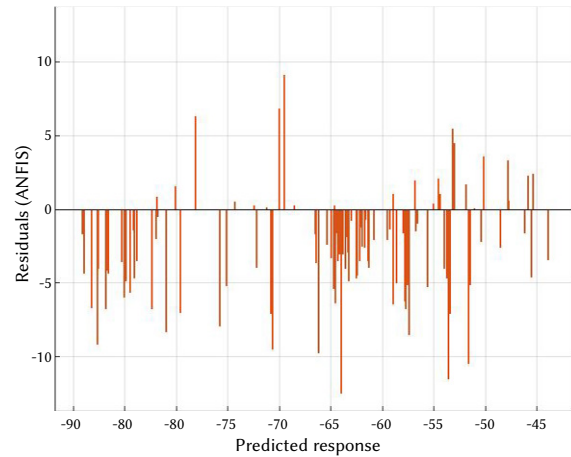Fig. 8. Predicted Vs True response of Gaussian Process.
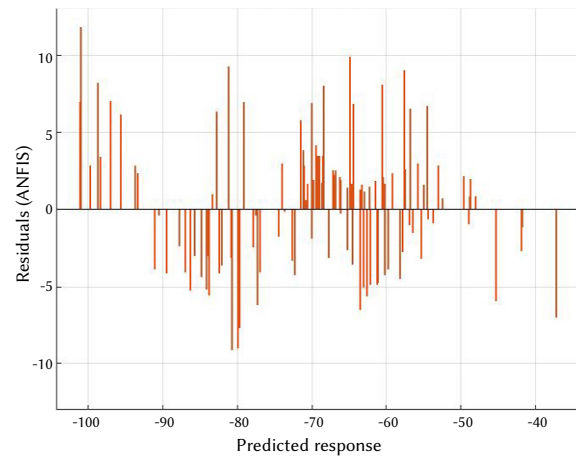

Fig. 9. Residual plot of Ensemble.

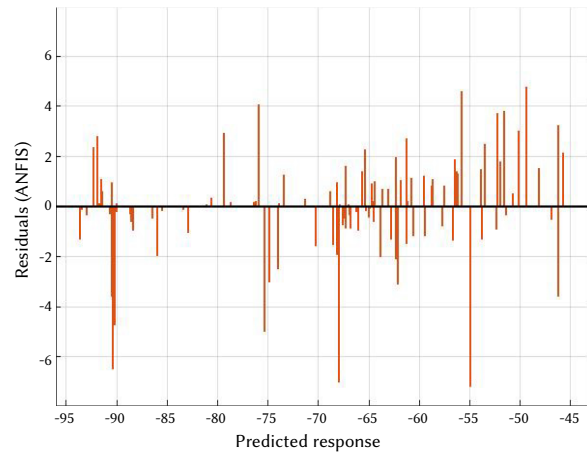Fig. 10. Residual plot of Linear Regression.

Fig. 11. Residual plot of Gaussian Process.

B. Comparison of Residual Plot Outlier

The residuals plot from Fig.9 to Fig. 12 displayed the deviation between the predicted and true responses. The predicted response variable was chosen among true response, predicted response, record number, or one of the predictors to plot on the x-axis. The efficient model had residuals distributed roughly symmetrically around 0. Fig.10 showed that the residual plot of Gaussian Process scattered roughly symmetrically around 0 and also clear patterns in the residuals are observed.
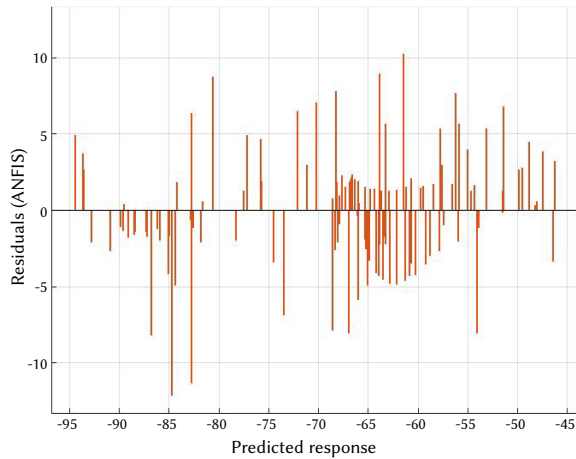
Fig. 12. Residual plot of SVM.

