IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 January 2024, accepted 28 January 2024, date of publication 1 February 2024, date of current version 8 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3361403

RESEARCH ARTICLE

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

HAJIAR YULIANA, ISKANDAR, AND HENDRAWAN
School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung 40132, Indonesia
Corresponding author: Hajiar Yuliana (33222316@mahasiswa.itb.ac.id)

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.

INDEX TERMS 5G coverage prediction, classifier algorithm, deep learning, performance evaluation, feature importance.

I. INTRODUCTION

The rapid development and deployment of 5G networks have brought about new opportunities and challenges in the telecommunications industry. One of the key challenges is accurately predicting the coverage of 5G networks, which is crucial for efficient network planning and optimization. In recent years, machine learning algorithms have emerged as a promising approach to address this challenge. 5G networks are expected to revolutionize the telecommunications industry by providing enhanced coverage, ultra-reliable low latency, high data rates, massive connectivity, and better

The associate editor coordinating the review of this manuscript and approving it for publication was Adnan Kavak.

support to mobility [1]. However, the successful deployment and optimization of 5G networks require accurate prediction of network coverage. The ability to predict coverage is crucial for efficient network planning and optimization, as it helps network operators to identify coverage gaps, optimize network resources, and improve the quality of service for end-users [2].

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 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 unlabeled data.

Many telecommunication domains have already achieved significant progress in ML-based research on topics such as propagation loss prediction, channel decoding, signal detection, and channel estimation [3], [4], [5]. Machine learning algorithms have emerged as a promising approach to address the challenge of 5G coverage prediction. These algorithms can analyze large amounts of data and identify complex patterns and dependencies that influence network coverage. By incorporating machine learning algorithms into the network planning and optimization process, network operators can improve the accuracy of coverage predictions and optimize network performance [6]. Previous research has explored the use of machine learning algorithms for 5G coverage prediction, including deep learning algorithms, decision tree algorithms, and support vector machines [7]. And also, in research [8], [9] also discusses coverage prediction in 4G technology using machine learning algorithms. In paper [8], a supervised machine learning algorithm is presented using several parameter features. Then the research is continued and presented in paper [9], where in this paper only the random forest model machine learning algorithm is used, which is claimed to have a higher level of accuracy performance. In paper [6], several parameter features are used in this research for 4G coverage prediction process.

Apart from the studies mentioned in these papers, the actual use of machine learning algorithms for coverage prediction has been widely implemented and used. From these various studies, with various uses of machine learning algorithms of both regression and classification types, even to the development to the ensemble learning stage, it shows that the use of machine learning algorithms for coverage prediction is highly considered. Because from various studies, it shows that the coverage prediction results produced using machine learning algorithms produce more accurate and more efficient prediction results, when compared to the prediction results using conventional methods. So that until now, research continues to be carried out and developed related to coverage prediction using various machine learning algorithms with various case studies, parameter feature variations and other variations, in order to obtain algorithms with the best prediction results and performance evaluation metrics.

In line with the use of machine learning algorithms, the use of deep learning algorithms has also continued to be developed and has not gone unnoticed in optimization and planning activities in telecommunications networks. Deep learning algorithms, known for their ability to extract deep and complex feature representations from data, are becoming a critical element in improving the efficiency, performance, and resilience of telecommunications networks. In the context of prediction, where signal coverage estimation, capacity planning, and traffic management are key parameters, deep learning offers solutions that are also quite adaptive. The importance of deep learning in telecommunications lies in its ability to automatically identify patterns, understand hierarchical relationships, and respond to environmental changes quickly and efficiently. Deep Reinforcement Learning (DR), Long Short Term Memory (LSTM), and Convolutional Neural Network (CNN) are the three most often used deep learning models [10]. They are getting more popular as a result of problems in all of the machine learning models that have been built up to this time [11]. These machine learning models are distinguished by their weakness as additional data is provided, which causes their accuracy to decline because they are prone to overfitting [12]. With architectures such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM), deep learning is able to process spatial and temporal data very well, opening the door for more accurate and adaptive modeling with a diversity of parameterized feature models that can be used later in the coverage prediction process.

Some data-driven technologies using deep learning or machine learning can help manage 5G infrastructure. For example, dynamic mobile traffic analysis can be used to predict user position, which enhances handover mechanisms [13]. Another example is the use of previous physical channel data to anticipate channel state information, which is a difficult analytical problem to solve [14]. Another example is the allocation of network slices based on user requirements, taking into account network state and available resources [15]. All of these examples are data-driven especially using deep learning algorithm. Some instances are based on historical data analysis and are used to forecast certain behaviors, while others are based on the current state of the environment and are used to aid in decision making.

While the use of deep learning algorithms for coverage prediction has also previously been used, although its use is still very limited and also has limitations on the results of evaluating the performance of the prediction results. The research presented in [16], shows the use of deep learning algorithms for several jobs in the telecommunications field, especially related to the adaptation of the use of 5G technology. In the study, of the three deep learning algorithms used in this study, namely DR, LSTM, and CNN, showed that the CNN algorithm showed much more optimal prediction results when compared to the other three algorithms. In addition, the research presented in papers [17], [18] recommends the use of the CNN algorithm for prediction. Whereas in the study presented in [19], where this study focuses on predicting the received reference signal power (RSRP) in cellular networks using deep learning techniques. According to the researchers, the current approach of using a path loss model for signal prediction was found to be inaccurate. The study, however, does not explicitly mention any specific algorithm as superior or recommended. However, it highlights the effectiveness of deep learning for RSRP prediction, suggesting that deep learning algorithms are suitable for this task. From various studies related to coverage prediction using deep learning algorithms, it turns out that the use of this algorithm is still very rare and there are still few who use it for coverage prediction, especially in 5G networks. This is also the supporting background for us to conduct research and also evaluate the performance related to coverage prediction on 5G networks using deep learning algorithms.

Features selected or generated from data play a crucial role in the predictive ability of machine learning models, including in the context of coverage prediction in telecommunications. The importance of features can be measured through the concept of ''feature importance''. Features that are relevant to the prediction target (coverage) will have a great influence on the prediction. If a feature has a strong relationship with the signal coverage, the model will tend to rely on that feature in making predictions [20]. Furthermore, important features can increase the reliability of the model. Models that rely heavily on less relevant or unimportant features may tend to be less reliable and may not generalize well to new data. There are several parameter features used in coverage prediction, which were previously taken and adapted from parameters used in conventional models. Such as parameters related to frequency, transmitting and receiving antenna height, distance between transmitter and receiver, and parameters related to transmitting antenna factors used in conventional prediction models are still very limited. By using machine learning algorithms, one of the advantages offered is being able to input and train data from various parameters that can affect the prediction results [21]. This is the case with coverage prediction, where the influential parameters are not only a matter of frequency, distance between transmitter and receiver, or the height of the transmitting and receiving antennas, but there are many parameters that can be considered.

Recent studies have begun to use a much larger and more varied set of parameterized features. The study conducted in [22], initiated the use of a large number and variety of parameter features. If in the previous study [23], [24], [25], the parameter features used were only limited to the distance of Tx and Rx, frequency, and height of Tx and Rx. Then in [22], additional parameter features are also used in addition to the commonly used ones, namely Tx tilt angle, Tx azimuth angle, transmit power, clutter, building height, vertical distance of Rx from the major lobe signal. Where data obtained from variations and additional parameter features used can be used as training data in training the machine learning algorithm used. Besides, in the study conducted in [8], several other parameter features that can be considered in coverage prediction are height ratio of Tx and Rx, tilt angle of transmitted signal to Rx, Rx elevation angle. Whereas in [9], new additional parameter features are used in the form of azimuth offset angle, tilting offset angle, elevation angle, and the status of the signal from base station antenna to receiver (UE). Looking at the various parameter features used in previous studies related to coverage prediction, in this current study use several parameter features that have been used in previous studies and also add other parameter features that have not been tried in previous studies. Some parameter features that have the possibility and ability to predict 5G coverage, such as parameter features about the horizontal and vertical distance of the receiver from the boresight of the transmitting antenna. It is hoped that these parameter features can become alternative parameter features used to predict coverage, especially in 5G networks and produce more accurate accuracy performance.

