ISSN 2347–3657

International Journal of Information Technology & Computer Engineering

Volume 13, Issue 2s, 2025

Comparative Analysis of Machine Learning Algorithms for 5G Coverage Prediction: Identification of Dominant Feature Parameters and Prediction Accuracy

Mohammed Ayaan[1], Mohammed Akram[2], Mohd Noman Ahmed[3] , Dr. Syed Asadullah Hussaini

[1,2,3]B.E Students, Department Of Artificial Intelligence & Data Science Engineering, ISL Engineering College, Hyderabad. India.

[4]Associate Professor in , Artificial Intelligence & Data Science Engineering, ISL Engineering College, Hyderabad, India.

mdayaan0050@gmail.com

ABSTRACT *5G technology is a key factor in delivering faster and more reliable wireless connectivity. One crucial aspect in 5G network planning is coverage prediction, which enables network providers to optimize infrastructure deployment and deliver high-quality services to customers. This study conducts a comprehensive analysis of machine learning algorithms for 5G coverage prediction, focusing on dominant feature parameters and accuracy. Notably, the Random Forest algorithm demonstrates superior performance with an RMSE of 1.14 dB, MAE of 0.12, and R2 of 0.97. The CNN model, the standout among deep learning algorithms, achieves an RMSE of 0.289, MAE of 0.289, and R2 of 0.78, showcasing high accuracy in 5G coverage prediction. Random Forest models exhibit near-perfect metrics with 98.4% accuracy, precision, recall, and F1-score. Although CNN outperforms other deep learning models, it slightly trails Random Forest in performance. The research highlights that the final Random Forest and CNN models outperform other models and surpass those developed in previous studies. Notably, 2D Distance Tx Rx emerges as the most dominant feature parameter across all algorithms, significantly influencing 5G coverage prediction. The inclusion of horizontal and vertical distances further improves prediction results, surpassing previous studies. The study underscores the relevance of machine learning and deep learning algorithms in predicting 5G coverage and recommends their use in network development and optimization. In conclusion, while the Random Forest algorithm stands out as the optimal choice for 5G coverage prediction, deep learning algorithms, particularly CNN, offer viable alternatives, especially for spatial data derived from satellite images. These accurate predictions facilitate efficient resource allocation by network providers, ensuring high-quality services in the rapidly evolving landscape of 5G technology. A profound understanding of coverage prediction remains pivotal for successful network planning and reliable service provision in the 5G era.*

The rapid development of 5G technology is transforming industries and enhancing connectivity through higher data speeds, ultra-low latency, and increased network capacity.

Efficient deployment of 5G networks requires accurate coverage prediction, ensuring seamless service across diverse geographical areas.

Traditional coverage prediction methods often rely on time-consuming and resource-intensive simulations, making them impractical for large-scale deployment.

Machine learning (ML) offers promising alternatives for real-time and scalable 5G coverage prediction by analyzing large datasets and identifying patterns that influence network performance.

ML algorithms can improve the accuracy of coverage predictions by considering complex environmental factors, user mobility, and signal propagation conditions.

I. INTRODUCTION

Machine learning (ML) is a set of methodologies for making predictions based on datasets and modeling algorithms. Methods based on machine learning have been used in a variety of fields, including speech recognition, image recognition, natural language processing, and computer vision. In general, machine learning techniques can be grouped into three main categories based on how they process and use data. This classification helps provide a better understanding of how machines learn from experience and make decisions. There are supervised learning, unsupervised learning, and reinforcement learning. Also, all machine learning methods rely on the type of information (input features) that is used for the training. Machine learning methods can be

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP531-540

classified as supervised learning and unsupervised learning. For classification or regression issues, supervised learning is used to learn a function or relationship between inputs and outputs. Unsupervised learning, on the other hand, is the process of extracting hidden rules or connections from unlabelled data.

PROBLEM STATEMENT

Accurate prediction of 5G coverage is crucial for optimizing network performance and resource allocation. Current methods are often computationally intensive and fail to incorporate real-time data, making them unsuitable for dynamic and large-scale networks. There is a need for efficient, scalable, and adaptive methods that can predict 5G coverage with high accuracy. This study explores machine learning approaches to address these limitations and improve 5G network deployment

EXISTING SYSTEM

These models may not account for dynamic changes in network environments, such as user density, obstacles, or signal interference, leading to suboptimal coverage predictions.Traditional prediction methods lack the ability to provide accurate updates and adaptive learning, limiting their usefulness in highly dynamic urban environments.Machine learning algorithms Navie Bayes , Linear regression algorithms with accuracy less than 90 percent. Existing 5G coverage prediction models often use ray-tracing or empirical models, which depend on detailed geographical data and propagation models but are computationally expensive.

