The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-4/W9-2024
GeoAdvances 2024 – 8th International Conference on GeoInformation Advances, 11–12 January 2024, Istanbul, Türkiye

SPATIAL PREDICTION OF RECEIVED SIGNAL STRENGTH FOR CELLULAR COMMUNICATION USING SUPPORT VECTOR MACHINE AND K-NEAREST NEIGHBOURS REGRESSION

G. M. Perihanoglu [1], H. Karaman [1],

[1] ITU, Civil Engineering Faculty, Dept. of Geomatic Engineering, 80626 Maslak Istanbul, Turkey - (perihanoglu, karamanhi)@itu.edu.tr

KEY WORDS: Received Signal Strength (RSS), Geographic Information Systems (GIS), Support Vector Regression (SVR), K-Nearest Neıghbours (KNN), Machine Learning

ABSTRACT:

Signal strength maps are of great importance for cellular system providers in network planning and operation. Accurate prediction of signal strength is important for solving problems such as link quality. In this study, Received Signal Strength (RSS) prediction model is proposed for the 900 MHz band in the Van Yüzüncü Yıl University campus environment by using machine learning regression methods such as K- Nearest Neıghbours (KNN) and Support Vector Regression (SVR) together with Geographic Information Systems. For the training of this model, signal strength values taken from the RF Spectrum Analyser at different locations and distances were used. In addition, spatial data sets such as the digital elevation model, location of base stations and measurement stations, building heights and location, and land use/cover were used in the model. The effect of these data sets on RSS power is included in the model. The model aims to predict RSS accurately, visualize the estimated signal strength, and analyze the signal field strength coverage. Different kernels from the SVR model such as Polynomial, , and Sigmoid were tested. To increase the success of the model, appropriate parameter values were selected and configured according to SVR and KNN methods. For 900 MHz, the performances of SVR and KNN models were compared and the results of the models were verified using root mean squares (RMSE). Among the measured data, the lowest prediction is found in KNN Manhattan. According to the results of the simulation was observed that the SVR model created with spatial data performs better for Signal Strength. Finally, the lowest RMSE value (1.71 dB) was obtained from the Sigmoid kernel in the best signal strength estimation SVR model. The SVR model is recommended for Campus Area signal strength estimation.

1. INTRODUCTION

Cellular planning and optimization have become essential in wireless communication because of their importance in signal characterization. The solution to many decisions and planning issues in modern communication networks depends on accurate coverage estimates (Ojo et al., 2022). According to several studies, predicting network coverage areas is challenging in theory (Amaldi et al., 2008; Y. Wang et al., 2010). Mobile network coverage prediction, on the other hand, can be solved mathematically by utilizing computer algorithms and beginning assumptions for the planning of communication networks (Erunkulu et al., 2020). In cellular propagation models used for coverage estimation, obstacles in the signal path between the transmitter and receiver significantly affect the received signal strength (Anderson et al., 2008). Radio waves cannot reach the receiver directly along the Line of Sight (LoS) path due to obstacles such as buildings, trees, hills, and mountains. This situation that occurs between the base stations, as well as the receiver, is known as signal strength loss (Anusha et al., 2017).

In recent years, machine learning methods have been considered to play an important role in cellular communication. Machine learning-based models have started to be used more for coverage estimation problem than experimental and deterministic signal path loss models due to the extensive data used to train the model (Moraitis et al., 2021a) and (Nuñez et al., 2023).

Signal strength estimation based on SVR and KNN methods in the literature, 1800 and 2100 MHz (Gideon et al., 2017), 853 MHz (A Timoteo et al., 2014) in urban areas as well as for indoor environments (Ault et al., 2005) and (Polak et al., 2021). On the basis of the current literature, a few research studies integrate spatial data into different machine-learning techniques for predicting signal strength in the campus area.

Geographic Information System (GIS) is a suitable tool to address the requirements of cellular systems and assess precise signal strength coverage areas. Furthermore, GIS is effective in solving complex issues that cellular network planning engineers deal with, such as optimal frequency planning by the terrain environment (Wagen and Rizk, 2003). Among the reviewed scientific literature, there are many studies on signal strength applied for coverage analysis in cellular network communication. In the literature, many researchers have used geostatistical interpolation techniques, especially the Kriging method (Braham et al., 2017) and (Mezhoud et al., 2020) and (Ponce-Rojas et al., 2011), for coverage estimation. In studies in the literature, GIS has been used to visualize the propagation and coverage of signals and to perform spatial analyses in the planning of networks (Q. Wang et al., 2020). Moreover, in the literature, some studies develop solutions to the problem of signal weakness caused by the unequal distribution of base stations (Chen et al., 2012) and show the spatial pattern of telecommunication signal quality and speed (Septian et al., 2021).

