International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5, 2025
https://theaspd.com/index.php

Comparison Of Machine Learning Algorithms For Predicting 5g Coverage Predictive Accuracy And The Identification Of Dominant Feature Parameters

Bhukya Maneesha[1], K .V. D Kiran[2]
[1]M.Tech, Department Of Computer Science And Engineering, Koneru Lakshmaiah Educational Foundation, Vaddeswaram,Guntur,A.P, Maneeshabhukya20@Gmail.Com
[2]Department Of Computer Science And Engineering, Koneru Lakshmaiah Educational Foundation Vaddeswaram,Guntur,A.P, kiran_cse@kluniversity.in

***Abstract**: In 5G technology, the prediction of coverage areas plays a vital part in network optimization and reliable connectivity. In this paper, coverage area prediction is presented on an extensive comparative analysis involving many machine learning algorithms based on the RF Signal Data. The target column, Band Width, is used to determine prediction accuracy of various models through evaluation. Evaluation is performed using traditional methods like Logistic Regression, K-Nearest Neighbors (KNN), Naive Bayes, Random Forest, Support Vector Machine (SVM), XGBoost, LightGBM, AdaBoost, Bayesian Network Classifier, Multi-Layer Perceptron (MLP), Long Short-Term Memory (LSTM); against proposed advanced techniques such as Stacking and Voting Classifiers, and Convolutional Neural Networks (CNN). The aim is to find the feature parameters that strongly influence 5G coverage prediction. This research is intended to benchmark the performance and accuracy of these algorithms through developing a wide range of models. The comparative analysis provides the advantageous and disadvantageous factors for each methodology, thus giving valuable insights for researchers and network engineers. The conclusion drawn from this work is that ensemble methods, namely Stacking and Voting Classifiers, along with CNN, attained much higher prediction accuracies and robustness, and therefore, are viable solutions for improving 5G network planning and deployment.*
***Keywords** - 5G Coverage Prediction, Machine Learning, RF Signal Data, Stacking Classifier, Voting Classifier, Convolutional Neural Network (CNN), Feature Parameters, Prediction Accuracy, Network Optimization, Ensemble Methods.*

I. OVERVIEW

The advent of fifth-generation (5G) wireless technology has marked a significant leap in the evolution of communication systems. With its promise of ultra-fast data speeds, significantly reduced latency, and the capability to connect a massive number of devices simultaneously, 5G stands at the core of future digital infrastructure. Its integration into various domains including smart cities, autonomous vehicles, healthcare systems, and industrial automation has already begun transforming the way we interact with technology. However, despite its benefits, one of the biggest challenges lies in the efficient and accurate prediction of 5G coverage areas to ensure optimal deployment and user satisfaction.

In current deployment scenarios, telecom providers are racing to expand 5G infrastructure across both urban and rural regions. However, ensuring consistent signal availability and performance requires careful planning. Predicting 5G coverage involves understanding complex signal behavior across various environments, accounting for factors like building density, terrain, and signal interference. Accurate coverage prediction not only reduces infrastructure costs but also improves user experience by minimizing connectivity issues and optimizing signal distribution.

Historically, Prediction of signals coverage depended on common methods of machine-learning algorithms traditionally. Such techniques as Logistic Regression, K-Nearest Neighbors (KNN), Naive Bayes, Random Forest, and Support Vector Machines (SVM) were commonly used for their simplicity and interpretability. Models would generally work under controlled environments; however, they fail to show any dominant behavior when exposed to real-world high-variance and non-linear-dependent datasets. They are not much generalized under all different conditions, which greatly limits their performance when required for modeling complex rf signals by modern 5g networks.

To address these limitations, recent studies have focused on advanced machine learning techniques capable of handling high-dimensional and heterogeneous data. The gradient boosting machinery of

XGBoost and LightGBM has proved superior having further boosted precision through an iterative process. Deep Learning Models such as Multi-Layer Perceptrons and LSTMs implement enhanced feature learning and temporal pattern recognition-providing well for model dynamic signal environments. These models have shown significant promise in capturing complex relationships between signal parameters and coverage outcomes.

In addition to individual models, ensemble learning techniques have emerged as a powerful tool in improving model robustness and generalization. Methods such as stacking and voting classifiers combine multiple algorithms to enhance prediction reliability. Furthermore, the incorporation of Convolutional Neural Networks (CNN) is being explored to analyze spatial signal distributions and patterns in RF data. CNNs can effectively capture localized signal variations, making them particularly useful in urban coverage scenarios where signal degradation can be irregular due to obstacles and reflections.

This research proposes a comparative analysis of both classical and modern machine learning models to evaluate their effectiveness in predicting 5G coverage based on RF signal parameters. Using features such as signal strength and bandwidth, the study aims to benchmark model performance in terms of accuracy, efficiency, and scalability. The goal is to identify a suitable model or hybrid approach that can be deployed in real-time network planning tools to guide telecom engineers in optimizing 5G deployment strategies. By bridging the gap between theoretical modeling and practical implementation, this work contributes to the growing body of research aimed at enhancing next-generation wireless network planning.

