

Channel State Information Prediction for 5G Wireless Communications: A Deep Learning Approach

Changqing Luo, Jinlong Ji, Qianlong Wang, Xuhui Chen, and Pan Li

Abstract—Channel state information (CSI) estimation is one of the most fundamental problems in wireless communication systems. Various methods, so far, have been developed to conduct CSI estimation, which usually requires high computational complexity. However, these methods are not suitable for 5G wireless communications due to many techniques (e.g., massive MIMO, OFDM, and millimeter-Wave (mmWave)) to be employed in 5G wireless communication systems. In this paper, we propose an efficient online CSI prediction scheme, called **OCEAN**, for predicting CSI from historical data in 5G wireless communication systems. Specifically, we first identify several important features affecting the CSI of a radio link and a data sample consists of the information of these features and the CSI. We then design a learning framework that is a combination of a CNN (convolutional neural network) and an LSTM (long short term with memory) network. We further develop an offline-online two-step training mechanism, enabling the prediction results to be more stable when applied in practical 5G wireless communication systems. To validate OCEAN's efficacy, we conduct extensive experiments by considering four typical case studies, i.e., two outdoor and two indoor scenarios. The experiment results show that **OCEAN** not only obtains the predicted CSI values very quickly but also achieves highly accurate CSI prediction with up to 2.650% - 3.457% average difference ratio (ADR) between the predicted and measured CSI.

Index Terms—Channel state estimation, 5G wireless communications, deep learning

1 INTRODUCTION

Channel state information (CSI) is one of the most fundamental concepts in wireless communications [1], [2]. It refers to known channel properties of a radio link. Specifically, CSI can characterize the combined effect of path loss, scattering, diffraction, fading, shadowing, etc., when a signal propagates from a transmitter to its corresponding receiver. As a result, this information can tell us whether a radio link is in a good or bad state.

Obtaining accurate CSI is of paramount importance in guaranteeing the performance of radio links in wireless communication systems. To be more specific, CSI highly determines the physical-layer parameters and schemes deployed for a radio communication link in a wireless communication system. For example, when CSI is low, the physical layer needs to employ a low-order modulation scheme for combating the poor channel to reduce bit error rate, and, contrariwise, the physical layer needs to deploy a high-order modulation scheme for high data transmission rate when CSI is high. We can see that improper deployment of the modulation scheme can incur high data error rate or low data transmission rate. Furthermore, it is well-known that CSI also has a significant impact on radio resources allocation [3] and interference management [4] in wireless communication systems. Thus, it is very important to find accurate CSI in wireless communication systems.

Wireless communication systems can obtain accurate

CSI by performing channel estimation. So far, researchers have proposed many channel estimation methods, such as maximum-likelihood (ML) estimation [5], least-square (LS) estimation [6], minimum-mean-square-error (MMSE) estimation [7]. For example, Du *et al.* [5] consider performing ML-based channel estimation to predict the CSI of macro-cellular OFDM uplinks in a time-varying wireless environment. Karami [6] considers employing an LS algorithm to perform MIMO channel estimation and shows the tracking performance. Ma *et al.* [7] propose a linear MMSE method to estimate the CSI of individual channels, in which using an LS channel estimator obtains the initialization point for the iterative linear MMSE. We notice that these algorithms need to perform matrix operations like matrix multiplication, matrix inverse, eigenvalue decomposition, and singular value decomposition. Performing such matrix operations takes high computational complexity. Therefore, such traditional methods are only suitable for small-scale CSI estimation in wireless communication systems.

On the other hand, past few years have witnessed that cellular communication systems are evolving to 5G wireless communication systems due to the exploding growth of mobile devices and mobile traffic [8]–[10]. According to Ericsson mobility report 2017 [11], global total mobile data traffic was already up to 8.8 ExaBytes in 2016, and this figure grew by 70% in 2017. A 5G wireless communication system proposes multiple techniques to boost its system capacity, such as massive MIMO (multiple-input-multiple-output), OFDM (orthogonal frequency-division multiplexing), and mmWave (millimeter-Wave) communications. Massive MIMO exploits spatial-domain resources for offering diversity gain, multi-

This work was partially supported by the U.S. National Science Foundation under grants CNS-1602172 and CNS-1566479.

C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li are with Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH 44106. Email: {cxl881, jxj405, qxw204, xxc296, lipan}@case.edu.

plexing gain, and power gain [12]. OFDM can use multiple sub-carriers to support high-speed data transmission. mmWave communication enables 5G wireless communication systems to utilize the spectrum ranging from 3 GHz to 300 GHz [13]. Making use of these techniques can improve the system capacity significantly, thus leading to accommodating the increased mobile traffic.

Due to the newly-developed techniques, conventional CSI estimation methods face a few challenges in estimating CSI in 5G wireless communication systems. Due to using massive MIMO, OFDM, and mmWave communications, 5G wireless communication systems have much more channels than current cellular communication systems, which makes the channel estimation problem much more complicated. Furthermore, using conventional CSI estimation methods needs to perform large-scale matrix operations with high computational complexity, which incurs a formidable challenge for mobile equipment with insufficient computational resources. As a result, obtaining the CSI needs to take a very long time. It is impractical for a practical 5G communication system because it must refresh the CSI at a rate depending on the correlation time of the channel. Therefore, we need to find a very efficient way of performing CSI estimation for 5G wireless communications.

In this paper, we propose an efficient online CSI prediction scheme, called OCEAN, for predicting CSI from historical data in 5G wireless communication systems. To be more specific, we first identify several very critical features (i.e., frequency band, location, time, temperature, humidity, and weather) affecting the CSI of a radio link and a data sample consists of the information of these features and the CSI. Then, we consider the spatial-temporal relationship of the CSI and design a learning framework that is a combination of a CNN (convolutional neural network) and an LSTM (long short term with memory) network. We further develop an offline-online two-step training mechanism, enabling the prediction results to be more stable when applied in practical 5G wireless communication systems. To validate OCEAN's efficacy, we conduct extensive experiments to test OCEAN's performance by considering four typical case studies, i.e., two outdoor environments (i.e., a parking Lot and outside a building) and two indoor environments (i.e., a workroom and a building).

We summarize our major contributions as follows:

- 1) We propose an online CSI prediction scheme, called OCEAN, for 5G wireless communications.
- 2) We design a learning framework, i.e., a combination of a CNN and an LSTM network, for exploiting the spatial-temporal relationship of the CSI.
- 3) We identify several very critical features (i.e., frequency band, location, time, temperature, humidity, and weather) affecting the CSI of a radio link and a data sample consists of the information of these features and the CSI.
- 4) We design an offline-online two-step training mechanism to improve the stability of the developed learning framework when applied to a practical 5G wireless communication system.
- 5) We conduct extensive experiments to evaluate OCEAN's performance by considering four typical case studies. In

particular, we collect the data and apply the developed learning framework to learn from the collected data. The experiment results show the excellent performance in CSI prediction achieved by our proposed scheme.