In addition to paying attention to the use of parameter features used in 5G coverage prediction, it turns out that the use of algorithms also needs to be considered about which algorithms can produce the best prediction accuracy performance. Machine learning algorithms that used and analyzed in this study is machine learning classification algorithm models, there are Logistic Regression, K-Nearest Neighbors (KNN), Naive Bayes, Random Forest, Support Vector Machine (SVM), XGBoost, LightGBM, AdaBoost, and Bayesian Network Classifier. Also, in addition to using machine learning algorithms, this study also using deep learning algorithms to predict coverage in 5G networks. The deep learning algorithms used include Multi Layer Perception (MLP), Long Short Term Memory (LSTM), and Convolutional Neural Network (CNN). The various algorithms used in this study will then be evaluated regarding the performance of these algorithms on the resulting prediction results. There are several parameters that become evaluation metrics in this study. Among them are evaluation metrics in the form of Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-Squared. In addition to these evaluation metrics, performance evaluation metrics of the accuracy of the prediction results generated from each algorithm are also used, namely accuracy, precision, recall, and f1-score. The accuracy performance parameters measured in this study related to the 5G coverage prediction process.

In conclusion, accurate prediction of 5G coverage is crucial for efficient network planning and optimization. Machine learning algorithms have the potential to improve the accuracy of coverage predictions and enhance the efficiency of 5G networks. These studies have highlighted the importance of feature parameters in predicting 5G coverage. By identifying 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 comparative analysis of machine learning algorithms and identifying the dominant feature parameters, this study aims to contribute to the advancement of 5G coverage prediction. The findings of this research can provide valuable insights for network operators and researchers in optimizing the deployment and performance of 5G networks.

In the following sections, we will discuss the related work on machine learning for 5G coverage prediction. We will then present the methods used in this study. Finally, we will analyze and compare the prediction accuracy of different machine learning algorithms and discuss the implications of the findings. Overall, this research aims to contribute to the understanding of machine learning algorithms for 5G coverage prediction, providing valuable insights for network operators and researchers in optimizing the deployment and performance of 5G networks.

The main contributions of this paper can be summarized as follows:

1) Insight into the trained machine learning classification and deep learning model behaviour especially to validate the impact of utilizing new types of features, i.e., horizontal and vertical distance of Receiver (Rx) from boresight of Transmitter (Tx) Antenna for 5G coverage prediction.
2) Evaluate the performance of machine learning classification algorithm: Logistic Regression, K-Nearest Neighbors (KNN), Naive Bayes, Random Forest, Support Vector Machine (SVM), XGBoost, LightGBM, AdaBoost, and Bayesian Network Classifier in training prediction data and evaluating the accuracy of the prediction results using machine learning algorithms on 5G networks. The best performance of the final trained model is evaluated against best model in previous works.
3) Also, in this paper will evaluate the performance of deep learning algorithm: Multi Layer Perception, Long Short Term Memory, and CNN in training prediction data for 5G coverage.
4) Identifying 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.

II. 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 [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 [27]. 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.

Logistic regression is a statistical method that is widely used in various disciplines, including telecommunications, and has proven its value in predicting and analyzing data. In telecommunications, the use of this algorithm is mostly used to predict and classify telecommunication customers based on their characteristics and behaviour. Logistic regression models provide accurate predictions, which can help decision makers gain a better understanding of individual interactions and group-level customer behaviour [28]. 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 [29].

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 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 algorithm employs a group of named points to mark another point. The data is sorted based on similarities, and K-NN can be used to fill in missing data values. After filling in the missing values, the data set is subjected to several prediction algorithms. Using varied combinations can help you improve your accuracy. K-NN is simple to implement. This algorithm is used for classification, regression, and search. 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 [30]. KNN can also be used in conjunction with other machine learning algorithms, such as deep learning, to boost performance [31].

The use of the KNN algorithm has previously been used in coverage prediction. In the research presented in [32], 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 [33], 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.

Based on previous existing research, which has not actually been carried out experiments and predictions of signal levels and coverage specifically specifically on 5G technology using the KNN algorithm, so that in this research a study will be carried out related to this matter. In the context of 5G coverage prediction, KNN can be used to forecast whether or not a specific location will have 5G coverage based on a variety of independencharacteristics such as population density, terrain, and proximity to current infrastructure.

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 [34], 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 [33], 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.

In addition, in research specifically conducted to predict coverage on cellular telecommunications systems in 4G networks, presented in papers [8], [9] 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.

In addition, in research [35], the ensemble learning model was also developed which became the Random Forest algorithm as the basis of the model used. The Extremely Randomized Trees Regressor (ERTR) algorithm model is used to predict coverage on 5G networks in dense urban environments around Victoria Island and Ikoyi in Lagos, Nigeria. This research also compares various algorithms, namely Random Forest itself and several other algorithms. From the performance evaluation, the resulting value for ERTR and RF is the best value when compared to other algorithms and also compared to previous research, which reaches around 2.23 dB.

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 [8], 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.

F. XGBOOST

XGBoost is a well-known machine learning method that can be used to predict 5G coverage. It is an ensemble learning method that integrates numerous decision trees in order to improve the model's accuracy and robustness. XGBoost is well-known for its scalability, speed, and capacity to handle large amounts of data.

The use of the XGBoost algorithm also turns out that in detail it has never been used for coverage prediction in cellular telecommunications systems. Most of the use of the XGBoost algorithm to date has been used to develop customer churn prediction models. In the paper [36], [37], these two studies discuss the challenges of unbalanced data sets in the telecommunications industry and the variations in real telecommunications data compared to publicly available data sets. By utilizing the application of XGBoost Algorithm on this dataset, it achieves 97% of accuracy evaluation performance result and 88% of F1 score. However, this may be different if applied to other predictions.

Therefore, since there are still limitations to the use of the XGBoost algorithm, especially in coverage prediction in mobile telecommunications systems, this study will consider and evaluate the performance of using the XGBoost algorithm for coverage prediction in 5G networks.

G. AdaBoost

AdaBoost is an ensemble learning method that combines numerous weak classifiers to generate a strong classifier. AdaBoost can forecast whether or not a specific place will get 5G coverage based on a variety of independent criteria such as population density, terrain, and proximity to current infrastructure. The approach works by training weak classifiers on the data iteratively and applying larger weights to misclassified data points. The final prediction is obtained by integrating the predictions of all weak classifiers and weighting them according to their accuracy. AdaBoost has been employed in a variety of wireless network applications, including traffic network prediction [38] and network performance forecasting [39]. However, there are still limited studies that specifically focus on using AdaBoost for 5G coverage prediction.

Similar to the XGBoost algorithm, the use of the AdaBoost algorithm also turns out to be rarely used in coverage prediction, especially in cellular communication systems. AdaBoost has been used as a comparison on coverage prediction in cellular communication systems in the paper [33]. In this study, this model is only used in experiments to measure predictions on unmanned aerial vehicles (UAVs). AdaBoost has evaluation performance results that are not very good when compared to other algorithms for coverage prediction, namely RMSE of 7.112 dB. The evaluation results show that, AdaBoost is not really optimal if used in coverage prediction in the study.

Even so, the use of the AdaBoost algorithm still needs to be considered in predicting various other things in the telecommunications system. This study will also use and evaluate the performance results of using the AdaBoost algorithm in the coverage prediction process, especially in 5G communication systems.

H. BAYESIAN NETWORK CLASSIFIER

Bayesian networks offer several benefits compared to classical methods such as Markov Chains, Fault Trees, and Petri Nets, including the ability to model complex systems, make predictions and diagnostics, calculate event probabilities, update calculations based on evidence, represent multi-modal variables, and provide a user-friendly graphical approach [40]. However, there are not many studies that use the Bayesian Network algorithm in prediction in 5G mobile network systems. Mostly, the use of the Bayesian Network algorithm is used to predict throughput and users in cellular network systems. And until now, there has been no research related to coverage and signal level predictions.

The research presented in paper [41], describes the use of Bayesian Network (BN) for throughput reliability prediction in 5G wireless networks. The BN algorithm is used in this study to predict future test results by estimating parameters such as base station load, user location, and moving speed, which affect the signal-to-noise ratio (SNR) received by users and the signal interference plus noise ratio (SINR). Computer simulation results show that the BN model can effectively infer user throughput under low-speed movement conditions.