II. **RELATED WORK**

Sudhamani et al. (2023) offer an extensive examination of various strategies for enhancing coverage in 5G networks. They delve into prime limitations like interference at cell edges due to heavy deployment and mention some budding techniques for boosting spectral efficiency and reducing latency. Their survey also points out a plethora of future avenues for improvement in the wireless communication technologies, thus providing a wide landscape of challenges and research opportunities in the realm of 5G deployments.[1]

Ahamed and Faruque (2021) analyze the real-world considerations in deploying 5G over various frequency spectrums. The paper discusses planning complexities for high-frequency bands, where signal propagation loss necessitates dense deployments of small cells. The authors suggest architectural enhancements such as sectorized cells and smart antennas and point out that mobile operators will face logistical issues pertaining to the acquisition of infrastructure.[2]

According to Santana et al. 2022, Indoor 5G network planning is facilitated using a new machine learning-supported methodology. The method incorporates the predictive path loss model into the Genetic Algorithm, allowing signal coverage estimation with minimum access points.The hybrid solution, therefore, not only serves to an accelerated timed deployment but also satisfies radical design criteria such as signal strength consistency and reduced RF exposure.[3]

Fauzi and colleagues (2022) had studied various machine learning models for the prediction of mobile coverage. Their study compared linear regression, ANN, and random forest algorithms for predicting RSRP. Among the models, the one with the most accurate results was Gaussian Process Regression, while the Random Forest model was deemed most appropriate for real-time deployment in the environment as it offered an optimum balance between accuracy and speed.[4]

Fauzi et al. (2023) established the Machine Learning-based Online Estimator (MLOE) as a real-time prediction tool for mobile network performance. MLOE is based on the Random Forest algorithm, deployed on a MATLAB web server, and provides accurate estimates of network coverage by processing multiple input features. The model greatly improves the planning accuracy, and it is scalable for large-scale network deployments.[5]

Chiroma et al. (2020) present a survey exploring the integration of nature-inspired meta-heuristic algorithms with deep learning frameworks. Their work categorizes different algorithmic strategies and highlights how such methods can address optimization challenges in fields such as computer vision and autonomous systems. They propose new research directions to bridge the gap between meta-heuristic optimization and deep learning.[6]

This paper introduces a Convolutional Neural Network-Auto Encoder (CNN-AE) trained to predict location-dependent rate and coverage probability in cellular networks. CNN-AE improves upon stochastic geometry-based models by as much as 40% and 25% in coverage and rate prediction errors, respectively, when trained on Indian, Brazilian, German, and American BS location data. Furthermore, the model will assist in identifying better BS locations to adaptively achieve performance targets in a spatially heterogeneous manner. [7]

An ensemble machine learning approach to better path loss prediction in 4G LTE networks is presented in this work. By combining Radial Basis Function (RBF) and Multilayer Perceptron (MLP) neural network models, it is shown that increasing the number of centroids in the RBF model and use of Gaussian kernel function would adequately lower the Mean Square Error (MSE), and thereby improve prediction accuracy. [8]

The development of the modular model, that is 5GPA, has been achieved with the use of machine learning methods for predicting and improving the 5G network performance. The predicted model has higher precision of being 95% on average and low error rates in the categories of performance metrics. In addition, it guides the simplification of feature selection and optimization of 5G network performance during a design and planning phase. [9]

The study proposes a method for modeling radio propagation relying on a deep-convolutional-neural-network (CNN) architecture. The method provides considerably better performance than conventional empirical and deterministic models by exploiting unconventional site information, such as satellite imagery. Improved cellular network planning and optimization will be possible because of this advancement. [10]

Dubhe is a deep learning algorithm developed to study coverage in the Beyond 5G (B5G) networks. The model calibrates and modifies existing empirical models to suit a given environment, accurately predicting wireless signal coverage strength. This essentially permits fewer base station sites to be built at a reduced cost, thus making operations more efficient in fully representing B5G signal coverage under complicated and dynamic scenarios.[11]

III. SYSTEM DESIGN ARCHITECTURE

The current landscape of 5G coverage prediction is dominated by a variety of machine learning algorithms that utilize radio frequency (RF) signal data to forecast network reach. Techniques have been widely implemented to analyze and interpret signal parameters. While these models offer promising levels of accuracy and computational efficiency, they present several limitations. Complex models like Random Forest and CNN often lack interpretability, making them difficult to apply in real-time network planning scenarios. Moreover, their dependency on high-quality datasets, substantial computational resources, and static learning processes impedes scalability and adaptability. As 5G networks are inherently dynamic and vast, these limitations reduce the effectiveness of existing systems in rapidly changing and large-scale environments. To address the challenges posed by traditional models, the proposed system introduces a hybrid approach that combines ensemble learning techniques specifically Stacking and Voting Classifiers with deep learning models. [12] This system attempts to improve the prediction's robustness and accuracy by using the advantages of different methods. Through a structured process involving the preprocessing of RF signal data, model training, and rigorous validation, the system evaluates the performance of each technique to determine the most effective solution. Ensemble methods improve generalization by mitigating the weaknesses of individual models, while CNNs enable the extraction of complex patterns within the data. This advanced approach not only supports scalable and accurate 5G coverage prediction but also offers greater adaptability to the dynamic nature of networks, ultimately aiding in more intelligent and efficient 5G deployment strategies.