The rest of this paper is organized as follows. We first briefly introduce the preliminary knowledge about the radio signal propagation in Section 2. Then, we describe the proposed CSI prediction scheme in detail, including the learning framework, the data, and the developed learning framework in Section 3. Afterwards, we present the results of the experiments thoroughly in Section 4. Subsequently, we present the most related works in Section 5, and finally conclude this paper in Section 6.

2 CSI ESTIMATION

In this section, we briefly present the CSI estimation in wireless communication systems.

In this section, we briefly present the CSI estimation in wireless communication systems. In wireless communication systems, radio signals are transmitted over transmission media, i.e., the air. In practice, electromagnetic wave propagates over the air without any protection, and thus a radio signal is prone to suffering from the attenuation of transmission power. The CSI is basically used to represent the attenuation of a radio signal over the air, i.e., the combined effect of path loss, scattering, fading, etc. It is well-known that radio channels suffer from path loss inevitably due to power dissipation radiated by the transmitter.

In free space, there is no obstacle between a transmitter and a receiver, and thus a transmitter can perform line-of-sight (LOS) signal transmissions to transmit radio signals to a receiver that is d meters away. Based on [14], we can have an approximation expression of the path loss as follows:

$$Pl = 10 \log_{10} \frac{P_t}{P_r} = -10 \log_{10} G_l \left(\frac{\lambda}{4\pi d} \right)^2, \quad (1)$$

where Pl is the path loss in decibel (dB), P_r is the receiving power, G_l is the product of the transmitter and receiver antenna field radiation patterns of LOS transmissions, and λ is the operating wavelength that can be calculated by $\lambda = 3 \times 10^8 / f$ (here, f is the frequency). Hence, we can see that path loss depends on the distance between a transmitter and its receiver for transmitting a radio signal over a specific channel through a specific antenna.

Moreover, scattering and fading also have an impact on the CSI. Scattering is natural in wireless communication systems due to the radiation of electromagnetic wave transmission in a medium and obstacles. When air density and humidity are different, the impact of the scattering on the CSI is also different. On the other hand, the fading of a radio communication link may be due to either multi-path propagation or shadowing from obstacles. Due to the changing rate of electromagnetic waves' magnitude and phase, we can usually have slow fading and fast fading. Besides, fading is also related to the frequency band. Signals transmitted over some frequency bands may suffer from severe performance degradation, which refers to frequency selective fading. Moreover, some frequency bands

have some other features. For example, as reported by [15], the atmosphere absorbs the power of electromagnetic waves, and absorption peaks occur at 24 GHz and 60 GHz when radio signals travel through the atmosphere, and there is stable loss window between two peaks.

Consider a flat block-fading MIMO system with t transmit and r receive antennas. To estimate the CSI, we need to send pilot signals and get feedback. Based on the knowledge of the transmitted and received signals, we then conduct CSI estimation. Let p_1, \dots, p_N denote pilot sequences sent to the MIMO system and p_i is the $t \times 1$ complex vector of the transmitted pilot signals. At the receiver, we can have received signal vector as follows:

$$y_i = H p_i + n_i, \quad (2)$$

where H is the $r \times t$ complex channel matrix and n_i is the $r \times 1$ complex vector representing the zero-mean white noise. Note that any estimator of H is to find an estimation of this channel matrix. In order to conduct channel estimation of finding H , we need to transmit $N \leq t$ training pilot signal vectors p_1, \dots, p_N and the received signals can be given by

$$Y = HP + N. \quad (3)$$

The task of a channel estimation algorithm is to find the channel matrix H based on the knowledge of Y and P . To recover H , researchers have developed several methods like ML, MMSE, and LS.

3 ONLINE CSI PREDICTION SCHEME FOR 5G WIRELESS COMMUNICATIONS

In this section, we describe online CSI prediction framework, data sample, learning framework, and offline-online two-step training mechanism, respectively.

3.1 Online CSI Prediction Framework

Based on the introduction in Section 2, we notice that the CSI of a radio communication link is related to path loss, specific frequency band characteristics, weather, atmosphere, and obstacles, etc. Thus, two questions raise:

- 1) Are there any patterns between such factors and the CSI?
- 2) Is it possible for us to employ certain machine learning techniques to extract the relationships from the dataset constructed by such factors and the CSI and perform CSI prediction?

After the careful analysis of all related factors, we note that the CSI has some certain patterns. Specifically, a specific frequency band naturally has particular electromagnetic propagation characteristics that do not change along time and locations, which implies that different frequency bands have different CSI even at the same place and at the same time. As a result, the CSI has some certain patterns in the frequency domain. Moreover, the atmosphere factors like water vapor, molecules of oxygen, and other gaseous atmospheric components are usually related to the time domain. For example, the measured values of such atmosphere factors in the daytime

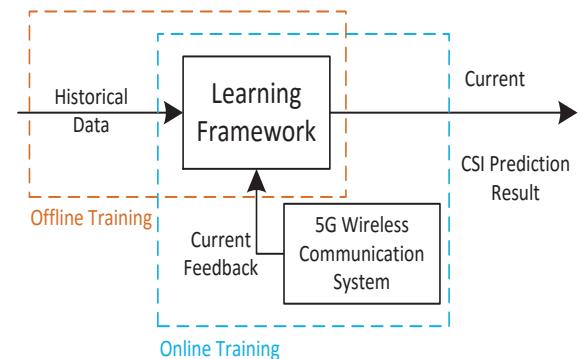


Fig. 1. The framework for online CSI prediction in 5G wireless communication systems.

are different from that during the night. Thus, the CSI tends to have certain patterns in the time domain. Furthermore, some obstacles like buildings and trees are usually static, while some other obstacles like walking people and cars are dynamic. For example, a crowd of students walking during the daytime of weekdays on campus and only a few students there during the evening and weekends. Thus, such dynamic obstacles tend to have some specific patterns as well. As a result, the CSI has some specific patterns in the space domain. Besides, different rain degrees can have a different effect on electromagnetic wave propagation, which is helpful for us to predict CSI for 5G wireless communications. To sum up, these factors related to the CSI have some particular patterns that we can extract. We can take advantage of such patterns to perform CSI prediction for a mobile device at a place during a time in 5G wireless communication systems.