DISADVANTAGES OF EXISTING SYSTEM

- Existing systems, particularly traditional algorithms like linear regression, often fail to capture non-linear relationships between feature parameter.
- limiting prediction accuracy in dynamic 5G network conditions.

PROPOSED SYSTEM

In proposed system we identify the dominant feature parameters and evaluating the prediction accuracy of different algorithms, we can gain insights into the effectiveness of these algorithms in predicting 5G coverage. By conducting a training using machine learning algorithms and use important features, this study aims to contribute to the advancement of 5G coverage prediction. The findings of this work can provide valuable insights for network operators and researchers in optimizing the deployment and performance of 5G networks.Evaluate the performance of machine learning classification algorithm : Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, Support Vector Machine (SVM), Gradiant boost, and CNN in training prediction data and evaluating the accuracy of the prediction results using machine learning and deep learning algorithms on 5G networks. The best performance of the final trained model is evaluated against best model.

ADVANTAGES OF PROPOSED SYSTEM

- The system incorporates a broader range of features, including terrain, urban density, weather, and real-time network data, allowing for more accurate and context-aware coverage predictions.
- Better performance

LITERATURE REVIEW

Based on the input data, supervised learning develops a function to predict a defined label. It might be either categorizing data (classification problem) or forecasting an outcome (regression algorithms). Predicting the coverage of 5G mobile networks also can be categorized as classification type problem [26]. Classification is a type of supervised machine learning in which the model attempts to predict the proper label of given input data. In classification, the model is fully trained on training data before being tested on test data and used to predict new unobserved data. The capacity of classification models to classify input data into several classes or categories based on patterns and correlations existing in the data makes them well-suited for this purpose. The objective of classifying various places or regions into groups is to forecast 5G coverage. The most significant characteristics that influence coverage can be found using classification models. This aids in network optimization by enabling a better knowledge of the variables that have a substantial impact on coverage quality. Between classification and regression techniques, there can occasionally be some ambiguity. Regression and classification can both be performed using a variety of algorithms, with classification simply being a regression model with a threshold applied. The number is classified as true when it exceeds the threshold and categorized as false when it is lower. A. LOGISTIC REGRESSION Logistic regression is a statistical method used to analyze the relationship between a dependent variable and one or more independent variables

METHODOLOGY OF PROJECT

As seen from the above figure, we can see how the data is divided into different sets and then trained for different models.

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP531-540

ISSN 2347–3657

Volume 13, Issue 2s, 2025

• The dataset was first divided into training set (80%) and pre-training set(20%). • The pre-training set was divided into pre-train(80%) and pre-test(20%) • Now, the training set is further is divided into train(80%) and validation set(20%). This train set is again divided into train(80%) and test set(20%). So, now I have train validation and test sets separate which are nonoverlapping. • The pretrain set was used to find the best models for the given dataset. I took best 4 models using pretest set. Their performance was compared based on their mean absolute errors. • Once the best 4 models were obtained, hyperparameters for these models were tuned and the best parameter was selected.

MODULE DESCRIPTION

SERVICE PROVIDER

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Train & Test 5G coverage Datasets, View Trained and Tested 5G coverage Datasets Accuracy in Bar Chart, View Trained and Tested 5G coverage Datasets Accuracy in Bar Chart, View Prediction Of 5G coverage Type, View 5G coverage Type Ratio, Download Predicted Data Sets, View v Type Ratio Results, View All Remote Users.

VIEW AND AUTHORIZED USERS

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

DATA COLLECTION

we will take data set created own dataset which has features as tweet data and labels as Forensic data or not.

DATA PROCESSING

Features are extracted from data set and stored in variable as xtrain variable and labels are stored in y train variable. Data is preprocessing by stscalar function and new features and labels are generated.

TESTING TRAINING

In this stage data is sent to testing and training function and divided in to four parts x test train, and y test train. Train variables are used for passing to algorithm where as test are used for calculating accuracy of the algorithm.

