SNR Prediction Based on Environmental Sensing Data: An Approach Using Machine Learning

Kanchanarat Chokruay and **Anya Apavatjrut**†, Non-members

ABSTRACT

Signal-to-noise ratio (SNR) is a critical metric for assessing wireless link quality and optimizing various aspects of wireless communication, such as modulation level, coding scheme, handover decisions, and antenna configuration. While prior research has primarily focused on SNR prediction based on channel state information using feedback channels, this approach has limitations in terms of applicability and efficiency. In this paper, we propose a novel machine learning-based approach for SNR prediction that leverages environmental sensing data, eliminating the need for feedback channels. Our methodology harnesses the untapped potential of environmental factors, such as soil and air characteristics, to enhance SNR prediction accuracy. By intelligently fusing these environmental parameters with machine learning algorithms, we develop an adaptable SNR prediction model that can effectively capture the dynamics of wireless environments. Experimental results demonstrate the potential to predict the SNR based on environmental data which open up new possibilities for efficient resource allocation, proactive network optimization, and seamless connectivity in dynamic wireless environments, without the constraints imposed by feedback channel availability.

Keywords: LoRaWAN, Machine Learning, SNR, SNR Prediction

1. INTRODUCTION

Signal-to-noise ratio (SNR) is widely used in communication systems as a channel quality indicator (CQI) [1]. SNR measurement is crucial for network configuration as it enables optimal parameter settings for wireless networks across multiple aspects. SNR values guide the adjustment of transmission parameters, including adaptive modulation and coding schemes, to enhance both throughput and reliability of the communication link. Recent mobile communication deployed SNR to assist handover decisions and beamforming [2]. Thus, being able to predict an SNR provides significant advantages for wireless system configuration [3].

Recent cellular technologies allow SNR to be predicted based on channel state information (CSI). The CSI feedback from user equipment on uplink channel is an integral part of the 5G standard [1]. This feedback mechanism facilitates effective sensing of multicast channel quality and supports the decision-making process involved in modulation and coding scheme selection and transmission mechanism determination [4]. However, channel state information is not always available in other wireless technologies. We want to find a way to predict the SNR without using channel state information or feedback channel by exploring other factors that could be deployed to predict the SNR values.

In this study, we investigate the correlation between SNR and environmental variables. Given the absence of a feedback channel for SNR measurement, we examine how environmental variables could serve as potential SNR indicators. The dataset used in this analysis is from [19], which comprises environmental measurements collected by LoRa nodes deployed in an agricultural field. These measurements include air temperature, air humidity, air pressure, soil temperature, and soil moisture. Using machine learning algorithms, we attempt to find the relationship among these variables and predict the SNR using environmental data.

Being able to accurately predict SNR provides significant advantages in configuring wireless systems for optimal performance, leading to more precise channel quality estimation and enhanced reliability in selecting appropriate modulation and coding schemes. SNR prediction can play a crucial role in refining power control, beamforming selection and handover processes, paving the way for enhanced communication efficiency for modern network systems.

2. BACKGROUND AND RELATED WORK

The signal-to-noise ratio (SNR) is a parameter used to measure the efficiency and quality of communication. Generally, the value of SNR is influenced by various factors such as communication distance, carrier frequency, signal interferences, transceiver characteristics, device configuration, and usage, etc[5, 6]. In cellular networks, channel state information or radio measurements from feedback channel are required for SNR prediction [1,2,7]. Several learning algorithms have been proposed for this task, including deep learning [8], K-nearest neighbor, support vector machine, random forest [9], artificial

Manuscript received on August 3, 2024; revised on December 25, 2024; accepted on February 6, 2025. This paper was recommended by Associate Editor Kampol Woradit.

The authors are with Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Thailand.

†Corresponding author: anya@eng.cmu.ac.th

Digital Object Identifier: 10.37936/ecti-eec.2525232.255314

neural network [10], and Gaussian process regression [2,7,9].

Beyond using channel state information, SNR prediction has also been proposed to derive from feedback signals of neighboring wireless nodes [2]. The efficiency of SNR prediction has been enhanced by considering dynamic signal characteristics due to route delay and Doppler effects [11]. Shadowing and blockage of signal transmission are also considered in channel modeling and characterization [10].

The consideration of environmental factors beyond channel characteristics for SNR prediction is relatively limited. For instance, the authors in [12] provide insights into how signals behave in communication between moving vehicles in densely forested suburban environments. Several studies have emphasized the effects of environmental factors on link quality, such as signal degradation due to tree density [13].

Regarding the effects of climate and weather on SNR, research has examined the impacts of temperature and humidity [14], and the effects of congestion in urban areas [15]. The influence of climate on SNR has been discussed; for example, high temperature and high relative humidity may result in decreased SNR [14], while low temperatures tend to yield higher SNR [16]. Increased humidity has been shown to affect wireless signal propagation [17]. It has been analyzed that higher temperatures lead to increased noise levels, thereby reducing SNR [18].

