

A Generalized CNN Model with Automatic Hyperparameter Tuning for Millimeter Wave Channel Prediction

Chengfang Yue, Hui Tang, Jun Yang, and Li Chai

Abstract—This paper focuses on millimeter wave (mmWave) channel prediction by machine learning (ML) methods. Previous ML-based mmWave channel predictors have limitations on requirements of the amount of training data, model generalization ability, robustness to noise, etc. In this paper, we propose a CNN model with a novel feature selection strategy for mmWave channel prediction. Automatic hyperparameter tuning (AHT) algorithms are embedded in the training process to iteratively optimize the predictive performance of the proposed CNN. The diversification strategy is leveraged to enhance the robustness of the AHT procedure against different communication scenarios. To improve the generalization ability of the prediction model, the input features are designed to capture the correlation between the physical environment and the channel characteristics. In parallel, the Cartesian coordinates of the transmitter (Tx) and receiver (Rx) are transformed into polar ones to reduce the model's sensitivity to coordinate noise. Numerical results demonstrate the effectiveness of the proposed CNN model in predicting mmWave channel characteristics in various communication scenarios.

Index Terms—Automatic hyperparameter tuning, channel prediction, CNN, diversification strategy, feature selection, mmWave communication.

I. INTRODUCTION

THE design of wireless communication systems in specific scenarios relies heavily on knowledge of the channels [1]. However, in millimeter wave (mmWave) communication systems, the implementation of hybrid analog/digital array architecture makes it challenging to obtain the channels directly [2]. Traditional channel modeling methodologies, such as deterministic or stochastic, face significant challenges when applied to mmWave channels, including high implementation overheads, heavy computational burden, and increased data traffic [3]. Alternatively, channels can be studied from the perspective of parameters since channel parameters aid in the design of mmWave communication systems. For instance, path loss helps understand the coverage of a base station, time

delay indicates communication latency, and angles of arrival and departure contribute to forming highly directional beams.

Machine learning (ML) is a powerful tool for channel modeling and parameter prediction due to its self-learning, non-linear fitting, and big data processing capabilities [4]–[11]. For example, [5] proposed an artificial neural network (ANN) based model to playback measured channels, where three-dimensional coordinates of antennas and carrier frequency were selected as input features, while cross-polarization ratio (XPR), amplitude, delay, and phase were chosen as output labels. [7]–[9] applied neural network (NN) [7], relevance vector machine (RVM) [8], and multi-layer perceptron (MLP) [9] to predict the large-scale path loss (PL) and received power (RP), respectively. In [7]–[9], the height of Tx and Rx antennas, the distance of Tx and Rx antennas, carrier frequency, and diffraction loss were selected as input features to learn and predict PL and RP. The RVM model was utilized to estimate the angle of arrival (AoA) [10]. The procedure was to estimate the rough positions of the signals by sparse RVM and then to obtain the precise estimated angles by searching the preset spatial grid. In [11], CNN was used to predict mmWave channel parameters using coordinates of transmitters (Tx) and receivers (Rx), highlighting the effect of data acquisition methods on prediction performance.

The generalization ability of most state-of-the-art channel prediction models is inadequate, which hampers the availability of ML-based channel prediction methods. Specifically, deploying a trained channel prediction model in new communication systems necessitates conducting repeated channel soundings to extract channel parameters and build new training databases. However, channel sounding is both costly and challenging to implement, particularly for mmWave communication systems. To reduce the labeling overhead associated with channel sounding and facilitate ML-based channel prediction, one approach is to devise a model that is adaptable to diverse communication scenarios.

On the other hand, it is well-known that the empirical performance of ML models relies significantly on hyperparameters, which should be carefully tuned. However, manual hyperparameter tuning is a challenging and tedious task that often leads to suboptimal solutions [12]. To address this issue, various automatic hyperparameter optimization (AHT) techniques were proposed, such as random search, Bayesian optimization, and Gaussian process [12]–[17]. In the field of ML-based channel prediction, the robustness of the AHT process over different datasets is relatively under-researched.

Manuscript received January 31, 2023; revised April 4, 2023; approved for publication by Chien-Ming Chen, Division 2 Editor, May 15, 2023.

This work was supported by the National Natural Science Foundation of China under grants 62173259.

C. Yue is with the Wuhan University of Science and Technology, Wuhan, 430081, China, email: yuechengfang@wust.edu.cn.

H. Tang and Jun Yang are with the Engineering Research Center of Metallurgical Automation and Measurement Technology, Wuhan University of Science and Technology, Wuhan, 430081, China, email: {htang, yangjun}@wust.edu.cn. Li Chai is with the College of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, China, email: chaili@zju.edu.cn.

H. Tang is the corresponding author.

Digital Object Identifier: 10.23919/JCN.2023.000024

Creative Commons Attribution-NonCommercial (CC BY-NC).

This is an Open Access article distributed under the terms of Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided that the original work is properly cited.

Coordinates of TxS and RxS were commonly selected as input features for channel modeling and prediction [5], [11], [18]. In situations where the positions of antennas fluctuate dynamically (e.g., antennas mounted on drones and unmanned surface vehicles), coordinate noise can not be ignored. However, these problems have not been well addressed in existing literature.

This paper introduces an automatic hyperparameter tuning algorithm-assisted CNN model (AHT-assisted CNN) to predict various channel parameters. Our work aims to address the challenges of reducing offline training efforts, improving the applicability of the prediction model to different communication systems with less channel sounding overhead, and enhancing the robustness of the framework to coordinate noise. The main contributions of this study can be summarized as follows:

- The AHT-assisted CNN model proposes the integration of AHT algorithms into the CNN training process to optimize the hyperparameters in an iterative manner by considering the tuning process as an optimization problem of the unknown black-box function [13], [14]. The AHT procedure is limited by inherent vulnerability, which means a single tuning algorithm might converge rapidly for one database but then perform poorly for another [17]. To overcome this limitation, we take advantage of the diversification strategy to make the AHT procedure robust across different databases, thus saving offline training effort.
- We save the labeling overhead associated with channel sounding by improving the generalization ability of the CNN model and reducing the amount of training data required. To achieve this purpose, the input features are designed according to the mmWave propagation mechanism to capture the intrinsic interaction between the physical environment and the corresponding channel. Consequently, the input features are determined as the received power, azimuth mean angle of departure, elevation mean angle of departure, and coordinates of TxS and RxS, which allows us to predict the path loss, azimuth mean angle of arrival, elevation mean angle of arrival, and delay mean. Numerical results show that such a feature selection strategy improves the generalization ability of the channel prediction model and requires less training data to achieve satisfactory prediction accuracy.
- The recorded Cartesian coordinates of the TxS and RxS antennas are transformed into polar ones to enable the model to be robust to coordinate noise. We demonstrate that the featured polar coordinates are less sensitive to coordinate noise, which improves the predictive performance of the model in situations where the actual coordinates of the antennas fluctuate.

To create simulation datasets, the ray tracing software Wireless InSite [19] is utilized. Ray tracing is a well-established deterministic technique that is commonly used for modeling radio propagation. This technique is based on the principles of geometrical optics (GO) and the uniform theory of diffraction (UTD). Various interactions between rays and objects,

TABLE I
ABBREVIATIONS AND CORRESPONDING FULL NAMES OF CHANNEL PARAMETERS.

