This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2023.3238013

Correlating the Ambient Conditions and Performance Indicators of the LoRaWAN via Surrogate Gaussian Process-based Bidirectional LSTM Stacked Autoencoder

Showkat Ahmad Bhat[1*], *Student Member, IEEE,* Nen-Fu Huang[2*], *Senior Member, IEEE,* Imtiyaz Hussain[3*], Uzair Sajjad[4]

***Abstract*— LoRa's biggest advantage is its flexibility, which is the ability to increase or decrease data rate and range while decreasing or increasing sensitivity. Whenever propagation conditions change frequently, this function allows the spreading factor to be modified accordingly. Despite their efficiency and scalability, adaptive data rate algorithms ignore and fail to factor in the complex correlation between ambient weather parameters influencing the communication channel design. In this research, a Bayesian surrogate Gaussian process-based bidirectional LSTM stacked autoencoder model (BSGP-BLSTM_SAE) is proposed to estimate the channel performance indicators such as received signal strength indicator (RSSI) and signal-to-noise ratio (SNR) and to determine the correlation between the ambient weather conditions and performance indicators for the LoRaWAN network. Bayesian optimization algorithm has been used to optimize the hyperparameters of the developed model. A LoRaWAN experimental multivariate time series dataset has been used for the evaluation of the developed model, which upon testing and validation produces high accuracy in predicting the channel performance indicators and ambient conditions of the experimental LoRaWAN network. The mean absolute error of the developed model was around 0.45. Thus, the proposed model can predict the link performance indicators and thereby assist in real-time optimization of the transmission parameters to enhance the network performance in LoRaWAN-based systems at different ambient conditions.**

Index Terms—LoRaWAN, Recurrent Neural Network, Bayesian Optimization, Ambient Weather Conditions, Internet of Things, Stacked Autoencoder, Adaptive Data Rate

I. INTRODUCTION

The LoRaWAN protocol is broadly utilized in the IoT for providing wireless connectivity for a variety of applications [1] and is getting deployed at an extremely high rate. There are several reasons for the growing popularity of the LoRaWAN, including the increasing number of industries engaged in the development process [2], the appropriate selection of a balance between the low complexity and high power efficiency of the technology, and its practicability to a diverse range of applications. An important feature of the LoRaWAN protocol [3] is that it does not require main back compatibility with underlying technologies unlike NB-IoT and Wi-Fi HaLow.

Therefore long-range (LoRa) technology, along with the LoRaWAN standard [4], is among the most reliable and efficient data communication methods in the field of internet of things (IoT) for sub-gigahertz systems. The success of LoRaWAN stems from the realization of three critical characteristics of IoT such as a long range of communication, power efficiency, and inexpensive device and network infrastructure. As a result, LoRaWAN-based systems are being deployed in challenging conditions in which the technical qualities of this system help to overcome certain limitations like the severity of the data collecting systems implementation location [5], where coverage is hard to ensure, and where access for battery replacement is difficult. An overview of LoRaWAN's architecture can be seen in Figure 1, which shows sensor nodes, gateways (GW), network servers (NS), and application servers (AS).

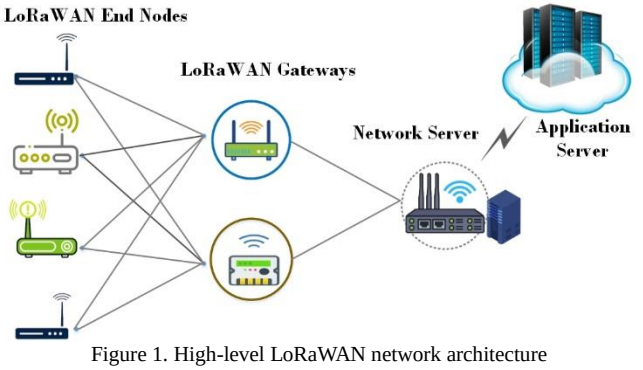


Figure 1. High-level LoRaWAN network architecture

However, environmental conditions influence the reliability of wireless channel transmission and connectivity between the nodes. The influence of temperature, rain, and humidity variations [6], pressure, snow cover, heavy vegetation, and other factors on the performance of different low-power wireless protocols [7] have been studied [8, 9]. Temperature fluctuations have a significant influence since they can induce clock drift,

- *S. A. Bhat is with the College of Electrical and Computer Science, National Tsing Hua University, Hsinchu, 1300044, Taiwan, (R.O.C), E-mail: showkatbhat1994@gmail.com.
- *N.-F. Huang is with the Department of Computer Science, National Tsing Hua University, Hsinchu, 1300044, Taiwan, (R.O.C), E-mail: nfhuang@cs.nthu.edu.tw.
- *I. Hussain is with the Graduate Program in Energy and Opto-Electronic Materials, National Taipei University of Technology, Taipei, 10608, Taiwan, (R.O.C), E-mail: imtiyazkou@gmail.com.
- U. Sajjad is with the Department of Energy & Refrigerating Air-conditioning Engineering, National Taipei University of Technology, Taipei, 10608, Taiwan, (R.O.C), E-mail: energyengineer01@gmail.com.

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: Southeast University. Downloaded on February 19,2023 at 08:24:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2023.3238013

capacity changes, and depletion of the battery's charge level [8], heavy rainfall along with fog causes significant packet loss rate and disruption of connectivity between the nodes [8], and humidity has also a direct correlation with the received signal strength [6]. Fundamental issues like time synchronization, communication reliability, and operating time of the network can also be affected by variations in weather conditions. For instance, communication systems built on IEEE 802.15.4 transceivers [10] have been widely employed in the previous one and a half decade. Following its deployment, researchers looked into the stability of radio connections, which led to improvements in routing algorithms and the functioning of duty-cycled Medium Access Control (MAC) protocols.

It is important to utilize radio and channel resources efficiently in an environment setting where a single base station links and supports large IoT end devices. For this purpose, LoRaWAN offers a range of user transmission parameters that helps to optimize resource usage and regulation of connection performance. LoRaWAN devices are typically configured to use a different range of parameters, such as spreading factors (SFs), transmission power (T_P), central frequencies (CFs), bandwidths (BW), and coding rates (C_Rs). To adjust the transmission parameters such as SF, BW, C_R, and T_P, LoRaWAN makes use of the adaptive data rate (ADR) technique [11]. But transmission parameters are allocated to the end devices by the ADR conservatively, which causes the medium they rely upon to get saturated very quickly. The other ADR methods [12] [13] [14] [15] that followed [12, 13] sought to reduce the risk of inter-SF collisions by spreading the load fairly across multiple SFs, whose orthogonally ensures optimal device fairness. Although SF was allocated equitably, energy efficiency and longevity of edge devices were not considered, meaning the devices had a short lifespan regardless of their size or connectivity. It has been discovered that they are only useful in tiny cells with access to all SFs. The hybrid ADR proposed in [14] allows SF and T_P to be allocated simultaneously to static and mobile applications. In [15], the authors suggested enhanced ADR (EADR) for the appropriate assignment of C_R and the tradeoffs that come with it. By utilizing the capture effect, EADR increases the probability that collision signals will survive. Even while the capture effect might boost the survival rate of collision signals, it is often overlooked or misunderstood. None of the existing ADR designs considers the tradeoff with the ambient conditions. In the case of the network with fixed-end nodes, the ADR is controlled by the NS side based on the history of the uplink (UL) packets that have been received, known as "Network-managed ADR or Static ADR". ADR using a network-based approach is not suitable for networks with mobile end nodes due to channel attenuation. Then, ADR is performed blindly on the mobile end node side referred to as "Blind ADR". We will indeed be able to change the settings for a specific connection situation if the performances of end devices can be accurately estimated while considering the ambient channel conditions even when the downlink is unavailable due to the harsh environment.

In this study, we have developed a Bayesian Surrogate Gaussian Process using bidirectional LSTM [16] sequence to sequence stacked autoencoder for multivariate multi-step time series forecasting of LoRaWAN channels performance indicators such as received signal strength indicator (RSSI) and signal-to-noise ratio (SNR) based on the ambient weather conditions employing a LoRaWAN experimental dataset. Bayesian Optimization Algorithm is used for tuning the hyperparameters of the proposed model to use it for multivariate time series prediction. In univariate time series, the time series is made up of sequential samples of single (scalar) observations. However, a Multivariate time series has more than one time-dependent variable and one sequential variable. Each variable depends not only on its past values but also has a dependency on other variables. The given problem is composed of multivariable input and multivariable output. Assuring a correct fit for these hyperparameters is an art. The DL framework employs a variety of hyperparameters tuning algorithms, such as grid search and random search, among others. However, the grid search approach does not scale well; an increase in the size of the hyperparameter search space will result in an exponential rise in run time and computation. Moreover, random search tests hyperparameter sets at random, it runs the risk of missing the ideal set of hyperparameters and forgoing peak model performance. In contrast to grid search and random search, in which each treats hyperparameter sets independently, Bayesian optimization is an informed search method that learns from previous iterations. The number of trials in this approach is determined by the user. Thus, Bayesian Optimization is a viable strategy for locating the extrema of a given multivariate time series objective function. The objective function is estimated as a Gaussian process and perceived as a proxy function. Bayesian Optimization performs well when the closed-form expression of the provided objective function is unknown, but specific observations may be derived from it. In our devised model, Bayesian optimization is employed to find the optimal hyperparameters for discovering the test or validation loss minima and minimizing MEA metric function. Moreover, the Bayesian surrogacy is a modern tool to optimize the hyper-parameters of the LSTM-based RNN architecture in a very short time (almost 10-15 hours) as compared to other automatic tuning methods such as evolutionary algorithms where the computation time is almost in days. Thus, the computational complexity of the proposed method is better than other optimization methods.

The contributions of this study are presented as follows; 1) developed a BSGP-BLSTM_SAE model to evaluate the system's transmission performance over an extended period and correlate it with the experiment site's ambient weather conditions; 2) the proposed method automatically selects the optimized model by considering different combinations of LSTM stacked autoencoder hyper-parameters for a given multivariate time series by using the Gaussian process-based Bayesian optimization method; 3) the developed model can be used for estimation of channel performance indicators (RSSI and SNR) as well as weather forecasting; 4) It has the capability of capturing nonlinear patterns in time series data while reflecting the fundamental properties of non-stationary data, and 5) deliberated how to use the proposed LSTM-based predictor to improve the current LoRaWAN's ADR method to optimize the end device settings. The results show the effect of environmental parameters on the LoRaWAN and how the channel performance can be predicted using deep learning techniques and proved that the proposed model has practical significance and innovation.