In this study, signal strengths were calculated using K- Nearest Neıghbours (KNN) and Support Vector Regression (SVR)

machine learning methods. Signal strength estimation is a supervised regression problem therefore it can be solved by supervised machine learning algorithms such as K- Nearest Neıghbours (KNN) and Support Vector Regression (SVR)(Zhang et al., 2019). Selecting the input data appropriately while setting up the model makes the model more efficient and flexible as well as helping to reduce the complexity of the solution. The first step is to create preliminary features based on knowledge of electromagnetic wave propagation. Given the influence of the base station, geographical location, survey stations, environment, and other factors on the received signal strength, a large number of characteristics are needed to characterize these factors. This study aims to support the planning of base stations in the campus area and to automatically calculate the signal strength with SVR and KNN regression methods by integrating spatial data such as signal measurements, land use, and building heights into the model. Van Yüzüncü Yıl University Campus area has been selected as the study area. In this study, Geographic Information Systems (GIS) and machine learning methods such as SVR and KNN are integrated. A comparison of the 900 MHz frequency band received signal strength estimates of the proposed models is presented. The contributions of this study are summarised as follows:
Support vector regression (SVR) and K- Nearest Neıghbours (KNN) models have been developed with comprehensive data sets to accurately predict signal strength in cellular systems. The tuning of various hyperparameters of the models is discussed to optimize the SVR and KNN machine learning algorithms for accurate system design. The performances of SVR and KNN machine learning are compared using RMSE and R-squared. It was observed that spatial results were obtained by including the spatial data created using Geographic Information Systems (GIS) in the model. In addition, a signal strength map of the campus area has been created using spatial analysis and visualization approaches provided by GIS. The rest of this study is organised as follows. In Section II, the study area, the data used and the methods applied has been discussed. In section III, the model development process, including data processing, model training, and model hyperparameter tuning has been described. Discussion and results have been presented in Section IV. Section V has summarized the conclusion of the study.

2. METHODOLOGY

2.1 Study Area and Measurement Setup

Van Yüzüncü Yıl University campus area has been selected as the study area. Signal measurements have been performed in the campus area to accurately train and test the machine learning model used in this study. Three different types of data sets were used as input data to the model in order to implement the simulation. Data set: Characteristics of base stations; location, height, antenna power, and gain and frequency, Geographic Information; Digital Elevation Model (DEM) and Land use/cover map, building height, Characteristics of mobile stations; location, height, distance from the base station and measured signal strength at the location.
Van Yüzüncü Yıl University is located in the city of Van in the Eastern Anatolia region of Turkey. The elevation of the campus area varies between 1642-1725 metres. There are four existing base stations in the study area and test measurements have been taken at 250 locations within the campus. At each location, 10 measurements have been taken for 900 MHz. The total number of samples is 2500. Information about the study area is shown in Figure 1. The land use/cover data used in the study has been obtained using CORINE (Coordination of Information on The Environment) infrastructure. The data obtained has been simplified from the CORINE classification system and the study area has been divided into 4 classes. In this study, Python software has been used to implement analyses using Support Vector Regression (SVR) and K-Nearest Neıghbours (KNN) methods. Additionally, ArcGIS 10.8 Geographic Information System software has been employed to visualize signal strength maps and perform spatial analyses.

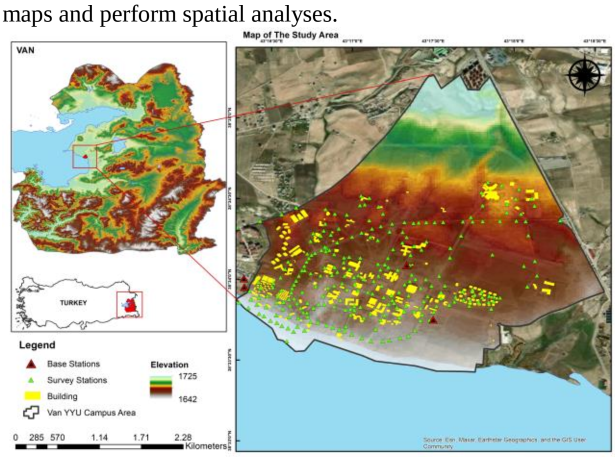

Figure 1. Location map of Van YYU campus area.

2.2 Model Developments

This section introduces the methodology and model developments. Two machine learning-based models were introduced: Support Vector Regression and K-Nearest Neighbours. These models have been adapted to signal strength prediction, as in a supervised learning regression problem. Both models have been trained using the same datasets to conduct a justifiable comparison of their performances.