C. Comparison of Response Plot Outlier

A regression model result was viewed in the response plot that displayed the prediction response against the record number. Predication error was displayed as vertical lines between predicted and new responses.
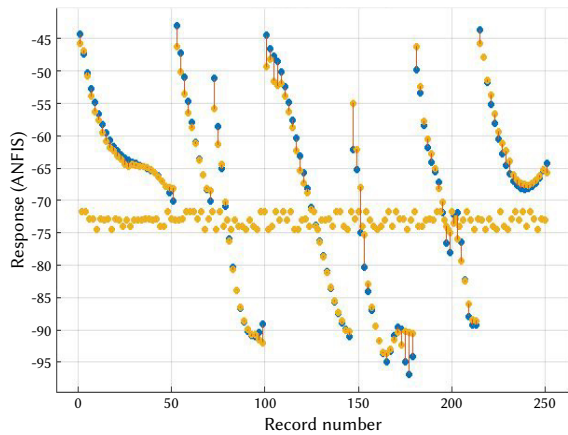

Fig. 13. Response plot of Gaussian Process.

Gaussian Process model performance was also evaluated using the residual plot after the model was trained. The difference between true and predicted response was displayed by a residual plot in Fig. 13. A variable predicted response had been plotted on the x-axis. For a good model, residuals had been scattered approximated identical around 0 and it changed considerably in size from left to right.

VI. Conclusion and Observation

The performance evaluation of machine learning algorithms on training data is tabulated in Table III.

We observed that the lowest value of RMSE was obtained from the Gaussian Process classifier, depicting the probability to correctly predict the propagation loss, while the highest value of RMSE (4.692%) was observed with the Ensemble boosted tree. However, SVM and linear regression classifiers hold intermediate values of RMSE 3.9823%, and 4.2946%, respectively. The lowest value of RMSE (1.9691%) was estimated by the Gaussian Process classifier. MSE and MAE for Gaussian Process are also a minimum of 3.8774 and 1.3202. Response plot Outlier curves for the proposed Gaussian Process classifier and other state-of-the-art algorithms are shown in Fig. 13 where residuals have been scattered approximately identical around 0. It is analyzed

that the Empirical signal coverage models which are univariate cannot predict signal coverage by using only one network parameter for coverage prediction, however machine learning-based signal coverage prediction model is multivariate and it could be designed on field RSSI measurement by considering two or more network parameters, hence predict signal coverage more accurately. Signal coverage prediction using the machine learning model requires training of best-fit machine learning classifier by hit and trial method and shortlisting machine learning classifiers with minimum RMSE error on RSSI field dataset. To validate it practically, the classifier-based signal mapping approach was applied to a real-time wireless network at the fringe area of Uttarakhand-India. However, the results of this application could encourage practitioners and researchers to validate further the practicality of the approach for similar real fringe area wireless networks.

TABLE III. Performance Evaluation of Machine Learning Algorithms on Training Data

Coverage Objective Threshold	Gaussian Process	Ensemble Boosted Trees	SVM	Linear Regression
RMSE	1.9691	4.692	3.9823	4.2946
R Squared	0.98	.89	0.92	0.91
MSE	3.8774	22.015	15.858	18.444
MAE	1.3202	3.8209	3.1148	3.5477
Prediction Speed (obs/Sec)	8600	1800	11000	4000
Training Time (Sec)	38.888	32.441	26.495	25.623

References

[1] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, V. Young, "Mobile edge computing-A key technology towards 5G," in ETSI White Paper, vol. 11, no. 11, pp. 1–16, Sept. 2015, ISBN No. 979-10-92620-08-5.

[2] K. Cichon, H. Bogucka, G. Molis, J. Adamonis, T. Krilavicius, "Learning and detection mechanisms of spectral-activity information towards energy efficient 5G communication," Baltic URSI Symposium (URSI), Poznan, Poland, May 2018,doi: 10. 23919/URSI. 2018.8406715.

[3] Y. Liu, Y. Zhang, R. Yu, S. Xie, "Integrated energy and spectrum harvesting for 5G wireless communications," in IEEE Network, vol. 29, no. 3 ,pp. 75 – 81, 2015,doi: 10. 1109/MNET. 2015.7113229.

[4] J. Kim, A. F. Molisch, "Fast millimeter-wave beam training with receive beamforming," in Journal of Communications and Networks, vol. 16, no. 5, pp. 512–522, Oct. 2014, doi: 10. 1109/JCN.2014.000090.

[5] E. Balevi, R. D. Gitlin, "Unsupervised Machine Learning in 5G Networks for Low Latency Communications," in proc. IEEE 36th international Performance computing and communication conference (IPCCC), ISSN: 2374-9628, pp. 1-2, 2017, doi: 10. 1109/PCCC.2017.8280492.