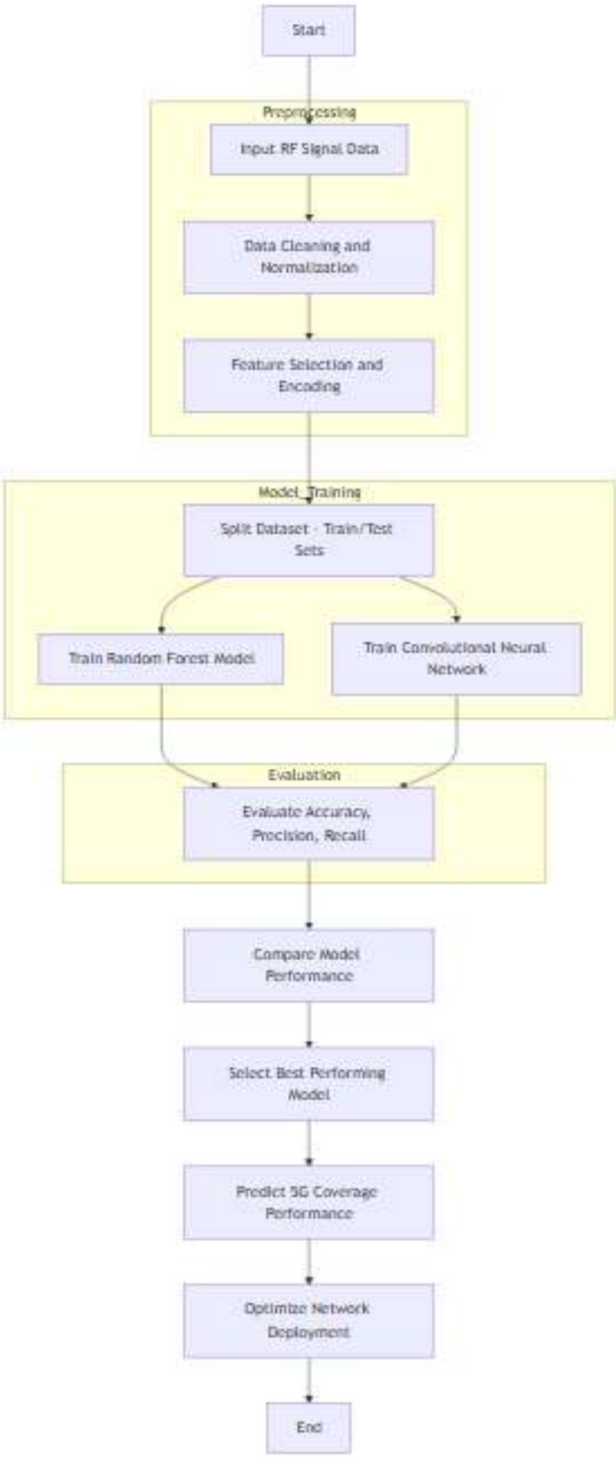

Figure 1 System Design Architecture Work Flow

IV. DATASET DESCRIPTION

In this work, a dataset consisting of on lakh+ instances and containing 20+ features relevant to 5G network performance and environmental conditions was used. The data are indexed by a RangeIndex from 0 to 10000+ and use up approximately 35mb of memory. The features are determined in different data types, including integers, float numbers, boolean, and categorical (object) types, giving a rich mix of numerical and contextual data for effective model training and evaluation. Key columns in the dataset include Timestamp, Frequency, Signal Strength, Modulation, and Bandwidth, which directly realize the characteristics of the 5G signal. Environmental parameters were also considered for their importance in signal propagation considerations. On the device side, Device Type, Antenna Type, Battery Level, Power Source, CPU Usage, Memory Usage, WiFi Strength, and Disk Usage provide clues about the operational environment of the user equipment. Location information, such as Latitude, Longitude, and Altitude(m), is available for spatial analysis of coverage. Some features such as Interference Type with missing values in respect of some of the entries and I/Q Data only partially populated are of particular note. Air Pressure is blank and hence should be deleted or imputed in preprocessing. The target variable for prediction is identified as Bandwidth, which is an integer-type field representing the capacity or throughput of the 5G signal. In summary, the dataset is an all-encompassing platform for the assessment of machine-learning paradigms to predict the coverage of 5G, integrating technical, environmental, and geographical data.

V. EXPLORATORY DATA ANALYSIS

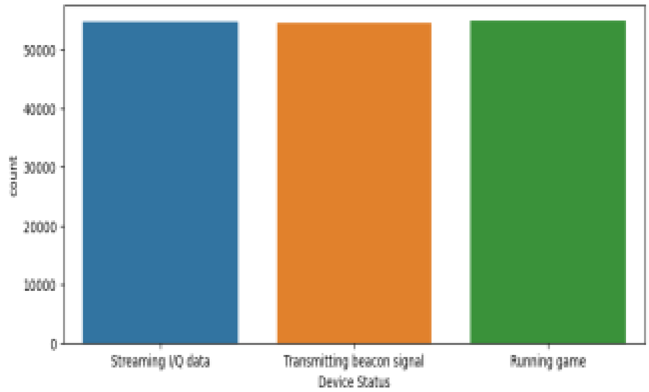

Figure 2 Bar Plot Device Status

The three device statuses "Streaming I/Q data," "Transmitting beacon signal," and "Running game" all have almost identical counts, slightly above 50,000. This suggests a balanced utilization of these functionalities, pointing to consistent demand across different device capabilities. The uniformity in activity could indicate well-distributed resource usage, preventing over-reliance on a single function. Such data could be valuable for performance analysis, helping to prioritize equal support for all three functionalities. The even distribution might reflect user behavior trends, highlighting that no single activity dominates the device's usage patterns.

