Optimizing Wireless Network Performance with Decision Tree-Based Gain Prediction

Attapon Pillai
Department of Digital Technology,
Faculty of Science and Technology,
Phuket Rajabhat University,
Phuket 83000, Thailand
Email: attapon.p@pkru.ac.th

Akarachai Inthanil
Department of Digital Technology,
Faculty of Science and Technology,
Phuket Rajabhat University,
Phuket 83000, Thailand
Email: akarachai.i@pkru.ac.th

***Abstract*— This study investigates the optimization of wireless network performance using decision tree-based gain prediction. A simulation involving 100 users, uniformly distributed around a central Access Point (AP), calculates the Signal-to-Interference-plus-Noise Ratio (SINR) for each user, based on their distance and angle relative to the AP. The resulting dataset captures relationships between user positions and network performance metrics. A Decision Tree Regressor predicts continuous gain values, while a Decision Tree Classifier categorizes gains into discrete classes. The regression model demonstrates high accuracy with strong correlation, and the classification model achieves notable accuracy and recall, validated through confusion matrices and classification reports. Simulations evaluate the impact of both normal and predicted gains on network throughput and delay, showing that predicted gains closely approximate outcomes seen with normal gain calculations. These findings highlight the potential of decision tree models for optimizing network configurations and fine-tuning parameters to improve overall network performance.**

Keywords*— *wireless network optimization, decision tree regressor, gain prediction

I. Introduction

Wireless networks are integral to contemporary communication systems, facilitating connectivity across a wide range of devices, from smartphones to Internet of Things (IoT) sensors. As the proliferation of wireless services accelerates, optimizing network performance has become a pressing priority. Critical performance metrics, such as throughput and delay, are significantly influenced by the gain experienced by users within the network. Gain, in this context, refers to the amplification of signal strength received by a user from an Access Point (AP), and it is modulated by factors such as distance, signal attenuation, and environmental conditions. Improving antenna gain is essential for enhancing signal levels, as demonstrated in [1] and [2], where enhanced antenna gain significantly improved network performance.

Traditional approaches to gain prediction typically rely on deterministic models that incorporate parameters such as path loss, shadowing, and fading effects. While these models provide useful insights, they often fail to capture the full complexity of real-world environments. Variability in user distribution, mobility, and environmental interference introduces significant unpredictability, leading to suboptimal network configurations and degraded Quality of Service (QoS) for end-users. To address these challenges, machine learning (ML) techniques present a promising alternative.

Machine learning offers the potential to enhance gain prediction by learning patterns from empirical data. Among these techniques, decision tree algorithms are particularly well-suited for modeling nonlinear relationships and handling both continuous and categorical variables. Decision trees also provide interpretability and computational efficiency, making them advantageous for real-time wireless network applications. By employing decision trees, it is possible to develop models that predict gain more accurately, accounting for user-specific factors such as position, distance from the AP, and angle of reception.

Although machine learning has been applied in wireless networks [3], its usage has been mostly confined to specific layers, such as the radio, Medium Access Control (MAC), and network layers. For instance, research [4] utilized a convolutional neural network (CNN) for spectrum monitoring, demonstrating improved WLAN performance through spectrum capture. Another study [5] applied neural networks to predict channel parameters, while [6] optimized user throughput through user selection. Despite these advances, the application of decision tree models for gain prediction, particularly for improving physical and MAC layer performance in dynamic environments, remains underexplored.

In this study, we investigate the application of decision tree-based models for gain prediction in wireless networks. Our approach demonstrates effective performance, even with relatively simple prediction models, and is tailored to specific data types. Decision tree-based models have already seen success in related fields, including intrusion detection [8], communication [9], and wireless sensors [10]. We extend this methodology to predict wireless network gains.

