Energy Efficiency Optimization in LoRa Networks—A Deep Learning Approach

Lam-Thanh Tu, Abbas Bradai, *Senior Member, IEEE*, Olfa Ben Ahmed, Sahil Garg, *Member, IEEE*, Yannis Pousset, and Georges Kaddoum, *Member, IEEE*

***Abstract*— The optimal transmit power that maximizes energy efficiency (EE) in Longe Range (LoRa) networks is investigated by using the deep learning (DL) approach. Particularly, the proposed artificial neural network (ANN) is trained two times; in the first phase, the ANN is trained by the model-based data which are generated from the simplified system model while in the second phase, the pre-trained ANN is re-trained by the practical data. Numerical results show that the proposed approach outperforms the conventional one which directly trains with the practical data. Moreover, the performance of the proposed ANN under both partial and full optimum architecture are studied. The results depict that the gap between these architectures is negligible. Finally, our findings also illustrate that instead of fully re-trained the ANN in the second training phase, freezing some layers is also feasible since it does not significantly decrease the performance of the ANN.**

***Index Terms*— Deep learning, energy efficiency, LoRa networks, Poisson cluster process, stochastic geometry.**

I. Introduction

ENERGY efficiency (EE) is one of the long-lasting problems in wireless communications systems. However, in the past, the network operator/planning primarily focused on maximizing the spectral efficiency (SE) as well as enhancing the coverage area, as a result, the maximum transmit power was typically considered one of the optimal solutions. Nonetheless, such an approach, of course, reduces dramatically the energy efficiency. The issue is even more serious since it is expected that by 2022 there will be 12.3 billion end-devices (EDs) to be connected to the wireless networks [1] and making the internet-of-thing (IoT) becomes feasible. As a consequence, minimizing the power consumption or maximizing the EE while guaranteeing the SE has been emerged as one of the most important issues in wireless networks. Despite the fact that the operation of the base stations (BSs) has been optimized to maximize the EE of the cellular networks [2], super-dense deployment makes it too bulky, thus, letting it less attractable for supporting such massive low power networks. Fortunately, the low power wide area network (LPWAN) is regarded as one of the most suitable technologies thanks to its properties, i.e., low power consumption, low cost, and wide coverage area [3]. Among all the available LPWAN techniques, i.e., SigFox, Weightless, Narrowband-IoT (NB-IoT), etc., Long Range (LoRa) is gained lots of attraction in both academia and industry [4]. By actively fine-tuning its parameters, i.e., the spreading factor (SF), the coding rate (CR), and the bandwidth (BW), LoRa is able to serve a wide range of IoT applications/devices, e.g., e-Health, smart city, smart home, that, in general, have different quality-of-service (QoS).

Deep learning (DL), on the other hand, is proven itself as one of the best ones among all machine learning (ML) techniques when a large number of data are available [5], [6]. Nevertheless, unlike other domains where mathematical modelling is primarily difficult to be performed and the pure data-driven approach like deep learning seems to be the sole solution, wireless communications, on the contrary, have always depended on robust mathematical modelling for system design, analysis, optimization and can be considered as a model-based approach. Nonetheless, with the dramatic evolution of the wireless networks, i.e., the exponential growth of end-devices with different applications are connected to wireless networks, making it more complex, hence, the mathematical modelling is steadily losing its accuracy as well as mathematical tractability. As a result, deep learning has recently been commenced applying to wireless communications. The application of DL in wireless networks, differently, does not mean that the prior mathematical modelling is ignored. In fact, the most feasible solution for applying DL in wireless communications is to combine the advantages of both data-driven and model-based approaches [7]. The main target of this combination is to synergistically exploit the deep and expert knowledge from theoretical models even though it may be inaccuracy and cumbersomeness to facilitate the use of DL in wireless networks [8]. This approach can also be considered as transfer learning [9], where the artificial neural network (ANN) is first trained by faults data and then re-trained by the empirical data. Of course, such an approach requires strong mathematical modelling which is as close as possible to the empirical data in order to attain the highest performance. As a result, in this paper, we maximize the energy efficiency in

Manuscript received November 21, 2021; revised April 6, 2022; accepted May 21, 2022. This work was supported by the CPER/FEDER NUMERIC Project "Intelligent Networks II" under Grant Agreement NUMERI06. The Associate Editor for this article was Z. Lv. *(Corresponding author: Sahil Garg.)*

Lam-Thanh Tu was with the Institute XLIM, University of Poitiers, 86000 Poitiers, France. He is now with the Communication and Signal Processing Research Group, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City 72915, Vietnam (e-mail: tulamthanh@tdtu.edu.vn).

Abbas Bradai, Olfa Ben Ahmed, and Yannis Pousset are with the Institute XLIM, University of Poitiers, 86000 Poitiers, France (e-mail: abbas.bradai@univ-poitiers.fr; olfa.ben.ahmed@univ-poitiers.fr; yannis.pousset@univ-poitiers.fr).

Sahil Garg is with the Resilient Machine Leaning Institute (ReMI), École de Technologie Supérieure, Montreal, QC H3C 1K3, Canada (e-mail: sahil.garg@ieee.org).

Georges Kaddoum is with the Electrical Engineering Department, École de Technologie Supérieure, Montréal, QC H3C 1K3, Canada (e-mail: georges.kaddoum@etsmtl.ca).

Digital Object Identifier 10.1109/TITS.2022.3183073

LoRa networks by combining the model-based and data-driven approaches.

The direct and indirect studies of the energy efficiency in LoRa networks were studied in [10]–[17]. In [10], the energy efficiency was investigated by considering other medium access schemes rather than pure ALOHA. By jointly exploiting user scheduling and SF assignment, the maximization of system EE was investigated in [11]. In [12], the EE of LoRa networks was studied by leveraging tools from stochastic geometry as well as taking into account the current consumption of all states of EDs such as wake up, radio preparation, transmission, wait for the 1st window, etc. Their outcomes illustrated that the EE is a pseudo-concave function with respect to the transmit power. They, however, assumed that the distribution of the EDs followed by the Poisson point process (PPP) is proven to be impractical [13]. The authors in [14] proposed a couple of routing schemes, i.e., next-ring-hop routing and variable-hop routing instead of single-hop routing to minimize the total energy consumption thus improving the EE of the networks. Numerical results showed that all multi-hop routing schemes outperform the single-hop scheme. Their work, nonetheless, focused on minimizing energy consumption without utilizing deep learning techniques. In [15], a novel media access control (MAC) protocol was proposed to improve LoRa network performance. These outcomes depicted that the novel protocol outperforms the conventional LoRaWAN in terms of throughput, energy efficiency, and delay. Nevertheless, their work concentrated on designing a novel MAC protocol while the present work focuses on studying the EE by combining both mathematics and deep learning techniques. Moises *et al.* in [16] analyzed and proposed different strategies to adapt LoRa parameters in order to optimize energy consumption. On the other hand, the work in [17] did the survey on the energy efficiency over multiple layers in LoRa networks.

Regarding the applications of DL in LoRa networks, it is still in the infancy stage despite receiving tremendous attention in cellular networks. In fact, there were solely a few works that applied ML/DL in LoRa networks, [19]–[26]. Particularly, in [19], the process of network configuration was formulated as a reinforcement learning (RL) problem. The time difference of the arrival positioning method in LoRa networks was improved by applying deep learning [20]. Sanghyun *et al.* in [21] proposed a DL-based transmission interval management to reduce the data measurement in LoRa networks. The results illustrated that the scalability of LoRa networks was enhanced by 31%. The work in [22], on the other hand, addressed the collision problem in the LoRa network. Particularly, the authors proposed a long short-term memory (LSTM) extended Kalman filter to predict the collision in LoRa. The results showed that their proposed approach outperformed the conventional recurrent neural network (RNN). Inaam in [23], on the other hand, employed reinforcement learning to reduce collisions and enhance network performance. DL was also employed to predict the large-scale path-loss in [24]. More precisely, a bidirectional-LSTM was utilized to develop a land-cover-aware path-loss model. Authors in [25] showed that the coverage area was able to be extended by 60.4% percent compared to the short-range communications inside a commercial building by applying a machine-learning-based approach. Guanxiong in [26] designed a DL-based radio frequency fingerprint identification (RFFI) for LoRa networks. Experimental results revealed that the proposed DL scheme was able to reach an accuracy of 96% with less complexity and training time.

The performance of energy efficiency by employing the DL approach in LoRa networks was addressed in [27]–[29]. Particularly, the work in [27] minimized the energy consumption at EDs to predict their lifespan. This work, nonetheless, assumed that the distance from the gateway (GW) to ED was deterministic and the impact of small-scale fading was skipped. Similarly, Kumari *et al.* in [28] also minimized the energy consumption of the smart metering system based on LoRa networks via DL and edge computing techniques. To improve the LoRa gateway energy efficiency, the author in [29] proposed novel resource management drawing on the reinforcement learning approach.

In the present work, different from these above-mentioned papers, we maximize the energy efficiency in LoRa networks with respect to the transmit power under practical scenarios where the distribution of the end-devices and the small-scale fading follow general distributions, i.e., the Poisson cluster processes (PCP) and the Nakagami-m distribution, combined with the imperfect power consumption at the end-devices. Of course, it is crystal clear that the optimal transmit power that maximizes the EE under such the general system model can not be attained based on the mathematical frameworks, thus, the deep learning approach is utilized instead. Our proposed DL approach (called the double training approach), however, requires a relatively small practical data to attain the high accuracy output and is different from the conventional approach (called direct training approach) where a large number of practical data is imperatively needed in order to fully train the ANN. To realize such well-trained neural networks with relatively small practical data, the proposed artificial neural networks is trained two times. In the first training phase or pre-trained phase, the proposed ANN is trained based on the model-based data and in the second training phase, the pre-trained ANN is re-trained based on the practical data. Here, the model-based data in the first training phase are generated based on the mathematical frameworks under the simplified system model that may be inaccuracy but may provide some expert knowledge and can be helpful for the second training; the practical data in the second training phase, diversely, are uniquely generated by Monte Carlo simulation that is resource-consuming compared to the mathematical frameworks. To be more specific, the main contributions and novelties of this paper are summarized as follows:

- We model the distribution of the end-devices and the small-scale fading according to the Poisson cluster processes and Nakagami-m distribution, respectively. Additionally, the power consumption at EDs is impaired by an additive noise that follows either Gaussian or uniform distribution.
- We derive the closed-formed expression of the energy efficiency under the simplified system model.

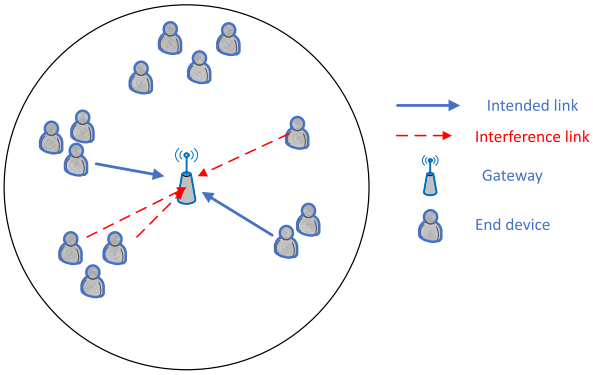

Fig. 1. An uplink single-gateway LoRa networks.

