

Received 15 Jul. 2020, Accepted 10 Aug. 2020, Date of publication 12 Aug., 2020, date of current version 12 Aug. 2020.

Digital Object Identifier (DOI): 10.46470/03d8ffbd.80623473

Massive MIMO Channel Prediction Using Recurrent Neural Networks

JOEL PONCHA LEMAYIAN¹, JEHAD M. HAMAMREH²

¹Department of Electrical and Computer Engineering, Antalya Bilim University, Antalya, Turkey (e-mail: lemajian.joel@std.antalya.edu.tr).

²Department of Electrical and Electronics Engineering, Antalya Bilim University, Antalya, Turkey (e-mail: jehad.hamamreh@antalya.edu.tr).

Authors are from WISLAB for Innovations in Wireless Communication Computation laboratory.

Corresponding author: Joel P. Lemayian (e-mail: lemajian.joel@std.antalya.edu.tr).

This work is funded by the scientific and technological research council of Turkey (TÜBİTAK) under grand 119E392. All the codes used can be found at researcherstore.com.

ABSTRACT Massive MIMO has been classified as one of the high potential wireless communication technologies due to its unique abilities such as high user capacity, increased spectral density, and diversity among others. Due to the exponential increase of connected devices, these properties are of great importance for the current 5G-IoT era and future telecommunication networks. However, outdated channel state information (CSI) caused by the variations in the channel response due to the presence of highly mobile and rich scattering is a major problem facing massive MIMO systems. Outdated CSI occurs when the information obtained about the channel at the transmitter changes before transmission. This leads to performance degradation of the network. In this work, we demonstrate a low complexity channel prediction method using neural networks. Specifically, we explore the power of recurrent neural network utilizing long-short memory cells in analyzing time series data. We review various neural network-based channel prediction methods available in the literature and compare complexity and performance metrics. Results indicate that the proposed methods outperform conventional systems by tremendously lowering the complexity associated with channel prediction.

INDEX TERMS Massive MIMO, Neural networks, RNN, Artificial Intelligence, Channel state information.

I. INTRODUCTION

Multiple input multiple output (MIMO) employs multiple antennas at the transmitter and/or receiver. This technology has highly desired properties such as high throughput, high spectral efficiency, and multiplexing gains [1]. MIMO has evolved from a mere research concept to a real-world application and has been integrated into state-of-the-art wireless network standards such as IEEE 802.11n, 3GPP long-term evolution (LTE) and LTE-Advanced (E-UTRA) [3]. In a massive MIMO (mMIMO) system, the number of antennas in MIMO increases to hundreds. mMIMO has been classified as one of the high potential wireless communication technologies with the ability to have high user capacity [4], which is a key requirement for 5G-IoT and beyond technologies [6] [7] [16].

Nevertheless, numerous challenges are facing mMIMO and the general wireless communication paradigms. This

is due to the population increase of users which causes increases in the number of connected devices [8] [9] [10] [11]. Outdated channel state information (CSI) caused by the variations in the channel response due to the presence of highly mobile and rich scattering is a major problem facing mMIMO systems [4]. Outdated CSI occurs when the information obtained about the channel changes before transmission. Proposed solutions in the literature to combat the outdated CSI problem are mainly divided into passive and sub-optimal methods [12]. Passive method passively compensates for the performance loss at the cost of wireless resources (frequency, time, power), while sub-optimal methods assumes imperfect CSI as a constraint and aims to acquire only partial performance [13].

For example, in time division duplex (TDD) systems the channel is reciprocal, where the same frequency is used for both uplink and downlink. Therefore the downlink

channel is estimated from the uplink channel in conventional TDD systems [19]. Nevertheless, the increase in the number of devices caused by the exponential human population increase, and the IoT era where everything is connected to everything has forced engineers to use ultra-high bands including millimeter and terahertz in wireless communication [5]. Channel coherence time is significantly reduced in ultra-high bands thus becoming shorter than the pilot transmission time. Consequently, the conventional uplink-downlink channel state acquisition method will provide an outdated CSI resulting in significant system performance degradation. In light of this, an efficient and reliable channel state prediction system is in demand.