INITIALIZING MULTIPLE ALGORITHMS AND TRANING WITH LOGISTICS REGRESSION:

In this stage machine learning algorithms are initialized and train values are given to algorithm by this information algorithm will know what are features and what are labels. Then data is modeled and stored as pickle file in the system which can be used for prediction.Data set is trained with multiple algorithms and accuracy of each model is calculated and best model is used for prediction

PREDICT DATA

In this stage new data is taken as input and trained models are loaded using pickle and then values are preprocessed and passed to predict function to find out result which is showed on web application

ALGORITHM USED IN PROJECT :

CLASSIFICATION ALGORITHMS IN MACHINE LEARNING FOR 5G COVERAGE PREDICTION

Based on the input data, supervised learning develops a function to predict a defined label. It might be either categorizing data (classification problem) or forecasting an outcome (regression algorithms). Predicting the coverage of 5G mobile networks also can be categorized as classification type problem . Classification is a type of supervised machine learning in which the model attempts to predict the proper label of given input data. In classification, the model is fully trained on training data before being tested on test data and used to predict new unobserved data.

A. LOGISTIC REGRESSION

Logistic regression is a statistical method used to analyze the relationship between a dependent variable and one or more independent variables . A statistical technique called logistic regression is employed to examine the relationship between a dependent variable and one or more independent variables. It is frequently used for binary classification issues where there are two alternative outcomes for the dependent variable.

The use of logistic regression algorithms allows the formulation of equations and the calculation of probabilities, which are essential for classifying customers into different groups. Also, this algorithm model can be used by organizations to conduct marketing research, understand customer needs, and produce goods accordingly, leading to sustainable brand and network loyalty .

However, the use of this logistic regression algorithm has never been used in research related to predicting signal levels or coverage in an area. Therefore, this research will also try to use this algorithm to predict signal coverage levels, especially in 5G networks.

B. K-NEAREST NEIGHBORS (KNN)

The K-Nearest Neighbors (K-NN) algorithm is a supervised classification algorithm that can be used for classification and regression problems. It

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP531-540

ISSN 2347–3657

Volume 13, Issue 2s, 2025

categorizes items based on their closest neighbors. This is an example of situational learning. The Euclidean distance is used to calculate the distance between an attribute and its neighbors. The technique operates by locating the K nearest data points to the new data point and allocating the new data point to the class with the highest frequency of occurrence among its K nearest neighbors. Cross-validation or other performance measures can be used to determine the value of K. KNN has been utilized in a variety of wireless network applications, including localisation, beamforming, MIMO, anomaly detection, and network slicing . KNN can also be used in conjunction with other machine learning algorithms, such as deep learning, to boost performance .

The use of the KNN algorithm has previously been used in coverage prediction. In the research presented in developed a machine learning model to predict radio signal strength in certain geographic areas based on transmitter placement. The dataset consists of simulated power at each point in the neighborhood for a given set of transmitter locations. Various machine learning models, including generalized linear models (GLM), neural networks (NN), and k-nearest neighbors (KNN), were tried. Feature engineering approaches are used to improve the predictive performance. In this research, the K-nearest neighbor (KNN) model has the best performance with an average mean absolute error (MAE) of 0.65 dB and is also much faster to train than other. However, it is not detailed that the prediction is done in what type of cellular network. So maybe the prediction results using KNN, will also produce different evaluation values for some other cellular network conditions.

Whereas the research presented in paper discusses the application and comparison of various machine learning techniques to predict received signal strength (RSS) in cellular communications. The training set was generated using experimental measurements from an unmanned aerial vehicle (UAV). This paper creates a prediction model for RSS using five basic learners, including the use of KNN in it. Compared to other algorithms, the RMSE result generated for the KNN algorithm is not good enough, which is about 6.993. There are several factors of parameter features that need to be considered, as well as the KNN concept itself that does not optimally produce predictions for UAV measurements.

C. NAIVE BAYES

Naive Bayes is a probabilistic classification algorithm that can be used for 5G coverage prediction. It works by calculating the probability of a new data point belonging to a certain class based on the probabilities of its features given that class.

In previous research, the use of Naive Bayes algorithm has never been used to predict signal level or coverage in a cellular telecommunication system. In the research presented in the paper , a proposed approach for customer churn prediction (CCP) using the Naïve Bayes classifier as the base model was conducted. It assumes that the features are conditionally independent given the class label, which is a simplifying assumption known as the "naive" assumption. The classifier calculates the probability of each class label given the input features and selects the class label with the highest probability as the predicted class. It uses the training data to estimate the probability of each feature value given each class label, and then combines these probabilities using Bayes' theorem to calculate the posterior probability of each class label given the input feature. The Naïve Bayes classifier is computationally efficient and works well with high-dimensional data, but may make incorrect assumptions about feature independence in some cases.

Until now, the use of this Naive Bayes algorithm is still very limited, especially its use in coverage prediction in cellular telecommunications systems. Therefore, this study will try to use the naive bayes algorithm and evaluate its performance on coverage prediction results, especially in 5G networks.