To efficiently find the CSI for a mobile device at a place during a period, we propose an Online CSI prEdiction scheme for 5G wireless communicAtioNs, called OCEAN. The detailed framework of OCEAN is shown in Fig. 1. Specifically, OCEAN employs a learning framework for accurate CSI prediction. Particularly, due to certain patterns existing in frequency, time, space domains, we develop a learning framework, a combination of a CNN network and an LSTM network, to explore the spatiotemporal relationship of the CSI. To improve the stability of the developed learning framework, we design an offline-online two-step training mechanism. The offline training step is to analyze the historical data, and the online training step is to integrate the predicted and measured CSI. When a CSI prediction request arrives, OCEAN first collects real-time data of all factors (Note: such data are called as side information in the rest of this paper). It then preprocesses the side information and inputs the information into the learning framework to perform CSI prediction. In the following, we describe the dataset construction, the learning framework, and the offline-online two-step training mechanism, respectively.

3.2 Data Sample

In this paper, we consider several critical factors affect the CSI, including frequency band, location, time, temperature,

humidity, and weather. In the following, we present all these features.

Frequency band: 5G wireless communications utilize the spectrum ranging from 3 GHz to 300 GHz. We divide the spectrum into a number of channels, each of which corresponds to a frequency band. Thus, in OCEAN, there are a set of channels denoted by $\mathcal{F} = \{f_1, f_2, \dots, f_N\}$, where f_n is the n th channel.

Location: For 5G wireless communications, a base station covers an area. At different locations, mobile devices can experience different CSI even for the same frequency band. In OCEAN, the coverage area is initially divided into several sub-regions based on the distance to the base station. However, due to the impact of obstacles, mobile devices in the same sub-region may have very different CSI. To solve this problem, we first calculate the mean *mean* and root-mean-square deviation (RMSD) *RMSD* for a sub-region. If *RMSD* is larger than a predefined value ϵ , we consider to separate a sub-region into three: the first sub-region contains all mobile devices with the values within $[mean - RMSD, mean + RMSD]$, the second one contains all mobile devices with the values less than $mean - RMSD$, and the third one contains all mobile devices with the values larger than $mean + RMSD$. Following this way, we check all the initial sub-regions. Then, we merge the sub-regions whose RMSD is less than ϵ . Thus, we can finally obtain the number of sub-regions that are denoted by $\mathcal{L} = \{l_1, l_2, \dots, l_M\}$, where l_m is the m th location.

Time: In practice, the atmosphere density is different in different season and also in different period during a day. Thus, we consider the time format as "xx:xx:xx, month/day/year". To reduce the amount of similar data, we collect data every 5 min and thus we have the time set $\mathcal{T} = \{tt_1, tt_2, \dots, tt_I\}$, where tt_i is the i th time slot.

Temperature: The different temperature can lead to the different density of the atmosphere, hence affecting the effect of scattering and fading of radio signal propagation. Hence, we collect the real-time temperature data denoted by $\{t_1, t_2, \dots, t_P\}$, where t_p is the p th temperature level.

Humidity: The humidity also has impact on the effect of scattering and fading. We collect the real-time humidity data that is denoted by $\{u_1, u_2, \dots, u_H\}$, where u_h is the h th humidity level.

Weather: As stated before, the weather has a very important impact on the CSI. Particularly, rain can absorb the power of radio signals transmitted over some specific frequency bands. In OCEAN, the set of the weather is {sun, sun with cloud, cloud, light rain, medium rain, heavy rain, light snow, medium snow, heavy snow} and the set is denoted by $Wr = \{w_1, w_2, \dots, w_R\}$, where w_r is the r th weather level.

Therefore, these features correspond to a CSI value and a data sample consists of the information of these features and the corresponding CSI. The format of a data sample is $CSI_{f_n, l_m, tt_i, t_p, u_h, w_r}$. It is obvious that all data samples form a tensor. Therefore, the input data to the learning framework is $(f_n, l_m, tt_i, t_p, u_h, w_r, CSI)$.

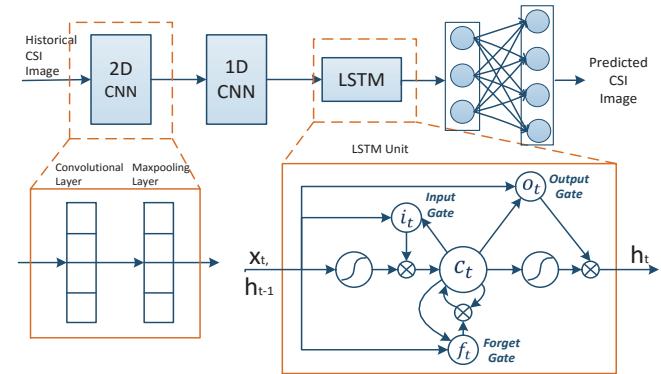


Fig. 2. The learning framework.

3.3 The Proposed Learning Framework

Our proposed deep learning framework is to conduct online CSI prediction for 5G wireless communications. In what follows, we describe the architecture of the learning framework and the data passing through the learning framework, respectively.

3.3.1 The Architecture of the Learning Framework

We develop the learning framework to analyze the historical data for online CSI prediction. Fig. 2 shows the architecture of this learning framework. Specifically, this learning framework consists of a 2D CNN network, a 1D CNN network, and an LSTM network.

This 2D CNN network is used for extracting frequency representative vector from CSI information images. It has a 2D convolutional layer, a 2D max-pooling layer, and a flatten layer. The convolutional layer is composed of several parallelized filters that connect to a local patch of CSI information images through a set of weights. These filters stride through the image along two dimensions, i.e., vertically and horizontally, to compute the convolutional products. Then, a 2D max-pooling layer is used to find the max values among the output results of the filters. Since the CSI information images become high-dimensional feature tensors, a flatten layer performs concatenating to reduce these tensors into one dimension vectors.

The following network is a 1D CNN network that is used for extracting state representative vector from frequency representative vectors. It contains a 1D convolutional layer, a 1D max-pooling layer, and a flattening layer. The functions of these layers are similar to those in a 2D CNN network. The main difference is that these filters stride through the one dimension vectors along one dimension, instead of two dimensions.

The last component is an LSTM network used for state vector prediction. The LSTM network consists of several LSTM units, each of which has an input gate, a forget gate, an output gate, and a memory cell. The mathematical description

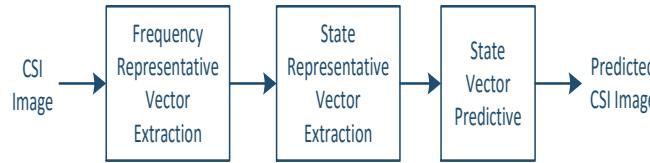


Fig. 3. The data flow in the learning framework.