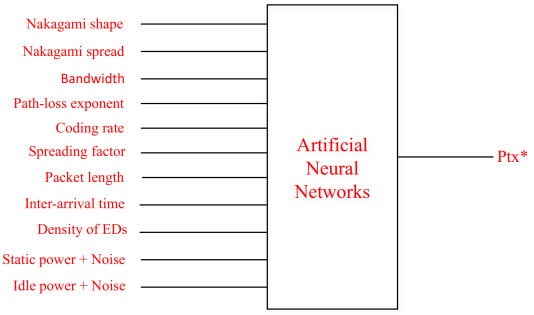

Fig. 2. Schematic of the considered artificial neural networks.

- We generate the first training set by deploying the mathematical frameworks under the simplified system model. Particularly, the optimal transmit power in this phase is attained by numerical solving a non-linear equation.
- We unveil that the proposed double training approach outperforms the direct training approach under all considered metrics, i.e., the mean square error (MSE) and R squared, $\mathcal{R}^2$, or the coefficient of determination.
- Our findings show that increasing training data monotonically ameliorate the performance of the neural networks. Nonetheless, enhancing the ANN performance by raising either the number of epochs or the number of neurons does not always benefit.
- The performance of the ANN under various optimal network architecture, i.e., partial and full optimum architecture, are investigated. The findings depict that the gap between these optimum architectures is negligible, thus, the partial optimum architecture is preferable provided that the resources consumption is taken into account.
- Our findings also illustrate that it is feasible to freeze some layers during the second training phase which does not significantly affect the performance of the ANN. We also unveil that relying on the cluster process, the impact of the additive noise followed by Gaussian distribution on the performance of the considered ANN is either worse or better than one followed by the uniform distribution. However, the gap between the two distributions generally is minor.

The rest of this paper is organized as follows. In Section II, the considered system model is presented. In Section III, we formulate the optimization problem and the design of the proposed neural networks is provided in Section IV. In Section V, the performance of the proposed ANN is evaluated and discussed under various scenarios. Finally, Section VI concludes the paper.

II. System Model

A. LoRa Networks Modeling

Let us consider an uplink single gateway LoRa networks as shown in Fig. 1 where the gateway is located at the center of the disc with area $A = \pi R^2$, here, R is the network radius and a number of EDs which follow Poisson cluster processes (or doubly Poisson point processes), with density $\lambda_{\mathcal{C}} > 0$, inside A and $\lambda_{\mathcal{C}} = 0$, otherwise. In particular, in this work we consider two notable Poisson cluster processes, i.e., the Matérn and the Thomas cluster process with corresponding density $\lambda_{\text{Mat}} = \lambda_P \overline{C}_{\text{Mat}}$ and $\lambda_{\text{Tho}} = \lambda_P \overline{C}_{\text{Tho}}$, where λ_P is the density of the parent point process and $\overline{C}_o$, $o \in \{\text{Mat}, \text{Tho}\}$, is the average number of offspring per cluster in the o cluster process [30]. The offspring of the Matérn cluster process are independently and uniformly distributed in a disc of radius r_{Mat} around the parent point while the offspring of the Thomas cluster process is scattered with variance σ^2_{Tho} around each parent point.

B. Channel Modelling

Let us consider a generic signal from an arbitrary ED to the gateway, it is impaired by both the small-scale fading and large-scale path-loss. It should be emphasized that the impact of the shadowing is implicitly studied by modifying the density of the EDs [2].

1) Small-Scale Fading: Let us denote h_q as the fading from an arbitrary node q to the gateway which follows Nakagami-m distribution with the corresponding shape and spread parameters, i.e., $m \geq 1/2$ and Ω. There is no doubt that Nakagami-m fading is one of the most general fading distributions that can represent other well-known distributions by properly adjusting its shape parameter, e.g., $m = 1$, is the Rayleigh fading, and $m = 1/2$, is the single-sided Gaussian distribution. Additionally, assuming that time is slotted (slotted ALOHA medium access [10]) and the fading remains constant during one time-slot and changes between time-slot.

2) Large-Scale Path-Loss: Considering a transmission link from a generic node q to the gateway, the large-scale path-loss is formulated as

$$\rho_q = K_0 r_q^{\beta}, \tag{1}$$

where $\beta > 2$ and $K_0 = (4\pi f_c/c)^2$ are the path-loss exponent and the path-loss constant, respectively; f_c is the carrier frequency and $c = 3 \times 10^8$ (in meter per second) is the speed of light; and r_q is the distance from node q to the gateway.

C. Power Consumption Modelling

Assuming that the ED is operated either in transmission mode or sleep mode. To be more precise, the ED is considered in the transmission mode providing that it transmits packets to the gateway and in the sleep mode otherwise. The power consumption under the transmission mode comprises of two parts, i.e., the transmit power, P_{tx}, and the static (circuit)

TABLE I
MAIN NOTATIONS AND MATHEMATICAL SYMBOLS

Symbol	Definition
$\mathbb{E}\{.\}$, $\Pr(.)$	Expectation and probability operators
$\exp(.)$, $\ln(.)$	Exponential and logarithm functions
$F_X(x)$	Cumulative distribution function of random variable X
$f_X(x)$	Probability density function of random variable X
$\mathbf{1}(x)$	Indicator function
$\dot{f}(x) = df(x)/dx$	First-order derivative of f with respect to x
$\max\{.\}$, $\min\{.\}$	Maximal and minimal functions
Pcov (γ_{D}), $\widetilde{P}(\gamma_{\mathrm{D}})$	Exact and approximated coverage probability
$\lambda_{\mathcal{C}}$, λ_P	Density of the Poisson cluster process and of the parent process
$\lambda_{\mathcal{A}}$, NF	Density of the set of active ED and noise figure
$\overline{C}_o$, $o \in \{$Mat, Tho$\}$	Average number of offspring per cluster of the o cluster process
r_{Mat}	Radius of the cluster in the Matérn cluster process
σ^2_{Tho}	Variance of the cluster in the Thomas cluster process
m, Ω	Shape and spread parameters of the Nakagami-m distribution
c, f_c	Speed of light and carrier frequency
β, K_0	Path-loss exponent and path-loss constant
p_A, L_{pac}	Active probability and packet length
T_{in}, R_{bit}	Average packet inter arrival time and bit rate
γ_{D}, γ_{I}	Quality-of-service threshold and interference rejection threshold
P_{tx}, σ^2	Transmit power of end-device and noise variance at gateway
$P_{\mathrm{tx}}^{\min}$, $P_{\mathrm{tx}}^{\max}$	Minimum and maximum transmit power of end-device
P_{tx}^{*}, R	Optimal transmit power of end-device and network radius
$\widetilde{P}_{\mathrm{sta}}$, P_{sta}, and $\overline{P}_{\mathrm{sta}}$	Practical, ideal, and average static power consumption
$\widetilde{P}_{\mathrm{idl}}$, P_{idl} and $\overline{P}_{\mathrm{idl}}$	Practical, ideal, and average idle power consumption
BW, SF, and CR	Bandwidth, spreading factor, and coding rate
z, z_{ori}	Input and output of the normalization process
$\mathcal{L}_{\min}$, $\mathcal{L}_{\max}$	Minimum and maximum of number of hidden layers
$\mathcal{N}_{\min}$, $\mathcal{N}_{\max}$	Minimum and maximum of number of neurons of each layer
$\mathcal{L}$, $\mathcal{N}$	Number of hidden layer and number of neurons per layer
$\mathcal{L}_a^*$, $a \in \{1, 1+2\}$	Optimal number of hidden layer under a optimal architecture scheme
$\mathcal{N}_a^*$, $a \in \{1, 1+2\}$	Optimal number of neurons per layer under a optimal architecture scheme
y, $\widehat{y}$	Observed and predicted output of ANN
$\overline{y}$	Mean of the observed output of ANN
PSE, Pcon	Potential spectral efficiency and average power consumption
$\Psi^{(u)}$, $u \in \{1,2\}$, $\lvert X\rvert$	Training set of u phase and the size of set X
$\zeta(.)$, b_n^l	Activation function and bias of node n in layer l
e_n^l	Output signal at neuron n in hidden layer l
$w_{i,n}^l$	Weight from node i in layer $l-1$ to node n in layer l
MSE, $\mathcal{R}^2$	Mean square error and coefficient of determination
SS_{res}, SS_{tot}	Residual sum of squares and total sum of squares

power, P_{sta}; while in the idle mode, the ED consumes P_{idl} power. Here, the static power consumption comprises all other power consumption excluding the transmit power in the transmission mode, i.e., wake up, radio preparation, wait and receive the 1st and 2nd windows, radio off, and so on. Furthermore, we also consider the impact of the hardware impairment on the static and idle power; indeed, the static and idle power are impaired by an additive noise that follows either Gaussian or uniform distribution. As a result, the practical static and idle power denoted as $\widetilde{P}_{\mathrm{sta}}$ and $\widetilde{P}_{\mathrm{idl}}$ are written as follows:

$$\widetilde{P}_{\mathrm{sta}} = P_{\mathrm{sta}} + \omega_{\mathrm{sta}}, \quad \widetilde{P}_{\mathrm{idl}} = P_{\mathrm{idl}} + \omega_{\mathrm{idl}}, \tag{2}$$

where ω_x, $x \in \{\mathrm{sta}, \mathrm{idl}\}$, is the random variable (RV) and P_{sta} and P_{idl} are the ideal static and idle power.