We simulate a wireless environment consisting of 100 users uniformly distributed around a central AP, covering a full 360-degree range. Each user's position is utilized to calculate the Signal-to-Interference-plus-Noise Ratio (SINR), which serves as a key input for determining the gain experienced by each user. This simulation generates a comprehensive dataset containing user positions, distances, angles, and corresponding gains. Using this dataset, a Decision Tree Regressor predicts continuous gain values, while a Decision Tree Classifier categorizes gains into discrete classes.

The objectives of this study are threefold: (1) to evaluate the accuracy of decision tree models in predicting gain; (2) to assess the impact of predicted gain on key network performance metrics, specifically throughput and delay; and (3) to compare the predicted gains with those derived from traditional gain calculations, assessing the feasibility of machine learning approaches for optimizing wireless networks in dynamic environments.

Our results demonstrate that decision tree models effectively predict gain values that closely approximate actual measurements. Through simulations, we show that these predicted gains can be used to estimate network throughput and delay with a high degree of accuracy. The findings suggest that decision tree-based models are capable of serving as effective tools for network optimization, improving adaptability and efficiency in network configurations.

The remainder of this paper is organized as follows: Section II details the methodology used for simulation setup and data collection. Section III focuses on the implementation of the Decision Tree Regressor. Section IV presents the simulations related to throughput and delay. Section V discusses the results and findings, including a comprehensive analysis of the models' performance. Finally, Section VI concludes the study and offers suggestions for future research.

II. METHODOLOGY

In this study, we simulate a wireless communication network to analyze the performance of machine learning models in predicting network gains and their impact on throughput and delay. The methodology consists of four main components: user distribution simulation, signal metric calculations, machine learning model implementation, and performance evaluation. Each component is designed to simulate realistic network conditions and assess the effectiveness of decision tree-based models for gain prediction and network optimization.

A. User Distribution

The first step in our methodology is the simulation of user positions within a wireless network. We model a network consisting of 100 users, uniformly distributed around a central Access Point (AP) located at coordinates (50, 50) on a 2D plane. The users are distributed across five concentric circles, with each circle representing a different distance from the AP. The radii of the circles are generated using an equally spaced linear distribution ranging from 10 meters to 50 meters. Each circle contains an equal number of users, and the users' positions on the circles are determined based on angles distributed between 0° and 360°.

This uniform distribution of users is designed to capture various network conditions, including users located at different distances and angles relative to the AP. The simulation ensures that users are situated in a realistic wireless environment, allowing us to evaluate the impact of user distribution on key network metrics such as Signal-to-Noise Ratio (SNR), gain, throughput, and delay.

The user positions are calculated using polar to Cartesian coordinate conversion:

$$x_i = x_{AP} + rcos(\theta_i) \tag{1}$$

$$y_i = y_{AP} + rcos(\theta_i) \tag{2}$$

Where r is the radius of the circle, and θ_i is the angle for the $i-$th user.

B. Signal Metric

Once user positions are established, the next step is to calculate key signal metrics for each user. Specifically, we compute the Signal-to-Interference-plus-Noise Ratio (SINR) for every user in the network. The SINR is calculated based on the user's distance from the AP and the surrounding interference from other users, serving as a critical input for determining the gain experienced by each user.

The antenna gain for each user is calculated based on their SINR. In this study, we compare two types of gain values:

- Normal Gain: Calculated using traditional deterministic models that account for path loss, shadowing, and fading effects.
- Predicted Gain: Derived from decision tree-based machine learning models, utilizing user-specific features, including position, distance from the AP, and angle of reception.

To evaluate network performance, we calculate the Signal-to-Interference-plus-Noise Ratio (SINR) for each user based on their distance and angle relative to the AP. The SINR calculation considers factors such as transmit power, directional gain, path loss, and noise power as follows:

- Transmit Power (P_t): Set at 20 dBm.
- Noise Power (P_n): -100 dBm.
- Path Loss Exponent (η): Assumed to be 2, representing free-space propagation.