[6] T. Bogale, X. Wang, L. Le, "Machine Intelligence Techniques for Next-Generation Context-Aware Wireless Network," in ITU Journal: ICT Discoveries, vol. no. 1, pp. 1-11, Feb. 2018, doi:arXiv:1801.04223.

[7] W. Lee, "Mobile Communication Design Fundamentals", New York, John Wiely&Sons, 1993.

[8] B. Li, Z. Fei, Y. Zhang, "UAV Communications for 5G and Beyond: Recent Advances and Future Trends," in IEEE Internet of Things Journal, vol. 6, no. 2, pp. 2241 – 2263, April 2019, doi: 10. 1109/JIOT.2018.2887086.

[9] P. Antonia, A. Markos, T. Anna, S. Dimitra, "Provisioning of 5G services employing machine learning techniques," in International Conference on Optical Network Design and Modeling (ONDM) Dublin, Ireland, May 2018,doi: 10. 23919/ONDM. 2018.8396131.

[10] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. C. Chen, L. Hanzo,"Machine learning paradigms for next-generation wireless networks," IEEE Wireless Communications., vol. 24, no. 2, pp. 98–105, 2017, doi: 10. 1109/MWC.2016.1500356WC.

[11] P. V. Klaine, M. A. Imran, O. Onireti, R. D. Souza, "A survey of machine

learning techniques applied to self-organizing cellular networks," in IEEE Communications Surveys and Tutorials, vol. 19, no. 4, pp. 2392–2431, 2017, doi: 10.1109/comst.2017.2727878.

[12] E. Bjorn, "Machine learning for beam based mobility optimization in NR," Master of Science dissertation, Department of Electrical Engineering, Linkoping University, Linkoping, Sweden, 2017.

[13] D. Krajzewicz, J. Erdmann, M. Behrisch, L. Bieker, "Recent development and applications of SUMO - Simulation of Urban Mobility," in International Journal On Advances in Systems and Measurements, vol. 5, no. 3&4, pp. 128–138, Dec. 2012, doi: 10.1.1.671.2113.

[14] J. G. Carbonell, R. S. Michalski, T. M. Mitchell, "An overview of machine learning," in book Machine learning, vol. 1, pp. 3-23, Elsevier Inc. 1983, doi: https://doi.org/10.1016/bs.host.2018.07.004

[15] M. Bkassiny, Y. Li, S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," in IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1136–1159, Oct. 2013. ,doi: 10.1109/SURV.2012.100412.00017.

[16] S. Osvaldo, "A Very brief introduction to machine learning with applications to communication systems," IEEE Transactions on Cognitive Communications and Networking, vol 4, no. 4, pp. 648 - 664 ,Dec. 2018, doi: 10.1109/TCCN.2018.2881442.

[17] H. Hajar, G. Mounir, Z. Syed, "A Machine Learning Approach to Predicting Coverage in Random Wireless Networks," in IEEE Globecom Workshops, pp. 1-6, 2018, doi: 10.1109/GCWkshps43968.2018.

[18] G. Amir, "Data-driven prediction of cellular networks coverage: an interpretable machine-learning model," in 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), pp. 604-608, 2018, doi: 10.1109/GlobalSIP.2018.8646338.

[19] R. Janne,M. Petri, "Machine learning for performance prediction in mobile cellular networks," in IEEE Computational Intelligence Magazine, pp. 51-60, vol 13, Issue:1,2018,doi: 10. 1109/MCI.2017.2773824.

[20] H. Braham, S. B. Jemaa, G. Fort, E. Moulines, B. Sayrac, "Fixed rank kriging for cellular coverage analysis," IEEE Transaction Vehicular Technology, vol. 66, no. 5, pp. 4212–4222, May 2017. doi:10.1109/TVT.2016. 2599842.

[21] O. Carlos, Z. Ziran, W. Thomas, G. Steven, "A Machine-Learning based connectivity model for complex terrain large-scale low-power wireless deployments," IEEE Transactions on Cognitive Communications and Networking (Vol. 3, No. 4, Dec. 2017). doi: 10.1109/TCCN.2017.2741468.

[22] R. Nihesh, S. Renu, S. Rajesh, "Coverage estimation in outdoor heterogeneous propagation environments," in IEEE Access, vol. 8,pp. 31660 – 31673, doi: 10.1109/ACCESS.2020.2972811.

[23] Z. Zhimeng, Z. Jianyao, L. Chao, "Outdoor-to-Indoor channel measurement and coverage analysis for 5G typical spectrums," in International Journal of Antennas and Propagation, Vol. 2019, pp. 1-10, doi: 10.1155/2019/3981678.

