Channel Estimation Using CNN-LSTM in RIS-NOMA Assisted 6G Network

Link to publication record in Ulster University Research Portal

Document Version
Publisher's PDF, also known as Version of record

IEEE ComSoc *IEEE Communications Society* | **IEEE COMPUTER SOCIETY** | **IEEE Signal Processing Society** | **VTS** *Connecting the Mobile World* | *IEEE Transactions on* **Machine Learning in Communications and Networking**

Channel Estimation Using CNN-LSTM in RIS-NOMA Assisted 6G Network

Chi Nguyen[1], Tiep M. Hoang[2], and Adnan A. Cheema[1]

[1]SenComm Research Lab, School of Engineering, Ulster University, Belfast, BT15 1AP, UK
[2]College of Engineering, Design and Computing, University of Colorado Denver, CO 80204, USA

Corresponding author: Adnan A. Cheema (e-mail: a.cheema@ulster.ac.uk)

ABSTRACT The combination of non-orthogonal multiple access (NOMA) and reconfigurable intelligent surface (RIS) technologies is proposed to meet the demands of data rate, latency, and connectivity in sixth-generation (6G) networks. The two techniques can support each other to increase the performance of the 6G system. In a RIS-aided system, channel estimation is a challenging problem, especially when applying passive RIS which has no signal processing. This paper proposes a deep learning (DL)-based channel estimation method using a convolutional long-short term memory (CNN-LSTM) model for RIS-NOMA wireless communication systems that integrate RIS and NOMA techniques. CNN-LSTM leverages both the benefits of convolutional neural network (CNN) as well as long-short term memory (LSTM), in which CNN can capture special features while LSTM can capture temporal features of time-series data. The simulation results indicate that the proposed CNN-LSTM model shows its robustness toward the variation of the RIS-NOMA system parameters, i.e., transmit signal-to-noise ratio (SNR), power allocation factor, and the number of RIS elements. The impacts of the RIS-NOMA system parameters on the prediction accuracy of the proposed DL-based channel estimation methods are evaluated via different performance metrics. The results reveal that the performance accuracy in terms of normalized root mean square error (NRMSE), coefficient of determination R-squared score (R2 score), mean absolute scaled error (MASE), and mean absolute percentage error (MAPE) increases with an increased transmit SNR, power allocation factor of the first user and the number of RIS elements. Additionally, the CNN-LSTM prediction performance shows its superiority as compared to those of the four benchmark models including the CNN1D-LSTM model using one-dimensional convolution layer (conv1D), CNN1D-BiLSTM model using bidirectional long-short term memory, CNN model and LSTM model.

INDEX TERMS Channel Estimation, CNN-LSTM, LSTM, NOMA, 6G Network, Reconfigurable Intelligent Surface (RIS).

I. INTRODUCTION

THE upcoming next-generation sixth-generation (6G) networks, with the explosion of up to trillions of intelligent devices and massive connections between devices, put forward requirements for fast speed, low power consumption, low-latency, and energy efficiency. Reconfigurable intelligent surface (RIS) emerges as one of the potential and innovative techniques, apart from multiple-input multiple-output (MIMO), millimeter-wave (mm-wave), and relays communications, for the future 6G networks. RIS can enhance the spectrum efficiency, energy efficiency and throughput of wireless communications by leveraging its ability to reflect and redirect signals in a controlled manner [1], low hardware footprint [2], and capability to expand the range of network coverage [3]. RIS is composed of elements, in which each element or group of elements independently adjust the amplitude and phase responses of incident signals in real-time to obtain energy toward an intended direction [4]. By appropriate adjustments and designs for phase shifts, RIS is able to manipulate the communication environment, increase the link quality, and improve environment coverage [5]. In addition, RIS can operate in a passive mode containing low-cost passive RIS elements with low energy consumption,

contributing to an increase in the energy efficiency for wireless communication systems.

Although the next-generation multiple access (NGMA) techniques are in the investigation, NOMA is recommended for the 6G networks [6]. Regarding new requirements, the key concept of NGMA is to enable massive devices to connect efficiently and intelligently in a given radio resource under extremely low latency, high-reliability communications, high capacity, and high data rate. The existing orthogonal multiple access (OMA) techniques might be infeasible for these extreme demands. The reason lies behind the fact that the current OMA access techniques only serve a single user in each orthogonal radio resource block, and thus, limit the system capacity and spectrum efficiency [7]. To provide a higher spectrum efficiency compared to OMA, NOMA technique introduces a mechanism in which the given orthogonal radio resource is attributed to multiple users simultaneously. NOMA is a promising research trend for the evolution of critical and massive machine-type communications in 6G [8], [9]. Regarding new technologies, the integration of NOMA with emerging techniques such as RIS, Terahertz communications, index modulation, random access, and visible light communications brings new research challenges in 6G. NOMA is combined with Terahertz communications for vehicle-to-vehicle and vehicle-to-infrastructure applications to improve performance [10]. In [11], the integration of NOMA and RIS is discussed as one of the potential scenarios for 6G and has been proposed by the research community [12].

In multi-user 6G networks with massive devices, NOMA is a potential technique integrated with RIS to meet the demands of data rate, latency, and connectivity [13]. On the one hand, NOMA systems are useful in tackling the problems of user (UE) explosion in future networks by allowing massive devices to connect simultaneously. A NOMA system serves multiple users (UEs) with different quality of service (QoS) in the same orthogonal resource (e.g., time, frequency, and code) block by exploiting a power domain. The power-domain-based technique exploits the difference of channels between UEs for multiplexing by using superposition combining (SC) at a transmitter and applying successive-interference-cancellation (SIC) at a receiver to eliminate the co-channel interference before decoding its own signal [14]. Additionally, a system using NOMA is superior to a conventional OMA system in terms of average sum-rate and outage probability [15]. On the other hand, with the aid of RIS, a NOMA system increases its performance by additional signal diversity with no more time slot or energy. Compared to a conventional NOMA system where channel conditions are determined by the propagation environment, the counterpart with the support of RIS is able to tune the channel quality of each UE by changing the reflection coefficient or the position of RIS. This flexibility of RIS toward channel quality provides a smart NOMA design and the NOMA system turns into a system based on QoS

instead of on channel gains from the environment [16], which is useful in 6G networks where different QoS devices are connected. Due to the excellent integration of RIS and NOMA, many studies recently have been paying attention to reconfigurable-combined-NOMA (RIS-NOMA) systems.

II. BACKGROUND

This section presents a review of RIS-NOMA wireless communication systems and the use of DL methods for channel estimation in conventional RIS-aided systems without NOMA. From this background, we outline the motivations and contributions in this paper.

A. RIS-NOMA wireless communication systems

Recent studies on RIS-NOMA systems mainly focus on system performance analysis by solving optimization problems. It was demonstrated that deploying RIS for a NOMA-assisted system or applying NOMA for a RIS-aided system can enhance the system performance compared to the conventional NOMA (without RIS) or OMA-based RIS systems, respectively. Specifically, by optimizing the rate performance of the system, the study in [17] shows that the integration of RIS and NOMA enables the system to achieve a higher rate than an OMA-based RIS system and a conventional NOMA system. In [18], in an effort to maximize the weighted sum rate and propose an UE order scheme for NOMA system, it is shown that an optimal deployment location of RIS can enlarge the disparities among channels of UEs in a RIS-NOMA system and increases channel gains of all UEs for a RIS-OMA system. With an aim of transmit power optimization, the study in [19] proves that the proposed a RIS-NOMA system using a 32-elements-RIS obtains a higher performance than a conventional MIMO system equipped with 64 antennas. Similarly, the paper in [20] and [21] studied maximizing the system throughput and energy efficiency of RIS-NOMA systems. Of these, the RIS-NOMA systems show their superiority in terms of system throughput and energy efficiency, as compared to systems only utilizing RIS or NOMA.

The aforementioned studies have highlighted the merits of the integration of NOMA and RIS with an assumption that the channel state information (CSI) of the network is available at the transmitter but no consideration has been given to the importance and challenge of estimating these channels. For NOMA systems, an accurate estimation of channel gains is a prerequisite to perform SIC at each receiver [17]. In practice, the challenge of channel estimation in a passive RIS-assisted network lies in that a RIS is composed of passive elements without any receive or transmit processing chains, thus channel estimation is only implemented at the transmitter or the receiver side. In addition, a large number of RIS elements in reality leads to an increased number of unknown channel parameters that needed to be estimated [22]. So far, channel estimations for RIS-NOMA systems

have been still a research gap on which existing works have rarely focused.

B. DL approaches for channel estimation in a conventional RIS-assisted system without NOMA

Recently, channel estimation for conventional RIS-aided systems has been widely studied. However, channel estimation methods, especially DL-based channel estimator, for passive RIS-assisted systems have been investigated in a few recent works. As the challenges of channels estimation for passive RIS-aided systems discussed above, our work only considers channel estimation methods for passive RIS-aided systems. A channel estimation-based DL algorithm, which is a data-driven approach aids RIS to overcome challenges caused by a conventional RIS-assisted channel estimation algorithms, i.e., minimum mean square error (MMSE) or least-square methods. Specifically, the pilot overhead when applying conventional channel estimation methods is prohibitively high [23], [24], resulting in a limitation of spectral efficiency and a delay in channel estimation. Although in [25] and [26], the authors proposed methods to tackle these issues, these approaches still required a reduction in beamforming capacity and assume channel sparsity in the network, respectively. DL-based channel estimators have shown their potential by reducing the high pilot overhead which is caused by an increased number of RIS elements in RIS-OMA systems [27], [28].

The channel estimation algorithm based on DL approaches for RIS-aided systems, particularly multi-users ones, has been investigated in recent studies. Compared to the conventional linear channel estimator, i.e. minimum-mean-square-error or least square method, DL-based approaches exploit non-linear inherent relationships between the input-output signals to produce more reliable channel estimations for RIS-assisted systems with a less computational cost [29]. A DL model with an orthogonal matching pursuit algorithm followed by the residual network is proposed for a RIS-aided multi-users massive system [30]. This method is built on the assumption of a quasi-sparse structure of the cascaded angular channel, therefore is not suitable for the scenarios of frequency-selective fading in a time-varying environment. In particular, the author in [4] applied generative adversarial networks based on convolutional blind denoising, a convolutional blind denoising network, and multiple residual dense networks to estimate cascaded uplink channels of a time division duplex (TDD) passive RIS-aided communication system. This study is under the assumption that the optimal phase shift matrix is different for one or more UEs in each time slot. In this paper, this assumption is unnecessary; as our proposed RIS-NOMA wireless system serves multiple UEs at the same time block, therefore the optimal phase shift is considered for the overall system at a time instead of for each UE. A CNN-based deep residual learning (CDReL) model is proposed in [31] to estimate cascaded channels in a RIS-assisted TDD multi-user system based on noisy pilot-based

observations. This study is based on an OMA technique, therefore an optimal scheme for the matrix of pilot sequence at all UEs is necessary to improve the received signal at base station (BS) before applying channel estimation. However, this optimal scheme can be ignored in our proposed downlink NOMA-aid system in this study owing to the fact that a NOMA system applies SC received signals at the transmitter and SIC at UEs.

Several studies have focused on channel estimation-based DL methods for RIS-assisted systems in time-varying environments. A time-varying channel estimation using a recurrent neural network combined with ordinary differential equations for a passive RIS-assisted single UE system is investigated in [28]. Although the mobility speed at the receiver side is taken into account, this study only observes the system for a single UE. In addition, it applies a control RIS in an on/off pattern during a pilot block to reduce the length of the cascaded channel with an increased the number of RIS elements. No benchmark channel estimation methods have been applied to compare with the proposed method in this study. A fast time-varying channel estimation approach is studied in [32] by applying novel sparse-connected LSTM along with a three-stages channel estimation approach for a full-duplex RIS-assisted multi-user multiple-input single-output (MU-MISO). The channel from BS to RIS is considered in a large-time scale, and the channels from RIS to each UE is considered on a small-time scale only. Hence, the cascaded channel, which is the combination of channel from BS to RIS and from RIS to each UE, disregards small time-scale variation of the former as well as large-scale variation of the latter. Most of the aforementioned studies on passive RIS-aided systems applying OMA technique, in which all devices are fixed or time-varying channels were derived from geometric channel models.

Recently, channel estimation based on DL algorithms for an integrated sensing and communication (ISAC) systems that are assisted with RIS has been investigated. The authors in [33] proposed a three-stage channel estimation using CNN architecture for a RIS-assisted ISAC system. This approach aims to overcome the inherent inference in a system containing both sensing and communications signals. The proposed estimation DL method is applied for both direct channels, including UE-BS and target-BS and reflected ones, consisting of UE-RIS-BS and BS-target-RIS-BS, and proved to outperform the baseline least square method. This approach requires turning on RIS components at second stage, which leads to battery consumption. In addition, the proposed method is specifically designed for uplink channels and is not easily extended to downlink. The focus of this investigation was on a scenario where the devices within the system were stationary, and more specifically, the link between the RIS and BS was established over a very limited distance of 2 meters.

