

Article

# Improving CSI Prediction Accuracy with Deep Echo State Networks in 5G Networks

Tommaso Pecorella , Romano Fantacci  and Benedetta Picano \* 

Department of Information Engineering, University of Florence, 50139 Firenze, Italy; tommaso.pecorella@unifi.it (T.P.); romano.fantacci@unifi.it (R.F.); benedetta.picano@unifi.it (B.P.)

\* Correspondence: benedetta.picano@unifi.it

Received: 22 September 2020; Accepted: 10 November 2020; Published: 12 November 2020



**Abstract:** The forthcoming fifth-generation networks require improvements in cognitive radio intelligence, going towards more smart and aware radio systems. In the emerging radio intelligence approach, the empowerment of cognitive capabilities is performed through the adoption of machine learning techniques. This paper investigates the combined application of the convolutional and recurrent neural networks for the channel state information forecasting, providing a multivariate scalar time series prediction by taking into account the multiple factors dependence of the channel state conditions. Finally, the system performance has been analyzed in terms of prediction accuracy expressed as absolute deviation error and mean percentage error, in comparison with an alternative machine learning method recently proposed in the literature with the aim at solving the same prediction problem.

**Keywords:** channel state information; recurrent neural networks; time series forecasting

## 1. Introduction

The emergence of the need for new network technologies and architectures has been launched by the diffusion of novel services and applications demanding high data rates and reliability, low latency, and congestion levels. Augmented reality, automatic driving, and tactile internet are only a few of the examples of fifth-generation (5G) applications. Furthermore, new era applications require remarkable improvements with regards to the intelligence core of the network, giving rise to self-organizing ecosystems able to contextualise awareness in all its different forms and facets. In fact, one of the most challenging issues in 5G networks is the empowering of the learning processes, recreating on devices actual human neural network behavior to obtain effective solutions to new era networks problems.

From previous network generations, the analysis and design of innovative technologies and methodologies to move towards wireless systems able to optimally exploit radio resources and actuate proper decision policies has gained attention. During the years, cognitive radio (CR) approaches have played a crucial role, aiming at designing context-driven strategies. In this regard, the advent of the 5G networks, which involve the usage of the millimeter wave (mmWave) band has resulted in a large variety of obstacles such as pathloss, blockage, and high oxygen absorption [1]. All these factors increase the complexity of detecting actual channel conditions and signals characteristics. Under these circumstances, the channel state information (CSI) represents one of the most important concepts in the wireless networks [2–4], providing a performance metric for the quality of the radio channel. The CSI describes the physical layer parameters arranged for wireless communication systems [2], being crucial in many decision making areas such as radio resource allocation [5], interference management [6], and so on. Consequently, an accurate value of the CSI plays a role of paramount importance in a wide range of practical applications, while inaccuracy on the CSI prediction value leads inevitably to substantial degradation on communication effectiveness.

Generally speaking, the channel prediction is an impressive approach to obtain accurate CSI values. Differently, during years, many CSI estimation methods have been proposed [2], from the maximum-likelihood estimation [2,7], up to the minimum mean squared error estimation [2,8], passing through the least square estimation [2,9].

Typically, these methods involve matrix operations as the matrix inversion or the eigenvalue decomposition, resulting in procedures characterized by high computational costs. Due to such hardness in channel condition knowledge or the presence of remarkable overhead costs associated with the traditional channel state estimation methods, the design of novel strategies has become a necessity in reference to upcoming 5G networks. In fact, 5G environments have to face highly complex scenarios, in which CSI prediction is a large-scale channel problem, due to the presence of massive multiple-input-multiple-output (MIMO), orthogonal frequency division multiplexing (OFDM), and mmWave communications [2]. As consequence, the matrix operations involved in the traditional CSI estimation methods, applied to 5G contexts, bring large-scale matrix operations with significant computational complexity and overheads, therefore becoming unviable approaches implying long convergence time [2].

Recently, the machine learning (ML) techniques have gained attention for their application in numerous research fields such as channel prediction [10], detection [11], modulation [12], and so on. Therefore, the ML provides modules and frameworks suitable to be contextualized to the CR environments, especially those in which several factors are involved. Indeed, the CSI parameter depends on a multitude of factors interrelated to each other as humidity, temperature, time, location, etc. [2], which contribute to the definition of the CSI. In this sense, scalar time series analysis and prediction based on the ML may represent a concrete strategy to provide accurate CSI forecast.

This paper proposes the application of the ML principles to the CSI prediction, by formulating that problem in terms of scalar time series forecasting [13].

Overall, the class of the prediction problems can be classified into short-, medium-, and long-term forecasting. In the first one, the predictions are performed over horizons from a few minutes up to a few days ahead, and the second one deals with time horizons from few days to few months ahead. Finally, predictions from quarters to years ahead [14] are named long-term forecasting. It is important to highlight that typically, medium and long-term forecasting involve the prediction about risk management and profitability planning. Differently, short-term forecasting is applied to traffic demands and mobility prediction [14,15]. Short-term forecasting has been extensively studied in the literature, and many different methods have been proposed [13,15]. Therefore, the aims of the paper are summarized as

- Design and development of an ML-based framework involving the usage of the convolutional neural networks (CNNs) and the recurrent neural network (RNN) to predict the CSI behavior;
- Extensive numerical simulations to provide the performance evaluation analysis as regards the prediction accuracy expressed as mean absolute deviation error and mean percentage error between the actual CSI value and those predicted;
- Testing of the proposed framework under different application scenarios, i.e., outdoor and indoor conditions, in comparison with the deep learning strategy proposed in [2] and that presented in [16].

The rest of the paper is organized as follows. Section 2 presents an in-depth review of the related literature, while in Section 3 the system model and the problem formulation is detailed. Section 4 presents the technique adopted to reach the main objective of the paper. Validations of the proposed scheme are presented in Section 5, and conclusions are drawn in Section 6.