of the LSTM structure is as follows:

$$\mathbf{i}_t = \sigma(\mathbf{W}_{ix}\mathbf{x}_t + \mathbf{W}_{ih}\mathbf{h}_{t-1} + \mathbf{W}_{ic}\mathbf{c}_{t-1} + \mathbf{b}_i), \quad (4)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_{fx}\mathbf{x}_t + \mathbf{W}_{fh}\mathbf{h}_{t-1} + \mathbf{W}_{fc}\mathbf{c}_{t-1} + \mathbf{b}_f), \quad (5)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \phi(\mathbf{W}_{cx}\mathbf{x}_t + \mathbf{W}_{ch}\mathbf{h}_{t-1} + \mathbf{b}_c), \quad (6)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_{ox}\mathbf{x}_t + \mathbf{W}_{oh}\mathbf{h}_{t-1} + \mathbf{W}_{oc}\mathbf{c}_t + \mathbf{b}_o), \quad (7)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \phi(\mathbf{c}_t), \quad (8)$$

where \mathbf{i}_t , \mathbf{f}_t , \mathbf{o}_t , \mathbf{c}_t , and \mathbf{h}_t are input gate, forget gate, output gate, memory cell, and hidden vector, respectively, $\mathbf{W}_{ix}, \mathbf{W}_{ih}, \mathbf{W}_{ic}, \mathbf{W}_{fx}, \mathbf{W}_{fh}, \mathbf{W}_{fc}, \mathbf{W}_{ox}, \mathbf{W}_{oh}, \mathbf{W}_{oc} \in \mathbb{R}^{2d}$ are weighted matrices, $\mathbf{b}_i, \mathbf{b}_f, \mathbf{b}_c, \mathbf{b}_o \in \mathbb{R}^d$, learned during training process, are biases of LSTM, σ is the sigmoid function, \odot stands for element-wise multiplication, d denotes the number of the LSTM units, and e is the input dimension.

The update of each LSTM unit can be briefly summarized as follows:

$$\mathbf{h}_t = LSTM(\mathbf{h}_{t-1}, \mathbf{x}_t, \theta). \quad (9)$$

where $LSTM(\cdot)$ is a combination of Eqs. (4)-(8), and θ represents all the parameters in the LSTM network.

3.3.2 The Data Flow in the Learning Framework

Input data pass through the proposed learning framework that outputs the predicted CSI, and the data flow is shown in Fig. 3. Specifically, raw data are first preprocessed and transformed into CSI information images. Then, the CSI information images are fed into a 2D CNN network to extract frequency representative vectors which are later transformed into a state representative matrix. Subsequently, this matrix is entered into a 1D CNN network to compress the state representative matrix as a state representative vector that is afterwards fed into an LSTM network. The LSTM network finally outputs the predicted state vector. In what follows, we describe CSI information image, frequency representative vector extraction, state representative vector extraction, and state vector prediction, respectively.

1) CSI Information Image: To use the proposed learning framework, the input data need to be preprocessed before feeding them into the learning framework. This is because the first component, i.e., the 2D CNN network, is initially developed for processing image data. As a result, to enable the proposed learning framework to work, we need to organize the input data.

To train color image data by a CNN network, the input data is usually a sequence of pixels that are the smallest addressable elements in an image. Each image is composed of pixels that are represented by a series of code. A channel is a gray-scale

image with the same size as the color image. Inspired by the color image representation, we organize raw data similarly. To be more specific, raw data are segmented into small cells, each of which corresponds to a pixel in an image. For each frequency band, the data of the CSI and side information are formed the pixels. Then, all pixels for a frequency band are formed as a channel. As a result, all the data can be transformed into N channels. Thus, the data in the dataset are preprocessed and converted into N channels which are fed into the learning framework.

2) Frequency Representative Vector Extraction: After receiving CSI information images, the learning framework first needs to extract features from such images. In some existing learning frameworks based on classic machine learning techniques, features are extracted and selected from datasets manually, but these operations heavily depend on the expertise in CSI analysis. Thus, we propose to explore a deep neural network for performing feature extraction and selection to obtain representative vectors from CSI information images.

In particular, we propose to utilize a CNN network to perform feature extraction and selection for extracting frequency representative vectors from CSI information images. To find such vectors, a CSI information image IMG is decomposed into multiple CSI information channels as follows:

$$[C_1, C_2, \dots, C_l] = IMG, \quad (10)$$

where $IMG \in \mathbb{R}^{h \times w \times l}$, $C_i \in \mathbb{R}^{h \times w}$, $i \in [1, l]$, w , h , and l are the width, height, and channel number of the image, respectively.

Generally, a normal CNN network has a number of convolutional filters to process these channels. For simple presentation, we describe the CNN network with one filter in this paper. Let w be a convolutional filter that is applied to a window of pixels belonged to a frequency information channel. Specifically, after a window of pixels x pass a filter, the corresponding feature pixels in the feature map x^* can be obtained as follows:

$$x^* = f(x \otimes w + b), \quad (11)$$

where b is a bias term and $f(\cdot)$ is the activation function.

Particularly, to capture the most salient features in x^* , a max-pooling operation is applied to yield a scalar as follows:

$$x^* = \max(x^*). \quad (12)$$

Then, the filters stride over each channel to obtain corresponding feature maps. As a result, a feature image IMG^* , a collection of features computed from all windows of pixels, can be expressed as follows:

$$IMG^* = [C_{11}^*, C_{12}^*, \dots, C_{1n}^*, \dots, C_{l1}^*, C_{l2}^*, \dots, C_{ln}^*], \quad (13)$$

where n is the number of filters.

At last, the output results from all convolutional filters are combined together and all of them are flattened into a vector, called frequency representative vectors.

$$V_{Fi} = concat(C_{i1}^*, C_{i2}^*, \dots, C_{in}^*), \quad (14)$$

where $concat(\dots)$ is a function flattening a feature image into a vector.

To sum up, this process can be briefly summarized as follows:

$$[V_{F1}, V_{F2}, \dots, V_{Fl}] = conv2D(IMG), \quad (15)$$

where $conv2D(\cdot)$ is used to represent the 2D convolution process.

3) State Representative Vector Extraction: After obtaining frequency representative vectors of all frequency bands, we need to construct a state representative matrix for an individual time slot by concatenating the obtained frequency representative vectors. Traditionally, the matrix needs to be fed into an RNN network like LSTM to extract temporal features from the data. However, sequentially concatenating vectors is unable to extract the relationship between frequency bands and other features. Moreover, the high-dimensional input data result in a significant number of parameters to be trained, which may incur infeasible training time. Besides, the training process may not converge until a large size of training samples are fed into the network.

To solve this problem, we propose a 1D-CNN network to compress the state representative matrix into a state representative vector. The compression process is similar to the process of extracting frequency representative vectors, except that the convolutional computation is only along the one dimension. The state representative vector of time t can be expressed as follows:

$$S_t = conv1D(V_{F1t}, V_{F2t}, \dots, V_{Flt}), \quad (16)$$

where $conv1D(\cdot)$ is a function used to represent the 1D convolution process.