Although there are many papers working on channel estimation for RIS-aided systems, most of the aforemen-

tioned studies on passive RIS-aided systems apply OMA technique. Furthermore, the fixed or time-varying channels have been mostly obtained through the use of geometric channel models, specifically, the Saleh-Valenzuela model which assumes that the RIS has a negligible size in comparison to the distance between the transmitter and receiver. This assumption could be impractical when RIS is placed in close proximity to BS.

Accurate channel estimation is more crucial for the NOMA-based SIC system than for the OMA system because the SIC decoder at the receiving end heavily depends on the CSI. Additionally, due to a large number of RIS elements and the multiuser nature of NOMA, there are numerous RIS-related channel coefficients that require estimation in a RIS-NOMA system. In channel estimation, it is expected that estimation may take a large number of time slots in a RIS-NOMA system which is usually directly related to the number of reflecting elements and users, and it has the potential to surpass the maximum duration of the channel coherence time [34]. Different from the traditional RIS system, a RIS-NOMA system differentiates between UEs using power allocation factors. When a UE has lower power allocation, it may receive a weaker signal as it moves farther away from the BS, which can lead to difficulties in channel estimation.

C. MOTIVATION AND CONTRIBUTIONS

As discussed in the previous section, DL-based channel estimation approaches for passive RIS-NOMA systems have not been explored sufficiently in existing studies yet. This paper is an early attempt to study a DL-based channel estimation method for a downlink multi-user RIS-NOMA wireless communication system. However, NOMA poses new challenges for a RIS-aided system in terms of channel estimation. Firstly, channel coefficient matrices become larger due to a huge number of channels produced from the nature of multiple users in NOMA combined with many RIS elements. This challenge is more complicated when RIS is fully passive which makes it difficult to estimate the direct channel from BS or UE to RIS. In this case, the cascaded channels are estimated alternatively [34]. However, the cascaded channels have complicated statistics and high dimensions due to the large number of channel coefficients in a RIS-NOMA system. In this situation, applying conventional methods such as least square (LS) or least minimum mean square errors (LMMSE) leads to an expensive computation with the possibility to cause more channel estimation errors [31]. DL-based channel estimation may be a good solution to cope with the problem of a large amount of channel dataset. In addition, when it comes to the time-series set of channel coefficients of the RIS-NOMA model system, none of the available time-series models gives the best results in various situations because it is not easy to identify the underlying data generation process of time-series. Using individual models is generally insufficient to

exploit all the characteristics of time-series data [35]. This demands a proper design of a DL-based channel estimation method to adapt to an unfamiliar wireless system model as well as propagation environment, like the proposed RIS-NOMA model in the paper. The contributions of the paper are summarized as follows:

- The paper proposes a potential RIS-NOMA wireless system model for 6G networks, for which the channel estimation is investigated. The proposed wireless system model is different from those in previous studies that also work on channel estimation for a RIS-aid wireless system. (i) Firstly, NOMA is considered as a multiple access technique of the proposed model; hence, the system exploits the power domain of the system that is defined by the power allocation factor assigned to each UE, instead of applying pilot schemes or time division schemes as in previous RIS-aided OMA system. (ii) Secondly, different from [29] where only single user mobility is observed, in this paper, all users slowly move further from RIS to avoid the reduction of transmission performance, e.g., outage probability, and the transmission rate of each UE. (iii) Thirdly, unlike other proposed system models that take control of RIS or use active RIS, this article suggests the passive RIS for the system, in which no signal processing is available at RIS. In addition, in contrast to [32], all channels in the proposed wireless system suffer from both large-scale fading and small-scale fading.

- A data generation algorithm is developed to generate a time-series dataset based on the parameters of the proposed RIS-NOMA system model. This dataset is utilized for the study of channel estimation in the RIS-NOMA system and can be further investigated to design and test new approaches to improve end-to-end system-level performance.

- A design of the CNN-LSTM algorithm is proposed for channel estimation in the RIS-NOMA system by exploiting all characteristics of the generated time-series dataset including spatial features and temporal features. The time-series dataset is very sensitive to the DL models; in other words, a small change in the dataset due to variation of the parameters in the RIS-NOMA system could lead to a failure of channel predictions. For this reason, the CNN-LSTM channel estimation algorithm is designed to work efficiently for a wide range of RIS-NOMA system parameters, including power allocation factor, signal-to-noise ratio (SNR), and the number of RIS, rather than for a fixed set of parameters from optimization problems.

- Extensive simulations are carried out to verify the performance accuracy of the proposed DL-based channel estimator for the proposed RIS-NOMA systems. The performance accuracy metrics of the proposed DL model are evaluated under variation of RIS-NOMA system parameters, i.e., a transmit SNR at BS, power
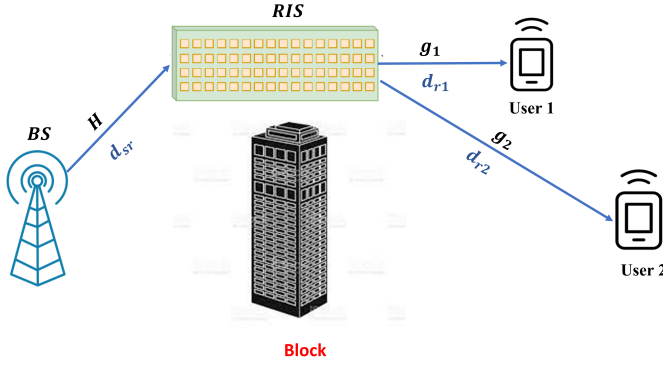
FIGURE 1: A system of RIS-aided downlink communication.

allocation factors for each UE, and the number of RIS elements. Through this evaluation, the robustness of CNN-LSTM against a large number of RIS-NOMA systems is verified.

- The paper provides a comprehensive comparison of the proposed CNN-LSTM using two-dimensional convolution layer (conv2D) convolution layers over the four benchmarks, including CNN1D-LSTM using conv1D layer, LSTM model, CNN1D-BiLSTM using bidirectional long-short term memory, conventional CNN, and conventional LSTM to demonstrate its superiority over other compared models in channel estimation for the proposed downlink RIS-NOMA system.

The rest of the paper is organized as follows. The RIS-NOMA system model and the problem formulation are included in Section III. In Section IV, we propose a CNN-LSTM architecture as a DL-based channel estimator for the proposed wireless communication system. Section V presents the simulation setup and simulation results of the proposed DL model. Finally, we conclude the paper in Section VI.

III. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, a NOMA-RIS system model is presented at first then a problem relating to channel estimation from this system model is defined. Finally, we discuss about dataset and data preparation in this section.

A. SYSTEM MODEL

We consider RIS-aided downlink communications in a RIS-NOMA system as in Fig.1. As shown in Fig. 1, the system composes of a BS equipped with N_t antennas, a RIS equipped with L reflecting elements, and M UEs where each UE is equipped with N_r antennas. Given RIS elements are assumed to be far apart from each other, resulting in mutually uncorrelated elements and avoiding mutual signal coupling among them. We assume the direct link from BS to j^{th} user (UE_j), denoted by $BS \rightarrow UE_j$ where $j \in \{1, 2, ..., M\}$, is blocked due to signal attenuation in a harsh propagation environment. Hence, the received signal at each UE_j comes

from the cascaded channel $BS \rightarrow RIS \rightarrow UE_j$. In practice, a BS and RIS are fixed while UE_j is mobile. In this paper, it is assumed that UE_j is slowly moving further away from BS. We assume that all users move at the same speed v meter per one time step t. The speed is slow enough to avoid causing Doppler effects. This assumption is to ensure the communication performance of the system, e.g., transmission rate, and outage probability. A such RIS-NOMA wireless system is proposed for several applications such as connected autonomous vehicles [10], ultra-high-definition video, virtual reality, and internet of things applications such as wearable devices and smart homes [36]. Due to the blockage of the direct channel, communication from a BS to UE_j is only defined by the cascaded channel $BS \rightarrow RIS \rightarrow UE_j$. The cascaded channel from BS to each UE_j at time t is the combination of the channel $BS \rightarrow RIS$ and $RIS \rightarrow UE_j$, which is denoted as G_j [32] and is calculated as

$$G_j(t) = g_j^H(t)\mathbf{\Phi}(t)\mathbf{H}(t). \quad (1)$$

All channels in this paper are considered approximately constant at each channel coherence block. We only focus on one channel coherence block with a duration of T. Each time step or time instant $t = 1, 2, \ldots, T$. Due to a slow-moving scenario and perfect CSI knowledge at BS, the channel coherence time in this paper is on the order of tens of milliseconds (ms) [37], [38].

Next, we describe the signal model of the proposed RIS-NOMA system. In a NOMA-based system, the technique is applied at the transmitter (BS), which means that the BS transmits a superposed signal $x(t) = \sqrt{P_S}[c_1 x_1(t) + c_2 x_2(t) + ... + c_M x_M(t)]$ at a time t to all UEs [39]. In this formula, the symbol P_s is the transmit power at BS, $x_j(t) \in \mathbb{C}^{N_t \times 1}$ is a signal for user UE_j, and c_j is power allocation factor assigned to UE_j which satisfies $\sum_{j=1}^{M} c_j^2 = 1$ and $c_1 < c_2 < ... < c_M$. We assume that the channel of a UE_j with a higher index has a weaker channel gain and thus requires more power allocation. Based on the cascaded channel from Eq.(1), The received signal at UE_j at time t is expressed as

$$y_j(t) = PL_j(t)g_j^H(t)\mathbf{\Phi}(t)\mathbf{H}(t)\gamma \frac{x(t)}{\sqrt{P_s}} + n_j(t), \quad (2)$$

where $PL_j(t)$ is path loss parameter at UE_j, $g_j(t) \in \mathbb{C}^{L \times N_r}$ and $\mathbf{H}(t) \in \mathbb{C}^{L \times N_t}$ denotes the channel matrices of the links $RIS \rightarrow UE_j$ and $BS \rightarrow RIS$, respectively. We assume that all channels are Rayleigh fading channels. The symbol $n_j(t) \in \mathbb{C}^{N_r \times 1}$ presents the noise at each UE_j [40], [19]. In this RIS-NOMA network, the noise n_i at UE_j follows Gaussian distribution $\mathcal{CN} \sim (\mu_n, N_0)$. In Eq. (1), the factor $\gamma = P_s/N_0$ is the transmit SNR with $\mathbb{E}[|n_j(t)|^2] = N_0$. We normalize the noise $n_j(t)$ by a standard normal Gaussian distribution with $\mu_n = 0$ and $N_0 = 1$. Diagonal coefficient matrix $\mathbf{\Phi}$ of RIS has main diagonal including elements from $\mathbf{\Phi} = [\phi_1, \phi_2, ..., \phi_L]^T$. Each element in the main diagonal takes into account both phase shift and amplitude reflection

coefficient on the incident signal [19] as follows

$$\phi_m = \alpha_m e^{j\theta m}, \tag{3}$$

where $\theta_m \in [0, 2\pi]$ is the phase shift and α_m is the amplitude reflection coefficient.

The path loss of each UE PL_j, is determined according to distances and path loss exponent (PLE) as follows

$$PL_j(t) = \sqrt{(\frac{d_{sr}}{d_0})^{-\alpha_{sr}}}\sqrt{(\frac{d_{rj}(t)}{d_0})^{-\alpha_{rj}}}, \tag{4}$$

where symbol d_0 presents the reference distance; α_{sr} and d_{sr} are the PLE of the environment and the distance of the link $BS \rightarrow RIS$; and α_{rj} and $d_{rj}(t)$ are the PLE of the environment and the distance of the link $RIS \rightarrow UE_j$ [41], [42]. It can be seen from Eq. (2), Eq. (3), and Eq. (4) that the received signal at each user is affected by the distance between devices and the parameters of RIS. Different users are allocated different power factors according to their CSI. In this paper, CSI of all cascaded channels of the proposed RIS-NOMA system are assumed to be perfectly known at BS. Based on the knowledge of the CSI, the NOMA system can separate the signal of each user by applying SIC at the receiver side.