II. Background and Related Works

A. Background

LoRaWAN has been widely utilized in IoT applications like

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: Southeast University. Downloaded on February 19,2023 at 08:24:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2023.3238013

agriculture, health, home automation, and many other applications since its inception in 2015, as they provide important IoT criteria including reliable bi-directional communication, mobility, and localization services [17]. The LoRaWAN provides critical characteristics for IoT such as enabling long-range and low-power communication across a large area. It defines the MAC layer protocol as well as the overall system architecture using the LoRa physical layer. In LoRa's physical layer, its M-ary chirp spread spectrum modulation (CSS), uses chirps, which are sine waves with linearly varying frequencies. Using f_0 as the center frequency of the sweep interval [$f_0 \pm$ ½ BW] and $t = 0$ as the signal beginning moment, a single LoRa chirp can be expressed mathematically as:

$$s(t) = A_0 \cos(2\pi t + 2\pi \int_0^t \Delta f(m,t)dt + \emptyset_0), \quad 0 \le t \le T_m \tag{1}$$

where $A_0 > 0$ denotes the amplitude of a chirp, $m \in 0, \dots, M-1$ represent the modulation symbol, $\Delta f(m,t)$ is an instantaneous frequency offset dependent on symbols, $\emptyset_0$ describes signal phase at t = 0, and T_s represents chirp duration.

Among the many different types of chirp used in LoRa, each corresponds to a different symbol of the modulation symbol $m = \{0, \dots, M-1$. A modulation symbol $m \in M$ causes the corresponding $f(m,t)$ to rise linearly at $-\frac{1}{2} BW$. The $f(m,t)$ wraps around to $-\frac{1}{2} BW$ after reaching maximum frequency offset $\frac{1}{2} BW$, then increases linearly until it reaches $\Delta f(m, t = T_m) = \Delta f(m, 0)$. As chirp duration reflects the symbol interval T_m, it is commonly known as the symbol interval.

Furthermore, the LoRaWAN standard has several features, like encryption and security of end-to-end (E2E) traffic, frequency regulation and optimization with ADR, and quality of service, as well as other advanced communication applications. LoRa nodes can handle simultaneous data transmission via dynamic data rates and pseudo-randomly modified channels [12].

To achieve different data rates (D_R) and network performance, LoRaWAN uses several transmission parameters. There are eight transmission channels each with different bandwidths: 125 kHz, 250 kHz, or 500 kHz. Bandwidths are chosen based on the particular mode of transmission such as long-range, transmission time (ToA), and sensitivity. The C_R utilized by the LoRaWAN transceiver may also be changed to any of the four C_Rs such as 4/5, 4/6, 4/7, or 4/8. Based on the selected value, every 4 bits will be encrypted by 5, 6, 7, or 8 transmission bits (bit rate (B_R)) [17]. However, longer transmission time (ToA) and lower C_R provide better reliability.

LoRa transceivers also can adjust the transmission power, allowing them to dramatically change their radio coverage range and energy requirements. For illustration consider the RFM_{95} transceiver [18], the power usage grows from 66 mW to 396 mW by varying the T_P from -4 to +20 dBm. As a result, a larger T_P improves SNR while enhancing the energy consumption of the transceiver along with increased interference and collision rates [19]. LoRa employs six programmable orthogonal SFs to achieve network separation. Setting each parameter directly impacts the D_R and ToA, along with the energy consumption of the LoRaWAN nodes.

B. Related Works

For LoRaWAN technology, research indicates that apart from the effects of variations in the combination of the parameters (as discussed in the previous section) on the performance of the LoRaWAN, rapidly changing ambient weather conditions also impact the communication between the LoRaWAN nodes [20]. When establishing outdoor commercial LoRaWAN-based sensor networks, LoRaWAN nodes must sustain continuous communication across a variety of environmental circumstances [8]. This is particularly challenging given the effect of high temperatures on sensor nodes and the radios they use. Researchers found that signal strength declined significantly during the warmest part of the day [21, 22], while other works also suggest that fog and rain have a severe impact on the transmission and connectivity of low-power wireless links [23]. However, the connectivity irregularities presented are usually minimal. It has been suggested that the longer link distances may have larger impacts.

There has been considerable packet loss as a result of high node temperatures, as described in reference [24]. An increase in hysteresis is observed during the warming-up or cooling-down of the receiver. An outdoor study correlates temperature, humidity, packet reception rate, and received signal intensity. This is similar to the experimental assessment in [25] which assesses the reliability of LoRa receivers under temperature fluctuations. In laboratory tests, it has been found that as ambient temperature increases by 10^O C, signal intensity decreases by about 1 dB. Also, a thorough study is conducted to demonstrate how you can increase the likelihood of receiving packets by selecting the appropriate LoRa parameters.

In [26], controlled lab equipment was used to study the sensitivity of LoRa to temperature and relative humidity. Furthermore, the authors set a temperature that makes LoRa communication unstable. In this case, however, the major focus was on the temperature of the LoRa sensors instead of the ambient temperature. The effect of temperature on LoRaWAN was also studied in Northern Sweden [21] having a subarctic climate and observed that lower temperatures reduce noise power, which improves SNR but it does not consider the impact of the other ambient parameters. In addition, [21] shows that snowfall adversely affects connection quality, particularly when long distances are covered. While [27] looked at the hypothetical coverage of LoRa in distinct SFs throughout the urban, suburban, and countryside situations using a single end node installed over a vehicle to calculate the achieved D_Rs at different distances from the base station with the same ambient conditions.

Additionally, [5] installs an offshore sensor node and two ashore gateways to demonstrate the feasibility of offshore LoRaWAN infrastructure. The correlation is drawn by plotting 70 days of data on transmission performance indicators with environmental and meteorological variables without any deep analysis. An analysis of RSSI and SNR was presented in [28] using a designed testbed to analyze the impact of temperature and relative humidity by varying D_R manually.

The performance analysis of LoRa based on physical layer parameters has been performed in different studies. For instance, in [29] path loss (PL) and RSSI are analyzed for indoor and outdoor settings of different SF, whereas in [30], RSSI and delay are analyzed for performance that can be observed over a broader area, such as a university campus. A large number of sensor nodes was investigated by [31] while adjusting the SF and modeling the results in terms of PL for indoor and outdoor RSSI and SNR in an industrial context with an adjustable SF. In their studies [32, 33], the researchers studied RSSI, SNR, and PL for different SFs, while [34] explored RSSI and ADR in

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: Southeast University. Downloaded on February 19,2023 at 08:24:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2023.3238013

rural areas within smart agriculture, moving between SFs, BWs, and C_Rs.

The efficiency of the LoRa link based on the above-discussed parameters has been demonstrated in the bulk of the published works: however, only a few studies examine the impact of ambient conditions on the RSSI and SNR such as [26] examines the temperature-RSSI correlation, whereas [35] assumes a 24-hour weather window to study the impact on the system's performance. The 1-day weather window is not sufficient to study the impact of different weather conditions on the channel performance. In addition to this, no comprehensive analysis of the effects of ambient and meteorological factors on LoRa/LoRaWAN channel performance indicators and how to optimize the configurable network parameters to reduce the impact of these changing environmental variables on network performance based on the forecasted weather conditions. To the best of our knowledge, there is not a single study that has studied the use of deep learning (DL) tools for predicting channel performance and finding the optimal LoRaWAN network configuration in different operational conditions based on the correlations among different networks and weather parameters. Since LoRaWAN is being deployed in a variety of environments around the world, understanding how environmental parameters impact data transmission is very important.

To address the above research gap, we have developed a DL model to find the correlations between the channel performance indicators and the environmental conditions that will enable better LoRaWAN device settings based on the measured RSSI and SNR values. LoRaWAN receivers may be immediately adjusted following weather station information or predicted RSSI and SNR without waiting for degradation or improvement in RSSI and SNR on the receiver side. Furthermore, due to dynamic ambient conditions at the network location, the developed model can facilitate the selection of optimal transmission parameters at the LoRaWAN end device when signals are disrupted by unfavorable weather.

III. METHODOLOGY

A. The Multivariate Time Series Problem Statement

The term time series refers to a collection of observations made over a span of time. In time series forecasting, prior data of a signal are used to estimate data for a subsequent time series (UTS forecasting) or previous data of multiple correlated signals (MTS forecasting). Consider $m_t \in Q^L$ signify the multivariate variable value of L dimension at time step t, and $m_t[j] \in Q$ symbolize the j^{th} variable value. Provided a series of historic K time steps of multivariate variable measurements, $M = \{m_{t_1}, m_{t_2}, \dots, m_{t_K}\}$, we aim to estimate the P-step in the future value of $N = m_{t_{K+1}}, m_{t_{K+2}}, \dots, m_{t_{K+P}}$. In general, the input variables can be combined with auxiliary characteristics like the time of the day, the date of the month, and the year. In the concatenation of input variables with the supporting characteristics, we consider the inputs to become $\mathcal{M} = \{S_{t_1}, S_{t_2}, \dots, S_{t_K}\}$ where $S_{t_i} \in Q^{L \times D}$ and D represents a dimension of features. By minimizing the mean square error with hyper-parameter optimization, we have developed a Bayesian Surrogate Gaussian Process model for multivariate multi-step time series forecasting.

B. Adaptive Data Rate Scheme

ADR is one of the vital features of the LoRaWAN specifications. This connection protocol aims to deliver fairly reliable and battery-friendly connections by adjusting SF and T_P based on changes in link conditions. End devices (EDs), as well as the network both, play a vital role in this process. The ED thinks a connection has been lost when it observes a sudden increase in uplink broadcasts without a downstream response from the network. The solution to the problem involves incrementally increasing TP before doing the same for SF. This procedure gradually improves the link's resilience. According to the LoRaWAN Specifications v1.1. If no downlink response is received, the ADR_ACK_LIMIT and ADR_ACK_DELAY settings determine how many uplink messages are sent before any ED must increase either TP or SF. In addition to the values of these parameters, the network size, deployment environment, and the degree of link fluctuation determine how long it takes ED to reach a condition where it can successfully re-establish a reliable link to the network.

A stable connection to the network is established by changing the EDs' communication settings, although not always energy-efficiently. An ED can, however, request that the network intervene and monitor the quality of recent uplink reception. As soon as the link quality measured over the last N packets exceeds the minimal receiver sensitivity threshold, the network cuts SF and/or T_P. By choosing new SF and T_P values, its objective is to make sure that the predicted SNR of future packets is above the receiver's minimum sensitivity threshold by a predetermined margin. Faster (higher data rates) and more energy-efficient transmissions would be possible with a reduction in SF and T_P. According to Semtech [33], the company that acquired LoRa in May 2012, the ADR algorithm can be implemented efficiently on the network server side, which is used by the Things Network, an open-source LoRaWAN network popular with crowdsourcing. However, due to shadowing, fading, bad ambient conditions, or movement, the ED and gateway (GW) channel conditions can be unstable. A significant volume of packets is lost due to this circumstance, increasing the ADR's convergence time. Therefore, we study the different ambient conditions affecting the performance of the LoRaWAN network and proposed an RNN-based deep learning model to find the correlation between ambient conditions and channel performance indicators as well as the prediction of RSSI and SNR of different EDs at the gateway.