Moreover, considering channel modeling in mMIMO channel prediction, Rayleigh distance between the transmitter and receiver in MIMO is defined as $2L^2/\lambda$, where L is the antenna dimension and λ is the carrier wavelength [15]. Unlike MIMO, mMIMO has a large number of antennas, which may cause the distance between the receiver and transmitter to be shorter than Rayleigh distance. Therefore, the far-field assumption defined in [20] cannot be used and in turn, we use the near field assumption defined in [21]. Moreover, non-stationarity is also considered in mMIMO due to the changing antennas and the varying physical environment [20] [21].

Neural Networks (NN) as an AI technique, is an effective recently proposed model for combating outdated CSI without wasting resources [14]. It is highly valued because it can avoid parameter estimation due to its data-driven nature. Channel prediction is viewed as a revolutionary future technology and hence it has attracted the attention of many researchers [4]. Additionally, instantaneously selecting transmit parameters (enabled by channel prediction) such as transmit power, coding rate, transmit antennas, and carrier frequency, depending on the instantaneous condition of the channel using channel prediction will tremendously improve the performance of adaptive wireless communications systems (Which are the future of effective wireless communication).

In this work, we concentrate on implementing channel state prediction in mMIMO using artificial intelligence, specifically recurrent neural networks (RNNs). Fig. 1 depicts a mMIMO BS with hundreds of antennas and multiple users.

RNNs are very effective when processing time series data. Since channel response data is closely related to time series data, we look at mMIMO channel prediction using RNNs as a technology with great future potential that will have a major impact in wireless technology. Moreover, we compare performance metrics of conventional CSI prediction processes with RNN-based prediction. lastly, an RNN model utilizing LSTMs is designed. The main contributions of this work are:

- 1) Propose a low cost mMIMO RNN-based CSI predic-

tor.

- 2) Provide quick answers about RNN-based mMIMO CSI prediction.
- 3) Demonstrate performance metrics between conventional CSI predictors and RNN-based predictors, in terms of complexity and cost.
- 4) Develop a mMIMO channel prediction method for 128 transmit antennas.
- 5) Review recent and common neural network channel prediction schemes.

The remainder of this work is organized as follows. In Section II we discuss the literature review related to this topic. In Section III we present the mMIMO system model. Section IV discusses about commonly used machine learning MIMO channel prediction models and their limitations, while Section V explains about RNN-based mMIMO channel prediction process using the RNN predictor. Section IV discusses about the mMIMO channel prediction process. In Section VII we discuss the simulation results and finally, the conclusion is presented in Section VIII.

TABLE 1. NOMENCLATURES

Abbreviations	Definitions
ACM	adaptive coded modulation
AI	artificial intelligence
AOA	angle of arrival
AOD	angle of departure
AP	access point
BER	bit error rate
BP	back-propagation
BS	Base Station
CNN	convolutional neural networks
CSI	channel state information
DL	downlink
ESPRIT	estimation of signal parameters via rotational invariant techniques
FDD	frequency division duplex
IEP	interval of effective prediction
LSTM	long short-term memory
LTE	long-term evolution
MIMO	multiple input multiple Output
MUSIC	multiple signal classification
NMSE	normalized mean square error
OFDM	orthogonal frequency division multiplexing
RNN	recurrent neural network
SCNet	sparse complex-valued neural network
TDD	time division duplex
UL	uplink
ULA	uniform liner array

II. RELATED WORKS

According to [18], frequency division duplexing (FDD) based networks is superior compared to TDD-based networks due to its low latency and high anti-interference properties. Nevertheless, its computation and feedback challenges for predicting downlink channel state information (DL-CSI) are the main constraints for advancing performance in FDD cellular networks. To solve this problem, the authors in [18] propose the use of deep