ABB	PL	RP
FN	Path loss	Received power
ABB	AoA	AoD
FN	Angle of arrival	Angle of departure
ABB	AMAAoA	EMAAoA
FN	Azimuth mean AoA	Elevation mean AoA
ABB	AMAAoD	EMAAoD
FN	Azimuth mean AoD	Elevation mean AoD

such as reflection, transmission, scattering, and diffraction, can be categorized using this approach. Numerical results show that the presented AHT-assisted CNN can achieve state-of-the-art performance in various aspects, including less training efforts, smaller training datasets, stronger generalization ability, and robustness to coordinate noise. We present the abbreviations (ABB) and corresponding full names (FN) of the involved channel parameters in Table I to avoid confusion.

II. SYSTEM MODEL

We consider an mmWave communication system with K randomly distributed RxS and the same number of TxS, where both TxS and RxS are equipped with a single antenna. The operating frequency is 100 GHz. M Tx-Rx pairs are established, where M equals $K \times K$. Then the received signal of the m th Tx-Rx pair can be written as

$$y_m = \mathbf{h}_m(f)x + n_m, m = 1, 2, \dots, M, \quad (1)$$

where x denotes the transmitted signal, $n_m \sim \mathcal{CN}(\mu, \sigma^2)$ is the additive white Gaussian noise at user m , σ^2 denotes the variance of noise, $\mathbf{h}_m(f)$ represent the channel for the m th Tx-Rx pair. Let f represent the carrier frequency, α_{ml} , τ_{ml} , β_{ml} , ϕ_{ml} , and θ_{ml} refer to the attenuation, time delay, phase shift, azimuth AoAs, and elevation AoAs associated with the l th path on the m th channel, respectively. Let L be the number of paths, and $\mathbf{a}_m(\phi_{ml}, \theta_{ml})$ be the m th antenna response vector. As in [20], the channel for the m th Tx-Rx pair is defined as

$$\mathbf{h}_m(f) = \sum_{l=1}^L \alpha_{ml} e^{-j2\pi f \tau_{ml} + j\beta_{ml}} \mathbf{a}_m(\phi_{ml}, \theta_{ml}), \quad (2)$$

where $m = 1, 2, \dots, M$. By denoting the path loss (PL), azimuth mean angle of arrival (AMAAoA), elevation mean angle of arrival (EMAAoA), and delay mean (DM) of the m th channel by ρ_m , $\bar{\phi}_m$, $\bar{\theta}_m$, and $\bar{\tau}_m$, respectively, we define the mapping from $\mathbf{h}_m(f)$ to channel characteristics according to (2) as follows,

$$\mathcal{F} : \mathbf{h}_m(f) \rightarrow [\rho_m, \bar{\phi}_m, \bar{\theta}_m, \bar{\tau}_m], m = 1, 2, \dots, M. \quad (3)$$

It is worth noting that we choose the mean arrival angles and mean time delay to define the above mapping function because the time delay and angles of each path are difficult to obtain and process during the model training procedure. Alternatively, statistical channel characteristics such as angle spread and delay spread can be used to model channels instead of mean

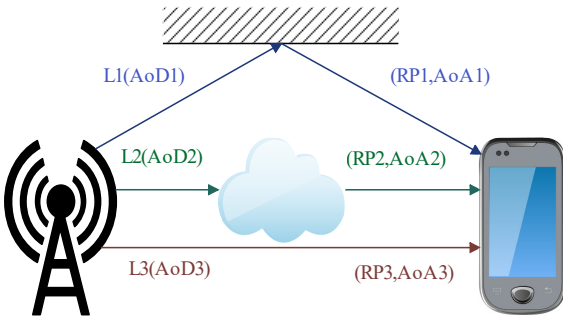


Fig. 1. Illustration for multipaths distribution.

angles and mean delay. Similarly, the mean values of the departure angles (AMaOD and EMaOD) are used in this paper. The phase shift β_i is not considered in (3) because multipath phases of the channel are immeasurable in practice and are assumed to be uniformly distributed in the range $[0, 2\pi)$ [21].

The channel $\mathbf{h}_m(f)$ in a certain communication scenario is affected by the distribution of multipaths, which is determined by the environment and antenna coordinates. As illustrated in Fig. 1, different multipaths are distinguished by AoDs, AoAs, and RP. In view of this, we choose coordinates, AoDs, AoAs, and RP as environment factors and define another mapping between the environment factors and the corresponding channel as

$$\mathcal{F}_* : [\mathbf{C}_m, \bar{\varphi}_m, \bar{\vartheta}_m, p_m] \rightarrow \mathbf{h}_m(f), \quad (4)$$

where \mathbf{C}_m denotes the vector of Cartesian coordinates of the m th Tx-Rx pair, that is $\mathbf{C}_m = [x_m^T, y_m^T, x_m^R, y_m^R]$, while $\bar{\varphi}_m$, $\bar{\vartheta}_m$, and p_m represent AMaOD, EMaOD and RP of the m th Tx-Rx channel, respectively. By combining (3) and (4), we obtain the mapping between the environment factors and the channel parameters as

$$\tilde{\mathcal{F}} : [\mathbf{C}_m, \bar{\varphi}_m, \bar{\vartheta}_m, p_m] \rightarrow [\rho_m, \bar{\phi}_m, \bar{\theta}_m, \bar{\tau}_m]. \quad (5)$$

As the mapping $\tilde{\mathcal{F}}$ indicated, the channel parameters vector $[\rho_m, \bar{\phi}_m, \bar{\theta}_m, \bar{\tau}_m]$ is closely related to the environment factors $[\mathbf{C}_m, \bar{\varphi}_m, \bar{\vartheta}_m, p_m]$.

When designing practical communication systems in different scenarios, important properties such as base station coverage, antenna orientation, and communication latency, can be investigated with the knowledge of path loss, arrival and departure angles, and time delay instead of repeatedly estimating or sounding channels. Our goal is to obtain such channel parameters from a new perspective in situations where channels are difficult to estimate or repeatedly sound. Specifically, we aim to construct a generalized CNN model which can predict channel characteristics for various different communication scenarios by learning the mapping $\tilde{\mathcal{F}}$ from datasets. Furthermore, the model is robust to coordinate noise.

III. CNN-BASED CHANNEL PREDICTION

The mapping $\tilde{\mathcal{F}}$ defined in Section II represents the interaction between the communication scenario and the corresponding channel parameters. $\tilde{\mathcal{F}}$ includes various aspects of

the communication process, such as the scenario geometry, materials, reflection, etc. Therefore, it is difficult to find an analytical model of $\tilde{\mathcal{F}}$.

In this section, we propose a CNN model for learning the mapping $\tilde{\mathcal{F}}$. We first introduce the procedure of the feature selection and preprocessing, and how it contributes to improving the model's generalization ability and robustness to coordinate noise. Afterwards, we explain the motivation and architecture of the CNN.