[24] V. Kristem, S. Sangodoyin, C. U. Bas, "3D MIMO outdoor-to-indoor propagation channel measurement," IEEE Transactions on Wireless Communications, vol. 16, no. 7, pp. 4600–4613, 2017.

[25] W. Hai, X. Su, L. Ke, M. Omair, "Big data-driven cellular information detection and coverage identification" in sensor journal, vol 19, no. 4, pp. 937-942, 2019, doi: 10.3390/s19040937.

[26] K. Aldebaro, B. Pedro, G. Nuria, W. Yuyang, H. Robert, "5G MIMO data for machine learning: application to beam-selection sing deep learning," in Information Theory and Applications Workshop (ITA), San Diego, CA, USA, October 2018, doi: 10.1109/ITA.2018.8503086.

[27] B. Tadilo, B. Long, "Massive MIMO and mmWave for 5G wireless HetNet: potential benefits and challenges," IEEE Vehicular Technology Magazine, vol. 11, no. 1, pp. 64 – 75, March 2016, doi: 10.1109/MVT.2015.2496240.

[28] M. Deussom, E. Tonye, "New propagation model optimization approach based on Particles Swarm Optimization algorithm," in International Journal of Computer Applications, vol. 118, no. 10, pp. 39-47, 2015 ,doi: 10. 5120/20785-3430.

[29] W. Chao, J. Shi, W. Kai-Kit, C. Jung, T. Pangan, "Channel estimation for massive MIMO using gaussian-mixture bayesian learning," in IEEE Transaction Wireless Communication, vol. 14, no. 3, pp. 1356–68, Mar. 2015, doi: 10.1109/TWC.2014.2365813.

[30] S. Sun, "Path loss models for 5g urban micro- and macro-cellular scenarios," in 2016 IEEE VTC-Spring 2016, May 2016. [Online]. Available: http://arxiv. org/abs/1511.07311.

[31] M. Hall, "Radiowave propagation effects on next-generation fixed-services terrestrial telecommunication systems," European Cooperation in Science and Technology (COST), Tech. Rep. ICT COST Action 235, 1996.

[32] A. Gupta, S. Sharma, S. Vijay, V. Gupta, "Secure path loss prediction using fuzzy logic approach," in 2008 Fourth International Conference on Wireless Communication and Sensor Networks, Allahabad, India, Dec. 2008, doi: 10.1109/WCSN.2008.4772717.

[33] J. Azevedo, F. Santos, "An empirical propagation model for forest environments at tree trunk level," IEEE Transactions on Antennas and Propagation, vol. 59, no. 6, pp. 2357–2367, 2011.

[34] I. T. U. (ITU), "Influence of terrain irregularities and vegetation on tropospheric propagation," in CCIR XVth Plenary Assembly, vol. V: Propagation in Non-Ionised Media, no. ITUR Report 236-6, Dubrovnik, Croatia, 1986.

[35] S. Kim, B. Guarino, T. Willis, V. Erceg, S. Fortune, R. Valenzuela, L. Thomas, J. Ling, J. Moore, "Radio propagation measurements and prediction using three-dimensional ray tracing in urban environments at 908 MHz and 1. 9 GHz," in IEEE Transactions on Vehicular Technology, vol. 48, no. 3, pp. 931–946, May 1999, doi: 10.1109/25.765022.

[36] E. Alpaydm, "Introduction to Machine Learning," 3rd edition, The MIT Press, Cambridge, Massachusetts, 2014.

[37] S. Maghsudi, S. Stanczak, "Channel selection for network assisted D2D communication via no-regret bandit learning with calibrated forecasting," in IEEE Transactions on Wireless Communications, vol. 14, no. 3, pp. 1309–22, 2015, doi: 10.1109/TWC.2014.2365803.

[38] V. Gupta, S. Sharma, M. Bansal, "Fringe area path loss correction factor for wireless communication," in International Journal of Recent Trends in Engineering ,vol. 1, no. 2, pp. 30-32, May 2009, ISSN(online): 2455–1457.

[39] T. Shea, J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," in IEEE Transactions on Cognitive Communications and Networking, vol. 3, no. 4, pp. 563–575, 2017,doi: 10.1109/tccn.2017.2758370.

[40] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning," in Nature, vol. 521, pp. 436–444, 2015,doi: 10.1038/nature14539.

[41] R. Atallah, C. Assi, M. Khabbaz, "Deep reinforcement learning- based scheduling for roadside communication networks," in Proceeding of International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Paris, France, pp. 1–8, May 2017, doi: 10.23919/wiopt.2017.7959912.