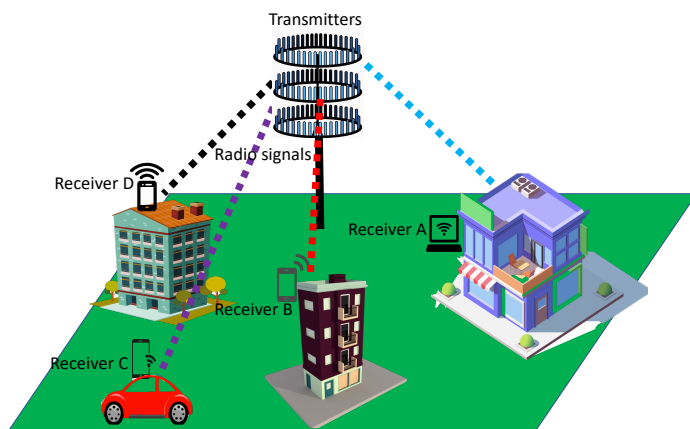


FIGURE 1. Massive multiple input multiple output (mMIMO) model.

learning convolutional Long Short-Term Memory network (convLSTM-net) to predict DL-CSI using uplink channel state information (UL-CSI). The proposed convLSTM-net consists of two modules. The first one is the feature extraction module, responsible for learning spatial and temporal correlations between DL-CSI and UL-CSI. The second part is the prediction model which maps the extracted features to the reconstruction of DL-CSI. Moreover, the authors compare the performance of convolutional neural networks (CNN) and long short-term memory networks (LSTM) to convLSTM-net. The performance simulation was further divided into two parts. In part one, the hyperparameters on the proposed convLSTM-net are analyzed to investigate their effects on the prediction performance. In part two the experiment is performed in both time and frequency domain to determine the proper environment for accurate prediction of DL-CSI. According to Wang et. al. [18], the results show that convLSTM-net outperforms both CNN and LSTM in DL-CSI prediction using UL-CSI in cellular FDD networks. However, the proposed convLSTM-net has high complexity evident from its design.

In addition, authors in [22] claim that in an FDD mMIMO system, the acquisition of downlink CSI is a complex task due to the overheads required for downlink training and uplink feedback. The authors propose a sparse complex-valued neural network (SCNet) system used to map uplink to downlink. Unlike the previous system, this paradigm is modeled in the complex domain and can learn the complex-valued mapping function by off-line training. According to the authors, numerical results suggest that SCNet outperforms conventional deep network-based channel prediction in terms of prediction accuracy and robustness over complex wireless channels.

Authors in [28] propose an ESPRIT-based parameter prediction model for narrow-band MIMO systems that exploit both temporal and spatial correlations in practical MIMO channels. The model estimates channel parameters by employing a vector transmit spatial signature

model and two-dimensional ESPRIT. According to the authors, the proposed scheme is suitable for both two-dimensional azimuths only and three-dimensional MIMO spatial channel models.

Authors in [23] utilize back-propagation (BP) in a multilayered neural network to model a multi-time channel prediction system. The paradigm is used to effectively predict CSI and enhance mMIMO performance, power control, and artificial noise physical layer security design. Additionally, the authors utilize a previous stopping criterion to prevent over-fitting of BP in neural networks. According to the authors, the results demonstrated by comparing the predicted normalized mean square error (NMSE), indicate that the performance of the proposed model has improved. Additionally, a sparse channel construction model used to save system resources without deteriorating the performance is proposed.

According to [24], a mMIMO channel is characterized by non-stationarity and quick variations. Therefore, using conventional methods to obtain CSI will result in an outdated CSI and consequently degrading the performance of the network. Authors in [24] propose a channel prediction algorithm used in massive MIMO. The authors propose a first-order Taylor expansion-based channel prediction model to handle channel characteristics. Moreover, a channel prediction model with estimation and prediction sections is further proposed to derive an interval of effective predictions (IEP). According to the authors, numerical simulations from the proposed algorithm indicate that a reliable channel prediction can be achieved within IEP.

By matching time-varying wireless fading channels to transmitter parameters, known as adaptive modulation, system throughput is considerably improved. Authors in [25] propose a channel prediction scheme using pilot symbol assisted modulation for MIMO Rayleigh fading channels. Moreover, the effect of the channel prediction error on the bit error rate of a transmit beam-former is analyzed. According to [25], the results obtained indicate a critical value, under which the adaptive modulator can consider the predicted channel as perfect, and above which categorical attention of the channel's imperfection must be accounted for at the transmitter.