C. Deep Learning Model Development Preliminaries

In deep learning, dense layers are sequentially stacked with nonlinear processing units. Separate layers process data, extract and classify features, convert features, interpret data, and predict. A certain amount of abstraction and interpretation occurs at every layer. The parameters and weights of the layers of the network are trained through backpropagation. Each layer receives the data from earlier layers and forwards it to the next layer after conducting various processing or training activities. Every deep network contains input and output layers. Initially, the input layer takes the data and then the output layer eventually produces the output.

Long-short-term memory (LSTM) and gradient recurrent unit (GRU) are interesting variants of the recurrent neural network architecture, an approach used to model sequential data under recursive neural networks. Among the recursive properties of an RNN that influence the explanation of dynamic performance is the network delay recursion. Moreover, RNNs can store a vector of activation functions for each sampling interval, making them very deep neural networks. RNNs are generally

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: Southeast University. Downloaded on February 19,2023 at 08:24:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2023.3238013

difficult to train to understand time series data owing to exploding and vanishing gradient problems [16] [36, 37]. LSTM and GRU are designed to solve the vanishing gradient problem. LSTM models feature memory cells capable of storing long-term correlations. The LSTM memory cell is made up of 4 gates that guide the interactions between distinct memory units. Input gates control whether a signal can change the state of a memory cell. The output gate, on the other hand, determines whether or not another memory cell's state may be changed. The forget gate, on the other hand, offers the option to forget (or remember) the previous condition. The LSTM equations are as follows:

$$I_t = \sigma(S_I Y_t + R_I H_{t-1} + T_I C_{t-1}) \quad (7)$$
$$F_t = \sigma(S_F Y_t + R_F H_{t-1} + T_F C_{t-1} \quad (8)$$
$$Z_t = \sigma(S_Z Y_t + R_Z H_{t-1} + T_Z C_t \quad (9)$$
$$\tilde{C}_t = tanh(S_C Y_t + R_C H_{t-1}) \quad (10)$$
$$C_t = F_t^l \odot C_{t-1} + I_t \odot \tilde{C}_t \quad (11)$$
$$H_t = Z_t \odot \tanh(C_t) \quad (12)$$

where I_t, F_t, and Z_t represent the input, forget, and output gates, respectively. Y_t, H_t, and C_t are used to denote the input state, hidden state, and memory cell state, respectively. In addition, (S_I, S_F, S_Z), (R_I, R_F, R_Z), and (T_I, T_F, T_Z) denoted the recurrent weights, input weights, and biases.

The GRU algorithm is an advanced version of LSTM. The architecture of the network uses gated recurrent units as gates to regulate memory resets and updates. In comparison to LSTM, GRU achieves comparable performance with fewer parameters and is, therefore, easier to train [38]. It only gates employed by the GRU are the update and reset gates. The GRU transition equations are expressed as follows:

$$H_t = (1 - u_t) \odot H_{t-1} + u_t \odot \tilde{H}_t \quad (13)$$
$$\tilde{H}_t = tanh(S_H y_t + R_H(v_t \odot H_{t-1}) + T_H) \quad (14)$$
$$u_t = \sigma(S_u y_t + R_u H_{t-1} + T_u) \quad (15)$$
$$v_t = \sigma(S_v y_t + R_v H_{t-1} + T_v) \quad (16)$$

where $H_t$represents the current output state and H_{t-1} represent the current prior output states. The reset and update gates are specified by v_t and u_t, respectively.

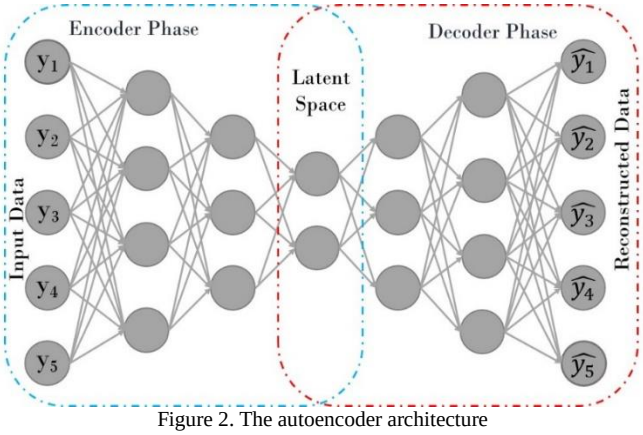


Figure 2. The autoencoder architecture

Autoencoder (AE) is an unsupervised artificial neural network that learns optimal data coding. A latent space representation is constructed by first transforming data from the input layer into a latent space representation (encoder phase), and then reconstructing the output of that reduced representation to the original input layer as closely as possible (decoder phase) as represented in Figure 2. The two phases of the autoencoder can be formulated by the mathematical equations as follows:

$$H(y) = f(S_1 y + T_1 \quad (17)$$
$$\hat{y} = g(S_2 H(y) + T_2 \quad (18)$$

where $H(y)$ denotes the hidden vector derived using input vector y on the encoder side and $\hat{y}$ is the reconstruction vector on the decoder side. Furthermore, f and g denoted the encoding and decoding functions, respectively. In addition, S_1 and S_2 represent weight metrics and T_1 and T_2 are the bias vectors on each phase, respectively.

The reconstruction error $|| y - \hat{y} ||$ refers to the difference between a reconstructed output and the original input. A reconstruction error is simply the difference between a reconstructed input (or output) and the original input (or output). A model's objective function represents this reconstruction error, which is aimed at being minimized during training. Multiple AE layers can be stacked to ensure desired high-level properties like abstraction and invariance. Therefore, the error reconstruction will be smaller, and as a result, an improved generalization is expected. An autoencoder can be built using different gated activation functions such as gated LSTM or gated GRU, or a combination of both functions.

D. Bayesian Optimization Algorithm

Bayesian optimization algorithm (BayesOpt) is a method for optimizing the parameters of any target function $P(g)$ sequentially. In BayesOpt, an objective function is modeled by a surrogate model which acquires all previous and examined data, and a decision process is used to select where the next sample should be taken. The model can be modified recursively once the results are available $O_t = g(I_t)$:

$$P(g(I_{1:t}, O_{1:t} = \frac{P(i_t O_t|g)P(g|I_{1:t-1}, O_{1:t-1})}{P(i_t O_t)} \quad (19)$$

For all t = 2 …T where input and output metrics are represented by $I_{1:t} = [i_1, i_2, \ldots, i_t]$ and $O_{1:t} = [o_1, o_2, \ldots, o_t]^T$. This strategy ensures that the data is always up to date.

BayesOpt calculates the optimum decision d of choosing the next trail $d = d_{t+1}$ by minimizing or maximizing a loss function

$$d(bayesOpt) = \arg d_{min} \int \delta_t(g, d) dP(g|I_{1:t}, Y_{1:t}) \quad (20)$$

where the optimality criterion that guides the optimization to the optimum I^* is denoted by $\delta_t(g, d)$.

TABLE 1
THE DATASET INPUT AND OUTPUT PARAMETER RANGE

Input Parameters	Count	Mean	Mini	Max
Unix timestamp	$2.029e^{+03}$	$1.609e^{+09}$	$1.605e^{+09}$	$1.612e^{+09}$
Temperature(^{0}C)	2029	4.571	-4.30	15.4
Humidity (%)	2029	75.695	1.0	86.8
Bar (hPa)	2029	1014.163	985.6	1035.2
Rain (mm)	2029	0.159	0.0	15.5
RSSI_01(dBm)	2029	-55.876	-58.2	-51.9
RSSI_02(dBm)	2029	-80.77	-92.7	-71.2
RSSI_03(dBm)	2029	-67.326	-70.2	-63.7
RSSI_04(dBm)	2029	-77.367	-84.4	-71.0
RSSI_05(dBm)	2029	-71.173	-81.8	-62.0
RSSI_06(dBm)	2029	-63.224	-80.8	-53.5
RSSI_07(dBm)	2029	-61.810	-67.4	-58.3
RSSI_08(dBm)	1879	-83.224	-92.6	-69.6
SNR_01(dB)	2029	9.127	8.4	9.8
SNR_02(dB)	2029	8.621	6.3	9.5
SNR_03(dB)	2029	8.54	6.4	9.4
SNR_04(dB)	2029	8.659	7.5	9.2

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: Southeast University. Downloaded on February 19,2023 at 08:24:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2023.3238013

SNR_05(dB)	2029	8.758	7.2	9.3
SNR_06(dB)	2029	8.821	6.8	9.5
SNR_07(dB)	2029	8.442	1.5	9.1
SNR_08(dB)	1879	8.203	6.8	9.0

E. Experimental Setup and Data Collection

A total of 19000 raw experimental data points from eight LoRaWAN sensor nodes and a LoRaWAN gateway were taken from the available open access literature [39]. The data of the LoRaWAN channel performance indicators such as RSSI and SNR for over 4 months from 16 September 2020 to 9 February 2021 were acquired through an experimental setup, which consists of eight LoRaWAN nodes (Tinovi PM-IO-5-SM) factory calibrated with soil temperature, moisture, and conductivity sensors but the current study does not use measurements from these sensors and MikroTik wAP LR8 gateway containing a pre-installed with UDP packet forwarded, 2.4 GHz WLAN, and an Ethernet port employed as backend. The SNR and RSSI parameter data were collected for each of the eight LoRaWAN nodes. Davis Vantage Pro2 weather stations were used to collect the data on meteorological parameters such as temperature (^{o}C), pressure (hPa), relative humidity (%), and rain (mm) for this study.

For an easier understanding of the data distribution used in this study, Table 1 summarizes the counts, mean, minimum, and maximum values. The table illustrates that RSSI_08 and SNR_08 have lower data counts than other parameters. The data contains some ‘NaN’ (not a number) values that produce an invalid input error if not removed from the data. Having data with ‘NaN’ alongside it is quite impossible to process, and manually changing 'NaN' to some specified values is not quite practical. To solve this problem, we calculated the mean of these parameters by using mean () and then employed the Pandas *fillna()* method to replace the ‘NaN’ values with the calculated mean values. Specifically, the Unix timestamp was also converted into *datetime* format using the *datetime.fromtimestamp()* of the *datetime* module to better understand the experiment period, a Unix timestamps are difficult to interpret. The following Table 2 contains information about the experiment system configuration and parameters.