III. PROBLEM FORMULATION

In this work, the principal objective is to identify the optimal transmit power denoted as P_{tx}^{*} that maximizes the energy efficiency (in bits/Joule) of whole networks given a set of input parameters, i.e., the fading parameters included both shape and spread factors, the transmission bandwidth, the spreading factor, the coding rate, the inter-arrival time between two packets, T_{in}, the packet length, L_{pac}, the path-loss exponent, the density of the PCP, $\lambda_{\mathcal{C}}$, and the imperfection static and idle power as shown in Fig. 2. Particularly, the problem in Fig. 2

can be formulated as follows [31]:

$$\max_{P_{\rm tx}\in\left[P_{\rm tx}^{\min},P_{\rm tx}^{\max}\right]} {\rm EE} = \frac{\rm PSE}{\rm Pcon} = \frac{\lambda_{\mathcal{A}}{\rm BW}\log_2\left(1+\gamma_{\rm D}\right){\rm Pcov}\left(\gamma_{\rm D}\right)}{\lambda_{\mathcal{A}}\left(P_{\rm tx}+\overline{P}_{\rm sta}\right)+\left(\lambda_{\mathcal{C}}-\lambda_{\mathcal{A}}\right)\overline{P}_{\rm idl}}. \quad (3)$$

It is noted that the range of the transmit power in (3), without loss of generality, can go from zero to infinity, i.e., $P_{\rm tx}^{\min}\rightarrow 0$ and $P_{\rm tx}^{\max}\rightarrow\infty$; potential spectral efficiency (PSE) (in bits/s/m^2) measures number of bit successfully transmitted per unit area and is formulated as $\lambda_{\mathcal{A}}{\rm BW}\log_2\left(1+\gamma_{\rm D}\right){\rm Pcov}\left(\gamma_{\rm D}\right)$; where ${\rm Pcov}\left(\gamma_{\rm D}\right)$ is the coverage probability (Pcov) of an arbitrary link and formulated as [32]

$${\rm Pcov}\left(\gamma_{\rm D}\right) = \Pr\left\{{\rm SIR}=\frac{S}{I_S}\geq\gamma_{\rm I},\,{\rm SNR}=\frac{P_{\rm tx}S}{\sigma^2}\geq\gamma_{\rm D}\right\}, \quad (4)$$

where SIR is signal-to-interference ratio and is computed as ${\rm SIR}=\frac{P_{\rm tx}S_0}{P_{\rm tx}I_S}=\frac{P_{\rm tx}h_0^2/\rho_0}{P_{\rm tx}\sum_{i\in\Phi^{\mathcal{A}}\setminus(0)}h_i^2/\rho_i}$; S_0 is the signals from the ED of interest to the gateway; I_S is the aggregate interference from all active EDs except for the desired one; assuming that packets with different spreading factor are perfectly orthogonal, hence, there is no inter spreading factor interference at the gateway [33]; h_0^2 and ρ_0 are the channel gain and large-scale path-loss of the ED of interest while h_i^2 and ρ_i are the channel gain and the large-scale path-loss from interferer i to the gateway; $\Phi^{\mathcal{A}}\setminus(0)$ is the set of active EDs except for the desired ED which density $\lambda_{\mathcal{A}}=p_A\lambda_{\mathcal{C}}$ under the considered area; where $p_A=\frac{1}{T_{\rm in}}\frac{L_{\rm pac}}{R_{\rm bit}}$ being active probability. In this work, assuming that the length of the packet, $L_{\rm pac}$, is identical among all EDs, and $R_{\rm bit}$ is the bit rate which is computed as $R_{\rm bit}={\rm SF}\frac{\rm BW}{2^{\rm SF}}{\rm CR}$ [3]. SNR is the signal-to-noise ratio and is formulated as ${\rm SNR}=P_{\rm tx}S_0/\sigma^2$ where $\sigma^2=10^{\left(-174+{\rm NF}+10\log_{10}{\rm BW}\right)/10}$ [33] is the noise variance of the additive white Gaussian noise (AWGN); NF is the noise figure (in dBm) of the receiver. $\gamma_{\rm D}$, $\gamma_{\rm I}$ are the quality-of-service threshold and the interference rejection threshold. In (3), Pcon is the average power consumption of the whole networks measured in Watt/m^2; the first term of Pcon, $\lambda_{\mathcal{A}}\left(P_{\rm tx}+\overline{P}_{\rm sta}\right)$, accounts for the power consumption under transmission mode and the remain term, $\left(\lambda_{\mathcal{C}}-\lambda_{\mathcal{A}}\right)\overline{P}_{\rm idl}$ is under idle mode; where $\overline{P}_x=P_x+\mathbb{E}\left\{\omega_x\right\}$, $x\in\{\rm sta,idl\}$, is the average dissipation of the static and idle power.

Remark 1: It is noted that the considered definition of the Pcov in (4) is more accurate compared with the state-of-the-art definition in [33]. The main reason is that the Pcov definition in [33] ignored the correlation between the SIR and SNR at the receiver while the considered one is able to capture this correlation [32].

It is apparent that the most intuitive approach to solving the optimization problem in (3) is to compute the EE in a closed-form expression followed by solving a non-linear equation to obtain $P_{\rm tx}^*$. However, as shown in the sequel, it is impossible to obtain the closed-form expression of the EE under the considered system model. Moreover, even the distribution of the distance from an arbitrary node to the gateway as well as the aggregate interference is also unfeasible to represent in the closed-form expressions. Especially, by direct inspection (3), it is evident that the framework of the Pcov is essential in order to compute the EE, let us formulate the Pcov as follows:

$${\rm Pcov}\left(\gamma_{\rm D}\right)=\int_{s=\sigma^2\gamma_{\rm D}/P_{\rm tx}}^{\infty}F_{I_S}\left(s/\gamma_{\rm I}\right)f_S\left(s\right)ds, \quad (5)$$

where $f_S\left(s\right)$ and $F_{I_S}\left(x\right)$ are the probability density function (PDF) of the intended signal and the cumulative distribution function (CDF) of the aggregate interference. Nonetheless, the CDF of I_S has never existed even with the simplest scenario, i.e., the EDs follow the Poisson point process (PPP); the PDF of the intended signal, on the contrary, can be computed in the closed-form expression under some special cases, for example, the EDs follow PPP combined with Rayleigh fading and a special value of the path-loss exponent. It, as a result, is unworkable to obtain the closed-form expression of both Pcov and EE even with the simplest case [34].

As a result, in this paper, we are going to find the $P_{\rm tx}^*$ by using a deep learning approach. Nevertheless, different from the conventional DL approach that requires enormous training data to fully train the ANN or simply called the data-driven approach. In this manuscript, we synergistically combine the model-based with the data-driven approach to derive the optimal transmit power in (3). To do so, we split the training process into two phases instead of one as the conventional approach. In the first training phase, the ANN is trained by the model-based data which may be simplified and inaccuracy but can bring some expert knowledge and be useful for the second phase. In the second training phase, the pre-trained ANN is re-trained by utilizing a small amount of practical data which is generated via Monte Carlo simulation. In the next section, the design of the proposed neural networks, as well as a detailed description of the training process, are provided.

IV. Design of the Neural Networks

In this section, the proposed ANN is designed from creating the data to selecting the optimal architecture of the ANN as well as identifying the performance metrics which are used to evaluate the performance of the ANN.

A. Generate Data Set

1) First Training Phase (Pre-Train Phase) Data Set: It is obvious that data is the most important element of any neural networks. Under the considered system model, it is unreasonable to obtain a large number of data from either Monte Carlo simulations due to the resources constraint or the frameworks owing to the mathematical intractability as discussed in Section III. Thus, in order to obtain an ultra-large data set under the limited resources, we, as a result, propose to create the data set in two phases. In the first phase, the data set is obtained by simplifying the system model in Section II so that the closed-form expression of the EE and the optimal transmit power, $P_{\rm tx}^*$, can be mathematically obtained.

Particularly, the simplified system model is described as follows: i) assuming that the ED is no longer followed by PCP but PPP with density λ_P; ii) the ideal static and idle power at EDs; and iii) the aggregate interference is approximated

by the dominant interferer [33] and the instantaneous fading by its average [34]. Other assumptions remain the same as Section II. Based on these assumptions, the approximated coverage probability denoted as $\widetilde{P}(\gamma_{\mathrm{D}})$ is computed in (6), shown at the bottom of the page. Here, $\alpha = 2/\beta$, $\mathcal{M} = p_A \lambda_P \pi R^2$, $\mathcal{A} = \frac{m\theta P_{\mathrm{tx}}}{R^{\beta} K_0 \sigma^2 \gamma_{\mathrm{D}}}$ and $\mathbf{1}(x)$ is the indicator function.

Having the approximated Pcov in hand, the optimal transmit power under the simplified system model denoted as $\widetilde{P}_{\mathrm{tx}}^{*,1}$, is obtained by solving the following nonlinear equation:

$$\dot{\widetilde{P}}(P_{\mathrm{tx}})\, Q(P_{\mathrm{tx}}) - \widetilde{P}(P_{\mathrm{tx}})\, \dot{Q}(P_{\mathrm{tx}}) = 0, \tag{7}$$

where $\dot{\widetilde{P}}(x = P_{\mathrm{tx}})$ is the first-order derivative of the approximated Pcov with respect to the transmit power and is computed in (8), shown at the bottom of the next page and $\dot{Q}(x) = p_A \lambda_P$ is the first-order derivative of the average power consumption respect to P_{tx}; $Q(x) = \lambda_{\mathcal{A}} \left(P_{\mathrm{tx}} + \overline{P}_{\mathrm{sta}}\right) + (\lambda_{\mathcal{C}} - \lambda_{\mathcal{A}})\, \overline{P}_{\mathrm{idl}}$ and $\dot{\mathcal{A}}(x = P_{\mathrm{tx}}) = d\mathcal{A}/dx = \frac{m\theta}{R^{\beta} K_0 \sigma^2 \gamma_{\mathrm{D}}}$, respectively.

2) Second Training Phase (Re-Train Phase) Data Set: The data in this phase are created imperatively by employing Monte Carlo simulation. However, owing to the resources constraint, a limited amount of data is created compared to the previous phase.

B. Data Normalization

The data set is normalized before putting into the ANN. In this work, the simple max-min normalization is applied as follows:

$$z = \frac{z_{\mathrm{ori}} - \min(z_{\mathrm{ori}})}{\max(z_{\mathrm{ori}}) - \min(z_{\mathrm{ori}})}, \tag{9}$$

where z_{ori} and $z \in [0, 1]$ are the input and output of the normalization process. In the sequel, without explicit explanation, we assume that the data set has already been normalized.

C. Networks Architecture

To maximize the performance of the ANN, optimizing the networks architecture, i.e., the optimal number of hidden layers and/or a number of neurons of each layer, is essential along with other hyper-parameter optimizations. As the considered ANN will be trained two times; it, in theory, exists two optimal typologies, the first one which merely optimizes from the 1st training set and another which optimizes from both the 1st and 2nd training set. In mathematics, the optimal architecture based on the 1st training set (partial optimum) and from both 1st and 2nd training set (full optimum) denoted as N_1^* and N_{1+2}^* as a function of the number of hidden layers, $\mathcal{L} \in \{\mathcal{L}_{\min}, \ldots, \mathcal{L}_{\max}\}$, the number of neurons of all hidden layers are formulated as follows:

$$\begin{aligned} N_1^*\left(\mathcal{L}_1^*, \mathbf{N}_1^*\right) &= \min\ L\left(\mathcal{L}, \mathbf{N} | \Psi^{(1)}\right) \\ N_{1+2}^*\left(\mathcal{L}_{1+2}^*, \mathbf{N}_{1+2}^*\right) &= \min\ L\left(\mathcal{L}, \mathbf{N} | \Psi^{(1)} + \Psi^{(2)}\right), \end{aligned} \tag{10}$$

where $\mathbf{N} = [\mathcal{N}_1, \ldots, \mathcal{N}_{\mathcal{L}}]$ being a vector which contains number of neurons of all hidden layers; $\mathcal{N}_l \in \{\mathcal{N}_{\min}, \ldots, \mathcal{N}_{\max}\}$, $l \in \{1, \ldots, \mathcal{L}\}$, is the number of neurons in l-th hidden layer; $\Psi^{(u)}$, $u \in \{1, 2\}$, is the training set of u phase; $L(.)$ is the objective function which needs to be minimized; $\left(\mathcal{L}_a^*, \mathbf{N}_a^*\right)$, $a \in \{1, 1{+}2\}$, is the optimal solution based on u training set; $\{\mathcal{L}_{\min}, \ldots, \mathcal{L}_{\max}\}$ and $\{\mathcal{N}_{\min}, \ldots, \mathcal{N}_{\max}\}$ are the minimum and maximum of number of hidden layers and number of neurons of each layer, respectively.