A comprehensive literature review indicates that current SNR prediction methods predominantly depend on channel state information (CSI) and feedback channels. However, not all wireless technologies, such as LoRaWAN, incorporate feedback channels, which limit the applicability of these techniques. Furthermore, measuring SNR levels often requires specialized hardware, which can be expensive and impractical in many situations. To overcome these challenges, we propose a novel SNR prediction methodology that leverages machine learning to eliminate the need for feedback channels. Our literature survey found no mention of machine learning (ML)-based SNR prediction utilizing climate data.

The ML-based SNR prediction method proposed in this study aims to harness the potential of incorporating environmental sensing data, particularly weather data, to improve prediction accuracy. By intelligently integrating this contextual information with advanced ML algorithms, we strive to develop a robust and adaptable SNR prediction model suitable for various wireless environments. This innovative approach not only bypasses the limitations of feedback channel dependency but also leads to optimizing wireless communication systems.

Considering the power constraints of wireless networks like LoRaWAN, which may face multiple limitations, ML applications within LoRaWAN should focus on less complex learning techniques. In this research, we concentrate on algorithms such as Random Forest, XGBoost, etc., which are known for their high performance and precise predictions, even with data of low complexity and limited processing capabilities.

***Table 1**: Summary description of variables of interest and their symbols.*

Variable	Description	Symbol
SNR	Signal-to-Noise Ratio in dB	SNR
Frequency	Transmitting frequency in MHz	f
Distance	Distance measured from the LoRaWAN gateway to the 8 sensor nodes in meters	d
RSSI	Received signal strength indicator in dBm	RSSI
Air temperature	Outdoor air temperature in °C	T_a
Air humidity	Outdoor relative air humidity in %	RH_a
Soil temperature	Soil temperature measured by the node in °C	T_s
Soil humidity	Relative soil humidity measured by the node in %	RH_s
Pressure	Barometric pressure in hPa	P_a

3. FRAMEWORK

In this study, we utilize a public dataset from [19]. The dataset comprises SNR measurements along with environmental variables including air temperature, air humidity, air pressure, soil temperature, and soil moisture, as well as technical parameters such as frequency bands and Received Signal Strength Indicator (RSSI), all collected from an agricultural field. The data were collected from eight Tinovi PM-IO-5SM LoRaWAN sensing nodes, paired with a MikroTik gateway. These devices were strategically placed in an agricultural field situated 30 meters above sea level. Each sensor node was mounted at a height of approximately 1.5 meters from the ground. The Tinovi PM-IO-5-SM sensor nodes are LoRa devices, version 1.0.3 Class A, equipped with the Semtech SX1276 LoRa transceiver, a BME280 indoor air temperature sensor, and a soil humidity sensor. Data collection spanned from November 16, 2020, to February 5, 2021. Additionally, a Davis Vantage Pro 2 weather station was installed to gather weather data for the area.

Basic correlation among these variables have been mentioned in [19] but the relationship among these variables were not analyzed. Therefore, we want to study the effects of the environmental variables by determining factors that have a high relationship with Signal-to-noise ratio (SNR) for prediction.

For our quantitative analysis, we selected the following variables: the signal-to-noise ratio (SNR) measured by the LoRaWAN gateway in decibels (dB), the transmitting frequency in megahertz (MHz), the distance from the LoRaWAN gateway to each sensor node in meters (m), and the received signal strength indicator

***Table 2**: Statistical measurements for each variable.*

Measure	Minimum	1st Quantile	Standard Deviation	Mean	3rd Quantile	Maximum
SNR	-9.00	7.75	1.10	8.65	9.50	11.75
Frequency	867.10	867.35	0.45	867.80	868.30	868.50
RSSI	-108.00	78.00	9.80	-69.98	-62.00	-45.00
Distance	16.0	40.0	120.18	159.79	240.0	340.0
Air Temperature	-4.30	2.0	3.29	4.54	7.10	15.40
Air Humidity	1.00	76.0	14.84	75.56	83.00	87.00
Soil Temperature	0.46	3.31	2.42	5.18	6.74	22.65
Soil Humidity	1.11	68.42	17.40	76.19	90.64	97.18
Pressure	985.00	1006.40	10.30	1014.02	1022.40	1035.20