Then in paper [40], discusses modeling the prediction of 5G wireless network service reliability using Bayesian networks. This model is used to predict network service reliability and infer the hidden status of the network. The use of Bayesian networks allows a compact representation of the joint probability distribution, making it easier to model the reliability of network services. This research offers a promising direction for designing next-generation networks that meet high quality of service requirements.

Most of the limited research that has been done using Bayesian Network, no one has discussed in detail the conditions and performance evaluation of this algorithm on the predictions made. Especially the use of this algorithm for coverage prediction in 5G cellular communication systems, no one has discussed and used it. So that in the current study, the Bayesian Network algorithm is also used in coverage prediction, especially in 5G networks.

III. DEEP LEARNING ALGORITHMS FOR 5G COVERAGE PREDICTION

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).

A. MULTI LAYER PERCEPTION (MLP)

Multi-Layer Perceptron (MLP) is a type of neural network architecture that consists of several interconnected layers. Each neuron in one layer is connected to a neuron in the next layer. MLP is used for tasks such as classification, regression, and pattern recognition in complex and non-linear data.

Multilayer Perceptron (MLP) is a type of feedforward artificial neural network that is suitable for modeling non-linear data and has been widely used in various prediction applications, including 5G coverage prediction. The study presented in [42] focuses on predicting path loss using multi-layer perceptron (MLP) neural networks for network planning and optimization in 5th generation and future communication systems. A path loss model based on MLP neural networks is created by combining measurement data and environmental features extracted using environmental characterization methods. Comparative analysis of data experiments shows that MLP neural networks can accurately predict path loss, and the inclusion of environmental features improves the performance of the model. To address the problem of interference clutter that reduces the accuracy of MLP-based path loss models and their sensitivity to environmental changes, the authors improved the environmental characterization method based on line-of-sight (LOs) and non-line-of-sight (NLOs) labels. This improves the stability and generalization ability of the MLP-based path loss model. This research also compares the use of MLP algorithm and conventional pathloss model algorithms, namely CI model and ABC model. The research shows that multi-layer perceptron (MLP) neural networks can accurately predict path loss in communication systems.

However, it turns out that the use of MLP algorithms is still very limited in various studies. Therefore, the MLP algorithm will be used in the coverage prediction process in the 5G network studied in this research. Through this research, the performance of the algorithm can also be evaluated against the resulting prediction results.

B. LONG SHORT TERM MEMORY (LSTM)

Long Short-Term Memory (LSTM) is a type of architecture in the field of deep learning specifically designed to overcome vanishing gradient and exploding gradient problems in ordinary neural networks. LSTMs are effective in understanding and modeling sequential relationships in data, such as time series, and have been widely used in various applications, including time prediction and natural language processing.

One of the main advantages of LSTMs is their ability to overcome the vanishing gradient problem commonly encountered in conventional neural networks. This allows LSTMs to capture information over long periods of time, which is particularly important in applications involving data sequentiality, such as in telecommunications where real-time data is critical. LSTMs have two main components, the state cell and the forget gate, which allow the model to store relevant information and ignore unnecessary information. This gives the model the ability to better manage and store information in data sequentiality. LSTMs can be used to forecast trends and patterns in real-time data, such as network traffic, service requests, or network utilization rates. By understanding the sequentiality of the data, LSTM can provide more accurate predictions compared to non-consequential models.

Research related to the use of the LSTM algorithm in the telecommunications field, one of which is presented in the research paper [43]. This study focuses on predicting throughput in LTE networks using the attention-based LSTM model. The researchers collected TCP and throughput logs in LTE networks and transformed them using CUBIC and BBR trace log data. They used the sliding window method to create input data for the prediction model. The LSTM model with attention mechanism was trained using the collected data. This study compares the proposed attention-based LSTM model with other methods for throughput prediction in LTE networks. From the presented research results, that the proposed LSTM-based model with attention mechanism achieves better throughput prediction performance. The results showed that the use of LSTM algorithm resulted in lower RMSE compared to other methods. This also shows its effectiveness in predicting future throughput in LTE networks.

In addition, in research in [16], LSTM algorithm is one of the deep learning models used in this study to predict the adoption of 5G technology. This study compares the performance of different deep learning models, including LSTM, DR, and CNN. The study also mentions that the LSTM model has been used in various 5G technology applications, such as anomaly detection in network flows and security threat detection in 5G core wireless networks. From the results, it shows that the LSTM model produces accurate prediction results, but the DR and CNN models are found to be more effective in predicting factors that will affect 5G adoption. Overall, the LSTM algorithm played a role in predicting the adoption of 5G technology in this study, but was found to have poorer data quality compared to other deep learning models such as DR and CNN. The specific reason for this difference in data quality was not explicitly mentioned in the source provided. However, it is important to note that the LSTM algorithm is known for its ability to handle long time lags and complex tasks. It is possible that the nature of the data used in this study, which includes channel metrics, context metrics, cell metrics, and throughput data, may not be suitable for the LSTM algorithm.

Based on the advantages and limitations of the LSTM algorithm model, this current study will evaluate the performance of using the LSTM algorithm and the prediction results of the 5G coverage prediction conducted in this study. Hopefully, this research can provide an overview of the performance of the LSTM algorithm in the coverage prediction process, especially in 5G networks.

C. 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 [17], 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 study is used as an algorithm to train the model to provide power estimates with good accuracy and real-time 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.

Whereas in research [18], this study proposes a Convolutional Neural Network-based Auto Encoder (CNN-AE) to predict the level of location dependence and coverage probability in cellular networks. It compares the performance of CNN-AE with stochastic geometry-based analytical models and shows significant improvement in coverage and rate prediction errors. In addition, this paper proposes a low complexity algorithm that uses trained CNN-AE to calculate locations for new base stations to meet specific performance objectives. The use of CNN-AE enables more accurate prediction of location-dependent rate and coverage probabilities in cellular networks. The improved performance of CNN-AE over SG-based models suggests that deep learning-based approaches can provide better network performance estimation in real-world scenarios. The CNN-AE model outperforms the stochastic geometry (SG)-based analytical model in terms of coverage and rate prediction errors. The coverage prediction error is improved by 40% and the rate prediction error is improved by 25% compared to the best-fit SG-based model.

In addition, the use of the CNN algorithm in research [16] also showed significant results. From this paper, CNN (Convolutional Neural Network) algorithm is one of the deep learning models used in this study to predict the adoption of 5G technology. The CNN model was found to be the most effective in predicting the factors that will influence 5G adoption, outperforming the LSTM and DR models in this study. The CNN model in this study uses layers such as input layer, convolution layer, pooling layer, and output layer to process input images and generate feature maps. The study concludes that the DR model and CNN model produce aesthetically pleasing results and can be used with existing data sets. The CNN model was highlighted as providing the most accurate 5G adoption rate estimates due to its ability to recognize and extract patterns from convolutions.

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.

IV. METHODOLOGY

To create an accurate Synchronization Signal-reference signal received power (SS-RSRP) prediction model, the features must be able to properly define the receiver (Rx) location in relation to the transmitter (Tx) antenna location. Aside from that, the features must be able to explain the signal propagation status as well as the operational environment's properties. As a result, the quantity of signal attenuation encountered before to reaching the Rx location can be predicted more precisely.

SS-RSRP reading measurement can be performed on 5G NR Network. 5G NR networks have distinct characteristics, but the objective of their use is the same, there is UE measures RSRP on a regular basis for cell selection/reselection and handover [44]

A. 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 intensity measurements, all of which contribute to coverage prediction. The dataset was meticulously preprocessed prior to analysis, including data cleansing, treatment of missing values, and scaling to ensure uniformity and correctness in later phases

For this study, dataset is constructed from a comprehensive measurement campaign conducted in Batununggal Area, a densely populated area in Bandung City, West Java, Indonesia. This area representing urban environment using hardware and software described in Table 1. The measurement campaign was conducted at a vehicle speed below 30 km/h to minimize the fast fading effect due to the Doppler shifts [45]. This area consist of 10 g-NodeB, and each g-NodeB has 3 tranceiver BS Antenna. The selection of Batununggal area was made because in the area there are already 5G networks that are quite dominant covering the area. Because, in Indonesia for the existence of 5G network is still in the stage of development by the provider of cellular network services, so it is not massive and not comprehensive. For drive test result in Batununggal Area, shown in Figure 1

The measurement campaign was carried out for two reasons, there are to generate the model training dataset and to generate a test dataset. The drive test data must be cleansed before any dataset preparation procedures can begin. The data acquired under static conditions must be deleted to verify that the data is error-free [46]. From base station specification we will have some parameters that can be used for the training model of the dataset and also to determine other parameters, there are antenna type, base station antenna height at above sea level (ASL), antenna tilting degree, base station position coordinate, and antenna direction. And also, from drive test results obtained in the Batununggal area, as shown in Figure 1, are subsequently extracted in the form of.csv and then processed. From the processing results based on the drive test results carried out, we will get some parameters data that can be used for the training model of the dataset, among them are 2D distance between eNodeB (eNB) and UE, antenna height of UE at above sea level (ASL), UE position coordinate. From correlation of base station specification and drive test result, we will get and find another generated parameter, there are elevation angle from g-NodeB to UE position, azimuth offset angle, tilting offselt angle, and horizontal and vertical distance of receiver from boresight of base station antenna. The input feature parameters are further summarized in Table 2 for a clear explanation. The selection of these ten factors as input data was based on previous work using electromagnetic wave propagation understanding.