4) State Vector Prediction: Our OCEAN aims at predicting the current CSI of all frequency bands based on historical input data and current feedback. Considering that the LSTM network can be excellent in sequential tasks learning, we adopt an LSTM network to predict the current state vector, since it is beneficial in learning long-term dependencies.

Based on Eq. (9), we can have the sequential feature of state representation vector as follows:

$$\mathbf{h}_{ST} = LSTM(\mathbf{h}_{ST-1}, \mathbf{S}_t, \theta). \quad (17)$$

Then, a fully-connected softmax layer is followed to generate a probability distribution of predicted state vector as follows:

$$\hat{\mathbf{S}}_T = softmax(\mathbf{W}_h \mathbf{h}_{ST} + b). \quad (18)$$

Thus, the prediction loss of our model is:

$$loss = \sum_{t=0}^T \frac{1}{2} \left(IMG_t - \widehat{IMG}_t \right)^2. \quad (19)$$

3.4 Offline-Online Two-step Training Mechanism

The CSI prediction may not be stable when using the learning framework in a practical 5G wireless communication system. To address this problem, we develop an offline-online two-step training mechanism to improve the stability of our proposed learning framework to guide the weights updating, as shown in Fig. 1.

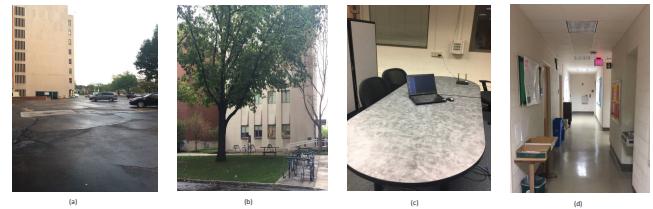


Fig. 4. The scenarios of all case studies.

For the offline training step, we first conduct the offline training on massive volume of historical data to capture the sequential pattern of CSI series along the time. Then, we reduce the system loss by using the stochastic gradient descent algorithm based on back propagation. As a result, we obtain a well-trained neural network whose weights can be represented by \mathbf{W}_h .

For the online training step, we consider that the up-to-date CSI information is important for improving the accuracy of CSI prediction. Thus, we first measure the CSI and get it from the 5G wireless communication system as feedback. Then, we build a temporal dataset that contains these recently measured CSI. Each time after update the temporal dataset, we train the model and calculate the gradients $\nabla \mathbf{W}$. Finally, we update the weights of the developed learning framework as follow:

$$\mathbf{W}_c = \mathbf{W}_h - \beta \nabla \mathbf{W}. \quad (20)$$

4 EXPERIMENTS

In this section, we evaluate the performance of the proposed CSI prediction scheme through four typical case studies: **Case I:** free space environment, **Case II:** outdoor environment, **Case III:** a workroom, and **Case IV:** a building. In what follows, we present the data collection and preprocessing, parameter settings for the learning framework, and experiment results, respectively.

4.1 Data Preparation

We collect data for the considered four case studies. To collect the CSI data, we use a Dell laptop equipped with an Intel Wireless Link 5300 network interface card as a mobile device and configure a TP-Link WiFi router as an access point. The mobile device sends successive packets to the WiFi router and extracts CSI information from its received packets. Moreover, we collect the weather data from OpenWeatherMap [16].

For **Case I:**, we consider a free space environment and collect data in a scenario as shown in Fig. 4(a). In this scenario, we place the WiFi router and the laptop in a Parking Lot and move the laptop around to collect CSI data at each point. In this scenario, there are no obstacles between the WiFi router and the laptop, and thus this is a LOS transmission. We use this scenario to simulate the free space.

For **Case II:**, we consider radio communication at a campus as an outdoor environment, as shown in Fig. 4(b). Particularly, we place the WiFi router near the window on the fourth floor and place a Laptop outside where trees and people are around.

For **Case III**, we consider a workroom as a typical indoor environment, as shown in Fig. 4(c). We consider a laboratory room in a building, and there are several cubicles and desks in this room. As shown in the figure, we place the wireless router on the desk. To collect the CSI data, we move the laptop in this room and stay each collecting point for a while.

For **Case IV**, we consider wireless communications in a building as another typical indoor environment, as shown in Fig. 4(d). Specifically, we place the wireless router on the desk of a room and move the laptop on one floor of this building. To collect the CSI data, we place the laptop in several rooms, and especially the laptop stays in each room for a period.

4.2 Parameter Settings for The Learning Framework

We set 6 measuring points in each case study as pixels of CSI image. Moreover, for each CSI image, it contains 7 channels which are CSI data in 2.4 GHz, CSI data in 5 GHz, location, time, temperature, humidity, and weather.

Based on the size of CSI image, we use a 2×2 filter for 2D CNN and a 3×1 filter for 1D CNN. For the LSTM network, there is an LSTM layer with 64 nodes. The output of the LSTM network connects two fully connected layers.

In the offline training process, the initial weights are set as random values between $[-0.1, 0.1]$, and we update them with RMSprop gradient descent algorithm [17]. For online training process, we set the update period as 5 minutes, and the $\beta = 0.1$.

4.3 Experiment Results and Discussions

We show experimental results case study by case study to validate OCEAN's performance. To show OCEAN's efficiency, we compare the computing time of obtaining the CSI information between the OCEAN and other conventional methods. Moreover, we also show the accuracy of the designed learning framework by comparing the prediction results at the training and testing process with the ground truth (i.e., the measured CSI). In addition, we also show the performance of our designed learning frame by comparing it with an artificial neural network (ANN). To show the accuracy, we show the comparison results by using figures whose horizontal axis represents time slots, and vertical denotes the CSI in dB. Furthermore, we use the red, green, and blue lines to represent the prediction results at the training process, prediction results at the testing process, and the ground truth, respectively, and 2/3 time slots are used to show the prediction results at the training process and the rest 1/3 time slots are used to show the prediction results at the testing process. To demonstrate the performance of the designed learning framework, we show the performance by tables where OCEAN represents our designed learning framework, and ANN is the one using ANN. To assess the performance, we use the average difference ratio (ADR) as the performance metric. ADR is calculated by $1/T \sum_1^T \frac{|Predictedvalue - Measuredvalue|}{Measuredvalue}$, where T is the total time slots we compared.

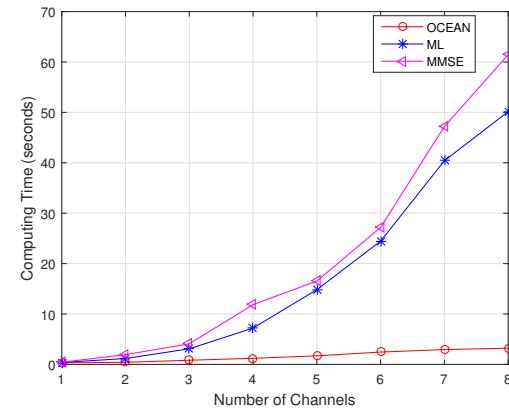


Fig. 5. Computing time comparison.