Authors in [19] propose a split-brain autoencoder system, which is a modified regular autoencoder for learning. The paradigm divides the network into two disjoint parts. Each part performs complex channel prediction tasks. According to the authors, the method produced state-of-the-art results on several large-scale learning standards. Moreover, authors in [1] provide a comprehensive survey on the application of recurrent neural networks (RNN) in channel prediction. The complexity and performance of predictors are relatively illustrated by numerical results. Adaptive coded modulation (ACM) is a promising method used to enhance spectral efficiency in a time-varying mobile channel with no effect on the targeted bit error

rate (BER) [26]. However, the transmitter must have perfect up-to-date channel characteristics for ACM to work. Authors in [26] propose a linear fading-envelope predictor to predict the CSI. Moreover, authors in [27] propose a CSI prediction algorithm for OFDM that determines time-delays and Doppler frequencies of each propagation path. According to the authors, the method requires less feedback information and has better mean-squared error performance than previous methods.

The model proposed in this work differs from the discussed reviews in that the model strives to provide a low-cost low-complex CSI predictor with the ability to support future wireless technologies such as mMIMO and mm-Wave. Key contributions are:

- 1) Propose a low-complexity low-cost mMIMO RNN-based CSI predictor.
- 2) Demonstrate performance metrics between conventional CSI predictors and RNN-based predictors, in terms of complexity and cost
- 3) Develop and RNN-based predictor for mMIMO.

In the next section, we will discuss the mMIMO system model.

III. mMIMO CHANNEL PREDICTION SYSTEM MODEL

In a time-varying mMIMO system, there are M_T transmitter antennas and M_R receiver antennas. Where, at any instance $M_T \leq M_R$ [24]. Each antenna transmits N time slots at a given transmission, where $N \geq M_T$. Considering an instantaneous signal transmit and receive in mMIMO, the base-band receiver can be shown as:

$$\mathbf{r}(t) = \mathbf{H}(t)\mathbf{s}(t) + \mathbf{n}(t), \quad (1)$$

where $\mathbf{r}(t) = [r_1(t), \dots, r_{N_r}(t)]^T$ is an $N_r \times 1$ vector of the receive signal at time t (N_r is the number of receive antennas) and $\mathbf{s}(t) = [s_1(t), \dots, s_{N_t}(t)]^T$ is an $N_t \times 1$ vector of the transmit signal at time t (N_t is the number of transmit antennas). $\mathbf{H}(t) = [h_{n_r n_t}(t)]_{N_r \times N_t}$ is the matrix of continuous channel impulse response and $h_{n_r n_t} \in \mathbb{C}^{1 \times 1}$ is the flat fading channel gain between transmitter antenna n_t and receiver antenna n_r . Moreover, $1 \leq n_r \leq N_r$ and $1 \leq n_t \leq N_t$. Due to the multipath fading, feedback and processing delays the obtained CSI at the transmitter may be outdated before it can be used. That is, $\mathbf{H}(t) \neq \mathbf{H}(t+\tau)$, consequently, this will lead to performance degradation of adaptive communication systems [3]. The goal of channel prediction is to estimate $\mathbf{H}(t+\tau)$ at time t to be as close as possible to the actual value at $(t+\tau)$. That is $\mathbf{H}(t+\tau) \rightarrow \hat{\mathbf{H}}(t+\tau)$. Closely related MIMO prediction techniques are briefly discussed in the subsequent section.

IV. RELATED PREDICTOR MODELS

Apart from RNN-base predictors, several other methods have been proposed for mMIMO CSI prediction. Parametric and autoregressive channel prediction models are the

most popular techniques [1]. In this section, the models are briefly discussed to provide the reader with a clear distinction to the proposed work.