B. MODEL TRAINING AND VALIDATION

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 [47]. It can also be used to test basic

FIGURE 1. Distribution of drive test data around Batununggal area.

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

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 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 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 [20].

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

where:

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

The smaller values of RMSE indicate a better prediction of the ML model. According to [48], 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) [49].

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$

where

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

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

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 Angle	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

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 [50].

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

where:

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

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 performance of the classification model machine learning algorithm. There are 4 performance evaluations used to evaluate the classification model machine learning algorithm, namely Accuracy, Precission, Recall, and F1-Score.

Accuracy, Precision, Recall, and F1-Score are evaluation metrics used to assess the performance of machine learning classification models. It is essential to evaluate the classifications model's performance before using it in production to solve real-world situations. Machine learning classification model performance measurements are used to analyze how well machine learning classification models perform in a given situation. This performance evaluation enables us to understand the models' strengths and limitations when generating predictions in new scenarios. As the severity of different types of errors differs among use cases, these measures are utilized to balance the classifier estimates as preferred. Accuracy, Precision, Recall, and F1-Score are significant evaluation factors in estimating coverage, especially in the context of 5G technology. Coverage prediction is the ability to forecast how well a 5G network will cover a certain geographic area and how well it will provide dependable services in that location.

The base metric used for model evaluation is often Accuracy, describing the number of correct predictions over all predictions. Accuracy can be used to measure the extent to which the coverage prediction of the 5G model is correct overall. In this context, accuracy will measure the extent to which the model can correctly predict the areas that will have 5G coverage. It is calculated as the ratio of the number of correct predictions to the total number of predictions.

Precision is the ratio of true positives to the total number of predicted positives. It measures the proportion of actual positive cases that were correctly identified. Precision measures the extent to which the model's positive predictions are correct. Precision measures the extent to which the model's positive predictions are correct. It is useful when we want to minimize false positives [51]. Precision in the context of 5G coverage prediction measures the extent to which areas predicted to have 5G coverage will actually have such coverage. In this case, precision will help minimize false positives, i.e. areas that are incorrectly predicted to have 5G coverage.

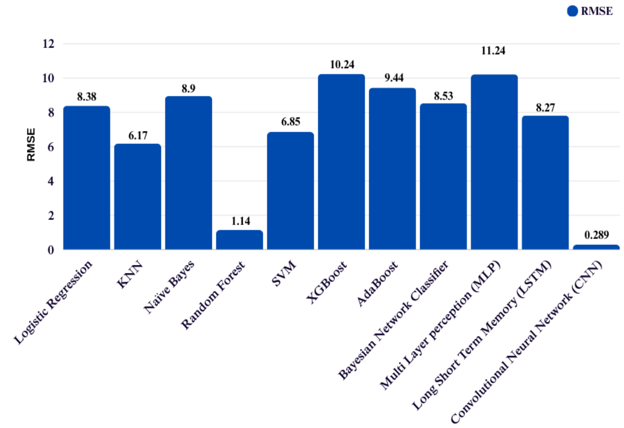

FIGURE 2. RMSE result value for each model algorithm.

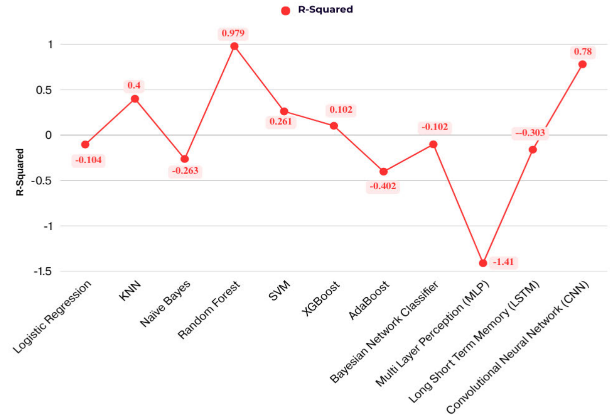

FIGURE 3. R^2 result value for each model algorithm.

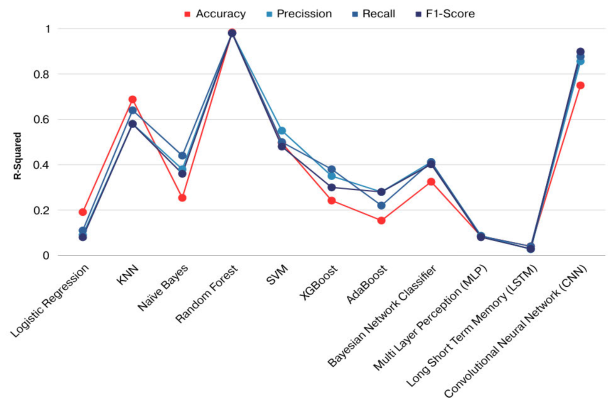

FIGURE 4. Performance evaluation metrics of machine learning algorithm on training data.

Recall is the ratio of true positives to the total number of actual positives. It measures the proportion of actual positive cases that were correctly identified. Recall can measure the extent to which the model is able to correctly identify areas that should have 5G coverage. In this case, recall will help minimize false negatives, which are areas that should have 5G coverage but are not predicted. [51]

The F1-score represents the model score as a function of precision and recall. F-score is a machine learning model performance metric that gives equal weight to Precision and Recall when measuring accuracy, making it a viable alternative to Accuracy metrics (it does not necessitate knowing the whole number of observations). It is typically used as a single value that provides high-level information about the model's output quality. This is a useful model metric when attempting to optimize either accuracy or recall score and the model performance suffers as a result. F1-Score can be used to find a balance between precision and recall in predicting 5G coverage. This will aid in optimizing the trade-off between false positives and false negatives. [52]

The final procedure in the model training session is hyperparameter tuning, in which a model is tuned in greater detail to achieve the best performance results. The process of determining the ideal settings of hyperparameters in a machine learning model in order to improve its performance is known as hyperparameter tuning [36]. Hyperparameters are external configuration variables that are used to regulate a machine learning model's training. The purpose of hyperparameter tuning is to discover the best combination of hyperparameters for minimizing the loss function and improving the model's accuracy or F1 score. Hyperparameter tuning is critical in machine learning because it ensures that the model performs optimally and does not overfit or underfit the data. The significance of hyperparameter tuning stems from the fact that it has the potential to greatly increase the performance of a machine learning model. We can verify that the model is not biased towards specific characteristics or parameters by determining the ideal values of hyperparameters. Tuning the hyperparameters can also help to lessen the danger of overfitting and improve the model's robustness. In this study, hyperparameter tuning optimization is only performed on the algorithm model that has the best performance evaluation value among other algorithm models.

C. DOMINANT PARAMETER IDENTIFICATION

The discovery of dominant feature characteristics driving 5G coverage prediction was one of our investigation. To evaluate the impact of each parameter on prediction accuracy, advanced approaches such as feature importance scores, correlation analysis, and permutation importance were used. The findings were thoroughly examined in order to identify the factors that had the greatest impact on coverage prediction.

The study found that selecting important features is more important than designing the prediction mode [53]. In another study that predicted mobile network coverage, the input parameters used were the 3D distance between Tx and Rx, Tx height, Rx height above sea level, and frequency [8]. There are many ways to estimate feature importance, including random forest measures, permutation feature importance, and tree-based feature importance [54]. Wrapper methods such as recursive feature elimination use feature importance to more efficiently search the feature space for a model.