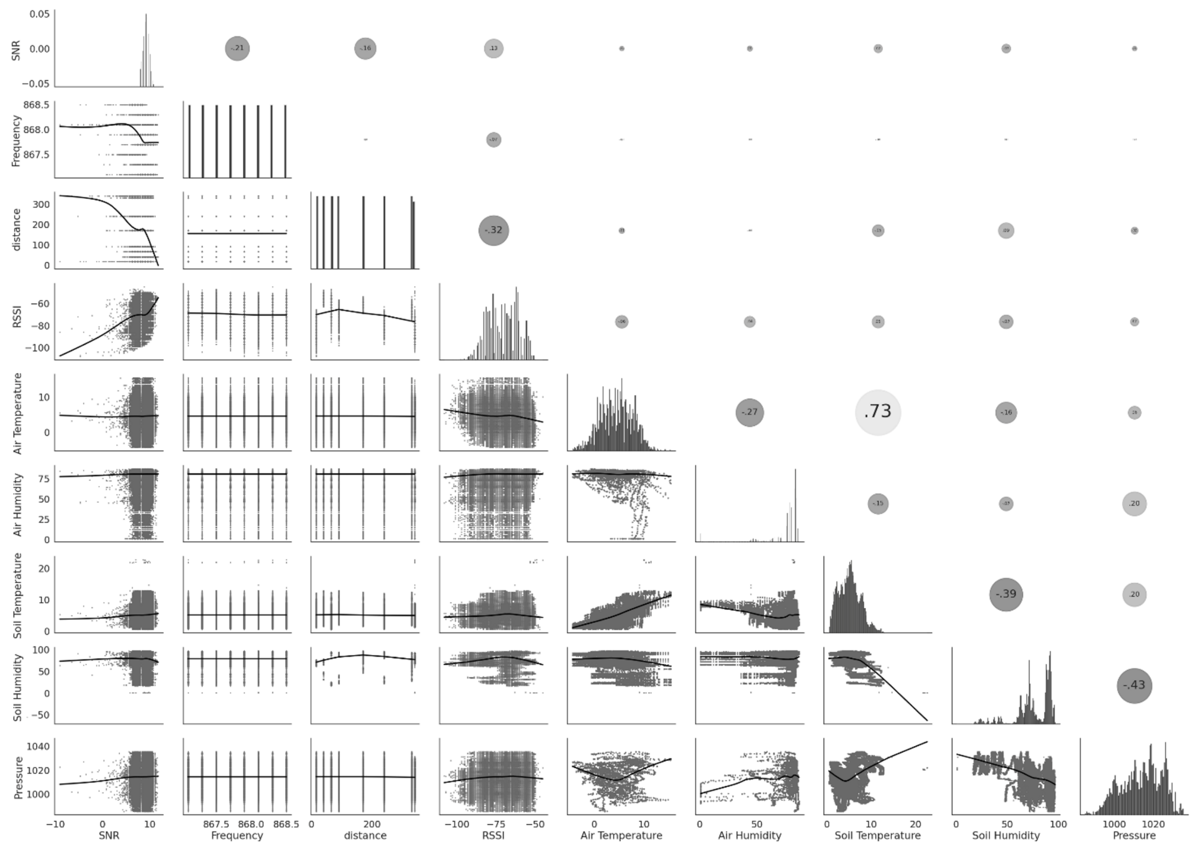

***Fig. 1**: Correlation analysis among SNR and other environmental variables.*

(RSSI) measured by the gateway in decibels-milliwatts (dBm). Environmental parameters included outdoor air temperature in degrees Celsius ($^{\circ}C$), outdoor relative air humidity in percentage (%), air barometric pressure in hectopascal (hPa), soil temperature measured by the node in degrees Celsius ($^{\circ}C$), and soil humidity, which is the moisture content of the soil measured by the node, also in percentage (%). A summary description of these variables, along with the symbols adopted in this work, is presented in Table 1. These variables are collectively referred to as the variables of interest.

The summary of statistical measures for each variable is detailed in Table 2. Preliminary correlation analysis, illustrated in Figure 1, revealed no significant linear relationship between the signal-to-noise ratio (SNR) and the other variables. The highest correlation coefficient was observed between SNR and frequency, at -0.21, indicating a weak inverse relationship. Conversely, the correlation between SNR and the received signal strength indicator (RSSI) was 0.13, suggesting a very slight positive relationship. Correlations between SNR and environmental variables, such as soil temperature and soil humidity, were 0.02 and -0.03, respectively, which are too weak to be considered indicative of significant linear relationship.

As a result, machine learning techniques were applied to discover potential nonlinear or multidimensional relationships within the dataset. A range of machine learning algorithms was employed to investigate the interconnections among the variables and to predict the Signal-to-Noise Ratio (SNR) using the dependent variables, which include the aforementioned environmental parameters.

4. MACHINE LEARNING ALGORITHMS

Machine learning algorithm comprises a wide range of techniques designed to extract insights and make predictions from data. Among these, *Linear Regression* (LR) and *Decision Trees* (DT) are two fundamental machine learning techniques that serve as baselines for more advanced algorithms. LR is a statistical approach that models the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data. While LR is straightforward and interpretable, its performance is limited as it can only capture linear relationships and is sensitive to outliers and multicollinearity. Similarly, DT partitions data into subsets based on feature values, forming a tree-like structure where each node represents a decision rule. Although DT are known for their interpretability, as their structure mirrors a logical decision-making process, making them easy to understand and explain, they are susceptible to overfitting and can be unstable with minor data variations.