TABLE 2
EXPERIMENTAL SYSTEM PARAMETERS

Parameters	Values	Parameter	Values
Transmission	Uplink and Downlink	Activation Methods	OTAA or ABP
Acknowledgment	Yes	Carrier Frequency	868MHz
CR	4/5	Deployment Area	400 m x 30 m
SF	7	LoRaWAN Nodes	8
BW	125 kHz	Gateways	1
Packet Size	26-byte UL and 5 bytes DL	Weather Station	1

F. Bayesian Surrogate Gaussian Process Bidirectional LSTM Stacked Autoencoder (BSGP-BLSTM-SAE)

As a variant of the autoencoder's fundamental structure, we propose a novel architecture to enable it to predict the features from MTS, or time series problems in general. This bidirectional LSTM-based stacked autoencoder (BLSTM-SAE) is a modified version of the original feed-forward neural network-based autoencoder (see Figure 2) that has been renamed from the feed-forward neural network-based autoencoder. Recurrent networks are better at modeling time series data, which is why Bayesian surrogate Gaussian process bidirectional LSTM-SAE was developed.

TABLE 3
RANGE OF DIFFERENT HYPER-PARAMETERS

MLRA	Hyper-parameters	Range of Hyper-parameter Values	Hyper-parameters	Range of Hyper-parameter Values
GP GBRT RF	Learning Rate	0.0001 to 0.1	Dropout Range %	0.1 to 0.9
	Encoder Layers	2 to 5	RNN Units	LSTM, Bidirectional, GRU, and Simple RNN
	Encoder Dense RNN Units	1 to 500	Activation and Recurrent Functions	ReLU, tanh, sigmoid, hard-sigmoid, and linear
	Decoder Layers	2 to 5	Initialization	Uniform, Lecun uniform, glorot_normal, glorot_uniform, he_normal, he_uniform
	Decoder Dense RNN Units	1 to 500	Optimizer	Adam, RMS Prop, SGD, and Adamax

The BayesOpt scheme was introduced for the developed model to optimize the hyper-parameters automatically. To obtain the optimum hyperparameter combinations, three machine learning regression algorithms (such as GP, GBRT, and RF) with different acquisition functions (like EI and PI) are employed. The target hyper-parameters selected for the tuning of the machine learning regression algorithms (MLRA) are the number of hidden layers and dense nodes on both the encoder and decoder sides, learning rate, batch size, activation function, initialization functions, and epochs. The range of investigated hyper-parameters for all three developed models is illustrated in Table 3. This research investigates a spectrum of possible hyper-parameter values to analyze and pick the best solution for the RNN model to be constructed.

Data is divided between training, validation, and testing sets during the data processing process using the concept of cross-validation. Before performing the model training and validation operations, the data needs to be pre-processed. In this study, this stage comprises the feature transformation and normalization like we have converted the Unix timestamp into the date-time format and the Min-Max scalar is used to scale each feature to the 0 and 1 range. Thus, the process involves converting time series data to the relevant data formats.

A cross-validation process usually uses 60% of a data set as training data, 20% as validation, and the next 20% as testing data. In the BayesOpt, the mean absolute error (MAE) of the validation set is used as an iterative indicator to analyze the developed model. After that, the training and validation data sets will be used as a new training set for the model with optimum mapping hyper-parameters to predict the results. The

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: Southeast University. Downloaded on February 19,2023 at 08:24:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2023.3238013

Bayesian optimization process with GP, GBRT, and RF to optimize the hyper-parameters of mapping relation employed in this study can be illustrated by a flow chart as shown in Figure 3.

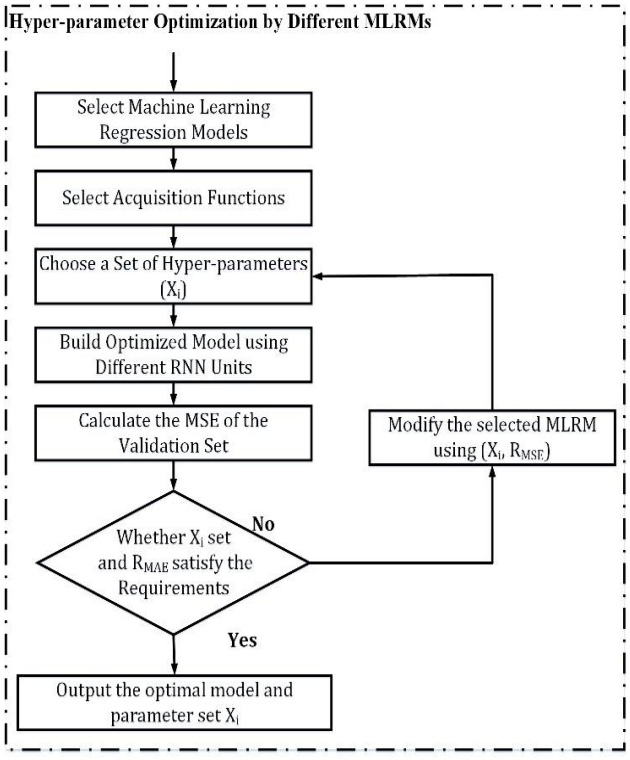


Figure 3. MLRM hyper-parameter optimization scheme

Three optimized models were developed from the Bayesian optimization process including GP, GBRT, and RF MLRM. The MAE of all the developed models was around 0.45. The convergence plot represents how the time step evolves during the solution process. In Figure 4, convergence plots for 3 different hyper-parameters optimization algorithms are illustrated. The MAE value changes to the number of iterations for the developed 3 optimization models are shown in Figure 4. It can be seen from the plots that the GBRT and GP algorithm converges quickly having a convergence range of 8-23. While as RF algorithm takes the longest number of iterations to converge (85 iterations).

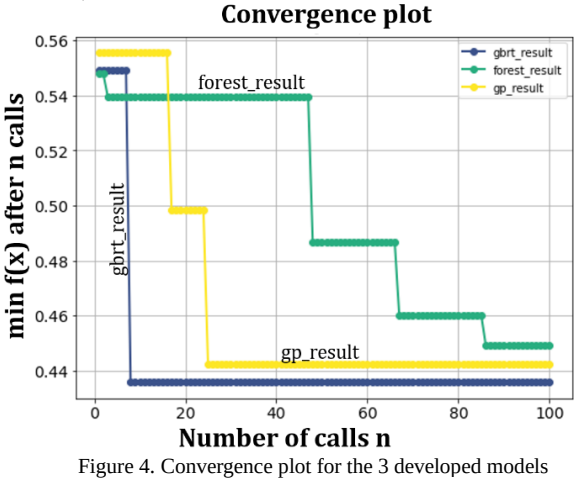


Figure 4. Convergence plot for the 3 developed models

The optimal bidirectional LSTM-based stacked autoencoder model selected for the forecasting of LoRaWAN channel performance indicators has 2 and 3 encoder and decoder layers and 324 and 275 encoder and decoder dense bidirectional units respectively as shown in Figure 5. Tanh, sigmoid, uniform, and Adamax were preferred as suitable activation, recurrent activation, initialization, and optimizer. We found that the best learning rate, batch size, and decay rate for the optimal model were 0.0013, 616, and 0.8, respectively. Ten data batches were used with a sampling window size of 10 during the training and validation of the developed model. The required accuracy for the optimal model was attained after 46 iterations but we have let the model run for maximum iterations to analyze the trend of the model.

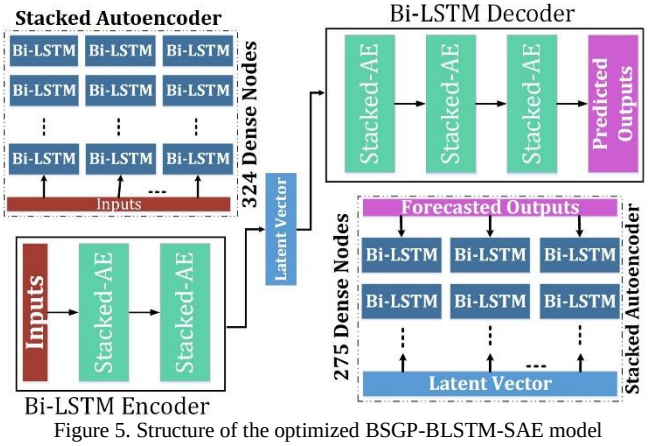


Figure 5. Structure of the optimized BSGP-BLSTM-SAE model

G. Model Complexity Analysis

A forward pass is normally used to test the complexity of a neural network. Essentially, the forward pass time complexity can be used to calculate the entire model's time complexity. Thus, the specifics of how a feedforward neural network without recurrent layers calculates the temporal complexity of its forward pass. An LSTM-layered neural network must also take into account recurrent connections in the LSTM units. An LSTM layer's forward pass can be calculated using the equations given above (Eqns. 7 to 12).

To determine the model's complexity, let's dissect these operations. Dimensions and object types (vectors and matrices) are critical. As a first step, we'll calculate I_t (the output of the input gate). As far as temporal complexity is concerned, F_t (The forget gate), Z_t (the cell), and $\tilde{C}_t$ (the output gate) have the same calculation.

From Equation (7), $S_I \in M^{H\times d}$ is the matrix used to denote the input gate forward connections, $Y_t \in M^d$ represents the input vector with the d feature, $R_I \in M^{H\times H}$ is the matrix used to denote the recurrent connections having different dimensionality than S_I, $H_{t-1} \in M^H$ are the hidden states, and $C_t \in M^H$ denotes the memory cell states. A time complexity estimate for operations inside the non-linearity is given below. $S_I Y_t$ and $R_I H_{t-1}$ are matrix and vector multiplications having a time complexity of $O(Hd)$ and $O(H^2)$ respectively. Whereas the $O(H)$ is the complexity of each of $S_I Y_t$ and $R_I H_{t-1}$ with bias.

For $N = \max(d, H)$, then the time complexity of the operations inside the non-linearity is $O(HN)$. Based on what activation function σ is, its time complexity can be determined, the σ for the proposed model is a $\tanh$ and $\sigma(x) = \tanh(x)$ operation is performed. Assuming $\tanh(x)$ has constant time complexity, then applying the tanh element-wise has a time complexity H.

Hence, I_t computation has time complexity $O(Hd)$. It's the

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: Southeast University. Downloaded on February 19,2023 at 08:24:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2023.3238013

same for F_t, Z_t, and $\tilde{C}_t$, so it multiplies the computational complexity of I_t by 4, but constant factors are not taken into account.

From Equation (11), $C_t = F_t^I \odot C_{t-1} + I_t \odot \tilde{C}_t$ where $C_t \in M^H$ is a cell vector, and $\odot$ denotes element-wise multiplication, so this operation has a time complexity of $O(2H)$. Similarly, Equation (12) has a time complexity of $O(2H)$, h for each tanh, and element-wise multiplication operations. Thus, a single LSTM layer's forward pass has an $O(4Hd)$ time complexity.

The autoencoder used in the development of the proposed model has 7 LSTM layers including 3 layers of each encoding and decoding side. Therefore, if L represents the number of LSTM layers in a single AE, then AE has the time complexity of $O(L(4Hd))$. Consequently, the overall time complexity of the proposed model can be calculated by multiplying the $O(L(4Hd))$ by the number of AE's (N_{AE}) in the stacked AE Model. Thus, the overall time complexity of the BSGP-BLSTM-SAE model is $O(N_{AE}L(4Hd))$.