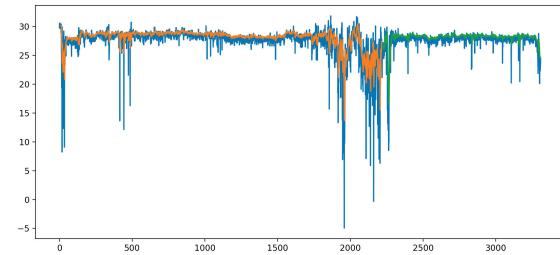


Fig. 6. Case study I: Results comparison between the predicted and measured CSI.

4.3.1 The Computing Time Comparison

To validate the efficiency of using our developed learning method, we show the computing time of obtaining the CSI and compare it with that using conventional CSI estimation methods (i.e., ML-based, MMSE-based). To compare the computing time, we change the number of channels.

Fig. 5 shows the comparison of the computing time. From this figure, we can observe that our developed learning method has the lowest computing time. For example, when the channel number is 8, the computing time are 3.213s, 50.152s, 61.436s for the developed method, ML-based method, and MMSE-based method, respectively. Moreover, we can also notice that the computing time is increasing with the increase in the number of channels. Particularly, the computing time of the ML-based and MMSE-based methods are polynomially increasing with the channel number. When having a large number of channels, the conventional CSI estimation methods are unable to obtain the CSI estimation results quickly, and thus fail catching the dynamics of the wireless environment.

4.3.2 Case Study I: A Free Space Environment

Fig. 6 shows the prediction results along the time. In this experiment, we consider a free space environment and collect CSI data in a Parking Lot. Specifically, we compare the prediction results at the training and testing process with the measured CSI collected in this case study. From this figure, we can find that both of them are very close to the measured CSI. This result implies that the designed learning framework is very beneficial in predicting CSI data.

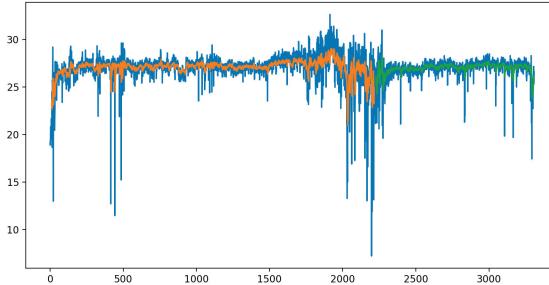


Fig. 7. Case study II: Results comparison between the predicted and measured CSI.

Table. 1 illustrates the prediction performance by comparing the designed learning framework with the ANN. From this table, we can find that the average ADR of the OCEAN and ANN are 2.730% and 6.834%, respectively. Specifically, we can see that OCEAN's ADR for different measuring points (MPs) are very low, which implies that OCEAN is very useful in predicting CSI. Moreover, we can also observe that the OCEAN has the lower ADR than the ANN, which shows us the need of considering the combination of a CNN network and an LSTM network. Therefore, both observations give us a sight that we need to explore spatiotemporal information to design an online CSI prediction scheme.

4.3.3 Case Study II: Outdoor Environment with Obstructions

Fig. 7 shows the predicted CSI along the time. In this experiment, we consider an outdoor environment and collect CSI data at the campus. In particular, we compare the prediction results at the training and testing process with the measured CSI collected in this case study. We can observe from this figure that the curves of the training and testing process are very close to the curve of the measured CSI. This result validates OCEAN's efficacy again.

Table. 2 illustrates the prediction performance of our designed learning framework and compares its results with ANN. To be more specific, the average ADRs of the OCEAN and ANN are 2.650% and 7.255%, respectively. Particularly, we can see that OCEAN's ADRs are very low, which implies that OCEAN is very useful in predicting CSI. Moreover, we can also observe that the OCEAN has the lower ADR than the ANN, which shows us the need of combining a CNN network and an LSTM network.

4.3.4 Case Study III: Within A Workroom

Fig. 8 shows the prediction results along the time. In this experiment, we consider an indoor environment and collect CSI data within a workroom. In particular, we compare the prediction results at the training and testing process with the measured CSI collected in this case study. From this figure, we can find that both of them are very close to the measured CSI, which implies that our proposed scheme can work well in predicting CSI data.

Table. 3 illustrates the prediction performance of the designed learning framework and compares its results with ANN.

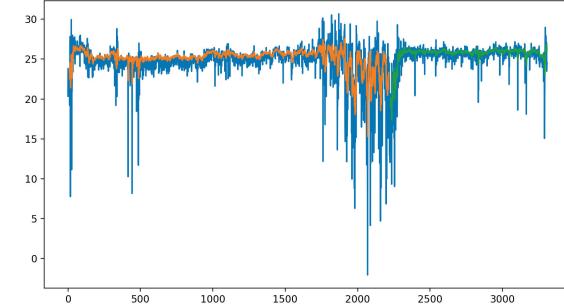


Fig. 8. Case study III: Results comparison between the predicted and measured CSI.

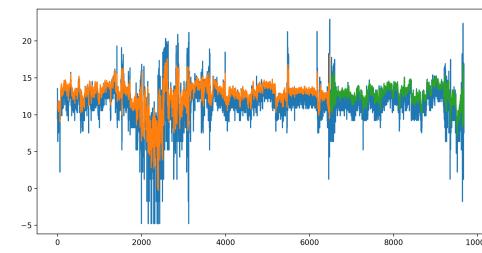


Fig. 9. Case study IV: Results comparison between the predicted and measured CSI.

We can find the average ADRs of the OCEAN and ANN are 3.457% and 6.837%, respectively. Furthermore, we can observe that OCEAN's ADRs are very low, which implies that OCEAN is very effective in predicting CSI. Moreover, we can also notice that the OCEAN has the lower ADR than the ANN, which shows us the need of exploring spatiotemporal relationship for CSI prediction.

4.3.5 Case Study IV: Within A Building

Fig. 9 shows the predicted CSI along the time. In this experiment, we consider an indoor environment and collect CSI data within a building. In particular, we compare the prediction results at the training and testing process with the measured CSI collected in this case study. We can observe from this figure that the curves of the training and testing process are very close to the curve of the measured CSI. This result validates OCEAN's efficacy as well.

Table. 4 illustrates the prediction performance of the designed learning framework and compares its results with ANN. We can see that the average ADRs of the OCEAN and ANN are 3.223% and 5.656%, respectively. Moreover, we can observe that OCEAN's ADRs are very low, which implies that OCEAN is very useful in predicting CSI. In addition, we can also find that the OCEAN has the lower ADR than the ANN, which further shows us the need of mining the spatiotemporal relationship to enhance the prediction precision of the CSI prediction scheme.