To overcome the limitations of these basic techniques, advanced ensemble methods such as *Random Forest* (RF), *Gradient Boosting* (GB), *XGBoost* (XGB), and *CatBoost* (CB) have been developed. These algorithms achieve higher predictive accuracy by combining multiple models. RF is an ensemble of decision trees, where each tree is trained on a random subset of the data and uses a random subset of features for splitting. The final prediction is obtained by aggregating the outputs of all trees, while GB is sequential ensemble technique where each tree is built to correct the errors of the previous trees by optimizing a loss function, XGB is an optimized implementation of gradient boosting that enhances speed and accuracy through parallel processing, tree pruning, and regularization techniques, while CB is specifically designed for datasets with categorical features, eliminating the need for extensive preprocessing. Despite their impressive performance, these ensemble methods can be prone to overfitting if not properly tuned and may require significant computational resources [20,21,22].

These machine learning algorithms were implemented in Python using the *sklearn*, *xgboost*, and *catboost* library as described in Table 3. These libraries were selected based on their robust implementation of the algorithms, wide adoption in the ML community, and extensive documentation.

For the ensemble methods, the number of base models (or estimators) was set to 100, the maximum depth of the decision trees was set to 5, and the random state was chosen as 42.

***Table 3**: Machine Learning Library.*

Technique	Library
RF	sklearn.ensemble.RandomForestRegressor
GB	sklearn.ensemble.GradientBoostingRegressor
XGB	xgboost.XGBRegressor
CB	catboost.CatBoostRegressor
DT	sklearn.tree.DecisionTreeRegressor
LR	sklearn.linear_model.LinearRegression

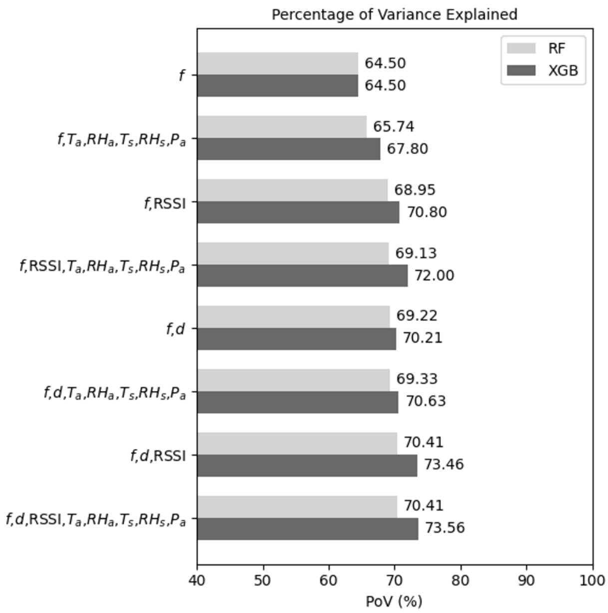

***Fig. 2**: PoV of SNR Prediction Compared BetweenRF and XGB Algorithm.*

5. RESULTS AND DISCUSSION

For predictive analysis, we compare the performance of different variables that are employed by the machine learning techniques to predict the signal-to-noise ratio (SNR) in terms of percentage of variance (PoV) and root mean square error (RMSE).

Percentage of Variance (PoV) is the percentage of total variance explained in the dependent variable in the training set by the independent variable(s) in the model that is constructed using a machine learning technique. The percentage of variance can be expressed as:

$$\text{PoV} = \left(1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}\right) \times 100 \quad (1)$$

Where:

- y_i : Actual value of the dependent variable for the i-th observation.

- $\hat{y}_i$: Predicted value of the dependent variable for the

Table 4: *Percentage of Variance from SNR Prediction"0307*

Parameters	RF	GB	XGB	CB	DT	LR
f	64.51	64.51	**64.51**	64.50	64.51	4.17
f, T_s, RH_s	65.73	66.63	**67.40**	66.18	65.66	4.26
f, T_a, RH_a, P_a	64.54	64.57	64.52	64.55	64.51	4.19
$f, T_a, RH_a, T_s, RH_s, P_a$	65.75	66.72	**67.80**	66.26	65.67	4.29
f, d	69.23	70.03	70.21	69.88	69.21	6.87
f, d, T_s, RH_s	69.28	70.17	70.51	69.92	69.23	6.90
f, d, T_a, RH_a, P_a	69.32	70.20	70.58	69.97	69.28	6.88
$f, d, T_a, RH_a, T_s, RH_s, P_a$	69.33	70.22	**70.64**	69.97	69.29	6.91
$f, RSSI$	68.95	70.03	70.80	69.33	68.71	5.53
$f, RSSI, T_s, RH_s$	69.13	70.60	71.82	69.85	68.82	5.57
$f, RSSI, T_a, RH_a, P_a$	68.96	69.84	70.85	69.33	68.71	5.54
$f, RSSI, T_a, RH_a, T_s, RH_s, P_a$	69.13	70.57	**72.01**	69.85	68.82	5.57
f, d	70.41	72.22	73.47	71.57	70.02	7.33
$f, d, RSSI, T_s, RH_s$	70.41	72.27	73.60	71.62	70.02	7.34
$f, d, RSSI, T_a, RH_a, P_a$	70.41	72.26	73.57	71.61	70.02	7.34
$f, d, RSSI, T_a, RH_a, T_s, RH_s, P_a$	70.41	72.23	**73.56**	71.59	70.02	7.35