A. Feature Selection and Preprocessing

The correlation between features and labels is considered to be highly relevant to the predictive performance of the model. One focus of this paper is to construct highly correlated input features and output labels according to the mmWave communication mechanism. In particular, it is well known that path loss is related to received power. The antenna response between the m th Tx-Rx pair $\mathbf{A}(\Psi_m, \Theta_m)$ is defined as follows,

$$\mathbf{A}(\Psi_m, \Theta_m) = \mathbf{a}_R(\phi_m, \theta_m) \mathbf{a}_T^*(\varphi_m, \vartheta_m), \quad (6)$$

where ϕ_m and θ_m are the azimuth and elevation AoAs of the m th Tx-Rx pair, respectively, while φ_m and ϑ_m refer to the azimuth and elevation AoDs of the m th Tx-Rx pair, respectively [22]. Based on the fact that the AoD and AoA enjoy reciprocity [23], an intuitive inference from (6) is that the AoAs (ϕ_m and θ_m) are coupled with AoDs (φ_m and ϑ_m). In addition, the coordinates indicate the positional relationship between the Tx and the corresponding Rx, which affects the distribution of channels. In summary, the environment factors $[\mathbf{C}, \bar{\varphi}, \bar{\vartheta}, p]$ are coupled with the channel characteristic vector $[\rho, \bar{\phi}, \bar{\theta}, \bar{\tau}]$.

Hence, different from the existing literature, we choose $[\mathbf{C}, \bar{\varphi}, \bar{\vartheta}, p]$ and $[\rho, \bar{\phi}, \bar{\theta}, \bar{\tau}]$ as the input features and output labels, respectively, to enable the model to exploit the correlation between the featured and labeled channel parameters. The communication scenario (geometry, scattering, communication settings, etc.) determines the featured environment factors. Thus the ML model can learn the mapping between the environment information and the corresponding labeled channel parameters, allowing the model to adapt to different scenarios.

It is necessary to clarify the reasonableness of the proposed feature selection strategy as follows. Although the path loss can be directly obtained via power calculation, we still predict the path loss using the received power for the following considerations. As depicted in Fig. 1, the received power is a metric of both the propagation distance and the scattering distribution. Specifically, multipath L1, L2, and L3 experience different propagation distances and scatters, where L1 is a reflective path, L2 is a non-LoS path, and L3 is an LoS path. The received powers resulting from the three paths (RP1, RP2, and RP3) are different, which means that we can distinguish one multipath from the other based on the received power. In other words, the received power is an indicator of the distribution of multipath. In addition, the received power is relatively easy to measure directly in practical communication systems.

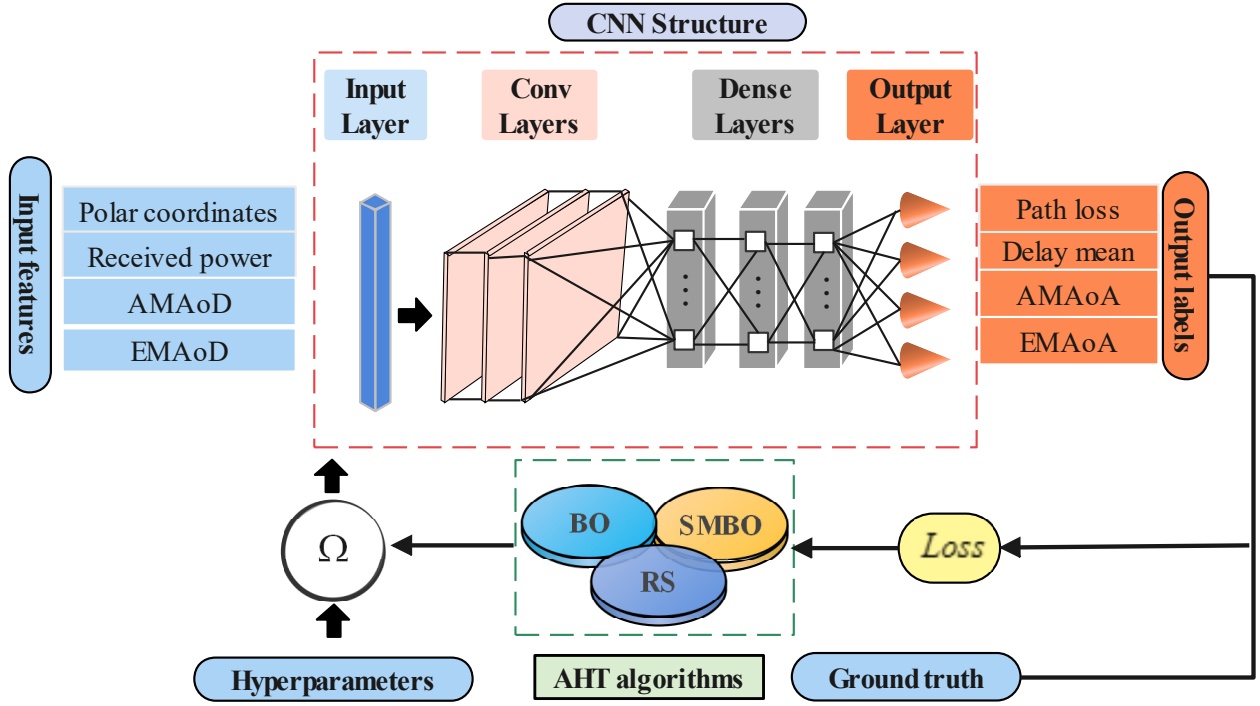


Fig. 2. Schematic layout of the AHT-assisted channel prediction CNN model. (Hyperparameters are not included in the figure for simplicity.)

Fig. 8 shows that the featured received power contributes to not only accurately predicting the path loss but also improving the adaptivity of predicting the AoAs and time delay. Thus, RP is a reasonable featured environment factor.

By the same token, elevation and azimuth angles of departure are chosen as environment factors. In this way, the proposed model can predict AoAs from known AoDs instead of jointly extracting AoAs and AoDs from the channel impulse response. For practical applications, the proposed model can be used to predict AoAs by certain predefined AoDs for mmWave communication systems equipped with directional antennas, which helps to arrange or adjust the direction of the receiving antennas in real time.

Inspired by the wide application of polar coordinates in antenna theory, we transform the Cartesian coordinates of TxS and RxS $\mathbf{C}_m = [x_m^T, y_m^T, x_m^R, y_m^R]$ to the polar coordinates $\mathbf{P}_m = [r_m^T, r_m^R, \delta_m^T, \delta_m^R]$ by the following equations,

$$\begin{cases} r_m^T = \sqrt{(x_m^T)^2 + (y_m^T)^2} \\ r_m^R = \sqrt{(x_m^R)^2 + (y_m^R)^2} \\ \delta_m^T = \arctan \frac{y_m^T}{x_m^T} \\ \delta_m^R = \arctan \frac{y_m^R}{x_m^R} \end{cases} \quad (7)$$

It is widely known that PL and DM depend on the distance d between Tx and Rx, which is given by

$$d = [\mathbf{C}_m \mathbf{C}_m^T - 2(x_m^T x_m^R - y_m^T y_m^R)]^{\frac{1}{2}}, \quad (8)$$

or

$$d = [(r_m^T)^2 + (r_m^R)^2 - 2r_m^T r_m^R \cos(\delta_m^T - \delta_m^R)]^{\frac{1}{2}}. \quad (9)$$

In (8), the distance is calculated by Cartesian coordinates. When $x_m^T x_m^R - y_m^T y_m^R = 0$, d relates to 4 variables. The

distance d calculated by polar coordinates in (9) depends on only 2 variables under the condition that $\delta_m^T - \delta_m^R = 0$. In other words, the labeled channel parameters are determined by fewer featured coordinate variables in situations where the mobile antennas move to abnormal positions, making the model less sensitive to coordinate noise.