At the receiver side, the SIC technique is applied to decode the signal of a specific UE by leveraging the difference between channel gains from a transmitter to each user UE_j. After receiving the signal, each user will perform SIC according to using the principle of NOMA. The principle of NOMA allows every UE to receive signals of all UEs in the RIS-NOMA system, then each user obtains its desired signal by decoding unwanted signals. Because UE_M is allocated the highest power allocation, it will directly decode its signal x_M by treating the other signal $x_{M-1}, x_{M-2}, ..., x_1$ as interferences. The signal-to-interference-plus-noise ratio (SINR) to detect signal x_M at user UE_M is

$$SINR_M = \frac{\left|\widehat{h}_M\right|^2 P_s(c_M)^2}{\sum_{l=1}^{M-1} \left|\widehat{h}_l\right|^2 P_s(c_l)^2 + N_0}, \tag{5}$$

where $\widehat{h}_j = PL_j g_H^j \Phi H$ with $j = 1, 2, ..., M$. When decoding UE with index ρ^{th} where $\rho \neq M$ and $\rho > 1$, all the previous signals, i.e., $M^{th}, (M-1)^{th}, ..., (\rho+1)^{th}$, should be cancelled, and the remaining signals, i.e., $(\rho-1)^{th}, (\rho-2)^{th}, ..., 2^{nd}, 1^{st}$ are considered as inter-user interferences [43]. For simplicity of equations, we omit the term t. Then, a perfect SIC at user UE_ρ, the SINR at UE_ρ to decode signal x_ρ can be expressed as

$$SINR_\rho = \frac{\left|\widehat{h}_\rho\right|^2 P_s(c_\rho)^2}{\sum_{l=1}^{\rho-1} \left|\widehat{h}_l\right|^2 P_s(c_l)^2 + N_0}. \tag{6}$$

Finally, UE_1 will decode its own signal after cancelling all signals from previous users with index, i.e, $M^{th}, (M-1)^{th}, ..., 2^{nd}$ [44] and the SINR at UE_1 to decode signal x_1

is described as

$$SINR_\rho = \frac{\left|\widehat{h}_1\right|^2 P_s(c_1)^2}{N_0}. \tag{7}$$

To determine the SINR in Eq. (5), Eq. (6), and Eq. (7), it is required the knowledge of cascaded channel at each UE.

B. Problem formulation

In this RIS-NOMA network, a channel between $BS \rightarrow RIS$ is fixed, while channels $RIS \rightarrow UE_j$ are slow time-variant channels. To perform the SIC at the receiver side, it is necessary to know the cascaded channel gains from $BS \rightarrow RIS \rightarrow UE_j$ in advance. Being equipped with passive elements, the RIS only plays a role as a signal reflector, thereby we estimate the cascaded channels based on the received signals at UE_j. Due to the slow mobility of UE_j, the cascaded channels slowly change versus time, introducing a time-variant channel estimation problem for the RIS-NOMA.

Algorithm 1 Generate dataset for two UEs

Input: $x_1(t)$, $x_2(t)$, $d_j{}^{sr}$, $d_{rj}(t)$, α_{sj}, α_{sr}, ϕ_m, α_m, $\mathbf{H}(t)$
Output: $\mathbf{Y}$, $\mathbf{G}$
    Initialisation :
1: $\mathbf{Y}$, $\mathbf{G}$ as empty matrices
    LOOP Process
    Dataset length or the number of time steps K
2: **for** $t = 1$ to K **do**
3:     Generate $x_1(t)$, $x_2(t)$, $\mathbf{H}(t)$, $\mathbf{n}(t)$
4:     Initialize $\mathbf{g}_t$, $\mathbf{G}_t$, $\mathbf{\Phi}_t$, $\mathbf{y}_t$ as empty matrices
        $M = 2$ the number of UEs
5:     **for** $j = 1$ to M **do**
6:         Generate $g_j(t), \theta_i(t)$
7:         Using $\theta_j(t)$ and α_m, calculate $\phi_j(t)$ from (3)
8:         Using $\phi_j(t)$, calculate $\Phi_i(t)$
9:         Calculate $d_{ri}(t) = d_{ri}(t)$ - v
10:        Calculate received signal $y_j(t)$ at j^{th} UE from (2)
11:        Calculate cascaded channel at j^{th} UE
          $G_j(t) = g_j^H(t)\mathbf{\Phi}_j(t)\mathbf{H}(t)$ from (1)
12:        Concatenate $y_j(t)$, $G_j(t)$ to $\mathbf{y}_t$, $\mathbf{G}_t$
13:     **end for**
14:     Concatenate $\mathbf{y}_t$, $\mathbf{G}_t$ to $\mathbf{Y}$, $\mathbf{G}$
15: **end for**
16: **return** $\mathbf{Y}$, $\mathbf{G}$

C. DATA PREPARATION

We assume that the channel gains follow independent and identically standard complex Gaussian (normal) distributions $\mathcal{CN} \sim (\mu, \sigma^2)$ with the mean μ and variance σ^2 [45], [46], [23]. The dataset is generated with K samples as in the Algorithm 1. In a time-series problem, each sample presents a time step t in the RIS-NOMA system. We assume that at each time step, UE_j slowly moves further away from the BS with the same velocity. The dimension of matrices in the Algorithm 1 are denoted as $\mathbf{Y} \in \mathbb{C}^{K \times N_r N_t \times M}$,

$\mathbf{g} \in \mathbb{C}^{K \times L \times M}$, $\mathbf{H} \in \mathbb{C}^{K \times LN_t \times M}$, $\mathbf{\Phi} \in \mathbb{C}^{\mathbf{K} \times \mathbf{N_r N_t} \times \mathbf{M}}$, $\mathbf{n} \in \mathbb{C}^{K \times N_r N_t \times M}$, and $\mathbf{G} \in \mathbb{C}^{K \times N_r N_t \times M}$. It can be seen that all matrices include dimensions which are the number of time steps K and the number of UEs M. The remaining dimensions depend on the number of antennas at the transmitters and the receivers and the elements of RIS. It is noted that in this study, the transmit signal at UE_j which is denoted by x_j follows the distribution $\mathcal{CN} \sim (\mu_{xj}, \sigma_{xj}^2)$ with the mean and variance μ_{xj} and σ_{xj}, where $j = 1, 2, ..., M$. The output of the Algorithm 1, including received signal $\mathbf{Y}$ and the cascaded channel $\mathbf{G}$, are processed and used as the input and output of the data in the proposed DL model.

IV. PROPOSED DEEP LEARNING MODEL

In this section, we propose a CNN-LSTM based model for addressing the channel estimation problem in the RIS-NOMA system. We also provide a complexity analysis of the proposed model and discuss the performance metrics to evaluate the effectiveness of the proposed model.

A. CNN-LSTM MODEL

In this section, an architecture of the proposed CNN-LSTM network is first briefly illustrated. Next, the flow of data in the DL model is presented. Then, the training process of the proposed DL-based channel estimation for the proposed RIS-NOMA communication system in Fig. 1 is illustrated.

This paper proposes a variant of an encoder-decoder LSTM model that composes of two modules, including CNN module at the encoder side and LSTM module at the decoder side [47] as in Fig. 2, which is called a CNN-LSTM model. This kind of variant achieves the advantage of both CNN model and LSTM model. Compared to a conventional LSTM which merely contains LSTM layers, the proposed DL model exploits the power of CNN in terms of feature extraction. In this paper, the CNN has been utilized to learn spatial correlated features from an input sequence [48] of the received signals at UEs. Then, the features from CNN are fed into LSTM for decoding, and in turn extract the temporal correlation to predict the output, the cascaded channels, at one or many future time steps.

As shown in Fig. 2, the proposed DL model consists of an input layer, a CNN module, a LSTM module, and an output layer. Each component of the model is introduced as follows.

1) Data processing for input layer

We need to process the output data from Algorithm 1 in Section B before feeding them into the input layer of the CNN-LSTM model. The first output of the Algorithm 1 is a time-series sequence of the received signals at UE_j of the RIS-NOMA network, which is $\mathbf{Y} = \{\mathbf{Y}(1), \mathbf{Y}(2), ..., \mathbf{Y}(K)\}$. In particular, the sample at time step t is $\mathbf{Y}(t) = [y_1(t), y_2(t), .., y_M(t)]$ and $y_M(t) \in \mathbb{C}^{N_r \times N_t}$, where $t \in \{1, 2, ..., K\}$. The received signal at UE_j which

is denoted as $y_j(t)$ is a complex number, but a conventional DL model is designed for real values. Thereby, we extract two parts of a complex number including amplitude and angle, and then concatenate them as a dimension; resulting in $y_M(t) \in \mathbb{R}^{2 \times N_r \times N_t}$ and $\mathbf{Y} \in \mathbb{R}^{K \times M \times 2N_r N_t}$. Likewise, the second output of the Algorithm 1 is the cascaded channels from BS at UE_j of the RIS-NOMA is a K time steps sequence, which is denoted as $\mathbf{G} = \{\mathbf{G}(1), \mathbf{G}(2), ..., \mathbf{G}(K)\}$. At each time step t with $t \in \{1, 2, .., K\}$, the cascaded channels of the RIS-NOMA system consist of individual cascaded channels from BS to each UE_j, and is presented by $\mathbf{G}(t) = \{G_1(t), G_2(t), ..., G_M(t)\}$. The goal of the channel estimation is to predict the cascaded channel gains, thus we extract the amplitudes of cascaded channels as the target of the CNN-LSTM model. In this paper, the training length or symbols count of the channel estimation duration is considered as long enough to estimate all channel coefficients of the RIS-NOMA wireless system.

For a time-series problem, it is necessary to transform both the input sequence and output sequence as an input-output time-series sequence. The total K samples of input $\mathbf{Y}(t)$ and target $\mathbf{G}(t)$ is transformed into a time-series sequence based on the input time steps t_i and future time steps t_o. After this transformation, the length of the dataset will be changed and is denoted by **Sample**. In turn, the time-series sequence is divided into a training dataset and a testing dataset before going to the input layer.

2) CNN module

CNN is commonly used for image processing applications such as feature extraction and classification, where an image can be divided into three channels, including red, green, and blue. In this paper, the matrix of the received signal could be considered a three-dimensional (3D) image as the matrix includes three dimensions including the time step, the number of users, and the power values of the received signal [49]. Plus, the matrix of the received signals becomes larger when the number of RIS increases. Inspired by the analogy between a 3D image and the matrix of received signals, this paper uses CNN to extract the features of the received signals from users. The first layer of CNN is the input layer, and the last layer is the output layer, with a few hidden layers in between. The input of the CNN model has three dimensions height, width, and depth, which correspond to the number of UEs, the received signals at UEs, and the time steps in the proposed RIS-NOMA system. In most used CNN models, hidden layers compose of several convolution layers preceding pooling layers, through which feature space at the input is matched with the output. Each convolution layer contains several kernels (or filters) which define the depth of the layer. Kernels are known as feature detectors that connect the input and a set of weights by filters, and the convolution layer calculates the dot product between them. An activation function is applied at the output of each
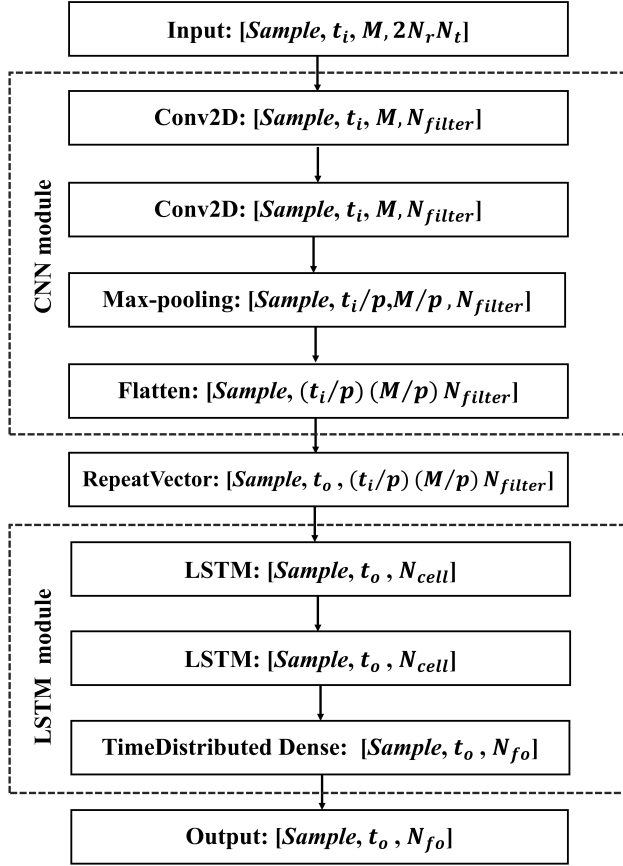
FIGURE 2: CNN-LSTM model

FIGURE 3: LSTM cell

connected kernels. Two 2D-conv layers apply the same hyperparameters and the number of filters is equal to N_{filter}. The extracted feature dimensions after convolutional layers are very high, thereby a max-pooling layer with a size $(p \times p)$ is added to minimize the size of extracted features and further extract the most important features as well as obtain more useful information. In this way, CNN module can capture the features of the data. Finally, a flatten layer is applied to flatten all features into a single vector before feeding into the LSTM.

convolution layer to generate a feature map. This feature map serves as the input of the next layer, for example, the next convolution layer, or pooling layer. Following the convolution layers, pooling (sub-sampling) layers are used to reduce the number of CNN parameters by performing down-sampling for each feature map. Finally, the output of the last pooling layer is fed into a fully connected layer. The input of the fully connected layer is flattened into the form of a vector. The output of the fully connected layer is the final output of CNN. The main benefit of using CNN is the weights sharing feature, where weights and bias are the same for all neutrons in each hidden layer, reducing the trainable parameters that enhance the generalization and avoid overfitting. Another advantage is the ability to extract the inference characteristics of features, however, CNN needs huge training data to learn all parameters [50], [51].