[42] T. Wang, C. Wen, H. Wang, F. Gao, T. Jiang, S. Jin, "Deep learning for wireless physical layer: Opportunities and challenges," in Journal China Communications, vol. 14, no. 11, pp. 92-111, Nov. 2017, doi: 10.1109/CC.2017. 8233654.

[43] R. Schapire, "The boosting approach to machine learning: An overview," in book Nonlinear Estimation and Classification, pp. 149–171, Springer, New York, NY, 2003, doi: 10.1007/978-0-387-21579-2.

[44] M. Prasad, L. Iverson, A. Liaw, "Newer classification and regression tree techniques: bagging and random forests for ecological prediction," in Springer International journal Ecosystems, vol. 9, no. 2, pp. 181-199, 2006, doi: 10.1007/s10021-005-0054-1.

[45] T. Shea, T. Erpek, T. C. Clancy, "Deep learning based MIMO communications," CoRR, vol. abs/1707. 07980, 2017. [Online]. Available: http://arxiv. org/abs/1707.07980.

[46] W. Loh, "Classification and regression trees," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 1, no. 1, pp 14–23, 2011, doi: 10.1002/widm.8.

[47] S. Salous, J. Kim, M. Sasaki, W. Yamada, X. Raimundo, A. Cheema, "Radio propagation measurements and modeling for standardization of the site general path loss model in International Telecommunications Union(ITU) recommendations for 5G wireless networks," In Radio Science Journal, (2020). vol 55, no. 1, Jan. 2020, doi: 10.1029/2019RS006924.

[48] S. Pattanayak, "A genetically trained neural network for prediction of path loss in outdoor microcell," in International Journal of Advanced Research in Engineering and Technology (IJARET), vol. 11, no. 4, pp. 346-351, 2020, doi: https://ssrn.com/abstract=3599786.

[49] A. Akinbolati, M. Ajewole, "Investigation of path loss and modeling for digital terrestrial television over Nigeria," in International Journal, Heliyon, 2020 Jun, vol. 6, no. 6, doi: 10.1016/j.heliyon.2020.e04101.

Akansha Gupta

Akansha Gupta received her B.Tech. degree in Electronics and Instrumentation Engineering from the UP Technical University, India, in 2005, and her M.Tech. degree in Computer Science Engineering from Uttarakhand Technical University India, in 2009. She is currently pursuing her Ph.D. degree in Computer Science engineering in developing future AI channel models for next generation 5G mobile cellular networks. Her research interests include Machine learning, wireless communication, IoT, 5G network, random matrix theory, and information theory.

Prof. (Dr.) Kamal Ghanshala

Prof. Kamal Ghanshala is an engineer, entrepreneur and a philanthropist with Bachelor's and Master's in Computer Science and Engineering. He has received his doctoral degree in Computer Science from Kumaon University, Nainital-India.He has received recognition for his research in conferences held at, Croatia, Denmark, Johannesburg, Turkey, London, Paris Germany and Thailand. He received the Visionary Edupreneur of India award 2017 from former president of India and handled many research projects. He has also received excellence award in the field of higher education in the international summit organized in Newyork, USA. He has founded two universities as a President at Uttarakhand-India, Graphic Era Deemed to be University and Graphic Era Hill University. His research interests center around the optimization in wireless multimedia networks, stochastic optimization method, and graphical approach for information processing.

Prof. (Dr.) R.C. Joshi

Dr. R.C. Joshi former Prof. E. & C.E. Department at IIT Roorkee and Chancellor at Graphic Era Deemed to be University Dehradun, received his B.E degree from NIT Allahabad in1967, M.E.1st Div. with Honors and Ph.D Degree from Roorkee University, now IIT Roorkee, in 1970 & 1980 respectively. He worked as a Lecturer in J.K Institute, Allahabad University during 1967-68. He had been Head of Electronics & Computer Engineering from Jan 1991-1994 & Jan. 1997 to Dec. 1999. He was also the Head of Institute Computer Centre, IIT Roorkee from March 1994 to Dec. 2005.He was on short visiting Professor's Assignment in University of Cincinnati, USA. University of Minnesota, U.S.A & Macquarie University Sydney Australia also visited France under Indo-France collaboration program during June 78 to Nov. 79. Dr. Joshi has guided 27 Ph.D, 250 M.Tech, Dissertation, 75 B.E Projects. He had taught more than 25 subjects in Computer Engineering, Electronics Engineering & Information Technology.