B. The Motivation and Architecture of CNN

We propose a CNN rather than a regular fully-connected neural network (FCN) model for the following reasons. First, as discussed in Subsection III-A, the featured environment factors are coupled with the labeled channel parameters. With this in mind, the convolutional layers are leveraged to extract the correlation. In parallel, considering the characteristic of the feature dimension, CNN is a straightforward choice because of the ability to compress redundant channel data [11]. Specifically, the dimension of the training data is $M \times N$, where M is the total number of Tx-Rx pairs, which is 90000 for the park and square scenarios and is 60000 for the lab scenario as detailed in Subsection IV-B. N denotes the number of input features.

The schematic layout of the proposed model is presented in Fig. 2. The first convolutional layer filters the $1 \times N$ input vector with ω_2 kernels. The output of the first convolutional layer is fed to the second convolutional layer and filtered with ω_3 kernels. The third convolutional layer processes the output of the second convolutional layer with ω_4 kernels. The kernel size is $1 \times S$. The output of the third convolutional layer is then connected to three dense layers with $\omega_5, \omega_6, \omega_7$ neurons, respectively. To speed up the convergence rate, the learning rate ω_1 , the activation function F , and the batch size N_{bs} are adjusted during the training process.

TABLE II
HYPERPARAMETERS AND CORRESPONDING SEARCHING RANGE.

Hyperparameters	Notations	Searching range
Learning rate	ω_1	10^{-9} – 10^{-1}
Filter number in convolutional layers	$\omega_2 \sim \omega_4$	1–200
Neuron number in dense layers	$\omega_5 \sim \omega_7$	1–200
Activation function	F	relu, sigmoid, tanh
Batch size	N_{bs}	1000–10000
Filter size	S	1–9

IV. AHT-ASSISTED TRAINING PROCESS

In this section, we first highlight the procedure of AHT, followed by introducing the generation of training databases. Then, we describe in detail the training process of the proposed CNN.

A. AHT Algorithms

1) *Formulation of AHT*: As shown in Fig. 2, AHT algorithms are incorporated into the CNN training process to adjust hyperparameters to improve the training efficiency and obtain optimal solutions. To formulate the AHT-assisted training process, we first define a function Δ to represent the mapping between the hyperparameters and the validation loss. This function can be written as

$$Loss_{\Omega} = \Delta(\Omega), \quad (10)$$

where $\Omega = [\omega_1, \omega_2, \dots, \omega_7, F, N_{bs}, S]$ is the hyperparameter vector. The procedure of finding optimal hyperparameters Ω_{opt} can be viewed as an optimization problem defined as

$$\Omega_{opt} \triangleq \arg \min_{\Omega \in \chi} \Delta(\Omega), \quad (11)$$

where χ denotes the search space of hyperparameters. The search space χ in this paper is investigated according to the experience of hyperparameter tuning and is given in Table II. It is worth noting that the search range can be arbitrarily modified for different application scenarios. The optimization problem in (11) is non-convex [17] and is solved by AHT algorithms instead of the tedious hand-tuning method.

2) *Procedure of AHT*: There are two components to the AHT-assisted training algorithm. The first component is the probabilistic surrogate model, which includes a prior distribution and an observation model. The prior distribution describes our beliefs about the behavior of the target black-box function. The observation model captures the interaction between the hyperparameters and resulting validation loss [14]. The second one is called the acquisition function, which evaluates the candidate hyperparameters for the next observation.

The AHT procedure is integrated with the diversification strategy to overcome the vulnerability of AHT algorithms to different databases. One of the popular diversification strategies is called ensemble, which employs a set of learned models rather than apply a single model [24]. The proposed AHT procedure adopts diversification at model level as in the ensemble to produce optimal convergence under different scenarios.

Algorithm 1 AHT-assisted training algorithms

Require: black-box function Δ ,

parameter space χ ,

surrogate models $\mathcal{M} = (M_1, M_2, \dots, M_T)$,

acquisition function $\tilde{\Delta}$

Ensure: optimal hyperparameters combination

```

1: for  $t = 1, 2, \dots, T$  do
2:   for  $n = 1, 2, \dots, N$  do
3:     Integrate out true function by the surrogate model  $M_t$ 
4:     Given observations  $\mathcal{H}_t \leftarrow (\Omega_n, Loss_{\Omega_n})$ 
5:     Calculate posterior belief with  $\mathcal{H}_t$ 
6:     Select  $\Omega_{n+1}$  according to (12) from  $\chi$ 
7:     Evaluate  $Loss_{\Omega_{n+1}} \leftarrow \Delta(\Omega_{n+1})$ 
8:     Augment  $\mathcal{H}_t \leftarrow \mathcal{H}_t \cup (\Omega_{n+1}, Loss_{\Omega_{n+1}})$ 
9:     Update probabilistic surrogate model
10:    Check for early stopping criteria
11:   end for
12:   Record the optimal observation  $\mathcal{H}_t$  for the  $t$ th surrogate model
13: end for
14:  $\mathcal{H} \leftarrow \mathcal{H}_1 \cup \mathcal{H}_2, \dots, \cup \mathcal{H}_T$ 

```

Specifically, we combine three surrogate models into the AHT algorithms, i.e., Gaussian processes (GP), gradient-boosted regression trees (GBRT), and random forests (RF), to create three AHT algorithms, including GP-based Bayesian optimization (BO-GP), GBRT-based sequential model-based optimization (SMBO-GBRT) and RF based SMBO (SMBO-RF). Considering the trade-off between the training time and performance, we employ three different models in this paper. However, additional surrogate models can be conveniently added to our AHT algorithms for more complicated situations. Random search (RS) is employed as a baseline instead of grid search because the latter is very time-consuming.

The acquisition function is based on the expectation improvement (EI) strategy. Let $\mu_n(\Omega)$, $\sigma_n(\Omega)$, $\Phi_C(\cdot)$, and $\Phi_P(\cdot)$ be the posterior mean function, posterior variance function, cumulative distribution function (CDF) and probability density function (PDF), respectively. Denote $\tilde{\Delta}_n^*$ as the best observation. The acquisition function is written as

$$\tilde{\Delta}_n(\Omega) = \begin{cases} (\mu_n(\Omega) - \tilde{\Delta}_n^*) \Phi_C(\Gamma) & \sigma_n(\Omega) > 0 \\ + \sigma_n(\Omega) \Phi_P(\Gamma), & \sigma_n(\Omega) = 0 \\ 0, & \sigma_n(\Omega) = 0 \end{cases}, \quad (12)$$

TABLE III
DIELECTRIC CONSTANTS OF DIFFERENT MATERIALS AT 100 GHz.