A. PARAMETRIC PREDICTOR MODEL

As stated by [16], [17], a single antenna channel is represented by overlaying a set of complex sinusoids in popular multipath fading models.

$$h(t) = \sum_{p=1}^P \alpha_p e^{j(\omega_p t + \phi_p)}, \quad (2)$$

where α_p is the complex amplitude, ω_p is the Doppler frequency shift in radians of the p^{th} sinusoid, and ϕ_p is the phase. $j^2 = -1$ denotes complex units and P represents the total number of scattered sinusoids. The single-antenna system depicted in equation (2) can be modeled to represent a MIMO propagation model shown by equation (3), by introducing spatial dimension parameters.

$$\mathbf{H}(t) = \sum_{p=1}^P \alpha_p \mathbf{a}_r(\theta_p) \mathbf{a}_t^T(\psi_p) e^{j(\omega_p t + \phi_p)}, \quad (3)$$

where θ_p and ψ_p are the angle of arrival (AOA) and angle of departure (AOD) respectively. \mathbf{a}_r and \mathbf{a}_t are the response vector of the receiver and transmitter antenna arrays respectively. \mathbf{a} can be represented as a uniform linear array (ULA) with M equally spaced antenna elements as follows:

$$\mathbf{a}(x) = [1, e^{-j(\frac{2\pi}{\lambda})d \sin(x)}, \dots, e^{-j(\frac{2\pi}{\lambda})(M-1)d \sin(x)}]^T, \quad (4)$$

where x can either be the angle of arrival or departure, λ is the wavelength of the sub-carrier frequency, and d is the distance between antennas. According to [1], multipath parameters change slowly compared to the channel fading rate. Therefore, future CSI up to a certain period can be obtained by simply extrapolating the known multipath parameters. Hence, channel prediction in mMIMO using equation (13) is reduced to parameter prediction. That is, a parameter prediction model to predict the total number of scatters, the angle of arrival and departure, and the Doppler shift for each path (i.e. $\{\hat{\alpha}_p, \hat{\omega}_p, \hat{\theta}_p, \hat{\psi}_p\}_{p=1}^{\hat{P}}$). In the following section, we model a prediction procedure for the mentioned parameters.

1) parametric model prediction procedure

1) Step 1:

We define k independent discrete-time channels $\{\mathbf{H}[k] | k = 1, \dots, K\}$, sampled from the continuous channel response $\mathbf{H}(t)$. We can therefore model a large matrix containing all the required translational invariance structure in all dimensions. According to [28], a block-Hankel matrix with dimensions $N_r Q \times N_t S$ is represented as follows:

$$\hat{\mathbf{D}} = \begin{bmatrix} \mathbf{H}[1] & \mathbf{H}[2] & \dots & \mathbf{H}[S] \\ \mathbf{H}[2] & \mathbf{H}[3] & \dots & \mathbf{H}[S+1] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}[Q] & \mathbf{H}[Q+1] & \dots & \mathbf{H}[K] \end{bmatrix}, \quad (5)$$

where Q is the size of the Hankel matrix and $S = K - Q + 1$. We use equation (5) to calculate a Spatio-temporal covariance matrix $\hat{\mathbf{C}}$.

$$\hat{\mathbf{C}} = \frac{\hat{\mathbf{D}}\hat{\mathbf{D}}^H}{N_r S}, \quad (6)$$

where $(\cdot)^H$ represents the Hermitian conjugate transpose.

2) Step 2:

We estimate the dominant scattering sources \hat{P} using the minimum description length (MDL) system described by [29].

$$\hat{P} = \underset{z=1, \dots, (N_r Q - 1)}{\text{arg min}} \left[S \log(\lambda_z) + \frac{(z^2 + z) \log(S)}{z} \right], \quad (7)$$

where λ_z represents the z^{th} eigenvalue of $\hat{\mathbf{C}}$.

3) Step 3:

Algorithms such as multiple signal classification (MUSIC) and estimation of signal parameters by rotational invariance techniques (ESPRIT) to find $\{\hat{\omega}_p, \hat{\theta}_p, \hat{\psi}_p\}_{p=1}^{\hat{P}}$ from $\hat{\mathbf{C}}$.