D. RANDOM FOREST

Random forest is a machine learning algorithm that can be used for 5G coverage prediction. It is an ensemble learning method that combines multiple decision trees to improve the accuracy and robustness of the model. The algorithm works by building multiple decision trees on random subsets of the data and features, and then aggregating their predictions to obtain a final prediction. Random forest can handle both categorical and continuous variables, and can also handle missing values and outliers.

The use of the Random Forest algorithm, especially in coverage prediction in cellular telecommunications systems, has been widely used and recommended, because this algorithm can produce quite good performance evaluation results when compared to other algorithms. The emergence of the use of the Random Forest algorithm for signal level prediction began with a comparative study conducted by who conducted measurements using experimental measurements from an unmanned aerial vehicle (UAV). From the results of comparisons made on various machine learning algorithms, Random Forest shows the best performance results when compared to other algorithms.

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP531-540

In addition, in research specifically conducted to predict coverage on cellular telecommunications systems in 4G networks, presented in papers it is conveyed about the limitations of current network planning techniques that are still conventional in the development of mobile digital connectivity, which hinders the development of sustainable Internet-oriented economies and technologies. In this research, a comparison and evaluation of several machine learning algorithms is carried out in predicting coverage. Of the several algorithms tried, the performance evaluation results show that the Random Forest algorithm is the algorithm that has the best performance evaluation value, which is indicated by the lowest RMSE value and is below 7.

From some of the considerations given in previous research, this study will also predict coverage, especially in areas that have 5G networks. The hope is that of course the use of RF in this study will also produce the best performance evaluation results, even better than previous research.

E. SUPPORT VECTOR MACHINE (SVM) SVM (Support Vector Machine) is one of the ML models tested and evaluated in the research for cellular network coverage prediction based on received signal strength. SVM is a kernel-based model that uses kernel functions to solve regression problems and can convert data sets to different dimensions to find the best hyperplane arrangement.

In the SVM model showed less predictive performance compared to other models, with an RMSE (Root Mean Square Error) of 6.62 dB and an R^2 (coefficient of determination) of 0.66. From these studies, the use of the SVM algorithm has limitations including model inefficiency when dealing with large data sets and noise.

DEEP LEARNING ALGORTIHMS FOR 5G COVERAGE PREDCITION

The application of deep learning algorithms in cellular communication system coverage prediction has become a major trend due to its ability to process complex and non-linear data. Deep learning is a branch of machine learning that uses deep neural networks to understand complex and deep patterns in data. The importance of using deep learning algorithms in coverage prediction lies in their ability to automatically extract relevant features from the large and diverse data generated by mobile communication systems. By involving layers in a neural network, deep learning algorithms can identify complex relationships between the various parameters used and influence the resulting prediction results.

The difference between deep learning algorithms and ordinary machine learning lies in the ability of deep learning algorithms to automatically extract more complex features without requiring manual extraction. Deep learning algorithms can better handle unstructured data, and the layers in neural networks allow for a deeper understanding of patterns. In addition to using a classification model machine learning algorithm, this research also uses a deep learning algorithm model to predict coverage and signal levels, especially for 5G networks. Deep learning algorithms that used include Multi Layer Perception (MLP), Long Short Term Memory (LSTM) and Convolutional Neural Network (CNN).

CONVOLUTIONAL NEURAL NETWORK (CNN)

Convolutional Neural Network (CNN) is commonly used for processing spatial data, such as images, and has also been applied in the context of 5G coverage prediction, especially when spatial data such as network maps are used as input. This is because CNNs have the ability to recognize spatial patterns and local features in data, making them highly effective in image and spatial data processing tasks.

The CNN algorithm is able to automatically identify and extract hierarchical features at various levels of abstraction. This enables the model to understand complex data structures such as images or spatial patterns in network data. CNNs have invariance to shifts and transformations in the data. This allows CNNs to remain effective in recognizing patterns even if the position or orientation of the pattern changes. In addition, CNNs can also model the spatial context of the data, which is useful in understanding the spatial relationships between elements in telecommunications data, such as the locations between base stations or the spatial distribution of users.

In addition to this, CNNs can be used to analyze satellite images or network maps to detect and understand important elements in telecommunications networks, such as cell towers, network topology, or user density. The advantages and benefits offered in the use of the CNN algorithm make CNN one of the deep learning algorithms that is often used in the implementation of various predictions, especially in the field of telecommunications. Even some studies related to coverage prediction in cellular communication systems have also used this CNN algorithm.