It is evident that optimizing DL topology in (10) is always a cumbersome task as it requires not only the skills and experience of the user but also the mastery of the data features [35]. As a result, the grid search method is applied to find out the optimal architecture with the help of [36]. In addition, to simplify the search space, we further assume that the number of neurons of all hidden layers is identical, i.e., $\mathbf{N} = \mathcal{N}$. Thus, the optimization problem in (10) can be re-written as follows:

$$\begin{aligned} N_1^*\left(\mathcal{L}_1^*, \mathcal{N}_1^*\right) &= \min\ L\left(\mathcal{L}, \mathcal{N} | \Psi^{(1)}\right) \\ N_{1+2}^*\left(\mathcal{L}_{1+2}^*, \mathcal{N}_{1+2}^*\right) &= \min\ L\left(\mathcal{L}, \mathcal{N} | \Psi^{(1)} + \Psi^{(2)}\right), \end{aligned} \tag{11}$$

In fact, the search space has massively reduced from $\mathcal{T} = \sum_{i=1}^{(\mathcal{L}_{\max} - \mathcal{L}_{\min})} (\mathcal{N}_{\max} - \mathcal{N}_{\min})^i$ in (10) to $\mathcal{T} = (\mathcal{L}_{\max} - \mathcal{L}_{\min})(\mathcal{N}_{\max} - \mathcal{N}_{\min})$ in (11).

Remark 2: It is apparent that the full optimum will theoretically provide better performance compared to the partial optimum. Thus, the main purpose is to clarify the gap between the partial optimum versus the full optimum. In case the gap is relatively small, it is more beneficial to employ the partial optimum than its counterpart provided that the consumed resources are taken into consideration.

D. Loss Function

As the considered problem belongs to the regression one, the typical mean square error, MSE, is deployed as the loss function of the ANN and is formulated as follows:

$$\min\ \mathrm{MSE} = (|\Psi|)^{-1} \sum\nolimits_{j \in \Psi} \left(y_j - \widehat{y}_j\right)^2, \tag{12}$$

where Ψ and $|\Psi|$ are the training set and its size; y is the observed output and $\widehat{y}$ is the predicted output.

E. Performance Metric

The mean square error and the R squared, $\mathcal{R}^2$, or the coefficient of determination are utilized as the main metrics to evaluate the performance of the ANN. As the MSE measures the average squared difference between the estimated and true values and is provided in (12); the $\mathcal{R}^2$, differently, measures how close two sets of data are in terms of the distribution. Mathematical speaking, $\mathcal{R}^2$ can be formulated as follows:

$$\mathcal{R}^2 = 1 - \frac{SS_{\mathrm{res}}}{SS_{\mathrm{tot}}} = 1 - \frac{\sum_{j \in \Psi} \left(y_j - \widehat{y}_j\right)^2}{\sum_{j \in \Psi} \left(y_j - \overline{y}\right)^2}, \tag{13}$$

$$\widetilde{P}(\gamma_{\mathrm{D}}) = \mathcal{M}^{-1} (\gamma_{\mathrm{I}})^{-\alpha} \left(1 - \exp\left(-\mathcal{M} \min\left\{(\mathcal{A}\gamma_{\mathrm{I}})^{\alpha}, 1\right\}\right)\right) + \exp(-\mathcal{M}) \left(\min\left\{\mathcal{A}^{\alpha}, 1\right\} - (\gamma_{\mathrm{I}})^{-\alpha}\right) \mathbf{1}(\mathcal{A}\gamma_{\mathrm{I}} - 1) \tag{6}$$

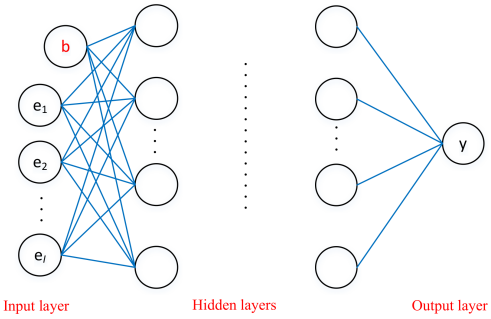

Fig. 3. The fully-connected feedforward neural networks.

where SS_{res} and SS_{tot} are the residual sum of squares and the total sum of squares. Here, SS_{res} measures the amount of variability that is left unexplained after performing the regression while SS_{tot} measures the total variance. As a consequence, if R squared towards 1, it explicitly means that most of the variability are explained by the model; $\overline{y} = (|\Psi|)^{-1} \sum_{j \in \Psi} y_j$ is the mean of the observed output.

F. Training Process

Considering a fully-connected feedforward neural networks with $\mathcal{L}$ hidden layers, $\mathcal{N}$ neurons per hidden layer, and one input layer with $\mathcal{I}$ input, and single output as shown in Fig. 3. The input of the ANN is the path-loss exponent, the fading parameters, the bandwidth, the coding rate, the spreading factor, the packet length, the inter-arrival time, the density of EDs, the practical static and idle power. The single output is the optimal transmit power, P_{tx}^*, which is obtained by (7) in the pre-trained phase or Monte Carlo simulation in the re-trained phase.

Given a training set, $\Psi^{(u)}$, $u \in \{1, 2\}$, the training process is commenced with normalizing the training set by using (9), followed by forward propagation. Particularly, let us denote e_n^l as the output signal at neuron n, $n \in \{1, \ldots, \mathcal{N}\}$, in hidden layer l, $l \in \{1, \ldots, \mathcal{L}\}$, and is computed as

$$e_n^l = \zeta\left(\sum_{i=1}^{\mathcal{V}} w_{i,n}^l e_i^{l-1} + b_n^l\right), \quad \mathcal{V} = \begin{cases} \mathcal{V} = \mathcal{I}, & l = 1 \\ \mathcal{V} = \mathcal{N}, & l \neq 1, \end{cases} \tag{14}$$

where $w_{i,n}^l$ is the weight from node i in layer $l-1$ to node n in layer l; e_i^{l-1} is output of neuron i in layer $l-1$; when $l = 1$, we have e_i^0, $i \in \{1, \ldots, \mathcal{I}\}$, is the i-th input of the ANN; ζ () is the activation function; b_n^l denotes the bias of node n in layer l. In this work, we utilize the sigmoid activation function as the training set is normalized to [0, 1]. The predicted output of the ANN denoted as $\widehat{y}$ is linearly combined the output of the last hidden layer with the weight from the last hidden layer to the output as $\widehat{y} = \sum_{i=1}^{\mathcal{N}+1} w_i^{\mathcal{L}+1} e_i^{\mathcal{L}} + b_y$, b_y is the bias of the output. The forward pass ends by computing the loss function given in (12). The next step is backward propagation where we compute the gradient of the loss function with respect to all the weights and bias of the networks. This gradient is then used to update the weights and bias to minimize the loss function by deploying the Adam algorithm with adaptive learning rate [37] and we finish one training epoch. The training process is repeated by re-computing the forward and backward propagation; updating the loss function, weights, and bias until reaching the stopping conditions, i.e., the maximum epoch or the MSE is smaller than a predefined threshold or the MSE starts increasing, etc.

Remark 3: In the step-back propagation during the second training phase, it is unnecessary to update all the weights and biases. In fact, the ANN can re-train partially instead of fully re-trained or the weight and bias of some layers can be kept constant during the second training phase.

V. Numerical Results and Discussion

In this section, numerical results are provided to confirm our findings. Unless otherwise stated, following values are used to generate the input training set in the first and second phases: $\beta \in [2, 6]$, $m \in [0.5, 30]$, $\Omega \in [1, 50]$, SF $\in \{7, \ldots, 12\}$, BW $\in \{1, \ldots, 500\}$ KHz, CR $= 4/(4+s)$, $s \in \{1, \ldots, 4\}$, $L_{\text{pac}} \in \{1, \ldots, 120\}$ (bytes), $T_{\text{in}} \in \{1, \ldots, 90\}$ (minutes), $P_{\text{cir}} \in [0.5, 5]$ (dBm), $P_{\text{idle}} \in [-10, -5]$ (dBm), $\lambda_P = \overline{N}/\pi R^2$ where $\overline{N} \in \{500, \ldots, 100000\}$ and $R =$ 2000 m; $\gamma_{\text{I}} = 1$ dBm, NF $= 6$ dBm, $f_c = 868$ MHz; γ_{D} is chosen from one of the following values: $\gamma_{\text{D}} \in \{-6, -9, -12, -15, -17.5, -20\}$ (dBm), which absolutely depends on the utilized SF, e.g., if SF $= 7$ then $\gamma_{\text{D}} = -6$ dBm and SF $= 12$ then $\gamma_{\text{D}} = -20$ dBm, respectively. The range of the output, P_{tx}^*, in the first and second phases, on the other hand, are different. Particularly, the range of the optimal transmit power in the first training set denoted as $P_{\text{tx}}^{*,1}$ is from -100 dBm to 60 dBm while the range of P_{tx}^* in the second training set denoted as $P_{\text{tx}}^{*,2}$, is only from -40 dBm to 50 dBm. The chief reason behind this difference is that $P_{\text{tx}}^{*,2}$ is obtained via Monte Carlo simulation which absolutely requires huge resources and efforts, thus, a limited search space is considered. $P_{\text{tx}}^{*,1}$, contrarily, is obtained by maths which is effortless compared to the Monte Carlo method thus the range can be arbitrarily meaningful numbers.