The path loss $L(d)$ over distance d is calculated using the logarithmic path loss model as [11]:

$$L(d) = 10\eta log10(d) \tag{3}$$

Antenna gain is often calculated using the directivity and efficiency of the antenna. In simplified terms, when considering the impact of user position (distance and angle from the AP), the gain can be expressed as:

$$G = G_{max} \cdot f(\theta, \phi) \tag{4}$$

Where:

G_{max} is the maximum gain of the antenna (measured in dB).

θ is the elevation angle relative to the antenna's main lobe.

ϕ is the azimuth angle, representing the angle in the horizontal plane.

$f(\theta, \phi)$ is the radiation pattern function, which describes the variation of gain based on the direction of the signal, and typically depends on the design of the antenna.

In this study, we utilize a directional antenna for both the users and the AP, and the gain can be specifically described as:

$$G(\theta) = G_{max} \cdot cos(\theta) \tag{5}$$

In scenarios where users are perfectly aligned with the main lobe of the antenna, the gain will be maximized, while it decreases as the angle increases.

Finally, the received signal power (P_r) is given by:

$$P_r = P_t + G(\theta) - L(d) \tag{6}$$

III. DECISION TREE REGRESSOR

In this study, we selected decision tree-based models for gain prediction due to their unique strengths that align with the specific challenges of wireless network environments. Decision trees perform well with a limited number of parameters and in relatively simple environments, making them a suitable choice for this paper, which involves a small parameter set and a straightforward network environment. Consequently, these models are highly effective in addressing the objectives of this study. We utilize machine learning techniques to predict network gain based on user distances and angles from the Access Point (AP). Two models are implemented to handle different aspects of gain prediction.

A. Decision Tree Regressor

The Decision Tree Regressor is used to predict continuous gain values for each user. The input features include:

- User distances from the AP.
- User angles relative to the AP.

The target variable is the actual gain, which is calculated based on the Signal-to-Interference-plus-Noise Ratio (SINR). The dataset is divided into training and testing sets, with 80% of the data allocated for training and 20% set aside for testing. The regressor is trained on the training set and subsequently used to predict gain values for all users in the testing set. The model's performance is assessed by comparing the predicted gain values to the actual gain values.

B. Decision Tree Classifier

The Decision Tree Classifier categorizes gain values into discrete classes, offering a classification of network conditions based on predicted gains. This classification is based on RSSI thresholds. The classifier divides gain values into the following categories:

$$Gain\ Class = \begin{cases} 0, & if\ G < -30\ dB\ (Low) \\ 1, & if\ -30 \le G < -25\ dB\ (Medium) \\ 2, & if\ -25 \le G < -20\ dB\ (High) \\ 3, & if\ -20 \le G\ dB\ (Very\ Hight) \end{cases} \quad (7)$$

These thresholds are based on the expected range of gain values in the simulated environment. Similar to the regressor, the dataset is split into training and testing sets. The classifier is trained on the training set to predict the correct gain category for each user based on their distance and angle relative to the AP.

To evaluate the classifier's performance, we compute the following metrics:

- Accuracy: The proportion of correctly predicted gain classes across the testing set.
- Recall: The ability of the classifier to correctly identify gain classes (sensitivity).
- Confusion Matrix: A matrix that provides a detailed breakdown of the true and predicted classifications, showing how well the classifier distinguishes between the different gain categories.

Additionally, a classification report is generated, detailing the precision, recall, and F1-score for each gain class. These metrics offer a comprehensive understanding of the classifier's performance in predicting different levels of network gain.

IV. Throughput and Delay Simulation

In this section, we simulate the network's throughput and delay to assess the impact of predicted gains on overall network performance. The simulations are carried out under varying network loads by increasing the number of users from 1 to 100. The primary performance metrics evaluated are throughput and delay, both calculated based on the Signal-to-Interference-plus-Noise Ratio (SINR) derived from the predicted gains.