In the research presented in research where this research presents the latest advances in the rapid prediction of signal power in mmWave communication environments using machine learning (ML). The use of the CNN algorithm in this

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP531-540

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

study is used as an algorithm to train the model to provide power estimates with good accuracy and realtime simulation speed. Improved training data pre-processing techniques. This study successfully extends the prediction to 3D, allowing for arbitrary transmitter heights. However, the rationale for using CNN algorithm in this study is not clearly and significantly explained. So it does not appear the advantages of using CNN in this study.

From some of the research results that have been conducted and presented in other studi and paper, it shows that the use of CNN algorithms can be one of the best considerations in predicting coverage, especially in 5G technology systems. So, in this research, it will also be evaluated regarding the use of the CNN algorithm in predicting 5G signal coverage levels.

PROJECT REQUIREMENT

HARDWARE REQUIREMENT

- System : Intel(R) Core(TM) i3-7020U CPU @ 2.30GHz
- Hard Disk : 1 TB.
- Input Devices : Keyboard, Mouse
- Ram : 4 GB.

SOFTWARE REQUIREMENTS**:**

- Operating system : Windows XP/7/10.
- Coding Language : Python
- Tool : Anaconda
- Interface : Django

DATA COLLECTION AND DATASET PREPARATION

To begin the study, a highly curated dataset including essential information about 5G network coverage was picked. This dataset includes geographical features, environmental variables, and signal

factors as input data was based on previous work using electromagnetic wave propagation understanding.

TABLE 1: Parameter of Measurement Campaign Setup

Hardware	
Smartphone	Xiaomi Poco F3 (M2012K11AG)
User Equipment (UE)	Category 20
3GPP Release Standard	Release 18 (5G NR)
Software	
G-NetTrack Pro	A non-rooted wireless network monitor and drive test tools

FIGURE 1: Distribution of drive test data around Batununggal In this study, to train and validate the ML-based SS-RSRP prediction model, we used Colaboratory by Google. Colaboratory by Google (Google Colab in short) is a Jupyter notebook based runtime environment which allows you to run code entirely on the cloud . It can also be used to test basic machine learning models, gain experience, and develop an intuition about deep learning aspects such as hyperparameter tuning, preprocessing data, model complexity, overfitting and more.

To assess each algorithm's predictive performance, the dataset was divided into training and testing subsets, with suitable cross-validation procedures used to reduce overfitting. In this study, all learners were examined using 10-fold cross-validation (CV). The use of 10-fold cross-validation (CV) in predicting coverage using machine learning algorithms, the available data will be divided into 10 equal parts or folds. The algorithm will then be trained on 9 of these folds and validated on the remaining fold. This process will be repeated 10 times, each time using a different fold as the validation set. Finally, the results from each validation step will be averaged to produce a more robust estimate of the model's performance. The use of 10-fold cross-validation can provide a more accurate estimate of the model's performance compared to other cross-validation techniques, especially when the dataset is large enough to support it. It is essential to determine the statistical error between the measurement and RSRP prediction values when examining and validating the performance of any ML model.

For this study, the given data will be divided into 10 equal sections or folds while employing 10-fold cross-validation (CV) in predicting coverage using machine learning algorithms with a total sample size of roughly 1500 signal level points. Each fold will

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP531-540

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

TABLE 2: List Feature and Explanations

Feature	Definition	Purpose
2D Distance Tx to Rx	The separation distance between Base Station (BS)	To estimate the position of UE away from the BS antenna on the x-axis
Frequency	The 5G spectrum sub-6 GHz bands in Indonesia. In this research we use 2300 MHz frequency	To estimate the level of attenuation experienced by the transmitted signal from the BS to the UE according to different spectrum band being utilized.
Tx Tilt Angle	The angle of transmitting the signal at the horizon axis from the g-NodeB antenna	To determine the tilt angle of signal travel from the gNB antenna to the location of UE
Tx Azimuth Angle	It specifies the horizontal angular position of the antenna's main lobe, indicating the direction in which the antenna is transmitting signals	To determine the specific direction in which a transmitting antenna is focused
Altitude	elevation of an object or location above a specific reference point or datum. It is represents how high or low something is with respect to sea level	By considering the elevation of the receiver and transmitter locations, network providers can identify the resulting coverage conditions
Elevation Angle	The angle formed between the horizontal plane and the observed line from UE to the BS antenna	To estimate the UE position with reference to position with reference to the position of BS antenna on z-axis
Azimuth Offset An- gle	The absolute value of the formed between the BS antenna pointing direction and the location of UE on the horizontal plane	To estimate the UE position with reference to the position of the BS antenna on y-axis. Also to estimate the UE position in the main lobe of the BS antenna on the horizontal plane
Tilting Offset Angle	The absolute value of the angle formed between the BS antenna boresight and the observed line from UE on the BS antenna	To estimate the UE position in the beam of the main lobe of BS antenna on the vertical plane
Horizontal Distance of Rx From Boresight of Tx Antenna	Lateral distance between a receiver's position (Rx) and the main axis (principal viewing line) of the transmitter's antenna on a base station	To measured vertically against the direct viewing direction of the transmitter antenna. It describes how the receiver is located from the point where the transmitter antenna is set to direct its signal at the highest intensity
Vertical Distance of Rx From Boresight of Tx Antenna	the distance along the vertical axis between the receptor's position and the maximum direction in which the transmitter's antenna is positioned to transmit a signal with the highest efficiency	To ensure that the receiver is in an optimal position to receive a strong and quality signal from the base station transmitter antenna