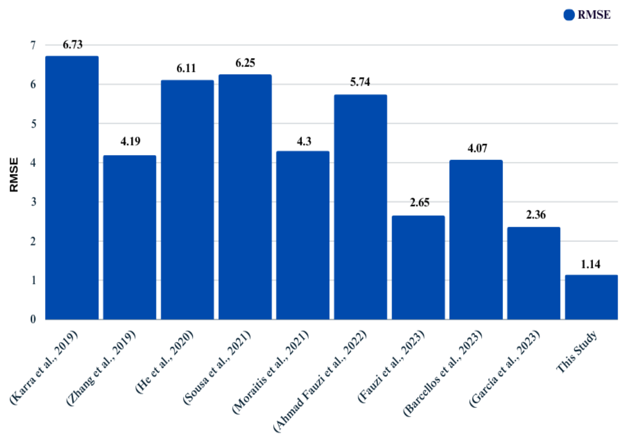

FIGURE 5. Comparison of RF model performance with the previous works.

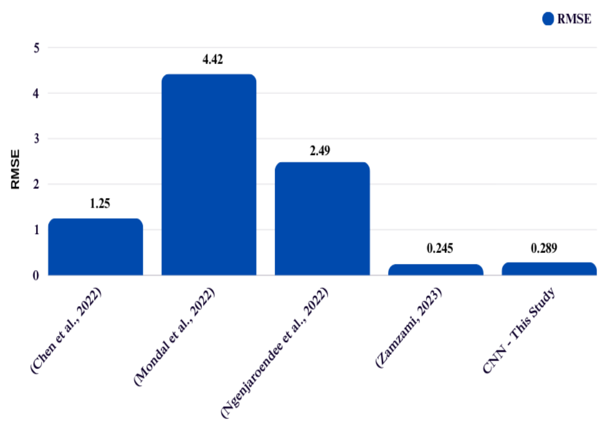

FIGURE 6. Comparison of CNN model performance with the previous works.

Model inspection and communication can also benefit from feature importance. Stakeholders may be curious about which features are most significant for prediction, and feature importance can assist in answering this issue. In this study, we evaluate feature importance using the feature importance algorithm in the classification model algorithm.

This highly organized methodology allowed for a thorough comparison of machine learning algorithms, allowing for the identification of dominating feature parameters and the measurement of prediction accuracy. The study's findings may greatly improve 5G network design and optimization tactics, paving the door for more efficient and resilient 5G deployments.

These studies have highlighted the importance of feature parameters in predicting 5G coverage. By identifying 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. It will be analyzed based on the trained data against several features used in the machine learning classifier algorithm. Each parameter feature will certainly bring out the dominant and most frequent role in each data training process. Each algorithm, of course, will also have which parameter features are more dominant. This is done and analyzed, hoping to choose the dominant parameter feature to be able to produce a more accurate prediction model.

TABLE 3. Performance evaluation metrics of machine learning and deep learning algorithm on training data.

Algorithm	RMSE	MAE	R^2
Logistic Regression	8.38	6.3	-0.104
KNN	6.176	2.86	0.4
Naïve Bayes	8.964	6.63	-0.263
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
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

V. RESULT AND DISCUSSIONS

A. FEATURE IMPORTANCE

The idea of "feature importance" emerges as a keystone in comprehending the intricate interplay between numerous input parameters and prediction accuracy in the field of 5G coverage prediction, where the efficacy of machine learning algorithms is analyzed. In essence, feature importance serves as a vital link between predictive models and real-world elements influencing 5G coverage. Its significance in our study. The significance of feature importance extends beyond network planning and optimization. It provides legislators and decision-makers with the data-driven insights they need to establish effective 5G infrastructure development strategies. Policymakers can direct investments and regulations to areas that demand attention by focusing on the most relevant feature factors.

Different machine learning algorithms may have different dominant feature parameters that affect the accuracy of coverage prediction. In this study comparing the accuracy of data classification models. From Figure 7, the figure shows the distribution of the dominance of parameter features used for each machine learning algorithm model used in this study. Each algorithm has a dominant feature to be able to produce

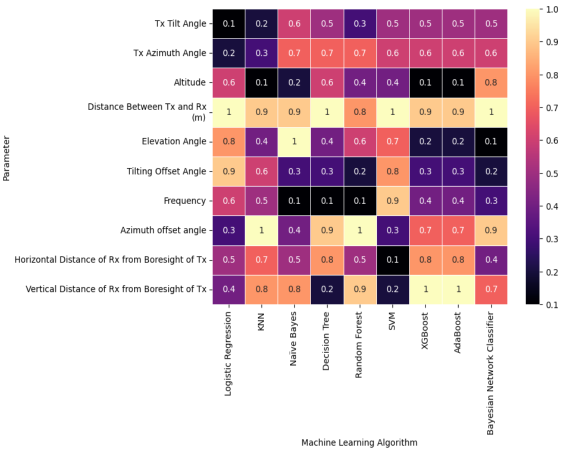

FIGURE 7. Heatmap of feature parameters in 5G coverage prediction.

accurate and better predictions. Of course, each algorithm will produce different feature choices.

Based on the results of this study, the 2D Distance Tx to Rx parameter is the dominant feature parameter in predicting coverage prediction in all algorithm models used in the research. However, other parameters still have their own different influences on each algorithm model. The operating frequency is also an important feature parameter in path loss prediction models. Ray-tracing models investigate hundreds of rays and their interactions with the environment, which is a valid alternative for radio coverage assessments.

B. MACHINE LEARNING PREDICTION MODEL AND COMPARATIVE RESULT

In this study, we evaluate the model predicted using Machine Learning Classifier Algortihm, there are Logistic Regression (LR), KNN, Naïve Bayes, Decision Tree, Random Forest (RF), Support Vector Machine (SVM), XGBoost, AdaBoost, Bayesian Network Classifier. And also using Deep Learning Algorithm, there are Multi Layer Perception, Long Short Term Memory, and Convolutional Neural Network (CNN). Table 3 and Table 4 report the comparative result for all algorithm. Table 3 is shown the performance evaluation metrics of machine learning algortihm on training data, such as RMSE, MAE, dan R^2 value and Table 4 is shown comparative result for performance evaluation for each machine learning and deep learning algorithm. The Table 3 and Table 4 are also represented in the form of a graph displayed in the figure in Figure 2, 3, and 4.

Figure 2 and 3 shown the visual comparative result RMSE and R^2 for all algorithm. These metrics help us to measure the accuracy of the model's predictions and to determine how well the model fits the data. RMSE and MAE are used to measure the accuracy of the model's predictions, while R^2 is used to measure how well the model fits the data. A good model should have a low RMSE and MAE and a high R^2 value. And also from Figure 4 shown the visual comparative result for performance evaluation for each algorithm. This performance evaluation allows us to identify the strengths and limitations of these models when generating predictions in new settings. Likewise, the performance evaluation results of each classification model machine learning algorithm and deep learning algorithm that used in coverage prediction in this study.

It can be seen the best result from the nine of machine learning algorithms classification models and three of deep learning algorithms used in this 5G coverage prediction. From machine learning algorithms classification models, 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. These two models show the best metric evaluation performance compared to other models. The evaluation results produced by CNN are the best results of the three deep learning algorithms used in this study. In addition, the RMSE value generated by CNN also shows better results when compared to the RMSE results from using the Random Forest algorithm. This can be seen in the RMSE results in Figure 2. However, for other performance evaluation results, Random Forest turns out to have better evaluation results than the CNN model.

As shown in Table 4 and Figure 4, the Random Forest models show an almost perfect performance evaluation, where the accuracy value is 98.4%, precision is 98%, recall and F1-score is also 98%. In addition, of the three deep learning algorithm models used in 5G coverage prediction in this study, the CNN algorithm shows much better evaluation results when compared to other models, namely 0.75 or 75% for accuracy value, 0.856 or 85.6% for precision value, 0.878% for recall value, and 0.899 or 89.9% for F1-score. These evaluation results are the best results of the three deep learning algorithms used, but the performance evaluation results are not as good as the evaluation results produced by the random forest algorithm.