Material	Dielectric constant
Concrete	6.3
Brick	6.5
Wood	3.3
Foliage	9.3
Glass	5.7
Water	81
Bookshelf	4.7
Electronic devices	4.0

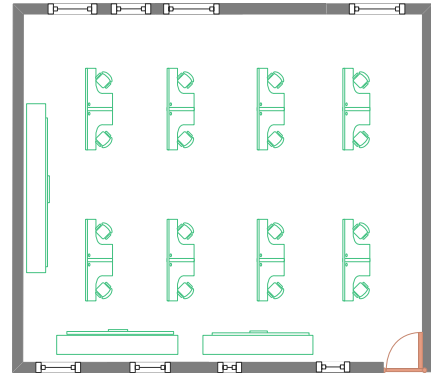
where

$$\Gamma = \frac{\mu_n(\Omega) - \tilde{\Delta}_n^*}{\sigma_n(\Omega)}. \quad (13)$$

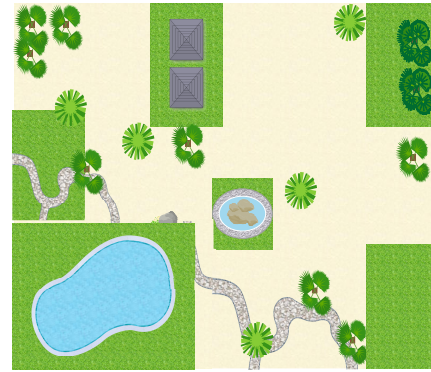
a) Pseudocode for AHT: Pseudocode for the AHT-assisted training algorithms is shown in Algorithm 1. During the training process, surrogate models are adopted in turn thanks to the sequential nature of the proposed AHT algorithms. The prior belief and the current observation \mathcal{H}_t are prescribed by the probabilistic surrogate model. The observation set \mathcal{H}_t consists of hyperparameters and the corresponding validation loss. For each surrogate model, the AHT algorithms incorporate prior belief to estimate $\tilde{\Delta}(\Omega)$. The prior distribution is updated by observing the performance of the model under the current hyperparameters to obtain the posterior distribution. The posterior distribution is then substituted into the EI-based acquisition function in (12) to determine the next hyperparameters Ω_{n+1} . Then evaluate $\Delta(\Omega_{n+1})$ with the real black-box function Δ , followed by updating the set of observations \mathcal{H}_t . The probabilistic surrogate model is then updated with the posterior distribution and the set of observations and iterates sequentially until the early stopping criterion is met. In this paper, the stopping criterion is to end the iteration when no performance improvement is achieved for 10 consecutive calls. The observations \mathcal{H}_t obtained by all T models are recorded in \mathcal{H} . The validation loss in each \mathcal{H}_t is compared, and the final output of the AHT algorithm is the hyperparameter combination corresponding to the minimum validation loss.

B. Training Data Generation

To generate the training data, we build an outdoor scenario (a virtual park) and an indoor scenario (a virtual lab) in the accurate ray-tracing simulator. The layouts of the park and the lab are shown in Fig. 3. The park is $400 \times 400 m^2$ and consists of a lake, a fountain, trees, grass, pavements, and constructions. The lab is furnished with tables, cabinets, and office furniture (bookshelf, electronic devices), and occupies an area of $30 \times 30 m^2$. In the outdoor scenarios (the park and the square), the constructions are made of concrete, and the pavement is constructed using bricks. For the indoor scenario (the lab), wooden tables, chairs, and cabinets are used, while the ceiling, floor, and walls are made of concrete. The dielectric constants of the materials used in the study, including concrete, brick, wood, foliage, glass, office furniture (bookshelves and electronic devices), and water, are given in Table III [25]–[27]. The simulation frequency and bandwidth



(a) The lab scenario



(b) The park scenario

Fig. 3. Layouts of the park and lab.

are set to 100 GHz and 2.5 GHz, respectively. A maximum of two orders of reflection and one order of diffraction are simulated due to the significant attenuation of mmWave propagation. Scattering due to surface roughness is neglected for the sake of the simplicity of simulation. The channel characteristics involved are calculated by

$$\begin{cases} \bar{\tau}_m = \frac{\sum_{l=1}^L \tau_{m,l}}{L} \\ \bar{\varphi}_m = \frac{\sum_{l=1}^L \varphi_{m,l}}{L} \\ \bar{\vartheta}_m = \frac{\sum_{l=1}^L \vartheta_{m,l}}{L} \\ \bar{\phi}_m = \frac{\sum_{l=1}^L \phi_{m,l}}{L} \\ \bar{\theta}_m = \frac{\sum_{l=1}^L \theta_{m,l}}{L} \end{cases}, \quad (14)$$

where l is the index of the propagation path and L denotes the total number of paths. We set the reference point at the lower left corner of each simulated communication area to obtain the 2-D relative coordinates of both Tx and Rx. The coordinates are assumed to be calculated by highly accurate positioning techniques (e.g., via [28]).

We randomly locate 300 Tx and 300 Rx within the park to create 90000 Tx-Rx pairs and collect 90000 input features $[\mathbf{C}_m, \bar{\varphi}_m, \bar{\vartheta}_m, p_m]$ and output labels $[\rho_m, \bar{\phi}_m, \bar{\theta}_m, \bar{\tau}_m]$. Due to the inherent sparsity of mmWave propagation, a maximum of 5 paths are considered. Each Tx (Rx) is equipped with a single half-wave dual-polarization antenna.

TABLE IV
DATABASES AND CORRESPONDING SCENARIOS.

Databases	Scenarios	Coordinates type	Remarks
Database i	The park	Cartesian	Original scenario
Database ii	The park	Polar	Original scenario
Database iii	The park	Cartesian	Add coordinates noise
Database iv	The park	Polar	Add coordinates noise
Database v	The square	Polar	Generalization validation
Database vi	The lab	Polar	Generalization validation

The scenarios and the resulting databases are described in Table IV. The original scenario is the virtual park, where database i is collected, and then database ii is generated by transforming the Cartesian coordinates in database i to polar coordinates. We add coordinates noise equally to database i and database ii to generate database iii and database iv, respectively, while keeping the actual antenna positions unchanged to validate the robustness of the model to coordinate noise. A virtual square is then built as a similar environment for the park to generate database v. The only difference between the park and the square is the distribution of scatterers. The virtual indoor environment (the lab) is constructed as a different scenario from the park to obtain database vi. 300 Tx's and 300 Rx's are randomly placed in the lab scenario to create 90000 Tx-Rx pairs. 60000 sets of valid featured and labeled channel data are collected. Both database v and database vi are built to validate the generalization ability of the model.

When features and labels are highly correlated, the model can more easily learn the correlation between them, resulting in better prediction performance with a smaller sample of data [29]. To verify that the proposed feature selection strategy can reduce the amount of training data, the percentage of the training set varies from 10 percent to 70 percent at intervals of 10 during the training stage. The test and validation proportions are set equal.

C. Training Process

The Adam optimizer is applied to the training of the network with a learning rate of ω_1 . The maximum iteration epoch is set to 3000, and early stopping with patience 100 is adopted to avoid overfitting. Also, the learning rate ω_1 decays by a factor of 0.5 whenever the validation loss no longer diminishes for a consecutive 100 epochs.