The architecture of the proposed model is described in detail in Fig. 2. Firstly, the sequential input is fed into the CNN module as in Fig. 2. The CNN module includes two 2-D convolutional (2D-conv) layers, one max-pooling layer and one flatten layer. A convolutional layer is a feature map that extracts the features of the input data by using a convolution operation with filters, which is a set of locally
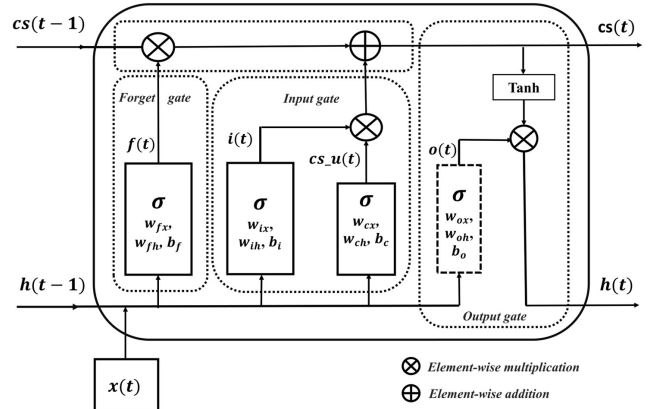
3) LSTM module

LSTM is considered a useful approach to capture long temporal features such as sequences like in natural language processing. Motivated by this, it is feasible to apply the LSTM module when coping with time-series sequences of received signals and channel coefficients. However, LSTM properly does not capture well other features such as spatial features [51]. This is the reason for considering the combination of the LSTM with the CNN module to capture not only time dependencies but also spatial features like the number of UEs and the received signal at each UE, for the proposed DL-based channel estimation method. To match the output of CNN as the input layer of LSTM, a RepeatVector layer is applied to repeat t_o times to make sure that the output contains the expected number of output time steps. The LSTM module consists of one or several LSTM layers and each LSTM layer is composed of LSTM cells. LSTM cell or LSTM internal memory cell is a key element of the LSTM module that establishes the temporal connections of a LSTM layer.

The topology of a LSTM cell is illustrated in Fig. 3 which consists of two different recurrent features, including hidden sate h and cell state cs. A LSTM unit uses a previous state, including previous hidden state and previous cell state ($h(t-1), cs(t-1)$), and a current input $x(t)$ to update the current

time step $(h(t), cs(t))$ which can be expressed as

$$(h(t), cs(t)) = \mathcal{L}(h(t-1), cs(t-1), x(t)), \qquad (8)$$

where $\mathcal{L}$ is a function which is capable of mapping input sequence and output sequence at any length. A basic LSTM unit contains three gates, including input gate, forget gate, and output gate to control the cell state. In particular, the input sequence $x(t)$ is fed into a LSTM cell at each time step t. In each cell, the input $x(t)$ and the hidden state of the current time step $h(t-1)$ are fed into the three gates (functions) before going through a sigmoid activation function σ. The update of a LSTM cell in terms of $(h(t), cs(t))$ at time step t is expressed as follows:

$$
\begin{aligned}
f(t) &= \sigma(x(t)w_{x,f} + h(t-1)w_{h,f} + b_f), \\
i(t) &= \sigma(x(t)w_{x,i} + h(t-1)w_{h,i} + b_i), \\
o(t) &= \sigma(x(t)w_{x,o} + h(t-1)w_{h,o} + b_o), \qquad (9)
\end{aligned}
$$

where $w_{x,f}$, $w_{x,i}$, $w_{x,o}$, $w_{h,f}$, $w_{h,i}$, $w_{h,o}$, $w_{x,c}$ and $w_{x,c}$ are weight parameters; and b_f, b_i, b_o and b_c are bias parameters in a LSTM cell. Both weights and biases are updated via the training process. Three gates, including input, forget and output, are presented by symbols i, f, and o. Forget gate $f(t)$ decides how much information in the previous state cell state $cs(t-1)$ should be retained. Input gate $i(t)$ determines how much new information is taken into account via cell update $cs_u(t)$. The computation for $cs_u(t)$ is similar to the three gates above, but uses Tanh activation function with a value range $(-1, 1)$ as

$$cs_u(t) = Tanh(x(t)w_{x,cs} + w_{h,cs}h(t-1) + b_c). \quad (10)$$

Output gate $o(t)$ controls how much content in the cell state $cs(t)$ should leave the cell to arrive the hidden state $h(t)$. The cell state at time t is presented as

$$cs(t) = f(t) \odot cs(t-1) + i(t) \odot cs_u(t). \qquad (11)$$

By using Tanh activation function, the hidden state of the current step $h(t)$ of a LSTM cell is always in interval $(-1, 1)$ and is expressed as

$$h(t) = o(t) \odot Tanh(cs(t)). \qquad (12)$$

The cell state in Eq. (11) and the hidden state in Eq. (12) is determined based on the input gate, output gate, and forget gate from Eq. (9), along with cell update from Eq. (10). A LSTM cell is flexible because it can obtain an input time-series sequence at any length, e.g., $x(t-1)$ and produce an output sequence at any length, e.g., $x(t+k)$ as in Fig. 4.

4) Output layer

The output of the last LSTM layer is fed into the input of a TimeDistributed layer which wraps the data for each time step of the sequential data. The number of the output of the CNN-LSTM is N_{fo}. The output is amplitude of the cascaded channel of the RIS-NOMA system at time step $(t_i + t_o)$. Based on the predicted cascaded channel at the output layer and the target cascaded channel in the RIS-NOMA system,
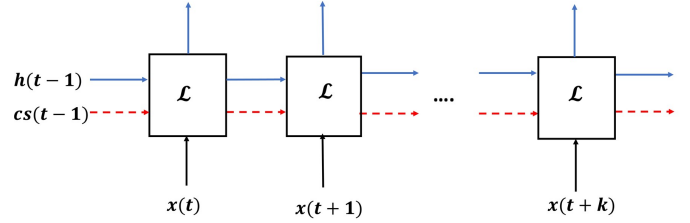

FIGURE 4: A LSTM cell at any length

the parameters of the CNN-LSTM model will be updated via an optimization algorithm. The objective of the optimization problem is to minimize the mean square error (MSE) loss function which is defined as

$$MSE = \frac{1}{N}\sum_{t=1}^{N}\left[\mathbf{G}(t_i + t_o) - \hat{\mathbf{G}}(t_i + t_o)\right]^2, \qquad (13)$$

where N is the total number of samples and $\mathbf{G}(t_i + t_o)$ and $\hat{\mathbf{G}}(t_i + t_o)$ are the amplitude of target cascade channels and predicted cascaded channels from BS to two UEs. The flow of data through each layer of the CNN-LSTM model is presented in Fig. 2.

B. Complexity analysis

Based on the proposed CNN-LSTM in Fig. 2, we determine the computational complexity by calculating the number of FLOPs (floating operations). The reason is the proposed CNN-LSTM using a backpropagation algorithm during the training stage, which normally works with matrices. To calculate the complexity of CNN-LSTM model at each time step, we are required to calculate the time complexity of CNN module and LSTM module separately [52], [53]. In addition, the computational complexity of each module comes from those in each layer. The complexity of CNN module mainly comes from those in convolutional layers. The max-pooling layer contributes marginally to the total computational complexity of CNN module. Flatten layer does not require the computational complexity. Therefore, the complexity of CNN module is calculated as $\mathcal{O}(\sum_{l=1}^{d} n_{l-1}s_w s_h n_l o_w o_h)$, where l is the index of a convolutional layer, d is the number of convolutional layers, n_l is the number of filters or the width in l^{th} layer, s_w and s_h are the width and the height of the filter respectively, and o_w and o_h are the width and height of the output feature map correspondingly [54]. Regarding LSTM module, the complexity of LSTM per time step comes from the computational complexity of input gate, output gate, forget gate and cell update in Eq.9 and Eq.10 which is calculated as $\mathcal{O}(4(\mathcal{D}_i\mathcal{D}_o + \mathcal{D}_o^2 + \mathcal{D}_o))$ where $\mathcal{D}_i$ is the input dimension and $\mathcal{D}_o$ is the output dimension. The LSTM updates its complexity according to time step [55]. At last, the computational of the dense layer is determined by $O(\mathcal{D}_o n_d)$ [56] where n_d is the number of neurons in dense layer. In sum, the computational complexity of the CNN-LSTM

will iterate via S time steps and I iterations is calculated as $\mathcal{O}(SI((\sum_{l=1}^{d} n_{l-1} s_w s_h n_l o_w o_h) + (4(\mathcal{D}_i \mathcal{D}_o + \mathcal{D}_o^2 + \mathcal{D}_o) + O(\mathcal{D}_o n_d)))$.

C. PERFORMANCE METRICS

The dataset of the paper is generated by Algorithm 1, and divided into training and testing dataset. The performance of the proposed CNN-LSTM model is evaluated using performance metrics on test dataset such as normalized root mean square error (NRMSE), mean absolute scaled error (MASE), mean absolute percentage error (MAPE), and R-squared score (R2 score). In this paper, we evaluate the DL model on different datasets generated from different RIS-NOMA system parameters, in which each dataset has its scale. Due to independent scales between different datasets, scale-free error metrics are applied instead of scale-dependent error metrics [57]. The MSE and mean absolute error (MAE) is good to compare the performance of different DL models on a dataset or same scale datasets. The MSE is defined as in (13) and the MAE is defined as follows:

$$MAE = \frac{1}{N} \sum_{t=1}^{N} |y_t - \hat{y}_t|. \qquad (14)$$

However, two metrics MSE and MAE are no longer feasible when comparing performance accuracy of a model for different datasets [58]. To evaluate the performance of a DL model against datasets with different scales, sale-free metrics should be applied. Since the generated datasets in this paper have different scales by adjusting the parameters of the RIS-NOMA system, we apply scale-free metrics including NRMSE and MASE instead of MSE in Eq. (13) and MAE in Eq. (14). The metric NRMSE is defined as:

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}}, \qquad (15)$$

where $RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2}$; $\hat{y}_t$ and y_t are the predicted channel gain and target channel gain at a given time t respectively; and y_{max} and y_{min} are maximum and minimum values of the target channel gains, respectively. The NRMSE measures the normalization of the deviation between the observed value as well as the actual value [59]. The metric MASE is defined as [57]:

$$MASE = \frac{N-1}{N} \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|}{\sum_{t=2}^{n} |y_t - y_{t-1}|}. \qquad (16)$$

The percentage error MAPE is defined as

$$MAPE = \frac{100\%}{N} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right|. \qquad (17)$$

Besides the metrics that evaluate the behavior of the difference between the predicted and the actual data, the decision coefficient R2 score which is known as a goodness of fit to evaluate the level of fit of a model [60]. The metric R2 score

is is expressed as

$$R_2 = 1 - \frac{SSE}{SST}, \qquad (18)$$

where

$$SSE = \frac{1}{N} \sum_{t=1}^{n} (y_t - \hat{y}_t)^2,$$

$$SST = \frac{1}{N} \sum_{t=1}^{t} (y_t - \mu_y)^2,$$

in which μ_y is the standard deviation of y.

This study applies NRMSE and R2 score in Eq. (15) and Eq. (18) to evaluate the cascaded channels prediction accuracy and the goodness of fit of the proposed CNN-LSTM respectively. This is because the root mean square error (RMSE) is a standard error which is extremely sensitive to extraordinarily large or small errors in the dataset [59]. Two metrics MASE and MAPE in Eq. (16) and Eq. (17) are added to evaluate the prediction performance of the proposed DL model in several scenarios of the RIS-NOMA system.

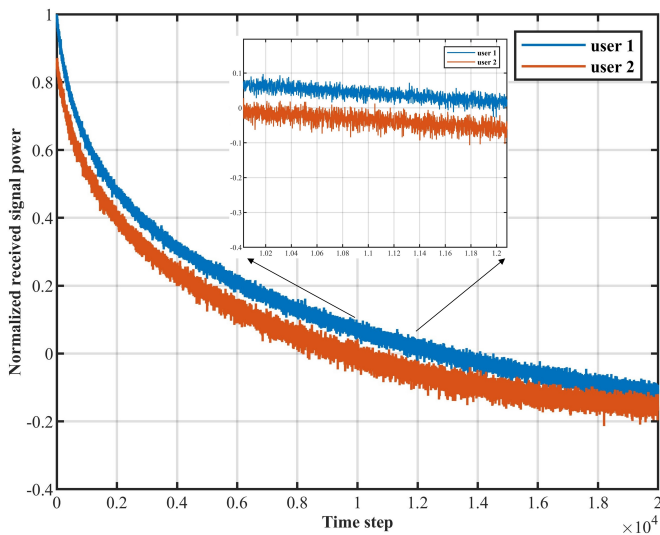## V. SIMULATION SETUP AND RESULTS

This section provides simulation results to verify the efficiency of the proposed CNN-LSTM model for different RIS-NOMA systems.