In addition, in the second training set, $\Psi^{(2)}$, we have the average number of offspring per cluster, $\overline{C}_u = \{2, 6\}$, $u \in \{\text{Tho, Max}\}$, the radius of each cluster in Matérn cluster process, $r_{\text{Mat}} = \{2, 6\}$ m, and the standard derivation in Thomas cluster process, $\sigma_{\text{Tho}}^2 = \{0.5, 3\}$. Moreover, also in the second training set, $\Psi^{(2)}$, both the static and idle power are impaired by an additive noise which follows either Gaussian distribution with zero mean and unit variance (in dBm) or uniform distribution from $[-0.5, 0.5]$ (dBm). The size of the 1st training set is equal to 30000, i.e., $|\Psi^{(1)}| = 30000$, while the size of the 2nd training set is not greater than 2000, i.e., $|\Psi^{(2)}| \leq 2000$, and the size of the validation and test set denoted as $|\Xi|$ and $|\Upsilon|$ are always fixed at 500, i.e., $|\Xi| = |\Upsilon| = 500$. Furthermore, the validation set is solely available for the second training phase. It means that we do not evaluate the performance of the ANN after the

$$\dot{\tilde{P}}(x) = \dot{\mathcal{A}}(x)\alpha(\mathcal{A}(x))^{\alpha-1}\left(\exp\left(-\mathcal{M}(\mathcal{A}(x)\gamma_{\text{I}})^{\alpha}\right)\mathbf{1}(1-\mathcal{A}(x)\gamma_{\text{I}}) + \exp(-\mathcal{M})\mathbf{1}(\mathcal{A}(x)\gamma_{\text{I}}-1)\mathbf{1}(1-\mathcal{A}(x))\right) \tag{8}$$

TABLE II
THE ANN PARAMETERS (UNLESS OTHERWISE STATED)

Parameters	Values
Epochs	150 or 300
$\mathcal{L}$	4
$\mathcal{N}$	55
r_{Mat}	$\{2, 6\}$
σ^2_{Tho}	$\{0.5, 3\}$
$\mathcal{L}_{\min}$, $\mathcal{L}_{\max}$	1, 6
$\mathcal{N}_{\min}$, $\mathcal{N}_{\max}$	20, 80
$\left\|\Psi^{(1)}\right\|$	30000
$\left\|\Psi^{(2)}\right\|$	≤ 2000
$\|\Xi\|$	500
$\|\Upsilon\|$	500

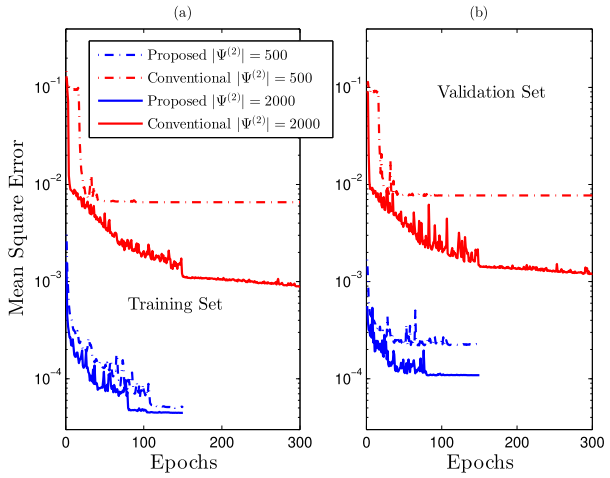

Fig. 4. Mean square error (MSE) versus number of epochs of training set (a) and test set (b) in the second training phase.

first training. Moreover, the test set is only evaluated by the best epoch from the validation phase. It should be noted that the training set in both phases is able to cover most of the practical environment, i.e., urban or rural areas, as well as the different applications/end-devices which are applied into various domains, i.e., smart home, smart city, and smart health, etc [38]. Unless otherwise stated, the general setup of the considered ANN is given in Table II.

A. Direct Training Versus Double Training Approach

In this section, the performance between the conventional approach (direct training with the second data set, $\Psi^{(2)}$) and the proposed approach are examined.

Fig. 4 illustrates the mean square error of both the training and validation set in the second training phase versus the number of epochs of the proposed approach and the conventional approach. The figure is stretched by considering the ED distribution followed by the Matérn cluster process with $\overline{C}_{\text{Mat}} = 2, r_{\text{Mat}} = 2$ and the static and idle power are impaired by Gaussian noise while the ANN architecture is of 4 layers and 55 neurons ($\mathcal{L} = 4, \mathcal{N} = 55$). To be more specific, the curves from the double training approach (the blue curves) are firstly trained by 30000 model-based data, i.e., $\left|\Psi^{(1)}\right| = 30000$, in 150 epochs then re-training by either $\left|\Psi^{(2)}\right| = 500$ (the dash-line blue curve) or $\left|\Psi^{(2)}\right| = 2000$

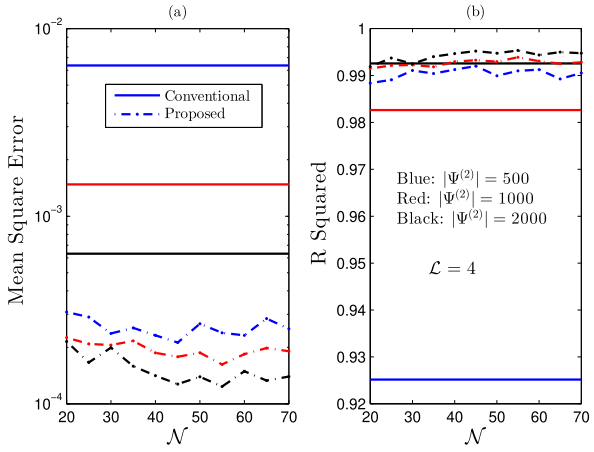

Fig. 5. MSE (a) and $\mathcal{R}^2$ (b) vs. number of neurons of per hidden layer, $\mathcal{N}$, of the validation set with different training data size of the 2nd training phase, i.e., $\left|\Psi^{(2)}\right|$.

(the solid blue curve), in another 150 epochs. The size of the validation set, on the other hand, is always fixed at 500, i.e., $|\Xi| = 500$. The conventional approach (the red curves), on the contrary, is direct training with the same amount of data of $\Psi^{(2)}$ in 300 epochs. We observe that our approach outperforms the conventional one in both the training set and test set. Particularly, our approach is better almost ten times compared to its counterpart. Additionally, it is expected that increasing the training set slightly ameliorates the MSE performance under the proposed method as its weights and bias have already been configured to the near-optimal values after the 1st training phase. The conventional one, differently, improves dramatically by the soaring number of training data. Fig. 4 also reveals that increasing the number of epochs is not a wise choice to enhance the performance of the MSE, especially in the direct training approach where the MSE starts constantly from around 50 epochs (case $\left|\Psi^{(2)}\right| = 500$). Moreover, we experience that 300 and 150 epochs are sufficient to obtain stable results for both the conventional and proposed approaches. As a result, in the sequel, this number of epochs is yielded if not explicitly stated.

Fig. 5 shows the performance of the MSE and $\mathcal{R}^2$ of both the conventional and the proposed methods versus the number of neurons, $\mathcal{N}$, with different training data sizes of the second phase, $|\Psi^{(2)}|$. Each point of all curves is chosen from the best epoch that provides the minimum MSE or the maximum $\mathcal{R}^2$ among all epochs. The specified setting of this figure is following: The ED distribution is followed by Thomas cluster process with $\overline{C}_{\text{Tho}} = 2, \sigma_{\text{Tho}} = 0.5$ while there is no impairment on the static and idle power. The number of layer of the ANN is fixed at 4, $\mathcal{L} = 4$ and the size of the validation set is of 500, i.e., $|\Xi| = 500$. The results are measured on the validation set and it should be clarified that the best epoch over the validation set is not necessarily the same over the training set. We see that the performance of the conventional approach (the solid curves) is constant with $\mathcal{N}$ and is regardless of the metrics, i.e., MSE and $\mathcal{R}^2$. It can be interpreted that the higher number of neurons, $\mathcal{N}$, is not necessarily the better performance under the direct training approach. The double training approach (the dash-dot

TABLE III

MSE, MAE AND $\mathcal{R}^2$ OF TEST SET WITH $\mathcal{L}=4$, $\mathcal{N}=55$, AND $\left|\Psi^{(2)}\right|=2000$; MATÉRN CLUSTER PROCESS WITH $\overline{C}_{\text{Mat}}=2$, $r_{\text{Mat}}=2$

	Gaussian distribution			Uniform distribution			Noise-free		
	MSE	MAE	$\mathcal{R}^2$	MSE	MAE	$\mathcal{R}^2$	MSE	MAE	$\mathcal{R}^2$
Conventional	$2.37\bullet10^{-4}$	0.00407	0.98603	$3.13\bullet10^{-4}$	0.00526	0.98272	$1.05\bullet10^{-4}$	0.00305	0.99458
Proposed	$\mathbf{8.85\bullet10^{-5}}$	**0.00146**	**0.99496**	$\mathbf{9.24\bullet10^{-5}}$	**0.00144**	**0.99389**	$\mathbf{8.34\bullet10^{-5}}$	**0.0013**	**0.99612**

curves), on the contrary, is fluctuated and the largest number or neurons, $\mathcal{N}=70$, is not necessary the best performance, the best one, however, is case $\mathcal{N}=55$ where the MSE $=1.236\times10^{-4}$ and $\mathcal{R}^2=0.9953$ for case $\left|\Psi^{(2)}\right|=2000$. It is interesting to point out that although an increasing number of neurons is not always beneficial, it is always true that the larger the training set, the better the performance of the ANN. Indeed, by rising $\left|\Psi^{(2)}\right|$ from 500 to 2000, the $\mathcal{R}^2$ of the conventional approach improves from 0.9252 to 0.9926 and is better than the $\mathcal{R}^2$ of proposed approach case $\left|\Psi^{(2)}\right|=500$. Nevertheless, the MSE of the double training approach is never worse than the direct training one. Even taking the worst case from the proposed method into comparison, i.e., $\left|\Psi^{(2)}\right|=500$, its performance is still better than the best one of the conventional approach, i.e., $\left|\Psi^{(2)}\right|=2000$. In addition, in Fig. 5(b), we also experience that the $\mathcal{R}^2$ of the double training one is never lower than 0.985. It means that only around 1.5% the variability of the test set can not be explained by the ANN.

Table III illustrates the MSE, $\mathcal{R}^2$, and mean absolute error (MAE), which refers to the arithmetic average of the absolute errors between the observed and predicted values, under the test set, $|\Upsilon|=500$, of the Matérn cluster process with $\overline{C}_{\text{Mat}}=2, r_{\text{Mat}}=2$ under different additive noises, i.e., Gaussian, uniform and the absence of noise. The ANN architecture for both the conventional and proposed approaches are of $\mathcal{L}=4$, $\mathcal{N}=55$. The results shown in Table III are obtained by employing the best epoch in the validation phase to evaluate the ANN with the test set. There is no doubt that the proposed approach consistently outperforms the conventional approach of all considered metrics under the impact of different kinds of noise power. Particularly, the MSE of the proposed ANN networks is better by approximately 10-fold compared to the MSE of the conventional training approach. However, the gap of the $\mathcal{R}^2$ between the two schemes is minor.

B. Partial vs. Full Optimum Architecture

In the following, we are going to study the performance of the proposed approach by measuring the performance of the partial optimum, N_1^*, versus the full optimum architecture, N_{1+2}^*. The partial optimum architecture means that we select the best topology based on $\Psi^{(1)}$, and solely utilize this architecture for $\Psi^{(2)}$. It is re-emphasized that the selection is based only on the training set. The full optimum architecture, N_{1+2}^*, selects the architecture that provides the best performance from both $\Psi^{(1)}$ and $\Psi^{(2)}$.