The CNN is trained to minimize the loss function formulated as

$$Loss = \frac{1}{4N_{BS}} \sum_{i=1}^4 \sum_{j=1}^{N_{BS}} (\hat{p}_{ij} - p_{ij})^2, \quad (15)$$

where N_{bs} is the batch size, \hat{p}_{ij} and p_{ij} are the j th predicted and true values of the i th channel parameter (one of $\rho, \bar{\phi}, \bar{\theta}$, and $\bar{\tau}$), respectively. During the network training process, the hyperparameters listed in Table II should be tuned repeatedly until the validation loss is satisfactory.

V. NUMERICAL RESULTS

We train three models with AHT algorithms to make comparisons, i.e., the enhanced CNN, the standard CNN, and a six-hidden-layer FCN without convolutional layers. The standard CNN is fed with Cartesian coordinates of Tx and Rx, which is the same as the model proposed in [11], while the enhanced one is the model presented in this paper, which is enhanced by the coordinate transformation and feature selection strategy. The input features and output labels of the FCN are the same as those of the enhanced CNN. We utilize a computer with an Intel Core i7-9700k central processing unit and an NVIDIA RTX 2070 graphics processing unit to train the network. The training time for each surrogate model amount to approximately 2 hours, while the average training time for a full model consisting of 3 surrogate models is less than 6 hours.

A. AHT Procedure Analysis

We train the enhanced model by database ii and database vi respectively, with the assistance of AHT algorithms and plot the convergence traces of:

- GP as the surrogate model for BO (BO-GP),
- GBRT as the surrogate model for SMBO (SMBO-GBRT),
- RF as the surrogate model for SMBO (SMBO-RF), and
- RS.

The plots given as Figs. 4 and 5 show the value of the determination coefficient R^2 (y-axis) between the predicted and true channel parameters as a function of the number of iterations (x-axis). R^2 is an indicator of the prediction accuracy calculated by

$$R^2 = 1 - \frac{\sum_i (\hat{p}_{ij} - p_{ij})^2}{\sum_i (p_{ij} - \bar{p}_i)^2}. \quad (16)$$

R^2 varies in the range of [0,1]. In (16), \hat{p}_{ij} and p_{ij} are the j th predicted and true values of the i th channel parameter (one of $\rho, \bar{\phi}, \bar{\theta}$, and $\bar{\tau}$), respectively, and \bar{p}_i denotes the average value of the i th output channel parameter, calculated by $\bar{p}_i = \sum_{v=1}^V \bar{p}_{iv} / V$, with V denoting the total number of the i th labeled channel parameters p_i in the validation set. The resulting R^2 scores of all predicted channel parameters are averaged with uniform weight to obtain the final R^2 .

The convergence rates and optimization performances of each tuning algorithm used vary for different databases, as shown in Figs. 4 and 5. This validates the necessity of the diversification strategy, which helps to determine the optimal

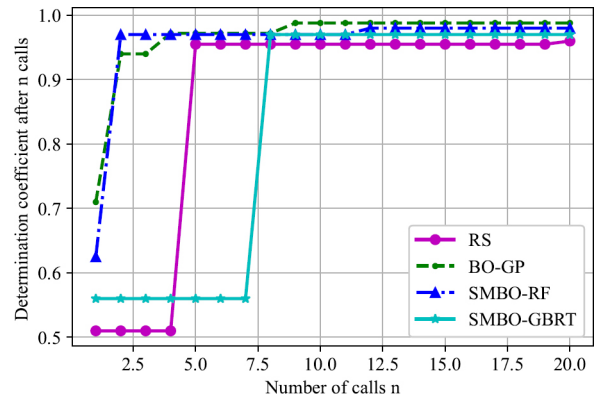
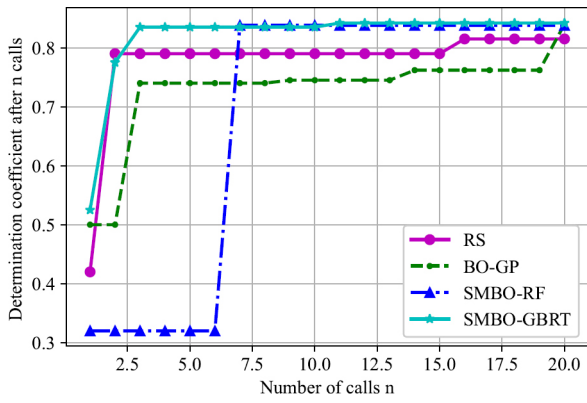
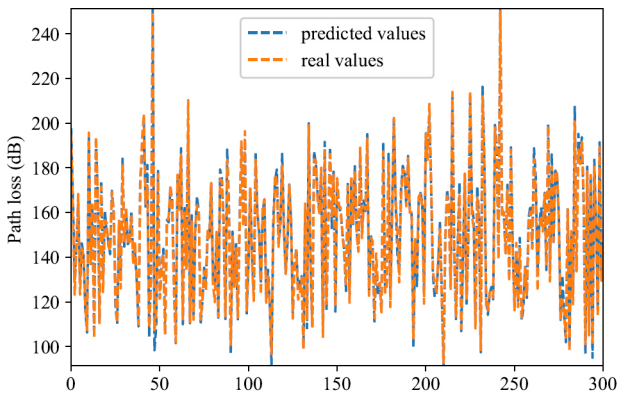
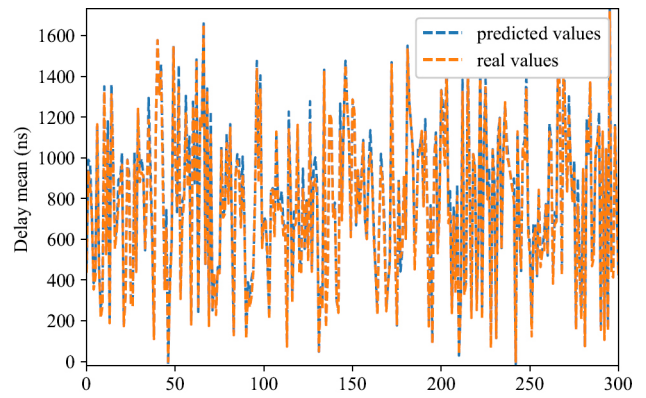


Fig. 4. Convergence plot of AHT algorithms for the indoor scenario. (One iteration is defined as one call.)

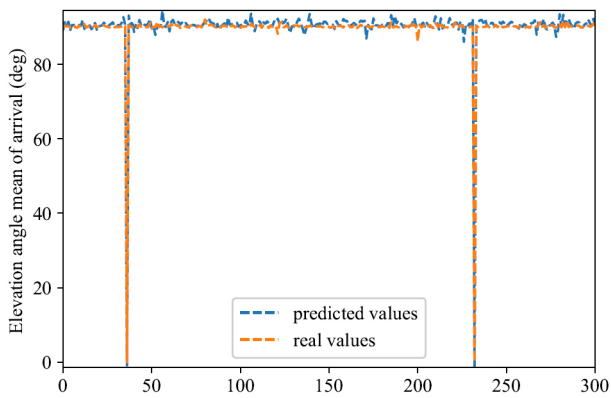
Fig. 5. Convergence plot of AHT algorithms for the outdoor scenario. (One iteration is defined as one call.)