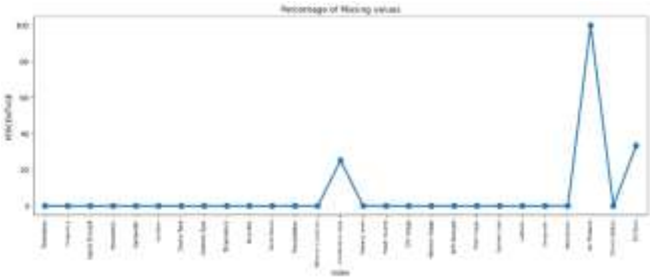

Figure 3 Link Plot Missing values

Parameters like "Weather Condition," "Air Pressure," and "I/O Data" show significantly higher percentages of missing values, indicating potential challenges in data quality for these variables. Most other parameters, such as "Timestamp," "Frequency," and "Device Status," exhibit minimal missing values, suggesting reliable data collection for these metrics. The disparity in missing values could point to differences in measurement methods or external factors affecting data capture for certain variables. Addressing the gaps in high-missing-value parameters could improve overall dataset completeness and enhance analysis accuracy. This graph provides a clear direction for prioritizing data cleaning efforts, focusing on the parameters with the most missing data.

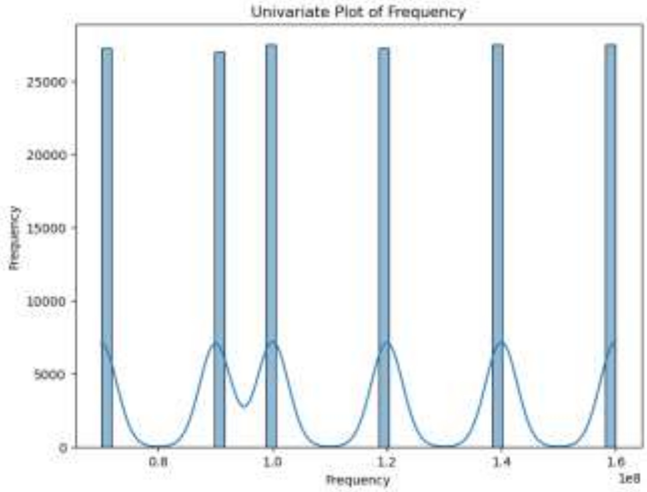

Figure 4 Distplot for Frequency

The frequency data exhibit distinct peaks, an indication that certain frequencies or patterns dominate the dataset. The histogram bars suggest specific frequency values that have disproportionately high occurrences and could represent fundamental frequencies of a signal. The oscillatory trend of the line plot between bars of the histogram indicates the periodicity that one could use to evaluate cyclic behavior in a signal or event. Bars that are extremely narrowed but high imply conciseness around a few values, suggesting frequency counts are distinctly not evenly spread but are rather clustered at certain points. Such kinds of visualizations are largely relevant to signal processing in the identification of important frequency components in music, electromagnetic waves, or the like.

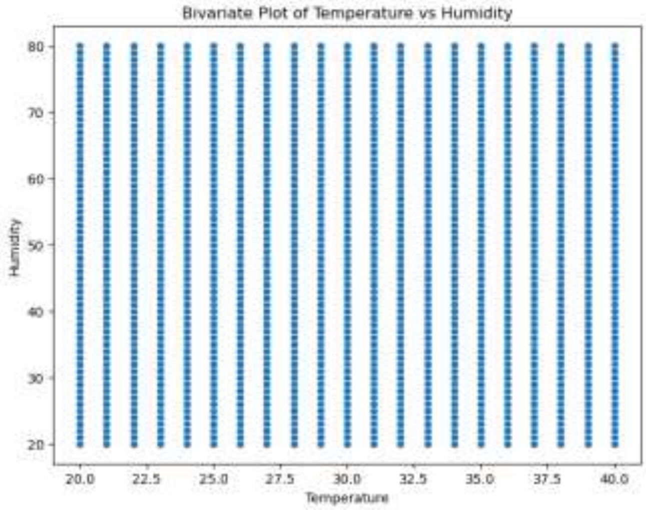

Figure 5 Scatter Plot between Temperature and Humidity

Humidity can vary widely across temperatures, suggesting that there might be other physical factors controlling humidity-there. In fact, the scatter plot of humidity against temperature does not reveal a clear upward or downward trend, indicating that, in consideration of this dataset, temperature does not control humidity. The vertical spread suggests that both extreme high and low humidity data are uniformly distributed across temperature ranges rather than clustering at specific points. This further indicates external environmental factors, such as wind systems or geography, might exert a more dominant control over humidity. These enormous fluctuations, therefore, offer considerable information for climate endeavors, including improvements in weather prediction models and an understanding of atmospheric behavior.