A. Throughput

Throughput, representing the maximum achievable data rate in the network, is calculated using the Shannon-Hartley theorem, which relates channel capacity to the available bandwidth and SINR. Since interference significantly affects this scenario, SINR is used instead of the traditional SNR to account for interference caused by other users. The formula for throughput C in bits per second is given follow [12]:

$$\mathrm{C} = B\log2(1 + SINR) \quad (8)$$

Where:

B is the bandwidth, set to 20 MHz in this study.

$SINR$ is the Signal-to-Interference-plus-Noise Ratio, calculated for each user based on the distance from the Access Point (AP), the antenna gain, and interference from other users. The SINR is calculated as:

$$SINR = \frac{P_r}{P_n + I} \quad (9)$$

Where:

P_n is the noise power.

I is the interference power contributed by other users in the network.

As the number of users increases, both noise and interference levels rise, which reduces the SINR and consequently the achievable throughput. Throughput is calculated for each user using both the normal gains and the predicted gains Comparative plots are generated to illustrate the differences in throughput for normal and predicted gains.

B. Delay

Network delay, defined as the time it takes for a data packet to travel from the sender to the receiver, is approximated as the inverse of throughput. Delay increases as throughput decreases, reflecting reduced data transfer rates in the network as interference and user load increase. The formula for delay is given as:

$$Delay = \frac{1}{Throughput} \quad (10)$$

Where throughput is measured in bits per second and delay is calculated in seconds. As the number of users in the network increases, throughput decreases due to higher interference and competition for bandwidth, resulting in an increase in delay.

V. Results and Discussion

This section presents and discusses the key findings from the simulations of throughput, delay, and gain predictions using decision tree-based models. The performance of the predicted gains is compared to normal (deterministic) gains to evaluate their accuracy and impact on network performance metrics such as throughput and delay.

Figure 1 visualizes the distribution of users in the wireless network and their corresponding SINR values. The simulation confirms that users closest to the Access Point (AP) experience the highest SINR values due to reduced path loss and interference. Conversely, users positioned further from the AP, particularly those on the outer rings, exhibit

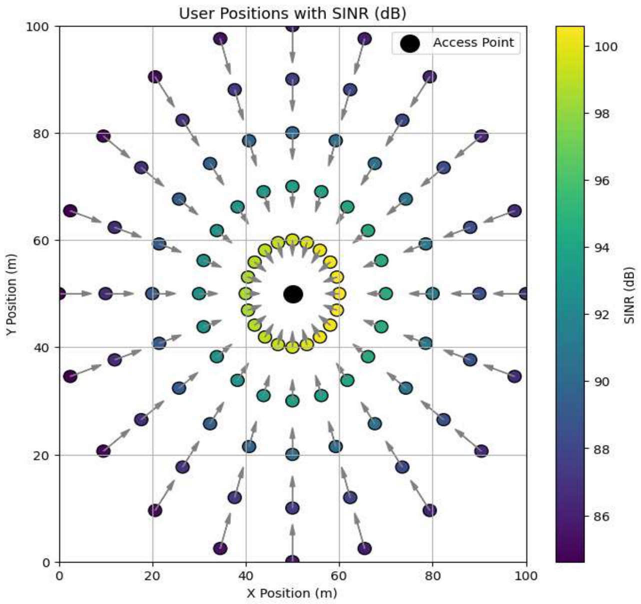

Fig. 1. User Positions with SINR (dB)