In this research, the use of parameter features is only limited to 10 parameter features which are numeric feature types. When processing and training training data for prediction using the CNN algorithm, these ten parameter features are arranged in such a way as to form a matrix shape so that it resembles the shape of a grid matrix or spatial data. Whereas in coverage prediction using the random forest algorithm, the trained data is not arranged in a matrix form resembling grid data or spatial data. In the processing and data training process, it is much easier to use the Random Forest algorithm than CNN if the dataset used is only in the form of ordinary numerical features.

The use of CNN is actually highly recommended for coverage prediction, especially if the dataset used not only uses numerical features, but also uses image features. CNNs are designed to automatically extract image features, eliminating the need to manually design features, thus increasing the generalizability of the method. CNNs are specifically designed for grid-based data processing, such as images or spatial data. This makes them more efficient and effective in handling tasks that involve understanding spatial patterns, such as in signal level coverage prediction. Whereas, Random Forest is less optimal for grid-based data processing and tends to give better results on data structured in different ways. Thus, the use of parameter features used for coverage prediction will also greatly affect the evaluation results and accuracy produced by the algorithms used. If the parameter features used are only numerical feature data, it is more advisable to use the Random Forest algorithm. However, if the parameter features used use satellite image features or other spatial image data, it is more advisable to use the CNN algorithm to get much better prediction results. CNN tends to perform better when using a large amount of training data, while RF tends to perform better on small data. However, there are also some challenges to consider in using CNNs, such as limitations in classification tasks, especially with limited training data. A combination of CNN for feature extraction and RF for classification can be a good option in some cases, especially when image data is involved.

In this study, analysis and comparisons are not only made between machine learning and deep learning algorithm models, comparisons are also made with previous research. This comparison is only done on Random Forest and CNN models, which have the best performance evaluation results and evaluation metrics compared to other algorithms. The comparison is shown in Figure 5 and Figure 6. Figure 5 shows the comparison of evaluation results in the form of RMSE between Random Forest used in this study and Random Forest used in other previous studies. While Figure 6 shows the comparison of evaluation results in the form of RMSE between CNN used in this study and CNN used in other previous studies. From these two results, it shows that the RMSE of Random Forest in this study has much better results when compared to other studies. As for the CNN evaluation results, it shows that the RMSE value of the CNN model used in this study has results that are close to the best results from previous research [16]. Although it has not produced better performance results, the use of CNN in this study can already produce evaluation performance that is close to the best results in previous studies. From both of this figure shows the best results when compared to other previous studies. The difference in these results can be caused by several things, such as the use of datasets that are different from previous studies, the use of variations in parameter features used, as well as the number of data samples trained for coverage prediction. In addition to producing the best RMSE performance evaluation results, this research was also conducted to predict coverage in 5G networks in Indonesia, which is still quite limited in research.

Random forest used in this research, produces a better performance evaluation value when compared to previous research. As shown in Figure 5 which is a graph showing the RMSE value of the results of the Random Forest model performed in this study compared to other previous studies. In other studies, the random forest model also shows the best performance evaluation results for coverage prediction when compared to other algorithms used. In this study, the performance evaluation of the Random Forest model used produced the lowest and best RMSE value, which was 1.14. This condition has met the criteria, where the RSME value measured in the prediction results for urban or urban areas is recommended to be below 7 value. The Random Forest RMSE results obtained in this study are the best evaluation results obtained from training using the Random Forest algorithm that has been optimized. The optimization performed is in the form of Hyperparameter Optimization (HPO). This HPO is a process to find the best and optimal configuration of the algorithm used in order to get more accurate and better prediction results, and also to improve the performance of the Random Forest model in the signal coverage prediction task. From the HPO process, the optimal hyperparameter configuration combination for the Random Forest algorithm for coverage prediction in this study is max_depth = 20, max_features = 'auto', and n_estimators = 300. This hyperparameter refers to the number of decision trees to be built in the Random Forest ensemble. By setting n_estimators to a value of 300, the model is built from 300 independent decision trees. The addition of decision trees can improve prediction accuracy and make the model more stable. However, there is a limit where increasing the number of trees will provide diminishing additional benefits. These HPO results reflect an attempt to strike a balance between the complexity of the model and its ability to generalize.

The usage of the CNN algorithm in this current study, shows quite good evaluation results, which are shown by the RMSE evaluation value of 0.289. The RMSE results generated in this study show better results when compared to previous studies related to the use of CNN algorithms on coverage prediction in cellular communication systems [17], [18], [19]. Although the RMSE results in research [16], showed a value of 0.245 and slightly better than the current research. This is due to differences in the datasets used, both in the number of data samples trained and the training data used. In addition, in research [16], the use of the number of features used and the methodological process of training the training data are also not clearly conveyed. In addition, the prediction performed in [16], does not specifically predict coverage, but also performs a combined prediction between utilizing channel data, context metrics, cell metrics, and throughput statistics. So that the resulting RMSE evaluation results still need to be reviewed.

The use of the CNN algorithm in predicting 5G coverage carried out in this study, produces a better RMSE value when compared to the random forest algorithm. CNN produces an RMSE value of 0.289, while random forest only produces an RMSE value of 1.14. In an effort to improve the CNN algorithm's performance in coverage prediction, we conducted a rigorous Hyperparameter Optimization (HPO) process. This optimization process involves experimentation and iterative adjustments to achieve optimal performance. From the HPO process, we obtained the optimal combination to get the most optimal RMSE value of 0.289.

In the HPO process performed on the CNN model used in this study, the CNN model is built using Sequential objects, which are linear containers for the layers of the model. This implies that our model will be built sequentially, one layer after another. The first layer added is a convolution layer (Conv1D) with 32 filters, a convolution window of size 2, and a ReLU activation function. This convolution layer is used to extract spatial features from the input data. After that, a dimensionality reduction layer (MaxPooling1D) with a window size of 2 is added. This is done to reduce the dimensionality of the convolution result and retain the most significant features. The same steps are repeated by adding a second convolution layer with 64 filters and a convolution window of size 2, followed by a dimensionality reduction layer. Then, a Flatten layer is added to flatten the output of the previous layers into a one-dimensional vector. This is followed by two Dense layers (fully connected layers), with 128 neurons and ReLU activation function for the hidden layer, and one neuron with linear activation for the regression task in the output layer. After building the model architecture, further configuration was performed. The model is compiled using 'adam' optimization, which is a moment method-based stochastic gradient optimization. The model was then trained using training data for 200 epochs (training cycles). Each batch of one sample is processed to reduce the gradient and update the model parameters.

Actually, the use of the CNN algorithm is still rarely used in the context of coverage prediction, especially in 5G network systems. This is due to several things, including the fact that there are still limited spatial data that can be used to process using the CNN algorithm, because the CNN algorithm requires a large volume of data for effective training, and in the context of the telecommunications world, obtaining sufficient data can be a challenge. In addition, CNN modeling requires considerable computational resources, especially if the model architecture is quite complex. Another thing is that CNNs, especially in very deep models, are often perceived as "black boxes" due to their complexity. This limitation of interpretability can be problematic in environments where better interpretation of results or intuitive understanding of predictions is important. With the limitations of the CNN algorithm, this algorithm is still recommended especially in coverage prediction, especially if the data used is image data or spatial data that is sufficient to produce more accurate coverage predictions.

From this study, it can be recommended the use of CNN and Random Forest algorithms to be used in 5G coverage prediction. Because both of them produce better evaluation performance when compared to other algorithms used in this study. In addition, the use of additional new parameter features in the form of Horizontal and Veritcal Distance of Rx From Boresight of Tx Antenna in this study, also helps to improve the prediction results produced in this study, so that it can produce better performance evaluation results from training data when compared to other previous studies. This research also proves that the use of machine learning and deep learning algorithms can be a recommendation and consideration in predicting coverage and RSRP signal levels in the process of planning and optimizing cellular networks in 5G networks. The use of this algorithm can still produce fairly accurate signal level prediction results and is more efficient and flexible when compared to using previous conventional methods.

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 F1-score of 98%. Furthermore, of the three deep learning algorithm models used in this study's 5G coverage prediction, the CNN algorithm outperforms the others, scoring 0.75 or 75% for accuracy, 0.856 or 85.6% for precision, 0.878% for recall, and 0.899 or 89.9% for F1-score. These assessment results are the best of the three deep learning algorithms employed, however the performance evaluation results are not as good as the random forest algorithm's evaluation results.