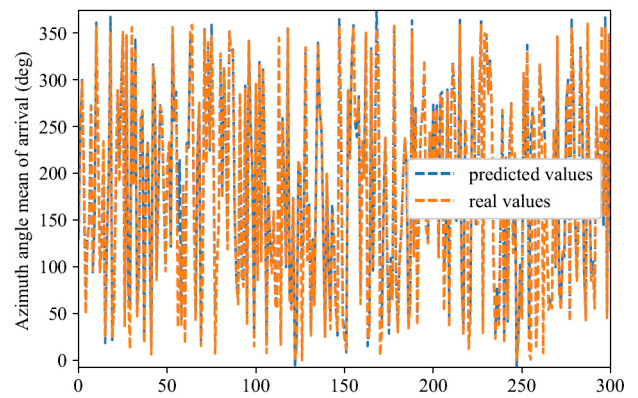
(a) PL



(b) DM



(c) EMAoA



(d) AMAoA

Fig. 6. Fitting plots for predicted channel parameters. The x-axis represents the number of samples.

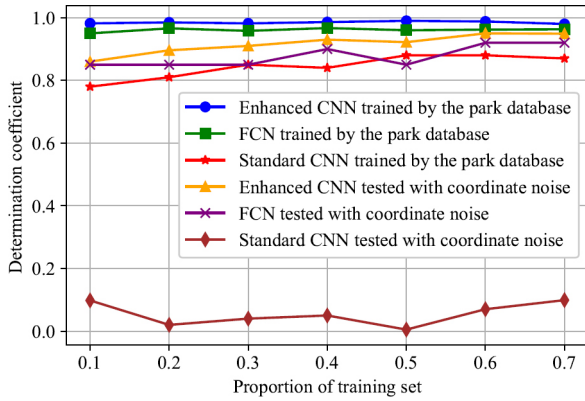


Fig. 7. Robustness analysis.

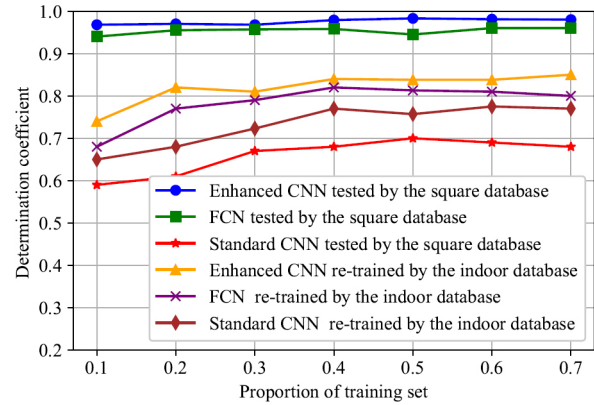


Fig. 8. Validation of generalization ability.

TABLE V
OPTIMAL HYPERPARAMETERS OBTAINED BY AHT ALGORITHMS FOR THE PARK SCENARIO.

Hyperparameters	Notations	Solutions
Learning rate	ω_1	0.0021
Filter numbers	$\omega_2 \sim \omega_4$	187,106,73
Neuron numbers	$\omega_5 \sim \omega_7$	97,76,111
Activation function	F	relu
Batch size	N_{bs}	3526
filter size	S	8

AHT algorithm for different communication scenarios. Furthermore, due to its fast convergence rate and satisfactory accuracy, the proposed AHT procedure is suitable for embedding in the training process of the channel prediction CNN. This contributes greatly to saving training effort and improving prediction performance.

Table V presents the optimal hyperparameters obtained by the AHT algorithms for the park scenario. The resulting CNN model is used to plot the fitting curves between the predicted and real channel parameters, as shown in Fig. 6. From this plot, it is evident that the predicted values fit the real ones well.

B. Impact of the Size of the Training Set and Robustness Analysis

The standard CNN model is trained by the database i, whereas both the enhanced CNN and FCN are trained using the database ii. Afterward, the trained models are tested on database iii (the standard CNN) and database iv (the enhanced CNN and FCN), respectively, to validate the robustness of the model against coordinate noise. The figure presented in Fig. 7 illustrates the relationship between the coefficient of determination and the size of the training data. The results indicate that the enhanced AHT-assisted CNN is capable of achieving a decent R^2 value (higher than 0.8) with less training data compared to the standard CNN. This characteristic of the enhanced model makes it more robust to data size, which

is highly advantageous for practical applications where data collection can be challenging.

Additionally, both the enhanced model and the FCN exhibit significantly higher accuracy. This improved performance is attributed to the feature selection and preprocessing procedure employed in the enhanced model. Furthermore, the enhanced CNN exhibits robustness to coordinate noise, while the standard CNN is completely unreliable under such conditions. This robustness allows the proposed enhanced model to detect changes in communication scenarios when coordinate fluctuations occur.

C. Generalization Ability of the Proposed Model

In this subsection, we evaluate the generalization ability of the enhanced and standard models by testing the trained models on database v (the square scenario) and database vi (the lab scenario), respectively. Numerical results are given by Fig. 8. The enhanced CNN shows adaptability to communication scenarios, indicating that the proposed model can accurately predict channel parameters when the scenario dynamically changes to a similar one. Consequently, our proposed model can be deployed in similar communication scenarios without the need for re-training.

As for the indoor scenario, all models require offline re-training. Nevertheless, the enhanced AHT-assisted CNN offers two significant advantages. Firstly, the AHT algorithms aid in the efficient and convenient re-training of the enhanced model by eliminating the need for hand-tuning hyperparameters. Secondly, as shown in Fig. 8, the enhanced model remains more accurate than the standard model when generalized to completely different mmWave communication scenarios, making it more reliable. In conclusion, the proposed model enhanced by the feature selection strategy produces higher accuracy and stronger generalization ability than the standard model. What's more, we can see from Figs. 7 and 8 that the enhanced CNN outperforms the FCN in terms of both accuracy and generality as well, suggesting that CNN is a better choice.

VI. CONCLUSION

In this paper, we propose a generalized AHT-assisted CNN model for mmWave channel prediction under different communication situations. We integrate the AHT procedure into the training process to simultaneously save training effort and improve predictive performance. The diversification strategy is applied to enhancing the stability of the hyperparameter optimization process under different databases. Moreover, we use a novel feature selection strategy to make the model generalizable to different communication scenarios, which contributes to avoiding repetitive channel sounding. In parallel, coordinate transformation is utilized to reduce the sensitivity of the model to coordinate noise. Numerical results show that the enhanced model achieves desirable performance, including being convenient to train, less training data required, robust to coordinate noise, and strong generalization ability. Channel sounding in mmWave communication is of great importance for channel modeling but it is difficult in practical implementations. In this context, our proposed model can serve as an alternative approach to channel modeling for different communication systems with limited channel sounding overhead.