Table 5: *Root Mean Squared Error from SNR Prediction.*

Parameters	RF	GB	XGB	CB	DT	LR
f	0.66	0.66	0.66	0.66	0.66	1.09
f, T_s, RH_s	0.65	0.64	0.64	0.65	0.65	1.09
f, T_a, RH_a, P_a	0.66	0.66	0.66	0.66	0.66	1.09
$f, T_a, RH_a, T_s, RH_s, P_a$	0.65	0.64	**0.63**	0.65	0.65	1.09
f, d	0.62	0.61	0.61	0.61	0.62	1.07
f, d, T_s, RH_s	0.62	0.61	0.6	0.61	0.62	1.07
f, d, T_a, RH_a, P_a	0.62	0.61	0.6	0.61	0.62	1.07
$f, d, T_a, RH_a, T_s, RH_s, P_a$	0.62	0.61	0.6	0.61	0.62	1.07
$f, RSSI$	0.62	0.61	0.6	0.62	0.62	1.08
$f, RSSI, T_s, RH_s$	0.62	0.60	0.59	0.61	0.62	1.08
$f, RSSI, T_a, RH_a, P_a$	0.62	0.61	0.6	0.62	0.62	1.08
$f, RSSI, T_a, RH_a, T_s, RH_s, P_a$	0.62	0.60	0.59	0.61	0.62	1.08
f, d	0.61	0.59	0.57	0.59	0.61	1.07
$f, d, RSSI, T_s, RH_s$	0.61	0.59	0.57	0.59	0.61	1.07
$f, d, RSSI, T_a, RH_a, P_a$	0.61	0.59	0.57	0.59	0.61	1.07
$f, d, RSSI, T_a, RH_a, T_s, RH_s, P_a$	0.61	0.59	0.57	0.59	0.61	1.07

i-th observation.

- $\overline{y}$: Mean of the actual values of the dependent variable.
- n : Number of observation.s

Root Mean Square Error (RMSE) is a commonly used measure of the differences between predicted values and observed values in a model. RMSE is given by:

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

Where:

- y_i : Actual value of the dependent variable for the i-th observation.
- $\hat{y}_i$: Predicted value of the dependent variable for the i-th observation.
- n : Number of observations.

These formulas provide a way to evaluate the performance of each machine learning model. Higher PoV and lower RMSE indicate better performance.

Various machine learning algorithms were utilized to explore the relationships among the variables and

***Table 6**: Computational Time from SNR Prediction.*

Parameters	RF	GB	XGB	CB	DT	LR
f	2.19	1.51	0.47	2.54	0.02	0.01
f, T_s, RH_s	9.66	9.39	0.6	1.57	0.13	0.02
f, T_a, RH_a, P_a	8.97	7.67	1.89	2.01	0.11	0.03
$f, T_a, RH_a, T_s, RH_s, P_a$	16.31	15.11	0.73	1.87	0.23	0.03
f, d	3.52	3.03	0.64	1.36	0.03	0.01
f, d, T_s, RH_s	10.98	9.59	2.28	1.61	0.16	0.02
f, d, T_a, RH_a, P_a	10.05	9.2	0.7	2.12	0.19	0.04
$f, d, T_a, RH_a, T_s, RH_s, P_a$	18.24	16.18	2.53	2.51	0.33	0.06
f, *RSSI*	4.24	3.23	0.53	1.36	0.04	0.02
f, *RSSI*, T_s, RH_s	11.55	10.91	0.57	1.55	0.15	0.02
f, *RSSI*, T_a, RH_a, P_a	10.96	9.86	1.94	1.67	0.14	0.02
f, *RSSI*, $T_a, RH_a, T_s, RH_s, P_a$	18	16.7	1.21	3.43	0.25	0.04
f, d	4.03	4.09	0.58	2.86	0.08	0.03
f, d, *RSSI*, T_s, RH_s	11.99	11.15	0.68	3.37	0.21	0.02
f, d, *RSSI*, T_a, RH_a, P_a	11.48	10.86	0.7	3.27	0.23	0.06
f, d, *RSSI*, $T_a, RH_a, T_s, RH_s, P_a$	19.97	17.79	2.19	2.24	0.27	0.04