To sum up, the above experiment results have shown the OCEAN's effectiveness and efficacy of our designed learning framework. In particular, we can see that our proposed scheme can achieve very low ADRs in each study case. Moreover, comparing with the learning framework only using an ANN

TABLE 1
Case study I: The comparison of ADR achieved by the OCEAN and ANN.

	MP 1	MP 2	MP 3	MP 4	MP 5	MP 6	The average
OCEAN	2.477%	2.753%	2.928%	2.395%	2.388%	2.4339%	2.730%
ANN	4.796%	6.226%	6.172%	10.090%	8.653%	5.060%	6.834%

TABLE 2
Case study II: The comparison of ADR achieved by the OCEAN and ANN.

	MP 1	MP 2	MP 3	MP 4	MP 5	MP 6	The average
OCEAN	2.501%	2.974%	2.884%	2.448%	2.791%	2.302%	2.650%
ANN	6.582%	7.317%	7.780%	6.366%	6.346%	9.139%	7.255%

TABLE 3
Case study III: The comparison of ADR achieved by the OCEAN and ANN.

	MP 1	MP 2	MP 3	MP 4	MP 5	MP 6	The average
OCEAN	2.426%	3.150%	3.122%	5.104%	4.377%	2.563%	3.457%
ANN	3.954%	4.826%	4.133%	6.997%	7.008%	5.504%	6.837%

TABLE 4
Case study IV: The comparison of ADR achieved by the OCEAN and ANN.

	MP 1	MP 2	MP 3	MP 4	MP 5	MP 6	The average
OCEAN	2.779%	3.148%	2.696%	2.760%	4.565%	3.595%	3.223%
ANN	5.338%	6.349%	6.156%	5.226%	5.958%	4.914%	5.656%

network, our proposed scheme has better performance, which sheds the light that the CSI prediction does not only depend on the spatial information but also the temporal information.

5 RELATED WORKS

So far, previous studies have examined how to apply machine learning techniques to 5G wireless communication networks. Specifically, these works have studied some critical issues, such as mobile traffic prediction, wireless network optimization, and user behavior prediction. We summarize them as follows.

Researchers have investigated the mobile traffic forecasting problem in wireless networks. Wang *et al.* [18] modeled spatiotemporal mobile traffic data based on the existing dataset and proposed to employ deep learning approaches for the traffic prediction. Sciancalepore *et al.* [19] considered network slicing traffic forecasting and developed a Holt-Winters theory based forecasting solution to predicting future traffic levels per network slice. Han *et al.* [20] also analyzed mobile traffic load for a customer and predicted busy hour for each month by exploiting the support vector machine (SVM) technique.

Another research line is to optimize wireless network design by employing machine learning techniques. Kato *et al.* [21] characterized mobile traffic in heterogeneous wireless networks, and proposed a deep learning based method to control mobile traffic flowing among multiple wireless access networks. Xu *et al.* [22] considered the problem of radio resource

allocation in cloud radio access networks and developed a radio resource allocation strategy based on a deep reinforcement learning technique that is used to approximate the action-value function. He *et al.* [23], [24] also utilized a deep reinforcement learning technique for jointly optimizing both caching and interference alignment for 5G wireless communications. Liu *et al.* [25] considered using the deep learning method to identify the link usage of the flow constrained optimization problem in 5G wireless networks.

Some other previous works have explored machine learning techniques to study several other vital issues in 5G wireless networks, such as user behaviors and quality of experience (QoE). Parwez *et al.* [26] used wireless network data, i.e., call detail record, to identify the anomalous behavior of a wireless network. Lin *et al.* [27] applied a machine learning method to predict users' QoE by considering users' features like user numbers and CSI experienced by a user.

From the above analysis, we notice that machine learning techniques have been applied in 5G wireless networks to enhance wireless network performance and improve users' QoE. However, few of them apply such techniques into predicting the CSI of a radio link in wireless networks.

6 CONCLUSIONS

In this paper, we studied the CSI estimation problem for 5G wireless communications. To efficiently acquire the CSI, we proposed an online CSI prediction scheme, called OCEAN,

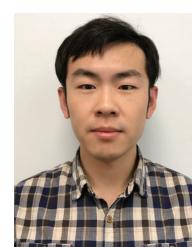
for predicting CSI from historical data in 5G wireless communication systems. To be more specific, we first identified several essential features affecting the CSI of a radio link and showed a data sample that is composed of the information of these features and the CSI. We then designed a learning framework that is a combination of a CNN (convolutional neural network) and an LSTM (long short term with memory) network. We further developed an offline-online two-step training mechanism, enabling the prediction results to be more stable when applied in practical 5G wireless communication systems. To validate OCEAN's efficacy, we conducted extensive experiments by considering four typical case studies and illustrated the experiment results to show that OCEAN not only obtains the predicted CSI values very quickly but also achieves highly accurate CSI prediction.