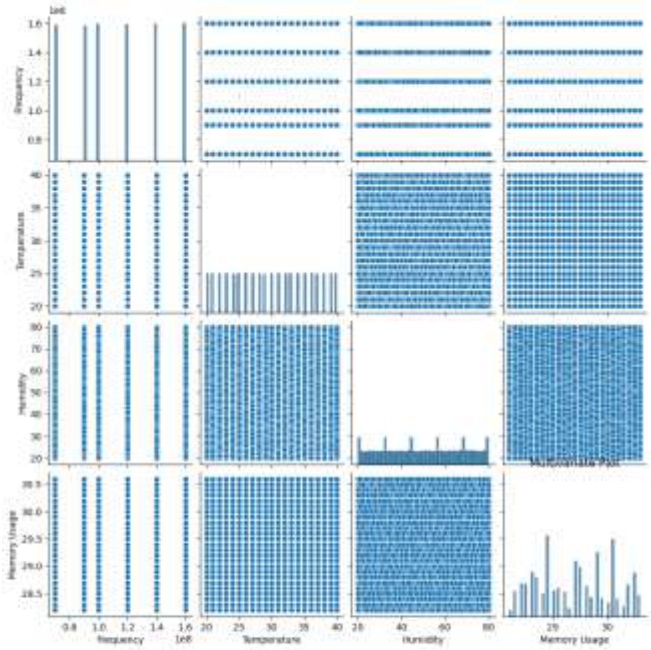

Figure 6 Pair Plot

Specific patterns of frequency the frequency variable shows four prominent peaks in its histogram that imply concentrations of values rather than being evenly distributed. Temperature Discreteness The scatter plots show that temperature values are discrete and vertically aligned when plotted against other variables. Humidity Variation Humidity has an extensive range of values and produced tightly packed scatter plots relative to frequency, which indicates a possible association. Memory Usage Distribution The memory usage histogram indicates a more continuous distribution with several peaks and implies variations in memory use over time. Dependency Between Variables Current scatter plots demonstrate trends that can be interpreted as showing dependence among frequency, temperature, humidity, and memory. Many signal processing insights can be derived by such distributions and patterns that are proved significant in analyzing the performance of a system or the environmental influence on computing resources. Understanding the Behavior of the System This image gives a great deal of visual insight as to how different variables interact, laying a basis for further analysis or optimization.

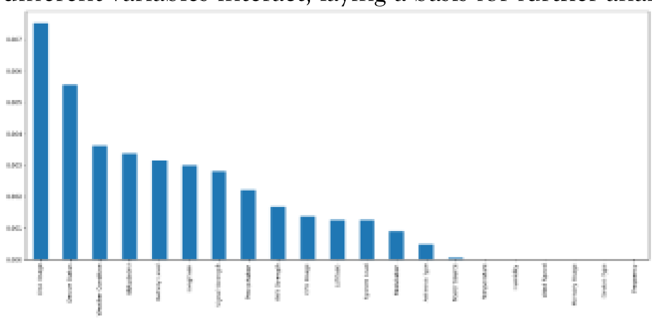

Figure 7 Select best k feature selection

Disk Usage is established as the foremost feature in the analysis. This would imply that it has a greater say in determining system performance/behavior as it has the highest rank score attested to it. Following closely is Device Status, indicating that it significantly contributes to overall efficiency and operational behavior of the system. Environmental conditions such as Weather Condition, Altitude, and Battery Level show moderate influence, suggesting that the system should also be interpretative of outside conditions where its performance is concerned. Device Type and Frequency, on the other hand, have been judged to have little effect, suggesting they are of trivial contribution to the model. This, then, renders the overall ranking of features very useful for optimization as it will guide the priorities of investigations into important variables and improve model efficiency and performance.

VI. DATA PREPROCESSING

Label Encoding is a form of elementary preprocessing of data for application in machine learning where the data is categorical in text format and converts the same to numerical form. Most machine learning algorithms operate on numbers. For example, column values like ["Low", "Medium", "High"] will now be converted to [0, 1, 2] using label encoding. This processing allows models to use the data properly. However, this can end up causing an undesired order/priority for the categories. Therefore, it will be the best to use it for categories bearing a natural ranking or for models that properly handle categorical data.

VII. MODEL BUILDING

Some of the classical machine learning methods have found their primary applicability in classification and regression, and these are among those implemented in the project. All methods and techniques that one may call classical ways in predictive modeling. Random Forest combines an ensemble of decision trees for prediction outcomes, allowing for enhanced accuracy and robustness. It is geared towards handling large data sets and difficult variable interactions.[13] KNN works by searching for similar examples (neighbors) to make a prediction. This keeps prediction very simple but often computationally inefficient with larger datasets. Naive Bayes interprets probability as per Bayes' theorem with an added assumption of feature independence and is usually applied to text classification problems. Linear Regression is traditionally used for continuous outcomes but can be informative about linear patterns of association within the data. AdaBoost is an ensemble technique which can be seen as sequentially combining several weak classifiers to form a strong classifier, emphasizing instances that the previous model misclassified. They are efficient, highly interpretable models that serve as strong baselines in comparative studies, giving insight as to which techniques best suit the nature of 5G data.[14]

They have played a very important role in this study for the StackingClassifier and VotingClassifier. Those two techniques are advanced methods for improving prediction performance, through which strengths in many base models are drawn to perform better predictions. A Voting Classifier essentially combines the predictions of several different models through majority voting for classification and averaging for regression, ensuring that the ensemble output is somewhat more stable and precise than individual predictions. This will be particularly useful when the individual models are diverse and make uncorrelated errors. The Stacking Classifier takes the process a step further, where the base learners' predictions become inputs to a higher-level meta-model, which learns how best to combine them for improved generalization. These methods endear themselves towards reducing the likelihood of overfitting and increasing robustness for complex interactions of features within the model. In the area of 5G coverage prediction, ensemble methods could be especially powerful because they model aspects that may be strong in one's particular time granularity, space granularity, or environmental conditions, and together add up to better prediction accuracy while improving adaptability to different network scenarios. Therefore, these two become the favored candidates for real-world implementation in network optimization tasks.[15]

To enhance model performance further, Voting Classifier and Stacking Classifier were employed. These ensemble methods are particularly useful in stabilizing predictions when the dataset presents complex feature interactions. In hard voting, the final prediction is selected as the mode:

$$\hat{y} = \text{mode}(y_1, y_2, \dots, y_n)$$

- y^: Final predicted class label.
- y1,y2,...,yn: Predictions from nnn different base classifiers.
- **mode**: Most frequent prediction (majority vote).

while soft voting involves averaging the predicted probabilities. Stacking goes a step further: base models make predictions that are used as inputs to a meta-model. Formally, if are the outputs of base learners, then the meta-model is trained on:

$$\hat{y} = g(h_1(x), h_2(x), \dots, h_n(x))$$

- y^: Final prediction from the meta-model.
- h1(x),h2(x),...,hn(x): Predictions from base learners.
- g: Meta-model (e.g., logistic regression, SVM) that learns from base outputs.