A. SIMULATION SETUP

The definition of 6G is evolving, there is no standard for system-level parameters in a specific setting or use-case in 6G so far [61]. Therefore, in this paper, we introduce a possible parameter setting for the potential RIS-NOMA system model in 6G. In this simulation, we model the RIS-NOMA system for $M = 2$ UE, and each user is equipped with a single antenna. We set the distance from $BS \rightarrow RIS$, $RIS \rightarrow UE_1$, and $RIS \rightarrow UE_2$ are $150\,\text{m}$, $30\,\text{m}$, and $40\,\text{m}$, respectively. The reference distance is $d_0 = 20\text{m}$. Both UE move at $v = 0.1$ meter per time step t to ensure the signal strength of two UE are not too low after K time steps moving further from BS the received signal at each user. It is assumed that the distribution of $x(t)$ follows $\mathcal{CN} \sim (1, 0.1)$ and that of $BS \rightarrow RIS$ channel follows $\mathcal{CN} \sim (1, 0.1)$. The distribution of links from $RIS \rightarrow UE_j$ are $\mathcal{CN} \sim (4, 1)$ and $\mathcal{CN} \sim (3, 1)$, correspondingly. In line with significant research progress on the development of loss-less metasurfaces [62], [63], this paper uses an ideal phase shift model in which the reflecting magnitude is considered as 1 without energy loss. We set up value of $\alpha_m = 1$ in (3) and $\alpha_{sr} = \alpha_{ri} = 2.2$ in (4). This is a time-series problem, so it is very sensitive to the change of the dataset, therefore, in this paper, we fix phase shift by generating fixed phase shifts in $[0.01\pi, 0.02\pi]$. We will observe the RIS-NOMA system versus the variation of system parameters including transmit SNR, the power allocation factor and the number of RIS elements. The details are illustrated in Table 1. Figure 5 shows the normalized (with respect to user 1) received signal
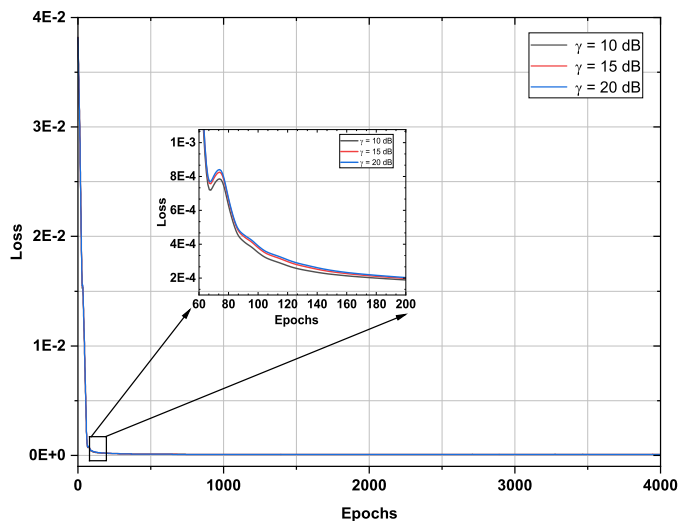
TABLE 1: Parameters setup.

Parameters	Names	Values
K	Number of time steps	20000
M	Number of users	2
v	Uniform speed per one time step	0.1 m
t_i	Number of input time steps	20
t_o	Number of output time steps	1
L	Number of RIS elements	20
γ	Transmit SNR	10 : 2.5 : 30 dB
d_0	Reference distance	20 m
d_{sr}	Distance from BS to RIS	150 m
d_{s1}	Distance from RIS to UE_1	30 m
d_{s2}	Distance from RIS to UE_2	40 m
α_m	Amplitude reflection coefficient	1
ϕ_m	Phase shift	$[0.01\pi : 0.02\pi]$
$\alpha_{ri},$	PLE from RIS to each UE	2.2
α_{sr}	PLE from BS to RIS	2.2
N_t	Number of antennas at BS	1
N_r	Number of antennas at each UE	1

FIGURE 6: Loss curve in case of the number of RIS elements $L = 20$ and power allocation factor $c = 0.25$ with different values of transmit SNR.

FIGURE 5: Magnitude of normalized received signals at two users.

power for both users over time in the case $\gamma = 20$, $L = 20$, and $c = 0.3$.

The **Sample** data points of the input-output time-series sequence will be divided into $0.3 : 0.7$ for training data and testing data respectively. The input applies max-min normalization and the target is scaled in the range of $[0, 1]$ before feeding into the DL model.

At the encoder side, we use the kernel size of convolution 2D (Conv2D) layer with $(3, 1)$ and apply a same padding for each Conv2D layer. A max-pooling layer with $p = 2$ and the size is 2×2 is applied. The CNN-LSTM model is tuned with $N_{filter} \in \{4, 8, 16, 64\}$. At the decoder side, each

LSTM layer consists of $N_{cell} = 16$ unit and at least one of two LSTM layers apply Tanh activation function. The output is finally fed into TimeDistributed Dense layer with the number of features at output layer $N_{fo} = 2$, which is for two cascaded channel gains of two users, before reaching the output layer. Adam optimizer is used to optimize the loss function in (13).

B. LOSS CURVE ON TRAINING DATA

The MSE loss function in (13) is used to train the proposed CNN-LSTM model, which has been effective in calculating the back-propagation of loss values. The advantage of MSE is that it ensures the model has a few outliers prediction with large errors. The MSE loss over 4000 epochs on the normalized training dataset in case $\gamma \in \{10, 15, 20\}$, $L = 20$ and $c = 0.25$ is illustrated in Fig. 6. In three scenarios, the MSE loss curves perform the same curve behavior. As shown in Fig. 6, the loss converges very well on the normalized training data. It can be seen that the loss reduces to approximately $2E-4$ for the first 200 epochs and reaches a factor of $1E-5$ at 4000 epochs. A large number of training epochs is to ensure that the proposed DL model provides a good fit to the dataset which obtains a high R2 score.

C. CONVERGENCE OF CNN-LSTM

Fig. 7 depicts the NRMSE on at each UE and the average NRMSE against the number of training epochs on test dataset in original scale in case $\gamma = 20$ dB, $c = 0.25$ and $L = 20$. We can find that NRMSE decrease when the number of epochs increases. The NRMSE on the test dataset achieve a steady state at 4000 epochs at observed scenarios, which shows the robustness of the proposed DL model. The R2 score at each user and average R2 score gradually increase
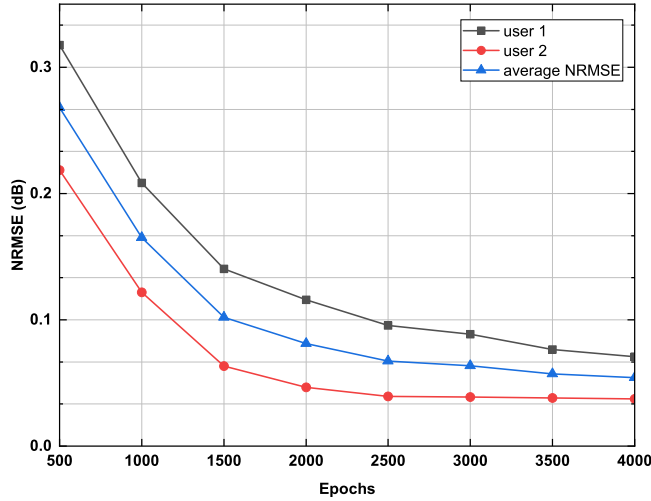
FIGURE 7: The NRMSE performance of the cascaded channels against number of epochs with transmit SNR at 15 dB and power allocation factor $c = 0.15$.
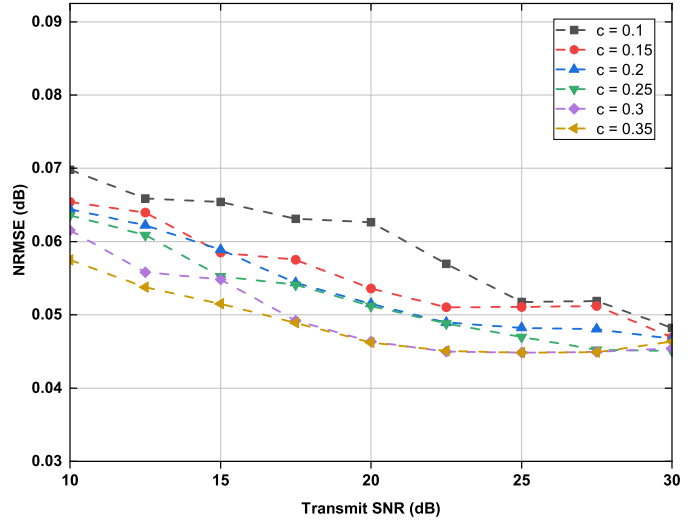
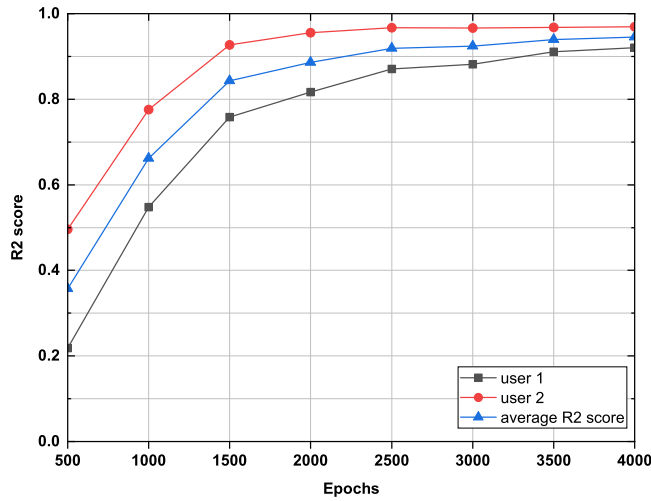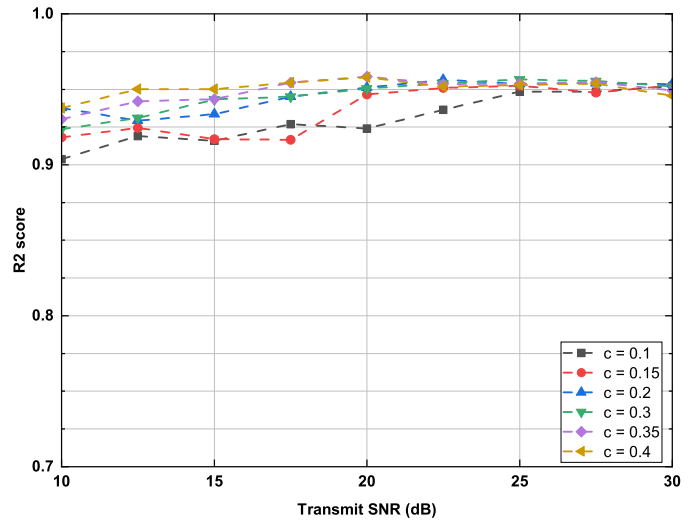**FIGURE 9:** The average NRMSE performance of cascaded channels according to transmit SNR in cases of the number of RIS elements $L = 20$ and different values of power allocation (c).

FIGURE 8: The R2 score performance of the cascaded channels against number of epochs with transmit SNR at 15 dB and power allocation factor $c = 0.15$.
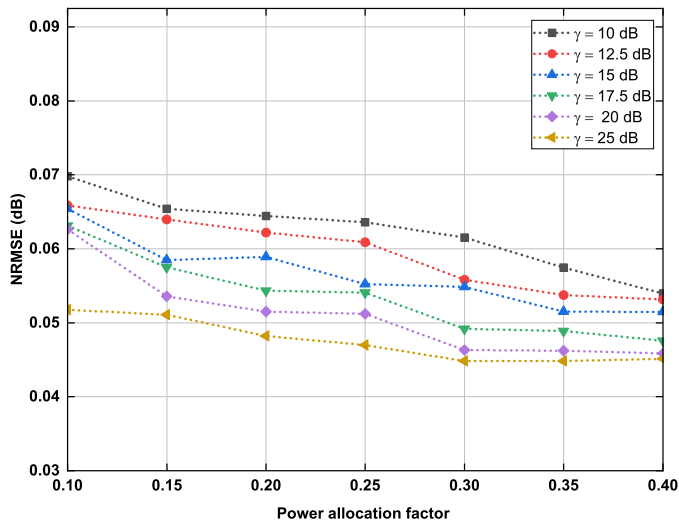

FIGURE 10: The average R2 score performance of the cascaded channels of the system with respect to transmit SNR in cases of the number of RIS elements $L = 20$ and different values of power allocation factor (c).

with respect to the number of epochs. At epoch 4000, the three values of R2 score are above 0.9.

D. TRANSMIT SNR

Fig. 9 and Fig. 10 study the average NRMSE performance and the average R2 score of the cascaded channel against the transmit SNR. In this simulation, we set the value of power allocation factor $c \in \{0.1 : 0.05 : 0.4\}$. Fig. 9 shows that the NRMSE decrease with the increase in transmit SNR of the system as a larger SNR contributes to better channel environments for both UEs. An increase in a transmit SNR can avoid a very low amplitude of the received signal at each user, which can degrade the channel estimation performance.