It was found that the performance of the final trained Forest (RF) and CNN model in this study is better than other model, and also the Random Forest (RF) and CNN model in this study have better performance than developed in previous study. The differences in these results can be explained as a variety of factors, including the use of datasets that differ from earlier studies, variations in parameter features used, and the quantity of data samples trained for coverage prediction. 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.

The results also reveal that the most dominant feature parameter in all algorithms is 2D Distance Tx Rx. This feature was found to have a significant influence in affecting the prediction of 5G coverage. Horizontal and vertical distances of Rx from boresight of Tx Antenna in this study contribute as well to improve prediction results, so that it may produce much better performance evaluation results from training data when compared to other previous studies. This study also shows that the use of machine learning and deep learning algorithms can be a recommendation and factor in predicting coverage and RSRP signal levels during the development and optimization of cellular networks in 5G networks. When compared to earlier conventional methods, our algorithm can still give reasonably accurate signal level prediction results while being more efficient and adjustable.

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 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] C. Sudhamani, M. Roslee, J. J. Tiang, and A. U. Rehman, "A survey on 5G coverage improvement techniques: Issues and future challenges," *Sensors*, vol. 23, no. 4, p. 2356, Feb. 2023, doi: 10.3390/s23042356.

[2] M. M. Ahamed and S. Faruque, "5G network coverage planning and analysis of the deployment challenges," *Sensors*, vol. 21, no. 19, p. 6608, Oct. 2021, doi: 10.3390/s21196608.

[3] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018, doi: 10.1109/LWC.2017.2757490.

[4] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep learning-based channel estimation," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 652–655, Apr. 2019, doi: 10.1109/LCOMM.2019.2898944.

[5] T. Gruber, S. Cammerer, J. Hoydis, and S. T. Brink, "On deep learning-based channel decoding," in *Proc. 51st Annu. Conf. Inf. Sci. Syst. (CISS)*, 2017, pp. 1–6, doi: 10.1109/CISS.2017.7926071.

[6] Y. H. Santana, R. M. Alonso, G. G. Nieto, L. Martens, W. Joseph, and D. Plets, "Indoor genetic algorithm-based 5G network planning using a machine learning model for path loss estimation," *Appl. Sci.*, vol. 12, no. 8, p. 3923, Apr. 2022, doi: 10.3390/app12083923.

[7] *2018 ITU Kaleidoscope Machine Learning for a 5G Future (ITU K)*, IEEE Staff, Piscataway, NJ, USA, 2018.

[8] M. F. A. Fauzi, R. Nordin, N. F. Abdullah, and H. A. H. Alobaidy, "Mobile network coverage prediction based on supervised machine learning algorithms," *IEEE Access*, vol. 10, pp. 55782–55793, 2022, doi: 10.1109/ACCESS.2022.3176619.

[9] M. F. A. Fauzi, R. Nordin, N. F. Abdullah, H. A. H. Alobaidy, and M. Behjati, "Machine learning-based online coverage estimator (MLOE): Advancing mobile network planning and optimization," *IEEE Access*, vol. 11, pp. 3096–3109, 2023, doi: 10.1109/ACCESS.2023.3234566.

[10] H. Chiroma, A. Y. Gital, N. Rana, S. M. Abdulhamid, A. N. Muhammad, A. Y. Umar, and A. Abubakar, "Nature inspired meta-heuristic algorithms for deep learning: Recent progress and novel perspective," in *Advances in Computer Vision*, K. Arai and S. Kapoor, Eds. Cham, Switzerland: Springer, 2020, pp. 59–70.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[12] 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, Aug. 2022, doi: 10.1016/j.anl.2021.10.003.

[13] 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, Aug. 2018, doi: 10.1109/LCOMM.2018.2841832.

[14] 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, Jan./Mar. 2020, doi: 10.1109/TNSE.2018.2848960.

[15] S. Joseph, R. Misra, and S. Katti, "Towards self-driving radios: Physical-layer control using deep reinforcement learning," in *Proc. HotMobile*, 2019, pp. 69–74, doi: 10.1145/3301293.3302374.

[16] I. F. Zamzami, "Deep learning models applied to prediction of 5G technology adoption," *Appl. Sci.*, vol. 13, no. 1, p. 119, Dec. 2022, doi: 10.3390/app13010119.

[17] M. Chen, M. Chateauvert, and J. Ethier, "Extending machine learning based RF coverage predictions to 3D," in *Proc. IEEE Int. Symp. Antennas Propag. USNC-URSI Radio Sci. Meeting (AP-S/URSI)*, Jul. 2022, pp. 205–206, doi: 10.1109/ap-s/usnc-ursi47032.2022.9887000.

[18] W. U. Mondal, P. D. Mankar, G. Das, V. Aggarwal, and S. V. Ukkusuri, "Deep learning-based coverage and rate manifold estimation in cellular networks," *IEEE Trans. Cognit. Commun. Netw.*, vol. 8, no. 4, pp. 1706–1715, Dec. 2022, doi: 10.1109/TCCN.2022.3201508.

[19] T. Ngenjaroendee, W. Phakphisut, T. Wijitpornchai, P. Areeprayoonkij, and T. Jaruvitayakovit, "Deep learning-based reference signal received power prediction for LTE communication system," in *Proc. 37th Int. Tech. Conf. Circuits/Syst., Comput. Commun. (ITC-CSCC)*, Jul. 2022, pp. 888–891, doi: 10.1109/ITC-CSCC55581.2022.9895098.

[20] J. Nagar, S. K. Chaturvedi, S. Soh, and A. Singh, "A machine learning approach to predict the k-coverage probability of wireless multihop networks considering boundary and shadowing effects," *Expert Syst. Appl.*, vol. 226, Sep. 2023, Art. no. 120160, doi: 10.1016/j.eswa.2023.120160.

[21] H. E. Hammouti, M. Ghogho, and S. A. Raza Zaidi, "A machine learning approach to predicting coverage in random wireless networks," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2018, pp. 1–6, doi: 10.1109/GLOCOMW.2018.8644199.

[22] R. He, Y. Gong, W. Bai, Y. Li, and X. Wang, "Random forests based path loss prediction in mobile communication systems," in *Proc. IEEE 6th Int. Conf. Comput. Commun. (ICCC)*, Dec. 2020, pp. 1246–1250, doi: 10.1109/ICCC51575.2020.9344905.

[23] Y. Zhang, J. Wen, G. Yang, Z. He, and J. Wang, "Path loss prediction based on machine learning: Principle, method, and data expansion," *Appl. Sci.*, vol. 9, no. 9, p. 1908, May 2019, doi: 10.3390/app9091908.

[24] M. Sousa, A. Alves, P. Vieira, M. P. Queluz, and A. Rodrigues, "Analysis and optimization of 5G coverage predictions using a beamforming antenna model and real drive test measurements," *IEEE Access*, vol. 9, pp. 101787–101808, 2021, doi: 10.1109/ACCESS.2021.3097633.

[25] N. Moraitis, L. Tsipi, and D. Vouyioukas, "Machine learning-based methods for path loss prediction in urban environment for LTE networks," in *Proc. 16th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Oct. 2020, pp. 1–6, doi: 10.1109/WiMob50308.2020.9253369.

[26] A. Gupta, K. Ghanshala, and R. C. Joshi, "Machine learning classifier approach with Gaussian process, ensemble boosted trees, SVM, and linear regression for 5G signal coverage mapping," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 6, no. 6, p. 156, 2021, doi: 10.9781/ijimai.2021.03.004.

[27] A. Volokita and B. Hereha, "Comparative analysis of machine learning algorithms for cardiovascular diseases prediction systems," *Tech. Sci. Technol.*, vol. 4, no. 30, pp. 130–139, 2022.

[28] R. J. M. Cabauatan and B. D. Gerardo, "Predicting classification of telecommunication subscribers based on polytomous logistic regression usage model," in *Proc. 9th Int. Conf. Mach. Learn. Comput.* New York, NY, USA, Feb. 2017, pp. 258–267, doi: 10.1145/3055635.3056615.