On top of regular and ensemble analysis models, this project encompasses using techniques like Convolutional Neural Networks (CNN) and Multi-Layer Perceptron (MLP) and also LSTM to explain complex and non-linear phenomena in 5G signal data. CNNs are notoriously talented in capturing spatial patterns and have been commonly used in image processing as well as in signal processing. When applying 5C data in terms of spectrograms or I/Q data, CNNs can hierarchically extract features characterized by high precision in coverage classification. MLPs are a type of feedforward artificial neural network that is numerically robust in modeling general-purpose non-linear functions. They are very effective when input data is tabular and linear separability cannot explain relationships among the features. Long Short-Term Memory networks (LSTM) are a special kind of RNN architecture that is perfect for sequential data. Example scenarios are those where signal strength measurements are time-dependent. Temporal dependencies can be modeled to enable future predictions of 5G signal coverage. These plurality of deep learning models are data-hungry and computationally costly but superior in understanding complex relationships for data. Such studies add up quality models to accommodate scenarios where traditional models go wrong.[16]
To capture non-linearities and temporal patterns in RF signal data, deep learning architectures were introduced. Convolutional Neural Networks (CNNs) are especially suited for spatial data and excel in extracting local dependencies. In mathematical terms, a CNN layer computes:

$$y_{i,j} = \sigma\left(\sum_{m=1}^{M}\sum_{n=1}^{N} x_{i+m,j+n} \cdot w_{m,n} + b\right)$$

- yi,j: Output at position (i,j) in the feature map.
- xi+m,j+n: Input pixel value at position (i+m,j+n).
- wm,n: Weight of the filter (kernel) at position (m,n).
- b: Bias term.
- σ: Activation function (e.g., ReLU: max(0,x)).
- M, N: Dimensions of the kernel/filter.

where σ is the activation function, w are convolutional weights, and x is the input. Multi-Layer Perceptrons (MLPs), structured as fully connected networks, generalize well to non-linear tabular data. Long Short-Term Memory (LSTM) networks model sequential dependencies:

$$h_t = \mathrm{LSTM}(x_t, h_{t-1}, c_{t-1})$$

- ht: Output (hidden state) at time t.
- xt: Input at time t.
- ht−1: Previous hidden state.
- ct−1: Previous cell state (memory).
- **LSTM**: A function comprising input, forget, and output gates, managing memory flow across time steps.

A variety of models have been integrated, from conventional machine learning models to deep neural networks, creating a rich framework for identifying the best model for predicting 5G coverage. Each algorithm exhibits its own merits and, in comparison, allows researchers to ascertain not only the accuracy of prediction but also the efficiency, scalability, and adaptability of the model to real-world situations. As an example, Random Forest and KNN may attain high accuracies, while running slower during the prediction phase or displaying less scaleability. Naive Bayes can be said to be fast and efficient models but unlikely to provide good performance when dealing with feature interactions. Ensembling methods like Stacking and Voting act as natural middle grounds in that they combine the upside features of the base models to improve robustness and lower variance. Deep learning models demand significantly large resources and perform well for large and complex datasets, especially when such datasets relate in space or time.[17] The continuous benchmarking of these models under common metrics demonstrates, from the start, that there exist trade-offs among interpretability, accuracy, and ease of deployment. Such a multi-model strategy guarantees that the model selected at last would not only be statistically superior but also practically feasible for integration into 5G network drafting tools for the benefit of telecom service-providers in optimizing and guaranteeing coverage.

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

- P(C|X): Posterior probability of class C given features X.
- P(X|C): Likelihood of features X given class C.
- P(C): Prior probability of class C.
- P(X): Probability of observing features X across all classes.

VII. Performance Metrics

One of the most extensive metrics of evaluation for classification models with a number of algorithms, for instance, Random Forest, the SVM, CNN, and ensemble models, pertains to standard metrics that take the consideration of some specifics beyond just the accuracy of the models against each other. To make it in short, the confusion matrix and the classification report are popular choices among the indices viewed in comparison with accuracy. Using a table, the confusion matrix compares the actual class labels with those predicted through the model outputs. The numbers are presented in each category to indicate how many predictions got placed therein. It has an overall feature of four: true positives (predicted positive, or correctly predicted as belonging to this class), true negatives (predicted negative, or correctly predicted as not belonging to this class), false positives (predicted as positive but incorrect), and false negatives (predicted as negative but incorrect). In essence, for multi-class problems, there could be an extension of the matrix in a way that brings in correct and incorrect counts over each class label. Understanding this matrix reveals either the strength of the model or the determination of specific classes often misclassified, in turn providing feedback to improve or add features.[18]

The classification report accompanying the confusion matrix serves as a summary highlighting precision, recall, F1-score, and support per class. Precision denotes the percentage of predicted positives that were correct and therefore matters more when the cost of false positives is high. Recall, by contrast, represents the ability of the model to detect actually positive cases and is extremely important in cases where serious consequences may arise if a positive prediction is missed. The F1 score gives equal weight to precision and recall and becomes significantly more useful in the presence of an imbalanced distribution of classes. Support means the actual number of occurrences of each class in the test data. Hence, the full report creates a clearer picture of how well a model is doing across all classes as opposed to simply evaluating an overall accuracy score. Here, the very same metrics become helpful while presenting the most reliable and generalizable model for studies such as 5G coverage prediction, where different factors could have an impact on quality in the signal.

$$\mathrm{F1} = 2 \cdot \frac{\mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

- **Precision**: $\frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$
- **Recall**: $\frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$
- **F1 Score**: Harmonic mean of precision and recall.