2.2.1 Support Vector Regression (SVR) Model

SVR is used successfully to address nonlinear regression issues for signal strength prediction. The fundamental concept involves transforming input data from a space with lower dimensions to a space with higher dimensions using non-linear functions. Subsequently, the objective is to identify an optimal hyperplane within this high-dimensional feature space, aiming to maximize the distance of the samples along this hyperplane. (Moraitis et al., 2021a). The following linear expression gives the specified hyperplane.

$$\hat{y}(x)=w^{T}\varphi(x)+b \tag{1}$$

where w = standard vector that controls the hyperplane's direction.
b = bias
x = input vector
φ(.)= function for non-linear mapping
After that, using Lagrange multipliers (i.e. support vectors), the estimation function can be determined as follows:

$$\hat{y}(x)=\sum_{i=1}^{N}(a_i-a_i^*)K(x_i,x)+b \tag{2}$$

Where $K(.,.)$ = Kernel function

a_i & a_i^* = Support vectors

The kernel function is the most important feature determining the model training effect in the Support Vector Regression (SVR) method. Therefore, the most appropriate kernel function should be selected according to the data set. This paper discusses the use of polynomial, sigmoid, and linear kernel types. The expressions for each kernel are presented below (3,4,5)

$$K_{Polynomial} = \left(\gamma . x_i^T x + c\right)^d \tag{3}$$

$$K_{Linear} = x_i^T x \tag{4}$$

$$K_{Sigmoid} = \tanh(\gamma (x_i^T x) + c) \tag{5}$$

where d= polynomial degree
c= the free constant term (c>0), and both are changeable.
γ = the input data scaling parameter.

2.2.2 K-Nearest Neighbours (KNN)

The k-Nearest Neighbours (KNN) technique is a straightforward supervised machine learning method that can deal with regression and classification issues. The basic concept involves using distance measures to identify the K training samples that are nearest to the sample being predicted. Subsequently, predictions are made based on the outcomes of these K neighbors (Moraitis et al., 2020). For regression issues, the final estimated value is obtained by averaging the K-nearest neighbours. There are various methods to calculate this distance. The most widely known methods are Euclidian, Manhattan, and Chebyshev (Vu Thanh Quang et al., 2022). The Euclidean distance is the most typical application of distance. The Euclidean distance is described as in equation (6).

$$D_{Euclidean}(x,y) = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{1/2} \tag{6}$$

Manhattan distance, often known as city block distance is a highly popular distance. Manhattan distance distance is given by (7):

$$D_{Manhattan}(x,y) = \sum_{i=1}^{n} \left| x_i - y_i \right| \tag{7}$$

Between two vectors, the Chebyshev distance, minimax metric, or infinity norm is the maximum of their absolute magnitudes along the vector dimension. The Chebyshev distance (Moghtadaiee and Dempster, 2015) is defined as (8):

$$D_{Chebyshev}(x,y) = max_i \left| x_i - y_i \right| \tag{8}$$

Where x, y = input vector

3. TRAINING AND EVALUATION

3.1 Data Preprocessing

A geographical database has been created in the ArcGIS environment for signal strengths measured in the study area and other data. Spatial join analysis has been applied with ArcGIS software to determine which terrain class of the base stations in the measurement test areas. The distances of each measurement location to the base stations have been calculated according to the spatial proximity analysis. In addition, the elevation of the measurement stations has been obtained by overlapping them with the vectorised digital elevation model. The heights of the buildings within the campus area between the measurement station and the base station have been averaged for building height data. In the database, the data collected from the field and the data obtained from the results of spatial analyses spatial analyses have been combined also converted into Csv. format. The converted data has been transferred to the Python software. Since machine learning algorithms do not work directly on categorical data, One Hot Encoding has been performed to convert the land classes in the data set into numerical data. The terrain class of the study area consists of 4 classes as discontinuous urban structure, pasture, non-irrigated agricultural areas, and water bodies. According to this technique, urban structure has been assigned as 1 and other classes as 0.

3.2 Model Training

The samples obtained for the implementation of received signal strength modeling based on machine learning methods are divided into training samples and test samples. Each sample contains a record of the signal strength from a terrain and the corresponding input features. Depending on the training examples and the selected algorithms, the models are trained and their optimal hyperparameters are tuned.
Differences in scales between input variables can increase the difficulty of the modeled problem. It is important to note that Support Vector Regression (SVR) and the K Nearest Neighbours (KNN) are affected by the size of the input space. Standardization is necessary for machine learning algorithms such as KNN and SVR as the data have input values at different scales. Accordingly, the standardization should take place before the model training. The flow chart of the study is given in Figure 2.