## 2. Related Works

The CSI prediction problem has been largely studied and, during the years, different approaches have been pursued. Recently, the involvement of ML techniques for estimating the CSI has gained

momentum. Authors in [17] address the problem of the short-term forecasting of the fading channel. In fact, they propose the application of both the machine learning and the Savitzky–Golay filter to increase the prediction accuracy. The paper [18] focuses on the application of the ML to the adaptive coding and modulation, considering satellite networks environments. More specifically, in [18] the prediction is performed on the signal-to-noise ratio series, and the remarkable forecasting accuracy derived by the application of the ML approach in comparison with the traditional alternatives. Furthermore, in [19], the authors expose the channel fading prediction for addressing the CSI aging problem to increase the network throughput and spectral efficiency. Differently, the sparse Bayesian linear regression combined with the support vector machine approach has been proposed in paper [20], whose main objective is the joint CSI prediction and the dynamic radio frequency slicing, considering the software-defined networks and the MIMO systems. Massive MIMO contexts have been also considered in [21], in which the ML has been applied to lower the overhead due to the downlink channel estimation. The proposed scheme consists of the combined usage of the linear regression and the support vector regression in ML. More in detail, the authors in [21] have involved realistic CSI samples to train the regression model. Then, the CSI antennas are online estimated by inputting the output data derived by the regression model. Similarly, the overheads due to the orthogonal pilot based estimation has been widely studied in paper [16], by designing an ML oriented scheme focusing on the channel time correlations analysis. In fact, the framework proposes the application of a CNN together with an auto-regressive model or an RNN auto-regressive network with exogenous inputs.

Differently, in the paper [22] highly mobile vehicular environments are considered. The authors affirm that, due to the almost identical propagation conditions, a deep learning approach may acquire and analyze the non-linear temporal correlation characterizing subsequent CSI samples, without the side effect of a remarkable overhead. In addition, in [22] a resource allocation for the network slicing problem is presented, by assuming vehicular users demanding enhanced mobile broadband (eMBB) and ultra-reliable low latency (URLLC) traffic slices. Finally, the CSI estimation within the free space optical communication networks is proposed in the paper [23], where also in this case an ML approach is pursued. Specifically, the authors involve the usage of both the maximum likelihood estimation and the Bayesian estimation methods, to estimate the channel coefficients of interest for the considered scenario.

### 3. Problem Statement

The CSI forecasting problem has been largely studied and it strongly correlated to the power attenuation characterizing the radio signals when transmitting over a distance. In fact, generally speaking, a radio channel is affected by phenomenons as scattering, fading, or path loss [2]. The CSI represents a parameter providing a measurement of the combined effect of all these factors. As is well known from literature, the path loss can be expressed as follows

$$\mathcal{P}\mathcal{L} = 10 \log_{10} \frac{P_{tx}}{P_{rx}}, \quad (1)$$

where  $P_{tx}$  and  $P_{rx}$  are the transmitting and the receiving power, respectively. Equivalently, Equation (1) can be expressed as a function of the distance as [2]

$$- 10 \log_{10} G \left( \frac{\lambda}{4\pi d} \right)^2, \quad (2)$$

in which the term  $G$  represents the product of the antenna field radiation pattern [2],  $d$  is the distance between a transmitter and the corresponding receiver, while  $\lambda = 3 \times 10^8 / f$  represents the wavelength and  $f$  the frequency. Wireless communications, in addition to being affected by path loss, suffer also from scattering and fading. Although the scattering is due to electromagnetic wave propagation, the fading is mainly dependent on the frequency and it is related to the obstacles and multi-paths

propagation [2]. Therefore, all these factors make it challenging to provide an accurate CSI prediction. This paper takes into account a MIMO system, in which multiple antennas at the transmitter and receiver are used. MIMO exploits the multiple antennas to perform the spatial dimension in addition to the time and frequency ones, avoiding changes in the bandwidth requirements of the system. More specifically, this paper focuses on a slowly-fading MIMO channel model, in which the channel state remains quasi-static within a fading block [24].

Furthermore, it is important to stress that we have that the CSI value results from a wide range of factors, such as weather, location, frequency band, time, and so on [2]. Considering a Wi-Fi channel and OFDM modulation, the raw CSI measure consists of complex numbers whose amplitude and phase are influenced by the propagation conditions of the communication that usually give rise to an amplitude attenuation and phase shift for the received signal. Such values depend on the frequency band used. For more details about the CSI extraction in Wi-fi MIMO systems in practice, we refer to [25,26]. It is important to highlight that this paper, inspired by [2], considers the CSI, as resulting from the combination of the following factors

- *Frequency band.* We consider a set of frequency bands for 5G denoted by  $\{f_1, \dots, f_N\}$ , hereafter used interchangeably with channels.
- *Location.* Due to the fluctuation of the factors previously introduced, different regions within the base station coverage may experience different CSI. Indeed, the locations are represented by  $\{l_1, \dots, l_M\}$ , where  $l_m$  is the  $m$ -th location, expressed as sub-region.
- *Time.* Due to the fact that the atmosphere density changes during different hours in a day and seasons [2], we slotted the time as  $\{\tau_1, \dots, \tau_p\}$ , where  $\tau_i$  represents the time slot  $i$ .
- *Weather.* The weather massively influences the CSI. In fact, as modeled in [2], we considered the following weather levels
  1. sunny;
  2. lightly cloudy;
  3. cloudy;
  4. lightly rainy;
  5. medium rainy;
  6. heavy rainy;
  7. light snow;
  8. medium snow;
  9. heavy snow.

Then, the corresponding weather set is denoted by  $W = \{w_1, \dots, w_R\}$ , in which  $w_r$  represents the corresponding weather level. As consequence, in this paper we consider the *CSI sample*, hereafter referred to STATE, given as input to the proposed learning strategy, having the following format

$$STATE = (f_n, l_m, \tau_p, w_r, CSI), \quad (3)$$

in which  $f_n$  represents the band,  $l_m$  the location,  $\tau_p$  the time, and  $w_r$  the weather level, as previously detailed. The main objective of the paper is short-term CSI forecasting.