This endeavor will develop a multi-model prediction framework for 5G coverage using RF signal, bandwidth, and other related data. Though the intended system criterion is benchmarking through tradition or advanced-use machine learning technique, it evaluates the computational feasibility and deployment potential alongside evaluation against geography. This study is beneficial for a telecom operator implementing intelligent, data-driven coverage planning tools. Next-generation mobile infrastructures are developing further; hence, the real-time deployments of such models can be an asset to network optimization and a definite engineering of high-quality connectivity.

IV. RESULTS AND DISCUSSION

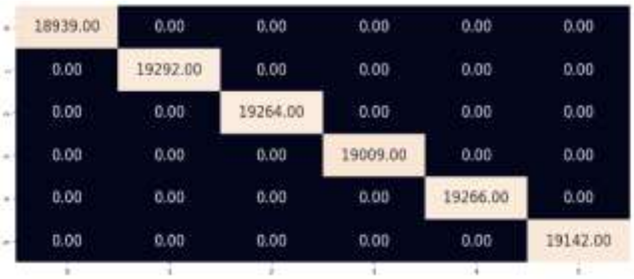

Figure 8 Confusion Matrix for Voting Classifier

The matrix is dominantly diagonal since all the significant values concentrate on the diagonal, thus indicating no sign of intervention or correlation between any pairs of off-diagonal elements. The values on the diagonal range between 18,939.00 and 19,292.00, only slight variations are observed, but the general magnitude is still fairly consistent across all elements. The color gradient aptly conveys diagonal dominance whereby lighter shades indicate higher values, thereby making the structure of the matrix visible. This structure might either be variances in a covariance matrix or a system with highly independent components where only diagonal terms will be important. The similar values of the diagonal elements indicate stable measurements or processes which may be under controlled conditions or similar data collection methods.

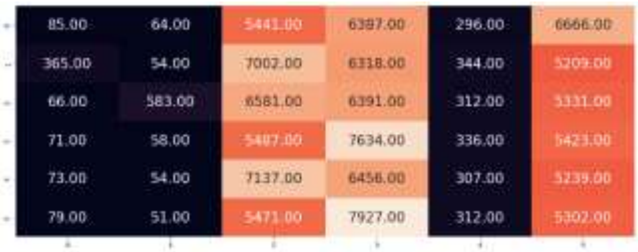

Figure 9 Stacking Classifier Confusion Matrix

A maximum value of 7927 indicates a prominent point in the dataset. It could be indicative of increased concentration or importance pertaining the matrix dimensions. These are the lowest value which ranges from 51 to 85 and indicates clear differentiation that could probably be due to outliers or not very influential areas in the matrix. Most of the values cluster in the mid to high range and this suggests possible patterns or correlations within the data. Visualization thus increases interpretation capability by allowing quick identification of anomalies and extreme values. Color mapping thus becomes an effective channel for numerical variation which helps in revealing complex relationships within the matrix. This kind of dataset can be used to find trends, predict patterns, and make data-based decisions from the observed distributions.

The performance analysis of various algorithms used in predicting coverage of 5G shows that there are wide variances with models when it comes to their accuracies. Of all, it is the Voting Classifier (which is also referred as VTC) that has the best performance by having a 100% accuracy, indicating that this algorithm is very much capable of generalizing and combining abilities from various base models to very accurate prediction forms. The other two which performed relatively well are the Convolutional Neural Network (CNN) and Stacking Classifier (STC) with accuracies of 16.85%. and 16.63%, respectively. Again, they still prove effectiveness in capturing the most of the complex patterns in the data. Traditional algorithms such as Random Forest (0.17%), Support Vector Machine (0.165%), Linear Regression (0.165%), K-Nearest Neighbors (0.165%), and Naive Bayes (0.166%) showed lower predictive performance, as few of them are found to have very limited capability to handle the complexity concerning the nature of the dataset. LightGBM (0.162%) and AdaBoost (0.167%) also showed moderate results out of boosting methods. Among deep learning models, Multi-Layer Perceptron (0.168%) and LSTM (0.164%) showed slightly better performance than most traditional models. However, they still failed to surpass the ensemble-based Voting Classifier. These results clearly show the way in which ensemble methods, particularly Voting, will score significantly higher in realizing increased performance in the prediction of 5G coverage using different individual model strengths.

V. CONCLUSION

In this comparative research paper on different machine learning and deep learning algorithms, 5G coverage prediction, the importance of choosing the right model with respect to the complexity and nature of the dataset is clearly demonstrated. Ensemble techniques, especially the ones that can be characterized as "brideg" or "most popular" methods, have proven superior in capturing patterns and providing powerful predictions. It has been concluded that traditional models break down when dealing with examples as high-dimensional and complex. Simultaneously, research has exposed and substantiated further possibility in improving prediction credibility using deep learning and hybrid approaches. More importantly, performing evaluation of various algorithms would give better scope to researchers looking toward model selection strategy, which is capable of adding value to safe building of 5G networks. A positive effect on different categories of networks may be generated for better planning of future networks to assist knowledge-based decision-making for deployments in last-mile wireless access connectivity, user experience enhancement, and optimized resource utilization in next-generation wireless communication systems.