A lower SNR leads to lower signal power or lower amplitude at each user. A low amplitude signal at each user will become a very low value after normalization at the input of the CNN-LSTM model. If the normalized input of the CNN-LSTM is too low, it can be treated as noise, which will degrade the prediction performance of the CNN-LSTM. As shown in Fig. 9 the NRMSE significantly reduces from $\gamma = 10$ dB to $\gamma = 22.5$ dB and gradually drops in very high SNR region from $\gamma = 22.5$ dB to $\gamma = 30$ dB. Especially, in the case of very high transmit SNR from 25 dB to 30 dB, the NRMSE slightly increases when $c = 0.35$. It could be due to differences in the received signal between two users which is

FIGURE 11: The average NRMSE performance of the cascaded channels of the system with respect to power allocation factor in cases of the number of RIS elements $L = 20$ and different values of transmit SNR.
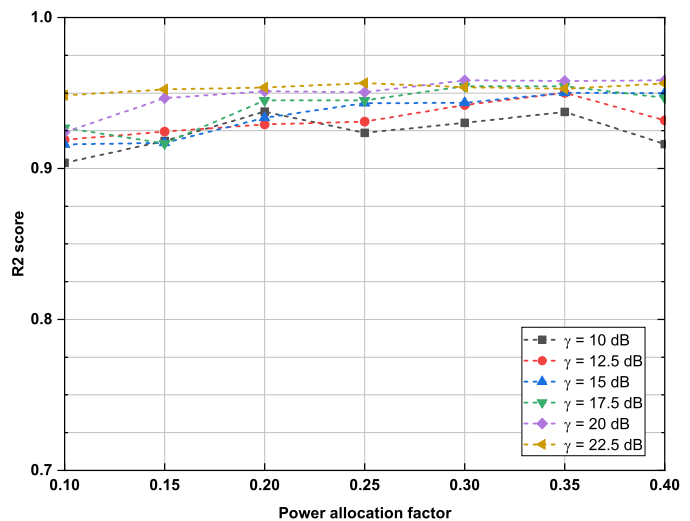
FIGURE 12: The average R2 performance of cascaded channel for the system against power allocation factor with the number of RIS elements $L = 20$ and different values of transmit SNR (γ).

significant, resulting in a huge difference between the scales of input features in the CNN-LSTM DL model. This is the reason for a slight reduction of the performance of the DL model in case $c = 0.35$ with a transmit SNR is 25 dB or 30 dB as in Fig. 9.

Fig. 10 plots the average R2 score curves vs. transmit SNR to evaluate the degree of fit [59] of the DL model. In the other words, the R2 score curve shows how well the predicted cascaded channels catch up with the trend of the target cascaded channels in the RIS-NOMA system. As can be seen from Fig. 10, the prediction performance in terms of average R2 score is over 0.9 overall. This means that the level of a good fit of the proposed CNN-LSTM model is not sensitive to the transmit SNR parameter of the RIS-NOMA system.

E. POWER ALLOCATION COEFFICIENT

In Fig. 11, average NRMSE performance is observed according to power allocation factors which are in $\{0.1 : 0.05 : 0.4\}$. It can be seen that the NRMSE decreases with an increased power allocation factor, which are in line with the results in Fig. 9. In Fig. 12, we can find that the average R2 score performance of the cascaded channel is higher than 0.9 overall with different power allocation factors. The proposed DL model shows its robustness to the change of the power allocation factor, the most important parameter that defines NOMA system. The results in Fig. 12 are in line with those in Fig. 10.

In this paper, the average NRMSE and average R2 score are selected to evaluate the prediction accuracy and the good fit of the proposed CNN-LSTM for cascaded channels of the RIS-NOMA system. We also consider other two

TABLE 2: Four performance metrics of CNN-LSTM models

c	γ	MASE	MAPE
0.2	15dB	1.8	10.45
	20dB	1.48	8.68
	25dB	1.36	7.97
0.25	15dB	1.74	10.13
	20dB	1.48	8.61
	25dB	1.29	7.65
0.3	15dB	1.62	9.33
	20dB	1.30	7.63
	25dB	1.20	7.20

metrics, including the average MASE and average MAPE in several cases to evaluate their behavior along with average NRMSE and average R2 score. The detail of the average MASE and average MAPE in several scenarios, in which $c \in \{0.2, 0.25, 0.3\}$, transmit SNR $\gamma \in \{15, 20, 25\}$ dB and $L = 20$, are presented in Table 2. As shown in Table 2, the behavior of average MASE and average MAPE are similar to that of average NRMSE in the observed scenarios. This means that average MASE and average MAPE decreases when power allocation factor c increases or when transmit SNR (γ) increases.

F. NUMBER OF RIS ELEMENTS

Fig. 13 studies the impact of the number of RIS elements on the average NRMSE performance. As seen in (3), the number of RIS elements affects the angle of the received signal at each UE, which in turn has an impact on the input features of the CNN-LSTM model. Therefore, a change in the number of RIS elements influences the prediction performance of the proposed DL for the cascaded channels.
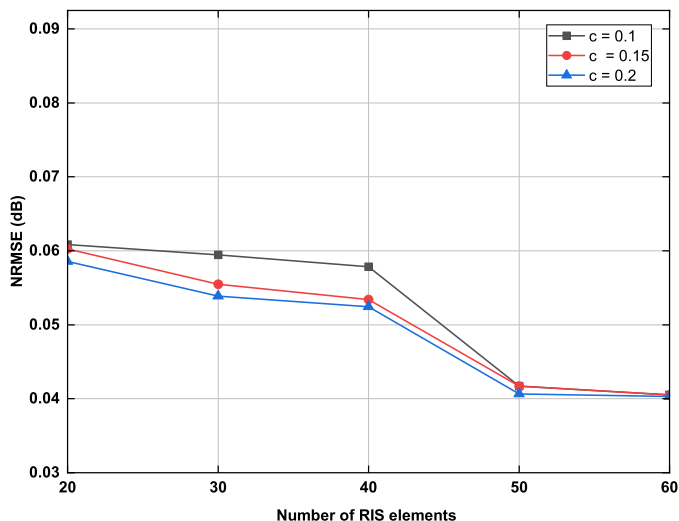
FIGURE 13: The average NRMSE performance of the cascaded channel for the system against the number of RIS elements in cases of transmit SNR at 15 dB and different values of power allocation factor (c).

FIGURE 15: Prediction on test dataset at user 1 (near user).

FIGURE 14: The average R2 score performance of the cascaded channel for two users against number of RIS elements of the system in cases of transmit SNR at 15 dB and different values of power allocation factor (c).
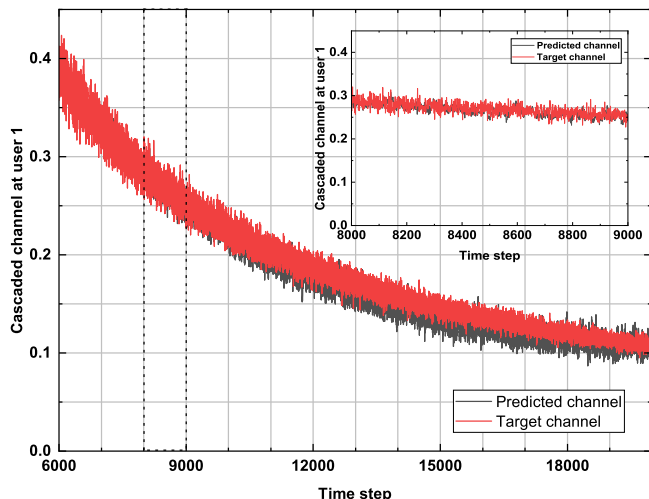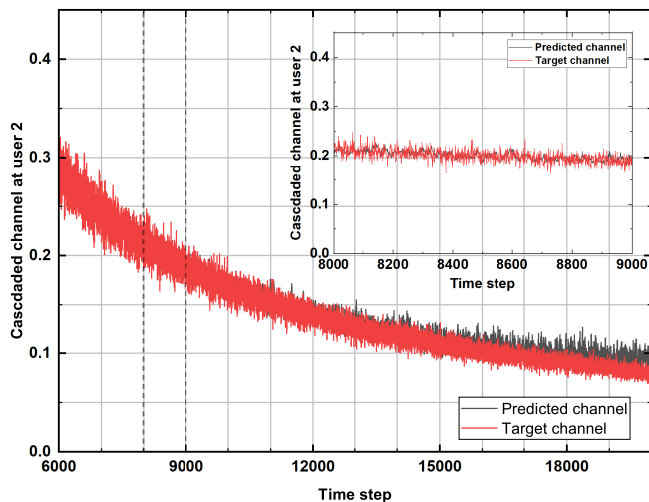

FIGURE 16: Prediction on test dataset at user 2 (far user).

trend of the cascaded channel of the RIS-NOMA system. We can see that the results in Fig. 13 and Fig. 14 are aligned with those in Section D and Section E.

G. EVALUATION OF CNN-LSTM

Fig. 15 and Fig. 16 show the predicted cascaded channels and the target (actual) cascaded channels on test dataset for each UE of the RIS-NOMA by applying the proposed CNN-LSTM. In this setting, we set $\gamma = 20$, $L = 20$ and $c = 0.3$. The figures prove that the proposed CNN-LSTM has a good prediction performance at both UE; which means that the predicted cascaded channel at each UE catches up very well with its target cascaded channel. The adopted channel estimation based CNN-LSTM model in the case of $N_{filter} = 64$ is illustrated in Fig. 17 in case the NOMA-RIS parameters system includes $\gamma = 10$ dB, $L = 20$, and $c = 0.3$.

Here, we set $c \in \{0.1, 0.15, 0.2\}$ and $\gamma = 15$. In Fig. 13, we can observe that the prediction accuracy in terms of average NRMSE improves with an increase in the number of RIS elements. The higher the number of RIS elements, the stronger reflected signals from RIS can reach each user. It becomes evident from Fig. 13 that the important role of increasing RIS elements in increasing the prediction accuracy for RIS-NOMA by applying the proposed CNN-LSTM model. Observe from Fig. 14 that the average R2 scores versus the number of RIS elements are above 0.92, which indicates that the CNN-LSTM is able to follow the
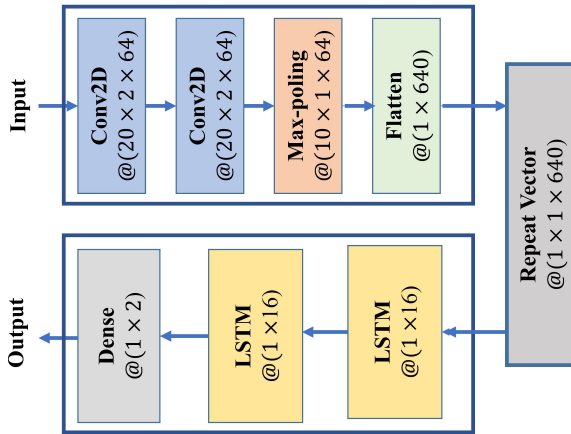
FIGURE 17: The adopted channel estimator based on CNN-LSTM model in the case of $\gamma = 20$, $L = 20$ and $c = 0.3$

H. COMPARISON OF THE PROPOSED DL MODEL WITH OTHER DL MODELS

To demonstrate the robustness of CNN-LSTM based channel estimation for the NOMA-RIS system, they are compared against previous study benchmarks, which are illustrated in Table. 3. In Table 3, we describe scenarios, multiple access techniques and the DL-based channel estimation methods and performance metrics. We select the benchmark channel estimation method that is based on/modified from CNN or LSTM architecture, as it allows for easier comparisons with our proposed CNN-LSTM architecture that we proposed. It can be seen from Table 3 recent studies generally apply orthogonal frequency-division multiple access (OFDMA) method, which is different from the NOMA approach as presented in this paper. Essentially, there are many differences between an OFDMA-based system and a NOMA-based system, such as variations in system design and parameters. As a result, there are different approaches to resource allocation among users, as well as different strategies for utilizing the capabilities of RIS to improve the signal quality.

Moreover, normalized mean square error (NMSE) is a commonly used metric in prior studies. However, our paper employs NRMSE, which is the square root of NMSE, for the sake of easy visualization and consistency with the dataset's scale. Results in [64], [65] indicate that certain studies on channel estimation, including those using machine learning, have employed NRMSE as a metric. Nonetheless, in cases where the NRMSE values are very similar between two scenarios, i.e., two sets of parameters in RIS-NOMA system, additional metrics are necessary to determine which scenario performs better. To make the evaluation more comprehensive, our paper incorporates two other metrics, namely MASE and MAPE. By using NRMSE, MASE, and MAPE, we only examine the extent of prediction errors in each scenario. On the other hand, the R2 score is a crucial metric that evaluates how well the predicted channel aligns with the pattern of the target channel.