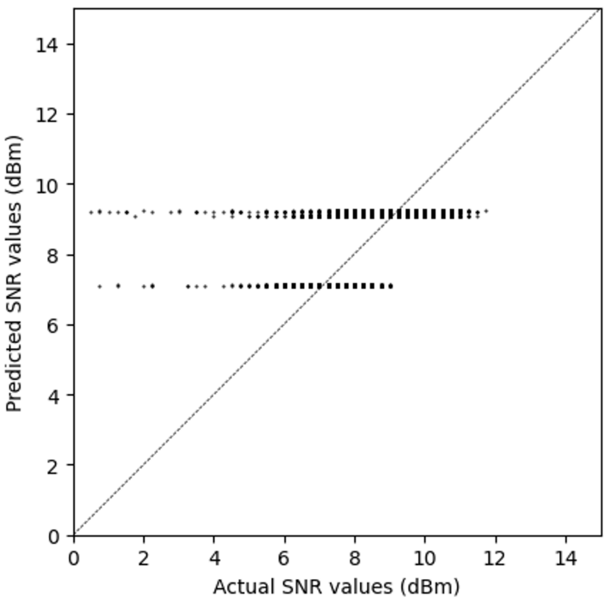

***Fig. 3**: SNR Prediction without Environmental Parameters.*

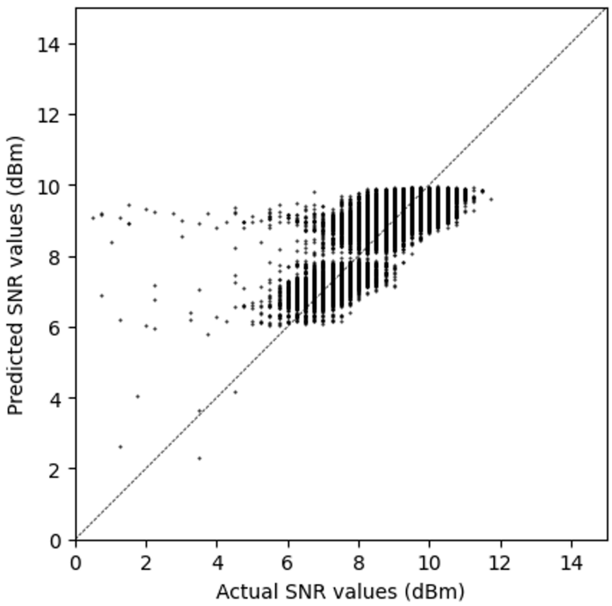

***Fig. 4**: SNR Prediction with Environmental Parameters.*

predict SNR based on the dependent variables including frequency (f), received signal strength indicator (RSSI), distance (d), soil temperature (T_s), soil relative humidity (RH_s), air temperature (T_a), air humidity (RH_a) and barometric pressure (P_a). Considering each of these parameters as a unique input parameter for SNR prediction, the analysis revealed that the carrier frequency f was the only unique parameter that enabled SNR prediction with a reasonable level of accuracy. The parameter f alone provides a baseline PoV of around 64.5% for most algorithms, except for linear regression (LR), which perform poorly at 4.17%. SNR prediction based on RSSI alone yields the PoV of 11.80%. For other unique environmental parameters when considered individually, the machine learning models failed to accurately predict SNR, as the PoV values were all inferior to 2%, indicating significantly low prediction accuracy. These preliminary results demonstrate that the relationship between SNR and other environmental variables might not be evident, as the correlation analysis failed to identify a connection between them.

Interestingly, incorporating multiple variables in the analysis improved prediction accuracy. Table 4 and Table 5 present the prediction results based on percentage of variance (PoV) and root mean square error (RMSE) between the actual and predicted values. When comparing the use of frequency alone as an input parameter against the combination of frequency with soil temperature and soil humidity, we observed significant improvements in SNR prediction across all ML algorithms. The XGBoost (XGB) algorithm achieved the highest prediction precision, with PoV increasing from 64.51% to 67.40%. This improvement of approximately 4.48% demonstrates the positive impact of including soil characteristics in the prediction model.

Further addition of environmental parameters continued to improve prediction accuracy. When all environmental parameters including air and soil cha-racteristics were included alongside frequency, the XGB algorithm reached its highest accuracy with PoV of 67.80% with the RMSE reduced to 0.63. This suggests that these additional environmental variables provide valuable information for SNR prediction, as the increase in PoV and the reduction in RMSE validate the consistency of the evaluation metrics used in the study.

Lastly, including the received signal strength indicator (RSSI) and the distance between two communicating devices as features further enhances the prediction accuracy. The PoV increases from 67.80% to 70.64% when distance is added, and to 72.01% when RSSI is added. Combining both parameters as features, the PoV increased to 73.56%.

Table 6 shows the computational time for each learning algorithm. Based on this table, the decision tree technique demonstrated advantages in computational speed. However, the XGBoost algorithm, despite being slightly slower than the decision tree-based methods, still provided satisfactory computational times compared to other machine learning techniques.

Figure 2 illustrates the percentage of variance (PoV) derived from the random forest (RF) and XGBoost (XGB) algorithms for SNR prediction. Although the accuracy of each technique exhibited slight variations, the overall trend indicated that incorporating certain environmental data enhanced the accuracy of SNR prediction. The simulation results suggest that soil characteristics, such as soil temperature and soil humidity, have a positive impact on improving SNR prediction accuracy. In contrast, air characteristics did not significantly contribute to the prediction accuracy, as their presence did not notably influence the model's performance.

Compared between various machine learning algorithm, the XGBoost algorithm consistently yielded the highest prediction accuracy while linear regression performs poorly. These results suggest that the proposed machine learning platform can help uncover complex, non-linear relationships between SNR and other variables. The XGBoost algorithm appears to be the most effective choice among the tested algorithms for this specific task.