include about 150 samples. The algorithm will then be trained on 9 of these folds before being validated on the final fold. This method will be repeated ten times, each time using a different fold as the validation set.

Finally, the validation results will be averaged to produce a more trustworthy approximation of the model's performance. Because it provides a fair balance between the bias and variance of the mode, 10-fold cross-validation is an excellent choice for a dataset with a sample size of roughly 1500 signal level points. It also ensures that the model is trained on a sufficient amount of data and is not overfitting or underfitting the data. The choice of the number of folds to use in cross-validation depends on the size of the dataset and the computational resources available.

The models were painstakingly trained on training data before being tested using a range of performance criteria such as accuracy, precision, recall, or F1-score. These measurements gave a thorough evaluation of each algorithm's prediction ability. And also, we need to evaluate the performance of the trained model using Root-mean-square error (RMSE), Mean Absolute Error (MAE), and

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP531-540

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

coefficient of determination (R^2). It is important to assess the statistical error between the measured and the predicted SS-RSRP values. RMSE, is shown in (1), is a commonly used metric to evaluate the performance of the regression prediction models. It is given, in decibels .

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

where:
n = total number of samples
y_i = actual value i $\hat{y}_i$ = predictive value i

The smaller values of RMSE indicate a better prediction of the ML model. According to predictive models with RMSE values less than 7 dB is considered acceptable, especially in an urban environment.

MAE measures the average absolute error between the true and predicted values of a model or algorithm. MAE is also used to compare the performance of different models or algorithms in making predictions. The smaller the MAE value, the better the model or algorithm is at making predictions. MAE can be expressed as in the equation (2)

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (2)$$

where n = total number of samples
y_i = actual value i $\hat{y}_i$ = predictive value i

On the other hand, as indicated in (3), we used the coefficient of determination (R^2) to determine the degree of performance of the prediction models. It is used to describe how effectively the model's input parameters explain the variability of the response variable. The model explains greater variability with higher R^2 values. It is given by

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (3)$$

where:
n = total number of samples
y_i = actual value i $\hat{y}_i$ = predictive value i
$\bar{y}$ = average of actual values

In addition to the measurements and analysis conducted to evaluate the data trained using the machine learning algorithm, an evaluation was also conducted on the

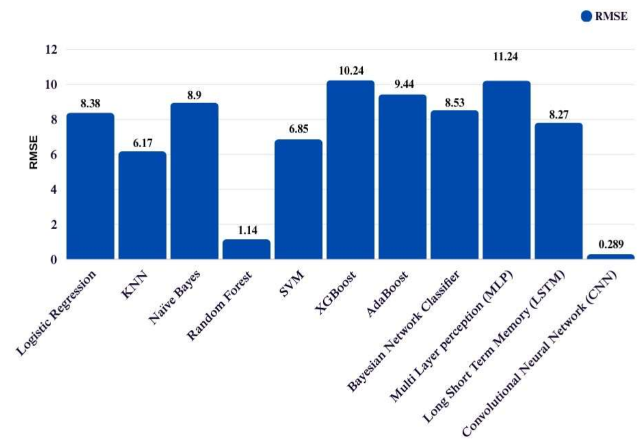

FIGURE 2: RMSE result value for each model algorithm.