Fig. 6 unveils the performance of the MSE in both the training set and validation set of the second phase versus the number of epochs with various optimum architecture, i.e., partial and full optimum. In Fig. 6, in order to have a fair comparison,

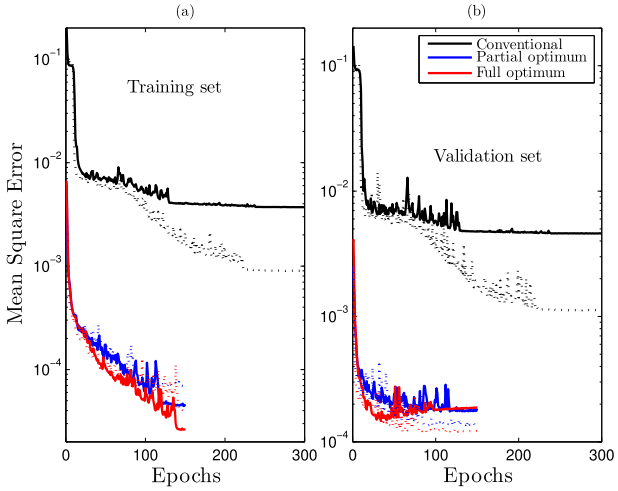

Fig. 6. MSE vs. number of epochs in training set (a) and validation set (b) of $\Psi^{(2)}$; Matérn cluster process with $\overline{C}_{\text{Mat}}=6$, $r_{\text{Mat}}=2$ and $\omega_x, x\in\{\text{sta,idl}\}$, follows uniform distribution.

the curves from the conventional approach is also chosen from the best architecture in the same search space as the proposed approach, i.e., $\mathcal{T}=(\mathcal{L}_{\max}-\mathcal{L}_{\min})(\mathcal{N}_{\max}-\mathcal{N}_{\min})$. The conventional approach still underperforms both optimum architecture from the double training approach though the best architecture has been selected. Here, the curves 'conventional' are the direct training approach and selected from the best architecture based on $\Psi^{(2)}$ while the curves 'partial optimum' and 'full optimum' are the proposed approach and selected from the best architecture based on the $\Psi^{(1)}$ and $\Psi^{(1)}+\Psi^{(2)}$, respectively. In Fig. 6(a), it is crystal clear that the full optimum architecture is superior compared to the partial one. Nevertheless, the gap between these curves, generally, is negligible, for example, 2.64×10^{-5} vs. 4.67×10^{-5} for $\left|\Psi^{(2)}\right|=500$ (the solid lines), 3.99×10^{-5} vs. 6.79×10^{-5} for $\left|\Psi^{(2)}\right|=1000$ (the dot lines), and 1.34×10^{-4} vs. 1.19×10^{-4} for $|\Xi|=500$. In addition, the expenses for this superior performance are quite expensive in terms of the consumed resources, namely, in order to attain the best architecture under the full optimum approach, each topology requires training exactly 300 epochs like the conventional approach while the partial optimum approach requires only half of each, i.e., 150 epochs. Mathematical speaking, assuming that the search space has $\mathcal{T}=(\mathcal{L}_{\max}-\mathcal{L}_{\min})(\mathcal{N}_{\max}-\mathcal{N}_{\min})$ architecture, then both the conventional and full optimum approach need $300\mathcal{T}$ epochs to attain the best architecture while the partial optimum approach, the epochs' requirements are only $150(\mathcal{T}+1)$. Again, we experience that surging the training data will monotonically increase the performance of the ANN. As a consequence, in the next figure, we investigate

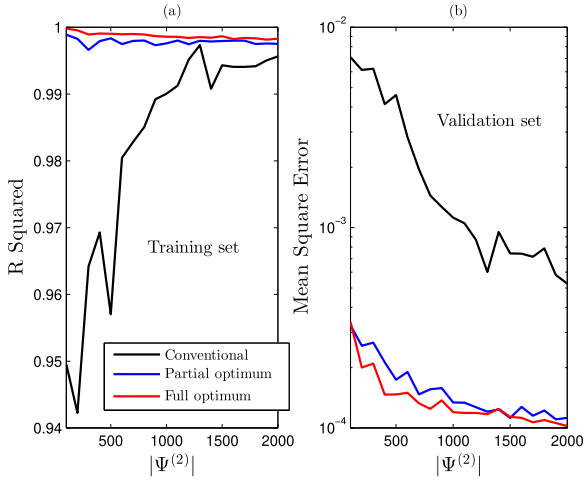

Fig. 7. $\mathcal{R}^2$ (a) and MSE (b) versus $\left|\Psi^{(2)}\right|$. Thomas cluster process with average $\overline{C}_{\text{Tho}} = 6$ and $\sigma_{\text{Tho}} = 0.5$ and $\omega_x, x \in \{\text{sta,idl}\}$, follows uniform distribution.

in detail the impact of training data on the performance of MSE and $\mathcal{R}^2$.

Fig. 7 illustrates the performance of $\mathcal{R}^2$ (a) and MSE (b) versus $|\Psi^{(2)}|$. We observe the same pattern as Fig. 5 that, in general, increasing training data will monotonically improve $\mathcal{R}^2$ and decline the MSE of all architectures, i.e., conventional, partial, and full optimum. However, the pace of improvement among these approaches is different, namely, the direct training enhances almost 10 times in terms of the MSE when training data rises from 100 to 2000. Concerning the partial and full optimum approaches, it merely increases around 3 times. It is noted that like Fig. 6, the conventional curves are direct training approach and selected from the best architecture based on $\Psi^{(2)}$ while the curves 'partial optimum' and 'full optimum' are proposed approaches and selected from the best architecture based on the $\Psi^{(1)}$ and $\Psi^{(1)} + \Psi^{(2)}$, respectively. As for the $\mathcal{R}^2$, the performance of the partial and optimum approaches are nearly stable and close to one while the remained one scales up around 5% in case the training data goes from 100 to 2000. From Fig. 7, it is apparent that in order to obtain the MSE equal to 10^{-4}, 2000 training data is sufficient for both optimum architecture. Furthermore, this figure also confirms the above statement that the gap between the partial and optimum approaches is negligible.

Fig. 8 shows the optimal transmit power, P^*_{tx}, of three different ways, i.e., the conventional, the partial and the full optimum architecture, versus the validation set where the samples are sorted in ascending order. The figure is stretched by assuming that the ED followed by the Matérn cluster process with $\overline{C}_{\text{Mat}} = 6, r_{\text{Mat}} = 2$ and the static and idle power consumption are impaired by an additive noise which follows uniform distribution. The data size for the second training and validation phase are $\left|\Psi^{(2)}\right| = 500$, and $|\Xi| = 500$, respectively. Here, the 'Observed' curves mean the observation output, and the 'predicted' curves mean the predicted output from the ANN. Looking at this figure, there is not surprising that the direct training approach experiences the highest error between the observed output and the predicted one while the error under the partial and full optimum are relatively small.

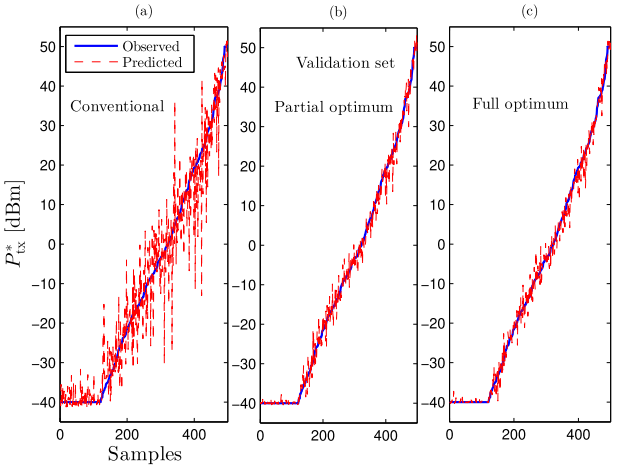

Fig. 8. P^*_{tx} versus the validation set (sorted in ascending order) under various case studies, e.g., conventional (a), partial optimum (b) and fully optimum (c).

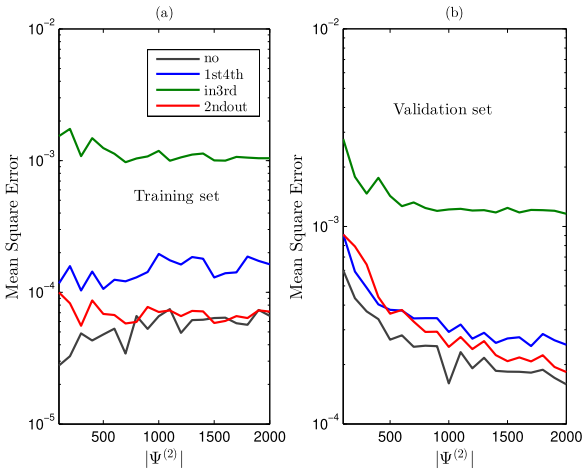

Fig. 9. MSE of the training set (a) and validation set (b) versus training data size; Thomas cluster process with $\overline{C}_{\text{Tho}} = 6$ and $\sigma_{\text{Tho}} = 0.5$ and ω_x, $x \in \{\text{sta,idl}\}$, follows Gaussian distribution.

The MAE under the conventional, the partial, and the full optimum approaches in Fig. 8 are 4.14, 1.47, and 1.35 dBm, respectively. It is noted that this MAE is computed based on the original range of the P^*_{tx}. Hence, if we compute the MAE after normalizing by (9), the MAE of these approaches is smaller, i.e., 0.046, 0.009, and 0.008 for the conventional, partial, and full optimum architecture.

C. Re-Train Whole vs. Parts of the ANN

Fig. 9 illustrates the MSE in both training set and validation set versus $|\Psi^{(2)}|$ by freezing some layers of the considered ANN during the second training phase. The partial optimum architecture with 4 layers and 55 neurons is adopted in this figure. In particular, we consider four distinct scenarios; i) re-train whole ANN or no freezing denoted as 'no'; ii) freezing all hidden layers denoted as '1st4th'; iii) freezing from the input layer to the 3rd hidden layer denoted as 'in3rd,' and iv) freezing from the 2nd hidden layer to the output layer denoted as '2ndout.' In the last 3 cases, the ANN only re-trains 2 layers, e.g., the input and output layers for case '1st4th,' the input and the 1st hidden layer for case '2ndout,' and the last hidden layer (the 4th hidden layer) and the output layer for case 'in3rd.' From the figure, we observe a fairly large gap

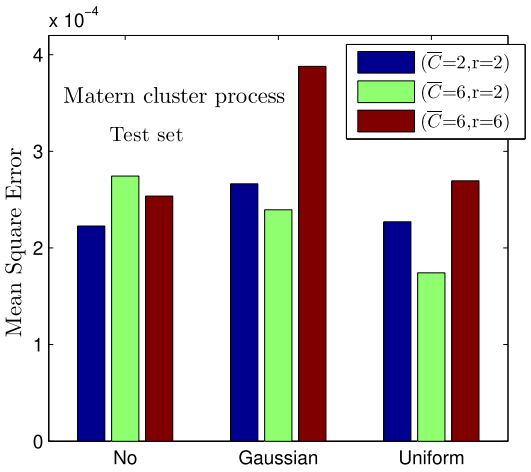

Fig. 10. The MSE of Matérn cluster process with various combinations of the mean of offspring, $\overline{C}_{\text{Mat}}$, and the radius of each cluster, r_{Mat} under different additive noises at static and idle power.