[29] S. Bharadwaj, B. S. Anil, A. Pahargarh, A. Pahargarh, P. S. Gowra, and S. Kumar, "Customer churn prediction in mobile networks using logistic regression and multilayer Perceptron(MLP)," in *Proc. 2nd Int. Conf. Green Comput. Internet Things (ICGCIoT)*, Aug. 2018, pp. 436–438, doi: 10.1109/ICGCIoT.2018.8752982.

[30] S. A. R. Zaidi, "Nearest neighbour methods and their applications in design of 5G & beyond wireless networks," *ICT Exp.*, vol. 7, no. 4, pp. 414–420, Dec. 2021, doi: 10.1016/j.icte.2021.01.003.

[31] R. K. Gupta, P. Pannu, R. Misra, and R. K. Gupta, "Towards ultra-latency using deep learning in 5G network slicing applying approximate k-nearest neighbor graph construction," *Res. Square*, 2021. [Online]. Available: https://doi.org/10.21203/rs.3.rs-162479/v1

[32] S. Mohammadjafari, S. Roginsky, E. Kavurmacioglu, M. Cevik, J. Ethier, and A. B. Bener, "Machine learning-based radio coverage prediction in urban environments," *IEEE Trans. Netw. Service Manage.*, vol. 17, no. 4, pp. 2117–2130, Dec. 2020, doi: 10.1109/TNSM.2020.3035442.

[33] D. Karra, S. K. Goudos, G. V. Tsoulos, and G. Athanasiadou, "Prediction of received signal power in mobile communications using different machine learning algorithms: A comparative study," in *Proc. Panhellenic Conf. Electron. Telecommun. (PACET)*, Nov. 2019, pp. 1–4, doi: 10.1109/PACET48583.2019.8956271.

[34] A. Amin, A. Adnan, and S. Anwar, "An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naïve Bayes," *Appl. Soft Comput.*, vol. 137, Apr. 2023, Art. no. 110103, doi: 10.1016/j.asoc.2023.110103.

[35] C. E. García and I. Koo, "Extremely randomized trees regressor scheme for mobile network coverage prediction and REM construction," *IEEE Access*, vol. 11, pp. 65170–65180, 2023, doi: 10.1109/ACCESS.2023.3287103.

[36] S. M. Shrestha and A. Shakya, "A customer churn prediction model using XGBoost for the telecommunication industry in Nepal," *Proc. Comput. Sci.*, vol. 215, pp. 652–661, Jan. 2022, doi: 10.1016/j.procs.2022.12.067.

[37] P. Senthan, R. Rathnayaka, B. Kuhaneswaran, and B. Kumara, "Development of churn prediction model using XGBoost–telecommunication industry in Sri Lanka," in *Proc. IEEE Int. IoT, Electron. Mechatronics Conf. (IEMTRONICS)*, Apr. 2021, pp. 1–7, doi: 10.1109/IEMTRONICS52119.2021.9422657.

[38] N. Deeban and P. S. Bharathi, "BDT: An ada boost classifier ensemble with decision tree for traffic network prediction," in *Proc. 5th Int. Conf. Contemp. Comput. Informat. (IC3I)*, Dec. 2022, pp. 1926–1931, doi: 10.1109/IC3I56241.2022.10072623.

[39] D. Ferreira, A. Braga Reis, C. Senna, and S. Sargento, "A forecasting approach to improve control and management for 5G networks," *IEEE Trans. Netw. Service Manage.*, vol. 18, no. 2, pp. 1817–1831, Jun. 2021, doi: 10.1109/TNSM.2021.3056222.

[40] P. Weber, G. Medina-Oliva, C. Simon, and B. Iung, "Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas," *Eng. Appl. Artif. Intell.*, vol. 25, no. 4, pp. 671–682, Jun. 2012, doi: 10.1016/j.engappai.2010.06.002.

[41] Q. Meng, X. Fang, W. Yue, Y. Meng, and J. Wei, "Bayesian network prediction of mobile user throughput in 5G wireless networks," in *Proc. 10th Int. Conf. Commun., Circuits Syst. (ICCCAS)*, Dec. 2018, pp. 291–295. [Online]. Available: https://api.semanticscholar.org/CorpusID:198932210

[42] L. Wu, D. He, B. Ai, J. Wang, K. Guan, and Z. Zhong, "Path loss prediction based on multi-layer perceptron artificial neural network," *Chin. J. Radio Sci.*, vol. 36, no. 3, pp. 396–404, 2021, doi: 10.12265/j.cjors.2020209.

[43] H. Na, Y. Shin, D. Lee, and J. Lee, "LSTM-based throughput prediction for LTE networks," *ICT Exp.*, vol. 9, no. 2, pp. 247–252, Apr. 2023, doi: 10.1016/j.icte.2021.12.001.

[44] T. Kim, K. Ko, I.-C. Hwang, D. Hong, S. Choi, and H. Wang, "RSRP-based Doppler shift estimator using machine learning in high-speed train systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 371–380, Dec. 2021, doi: 10.1109/TVT.2020.3044175.

[45] S. I. Popoola, S. Misra, and A. A. Atayero, "Outdoor path loss predictions based on extreme learning machine," *Wireless Pers. Commun.*, vol. 99, no. 1, pp. 441–460, Mar. 2018, doi: 10.1007/s11277-017-5119-x.

[46] V. Raida, P. Svoboda, and M. Rupp, "On the inappropriateness of static measurements for benchmarking in wireless networks," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–5, doi: 10.1109/VTC2020-Spring48590.2020.9128867.

[47] Neptune.ai. *How to Use Google Colab for Deep Learning—Complete Tutorial*. Accessed: Sep. 5, 2023. [Online]. Available: https://neptune.ai/blog/how-to-use-google-colab-for-deep-learning-complete-tutorial

[48] N. Moraitis, L. Tsipi, D. Vouyioukas, A. Gkioni, and S. Louvros, "Performance evaluation of machine learning methods for path loss prediction in rural environment at 3.7 GHz," *Wireless Netw.*, vol. 27, no. 6, pp. 4169–4188, Aug. 2021, doi: 10.1007/s11276-021-02682-3.

[49] S. Dridi, V. Machine, D. Tree, R. Forest, and L. Regression, "Supervised learning—A systematic literature review," 2021. [Online]. Available: https://osf.io/preprints/osf/tysr4

[50] H.-S. Jo, C. Park, E. Lee, H. K. Choi, and J. Park, "Path loss prediction based on machine learning techniques: Principal component analysis, artificial neural network, and Gaussian process," *Sensors*, vol. 20, no. 7, p. 1927, Mar. 2020, doi: 10.3390/s20071927.

[51] K. P. Shung. (2018). *Accuracy, Precision, Recall or F1?* Accessed: Sep. 10, 2023. [Online]. Available: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9

[52] A. Kumar. (2023). *Accuracy, Precision, Recall and F1-Score—Python Examples*. Accessed: Sep. 10, 2023. [Online]. Available: https://vitalflux.com/accuracy-precision-recall-f1-score-python-example/

[53] R.-C. Chen, C. Dewi, S.-W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *J. Big Data*, vol. 7, no. 1, p. 52, Dec. 2020, doi: 10.1186/s40537-020-00327-4.

[54] J. Brownlee. (2020). *How to Calculate Feature Importance With Python*. Accessed: Sep. 10, 2023. [Online]. Available: https://machinelearningmastery.com/calculate-feature-importance-with-python/

HAJIAR YULIANA received the B.S. degree in electrical engineering and in telecommunication engineering expertise from Universitas Jenderal Achmad Yani, Indonesia, in 2013, and the master's degree in telecommunication engineering from Institut Teknologi Bandung, in 2017, where she is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Informatics. She is also a Lecturer in electrical engineering with Universitas Jenderal Achmad Yani. Her research interests include wireless communication, cellular technology, data science, and applied machine learning algorithms for mobile networks.

ISKANDAR received the B.S. degree in electrical engineering and the master's degree in electrical and telecommunication engineering from Institut Teknologi Bandung, in 1995 and 2000, respectively, and the Ph.D. degree from Waseda University, Japan, in 2007. His research interest includes wireless communication.

HENDRAWAN received the B.S. degree in electrical engineering from Institut Teknologi Bandung, in 1985, and the master's degree in telecommunication engineering and the Ph.D. degree in electronics systems engineering from the University of Essex, U.K., in 1990 and 1995, respectively. His research interests include queuing theory, data communication, multimedia communication, telecommunication networks, network management, big data and machine learning, and artificial intelligent.