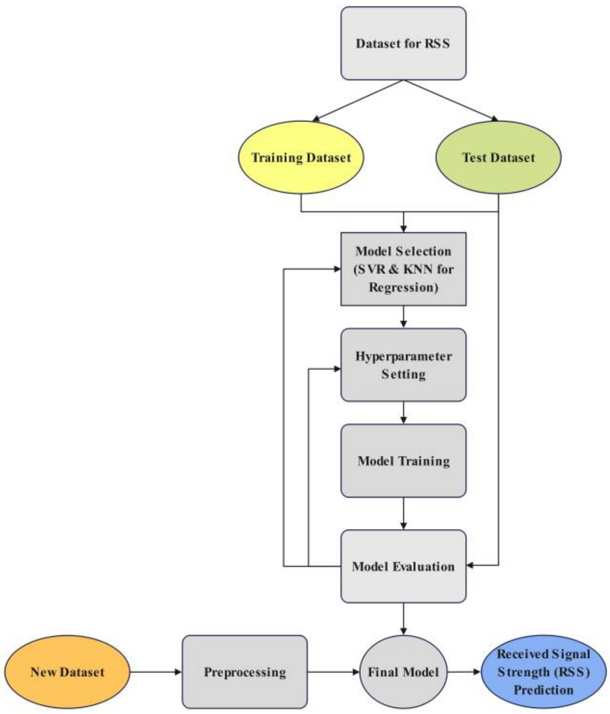

Figure 2. Procedure of SVR & KNN for regression-based Received Signal Strength (RSS) prediction.

Root Mean Square Error (RMSE) and R Square are common measurement metrics for evaluating the performance of regression-based algorithms (Moraitis et al., 2020) Root Mean Square Error (RMSE, R-squared and the Mean Absolute Error (MAE) have been selected from the model evaluation metrics to ensure that the signal strength values, which are the output of the model, are closest to the real values. In equations 9,10 and 11 below, the expression of RMSE, R squared and MAE are given.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(RSS_i^{sim} - RSS_i^{pred})^2} \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{i}(RSS_i^{pred} - RSS_i^{sim})^2}{\sum_{i=1}^{i}(RSS_i^{pred} - RSS_i^{mean})^2} \tag{10}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|RSS_i^{sim} - RSS_i^{pred}\right| \tag{11}$$

Where RSS_{sim}= Simulated received signal strength values

RSS_{pred} = Predicted received signal strength values

i = The index of the measured sample

N= Total number of the sample

3.3 Hyperparameter Tuning of SVR and KNN

Parameters that are not explicitly learned in predictive machine learning models can be tuned by searching a parameter space for the best cross-validation value. These types of parameters are known as hyperparameters. Techniques for tuning hyperparameters play an important role in the search for a suitable hyperparameter in machine learning techniques. Furthermore, machine learning methods are based on complex hyperparameters leading to optimization challenges. Furthermore, if we try for any possible combination of hyperparameters, it can be time-consuming to determine the values of the hyperparameters. In this study, the use of sigmoid, linear and polynomial kernels for the SVR model is considered. Tuning for this technique involves optimizing the hyperparameters: C represents the regularisation parameter. It must be entirely positive. $\gamma > 0$ indicating that the kernel coefficient. Epsilon (ε) controls the complexity of the regression functions. The data to be included in the regression should be at a certain epsilon distance and this epsilon value should not be negative.

KNN is an algorithm that works by assuming that similar things are close to each other. The approach known as KNN works by measuring the distances between a query and all points in the data, then selecting the nearest neighbours (k) and averaging the most common ones. Choosing the proper number of neighbours (k) can result in the best fit in the regression issue, which can be accomplished by experimenting with different k's and selecting the one that produces optimal outcomes. In other words, the number k for KNN is necessary for optimal performance. If the value of k is set too low, the model becomes more complex, which in turn raises the risk of overfitting, particularly when the nearest neighbour exhibit noise-like behaviour. In contrast, large values of k simplify the form of the model but influence the prediction accuracy between neighbouring samples(Moraitis et al., 2021b).

Cross-validation serves as a statistical resampling technique employed to assess a machine learning model's performance on distinct data sets in a manner that is both objective and precise The k-fold cross-validation method is a typical type of cross-validation that is commonly used to assess model accuracy. The k-fold cross-validation procedure is important in determining the stability of a machine learning model. Using k-fold cross-validation, the dataset is divided into k subsets, and the model is trained and evaluated k times. During each iteration, a different subset serves as the test set, while the remaining k-1 subsets are used for training (A Timoteo et al., 2014). This process generates k distinct performance metrics, typically averaged for a more robust estimate of the model's performance. The technique mitigates variability associated with a single train-test split, offering a comprehensive evaluation of the model's likely performance on unseen data. In this study, k is set to 10 for the k fold Cross Validation (k-fold CV) method.

Finally, Table 1 summarises the hyperparameters applied during the training process for SVR, KNN methods.