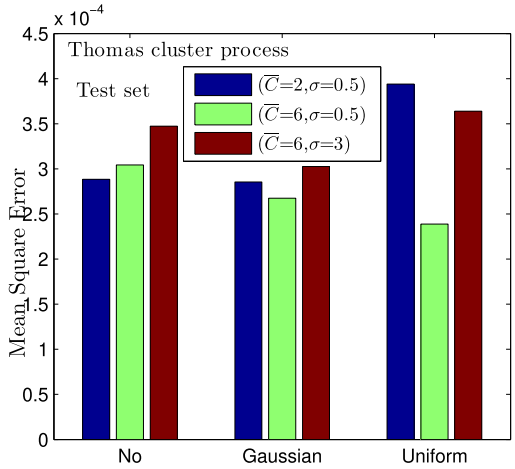

Fig. 11. The MSE of Thomas cluster process with various combinations of the mean of offspring, $\overline{C}_{\text{Tho}}$, and the standard derivation of each cluster, σ_{Mat}, under different additive noises at static and idle power.

between case 'in3rd' and others, in fact, its performance is worse than almost 10 times in both training and validation sets. The second worst case is '1st4th' where the ANN only updates the weights and bias of the input and output layers. The best scenario is, of course, the case without freezing any layers. However, the difference between the best and the second-best case, i.e., case '2ndout,' is significantly small, especially when the training size is adequately large, i.e., around 2000 data. From Fig. 9, we conclude that the impact of freezing the input layer and/or hidden layers close to input on the performance of the ANN is more serious than ones close to the output layer. Moreover, by carefully freezing some layers, the performance of the ANN slightly reduces compared with fully re-trained networks, thus, it is feasible to freeze some layers in order to save time and resources when re-training the neural networks.

D. The Impact of Different Kinds of Additive Noise Distributions and Point Processes

Fig. 10 depicts the MSE of the differences kinds of noises at the static and idle power, i.e., Gaussian, uniform, and without noise under the Matérn cluster process with various combinations of the expected number of offspring, $\overline{C}_{\text{Mat}}$, and the radius of each cluster, r_{Mat}, namely, $\left(\overline{C}_{\text{Mat}}=2, r_{\text{Mat}}=2\right)$, $\left(\overline{C}_{\text{Mat}}=6, r_{\text{Mat}}=2\right)$ and $\left(\overline{C}_{\text{Mat}}=6, r_{\text{Mat}}=6\right)$. The training data used during the second training phase is $\left|\Psi^{(2)}\right|=500$. The results are examined on the test set with $|\Upsilon|=500$. As noted above, the test set is examined based solely on the best epoch in the validation phase. It is evident that increasing the $\overline{C}_{\text{Mat}}$ while fixing r_{Mat} will improve the MSE under Gaussian and uniform distributions; the remaining case, on the other hand, is slightly decreasing. Next, the MSE performance of both Gaussian and uniform cases goes down providing that r_{Mat} is scaling up and $\overline{C}_{\text{Mat}}$ is kept constant. Thirdly, the performance of MSE monotonically plunges if both $\overline{C}_{\text{Mat}}$ and r are rising. Interestingly, the absence of additive noise does not always achieve the best performance among the three scenarios. Specifically, under the $\left(\overline{C}_{\text{Mat}}=6, r_{\text{Mat}}=2\right)$ setup, the best performance is the uniform distribution. Regarding another setup, the best performance is always the absence of additive noise scenario and the Gaussian noise is always the worst performance.

Fig. 11 plots the MSE of the different kinds of noises at the static and idle power, i.e., Gaussian, uniform, and without noise under the Thomas cluster process with various combinations of the expected number of offspring, $\overline{C}_{\text{Tho}}$, and the standard derivation of each cluster, σ_{Tho}. The figure is also plotted on the test set, $|\Upsilon|=500$, by employing the best epoch in the validation phase. The training data used in the second training phase is $\left|\Psi^{(2)}\right|=500$. We realize that the uniform case provides the worst performance under the set up, i.e., $\left(\overline{C}_{\text{Tho}}=2, \sigma_{\text{Tho}}=0.5\right)$ and $\left(\overline{C}_{\text{Tho}}=6, \sigma_{\text{Tho}}=3\right)$. It, however, achieves the best performance under the set up $\left(\overline{C}_{\text{Tho}}=6, \sigma_{\text{Tho}}=0.5\right)$. Furthermore, raising the average number of offspring, $\overline{C}_{\text{Tho}}$, while fixing the standard derivation, σ_{Tho}, will boost the performance of the MSE under the Gaussian and uniform distributions and slightly increases the MSE of case noise-free. This observation is similar to the Matérn cluster process in Fig. 10. Moreover, we also experience the same behavior as Matérn cluster process when increasing both $\overline{C}_{\text{Tho}}$ and σ_{Tho} that declines the performance of the MSE. Interestingly, different from Fig. 10 where the MSE performance of the Gaussian noise is considered the worst case, the MSE of the Gaussian noise in Fig. 11 is consistently better than the absence of noise scenario and only worse than the uniform distribution for the $\left(\overline{C}_{\text{Tho}}=2, \sigma_{\text{Tho}}=0.5\right)$ setup.

In general, the gap between all setups as well as all kinds of additive noises under both Matérn and Thomas cluster process are relatively small and the worst performance is also less than 4×10^{-4} which is generally acceptable for most of the applications in LoRa networks. Additionally, relying on the cluster process, the impact of additive noise that follows Gaussian distribution may outperform or underperform ones followed by uniform distribution and the noise-free case is not necessarily the best.

VI. Conclusion

In this paper, the maximization of the EE with respect to the transmit power is studied by combining the advantages of both model-based in wireless networks and the data-driven of deep learning technique. Our findings show that the proposed

training approach outperforms the conventional one almost 10 times under some scenarios. Moreover, the re-training process can be done either in whole or parts of the ANN. Finally, the application of the proposed approach is generally to cover most of EDs distributions as well as to overcome the hardware impairment at EDs.

REFERENCES

[1] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022 White Paper*, Cisco, San Jose, CA, USA, 2019.

[2] T. T. Lam and M. Di Renzo, "On the energy efficiency of heterogeneous cellular networks with renewable energy sources—A stochastic geometry framework," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6752–6770, Oct. 2020.

[3] C. Goursaud and J. M. Gorce, "Dedicated networks for IoT: PHY/MAC state of the art and challenges," *EAI Endorsed Trans. Internet Things*, vol. 1, no. 1, Oct. 2015, Art. no. 150597.

[4] L.-T. Tu, A. Bradai, and Y. Pousset, "Coverage probability and spectral efficiency analysis of multi-gateway downlink LoRa networks," 2022, *arXiv:2202.05014*.

[5] A. Jindal, G. S. Aujla, N. Kumar, R. Chaudhary, M. S. Obaidat, and I. You, "SeDaTiVe: SDN-enabled deep learning architecture for network traffic control in vehicular cyber-physical systems," *IEEE Netw.*, vol. 32, no. 6, pp. 66–73, Nov. 2018.

[6] A. Jindal, N. Kumar, and J. J. P. C. Rodrigues, "A heuristic-based smart HVAC energy management scheme for university buildings," *IEEE Trans. Ind. Informat.*, vol. 14, no. 11, pp. 5074–5086, Nov. 2018.

[7] A. Zappone, M. Di Renzo, M. Debbah, T. T. Lam, and X. Qian, "Model-aided wireless artificial intelligence: Embedding expert knowledge in deep neural networks for wireless system optimization," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 60–69, Sep. 2019.

[8] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?" *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 7331–7376, Oct. 2019.

[9] L. Y. Pratt, "Discriminability-based transfer between neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 5, 1992, pp. 204–211.

[10] L. Beltramelli, A. Mahmood, P. Osterberg, and M. Gidlund, "LoRa beyond ALOHA: An investigation of alternative random access protocols," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3544–3554, May 2021.

[11] B. Su, Z. Qin, and Q. Ni, "Energy efficient uplink transmissions in LoRa networks," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4960–4972, Aug. 2020.

[12] L.-T. Tu, A. Bradai, Y. Pousset, and A. I. Aravanis, "Energy efficiency analysis of LoRa networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 9, pp. 1881–1885, Sep. 2021.

[13] M. Di Renzo, S. Wang, and X. Xi, "Inhomogeneous double thinning-modeling and analysis of cellular networks by using inhomogeneous Poisson point processes," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5162–5182, Aug. 2018.

[14] B. Paul, "A novel energy-efficient routing scheme for LoRa networks," *IEEE Sensors J.*, vol. 20, no. 15, pp. 8858–8866, Aug. 2020.

[15] N. El Rachkidy, A. Guitton, and M. Kaneko, "Collision resolution protocol for delay and energy efficient LoRa networks," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 2, pp. 535–551, Jun. 2019.

[16] M. N. Ochoa, A. Guizar, M. Maman, and A. Duda, "Evaluating LoRa energy efficiency for adaptive networks: From star to mesh topologies," in *Proc. IEEE WiMob*, Oct. 2017, pp. 1–8.

[17] I. Cheikh, R. Aouami, E. Sabir, M. Sadik, and S. Roy, "Multi-layered energy efficiency in LoRa-WAN networks: A tutorial," *IEEE Access*, vol. 10, pp. 9198–9231, 2022.

[18] C. Rotsos *et al.*, "Ukko: Resilient DRES management for ancillary services using 5G service orchestration," in *Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm)*, Nov. 2020, pp. 1–6.

[19] R. M. Sandoval, A.-J. Garcia-Sanchez, and J. Garcia-Haro, "Optimizing and updating LoRa communication parameters: A machine learning approach," *IEEE Trans. Netw. Service Manage.*, vol. 16, no. 3, pp. 884–895, Sep. 2019.

[20] J. Cho, D. Hwang, and K.-H. Kim, "Improving TDoA based positioning accuracy using machine learning in a LoRaWan environment," in *Proc. IEEE ICOIN*, Kuala Lumpur, Malaysia, Jan. 2019, pp. 469–472.

[21] S. Lee, J. Lee, J. Hwang, and J. K. Choi, "A novel deep learning-based IoT device transmission interval management scheme for enhanced scalability in LoRa networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 11, pp. 2538–2542, Nov. 2021.

[22] S. Cui and I. Joe, "Collision prediction for a low power wide area network using deep learning methods," *J. Commun. Netw.*, vol. 22, no. 3, pp. 205–214, Jun. 2020.