REFERENCES

- [1] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 2491–2509, Jul. 2005.
- [2] M. Li, Y. Meng, J. Liu, H. Zhu, X. Liang, Y. Liu, and N. Ruan, "When CSI meets public WiFi: Inferring your mobile phone password via WiFi Signals," in *Proc. ACM CCS'16*, (Vienna, Austria), pp. 1068–1079, Oct. 24 - 28 2016.
- [3] N. Mokari, K. Navaie, and M. G. Khoshkholgh, "Downlink radio resource allocation in OFDMA spectrum sharing environment with partial channel state information," vol. 10, pp. 3482–3495, Oct. 2011.
- [4] A. G. Davoodi and S. A. Jafar, "Generalized degrees of freedom of the symmetric K user interference channel under finite precision CSIT," *IEEE Trans. Inform. Theory*, vol. 63, pp. 6561–6572, Oct. 2017.
- [5] Z. Du, X. Song, J. Cheng, and N. C. Beaulieu, "Maximum likelihood based channel estimation for macrocellular OFDM uplinks in dispersive time-varying channels," vol. 10, pp. 176–187, Jan. 2011.
- [6] E. Karami, "Tracking performance of least squares MIMO channel estimation algorithm," *IEEE Trans. Commun.*, vol. 55, pp. 2201–2209, Nov. 2007.
- [7] J. Ma, S. Zhang, H. Li, N. Zhao, and A. Nallanathan, "Iterative LMMSE individual channel estimation over relay networks with multiple antennas," *IEEE Trans. Veh. Technol.*, vol. PP, pp. 1–13, Aug. 2017.
- [8] C. Luo, G. Min, F. R. Yu, M. Chen, L. T. Yang, and V. C. M. Leung, "Energy-efficient distributed relay and power control in cognitive radio cooperative communications," *IEEE J. Select. Areas Commun.*, vol. 31, pp. 2442–2452, Nov. 2013.
- [9] S. Khan, M. Alam, and M. Franzle, "A hybrid MAC scheme for wireless vehicular communication," in *Proc. IEEE EUROCON'17*, (Ohrid, R. Macedonia), pp. 889–895, 6 - 8 Jul. 2017.
- [10] C. Luo, G. Min, F. Yu, Y. Zhang, L. T. Yang, and V. C. M. Leung, "Joint relay scheduling, channel access, and power allocation for green cognitive radio communications," vol. 33, pp. 922–932, May 2015.
- [11] Ericsson, "Ericsson mobility report, 2017," <https://www.ericsson.com/assets/local/mobile-report/documents/2017/ericsson-mobility-report-june-2017.pdf>.
- [12] M. Chowdhury, A. Manolako, and A. Goldsmith, "Multiplexing and diversity gains in noncoherent massive MIMO systems," vol. 16, pp. 265–277, Jan. 2017.
- [13] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shakir, "MmWave massive-MIMO-based wireless backhaul for the 5G ultra-dense network," *IEEE Wireless Commun. Mag.*, vol. 22, pp. 13–21, Oct. 2015.
- [14] D. Parsons, *The Mobile Radio Propagation Channel*. New York, USA: Wiley, 1994.
- [15] M. Marcus and B. Pattan, "Millimeter wave propagation: Spectrum management implications," *IEEE Microwave*, vol. 6, pp. 54–62, Jun. 2005.
- [16] "OpenWeatherMap," <https://openweathermap.org/api>.
- [17] M. Andrychowicz, M. Denil, S. G. Colmenarejo, M. W. Hoffman, D. Pfau, T. Schaul, B. Shillingford, and N. Freitas, "Learning to learn by gradient descent by gradient descent," in *Proc. NIPS'16*, (Barcelona, Spain), pp. 1–9, Dec. 5 - 10 2016.
- [18] J. Wang, J. Tang, Z. Xu, Y. Wang, G. Xue, X. Zhang, and D. Yang, "Spatialtemporal modeling and prediction in cellular networks: A big data enabled deep learning approach," in *Proc. IEEE INFOCOM'17*, (Atlanta, GA, USA), pp. 1323–1331, 1 - 4 May 2017.
- [19] V. Sciancalepore, K. Samdanis, X. Costa-Perez, D. Bega, M. Gramaglia, and A. Banchs, "Mobile traffic forecasting for maximizing 5G network slicing resource utilization," in *Proc. IEEE INFOCOM'17*, (Atlanta, GA, USA), pp. 2583–2591, 1 - 4 May 2017.
- [20] R. Han, Z. Jia, X. Qin, C. Chang, and H. Wang, "Application of support vector machine to mobile communications in telephone traffic load of monthly busy hour prediction," in *Proc. Fifth International Conference on Natural Computation*, (Tianjin, China), pp. 349–353, 14 - 16 Aug. 2009.
- [21] N. Kato, Z. Fablullah, B. Mao, F. Tang, O. Akashi, T. Inoue, and K. Mizutani, "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective," *IEEE Wireless Commun. Mag.*, vol. 24, pp. 146–153, Jun. 2017.
- [22] Z. Xu, Y. Wang, J. Tang, and M. C. Gursoy, "A deep reinforcement learning based framework for power-efficient resource allocation in cloud RANs," in *Proc. IEEE ICC'17*, (Paris, France), pp. 1–6, May 21 - 25 2017.
- [23] Y. He, C. Liang, F. R. Yu, N. Zhao, and H. Yin, "Optimization of cache-enabled opportunistic interference alignment wireless networks: A big data deep reinforcement learning approach," in *Proc. IEEE ICC'17*, (Paris, France), pp. 1–6, May 21 - 25 2017.
- [24] Y. He, Z. Zhang, F. R. Yu, N. Zhao, H. Yin, V. C. M. Leung, and Y. Zhang, "Deep reinforcement learning-based optimization for cache-enabled opportunistic interference alignment wireless networks," *IEEE Trans. Veh. Technol.*, vol. PP, pp. 1–15, Sept. 2017.
- [25] L. Liu, Y. Cheng, L. Cai, S. Zhou, and Z. Niu, "Deep learning based optimization in wireless network," in *Proc. IEEE ICC'17*, (Paris, France), pp. 1–6, May 21 - 25 2017.
- [26] M. S. Parwez, D. B. Rawat, and M. Garuba, "Big data analytics for user-activity analysis and user-anomaly detection in mobile wireless network," *IEEE Trans. Industrial Informatics*, vol. 13, pp. 2058–2065, Aug. 2017.
- [27] Y. Lin, E. M. R. Oliveira, S. B. Jemaa, and S. E. Elayoubi, "Machine learning for predicting QoE of video streaming in mobile networks," in *Proc. IEEE ICC'17*, (Paris, France), pp. 1–6, May 21 - 25 2017.



Changqing Luo received his B.E. and M.E. degree from Chongqing University of Posts and Telecommunications in 2004 and 2007, respectively. He is currently working towards the Ph.D. degree in the Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH. His research interests include cybersecurity, big data, and resource management and network optimization in cyber-physical systems.



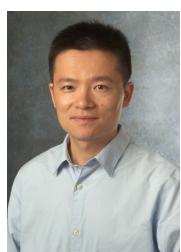
Jinlong Ji received the B.E. degree and M.E. degree in Telecommunication Engineering from Xidian University, Xiān, China, in 2010 and 2013, respectively. He is currently pursuing the Ph.D. degree in the Department of Electrical Engineering and Computer Science, Case Western Reserve University. His current research interests include deep learning, smart health, natural language processing, and cloud computing.



Qianlong Wang received the B.E. degree in Electrical Engineering from Wuhan University, China, in 2013, and the M.E. degree in Electrical and Computer Engineering from Stevens Institute of Technology, Hoboken, NJ, in 2015, respectively. He is currently working toward the PhD degree in the Department of Electrical Engineering and Computer Science, Case Western Reserve University. His research interests include big data, and deep learning. He is a student member of the IEEE.



Xuhui Chen received the B.E. degree in Information Engineering from Xidian University, XiāZan, China, in 2012, and the M.E. degree in Electrical and Computer Engineering from Mississippi State University, Starkville, in 2015, respectively. She is currently pursuing the Ph.D. degree in the Department of Electrical Engineering and Computer Science, Case Western Reserve University. Her current research interests include machine learning, smart health, security and privacy.



Pan Li (S'06-M'09] received the B.E. degree in Electrical Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2005, and the Ph.D. degree in Electrical and Computer Engineering from University of Florida, Gainesville, in 2009, respectively. He is currently an Associate Professor in the Department of Electrical Engineering and Computer Science, Case Western Reserve University. He was an Assistant Professor in the Department of Electrical and Computer Engineering at Mississippi State University between August 2009 and August 2015. His research interests include network science and economics, energy systems, security and privacy, and big data. He is a member of the IEEE and ACM.