#### 4. Forecasting Strategy

Recently, CNNs have been largely applied in many image processing and recognition problems. In this paper, we exploit a framework architecture yet presented in literature based on the CNNs [2], and then we combined such neural network architecture to a particular type of recurrent neural network (RNN) particularly suitable for its low complexity, named echo state network (ESN). The RNNs allows the exploitation of the correlation between consecutive samples within the time series.

As detailed in [2], the CNN based architecture consists of a 2D CNN and a cascade 1D CNN. In fact, the first 2D convolution module, as depicted in Figure 1 the CNN modules take as input the historical CSI images, i.e., the images of the data collected along time on which the training is performed. More specifically, the 2D CNN is responsible for the frequency extraction, in which there is

a composed of a set of filters in parallel and then a set of weights. Consequently, the state representative vector is extracted by the 1D network from the frequency representative vectors [2]. Therefore, the last component of the proposed framework is the ESN module. The ESN is a particular RNN, based on the Reservoir Computing (RC) paradigm [27,28] which, recently has been considered an effective approach for efficiently training the RNNs. Despite many instances of the RC methodology have been presented in the literature [29], the ESN [30] is one of the most known models [31–35]. Roughly speaking, the ESNs belong to a subclass of the RNN, named as the recurrent randomized neural networks [36], whose state dynamics are implemented by an untrained recurrent hidden layer [37].

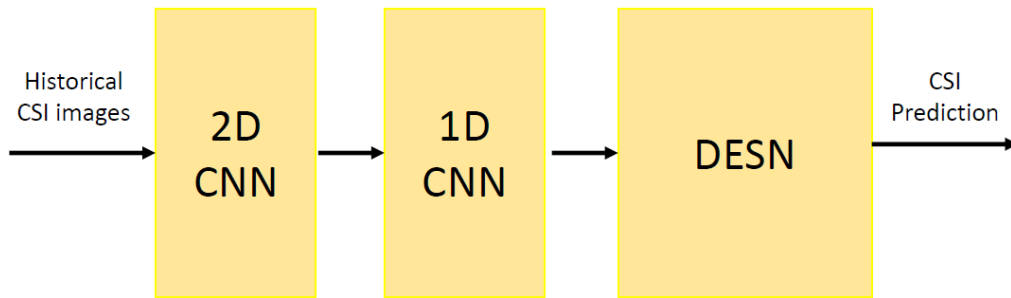


Figure 1. The forecasting framework.

The Deep Echo State Network (DESN) [38] emerged as a promising model for efficiently training deep neural networks for temporal data [37]. The DESN is mainly composed of the following characteristics [37]

- a dynamical reservoir component, transforming the input history into a state representation;
- a feed-forward readout component, which computes the output.

As widely discussed in the literature [37], the reservoir of a DESN is characterized by a hierarchy of cascade recurrent layers, where the output of each layer represents the input for the next one, as illustrated in Figure 2. At each time step  $t$ , let  $N_U$  be the external input dimension, and let  $N_L$  be the number of reservoir layers. Furthermore, each reservoir layer has  $N_R$  recurrent units. Then, the external input at time  $t$  is defined by  $\mathbf{u}(t) \in \mathbb{R}^{N_U}$ , while  $\mathbf{x}^{(i)}(t) \in \mathbb{R}^{N_R}$  represents the state of the reservoir layer  $i$ -th at time slot  $t$ . At each time slot  $t$ , the global state of the DESN, i.e., the state in all the reservoir layers is expressed by

$$\mathbf{x}(t) = (\mathbf{x}^{(1)}(t), \dots, \mathbf{x}^{(N_L)}(t)) \in \mathbb{R}^{N_R N_L}. \quad (4)$$

In formal terms, the reservoir dynamics for the first layer is described as in (5), while

$$\mathbf{x}^{(1)}(t) = F(\mathbf{u}(t), \mathbf{x}^{(1)}(t-1)) = (1 - a^{(1)})\mathbf{x}^{(1)}(t-1) + \mathbf{f}(\mathbf{W}^{(1)}\mathbf{u}(t) + \widehat{\mathbf{W}}^{(1)}\mathbf{x}^{(1)}(t-1)), \quad (5)$$

the  $i$ -th update is given by (6), in which  $\mathbf{W}^{(1)}$  and  $\mathbf{W}^{(i)}$  are the input and the inner weight matrix, respectively. More in depth,  $\mathbf{W}^{(i)}$  represents the weight matrix for the connection of the  $(i-1)$ -th inner layer to the  $i$ -th. Similarly,  $\widehat{\mathbf{W}}^{(i)}$  defines the recurrent weights matrix for layer  $i$ , while  $a^{(i)}$  is the leaking rate and  $\mathbf{f}$  is the activation function here selected as the hyperbolic tangent.

$$\mathbf{x}^{(i)}(t) = F(\mathbf{x}^{(i-1)}(t), \mathbf{x}^{(i)}(t-1)) = (1 - a^{(i)})\mathbf{x}^{(i)}(t-1) + \mathbf{f}(\mathbf{W}^{(i)}\mathbf{x}^{(i-1)}(t) + \widehat{\mathbf{W}}^{(i)}\mathbf{x}^{(i)}(t-1)), \quad (6)$$

Accordingly to the RC framework, the reservoir parameters, i.e., the matrices weights, are left untrained and constrained through stability conditions [37].

In reference to the computation of the output, typically the global state of the network feeds, at each time slot  $t$ , the output layer. Consequently, the readout gives different weights to the dynamics

developed at different layers, favoring the diversity exploitation due to the representations in the cascade reservoir.