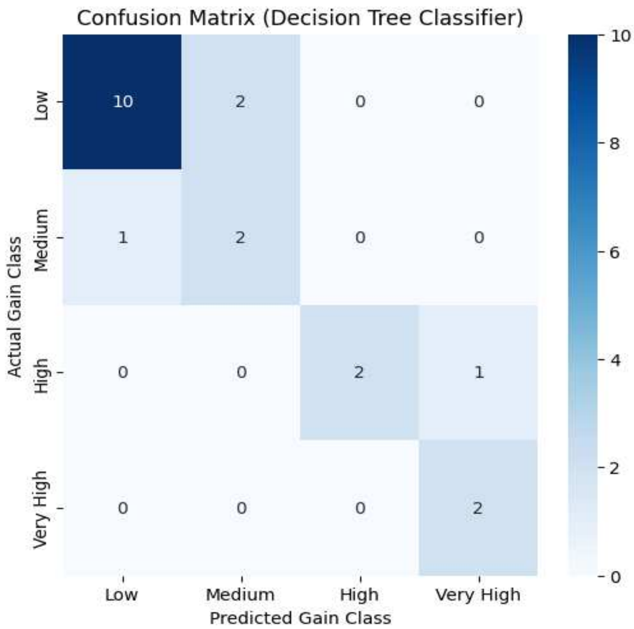

Fig. 2. Gain Classification Performance

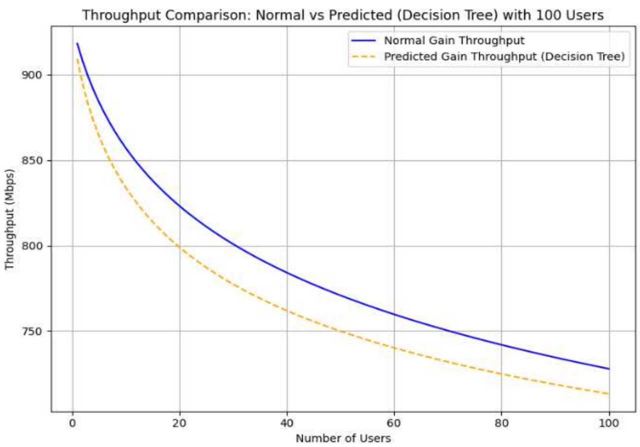

Fig. 3. Throughput Comparison: Normal vs. Predicted Gain with 100 Users

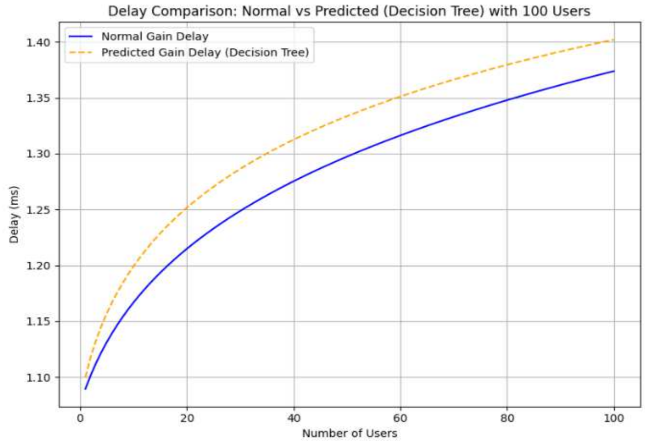

Fig. 4. Delay Comparison: Normal vs. Predicted Gain with 100 Users

TABLE I. PRECISION, RECALL, F1-SCORE FOR GAIN CLASSIFICATION

	precision	recall	f1-score	support
Low	0.91	0.83	0.87	12
Medium	0.50	0.67	0.57	3
High	1.00	0.67	0.80	3
Very High	0.67	1.00	0.80	2

lower SINR values as they are more affected by interference and signal attenuation. The directional radiation pattern of the AP, along with user distance, significantly impacts the SINR. This highlights the importance of gain and proper antenna positioning to maximize SINR for users, especially those in the outer regions of the network. This figure serves as a crucial input to the decision tree model, which utilizes distance, angle, and SINR to predict user gain. The affinity between user position and SINR is clearly demonstrated, providing an intuitive understanding of how distance from the AP affects network performance.