The time taken by Bayesian optimization is for finding the optimal hyperparameters during the training and validation process on the collected multivariate time series dataset of 19000 data points performed before the actual deployment in the LoRaWAN networks. Moreover, only the optimized model will be used for real-time deployment which only takes only 121 milliseconds for a given sampling window size. This shows that the proposed model is suitable for real-time deployment scenarios.

IV. Results And Discussion

The performance of the model is evaluated by employing MAE between the original and predicted values of the ambient conditions and channel performance indicators. Figure 6 represents the loss curve achieved through iterative reduction of the mean absolute error for the training and the testing datasets. The losses associated with training and validation tend to decrease as the number of epochs increases, i.e. they decrease with epoch count increasing. Losses are converging around the same position, so a good fit is apparent. According to the preceding section, when the optimal parameters are used, the AI model can learn the data well. In this case, after 46 iterations, the AI model nears a mean absolute error value of zero; after that, the error value remains nearly constant until all 100 iterations, when the repeat loop is stopped, which shows that the model was not overfitting.

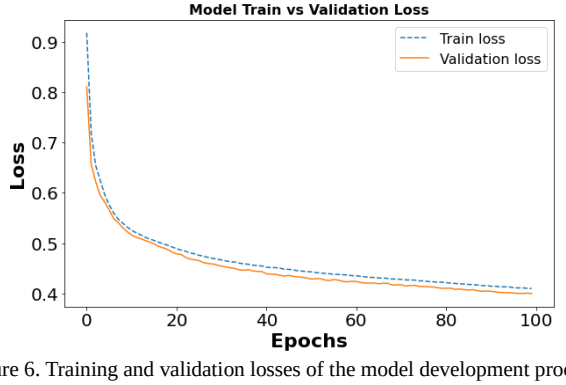


Figure 6. Training and validation losses of the model development process

The performance of the Bayesian Surrogate Gaussian process bidirectional LSTM stacked autoencoder is evaluated through the prediction results of two-channel performance indicators (SNR and RSSI) as well as four ambient weather parameters (temperature, pressure, humidity, and rain). Thus, this study aims to predict both the LoRaWAN channel performance metrics (RSSI and SNR) and ambient weather conditions precisely by using the proposed BSGP-BLSTM-SAE model.

A. Impact of Varying Ambient Parameters on RSSI and SNR

Based on the RSSI, SNR, and other ambient metrics collected over the 4 months, we can draw interesting conclusions about LoRaWAN transmission channel performance. There was indeed a wide variation in relative humidity during the data collection period, ranging from 1% to 86.8%. Air pressure was equally variable, with values ranging from 985.6 to 1035.2 hPa. The tests were conducted in the fall and winter, when temperatures were noticeably lower (i.e., -4.3^{0}C to 15.4^{0}C); it is generally understood that higher temperatures generate more thermal noise, and therefore, SNR values are less than those conducted in summer. Since 15.4 ^{0}C is considered the maximum possible temperature at the experiment site, this indicates the transmission channel was not evaluated under worst-case scenarios.

Results of the experiment show that the weather has an impact on transmission performances. Pearson correlation coefficient (r) [40] has been used to calculate the correlation between the channel performance metrics and weather conditions. An evaluation of Figure 7 illustrates how temperature affects LoRaWAN RSSI and SNR over the experiment time. A negative correlation between temperature, RSSI, and SNR is observed for the plotted period. As for RSSI change, the Pearson coefficients of temperature for LoRaWAN ED_01 and ED_04 are -0.5 and -0.43, respectively. As can be seen, with the rise in temperature signal strength (RSSI) decreases. The SNR values were also higher for low temperatures than for high temperatures but the magnitude of the impact differed from that of the RSSI. Similarly, the Pearson coefficients of temperature for ED_01 and ED_05 are -0.064 and -0.081, respectively. It is evident from the SNR and temperature plots that the SNR magnitude of the impact did not differ significantly between the various periods.

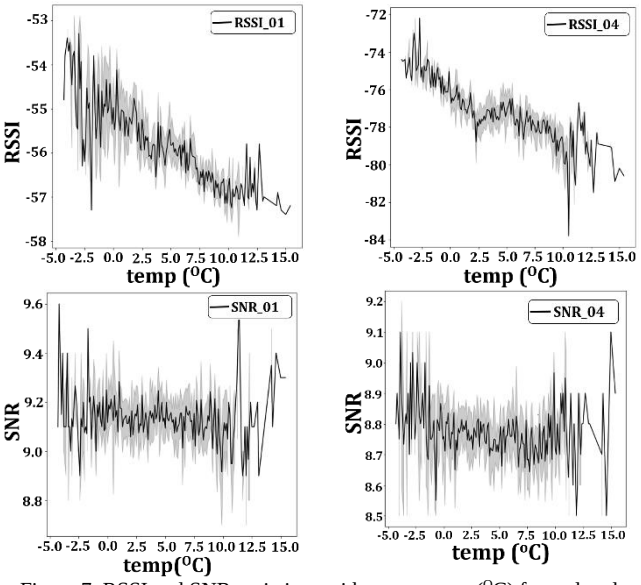


Figure 7. RSSI and SNR variations with temperature (^{O}C) for end nodes ED_01 and ED_04

Another factor that could influence signal strength is hu-

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: Southeast University. Downloaded on February 19,2023 at 08:24:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2023.3238013

midity. The RSSI and SNR signals were further investigated to determine their impact on humidity. Humidity is a measurement of the quantity of water vapor in the atmosphere. The relative humidity is defined as the difference between the quantity of water vapor in the air (absolute humidity) and the amount of water vapor that would be present in the air at the same pressure and temperature (saturated humidity), and it is expressed as a percentage (%). Figure 8 illustrates the relationship between relative humidity (RH) and signal strength over time. RSSI and relative humidity are positively correlated, meaning that RSSI and relative humidity are increasing and decreasing at the same time. For LoRaWAN ED_01 and ED_04, Pearson coefficients of humidity for RSSI are 0.26 and 0.24, respectively. While the humidity has also produced a very low positive correlation at ED_04 (r = 0.12) and ED_06 (r = 0.3).

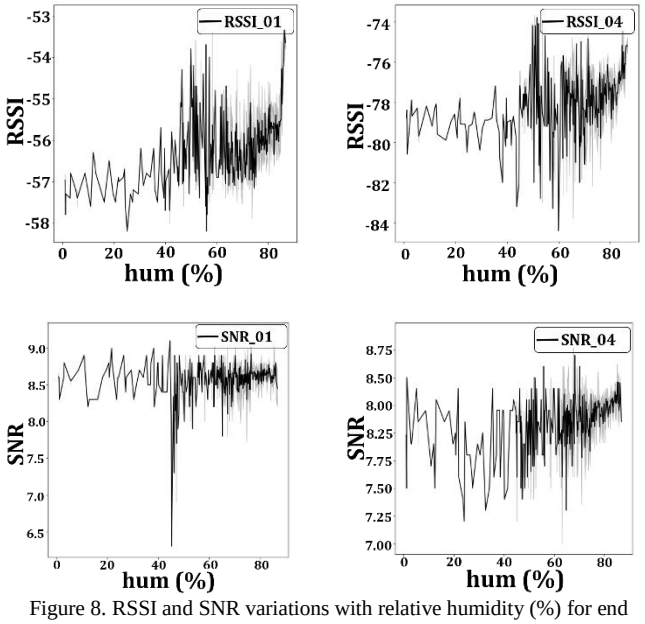
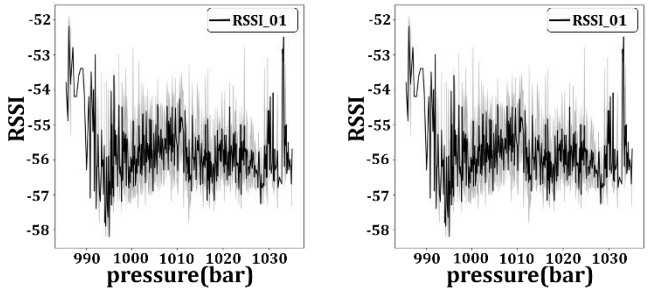


Figure 8. RSSI and SNR variations with relative humidity (%) for end nodes ED_01 and ED_04

The impact of the pressure can be observed in Figure 9. From Figure 9, atmospheric pressure (hPa) has only a minor role, since no significant variation in performance can be observed, except for a slight decline of RSSI values at high pressures. This is also confirmed by the calculated negative values of Pearson coefficients at given nodes for both RSSI (ED_01: -0.85, ED_04: r = -0.34) and SNR (ED_01: -0.011, ED_04: r = -0.098). Pressure has also shown a very low positive correlation at nodes ED_03 and ED_08 for both RSSI and SNR.



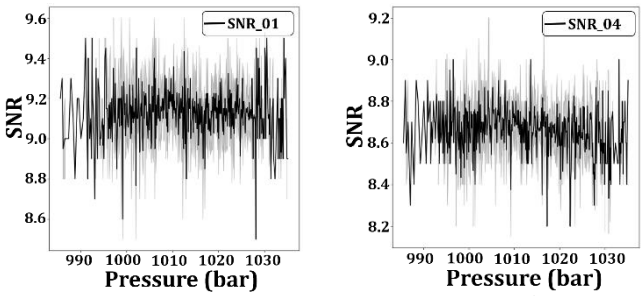


Figure 9. RSSI and SNR variations with pressure/bar (hPa) for end nodes ED_01 and ED_04

Radio transmission is affected by rain attenuation, particularly in the tropical and equatorial regions having a high level of precipitation. Frequency, path-averaged rainfall rate, raindrop size distribution (DSD), and temperature are the factors that contribute to rain attenuation. Attenuation due to rain can be expressed as cR_t^d, where R_t represents the rain rate over time t, while c and d are related to the frequency of radio waves and the temperature of the rain, respectively. In particular, at frequencies above 10GHz, DSD is recognized as the main cause of signal degradation and propagation attenuation by producing scattering and attenuation of electromagnetic waves. However, for the current study, the rain varies from 0 to 15.5 mm for a very short period. It can be seen from Figure 10, that the performance metrics (RSSI and SNR) trends are the same for the given nodes during the rainfall. Moreover, similar results are observed at all other nodes. Thus, the overall impact of the rain on the current experiment is negligible. As can be confirmed by the calculated Pearson coefficients of rain for RSSI and SNR for given nodes (ED_01: 0.065, ED_04: r = 0.29) and SNR (ED_01: -0.013, ED_04: r = 0.028).