Therefore, by supposing the output size given by  $N_Y$ , at time slot  $t$  the output can be expressed as [37]

$$\mathbf{y}(t) = \mathbf{W}_{out}\mathbf{x}(t) = \mathbf{W}_{out}(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N_L)}), \quad (7)$$

in which  $\mathbf{W}_{out} \in \mathbb{R}^{N_L \times N_R N_L}$  is the readout weight matrix. To predict the CSI value, we have to transform the input STATE into a format processable by the proposed CNN based strategy. Therefore, the approach proposed in [2], in which the CSI is represented as a pixel image  $\mathcal{I}$ , has been pursued resulting in

$$\mathcal{I} = [C_1, \dots, C_l], \quad (8)$$

in which  $C_i$  represents the *information segment* of the CSI, in accordance with [2]. At this point, the input CSI image is ready to be processed by the CNN, in which the convolutional filters are applied to the image giving rise to the feature image  $\mathcal{I}^*$ , which represents the collection of the outputs stem by the application of the filters. Once the convolutional filters are combined together, the representative vectors are build by concatenating the information segment. Then, the representative vectors, each of which corresponds to a single frequency, are concatenated to obtain the state representative matrix of the considered time slot. Finally, the 1D CNN is applied to catch the temporal features from the data matrix. Finally, the DESN is applied to obtain the CSI prediction considering the minimization of the mean percent error (MAPE), defined as

$$MAPE = \frac{1}{M} \sum_{i=1}^M \left| \frac{\widehat{\mathcal{I}}_{i+\delta} - \mathcal{I}_{i+\delta}}{\mathcal{I}_{i+\delta}} \right| \cdot 100, \quad (9)$$

where  $M$  represents the number of the samples in test data, and  $\widehat{\mathcal{I}}_{i+\delta}$  and  $\mathcal{I}_{i+\delta}$  are the actual and the predicted values at time  $i + \delta$ , where  $\delta$  expresses the increment on the time horizon of the predictive strategy. In addition, to provide an exhaustive analysis, we have also considered the mean absolute deviation (MAD) given by

$$MAD = \frac{1}{M} \sum_{i=1}^M |\widehat{\mathcal{I}}_{i+\delta} - \mathcal{I}_{i+\delta}|. \quad (10)$$

It is important to note that metric (10) highlights the variability of the forecasting error, while (9) expresses the error in terms of percentage on the actual data.

The proposed procedure acts throughout two phases: the online and the offline stages. The offline module corresponds to the learning phase, performed over the massive volume data collected over time, to catch the CSI behavior trend over time. In this framework stage, the stochastic descent gradient method is applied, together with the backpropagation technique, aiming at training the matrix weights  $W$ . Then, the prediction is improved during the online phase, in which the new CSI measurements are used as feedback for the proposed neural network framework. The online phase update is periodic and the model is updated following the standard law given by [2]

$$W = W - \alpha \nabla W, \quad (11)$$

where  $\alpha$  expresses the learning rate. For the offline phase,  $\alpha$  has been selected within the interval  $[-0.1, 0.1]$ , while the online phase exhibits  $\alpha = 0.1$ . Finally, to perform the prediction, we have assumed a  $2 \times 2$  filter for the 2D CNN, and a  $3 \times 1$  filter for the 1D CNN. Then, the DESN has been set considering 128 hidden nodes, in which the connection has been randomly established.

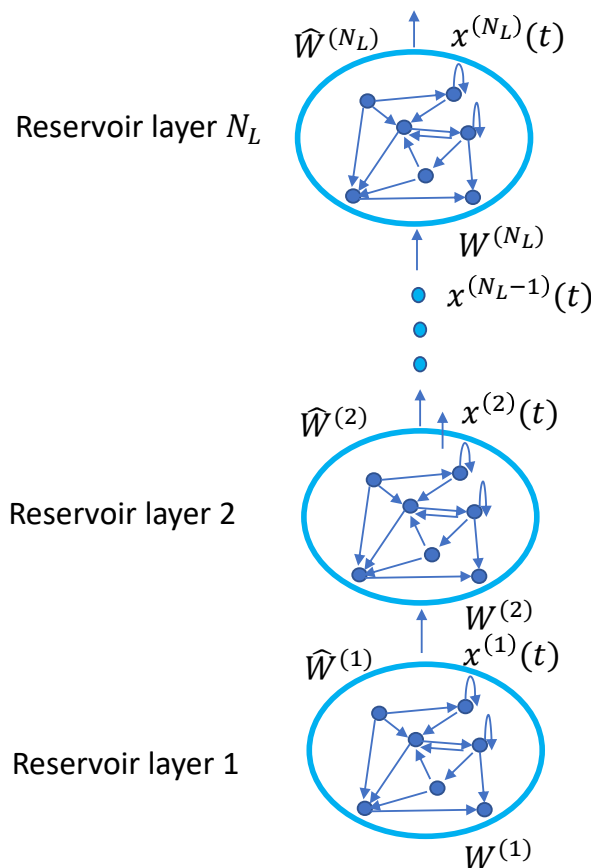


Figure 2. Reservoir architecture.

## 5. Numerical Results

In this paper, we analyzed two sets of data related to the nonlinear CSI time series corresponding to the realistic conditions of indoor and outdoor environments by collecting actual CSI samples on the campus with regards to the outdoor data, and within a building for the indoor scenario. Furthermore, it is important to highlight that the outdoor environment considers free space conditions, while the indoor gathering procedure has been performed in presence of several obstacles such as desks, chairs, walls, doors, and other ordinary office objects. The data have been collected by using two Nexus and three Raspberry Pi devices. Then, the data sets collected contain over 1500 non-duplicating samples, in which two-tiers of data are used for offline learning and the remaining data are used for validating the procedure. In fact, the obtained performance results derived by the usage of the collected CSI samples for both the framework training and testing, properly split among the training and testing sets. Furthermore, to provide an exhaustive analysis, we compare the proposed DESN method with the OCEAN scheme recently proposed in [2] and the method proposed in [16], hereafter referred to as the LSTM.

### 5.1. Discussion

The comparison between the collected data and the data predicted are reported in Figures 3 and 4, with regards to the indoor and outdoor conditions scenario, respectively. Figure 3 exhibits the closeness between the actual data collected within a building and the predicted values. Similarly, Figure 4 shows that the forecast CSI are remarkably close to those collected within an outdoor environment.