To evaluate the effectiveness of the proposed CNN-LSTM model, this paper compares the prediction performance of four DL models with four other DL models including CNN1D-LSTM, CNN1D-BiLSTM, CNN and LSTM. It is noted that we use symbol CNN1D- and CNN2D- to distinguish between models using conv1D and conv2D while making a comparison. Among them, CNN1D-LSTM uses conv1D layer while the proposed CNN2D-LSTM applies conv2D layer at their CNN modules. The difference between conv1D layer and conv2D layer is the dimension of input data. The input data of conv1D is two-dimensional data while of conv2D is three-dimensional data. The conv1D combines the last two features of the input data of conv2D as one feature, in which the dimension of input sequence of conv1D becomes $[\boldsymbol{Sample}, t_i, \mathbf{2} * \mathbf{M} * \mathbf{N_r} * \mathbf{N_t}]$. For performance comparison, we also use CNN1D-BiLSTM [67], which combines CNN1D and Bi-LSTM. In [67], CNN1D-BiLSTM is proposed for CSI estimation in an OFDMA assisted wireless system without using RIS in a high-speed mobile scenario. Furthermore, we apply the CNN model which uses conv2D [33], [66] and LSTM model [68] for the comparison. We use similar hyperparameters set up for CNN model and LSTM model as those in the CNN module and the LSTM module of the proposed CNN2D-LSTM, respectively. To analyze the prediction performance of the four DL models, we set the parameters of the RIS-NOMA with $L = 20$, $\gamma \in \{10, 15, 20\}$ dB, and $c = 0.3$.

In order to evaluate the behaviour of different models on a dataset, we use four scale-dependent metrics which consist of average RMSE, MAE, MAPE and R2 score as in Table 4. Table 4 shows the accuracy superiority in terms of four metrics of the proposed CNN2D-LSTM model over four remaining models. In case $\gamma = 10$ dB, the prediction accuracy in terms of RMSE and MAE values are similar for the CNN2D-LSTM model and CNN model. However, CNN2D-LSTM model is superior in terms of MAPE and R2 score compared to CNN model. Table 4 shows that using the conv2D layer outperforms using the conv1D layer regardless conv1D is combined with LSTM or Bi-LSTM. For example, e.g, the $RMSE = 0.0142$, $RMSE = 0.0391$, and $RMSE = 0.2504$ for CNN2D-LSTM, CNN1D-LSTM, and CNN1D-BiLSTM, respectively. CNN model shows a relatively good performance in terms of four performance metrics in case of $\gamma = 10$ dB and $\gamma = 15$ dB. Although in these two cases, its prediction accuracy is better than that of CNN1D-LSTM, CNN model fails to follow the trend of the cascaded channel of the RIS-NOMA system with $R2 = -4.2541$ in the case of $\gamma = 20$ dB. Clearly, CNN is sensitive to the change of the transmit SNR. Although CNN has the ability of feature extraction, its performance for time-series sequences of the cascaded channel is not robust to the change of the transmit power parameter in the RIS-NOMA. Different from the three remaining models, LSTM models failed to predict the cascaded channels of the RIS-NOMA system for all observed scenarios with very high values at

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited ar
content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2023.3278232

Chi Nguyen : Channel Estimation Using CNN-LSTM in RIS-NOMA Assisted 6G Network:

TABLE 3: Benchmarks to compare with the proposed CNN-LSTM channel estimation methods.

Ref	Scenarios	Multiple access technique	RIS included in the system	Proposed channel estimation method	Performance metrics
[33]	RIS-assisted integrated sensing and communication system	OFDMA	Yes	CNN	NMSE
[66]	Large intelligent surfaces aided Mm-wave massive multiple-input-multiple-output system	OFDMA	Yes	CNN	NMSE
[67]	High-speed mobile communication	OFDMA	No	CNN1D-BiLSTM	NMSE
[68]	IEEE 802.11 p Standard	NOMA	No	LSTM	NMSE
This study	RIS-aided NOMA system in slow moving environment	**NOMA**	Yes	**CNN2D-LSTM**	**NRMSE, MASE, MAPE, R2**

TABLE 4: Performance metrics of different DL models.

Metrics	Models	10dB	15dB	20dB
RMSE	CNN2D-LSTM	**0.0186**	**0.0142**	**0.0136**
	CNN1D-LSTM	0.0271	0.0391	0.0369
	CNN1D-BiLSTM	0.0968	0.2504	0.5667
	CNN	0.0186	0.0166	0.5667
	LSTM	0.9737	0.9789	0.9813
R2	CNN2D-LSTM	**0.9302**	**0.9436**	**0.9584**
	CNN1D-LSTM	0.8427	0.7393	0.7654
	CNN1D-BiLSTM	-1.0348	0.2068	-0.3596
	CNN	0.9159	0.9338	-4.2541
	LSTM	-951	-1096	-1395
MAE	CNN2D-LSTM	**0.0163**	**0.0142**	**0.0114**
	CNN1D-LSTM	0.0235	0.0342	0.0320
	CNN1D-BiLSTM	0.0698	0.0545	0.0530
	CNN	0.0163	0.0142	0.5322
	LSTM	0.9729	0.9781	0.9804
MAPE	CNN2D-LSTM	**5.8969**	**9.3270**	**7.6342**
	CNN1D-LSTM	18.4884	27.5424	25.6169
	CNN1D-BiLSTM	123.81	460.85	258.57
	CNN	15.5540	13.5168	60.8110
	LSTM	658	662	664

remaining models in the cascaded channel estimation problem in the RIS-NOMA system. This is because the proposed model takes advantage of the feature extraction characteristic of CNN model and the time-series prediction model LSTM model.

VI. CONCLUSION AND FUTURE WORKS

In this paper, we proposed a CNN-LSTM model to predict the cascaded channel of the RIS-NOMA network in which both users are slowly moving further from the BS. The CNN-LSTM shows its robustness to the variation of the RIS-NOMA system parameters. In other words, the CNN-LSTM provides a good channel prediction in terms of NRMSE, R2 score, MASE and MAPE. To evaluate the efficiency of the CNN-LSTM, we compare its performance accuracy with that of a CNN-LSTM model using conv1D, CNN model and LSTM model. The CNN-LSTM provides better prediction accuracy in terms of RMSE, MAE, MAPE, and R2 score compared to the other three models. In future work, we can consider the Doppler shift effect on the RIS-NOMA system. In order to deal with practical scenarios, a more general fading channel model, such as Rician fading, can be applied to all channels in the wireless system in future works. The RIS-NOMA system could include the direct channels from BS to each user to consider the effect of the proposed CNN-LSTM model, however, required further investigation. Plus, it is worth noting that an ideal phase shift is infeasible for practical hardware because of unavoidable energy dissipation [69], our future research will consider a practical phase shift model for the proposed RIS-NOMA system. In addition, to reduce the complexity of channel estimation and improve the performance of a system based on RIS, an optimization for the channel estimation of training length or symbols count can be taken into account as well where channel data can be provided by proposed model [70]. The proposed

average values of root mean square error (RMSE), MAE, and MAPE, especially with minus values of average R2 score. The LSTM model cannot exploit the feature extraction of CNN module as in CNN2D-LSTM, therefore it fails to predict the cascaded channel of the RIS-NOMA system. Through the comparative analysis, the proposed CNN2D-LSTM proves its feasibility and effectiveness over the three

channel estimation technique can be further considered and investigated in a scenario of integrated sensing and communication as in [33] in which RIS can potentially work for both sensing and communications signals from UE.

Data Availability Statement: The data can be requested by contacting the corresponding author

References

[1] N. Mensi and D. B. Rawat, "On the performance of partial RIS selection vs. partial relay selection for vehicular communications," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 9, pp. 9475–9489, 2022.

[2] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, 2019, doi: 10.1109/TWC.2019.2922609.

[3] W. U. Khan, E. Lagunas, A. Mahmood, B. M. ElHalawany, S. Chatzinotas, and B. Ottersten, "When RIS meets GEO satellite communications: A new sustainable optimization framework in 6G," in *2022 IEEE 95th Vehicular Technology Conference:(VTC2022-Spring)*. IEEE, 2022, pp. 1–6.

[4] Y. Jin, J. Zhang, C. Huang, L. Yang, H. Xiao, B. Ai, and Z. Wang, "Multiple residual dense networks for reconfigurable intelligent surfaces cascaded channel estimation," *IEEE Trans. Veh. Technol.*, 2021, doi: 10.1109/TVT.2021.3132305.

[5] S. Zhang, S. Zhang, F. Gao, J. Ma, and O. A. Dobre, "Deep learning optimized sparse antenna activation for reconfigurable intelligent surface assisted communication," *IEEE Trans Commun.*, vol. 69, no. 10, pp. 6691–6705, 2021, doi: 10.1109/TCOMM.2021.3097726.

[6] A. Shahraki, M. Abbasi, M. Piran, A. Taherkordi, *et al.*, "A comprehensive survey on 6G networks: Applications, core services, enabling technologies, and future challenges," *arXiv preprint*, 2021, doi: arXiv.2101.12475.

[7] M. Alsabah, M. A. Naser, B. M. Mahmmod, S. H. Abdulhussain, M. R. Eissa, A. Al-Baidhani, N. K. Noordin, S. M. Sait, K. A. Al-Utaibi, and F. Hashim, "6G wireless communications networks: A comprehensive survey," *IEEE Access*, vol. 9, pp. 148 191–148 243, 2021, doi: 10.1109/ACCESS.2021.3124812.

[8] N. H. Mahmood, S. Böcker, A. Munari, F. Clazzer, I. Moerman, K. Mikhaylov, O. Lopez, O.-S. Park, E. Mercier, H. Bartz, *et al.*, "White paper on critical and massive machine type communication towards 6G," *arXiv preprint*, 2020, doi: arXiv:2004.14146.

[9] N. Ye, J. An, and J. Yu, "Deep-learning-enhanced NOMA transceiver design for massive MTC: Challenges, state of the art, and future directions," *IEEE Wirel. Commun.*, vol. 28, no. 4, pp. 66–73, 2021, doi: 10.1109/MWC.001.2000472.

[10] J. He, K. Yang, and H.-H. Chen, "6G cellular networks and connected autonomous vehicles," *IEEE Network*, vol. 35, no. 4, pp. 255–261, 2020, doi: 10.1109/MNET.011.2000541.

[11] Y. Liu, W. Yi, Z. Ding, X. Liu, O. Dobre, and N. Al-Dhahir, "Developing NOMA to next generation multiple access (NGMA): Future vision and research opportunities," *arXiv preprint*, 2021, doi: arXiv:2103.02334.

[12] S. Basharat, S. A. Hassan, H. Pervaiz, A. Mahmood, Z. Ding, and M. Gidlund, "Reconfigurable intelligent surfaces: Potentials, applications, and challenges for 6G wireless networks," *IEEE Wirel. Commun.*, vol. 28, no. 6, pp. 184–191, 2021, doi: 10.1109/MWC.011.2100016.

[13] T. Hou, Y. Liu, Z. Song, X. Sun, Y. Chen, and L. Hanzo, "Reconfigurable intelligent surface aided NOMA networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2575–2588, 2020, doi: 10.1109/JSAC.2020.3007039.

[14] S. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surv. Tutor.*, vol. 19, no. 2, pp. 721–742, 2016, doi: 10.1109/COMST.2016.2621116.

[15] V. Kumar, Z. Ding, and M. F. Flanagan, "On the performance of downlink NOMA in underlay spectrum sharing," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4523–4540, 2021, doi: 10.1109/TVT.2021.3069440.

[16] Y. Liu, X. Mu, X. Liu, M. Di Renzo, Z. Ding, and R. Schober, "Reconfigurable Intelligent Surface (RIS) aided multi-user networks: Interplay between NOMA and RIS," *arXiv preprint*, 2020, doi: arXiv:2011.13336.

[17] G. Yang, X. Xu, Y.-C. Liang, and M. Di Renzo, "Reconfigurable intelligent surface-assisted non-orthogonal multiple access," *I IEEE Trans. Wirel. Commun.*, vol. 20, no. 5, pp. 3137–3151, 2021, doi: 10.1109/TWC.2020.3047632.

[18] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Joint deployment and multiple access design for intelligent reflecting surface assisted networks," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 10, pp. 6648–6664, 2021, doi: 10.1109/TWC.2021.3075885.

[19] Y. Li, M. Jiang, Q. Zhang, and J. Qin, "Joint beamforming design in multi-cluster MISO NOMA reconfigurable intelligent surface-aided downlink communication networks," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 664–674, 2020, doi: 10.1109/TCOMM.2020.3032695.

[20] J. Zuo, Y. Liu, Z. Qin, and N. Al-Dhahir, "Resource allocation in intelligent reflecting surface assisted NOMA systems," *IEEE Trans Commun.*, vol. 68, no. 11, pp. 7170–7183, 2020, doi: 10.1109/SPAWC48557.2020.9154252.