The scatter plots in Figure 3 and Figure 4 compare the actual and predicted SNR values. As observed in these figures, the accuracy of SNR prediction was significantly lower when the actual SNR values were low. However, when the actual SNR values were high, the predicted results exhibited better accuracy and were more satisfactory. This observation suggests that the machine learning models perform better in predicting SNR values within a higher range, while their accuracy diminishes for lower SNR values.

Overall, the incorporation of environmental data, particularly soil temperature and humidity, alongside the carrier frequency, improved the accuracy of SNR prediction using machine learning techniques. The XGBoost algorithm consistently outperformed other models in terms of prediction accuracy, while the decision tree-based methods offered advantages in computational efficiency. These findings highlight the potential of leveraging environmental data to enhance the performance of wireless communication systems and open venues for further research in this domain.

Although no strong correlation was identified for individual environmental variables, their combined influence in conjunction with technical parameters demonstrated improvements in predictive performance. This suggests that while individual environmental parameters might not strongly correlate with SNR, their collective impact contributes to a more robust predictive model.

The study acknowledges the challenges in using environmental data for SNR prediction. External interferences, varying terrain, and seasonal changes can introduce noise into the dataset, potentially reducing the reliability of the model. Future work should consider integrating additional external factors, such as wind speed or precipitation as well as other environmental factors, to refine the predictive capabilities further.

6. CONCLUSION AND PERSPECTIVE

In this paper, we analyzed the impact of environmental parameters to enhance SNR prediction. While no strong correlation was found among the individual variables, their combined influence significantly improved the accuracy of SNR prediction when using machine learning. This improvement is valuable for optimizing network configurations, as discussed earlier in the paper.

The results clearly indicate that frequency has a substantial impact on SNR prediction. When combined with soil temperature and soil humidity, the prediction accuracy improves, resulting in a measurable increase of up to 4.48% in the percentage of variance (PoV). This underscores the importance of integrating contextual environmental data, such as soil temperature and humidity, alongside traditional variables like frequency. By incorporating these factors, the machine learning model effectively captures the complex dynamics of the wireless environment, enhancing predictive accuracy and validating the hypothesis that environmental data substantially contributes to SNR prediction.

The improvement is primarily attributed to the role of soil characteristics such as moisture content, composition, conductivity in influencing the propagation and reflection of electromagnetic waves through the ground.

These factors affect signal attenuation and multi-path effects, which can distort the desired signal. In contrast, air characteristics in this experiment represented more uniform and incur less accuracy improvement compared to soil or other dense materials. While air can introduce noise through precipitation, turbulence, or similar factors, it may not attenuate or distort signals as significantly as the ground.

As a result, this paper has shown promise for novel SNR prediction using machine learning. By eliminating the reliance on feedback channels and leveraging the untapped potential of environmental sensors, our methodology establishes a foundation for more efficient, accurate, and adaptable SNR prediction techniques. Having advanced knowledge of the expected SNR in LoRaWAN network can enable nodes to optimize their link budgets, transmission parameters and energy consumption though techniques like adaptive data rate or interference mitigation. It also aids in network planning by predicting coverage or guiding gateway. This proposed approach has the potential to enhance network performance and reliability.

One of the notable insights from this research is adaptability of the proposed method. Although the focus of this study was on SNR prediction, the methodology is not limited to this metric alone. For instance, it could be extended to predict other important wireless communication metrics, such as the received signal-strength-indicator (RSSI). The principles of leveraging environmental data, combined with machine learning, remain applicable for various metrics as long as they are susceptible by external environmental factors.

Overall, this study contributes to the advancement of SNR prediction by leveraging environmental data in machine learning models, paving the way for further research and practical implementations.

One key limitation of the study is the dataset's geographical specificity. The environmental data used in this study were collected from an agricultural field, meaning the model may not generalize well to urban or industrial environments. Additionally, the study primarily relies on classical machine learning models. While the results indicate that XGBoost performs best, deep learning techniques such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs) could be explored to capture more complex relationships within the dataset.

Future work can enhance the adaptability and responsiveness of the network by introducing real-time data analysis to provide immediate adjustments based on current network conditions for real-time SNR prediction. Adaptive learning algorithms can also be introduced to adjust the parameters based on the current network state to ensure consistent prediction accuracy. Expanding data collection to multiple locations will help assess model adaptability across different terrains and climates, ensuring a broader applicability of the proposed approach.

ACKNOWLEDGEMENT

This work was supported by the New Researcher Development Fund from the Faculty of Engineering, Chiang Mai University.

REFERENCES

[1] K. Saija, S. Nethi, S. Chaudhuri and R. M. Karthik, "A Machine Learning Approach for SNR Prediction in 5G Systems," *2019 IEEE International Conference on Advanced Networks and Telecommunications Systems(ANTS)*, Goa, India, 2019, pp.1-6.

[2] P. Kazemi, H. Al-Tous, T. Ponnada, C. Studer and O. Tirkkonen, "Beam SNR Prediction Using Channel Charting," *in IEEE Transactions on Vehicular Technology*, vol. 72, no. 10, pp. 13130-13145, Oct. 2023.