Aiming at analyzing the performance of the strategies taken into account, the MAD has been represented in Figures 5 and 6 considering, also in this case, both the indoor and outdoor application contexts. As it is clearly evident from both the Figures, the prediction accuracy is greater when the

proposed algorithms are applied. In addition, it is important to highlight that one slot is assumed equal to 4 ms. Furthermore, in all the strategies the MAD grows by increasing  $\delta$  hereafter expressed in milliseconds, and denoting the time horizon considered to perform the prediction. It is a direct consequence of the complexity in predicting system dynamics over long interval times. To extensively analyze the framework designed, Figures 7 and 8 depict the MAPE metric trend, for the indoor and outdoor cases, respectively. The results confirm the good performance of the proposed approach in comparison with the alternatives taken into account.

In conclusion, the proposed DESN approach provides a valuable framework for short term prediction, within both the indoor and outdoor system conditions.

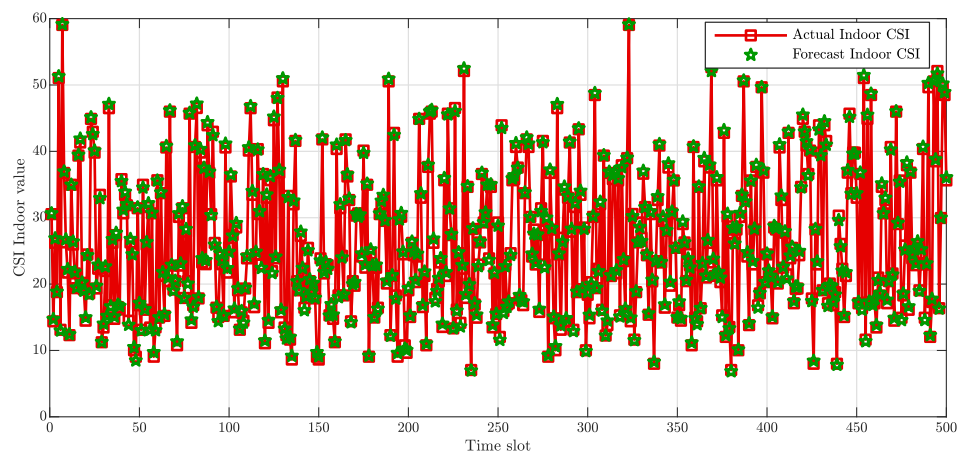


Figure 3. Indoor channel state information (CSI) amplitude value comparison in dB.

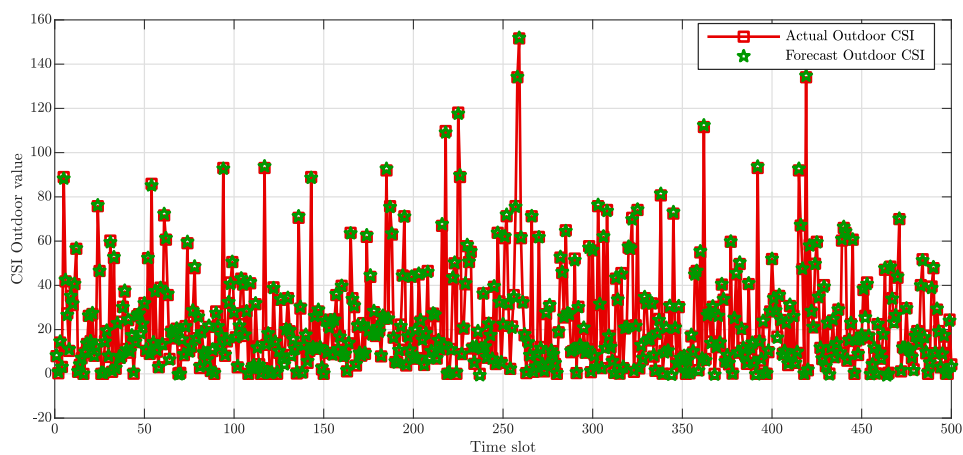


Figure 4. Outdoor CSI amplitude value comparison in dB.

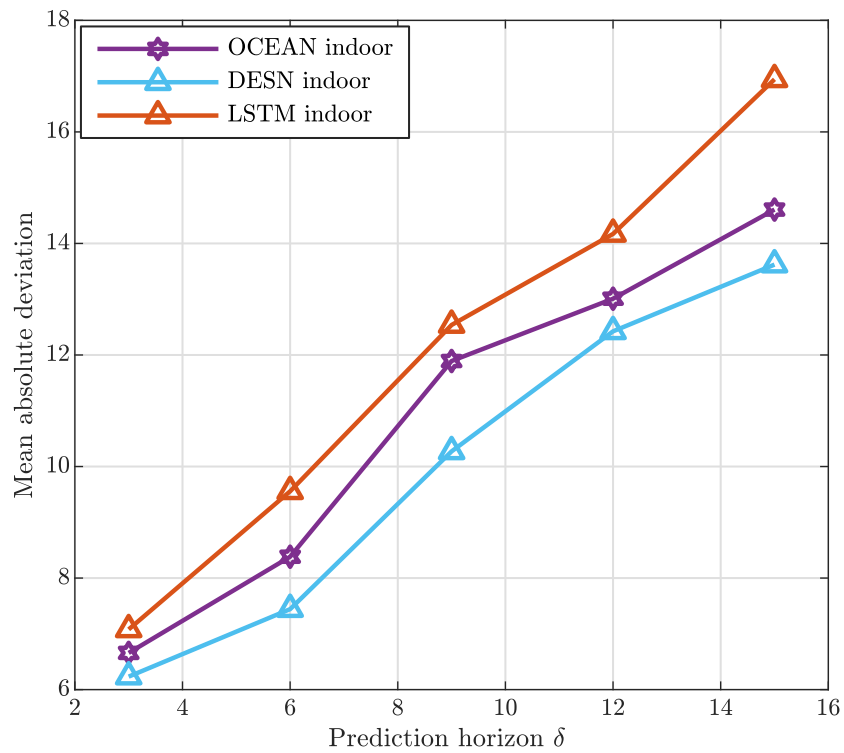


Figure 5. Indoor mean absolute deviation.