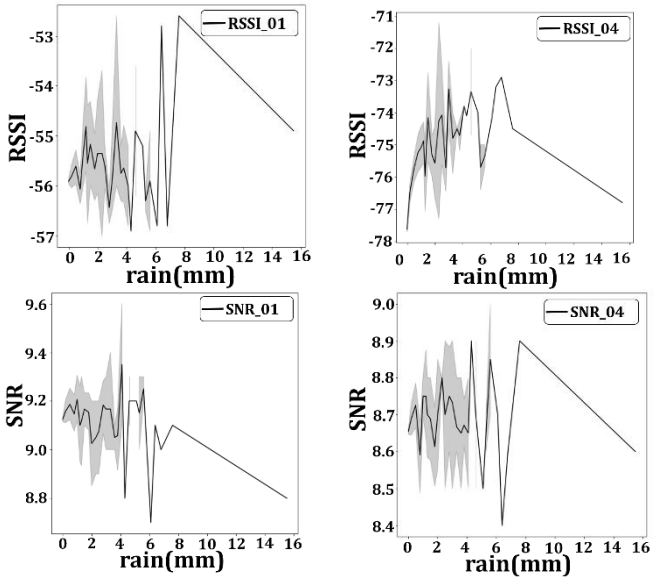


Figure 10. RSSI and SNR variations with rain (mm) for end nodes ED_01 and ED_04

When comparing the RSSI and SNR correlation results given in Figure 7 to 10, it is observed that temperature and humidity have the most impact on these channel performance indicators during the period of study. SNR and RSSI for gateways with shorter air distances from sensor nodes performed better compared to gateways with longer air distances. Additionally, the mean values of SNR and RSSI decreased during

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: Southeast University. Downloaded on February 19,2023 at 08:24:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2023.3238013

rains, which indicates that the quality of the links has deteriorated while the standard deviation increase deters the network from being resilient. However, due to a bit of dry weather during the experiment period, the rain has no impact on the channel performance metrics in the current study.

The correlation plots from Figure 7 to Figure 10 represent the X and Y pair of variables such as channel performance indicators and ambient conditions. For illustration, the correlation coefficients (r) of temperature and SNR are -0.064 and -0.081, respectively for end nodes (ED) ED_01 and ED_04. These are not perfect correlations, but remember that there are many other factors besides temperature that can affect SNR, such as transmission power, coding rate, bandwidth, and even other ambient conditions like humidity, pressure, and rain. A variable might be a weak, but significant predictor if it is just one of many factors (parameters) that determine the outcome (Y). Whether the temperature is a statistically significant predictor of SNR depends on both the strength of the correlation coefficient and the number of observations (n).

The statistical significance of the Pearson correlation coefficient of channel performance indicators with the ambient conditions presented in this section is evaluated based on the following equation: $SS_{obs} = \frac{r}{\sqrt{1-r^2/n-1}}$, $\alpha = 0.05$, and SS^*_{obs} where SS_{obs} is the absolute significance test value, $df = n - 1$ is referred to as the degree of freedom, α is the significance level, and SS^*_{obs} is the critical value calculated using the software based on the above variables. If the $SS^*_{obs} < \alpha$, then the correlation is statistically significant and vice versa. Based on the statistical significance evaluation of RSSI and Temperature correlation have SS^*_{obs} value of 0.00001 at both ED_01 and ED_04. Similarly, at ED_01 and ED_04, SS^*_{obs} values for SNR and Temperature correlation are 0.00602 and 0.00026, respectively. For both ED_01 and ED_04, SS^*_{obs} value for RSSI and Humidity correlation is 0.00001. Similarly, the SS^*_{obs} values for SNR and Humidity correlation coefficients are 0.00001 at both ED_01 and ED_04. Moreover, SS^*_{obs} value for both RSSI and SNR vs pressure correlations is 0.00001. Thus, the correlation results presented in Figure 7 to Figure 10 are also statistically significant. However, the correlation coefficient of rain with SNR is not statistically significant $SS^*_{obs} > \alpha$. This might be due to very less rain (0 to 15.5 mm) for a very short time during this experimental period.

B. Predictions of Channel Performance Indicators

To evaluate the performance of the proposed model for the estimation of the channel performance indicators using the multivariate time series data, we have plotted the prediction and the calculated values (test data) of the performance metrics (RSSI and SNR) for all eight LoRaWAN nodes. The actual and the prediction values of RSSI of 6 nodes are represented in Figure 11. To simplify the presentation, we have shown the data for the first 100 hours from the test and prediction data sets in the line graphs. These channel performance indicators (RSSI and SNR), which are input parameters that are used to develop the developed model, are the most influential channel performance indicators for wireless communication networks. Plotting the experimental and estimated RSSI and SNR with time (hours) enabled the evaluation of the predictive model's performance.

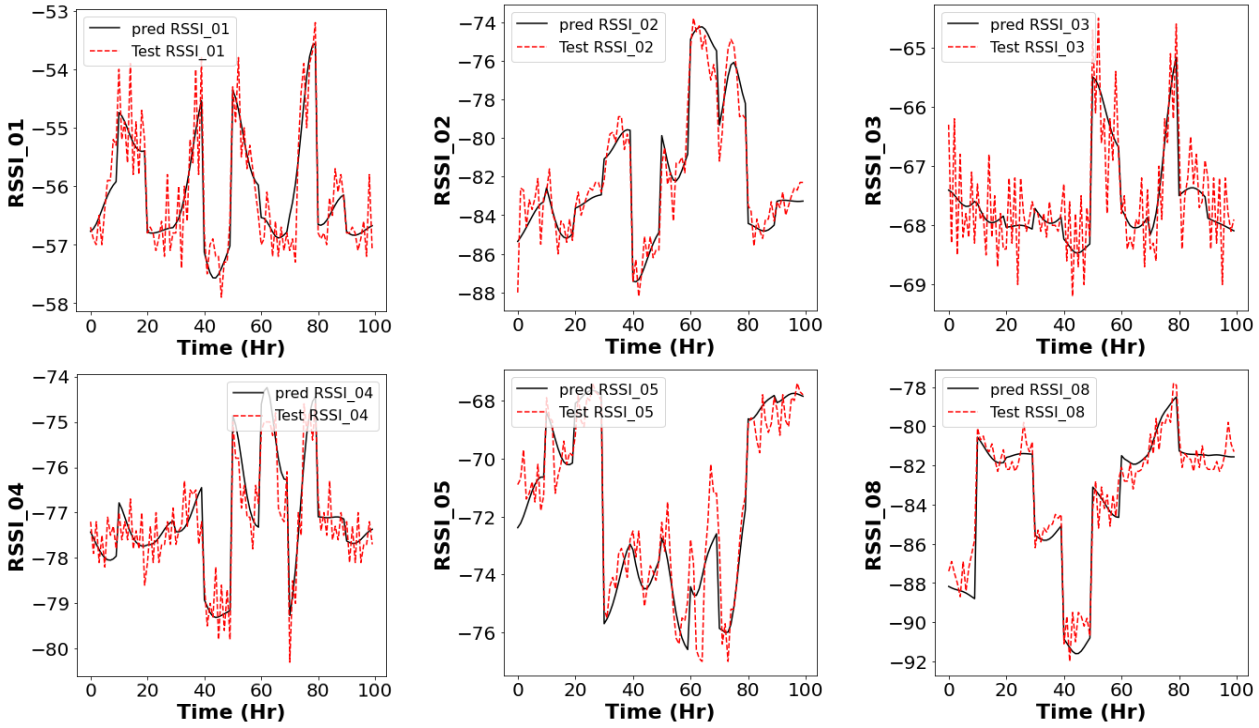


Figure 11. Prediction and test values of RSSI at 6 LoRaWAN nodes for given weather conditions

From Figure 11, it can be observed that the proposed model, fits the prediction and actual RSSI values most accurately. The model provides an MAE of 0.44 for both test and validation datasets. The same channel performance indicator RSSI test and validation trend have been observed for the other 2 end nodes. However, it can be observed that there are different RSSI values for each LoRaWAN node at the same ambient conditions for the experiment period. This might be due to the distance between the LoRaWAN nodes and the gateway, and the shadowing effects due to different obstructions in the communication chan-

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: Southeast University. Downloaded on February 19,2023 at 08:24:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2023.3238013

nel path. Depending upon the distance from the gateway, the transmitted signal from a particular node has different TOAs. Moreover, the SNR prediction performance of the proposed model is presented in Figure 12. However, from the line graphs in Figure 12, the proposed model can predict the SNR quite accurately. Based on the experimental data, the predictions for SNR for the 6 LoRaWAN end nodes show a similar trend. The same RSSI and SNR trend have been observed for the other LoRaWAN end nodes at different variable weather conditions in the current study. The proposed model can provide an accurate estimation of the channel performance metrics under different weather conditions.

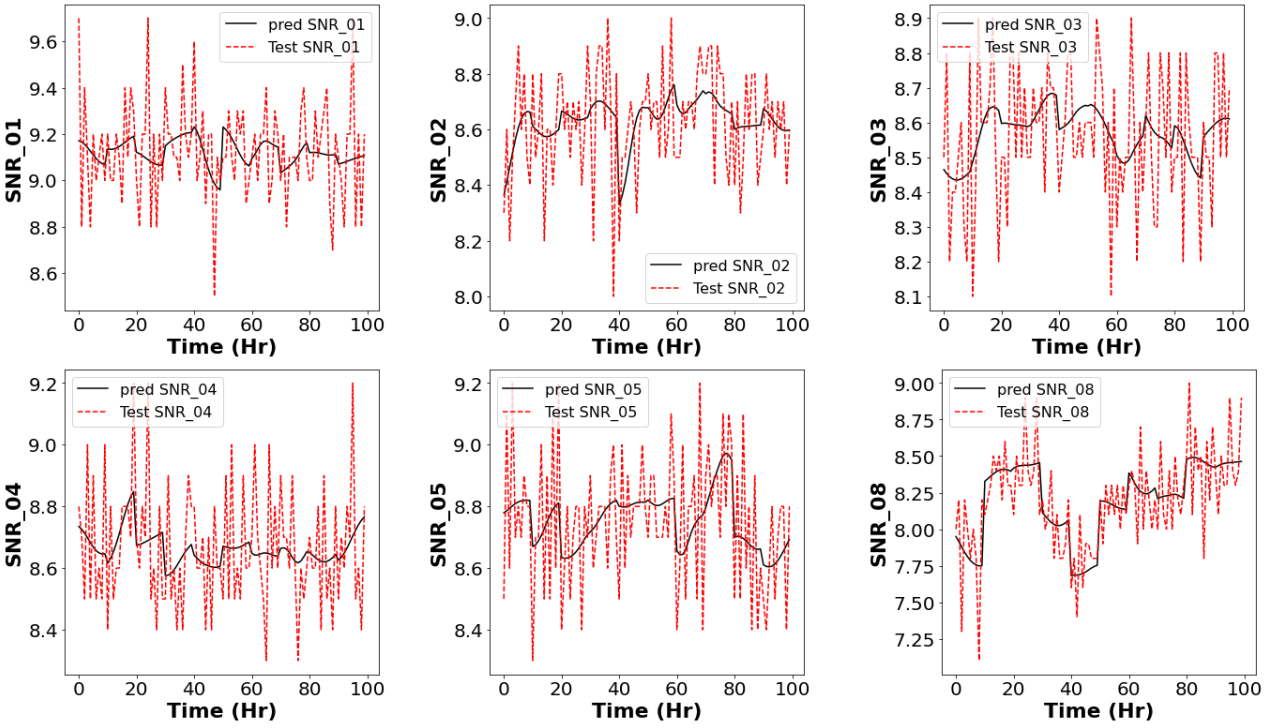
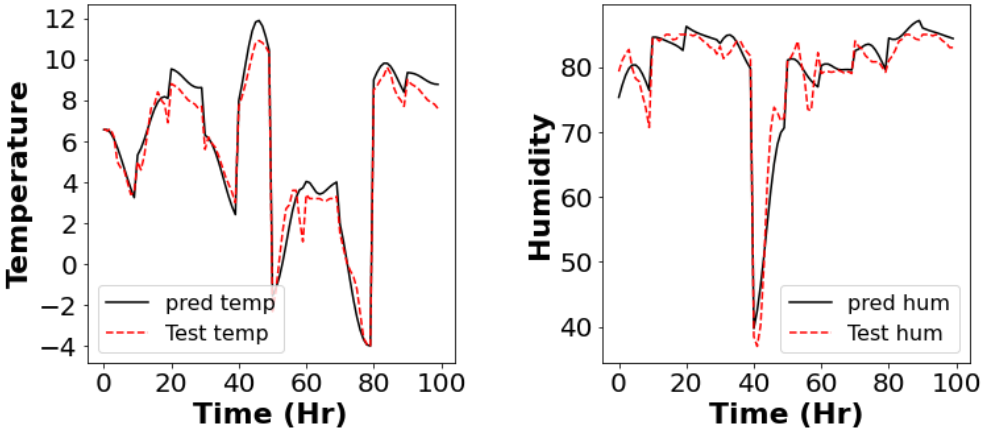