REFERENCES

- [1] H. Kim and J. Choi, "Channel AoA estimation for massive MIMO systems using one-bit ADCs," *J. Commun. Netw.*, vol. 20, no. 4, pp. 374–382, 2018.
- [2] D. Zhu, J. Choi, and R. W. Heath, "Two-dimensional AoD and AoA acquisition for wideband millimeter-wave systems with dual-polarized MIMO," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 7890–7905, 2017.
- [3] S. Moon, H. Kim, and I. Hwang, "Deep learning-based channel estimation and tracking for millimeter-wave vehicular communications," *J. Commun. Netw.*, vol. 22, no. 3, pp. 177–184, 2020.
- [4] A. Misra, M. P. Sarma, K. K. Sarma, and N. Mastorakis, "Temporal deep learning assisted UAV communication channel model for application in EH-MIMO-NOMA set-up," *J. Commun. Netw.*, vol. 24, no. 2, pp. 1–18, 2022.
- [5] X. Zhao *et al.*, "Playback of 5G and beyond measured MIMO channels by an ANN-based modeling and simulation framework," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 1945–1954, 2020.
- [6] H. Huang, S. Guo, G. Gui, Z. Yang, J. Zhang, H. Sari, and F. Adachi, "Deep learning for physical-layer 5G wireless techniques: Opportunities, challenges and solutions," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 214–222, 2019.
- [7] X. Zhao *et al.*, "Neural network and GBSM based time-varying and stochastic channel modeling for 5G millimeter wave communications," *China Commun.*, vol. 16, no. 6, pp. 80–90, 2019.
- [8] J. A. Cal-Braz, L. J. Matos, and E. Cataldo, "The relevance vector machine applied to the modeling of wireless channels," *IEEE Trans. Antennas Propag.*, vol. 61, no. 12, pp. 6157–6167, 2013.
- [9] G. P. Ferreira, L. J. Matos, and J. M. Silva, "Improvement of outdoor signal strength prediction in UHF band by artificial neural network," *IEEE Trans. Antennas Propag.*, vol. 64, no. 12, pp. 5404–5410, 2016.
- [10] Z.-M. Liu, Z.-T. Huang, and Y.-Y. Zhou, "An efficient maximum likelihood method for direction-of-arrival estimation via sparse Bayesian learning," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 1–11, 2012.
- [11] L. Bai *et al.*, "Predicting wireless mmWave massive MIMO channel characteristics using machine learning algorithms," *Wireless Commun. Mobile Comput.*, vol. 2018, no. 08, pp. 1–11, 2018.
- [12] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyperparameter optimization," in *Proc. NeurIPS*, 2011.
- [13] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," *Advances Neural Inf. Processing Syst.*, vol. 25, no. 01, pp. 1–12, 2012.
- [14] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas, "Taking the human out of the loop: A review of Bayesian optimization," *Proc. IEEE*, vol. 104, no. 1, pp. 148–175, 2015.
- [15] L. Kouhalvandi and L. Matekovits, "Hyperparameter optimization of long short-term memory-based forecasting DNN for antenna modeling through stochastic methods," *IEEE Antennas Wireless Propag. Lett.*, vol. 21, no. 4, pp. 725–729, 2022.
- [16] A. Xie, F. Yin, Y. Xu, B. Ai, T. Chen, and S. Cui, "Distributed Gaussian processes hyperparameter optimization for big data using proximal ADMM," *IEEE Signal Process. Lett.*, vol. 26, no. 8, pp. 1197–1201, 2019.
- [17] H. Cho, Y. Kim, E. Lee, D. Choi, Y. Lee, and W. Rhee, "Basic enhancement strategies when using Bayesian optimization for hyperparameter tuning of deep neural networks," *IEEE Access*, vol. 8, pp. 52 588–52 608, 2020.
- [18] C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, "Indoor signal focusing with deep learning designed reconfigurable intelligent surfaces," in *Proc. IEEE SPAWC*, 2019.
- [19] Remcom, "Wireless Insite." [Online]. Available: <http://www.atm.com>
- [20] Y. Yang, F. Gao, G. Y. Li, and M. Jian, "Deep learning-based downlink channel prediction for FDD massive MIMO system," *IEEE Commun. Lett.*, vol. 23, no. 11, pp. 1994–1998, 2019.
- [21] F. Gomez-Cuba and A. J. Goldsmith, "Sparse mmWave OFDM channel estimation using compressed sensing," in *Proc. IEEE ICC*, 2019.
- [22] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 852–855, 2018.
- [23] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, 2016.
- [24] T. G. Dietterich, "Ensemble methods in machine learning," in *Proc. MCS*, 2000.
- [25] D. Ferreira, I. Cuiñas, R. F. Caldeirinha, and T. R. Fernandes, "A review on the electromagnetic characterisation of building materials at micro- and millimetre wave frequencies," in *Proc. EuCAP*, 2014.
- [26] K. Korolev and M. Afsar, "Complex dielectric permittivity measurements of materials in millimeter waves," in *Proc. IEEE IRMMW-THz*, 2005.
- [27] P. Series, "Effects of building materials and structures on radio-wave propagation above about 100 MHz," *Recommendation ITU-R*, pp. 2040–1, 2015.
- [28] W. Jiang, Z. Cao, B. Cai, B. Li, and J. Wang, "Indoor and outdoor seamless positioning method using UWB enhanced multi-sensor tightly-coupled integration," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 10633–10645, 2021.
- [29] Y. Zhu, J. T. Kwok, and Z.-H. Zhou, "Multi-label learning with global and local label correlation," *IEEE Trans. Knowl Data Eng.*, vol. 30, no. 6, pp. 1081–1094, 2018.



Chengfang Yue received the B.E. degree in Electrical and Electronic Engineering from University of Sheffield, UK, in 2015. He is currently working toward the Ph.D. degree in Control Science and Engineering at Wuhan University of Science and Technology. His current research interests include wireless communication technology and machine learning.



Hui Tang received the B.E. degree in Electrical and Electronic Engineering and the Ph.D. degree in Radio Physics from Wuhan University, China, in 2011 and 2016, respectively. She was a Visiting Ph.D. student at Arizona State University, USA, from September 2014 to August 2016. She is currently an Associate Professor of Wuhan University of Science and Technology. Her research interests include machine learning, graph neural networks and radar signal processing.



Jun Yang received the B.E. degree in Computer Science and Technology and the Ph.D degree in Control Science and Engineering from Wuhan University of Science and Technology, China, in 2003 and 2016, respectively. She is currently an Associate Professor of Wuhan University of Science and Technology. Her research interests include machine learning, channel modeling and signal processing.



Li Chai (S'00-M'03) received the BS degree in Applied Mathematics and the MS degree in Control Science and Engineering from Zhejiang University, China, in 1994 and 1997, respectively, and the PhD degree in Electrical Engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2002. From August 2002 to December 2007, he was at Hangzhou Dianzi University, China. He worked as a Professor at Wuhan University of Science and Technology, China from 2008 to 2022. In August 2022, he joined Zhejiang University,

China, where he is currently a Professor at the College of Control Science and Engineering. He has been a Postdoctoral Researcher or Visiting Scholar at Monash University, Newcastle University, Australia and Harvard University, USA.

His research interests include distributed optimization, filter banks, graph signal processing, and networked control systems. Professor Chai is the recipient of the Distinguished Young Scholar of the National Science Foundation of China. He has published over 100 fully refereed papers in prestigious journals and leading conferences. He serves as the associate editor of *IEEE Transactions on Circuit and Systems II: Express Briefs*, *Control and Decision* and *Journal of Image and Graphs*.