[23] I. Ilahi, M. Usama, M. O. Farooq, M. U. Janjua, and J. Qadir, "LoRaDRL: Deep reinforcement learning based adaptive PHY layer transmission parameters selection for LoRaWAN," in *Proc. IEEE LCN*, Nov. 2020, pp. 457–460.

[24] L. Liu, Y. Yao, Z. Cao, and M. Zhang, "DeepLoRa: Learning accurate path loss model for long distance links in LPWAN," in *Proc. IEEE INFOCOM*, May 2021, pp. 1–10.

[25] A. Javed, H. Larijani, and A. Wixted, "Improving energy consumption of a commercial building with IoT and machine learning," *IT Prof.*, vol. 20, no. 5, pp. 30–38, Sep./Oct. 2018.

[26] G. Shen, J. Zhang, A. Marshall, L. Peng, and X. Wang, "Radio frequency fingerprint identification for LoRa using deep learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2604–2616, Aug. 2021.

[27] Y. Yazid, I. Ez-Zazi, M. Arioua, and A. El Oualkadi, "A deep reinforcement learning approach for LoRa WAN energy optimization," in *Proc. MTTW*, Oct. 2021, pp. 199–204.

[28] P. Kumari, R. Mishra, H. P. Gupta, T. Dutta, and S. K. Das, "An energy efficient smart metering system using edge computing in LoRa network," *IEEE Trans. Sustain. Comput.*, early access, Jan. 6, 2021, doi: 10.1109/TSUSC.2021.3049705.

[29] R. Hamdi, E. Baccour, A. Erbad, M. Qaraqe, and M. Hamdi, "LoRa-RL: Deep reinforcement learning for resource management in hybrid energy LoRa wireless networks," *IEEE Internet Things J.*, vol. 9, no. 9, pp. 6458–6476, May 2022.

[30] M. Haenggi, *Stochastic Geometry for Wireless Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[31] M. Di Renzo, A. Zappone, T. T. Lam, and M. Debbah, "System-level modeling and optimization of the energy efficiency in cellular networks—A stochastic geometry framework," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2539–2556, Apr. 2018.

[32] L.-T. Tu, A. Bradai, Y. Pousset, and A. I. Aravanis, "On the spectral efficiency of LoRa networks: Performance analysis, trends and optimal points of operation," *IEEE Trans. Commun.*, early access, Feb. 2, 2021, doi: 10.1109/TCOMM.2022.3148784.

[33] O. Georgiou and U. Raza, "Low power wide area network analysis: Can LoRa scale?" *IEEE Wireless Commun. Lett.*, vol. 6, no. 2, pp. 162–165, Apr. 2017.

[34] L.-T. Tu, A. Bradai, and Y. Pousset, "A new closed-form expression of the coverage probability for different QoS in LoRa networks," in *Proc. ICC*, Dublin, Ireland, Jun. 2020, pp. 1–6.

[35] E. Talbi. *Optimization of Deep Neural Networks: A Survey and Unified Taxonomy*. Accessed: Apr. 4, 2022. [Online]. Available: https://hal.inria.fr/hal-02570804v2/document

[36] V. Subramanian, *Deep Learning With PyTorch: A Practical Approach to Building Neural Network Models Using PyTorch*. Birmingham, U.K.: Packt Publishing,2018.

[37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[38] *IEEE 802.16p-11/0014, IEEE 802.16p Machine to Machine (M2M) Evaluation Methodology Document*. Accessed: Apr. 4, 2022. [Online]. Available: http://ieee802.org/16/m2m/index.html

Lam-Thanh Tu was born in Ho Chi Minh City, Vietnam. He received the B.Sc. degree in electronics and telecommunications engineering from the Ho Chi Minh City University of Technology, Vietnam, in 2009, the M.Sc. degree in telecommunications engineering from the Posts and Telecommunications Institute of Technology, Vietnam, in 2014, and the Ph.D. degree from the Laboratory of Signals and Systems, Paris-Saclay University, Paris, France, in 2018. From 2015 to 2018, he was with the French National Center for Scientific Research (CNRS), Paris, as an Early Stage Researcher of the European-Funded Project H2020 ETN-5Gwireless. From 2019 to 2021, he was with the Institute Xlim, University of Poitiers, Poitiers, France, as a Post-Doctoral Research Fellow. He is currently with the Department of Telecommunications, Ton Duc Thang University, Ho Chi Minh City. His research interests include stochastic geometry, LoRa networks, physical layer security, energy harvesting, and machine learning applications for wireless communications. He was an IEEE TRANSACTIONS ON COMMUNICATIONS Exemplary Reviewer for 2016 and a recipient of the 2017 IEEE SigTelCom Best Paper Award. He was the Assistant Project Manager of the H2020 MCSA 5Gwireless and 5Gaura projects.

Abbas Bradai (Senior Member, IEEE) received the Ph.D. degree from the University of Bordeaux, France, in 2012. He has been working as an Associate Professor with the University of Poitiers and a member of the XLIM Research Institute (CNRS UMR 7252) since September 2015. He is/was involved in many French and European projects (FP7 and H2020) such as ENVISION, VITAL, SAFE, EcoNaviConnect, and SAMBAS. His research interests include multimedia communications over wired and wireless networks, the IoT, software defined networks, and virtualization.

Olfa Ben Ahmed received the Ph.D. degree in computer science from the LaBRI, University of Bordeaux, Talence, France, in January 2015. She is currently an Associate Professor with the University of Poitiers, Poitiers, France; and the XLIM Research Center. She served as a Research and Teaching Assistant with the University of Bordeaux, ENSEIRB MATMECA, and the University of Limoges. She was a Post-Doctoral Researcher with the, EURECOM Research Center, Data Science Department, Nice, France. Her research interests include artificial intelligence, computer vision, image processing, machine learning, and data classification and retrieval. She was the recipient of the Best Poster Presentation Award at the Bioimaging Conference in 2017. She was awarded with her team with the MediaEval Distinctive Mention 2017 for best achieved results in video interestingness prediction task in Benchmarking Initiative for Multimedia Evaluation.

Sahil Garg (Member, IEEE) received the Ph.D. degree from the Thapar Institute of Engineering and Technology, Patiala, India, in 2018. He is currently a Research Associate with the Resilient Machine Learning Institute (ReMI) in correlation with École de Technologie Supérieure (ÉTS), Montreal. Prior to this, he worked as a Post-Doctoral Research Fellow with ÉTS, Montreal; and the MITACS Researcher at Ericsson, Montreal. He has over 80 publications in high ranked journals and conferences, including over 50 top-tier journal articles and over 30 reputed conference articles. Some of his research findings are published in top-cited journals such as IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON CLOUD COMPUTING, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, IEEE TRANSACTIONS ON SUSTAINABLE COMPUTING, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTING, IEEE INTERNET OF THINGS JOURNAL, IEEE SYSTEMS JOURNAL, *IEEE Communications Magazine*, *IEEE Network Magazine*, IEEE WIRELESS COMMUNICATIONS, *IEEE Consumer Electronics Magazine*, *FGCS*, *JPDC*, *Applied Soft Computing*, and *Information Sciences* including various international conferences of repute such as IEEE Globecom, IEEE ICC, IEEE WCNC, IEEE CCNC, IEEE VTC, IEEE Infocom Workshops, ACM MobiCom Workshops, and ACM MobiHoc Workshops. He has many research contributions in the area of machine learning, big data analytics, knowledge discovery, cloud computing, the Internet of Things, software defined networking, and vehicular ad-hoc networks. He is a member of the IEEE Communications Society, IEEE Industrial Electronics Society, IEEE Software Defined Networks Community, IEEE Smart Grid Community, ACM, and IAENG. He was the recipient of the prestigious Visvesvaraya Ph.D. Fellowship from the Ministry of Electronics and Information Technology under Government of India from 2016 to 2018. He has been awarded the 2021 IEEE SYSTEMS JOURNAL Best Paper Award; the 2020 IEEE TCSC Award for Excellence in Scalable Computing (Early Career Researcher); and the IEEE ICC Best Paper Award at Kansas City, MO, USA, in 2018. He also serves as the Workshops and Symposia Officer for the IEEE ComSoc ETI on Aerial Communications. He guest edited a number of special issues in top-cited journals, including IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE INTERNET OF THINGS JOURNAL, *IEEE Network Magazine*, *FGCS*, and *NCAA*. He also served as the TPC co-chair/publicity co-chair/special sessions chair/publication chair for several conferences. He also served as the Workshop Co-Chair for different workshops in IEEE/ACM conferences including IEEE Infocom, IEEE Globecom, and ACM MobiCom. He also served as the Symposium Chair for Aerial Communications Track in IEEE ICC 2022, Seoul, South Korea. He is currently a Managing Editor of *Human-Centric Computing and Information Sciences* (HCIS) (Springer); and an Associate Editor of *IEEE Network Magazine*, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, *Applied Soft Computing* (ASoC) (Elsevier), and *International Journal of Communication Systems* (IJCS) (Wiley).

Yannis Pousset was born in 1971. He received the Ph.D. degree in mobile radio communication from the University of Poitiers in 1998. Since 2012, he has been a Professor with the Department of Electrical Engineering, University of Poitiers. He develops its research activities in the XLIM Laboratory. His research interest includes the study of adaptive links related to the optimal transmission of data over wireless spatio-temporal radio channel.

Georges Kaddoum (Member, IEEE) received the bachelor's degree in electrical engineering from the École Nationale Supérieure de Techniques Avancées (ENSTA Bretagne), Brest, France, in 2004, the M.S. degree in telecommunications and signal processing (circuits, systems, and signal processing) from the Université de Bretagne Occidentale and Telecom Bretagne (ENSTB), Brest, in 2005, and the Ph.D. degree (Hons.) in signal processing and telecommunications from the National Institute of Applied Sciences (INSA), University of Toulouse, Toulouse, France, in 2009. He is currently an Associate Professor and the Tier 2 Canada Research Chair with the École de Technologie Supérieure (ÉTS), Université du Québec, Montreal, Canada. In 2014, he was awarded the ÉTS Research Chair in physical-layer security for wireless networks. Since 2010, he has been a Scientific Consultant in the field of space and wireless telecommunications for several U.S. and Canadian companies. He has published over 200 journals and conference papers and has two pending patents. His recent research activities cover mobile communication systems, modulations, security, and space communications, and navigation. He received the Best Papers Awards at the 2014 IEEE International Conference on Wireless and Mobile Computing, Networking, Communications (WIMOB), with three coauthors; and at the 2017 IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), with four coauthors. He received IEEE TRANSACTIONS ON COMMUNICATIONS Exemplary Reviewer Award for the years 2015, 2017, and 2019. He received the Research Excellence Award of the Université du Québec in the year 2018. In the year 2019, he received the Research Excellence Award from the ÉTS in recognition of his outstanding research outcomes. He is currently serving as an Associate Editor for IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY and IEEE COMMUNICATIONS LETTERS.