Model	Hyperparameters
SVR Linear	γ =1 , ε =0.01 , c =2
SVR Polynomial	γ = 0.1, ε = 0.1, c = 3
SVR Sigmoid	γ = 0.1 , ε =0.01 , c =1
KNN	metric = Euclidean, k = 6
KNN	metric = Manhattan, k = 6
KNN	metric = Chebyshev, k = 6

Table 1. Selected hyperparameters for SVR and KNN

4. RESULTS AND DISCUSSIONS

This section evaluates the performance of proposed machine learning regression algorithms based on a simulation conducted in a campus area located in the city of Van, Turkey. The evaluation is carried out using real signal strength measurements at 900 MHz frequency and spatial data. The data set used in this study consists of signal strength data, carrier frequency, land use data, information about base stations, terrain elevation, altitude, and building height data collected on the Van Yüzüncü Yıl University campus. In the study, Python JupyterLab has been used as the programming language for model development and analysis, and ArcGIS software was used for spatial analysis and mapping. The data set includes categorical and numerical data types. The terrain classes of the study area are; discontinuous urban structure, pasture, non-irrigated arable land, and water bodies, and the trained model was tested over these classes. The categorical data in the study area consists of land classes such as discontinuous urban structure, pasture, non-irrigated arable land, and water bodies. Categorical data were converted to numerical form using One Hot Encoding transformation, and the effect of this data was incorporated into the signal strength estimation model. For the K-Nearest Neighbours (KNN) and Support Vector Regression (SVR) methods used in the study, 80% of the data set is divided into 80% for training and 20% for prediction.

For each machine learning method analyzed in the study, all statistical error measures are calculated, and the results are presented. The Python JupyterLab software Scikit Learn library has been used, providing an integrated environment for all training and testing procedures, data preparation, machine learning, and visualization of results. Root Mean Square Error (RMSE), R Square (R^2) and the mean absolute error (MAE) performance metrics have been used to select the most appropriate parameters for KNN and SVR algorithms in the data set.

The performance of SVR and KNN algorithms is evaluated by computer simulations for the hyperparameters mentioned in Table 1. The GridsearchCV method available in the scikit learn library was applied for this hyperparameter optimization. The performance of SVR and KNN algorithms, which are machine learning methods, is evaluated in the testing phase. The numerical results of the corresponding statistical metrics for the adopted methods are listed in Table 2.

Model	RMSE (dB)	R^2 (dB)	MAE(dB)
SVR Linear	3.16	0.81	2.66
SVR Polynomial	2.91	0.84	2.27
SVR Sigmoid	1.71	0.94	1.38
KNN Euclidean	3.13	0.81	2.31
KNN Manhattan	3.60	0.75	2.69
KNN Chebyshev	3.50	0.77	2.43

Table 2. Performance metrics of the examined machine learning methods in the campus environment

The performance of the SVR algorithms was evaluated for the three kernels mentioned in Table 1. Table II provides a statistical analysis of the SVR algorithms, where the RMSE, R square, and MAE. Table I shows the best configuration settings of the SVR algorithms (C, ε, and γ) as well as the parameters of each kernel. According to what is explained about SVR for nonlinear learning problems in Section III and the results in Table 2, the Sigmoid kernel is the most appropriate for our data set. Figure 3 illustrates a comparison of the measurements corresponding to the sample set obtained from the driving test and the predictions using the SVR algorithms for Polynomial, Linear, and Sigmoid kernels. It can be seen that the sigmoid kernel is the best choice among the three kernels.

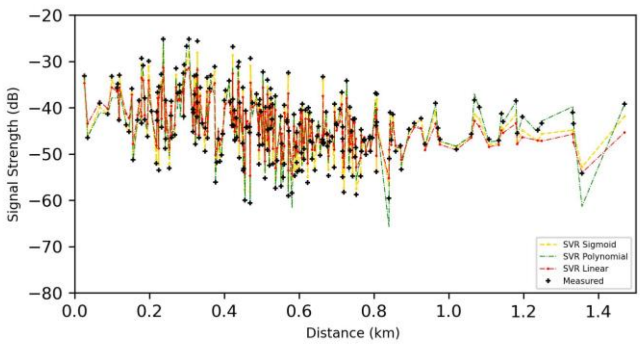

Figure 3. Comparison of Received Signal Strength (RSS) prediction versus distance using SVR algorithms.

We can see that Linear SVR shows an RMSE of 3.16 dB, MAE of 2.66 dB, and R-squared of 0.81 dB. Polynomial and Sigmoid SVR offer an RMSE of 2.91 dB, MAE of 2.27 dB and R-squared of 0.84 dB, and RMSE of 1.71 dB, MAE of 1.38 dB and R-squared of 0.94 dB, respectively. SVR with Sigmoid kernel has higher performance compared to the other evaluated kernel SVR models.

The SVR results mentioned above can be compared with those presented in (A Timoteo et al., 2014); Here, Polynomial, Gaussian, and Laplace kernels are compared using signal strength, height, as well as distance data at 853 MHz in the urban environment. In the study, an RMSE value of 3.47 dB was obtained using Polynomial SVR. In addition, in similar studies in urban areas, an R-squared score of 0.874 dB (Gideon et al., 2017) was obtained by adjusting the optimal kernel and hyperparameters using the SVR method. On the other hand, signal path loss studies (Moraitis et al., 2021a) reported MAE 5.1, RMSE 6.5, and R square 0. 74 with polynomial SVR.