Figure 12. Prediction and test values of SNR at 6 LoRaWAN nodes for given weather conditions

C. Weather Estimation

Furthermore, ambient weather estimation was also done using the same model and experimental data. From Figure 13, it can be observed that the proposed model is also able to predict ambient conditions such as temperature, humidity, pressure, and rain with high accuracy. Based on the dataset of 100 hours, these results were obtained. Predicted and experimental data trends are well tracked with minimum deviations at some points, which could be explained by the limited dataset of 2026 data points. Thus, from the results, it can be concluded that the proposed model shows the ability and suitability for weather estimation.



© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: Southeast University. Downloaded on February 19,2023 at 08:24:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2023.3238013

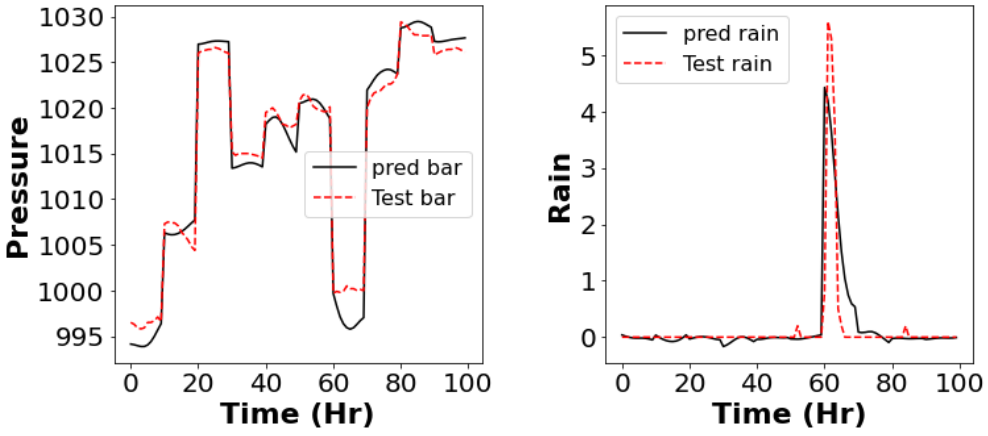


Figure 13. Test and prediction values of weather parameters (a) temperature, (b) humidity, (c) pressure, and (d) rain

Therefore, from the prediction results (cf. Figure 11 to 13), it can be observed that the proposed BSGP-BLSTM-SAE model performs effectively well for the multivariate time series. Thus, the results of the prediction confirm that the BSGP-BLSTM-SAE model can be accurately used for the estimation of both channel performance indicator problems and ambient weather parameters.

D. Discussion

The effect of ambient conditions on the channel performance indicators such as RSSI and SNR is mitigated by LoRaWAN using several configurable parameters such as SF, C_R, BW, and D_R that allow it to handle a wide range of use cases with varying operational conditions. These parameters are currently configured through the ADR algorithm [41] by regulating the end devices' D_R and P_T to improve the quality of service (QoS) and increase the node's battery life. While ADR's performance has been proven to be efficient and scalable, it ignores and does not consider the intricate correlation of the environmental parameters involved in its design. But ADR is significantly harmed by a highly unstable radio channel, like the existence of moving obstructions and variable weather conditions, which causes the SNR and RSSI values to fluctuate, resulting in configuration allocation problems [42]. As long as the channel variance is low, the fundamental ADR system is adequate, since it decreases interference in comparison to the static data rate. Using the highest RSSI from the past 20 packets does not provide a good indication of connection quality for a highly dynamic channel, like one affected by moving obstacles or changing weather. Furthermore, selecting the top 20 packets may take a long time for IoT applications that don't require a lot of uplink traffic. As for the transmission settings, they are adjusted based solely on the connection of one node. In the event of congestion, all LoRa nodes connected to this gateway will use the fastest data rate. As a result, the number of LoRa nodes using the same data rate will increase along with the likelihood of collisions.

The model developed in this study establishes the correlation between the ambient weather conditions and performance indicators such as RSSI and SNR for the LoRaWAN network and has a very high prediction accuracy for both ambient conditions and performance indicators. Therefore, the proposed artificial RNN based on bidirectional LSTM network units can be used to realize an optimum mode of ADR operation for network performance by predicting the LoRaWAN channel performance indicators based on the current ambient conditions. The forecasting quality of the model developed in this paper can be used to address the cases where an ED should adjust its transmission parameters based on the changes in current and past parameter values if the downlink control messages are lost. Furthermore, the developed model if integrated with ADR could enable the real-time configuration of LoRa transmission parameters to decrease the message ToA, maintain high SNR and RSSI values, and decrease the energy consumption in a variety of uses according to various operational conditions. With the proposed model, a large number of packets would not be required to be lost for the EDs to move to higher SF or T_P values. As a result, ADR will be able to determine the best SF and T_P configurations by adjusting the signal strength. Additionally, the proposed model can include the power associated with each packet, which can assist in choosing an appropriate SF for the ED, increasing the likelihood of packet loss falling below the SF sensitivity limits.

However, to establish the applicability of the developed model and incorporate it with ADR or develop it as a deep learning-based ADR, the developed model needs to be evaluated on new data sets with variable transmission parameters such as D_R, SF, and C_R along with link performance metrics (SNR, RSSI, and packet reception rate) and ambient weather conditions. The data set used in this study to evaluate the developed model has fixed transmission parameters. Using the variable transmission parameters will assist in finding the impact of different values of transmission parameters along with the ambient conditions on the channel performance indicators. And this can help to achieve the optimal configuration of LoRa transmission parameters for variable link conditions. In addition, the use of higher values of SF, P_T, ToA, and retransmission in ADR is to blame for excessive energy usage. The suggested model on the other hand can help in reducing the energy consumption at both the server and end node side by using appropriate parameter configurations.

Regardless of the impact of ambient conditions such as temperature, humidity, etc., on signal strength, it might have disastrous consequences for various sensor network protocols. The accuracy of RSSI-based range and localization suffers greatly when ambient conditions are not considered. So, when applying RSSI-based ranges, ambient factors need to be considered to adjust to current weather conditions. The RSSI and SNR differences that are caused by ambient conditions can be ac-

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: Southeast University. Downloaded on February 19,2023 at 08:24:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2023.3238013

counted for through our findings and thereby improve range accuracy. In SNR and RSSI-based ranging and localization, it may be better to use frequency diversity instead of a single channel and range of other user parameters based on the results of this developed RNN model.

V. Conclusion

This study contributes by developing surrogate models for forecasting the channel performance indicators of a LoRaWAN network based on ambient conditions with Bayesian surrogate models. The resulting RNN model describes the intricate correlation among the LoRaWAN parameters, underlining the importance of using artificial intelligence with the ADR mechanism in an adaptive manner. From both the prediction and test results, it is apparent that the model best fits the prediction and test data sets and the validation test data sets. The simulation results demonstrate that layer-wise pre-training is more suitable for multivariate time series problems when used with the bidirectional LSTM-based stacked autoencoder. Based on the forecasting capability of the model designed in this paper, it can be adapted to address the case in which an ED is to adjust its parameters autonomously based on the changes in current and past values of different parameters after downlink control signals are lost. As an experiment for testing the model developed for ADR design that makes better use of other LoRaWAN parameters, we will collect data on channel performance indicators and ambient conditions by varying the user parameters to evaluate how well the developed model performs. The prediction accuracy of the developed model assessed in terms of MAE is 0.44. However, the current model requires a longer training time to sequentially implement the pre-training and fine-parameter tuning phases. For future studies, it is recommended to combine appropriate mechanisms to reduce training time and increase its effectiveness.

Acknowledgment

This work is supported by the Ministry of Science and Technology (MOST 110-2622-8-007-009-TE2 and MOST 110-2923-E-007-008-) and the Ministry of Education (subsidy for talentCultivation in international competition—College of Electrical Engineering and Computer Science—5G-AloT Technology and Application Research Center) in Taiwan.

*Corresponding Authors: Showkat Ahmad Bhat (Showkatbhat1994@gmail.com), Imtiyaz Hussain (imtiyazkou@gmail.com), and Nen-Fu Huang (nfhuang@cs.nthu.edu.tw)

References

[1] N. C.-Romero et al., "Collision Avoidance Resource Allocation for LoRaWAN", Sensors, vol. 21, no. 4, pp. 1-19, 2021.

[2] H. Mroue et al., "LoRa+: An extension of LoRaWAN protocol to reduce infrastructure costs by improving the Quality of Service", Internet of Things, vol. 9, no. 2020. 9: p. 1-13.

[3] D. Bankov, E. Khorov, and A. Lyakhov, "LoRaWAN modeling and MCS allocation to satisfy heterogeneous QoS requirements", Sensors, vol. 19, no. 19, pp. 1-23, 2019.