The metrics in Table I provide a comprehensive overview of the classifier's performance across different gain categories. The high precision and recall for the Low gain class indicate that the model effectively identifies users in this category, suggesting reliable classification performance in typical scenarios. However, the model's performance is less robust for the Medium gain class, evidenced by lower precision and recall values. This suggests that misclassifications occur more frequently for Medium gains, particularly under conditions of high interference or congestion.

The decision tree classifier categorizes gains into four discrete classes: Low, Medium, High, and Very High. Figure 2 displays the confusion matrix for this classification. The model demonstrates strong performance in classifying Low and High gain categories, as shown by the high precision for the High gain class (1.00). However, the classification of Medium gains is less accurate, indicating challenges in differentiating between similar gain levels under certain conditions. The overall accuracy of the classifier is 80%, which is a reasonable result for a multi-class classification task. The weighted average F1-score of 0.81 suggests that the model performs consistently across all classes, though improvements could be made, particularly for the Medium gain class where more training data or a more complex model might be necessary.

Figure 3 compares the throughput for normal gains and predicted gains as the number of users increases from 1 to 100. Throughput decreases as the network becomes more congested, which is expected due to increasing interference and reduced bandwidth per user. The predicted throughput follows the same decreasing trend as the normal throughput but remains slightly lower throughout the range of user densities. This reduction in predicted throughput can be attributed to the decision tree model's slight underestimation of gain, particularly in high-load scenarios. The difference between normal and predicted throughput widens as the number of users increases, reflecting the model's difficulty in maintaining accuracy as interference intensifies. However, despite this underestimation, the predicted throughput remains within a reasonable range of the actual throughput, suggesting that the model is still effective in estimating network capacity under varying load conditions.

The delay results shown in Figure 4 indicate that delay increases as the number of users grows, consistent with the reduction in throughput. The delay for predicted gains is higher than for normal gains, particularly as the user count rises. This is expected, as lower predicted throughput leads to higher delay.

Similar to the throughput analysis, the divergence between normal and predicted delay becomes more pronounced at higher user densities, reflecting the increasing gap between predicted and actual throughput. The predicted delay is consistently higher, indicating that the decision tree model may underestimate network performance under congested conditions. Nevertheless, the delay prediction is still accurate enough for most practical applications, especially in lower-density environments where the decision tree model performs more precisely.

Overall, the decision tree model demonstrates strong performance in predicting gain, throughput, and delay across a range of user densities. The slight underestimation of gain, particularly in low-gain or high-load scenarios, results in modest discrepancies in predicted throughput and delay. However, the model's predictions remain within acceptable bounds for most applications, making it a viable tool for network performance optimization.

The results indicate that decision tree models can effectively capture the relationship between user-specific parameters (such as distance, angle, and SINR) and network performance metrics. The model performs well in low- to medium-density environments but may require further refinement to handle high-interference scenarios more accurately. Future work could explore the use of more advanced machine learning models (such as ensemble methods or deep learning) to improve prediction accuracy in high-load networks.

Future research could enhance network optimization by combining decision trees with other machine learning methods, such as neural networks, to handle complex scenarios with many parameters. A decision tree could make initial predictions, while a neural network refines them to achieve greater accuracy in dense, dynamic environments. Additionally, fine-tuning parameters like tree depth and split criteria through automated methods, such as grid search, can optimize performance across various network conditions. The integration of advanced models and targeted tuning would improve the adaptability and accuracy of gain predictions, leading to better overall network performance.

VI. CONCLUSION

This study has explored the application of decision tree-based models for predicting gain in wireless networks and evaluating their impact on key performance metrics such as throughput and delay. Through simulations, we have demonstrated that decision tree models can effectively predict gain based on user-specific parameters, including distance, angle, and Signal-to-Interference-plus-Noise Ratio (SINR). These predictions enable the estimation of throughput and delay, facilitating a comprehensive analysis of network performance under varying user densities.