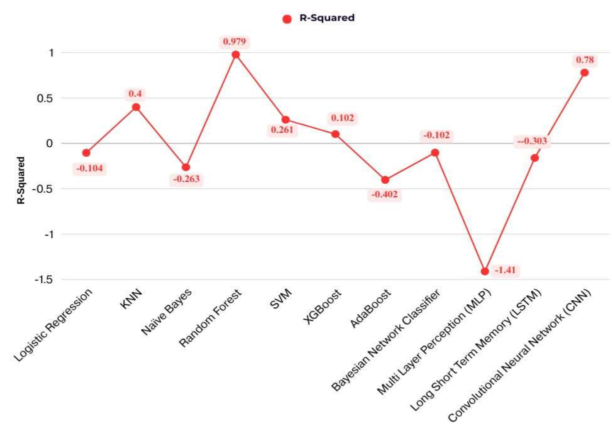

FIGURE 3: R^2 result value for each model algorithm

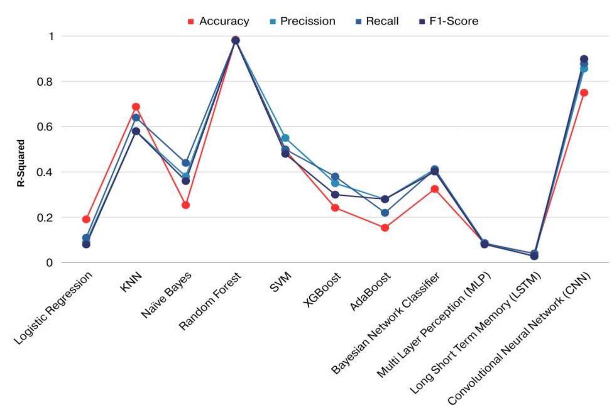

FIGURE 4: Performance Evaluation Metrics of Machine
Learning Algorithm on Training Data

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP531-540

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

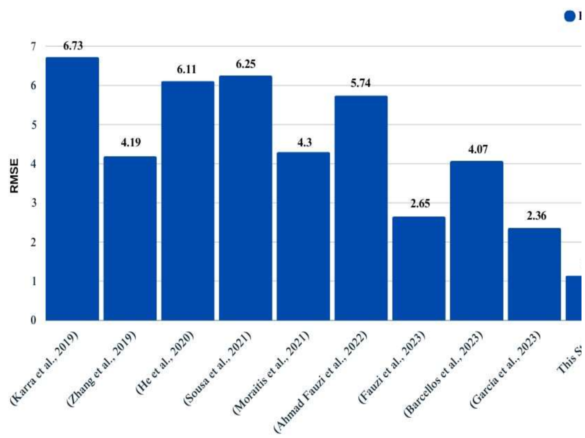

FIGURE 5: Comparison of RF model performance with the previous works

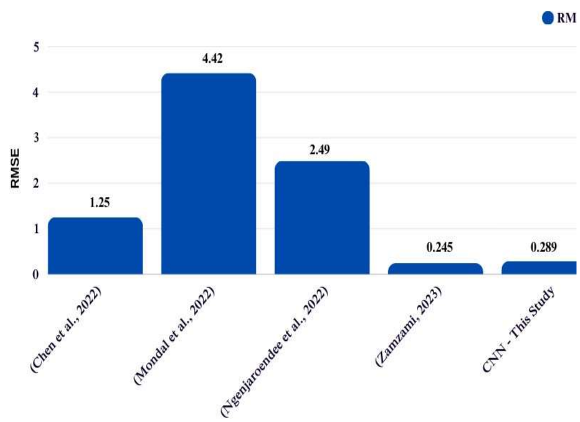

FIGURE 6: Comparison of CNN model performance with the previous works

TABLE 3: Performance Evaluation Metrics of Machine Learning and Deep Learning Algorithm on Training Data

Algorithm	RMSE	MAE	R2
Logistic Regression	8.38	6.3	-0.104
KNN	6.176	2.86	0.4
Naïve Bayes	8.964	6.63	-0.263
Decision Tree	0.66	0.05	0.993
Random Forest	1.14	0.12	0.979
SVM	6.854	4.03	0.261
XGBoost	10.244	8.34	0.102
AdaBoost	9.444	7.12	-0.402
Bayesian Network Classifier	8.534	6.12	-0.102
Multi Layer Perception(MLP)	10.223	9.44	-0.99
Long Short Term Memory(LSTM)	7.82	5.89	-0.16
Convolutional Neural Network (CNN)	0.289	0.289	0.78