[3] C. Beard, W. Stallings and M.P. Tahiliani, *Wireless Communication Networks and Systems*, Pearson, 2015.

[4] H. Peng, C. Feng, Q. Li, "User State Prediction Based on Channel Quality Feedback in 5G Broadcast," *2023 IEEE 13th International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, Qinhuangdao, China, 2023, pp. 547-551.

[5] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed., Prentice Hall, 2001.

[6] S. Haykin, *Communication Systems*, 5th ed., John Wiley & Sons, Inc., 2009.

[7] P. Kazemi, T. Ponnada, H. Al-Tous, Y. -C. Liang and O. Tirkkonen, "Channel Charting Based Beam SNR Prediction," *2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, Porto, Portugal, 2021, pp. 72-77.

[8] Q. Zhou, W. Jiang, D. Wang, and H. D. Schotten, "Deep Learning-Based Signal-to-Noise Ratio Prediction for Realistic Wireless Communication," *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, Helsinki, Finland, 2022, pp. 1-5.

[9] P. M. and B. R. Reddy, "Machine Learning based Prediction for Channel Stability in a 5G Network," *2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon),* Vijaypur, India, 2022, pp. 1-4.

[10] Z. Zhao, B. Deng, F. N. Khan, and H. Y. Fu, "Deep Learning-Based Signal-to-Noise Ratio Estimation for Underwater Optical Wireless Communication," *2022 IEEE 14th International Conference on Advanced Infocomm Technology (ICAIT)*, Chongqing, China, 2022, pp. 120-123.

[11] T. Ngo, B. Kelley, and P. Rad, "Deep Learning Based Prediction of Signal-to-Noise Ratio (SNR) for LTE and 5G Systems," *2020 8th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Reims, France, 2020, pp. 1-6.

[12] C. M. M. Cardoso, F. J. B. Barros, J. A. R. Carvalho, A. A. Machado, H. A. O. Cruz, N. M. C. Alcantara,

J. P. L. Araújo, "SNR Prediction with ANN for UAV Applications in IoT Networks Based on Measurements," *Sensors*, vol. 22, no. 14, p. 5233, 2022.

[13] M. Ransah, R. A. Sowah, J. Meli-Segu, F. A. Katsriku, X. Vilajosana, and W. Owusu-Banahene, "Characterizing Foliage Influence on LoRaWAN Pathloss in a Tropical Vegetative Environment," *IET Wireless Sensor Systems*, vol. 10, no. 5, pp. 198-207, 2020.

[14] N. Jeftenic, M. Simic, and Z. Stamenkovic, "Impact of Environmental Parameters on SNR and RSSI in LoRaWAN," *International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, Istanbul, Turkey, 2020, pp. 1-6.

[15] Gonzalez-Palacio, M., Tobón-Vallejo, D., Sepúlveda-Cano, L., Roa Pérez, S., Pau, G., Le, L., "LoRaWAN Path Loss Measurements in an Urban Scenario Including Environmental Effects," *Data*, vol. 8, no. 4, 2022.

[16] N. S. Bezerra, C. Åhlund, S. Saguna, V. A. Jr. de Sousa, "Temperature Impact in LoRaWAN: A Case Study in Northern Sweden," *Sensors*, vol. 19, p. 4414, 2019.

[17] S. Calvert, E. Popovici, and P. Leahy, "Using Environmental Data-based Communication Protocol for Improved Quality of Service in LoRaWAN Applications," *The 32nd Irish Signals and Systems Conference (ISSC)*, Athlone, Ireland,2021, pp. 1-6.

[18] L. Parri, S. Parrino, G. Peruzzi, and A. Pozzebon, "Offshore LoRaWAN Networking: Transmission Performances Analysis Under Different Environmental Conditions," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-10, 2021.

[19] E. Goldoni, P. Savazzi, L. Favalli, A. Vizziello, "Correlation Between Weather and Signal Strength in LoRaWAN Networks: An Extensive Dataset," *Computer Networks*, vol. 202, 2022.

[20] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 5-32, 2001.

[21] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, Association for Computing Machinery, New York, NY, USA, 2016, pp. 785-794.

[22] L. Prokhorenkova, G. Gusev, A. Vorobev, A. Dorogush, and A. Gulin, "CatBoost: Unbiased Boosting with Categorical Features," *In Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18)*, canada, USA, 2018, pp. 6639-6649.

Kanchanarat Chokruay received her Bachelor's degree in Sciences (Statistics) from department of Mathematics and Statistics at Maejo University, Thailand (2021). She is pursuing a master's degree in Computer Engineering at Chiang Mai University.

Anya Apavatjrut received the Engineering degree (2007), Master's degree in Computer Science (2007) and Ph.D. in Telecommunications (2011) from INSA de Lyon, France. She has been with the Faculty of Computer Engineering, Chiang Mai University, Thailand, since 2011, where she is currently an Associate Professor. Her research interests include computer networks, signal characteristics, propagation and localization.