[4] A. J. Wixted, et al., "Evaluation of LoRa and LoRaWAN for wireless sensor networks". IEEE SENSORS, IEEE, 2016, pp. 1-3.

[5] L. Parri et al., "Offshore LoRaWAN Networking: Transmission Performances Analysis under Different Environmental Conditions". IEEE Trans. on Instrumentation and Measurement, vol. 70, p. 1-10, 2020.

[6] J. Luomala and I. Hakala, "Effects of temperature and humidity on radio signal strength in outdoor wireless sensor networks". Fed. Conf. on Comp. Sci. and Inform. Syst. (FedCSIS), IEEE, 2015, pp. 1247–1255.

[7] C. A. Boano, K. Römer, and N. Tsiftes, "Mitigating the adverse effects of temperature on low-power wireless protocols". Int. Conf. on Mobile Ad Hoc and Sensor Syst., IEEE, 2014, pp. 336-344.

[8] C. A. Boano et al., "Low-power radio communication in industrial outdoor deployments: The impact of weather conditions and ATEX-compliance". Int. Conf. Sensor Apps, Exp. and Logistics. Springer, pp. 159-176, 2009.

[9] R. Marfievici et al. "How environmental factors impact outdoor wireless sensor networks: a case study". Int. Conf. Mobile ad-hoc and sensor Syst., IEEE, 2013, pp. 565-573.

[10] R. Zitouni et al., "IEEE 802.15. 4 transceivers for the 868/915 MHz band using Software Defined Radio". arXiv preprint arXiv:1304.8028, pp. 1-5, 2013.

[11] N. Benkahla et al., "Review and experimental evaluation of ADR enhancements for LoRaWAN networks". Telecomm. Syst., vol. 77, no. 1, pp. 1-22, 2021

[12] F. Cuomo, M. Campo, A. Caponi, G. Bianchi, G. Rossini, and P. Pisani, "Explora: Extending the performance of LoRa by suitable spreading factor allocations," in Proc. IEEE 13th Int. Conf. Wireless Mobile Comput. Netw. Commun. (WiMob), 2017, pp. 1–8.

[13] K. Q. Abdelfadeel, V. Cionca, and D. Pesch, "Fair adaptive data rate allocation and power control in LoRaWAN," in Proc. IEEE 19th Int. Symp. World Wireless Mobile Multimedia Netw. (WoWMoM), 2018, pp. 14–15.

[14] J. Park, K. Park, H. Bae, and C. K. Kim, "EARN Enhanced ADR with coding rate adaptation in LoRaWAN," IEEE Internet of Things Journal, no. 7, vol. 12, pp. 11873-11883, Dec. 2020.

[15] A. Farhad and J. Y. Pyun, "HADR: A Hybrid Adaptive Data Rate in LoRaWAN for Internet of Things," ICT Express, vol. 8, no. 2, pp. 1-7, 2020.

[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory". Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[17] N. Jeftenić, M. Simić, and Z. Stamenković, "Impact of Environmental Parameters on SNR and RSS in LoRaWAN". Int. Conf. on Elect. Comm., and Compt. Eng. (ICECCE), IEEE, pp.1-6, 2020.

[18] R. F. Hope, Microelectronics, RFM95/96/97/98 (W)-Low Power Long Range Transceiver Module. China, 2016.

[19] T. Ameloot, P. V. Torre, H. Rogier, "Variable link performance due to weather effects in a long-range, low-power Lora sensor network," Sensors, vol. 21, no. 9, pp. 1-22, Jan-2021.

[20] K. Bannister, G. Giorgetti, and S. Gupta, "Wireless sensor networking for hot applications: Effects of temperature on signal strength, data collection, and localization". Proc. workshop on embedded networked sensors (HotEmNets'), 2008, pp. 1-5.

[21] N. S. Bezerra et al., "Temperature impact in LoRaWAN—A case study in Northern Sweden". Sensors, vol. 19, no. 20, pp. 1-30, 2019.

[22] L. Parri, S. Parrino, G. Peruzzi, and A. Pozzebon, "Offshore lorawan networking: Transmission performances analysis under different environmental conditions," IEEE Trans. on Instrumentation and Measurement, no. 70, pp. 1-10, Oct-2020.

[23] M. Cattani, C.A. Boano, and K. Römer, "An experimental evaluation of the reliability of Lora long-range low-power wireless communication". J. of Sensor and Actuator Netw., vol. 6, no. 6, pp. 1-19, 2017.

[24] R. S.-Iborra et al., "Performance evaluation of LoRa considering scenario conditions". Sensors, vol. 18, no. 3, pp. 1-19, 2018.

[25] N. Jeftenić, M. Simić, and Z. Stamenković, (2020, June). "Impact of Environmental Parameters on SNR and RSS in LoRaWAN," In Int. Conf. on Electrical, Communication, and Computer Engineering (ICECCE), IEEE, Jun-2020, pp. 1-6.

[26] E. D. Ayele et al., "Performance analysis of LoRa radio for indoor IoT applications". Int. Conf. on Internet of Things for the Global Community (IoTGC), IEEE, 2017, pp. 1-8.

[27] A. Rahman and M. Suryanegara, "The development of IoT LoRa: A performance evaluation on LoS and Non-LoS environment at 915 MHz ISM frequency". Int. Conf. on Sig. and Syst. (ICSigSys), IEEE, 2017, pp. 163-167.

[28] J. Haxhibeqiri et al., "LoRa indoor coverage and performance in an industrial environment: A case study". Int. Conf. on Emerg. Techn. and Factory Autom. (ETFA), IEEE, 2017, pp. 1-8.

[29] T. Petric et al., "Measurements, performance, and analysis of LoRa FABIAN, a real-world implementation of LPWAN". Annual Int. Symp. Personal, Indoor, and Mobile Radio Comm. (PIMRC), IEEE, 2016, pp. 1-7.

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: Southeast University. Downloaded on February 19,2023 at 08:24:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2023.3238013

[30] A. Augustin et al., "A study of LoRa: Long range & low power networks for the internet of things". Sensors, vol. 16, no. 9, pp. 1-18, 2016.
[31] D. Yim et al., "An experimental LoRa performance evaluation in tree farm". Sensors Apps. Symp. (SAS), IEEE, 2018, pp. 1-6.
[32] S.-Y. Wang et al., Performance of LoRa-based IoT applications on campus". Vehicular Tech. Conf. (VTC-Fall), IEEE, 2017, pp. 1-6.
[33] L. Slats, "A Brief History of LoRa® : Three Inventors Share Their Personal Story at The Things Conference," 2020.
[34] N. Jeftenic, M. Simic, and Z. Stamenkovic, "Impact of Environmental Parameters on SNR and RSS in LoRaWAN," Int. Conf. on Electrical, Communication, and Computer Engineering (ICECCE), IEEE, June 2020, pp. 1-6.
[35] O. Elijah, S. K. A. Rahim, V. Sittakul, A. M. Al-Samman, M. Cheffena, J. B. Din, and A. R. Tharek, "Effect of weather condition on LoRa IoT communication technology in a tropical region: Malaysia." IEEE Access, vol. 9, pp. 72835-72843, May-2021.
[36] W. Kong et al., "Short-term residential load forecasting based on LSTM recurrent neural network". "Trans. on Smart Grid, IEEE, vol. 10, no. 1, pp. 841-851, 2017.
[37] Y. Yan et al. "A network intrusion detection method based on stacked autoencoder and LSTM". Int. Conf. on Comm. (ICC), IEEE, 2020, pp. 1-6.
[38] R. Dey and F.M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks". Int. Midwest Symp. Circuits and Syst. (MWSCAS), IEEE, 2017, pp. 1597-1600.
[39] E. Goldoni et al., "Correlation between weather and signal strength in LoRaWAN networks: An extensive dataset". Compt. Networks, vol. 202, pp. 1-4, 2022.
[40] Saidi, R., W. Bouaguel, and N. Essoussi, Hybrid feature selection method based on the genetic algorithm and Pearson correlation coefficient, in Machine Learning Paradigms: Theory and Application. 2019, Springer. p. 3-24.
[41] R. Kufakunesu, G. P. Hancke, and A.M. A-Mahfouz, "A survey on adaptive data rate optimization in lorawan: Recent solutions and major challenges". Sensors, vol. 20, no. 18, pp. 1-25, 2020.
[42] M. Slabicki, G. Premsankar, and M. Di Francesco, "Adaptive configuration of LoRa networks for dense IoT deployments". Netw. Operations and Manag. Symp., IEEE/IFIP, pp. 1-8, 2018.

Showkat Ahmad Bhat received his bachelor's degree in Electronics and Communication from Punjab Technical University, Punjab, India, and his master's degree from Lovely Professional University, Punjab, India. Currently pursuing a Ph.D. in High-Speed Network (HSNL) Laboratory, at National Tsing Hua University under the supervision of Prof. Nen-Fu Huang. His primary research interests include Edge Computing, Agriculture Engineering, IoT sensor networks, LPWAN (LoRaWAN/NB-IoT) transmission technologies, and applications, Blockchain Technology, Artificial Intelligence, Big Data, and Wireless Sensor Networks. He is also a student member of the IEEE.

Nen-Fu (Fred) Huang received the Ph.D. degree in computer science from National Tsing Hua University (NTHU), Taiwan, in 1986. He is now serving as the Dean of EECS College of NTHU. From 1997-2000, he was the Chairman of the Department of Computer Science, at NTHU and since 2008, he is a Distinguished Professor at NTHU. From 2015-2016, he was the director of the Computer and Communication Research Center (CCRC) of NTHU. His current research interests include Internet of Things (IoT) sensor networks, Smart Agriculture and agriculture blockchain, AI-based crops classification technologies, LPWAN (LoRaWAN/NB-IoT) transmission technologies, and applications, AI-based Big Data Analysis for MOOCs (Massive Open Online Courses), network security, SDN/NFV networks, and network applications/flow identification technologies.

Imtiyaz Hussain received his bachelor's degree in Mechanical Engineering from the University of Kashmir, J & K, India, and his master's degree from National Tsing Hau University, Hsinchu, Taiwan. Currently pursuing a Ph.D. from the National Taipei University of Technology, Taiwan. His primary research interests include Artificial Intelligence, Big Data, Boiling, and Heat and Mass Transfer.

Uzair Sajjad received the Ph.D. degree in Mechanical Engineering from National Yang-Ming Chiao Tung University (NYCU), Taiwan, in 2021. He is now serving as the Research Assistant Professor at the National Taipei University of Technology. His current research interests include Energy, Heat Transfer, Evaporating Cooling, Atmospheric Water harvesting, Surface Engineering, and Artificial Intelligence.

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: Southeast University. Downloaded on February 19,2023 at 08:24:06 UTC from IEEE Xplore. Restrictions apply.