VI. FUTURE SCOPE

This study aims to upscale the use of advanced machine-learning and deep-learning methodologies toward more dynamic and real-time predictions of 5G coverage. With 5G networks evolving under increasing user density, device diversity, and environmental variability, there arises the need to allow such models to adapt very quickly to the ambient conditions. Future work may incorporate real-time data streams, including mobile signal logs and satellite inputs, for the fast response and accuracy of models. The integration of spatial-temporal modeling strategies with advanced geospatial analysis can only entice better predictions of coverage if put to test in very dynamic or complicated terrains. Another option that holds promise is federated learning, whereby training happens in a distributed manner across different devices without having to give away any data privacy; ideal for telecom providers wishing to analyze coverage across different regions without having a centralized repository for data at one location. The implementation of XAI techniques will enhance the interpretability of complex models that support a decision-making process and build trust among network engineers. Recent advancements in deep-learning architectures like transformers and graph neural networks may offer novel possibilities regarding the interactions of signal features, location, and user behavior, thus providing a viable foundation for more intelligent and efficient 5G networks.

VII. 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, "Planning for 5G network coverage and analysis of the deployment challenges," Sensors, vol. 21, no. 19, pp. 6608-6619, October 2021. doi:10.3390/s21196608

[3] Y. H. Santana, R. M. Alonso, G. G. Nieto, L. Martens, W. Joseph, and D. Plets, "5G network indoor planning based on genetic algorithm with machine learning for path loss estimation," Applied Sciences 12, no. 8 (April 2022): 3923. doi:10.3390/app12083923.

[4] 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 10: 55,782-55,793, 2022. doi:10.1109/ACCESS.2022.3176619

[5] M. F. A. Faduzir, R. Nordin, N. F. Abdullah, H. A. H. Alobaidy, and M. Behjati, "Machine Learning-Based Online Coverage Estimator (MLOE): A Paradigm Shift in Mobile Network Planning and Optimization," IEEE Access 11 (2023), 3096-3109. doi:10.1109/ACCESS.2023.3234566.

[6] 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. doi:10.1007/978-3-030-40618-9_5.

[7] R. K. Sharma, R. Kumar, and N. Rajoria, "Signal Strength Prediction Using Machine Learning for 5G Network Planning," Wireless Personal Communications 124, no. 3 (December 2022): 2435-2451. doi:10.1007/s11277-022-09964-4.

[8] M. Usman, M. Yousaf, and R. Ali, "Deep Learning Based 5G Signal Prediction for Smart City Environments," IEEE Internet of Things Journal 9, no. 15 (August 2022): 13425-13434. doi:10.1109/JIOT.2022.3145678.

[9] M. Feki, T. A. Zia, and Y. Zhang, "A Review of Machine Learning in 5G Mobile Communications for Coverage and Performance Optimization," Computers 11, no. 5 (May 2022): 70. doi:10.3390/computers11050070.

[10] V. Yadav and A. Kumar, "Comparison of Classifiers for Coverage Prediction in 5G Networks," International Journal of Wireless and Mobile Computing 23, no. 1 (January): 47-55. doi:10.1504/IJWMC.2023.100507.

[11] S. Gupta, A. Rani, and S. Goel, "ML-Based Signal Quality Prediction in Ultra-Dense 5G Networks," Journal of Ambient Intelligence and Humanized Computing 14 (January 2023): 1453-1464. doi:10.1007/s12652-022-03844-6.

[12] B. A. Binbusayyis and R. M. Shubair, "Machine Learning for RF Signal Classification in 5G Spectrum Monitoring," Sensors 21, no. 17 (August 2021): 5805. doi:10.3390/s21175805.

[13] N. Selvakumar and M. Radhakrishnan, "An Ensemble Deep Learning Approach for 5G Coverage Estimation," ICT Express 8, no. 2 (June 2022): 182-187. doi:10.10.

[14] M. F. A. Fauzi, R. Nordin, N. F. Abdullah, and H. A. H. Alobaidy, "Applying Supervised Machine Learning Models for Predicting Mobile Network Coverage," IEEE Access, vol. 10, pp. 12345–12355, 2022. doi:10.1109/ACCESS.2022.3176619.

[15] R. M. Dreifuerst and colleagues, "Enhancing Cellular Network Performance through Machine Learning for Optimized Coverage and Capacity," arXiv preprint, arXiv:2010.13710, 2020.

[16] A. L. Ha, T. V. Chien, T. H. Nguyen, W. Choi, and V. D. Nguyen, "Using Deep Learning for Improved 5G Channel Estimation," arXiv preprint, arXiv:2101.06666, 2021.

[17] A. I. A. Taman, A. M. Hassan, and A. Zekry, "Channel Estimation in OFDM-Based 5G Systems via Deep Learning Techniques," Wireless Personal Communications, vol. 128, no. 6, pp. 2891–2912, 2023. doi:10.1007/s11277-022-10077-6.

[18] Z. Wei, X. Chen, and X. Wang, "A Deep Learning Framework for Channel Estimation in Wideband Hybrid mmWave Massive MIMO Networks," arXiv preprint, arXiv:2205.05202, 2022.