Figure 4 shows the signal strength predictions of KNN methods according to different distance measurements. According to comparable error metrics, the three models perform similarly, as can be observed in Table 2. Figure 4 presents a comparison between the measurements from the sample set obtained during the driving test and the predictions made using different distance metrics such as Euclidean, Manhattan, and Chebyshev. When Table 2 and Figure 4 are analyzed, it is seen that the signal strength estimated with the KNN method using the Euclidean distance shows the best performance among the three different distance measurements.

More specifically, we can see that the KNN Euclidean shows RMSE of 3.13 dB, MAE of 2.31 dB, and R-square of 0.81 dB. The KNN method using Manhattan and Chebyshev distance metrics yields an RMSE of 3.60 dB, MAE of 2.69 dB, an R-square of 0.75 dB, and an RMSE of 3.50 dB, MAE of 2.43 dB as well as R-square of 0.77 dB, respectively. Considering the KNN methods calculated according to these three distance metrics, although the values are statistically close to each other, KNN Euclidean shows the best performance.

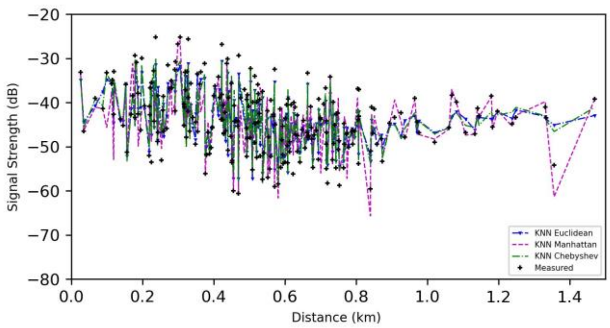

Figure 4. Comparison of Received Signal Strength (RSS) prediction versus distance using KNN algorithms.

The KNN results mentioned above can be compared with findings from other studies in the literature. Among the studies in the literature, for 3.7 GHz (Moraitis et al., 2021b) 4.2 dB RMSE, 3.2 dB MAE, and 0.92 R-squared value were obtained according to the KNN method based on k = 5 and Euclidean distance (as in this study). In addition, considering the measurements in a mixed urban environment (including LOS and NLOS conditions) at frequencies between 2120 and 2160 MHz, an RMSE value of 2.1 dB for LOS and 3.4 dB for NLOS

https://doi.org/10.5194/isprs-archives-XLVIII-4-W9-2024-291-2024 | © Author(s) 2024. CC BY 4.0 License.

was observed (Moraitis et al., 2020). On the other hand, it has been reported that the best performance is Manhattan distance when the KNN method is applied according to Euclidean, Manhattan, and Chebyshev distance metrics (Moghtadaiee and Dempster, 2015) (the metrics used in this study) for indoor signal strength estimation. Finally, in the air-to-air UAV scenario, when applying KNN with k=5 and using the Euclidean distance metric, the results based on simulated data at 2.4 GHz in urban locations show an 8.90 dB RMSE and a 4.56 dB MAE (Zhang et al., 2018).

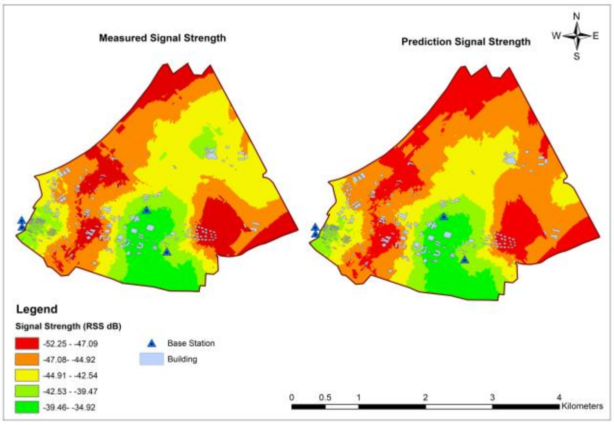

Figure 5. Comparison of the Received Signal Strength (RSS) map with the Signal Strength map using the SVR sigmoid kernel model for 900 MHz.

Figures 5 and 6 show the signal strength maps based on spatial covering the entire area of the dataset according to the Kriging method of the signal strength values estimated by the Support vector regression (SVR) and K Nearest Neighbour (KNN) algorithms and the actual signal strength values measured from the field. Figure 5 shows the coverage map of the SVR model according to the values obtained from the Sigmoid kernel. Figure 6 shows the coverage map according to the values obtained from the KNN Euclidean distance metric.