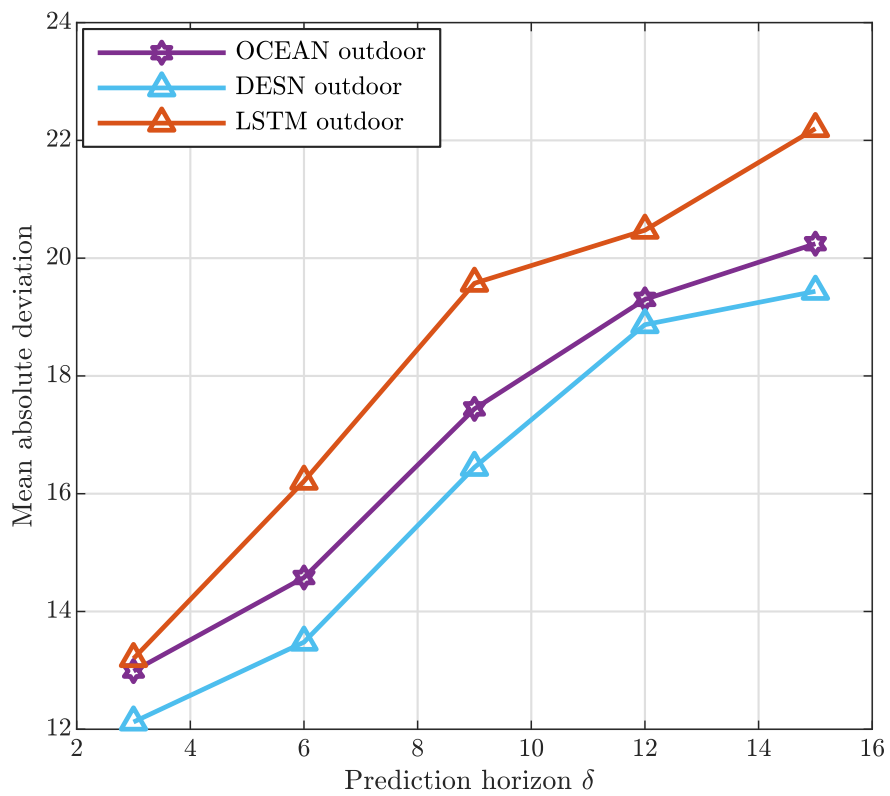


Figure 6. Outdoor mean absolute deviation.

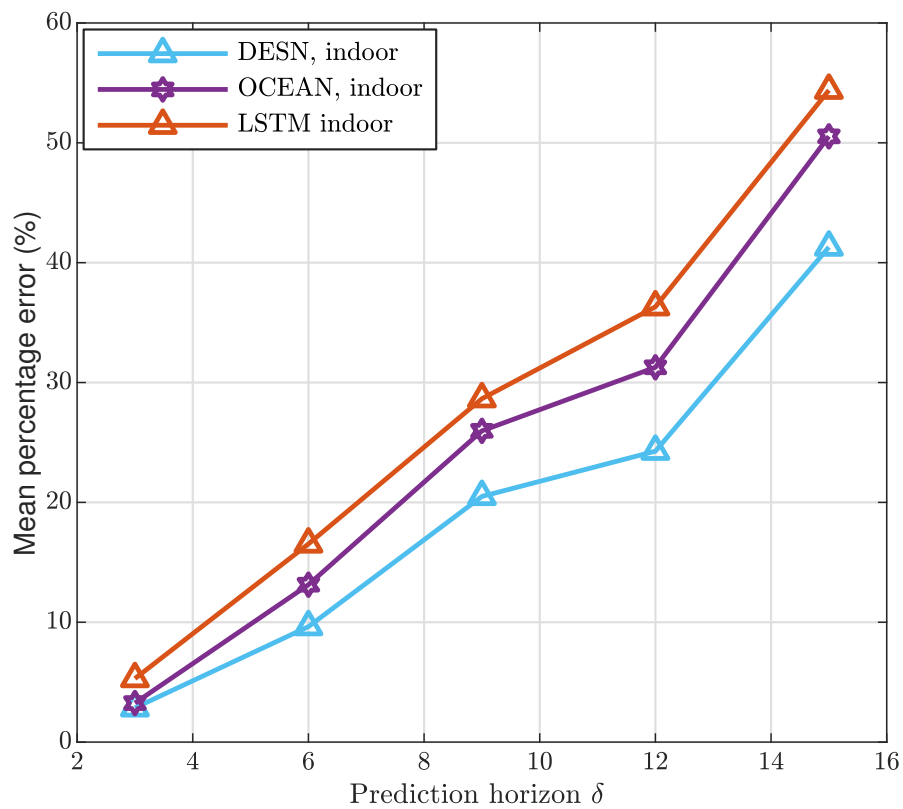


Figure 7. Indoor mean absolute percentage error.