TABLE 4: Performance Evaluation of Examined Models

Algorithm	Accuracy	Precission	Recall	F1-Score
Logistic Regression	0.1909	0.09	0.11	0.08
KNN	0.688	0.58	0.64	0.58
Naïve Bayes	0.254	0.38	0.44	0.36
Decision Tree	0.992	0.99	0.99	0.99
Random Forest	0.984	0.98	0.98	0.98
SVM	0.496	0.55	0.5	0.48
XGBoost	0.242	0.35	0.38	0.3
AdaBoost	0.154	0.28	0.22	0.28
Bayesian Network Classifier	0.325	0.412	0.412	0.402
Multi Layer Perception (MLP)	0.0850	0.0872	0.0850	0.0804
Long Short Term Memory (LSTM)	0.0400	0.0272	0.0400	0.0289
Convolutional Neural Network (CNN)	0.75	0.856	0.878	0.899

VI. CONCLUSION

In this study, we have conducted a comprehensive comparative analysis of various machine learning algorithms to predict 5G coverage. From usage of machine learning and deep learning algorithm in this study, the results show that the Random Forest and CNN algorithm models have the best results and performance when compared to other models used in this study. Random Forest algorithm model which shown the best result with an RMSE of 1.14 dB, MAE of 0.12, and R^2 value of 0.97. And from deep learning algorithm, CNN model shown the best result with 0.289 for RMSE value, 0.289 for MAE value, and 0.78 for R^2 value. This indicates their ability to predict 5G coverage with very high accuracy.

The Random Forest models perform almost perfectly, with accuracy of 98.4%, precision of 98%, recall of 98%, and F1score of 98%.This research was designed not only to produce the greatest RMSE performance evaluation results, but also to prepare for coverage in 5G networks in Indonesia, which is still fairly limited in research.

In conclusion, the use of Random Forest algorithms may be the best option for 5G coverage prediction with optimal accuracy. However, the utilization of deep learning algorithms also needs to

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP531-540

ISSN 2347–3657

Volume 13, Issue 2s, 2025

be considered. Because deep learning algorithms, especially CNN, can also be used for coverage prediction, especially on 5G networks. Looking at the evaluation results and performance of the algorithms shown from this study also shows that CNN has a fairly good performance when compared to other algorithms. In addition, the use of the CNN algorithm is highly recommended for coverage and signal level prediction if the parameter features used are spatial data derived from satellite images. These accurate predictions have a great impact in 5G network planning, allowing network providers to allocate resources more efficiently and provide high-quality services to customers. In the era of ever-evolving 5G technology, a deep understanding of coverage prediction is key to successful network planning and reliable service provision to end users.

REFERENCES:

[1] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[2] Y. Du, L. Ren, X. Liu, and Z. Wu, “Machine learning method intervention: Determine proper screening tests for vestibular disorders,” Auris Nasus Larynx, vol. 49, no. 4, pp. 564–570, 2022, doi: 10.1016/j.anl.2021.10.003.

[3] C. Zhang, H. Zhang, D. Yuan, and M. Zhang, “Citywide Cellular Traffic Prediction Based on Densely Connected Convolutional Neural Networks,” IEEE Commun. Lett., vol. 22, no. 8, pp. 1656–1659, 2018, doi: 10.1109/LCOMM.2018.2841832.

[4] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, “Channel State Information Prediction for 5G Wireless Communications: A Deep Learning Approach,” IEEE Trans. Netw. Sci. Eng., vol. 7, no. 1, pp. 227–236, 2020, doi: 10.1109/TNSE.2018.2848960.

[5] S. Joseph, R. Misra, and S. Katti, “Towards self-driving radios: Physicallayer control using deep reinforcement learning,” HotMobile 2019 - Proc. 20th Int. Work. Mob. Comput. Syst. Appl., pp. 69–74, 2019, doi: 10.1145/3301293.3302374.

[6] Nausheen Fathima, Dr. Mohd Abdul Bari , Dr. Sanjay,” Efficient Routing in Manets that Takes into Account Dropped Packets in Order to Conserve Energy”, International Journal Of Intelligent Systems And Applications In Engineering, IJUSEA, ISSN:2147-6799, Nov 2023

[7] Afsha Nishat, Dr. Mohd Abdul Bari, Dr. Guddi Singh,” Mobile Ad Hoc Network Reactive Routing Protocol to Mitigate Misbehavior Node”, International Journal Of Intelligent Systems And Applications In Engineering, IJUSEA, ISSN:2147-6799, Nov 2023

[8]) Ijteba Sultana, Dr. Mohd Abdul Bari ,Dr. Sanjay,” Routing Performance Analysis of Infrastructure-less Wireless Networks with Intermediate Bottleneck Nodes”, International Journal of Intelligent Systems and Applications in Engineering, ISSN no: 2147-6799 IJISAE,Vol 12 issue 3, 2024, Nov 2023

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP531-540