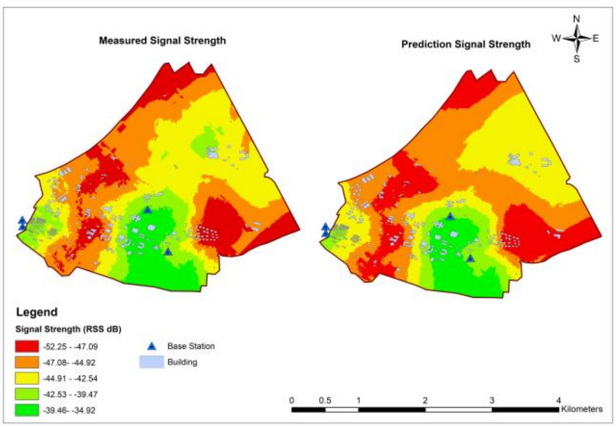

Figure 6. Comparison of the Received Signal Strength (RSS) map with the Signal Strength map using the KNN Euclidean distance model for 900 MHz.

When looking at the maps in Figures 5 and 6, it is observed that the signal strength is higher in areas close to the base station, indicated by green regions. However, in locations associated with the discrete urban structure characterized by building density, the signal strength decreases. In addition, the signal strength decreases as you distance far from the base stations (north of the study area, yellow and red regions).

According to Figures 3, 4, 5, and 6, the SVR sigmoid model exhibits lower performance than the KNN Euclidean model as the distance from the base stations increases. In addition, there is a decrease in signal strength due to the high number of buildings and the height of the buildings in the red regions close to the base station but where the signal strength decreases.

5. CONCLUSIONS

This study proposes a Support Vector and K Nearest Neighbour-based model based on Geographic Information Systems and machine regression algorithm to predict the signal strength in a campus environment. For this purpose, SVR and KNN models were applied and evaluated. The signal strength data measured in the field has been used to create a database through a geographic information system, and the spatial merging process of which base station is located within which land class.

SVR and KNN models were generated according to the modelled data set signal strengths at 900 MHz in the Van Yüzüncü Yıl Campus area were taken and used in the training phase of the model. Features with a strong correlation have been selected to estimate the signal strength, aiming to increase the efficiency of the model. Sigmoid, Linear, and Polynomial kernels in the SVR algorithm and Euclidean, Manhattan, and Chebyshev distance metrics for KNN have been tested for the performance of the models. Then, according to this problem, appropriate model parameters have been selected to improve the capability of the model. The results confirmed that the Sigmoid kernel is the best option among the analyzed kernels in the case of SVR. The poorest performance has been seen at KNN Manhattan. The prediction values obtained from the models with high-performance values from SVR and KNN methods have been used to create a signal-strength surface for the campus environment. For this purpose, the Kriging interpolation method has been used by utilizing the Geographical Information System. This method can be recommended and practically used for signal strength estimation in cellular planning tools. In this sense, GIS can be used to generate the necessary spatial parameters for cellular coverage prediction models and is also an important tool for the assessment of cellular network propagation quality. Finally, future work will utilize GIS and different machine learning models by incorporating spatial data in urban areas and will be able to further improve predictability using these hybrid methods.

REFERENCES

A Timoteo, R. D., Cunha, D. C., C Cavalcanti, G. D. 2014. A Proposal for Path Loss Prediction in Urban Environments using Support Vector Regression. *The Tenth Advanced International Conference on Telecommunications*.

Amaldi, E., Capone, A., Malucelli, F. 2008. Radio planning and coverage optimization of 3G cellular networks. *Wireless Networks*, *14*(4), 435–447. https://doi.org/10.1007/s11276-006-0729-3

Anderson, H., Hicks, T., Kirtner, J. 2008. The Application of

Land Use/Land Cover (Clutter) Data to Wireless Communication System Design. *EDX Wireless*, 2–11.

Anusha, V. S., Nithya, G., Rao, S. N. 2017. Comprehensive survey of electromagnetic propagation models. *Proc. Int. Conf. Commun.Signal Process. (ICCSP)*, 1457–1462.

Ault, A., Coyle, E., Zhong, X., Coyle, E. J. 2005. K-nearest-neighbor analysis of received signal strength distance estimation across environments. *Proceedings of the First Workshop on Wireless Network Measurements* .

Braham, H., Jemaa, S. Ben, Fort, G., Moulines, E., Sayrac, B. 2017. Fixed Rank Kriging for Cellular Coverage Analysis. *IEEE Transactions on Vehicular Technology*, *66*(5). https://doi.org/10.1109/TVT.2016.2599842

Chen, X., Wu, H., Trı, T. M. 2012. Field strength prediction of mobile communication network based on GIS. *Geo-Spatial Information Science*, *15*(3), 199–206. https://doi.org/10.1080/10095020.2012.720457

Erunkulu, O. O., Zungeru, A. M., Lebekwe, C. K., Chuma, J. M. 2020. Cellular Communications Coverage Prediction Techniques: A Survey and Comparison. In *IEEE Access* 8, 113052–113077.