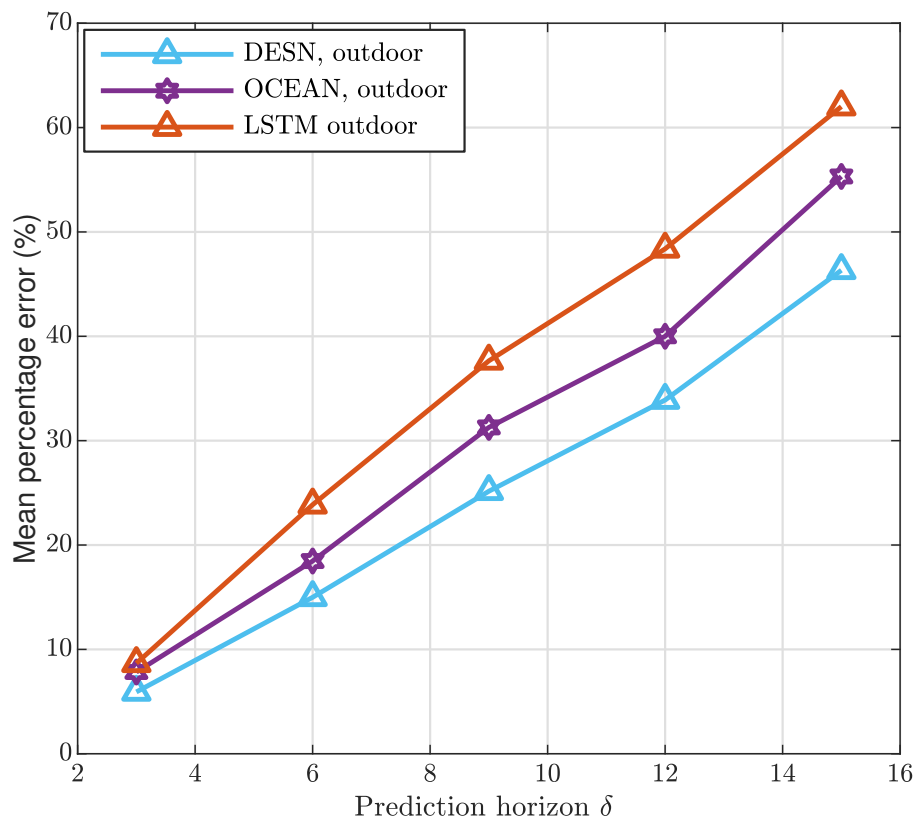


Figure 8. Outdoor mean absolute percentage error.

## 6. Conclusions

This paper has dealt with the CSI prediction problem, by proposing a forecasting strategy combining both the CNNs and the RNNs approaches and providing a method for predicting the CSI at the next time slot. This paper improves the CNN-based strategy recently proposed in literature by [2] and overcomes that proposed in [16].

Furthermore, the validation of the proposed scheme has been provided considering both the indoor and outdoor conditions, and the forecasting performance has been depicted in terms of mean forecasting errors to express the prediction accuracy. Finally, the CSI prediction accuracy resulting from the application of the proposed framework has been compared with that obtained with the strategy proposed in [2] and [16]. As it is evident to note from the simulation results, the proposed strategy overcomes the performance reached by applying the methods presented in [2] and [16], and constitutes a suitable prediction approach for the 5G networks. Such remarkable performance is due to the introduction of a further deep learning phase given by the DESN in comparison to the alternatives taken into account. In fact, the DESN, belonging to the RNN class, is particularly suitable in catching the temporal patterns characterizing data, resulting in a better fit of the data behavior.

**Author Contributions:** All the authors contributed equally to the paper. R.F. coordinated the research activities, T.P. and B.P. mainly worked on the system model and the development of the proposed framework. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Sim, M.S.; Lim, Y.; Park, S.H.; Dai, L.; Chae, C. Deep Learning-Based mmWave Beam Selection for 5G NR/6G with Sub-6 GHz Channel Information: Algorithms and Prototype Validation. *IEEE Access* **2020**, *8*, 51634–51646.
2. Luo, C.; Ji, J.; Wang, Q.; Chen, X.; Li, P. Channel State Information Prediction for 5G Wireless Communications: A Deep Learning Approach. *IEEE Trans. Netw. Sci. Eng.* **2018**, *7*, 227–236, doi:10.1109/TNSE.2018.2848960.
3. Tulino, A.M.; Lozano, A.; Verdu, S. Impact of antenna correlation on the capacity of multi-antenna channels. *IEEE Trans. Inf. Theory* **2005**, *51*, 2491–2509.
4. Li, M.; Meng, Y.; Liu, J.; Zhu, H.; Liang, X.; Liu, Y.; Ruan, N. When CSI Meets Public WiFi: Inferring Your Mobile Phone Password via WiFi Signals. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16), Vienna, Austria, 24–28 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 1068–1079, doi:10.1145/2976749.2978397.
5. Mokari, N.; Navaie, K.; Khoshkholgh, M.G. Downlink Radio Resource Allocation in OFDMA Spectrum Sharing Environment with Partial Channel State Information. *IEEE Trans. Wirel. Commun.* **2011**, *10*, 3482–3495.
6. Gholami Davoodi, A.; Jafar, S.A. Generalized Degrees of Freedom of the Symmetric  $K$  User Interference Channel Under Finite Precision CSIT. *IEEE Trans. Inf. Theory* **2017**, *63*, 6561–6572.
7. Du, Z.; Song, X.; Cheng, J.; Beaulieu, N.C. Maximum Likelihood Based Channel Estimation for Macrocellular OFDM Uplinks in Dispersive Time-Varying Channels. *IEEE Trans. Wirel. Commun.* **2011**, *10*, 176–187.
8. Ma, J.; Zhang, S.; Li, H.; Zhao, N.; Nallanathan, A. Iterative LMMSE Individual Channel Estimation Over Relay Networks With Multiple Antennas. *IEEE Trans. Veh. Technol.* **2018**, *67*, 423–435.
9. Karami, E. Tracking Performance of Least Squares MIMO Channel Estimation Algorithm. *IEEE Trans. Commun.* **2007**, *55*, 2201–2209.
10. Wei, X.H.; Hu, C.; Dai, L. Knowledge-Aided Deep Learning for Beamspace Channel Estimation in Millimeter-Wave Massive MIMO Systems. *arXiv* **2019**, arXiv:abs/1910.12455.
11. Samuel, N.; Diskin, T.; Wiesel, A. Deep MIMO detection. In Proceedings of the 2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Sapporo, Japan, 3–6 July 2017; pp. 1–5.