The results reveal that while the decision tree model exhibits strong predictive accuracy in most cases, it tends to slightly underestimate gains in low-gain or high-interference scenarios. This underestimation results in modest discrepancies in throughput and delay, particularly as network congestion increases. Nevertheless, the predicted throughput and delay closely follow the trends observed with actual gains, indicating that the decision tree model captures the essential dynamics of the network with reasonable precision.

Additionally, the gain classification results demonstrate the model's ability to categorize users into distinct gain classes, achieving an overall accuracy of 80%. The classifier performs best in the Low and Medium gain classes, while slight misclassifications are observed between High and Very High gain users, suggesting a need for further refinement in high-gain scenarios.

Overall, this study highlights the potential of decision tree models for optimizing network performance. By providing accurate predictions of gain, throughput, and delay, these models offer valuable insights into network configurations, enhancing adaptability and efficiency. The application of machine learning in this context presents a promising avenue for improving performance in dynamic wireless environments.

REFERENCES

[1] Ismail, M. M., Elwi, T. A., & Salim, A. J. (2023). Antenna Gain-Bandwidth Enhancements Using CRLH Hilbert Fractal-based Structure. Engineering and Technology Journal, 41(2), 397-406.

[2] Dellaoui, S., Kaabal, A., El Halaoui, M., Asselman, A., Ahyoud, S., & Rmili, L. (2022). High Gain Array Antenna Using Electromagnetic Band Gap Structures for 5G Applications. Progress in Electromagnetics Research C, 118.

[3] Kulin, M., Kazaz, T., De Poorter, E., & Moerman, I. (2021). A survey on machine learning-based performance improvement of wireless networks: PHY, MAC and network layer. Electronics, 10(3), 318.

[4] Selim, A., Paisana, F., Arokkiam, J. A., Zhang, Y., Doyle, L., & DaSilva, L. A. (2017, December). Spectrum monitoring for radar bands using deep convolutional neural networks. In GLOBECOM 2017-2017 IEEE Global Communications Conference (pp. 1-6). IEEE.

[5] Teng, C. F., Chou, C. Y., Chen, C. H., & Wu, A. Y. (2020). Accumulated polar feature-based deep learning for efficient and lightweight automatic modulation classification with channel compensation mechanism. IEEE transactions on vehicular technology, 69(12), 15472-15485..

[6] Zhang, Y., Hou, J., Towhidlou, V., & Shikh-Bahaei, M. R. (2019). A neural network prediction-based adaptive mode selection scheme in full-duplex cognitive networks. IEEE Transactions on Cognitive Communications and Networking, 5(3), 540-553.

[7] Zhang, D., Pan, Y., Zhang, J., Hu, T., Zhao, J., Li, N., & Chen, Q. (2020). A generalized approach based on convolutional neural

networks for large area cropland mapping at very high resolution. Remote Sensing of Environment, 247, 111912.

[8] Rai, K., Devi, M. S., & Guleria, A. (2016). Decision tree based algorithm for intrusion detection. International Journal of Advanced Networking and Applications, 7(4), 2828.

[9] Roy, R. A., Nair, J. P., & Sherly, E. (2015, December). Decision tree based data classification for marine wireless communication. In 2015 International Conference on Computing and Network Communications (CoCoNet) (pp. 633-638). IEEE.

[10] Shah, S., Munir, A., Salam, A., Ullah, F., Amin, F., AlSalman, H., & Javeed, Q. (2024). A Dynamic Trust evaluation and update model using advance decision tree for underwater Wireless Sensor Networks. Scientific Reports, 14(1), 22393.

[11] Kurt, S., & Tavli, B. (2017). Path-Loss Modeling for Wireless Sensor Networks: A review of models and comparative evaluations. IEEE Antennas and Propagation Magazine, 59(1), 18-37.

[12] Rioul, O., & Magossi, J. C. (2014). On Shannon's formula and Hartley's rule: Beyond the mathematical coincidence. Entropy, 16(9), 4892-4910.