Gideon, K., Nyirenda, C., Temaneh-Nyah, C. 2017. Echo state network-based radio signal strength prediction for wireless communication in Northern Namibia. *IET Communications*, *11*(12), 1920–1926. https://doi.org/10.1049/IET-COM.2016.1290

Mezhoud, N., Oussalah, M., Zaatri, A., Hammoudi, Z. 2020. Hybrid Kriging and multilayer perceptron neural network technique for coverage prediction in cellular networks. *International Journal of Parallel, Emergent and Distributed Systems*, *35*(6), 682–706. https://doi.org/10.1080/17445760.2020.1805609

Moghtadaiee, V., Dempster, A. G. 2015. *International Global Navigation Satellite Systems Society IGNSS Symposium*.

Moraitis, N., Tsipi, L., Vouyioukas, D. 2020. Machine learning-based methods for path loss prediction in urban environment for LTE networks. *International Conference on Wireless and Mobile Computing, Networking and Communications*. https://doi.org/10.1109/WiMob50308.2020.9253369

Moraitis, N., Tsipi, L., Vouyioukas, D., Gkioni, A., Louvros, S. 2021a. Performance evaluation of machine learning methods for path loss prediction in rural environment at 3.7 GHz. *Wireless Networks*, *27*(6), 4169–4188. https://doi.org/10.1007/S11276-021-02682-3/TABLES/1

Moraitis, N., Tsipi, L., Vouyioukas, D., Gkioni, A., Louvros, S. 2021b. Performance evaluation of machine learning methods for path loss prediction in rural environment at 3.7 GHz. *Wireless Networks*, *27*(6), 4169–4188. https://doi.org/10.1007/S11276-021-02682-3/FIGURES/6

Nuñez, Y., Lovisolo, L., da Silva Mello, L., Orihuela, C. 2023. On the interpretability of machine learning regression for path-loss prediction of millimeter-wave links. *Expert Systems with Applications*, *215*, 119324. https://doi.org/10.1016/J.ESWA.2022.119324

Ojo, S., Murat Akkaya, Sopuru, J. C. 2022. *An ensemble machine learning approach for enhanced path loss predictions for 4G LTE wireless networks*. *35*(7), 1–30. https://doi.org/10.1002/dac.5101

Polak, L., Rozum, S., Slanina, M., Bravenec, T., Fryza, T., Pikrakis, A. 2021. Received signal strength fingerprinting-based indoor location estimation employing machine learning. *Sensors*, *21*(13). https://doi.org/10.3390/s21134605

Ponce-Rojas, J., Vidal-Beltrán, S., Acevedo-Mosqueda, M. A., Jimenez-Licea, M. 2011. A Geographic Information System Applied to Coverage Maps of 3G Cellular Communications Networks. *Journal of Geographic Information System*, *3*, 140–144. https://doi.org/10.4236/jgis.2011.32010

Septian, M. R., Susilowati, M. H. D., Zulkarnain, F. 2021. *Spatial patterns of telecommunication signal quality, internet speed, and online game latency in East Jakarta*. *5*(1).

Vu T. Q., Dinh V. L., To T. T., 2022. Propagation Path Loss Models at 28 GHz Using K-Nearest Neighbor Algorithm. *Journal of Communication and Computer*, *19*(1). https://doi.org/10.17265/1548-7709/2022.01.001

Wagen, J. F., Rizk, K. 2003. Radiowave propagation, building databases, and GIS: Anything in common? A radio engineer's viewpoint. *Environment and Planning B: Planning and Design*, *30*(5), 767–788. https://doi.org/10.1068/b2958

Wang, Q., Zhao, X., Lv, Z., Ma, X., Zhang, R., Lin, Y. 2020. Optimizing the ultra-dense 5G base stations in urban outdoor areas: Coupling GIS and heuristic optimization. *Sustainable Cities and Society*, *63*. https://doi.org/10.1016/j.scs.2020.102445

Wang, Y., Si, Y., Leung, H. 2010. A novel inversion method for outdoor coverage prediction in wireless cellular network. *IEEE Transactions on Vehicular Technology*, *59*(1), 36–47. https://doi.org/10.1109/TVT.2009.2033702

Zhang, Y., Wen, J., Yang, G., He, Z., Luo, X. 2018. Air-to-Air Path Loss Prediction Based on Machine Learning Methods in Urban Environments. *Wireless Communications and Mobile Computing*, *2018*. https://doi.org/10.1155/2018/8489326

Zhang, Y., Wen, J., Yang, G., He, Z., Wang, J. 2019. *Path Loss Prediction Based on Machine Learning: Principle, Method, and Data Expansion*. https://doi.org/10.3390/app9091908

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-4-W9-2024-291-2024