12. O'Shea, T.; Hoydis, J. An Introduction to Deep Learning for the Physical Layer. *IEEE Trans. Cogn. Commun. Netw.* **2017**, *3*, 563–575.
13. Montgomery, D.C.; Jennings, C.L.; Kulahci, M. *Time Series Analysis and Forecasting*, 2nd ed.; Wiley: Cambridge, UK, 2015, doi:10.1017/CBO9780511755798.
14. Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.* **2014**, *30*, 1030–1081, doi:10.1016/j.ijforecast.2014.08.008.
15. Liu, X.; Fang, X.; Qin, Z.; Ye, C.; Xie, M. A Short-Term Forecasting Algorithm for Network Traffic Based on Chaos Theory and SVM. *J. Netw. Syst. Manag.* **2011**, *19*, 427–447, doi:10.1007/s10922-010-9188-3.
16. Jiang, W.; Schotten, H.D. Deep Learning for Fading Channel Prediction. *IEEE Open J. Commun. Soc.* **2020**, *1*, 320–332.
17. Sui, Y.; Yu, W.; Luo, Q. Jointly Optimized Extreme Learning Machine for Short-Term Prediction of Fading Channel. *IEEE Access* **2018**, *6*, 49029–49039.
18. Wang, X.; Li, H.; Wu, Q. Optimizing Adaptive Coding and Modulation for Satellite Network with ML-based CSI Prediction. In Proceedings of the 2019 IEEE Wireless Communications and Networking Conference (WCNC), Marrakesh, Morocco, 15–18 April 2019; pp. 1–6.
19. Konstantinov, A.S.; Pestryakov, A.V. Fading Channel Prediction for 5G. In Proceedings of the 2019 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO), Yaroslavl, Russia, 1–3 July 2019; pp. 1–7.
20. Sapavath, N.N.; Rawat, D.B.; Song, M. Machine Learning for RF Slicing using CSI Prediction in Software Defined Large-scale MIMO Wireless Networks. *IEEE Trans. Netw. Sci. Eng.* **2020**, doi:10.1109/TNSE.2020.2993984.
21. Dong, P.; Zhang, H.; Li, G.Y. Machine Learning Prediction Based CSI Acquisition for FDD Massive MIMO Downlink. In Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, UAE, 9–13 December 2018; pp. 1–6.
22. Khan, H.; Butt, M.M.; Samarakoon, S.; Sehier, P.; Bennis, M. Deep Learning Assisted CSI Estimation for Joint URLLC and eMBB Resource Allocation. In Proceedings of the 2020 IEEE International Conference on Communications Workshops (ICC Workshops), Dublin, Ireland, 7–11 June 2020; pp. 1–6.
23. Mishra, P.; Sonali; Dixit, A.; Jain, V.K. Machine Learning Techniques for Channel Estimation in Free Space Optical Communication Systems. In Proceedings of the 2019 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), Goa, India, 16–19 December 2019; pp. 1–6.
24. Lau, V.; Liu, Y.; Chen, T. On the design of MIMO block-fading channels with feedback-link capacity constraint. *IEEE Trans. Commun.* **2004**, *52*, 62–70.
25. Xie, Y.; Li, Z.; Li, M. Precise Power Delay Profiling with Commodity WiFi. In Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom '15), Paris, France, 7–11 September 2015; ACM: New York, NY, USA, 2015; pp. 53–64, doi:10.1145/2789168.2790124.
26. Schulz, M.; Link, J.; Gringoli, F.; Hollick, M. Shadow Wi-Fi: Teaching Smartphones to Transmit Raw Signals and to Extract Channel State Information to Implement Practical Covert Channels over Wi-Fi. In Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '18), Munich, Germany, 10–15 June 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 256–268, doi:10.1145/3210240.3210333.
27. Verstraeten, D.; Schrauwen, B.; D'Haene, M.; Stroobandt, D. An experimental unification of reservoir computing methods. *Neural Netw.* **2007**, *20*, 391–403, doi:10.1016/j.neunet.2007.04.003.
28. Lukoševičius, M.; Jaeger, H. Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.* **2009**, *3*, 127–149, doi:10.1016/j.cosrev.2009.03.005.
29. Maass, W.; Natschläger, T.; Markram, H. Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations. *Neural Comput.* **2002**, *14*, 2531–2560, doi:10.1162/089976602760407955.
30. Jaeger, H.; Haas, H. Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication. *Science* **2004**, *304*, 78–80.
31. Gallicchio, C.; Micheli, A. Architectural and Markovian factors of echo state networks. *Neural Netw.* **2011**, *24*, 440–456, doi:10.1016/j.neunet.2011.02.002.
32. Yildiz, I.B.; Jaeger, H.; Kiebel, S.J. Re-visiting the echo state property. *Neural Netw.* **2012**, *35*, 1–9, doi:10.1016/j.neunet.2012.07.005.

33. Manjunath, G.; Jaeger, H. Echo State Property Linked to an Input: Exploring a Fundamental Characteristic of Recurrent Neural Networks. *Neural Comput.* **2012**, *25*, doi:10.1162/NECO\_a\_00411.
34. Massar, M.; Massar, S. Mean-field theory of echo state networks. *Phys. Rev. E* **2013**, *87*, 042809, doi:10.1103/PhysRevE.87.042809.
35. Tiño, P. Fisher memory of linear Wigner echo state networks. In Proceedings of the 25th European Symposium on Artificial Neural Networks (ESANN 2017), Bruges, Belgium, 26–28 April 2017.
36. Gallicchio, C.; Micheli, A.; Tiño, P. Randomized Recurrent Neural Networks. In Proceedings of the 26th European Symposium on Artificial Neural Networks (ESANN 2018), Bruges, Belgium, 25–27 April 2018.
37. Gallicchio, C.; Micheli, A. Deep Echo State Network (DeepESN): A Brief Survey. *arXiv* **2017**, arXiv:abs/1712.04323.
38. Gallicchio, C.; Micheli, A.; Pedrelli, L. Deep reservoir computing: A critical experimental analysis. *Neurocomputing* **2017**, *268*, 87–99, doi:10.1016/j.neucom.2016.12.089.

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).