[21] X. Liu, Y. Liu, Y. Chen, and H. V. Poor, "RIS enhanced massive Non-orthogonal multiple access networks: Deployment and passive beamforming design," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1057–1071, 2020, doi: 10.1109/JSAC.2020.3018823.

[22] Z. Mao, M. Peng, and X. Liu, "Channel estimation for reconfigurable intelligent surface assisted wireless communication systems in mobility scenarios," *China Communications*, vol. 18, no. 3, pp. 29–38, 2021, doi: 10.23919/JCC.2021.03.003.

[23] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISOwireless energy transfer," in *Proc. ICASSP 2019, Brighton, UK*, IEEE, 2019, pp. 4659–4663, doi: 10.1109/ICASSP.2019.8683663.

[24] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, "Channel estimation for RIS-empowered multi-user MISO wireless communications," *IEEE Trans Commun.*, 2021, doi: 10.1109/TCOMM.2021.3063236.

[25] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *IEEE Commun. Lett.*, vol. 9, no. 4, pp. 518–522, 2019, doi: 10.1109/LWC.2019.2961357.

[26] J. Chen, Y. Liang, H. Cheng, and W. Yu, "Channel estimation for reconfigurable intelligent surface aided multi-user mmwave MIMO systems," *IEEE Trans. Wireless Commun.*, 2021.

[27] Z. Chen, J. Tang, X. Zhang, Q. Wu, Y. Wang, D. K. C. So, S. Jin, and K.-K. Wong, "Offset learning based channel estimation for intelligent reflecting surface-assisted indoor communication," *IEEE J. Sel. Top.*, 2021.

[28] A. Alkhateeb, "Deepmimo: A generic deep learning dataset for millimeter wave and massive MIMO applications," *arXiv preprint*, 2019.

[29] M. Xu, S. Zhang, J. Ma, and O. A. Dobre, "Deep learning-based time-varying channel estimation for RIS assisted communication," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 94–98, 2021, doi: 10.1109/LCOMM.2021.3127160.

[30] Z. Mao, X. Liu, and M. Peng, "Channel estimation for intelligent reflecting surface assisted massive MIMO systems—A deep learning approach," *IEEE Communications Letters*, vol. 26, no. 4, pp. 798–802, 2022.

[31] C. Liu, X. Liu, D. W. K. Ng, and J. Yuan, "Deep residual learning for channel estimation in intelligent reflecting surface-assisted multi-user communications," *IEEE Trans. Wirel. Commun.*, 2021, doi: 10.1109/TWC.2021.3100148.

[32] W. Xu, J. An, Y. Xu, C. Huang, L. Gan, and C. Yuen, "Time-varying channel prediction for RIS-Assisted MU-MISO networks via deep learning," *arXiv preprint*, 2021, doi: arXiv:2111.04971.

[33] Y. Liu, I. Al-Nahhal, O. A. Dobre, and F. Wang, "Deep-learning channel estimation for IRS-assisted integrated sensing and communication system," *IEEE Transactions on Vehicular Technology*, pp. 1–14, 2022.

[34] Z. Ding, L. Lv, F. Fang, O. A. Dobre, G. K. Karagiannidis, N. Al-Dhahir, R. Schober, and H. V. Poor, "A state-of-the-art survey on reconfigurable intelligent surface-assisted non-orthogonal multiple access networks," *Proceedings of the IEEE*, vol. 110, no. 9, pp. 1358–1379, 2022.

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2023.3278232

Chi Nguyen : Channel Estimation Using CNN-LSTM in RIS-NOMA Assisted 6G Network:

[35] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and Neural Network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003, doi: 10.1016/S0925-2312(01)00702-0.

[36] X. Xu, Q. Chen, X. Mu, Y. Liu, and H. Jiang, "Graph-embedded multi-agent learning for smart reconfigurable THz MIMO-NOMA networks," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 259–275, 2021, doi: 10.1109/JSAC.2021.3126079.

[37] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, "Capacity characterization of intelligent reflecting surface assisted NOMA systems," in *Proc. ICC*. IEEE, 2021, pp. 1–6, montreal, Canada.

[38] K. D. Wong, *Fundamentals of wireless communication engineering technologies*. John Wiley & Sons, 2011.

[39] A. Benjebbour, K. Saito, A. Li, Y. Kishiyama, and T. Nakamura, "Non-orthogonal multiple access (NOMA): Concept, performance evaluation and experimental trials," in *in Proc. WINCOM, Marrakech, Morocco*. IEEE, 2015, pp. 1–6, doi: 10.1109/WINCOM.2015.7381343.

[40] Z. Ding, R. Schober, and H. V. Poor, "On the impact of phase shifting designs on IRS-NOMA," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1596–1600, 2020, doi: 10.1109/LWC.2020.2991116.

[41] S. Zhang and R. Zhang, "Capacity characterization for intelligent reflecting surface aided MIMO communication," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1823–1838, 2020, doi: 10.1109/JSAC.2020.3000814.

[42] K. Odeyemi, P. Owolawi, and O. Olakanmi, "Reconfigurable Intelligent Surface in Wireless-Powered Interference-Limited Communication Networks," *Symmetry*, vol. 13, no. 6, p. 960, 2021, doi: 10.3390/sym13060960.

[43] S. Arzykulov, G. Nauryzbayev, T. A. Tsiftsis, B. Maham, and M. Abdallah, "On the outage of underlay CR-NOMA Networks with detect-and-forward relaying," *IEEE Trans. Cogn. Commun.*, vol. 5, no. 3, pp. 795–804, 2019, doi: 10.1109/WCSP.2018.8555600.

[44] T. P. Huu, T. N. Thi-Thanh, C. Nguyen-Yen, H. Tran, V. N. Dinh, *et al.*, "Secrecy outage probability and fairness of packet transmission time in a NOMA system," *IEEE Access*, vol. 8, pp. 79637–79649, 2020, doi: 10.1109/ACCESS.2020.2989531.

[45] H.-C. Chen, A. M. Widodo, and A. Wisnujati, "The performance analysis of RIS-Aided Cooperative NOMA network for user relaying without the direct link," 2021.

[46] J. Lin, G. Wang, R. Fan, T. A. Tsiftsis, and C. Tellambura, "Channel estimation for wireless communication systems assisted by large intelligent surfaces," *arXiv preprint*, 2019, doi: arXiv:1911.02158.

[47] S. Mehtab and J. Sen, "Stock price prediction using CNN and LSTM-based deep learning models," in *Proc. DASA*. IEEE, 2020, pp. 447–453, doi: 10.1155/2020/6622927.

[48] Q. Wang, S. Bu, Z. He, and Z. Y. Dong, "Toward the prediction level of situation awareness for electric power systems using CNN-LSTM network," *IEEE Trans. Industr. Inform.*, vol. 17, no. 10, pp. 6951–6961, 2020.

[49] S. Liu, Z. Gao, J. Zhang, M. Di Renzo, and M.-S. Alouini, "Deep denoising neural network assisted compressive channel estimation for mmWave intelligent reflecting surfaces," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9223–9228, 2020.

[50] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, pp. 1–74, 2021, doi: 10.1186/s40537-021-00444-8.

[51] B. Farsi, M. Amayri, N. Bouguila, and U. Eicker, "On short-term load forecasting using machine learning techniques and a novel parallel deep lstm-cnn approach," *IEEE Access*, vol. 9, pp. 31191–31212, 2021, doi: 10.1109/ACCESS.2021.3060290.

[52] E. Tsironi, P. Barros, C. Weber, and S. Wermter, "An analysis of convolutional long short-term memory recurrent neural networks for gesture recognition," *Neurocomputing*, vol. 268, pp. 76–86, 2017.

[53] C. Liu, W. Yuan, S. Li, X. Liu, H. Li, D. W. K. Ng, and Y. Li, "Learning-based predictive beamforming for integrated sensing and communication in vehicular networks," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 8, pp. 2317–2334, 2022.

[54] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 5353–5360.

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

[56] J. Gao, X. Yi, C. Zhong, X. Chen, and Z. Zhang, "Deep learning for spectrum sensing," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1727–1730, 2019.

[57] R. J. Hyndman *et al.*, "Another look at forecast-accuracy metrics for intermittent demand," *Foresight: The International Journal of Applied Forecasting*, vol. 4, no. 4, pp. 43–46, 2006.

[58] Ü. Ç. Büyükşahin and Ş. Ertekin, "Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition," *Neurocomputing*, vol. 361, pp. 151–163, 2019, doi: 10.1016/j.neucom.2019.05.099.

[59] C. Ma, G. Dai, and J. Zhou, "Short-term traffic flow prediction for urban road sections based on time series analysis and LSTM_BiLSTM method," *IEEE trans Intell Transp Syst.*, 2021, doi: 10.1109/TITS.2021.3055258.

[60] C. Ma and R. He, "Green wave traffic control system optimization based on adaptive genetic-artificial fish swarm algorithm," *Neural. Comput. Appl.*, vol. 31, no. 7, pp. 2073–2083, 2019.

[61] C.-X. Wang, J. Huang, H. Wang, X. Gao, X. You, and Y. Hao, "6G wireless channel measurements and models: Trends and challenges," *IEEE Veh. Technol. Mag.*, vol. 15, no. 4, pp. 22–32, 2020, doi: 10.1109/MVT.2020.3018436.

[62] A. Epstein and G. V. Eleftheriades, "Synthesis of passive lossless metasurfaces using auxiliary fields for reflectionless beam splitting and perfect reflection," *Phys. Rev. Lett.*, vol. 117, no. 25, p. 256103, 2016, doi:10.1103/PhysRevLett.117.256103.

[63] T. Badloe, J. Mun, and J. Rho, "Metasurfaces-based absorption and reflection control: perfect absorbers and reflectors," *Journal of Nano-materials*, vol. 2017, 2017, doi: 10.1155/2017/2361042.

[64] Y. Sun, J. A. Zhang, K. Wu, and R. P. Liu, "Frequency-domain sensing in time-varying channels," *IEEE Wireless Communications Letters*, vol. 12, no. 1, pp. 16–20, 2023.

[65] Y. Ye and Y. Chen, "Machine-learning-based pilot symbol assisted channel prediction," *IET Communications*, vol. 16, no. 8, pp. 866–877, 2022.

[66] A. M. Elbir, A. Papazafeiropoulos, P. Kourtessis, and S. Chatzinotas, "Deep channel learning for large intelligent surfaces aided mm-Wave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1447–1451, 2020, doi: 10.1109/LWC.2020.2993699.

[67] Y. Liao, Y. Hua, X. Dai, H. Yao, and X. Yang, "Chanestnet: A deep learning based channel estimation for high-speed scenarios," in *Proc. ICC, Shanghai, China*. IEEE, 2019, pp. 1–6, doi: 10.1109/ICC.2019.8761312.

[68] A. K. Gizzini, M. Chafii, S. Ehsanfar, and R. M. Shubair, "Temporal averaging lstm-based channel estimation scheme for ieee 802.11 p standard," in *2021 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2021, pp. 01–07.

[69] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5849–5863, 2020, doi: 10.1109/ACCESS.2019.2935192.

[70] C. You, B. Zheng, and R. Zhang, "Channel estimation and passive beamforming for intelligent reflecting surface: Discrete phase shift and progressive refinement," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2604–2620, 2020, doi: 10.1109/JSAC.2020.3007056.

Chi Nguyen received the B.S. degree in electronics and telecommunications engineering from the University of Transport and Communications, Hanoi, Vietnam, in 2011, and the M.S. degree in telecommunication engineering from the Hanoi University of Science and Technology, in 2016. She is currently a Ph.D. student at University of Ulster. Her research interests include channel modeling, DL, non-orthogonal multiple access (NOMA), reconfigurable intelligent surface (RIS), physical layer security, underwater communications, and visible light communications.

Tiep M. Hoang received the B.Eng. degree from the HCMC University of Technology, Vietnam, in 2012, the M.Eng. degree from Kyung Hee University, South Korea, in 2014, and the Ph.D. degree from the Queen's University of Belfast, United Kingdom, in 2019. From 2020 to 2022, he was a (postdoctoral) Research Fellow with the School of Electronics and Computer Science, the University of Southampton, United Kingdom. Since May 2022, he has been a Postdoctoral Fellow with the Department of Electrical Engineering, the University of Colorado Denver, United States. His current research interests include 5G/6G wireless communications, wireless security and authentication, reconfigurable intelligent surface (RIS), convex optimization, and machine learning.

Adnan A. Cheema received the B.Sc. degree from the COMSATS University, Pakistan., M.Sc. degree from the King's College London, U.K., and the Ph.D. degree from the Durham University, U.K., in 2006, 2008 and 2015, respectively. From 2015 to 2017, he was a postdoctoral Research Associate at Durham University and was involved in 5G channel measurements and modelling (sub-6 GHz and 24-90 GHz). In 2017, he joined Ulster University, U.K., as a Lecturer in electronics engineering and currently leading the SenComm research lab. His research interests include channel measurements and modelling, reconfigurable intelligent surface and AI for wireless communications (5G/6G).