4) Step 4:

Now that we have $\{\hat{\omega}_p, \hat{\theta}_p, \hat{\psi}_p\}_{p=1}^{\hat{P}}$, we calculate $\{\hat{\alpha}_p\}_{p=1}^{\hat{P}}$ by substituting the former parameters to equation (13) and obtain the equation below.

$$\mathbf{H}(\tau) = \sum_{p=1}^{\hat{P}} \alpha_p \mathbf{a}_r(\theta_p) \mathbf{a}_t^T(\psi_p) e^{j(\omega_p \tau + \phi_p)}, \quad (8)$$

where τ is the time steps predicted.

From the process described above, it is evident that this model experiences some constraints. The estimation process is tedious and has high complexity due to the manipulation of high order matrices. Moreover, equations (3) and (4) are highly dependent on the type of array used. That is, if a different kind of array is used then equations (3) and (4) must be adjusted accordingly. Finally, the obtained prediction outdates faster especially in a continuously changing environment compared to a constant environment.

B. AUTOREGRESSIVE (AR) PREDICTOR MODEL

The mMIMO time varying channel can alternatively be formulated using an autoregressive process where Kalman filters (KF) are used to compute AR coefficients used to build a liner predictor used to predict future CSI using current and past CSI [1]. the AR predictor for mMIMO can be represented as:

$$\hat{\mathbf{H}}(t-1) = \sum_{p=1}^P \mathbf{A}_p \circ \mathbf{H}[t-p+1], \quad (9)$$

where $\mathbf{A}_p = \{a_{n_r, n_t}^p\}$ is an $n_r \times n_t$ AR coefficient matrix, such that a_{n_r, n_t}^p is the p^{th} AR coefficient of the channel between transmitter n_t and receiver n_r . Other predictors proposed in literature include maximum-likelihood (ML) estimation, least-square (LS) estimation, and minimum-mean-square-error (MMSE) estimation

CSI prediction using the above models is faced by a number of challenges such as high complexity, low accuracy, lack of generality, single-step prediction limitation, and unreliability, hence such methods are only suitable for small scale estimation [2]. CSI prediction process in RNNs is summarised in the subsequent section.

V. Proposed method: RECURRENT NEURAL NETWORKS (RNNS)

In this section, we will discuss the RNN model which is a powerful machine learning technique that has shown great potential in predicting time series data. RNNs are superior because they not only use training data for learning but also learn from historical data of past events. There are different models of RNNs, Fig. 3 depicts the commonly known Jordan network. A simplified RNN network consists of an input layer with N_i neurons, a hidden layer with N_h neurons, and an output layer with N_o outputs. Each connection between the input layer, the hidden layer, and the output layer is assigned a weighted value. Let w_{ln} denote the weight between the l^{th} input and the n^{th} hidden neuron, and v_{ol} represent the weight of the l^{th} hidden neuron and the o^{th} output neuron. Such that $1 \leq n \leq N_i$, $1 \leq l \leq N_h$, and $1 \leq o \leq N_o$. Therefore, we can construct an $N_h \times N_i$ matrix \mathbf{W} of weights shown by (10).

$$\mathbf{W} = \begin{bmatrix} w_{11} & \dots & w_{1N_i} \\ \vdots & \ddots & \vdots \\ w_{N_h 1} & \dots & w_{N_h N_i} \end{bmatrix}. \quad (10)$$

The input activation vector is denoted as $\mathbf{x}(t) = [x_1(t), \dots, x_{(N_i)}(t)]^T$ while the recurrent or feedback component is modeled as: $\mathbf{f}(t) = [f_1(t), \dots, f_{(N_i)}(t)]^T$. Therefore, the input to the hidden layer can be represented by the following equation:

$$\mathbf{z}_h(t) = \mathbf{W}\mathbf{x}(t) + \mathbf{f}(t) + \mathbf{b}_h, \quad (11)$$

where $\mathbf{b}_h = [b_1^h, \dots, b_{(N_h)}^h]^T$ represents the bias in the hidden layer. In addition, we use a matrix \mathbf{F} to map the previous output $y(t-1) = [y_1(t-1), \dots, y_{N_o}(t-1)]^T$ to the current input component. Hence,

$$\mathbf{f}(t) = \mathbf{F}\mathbf{y}(t-1). \quad (12)$$

The behavior of a neuron network is determined by an activation function (AF). The activation function introduces some nonlinearity characteristics to the system,

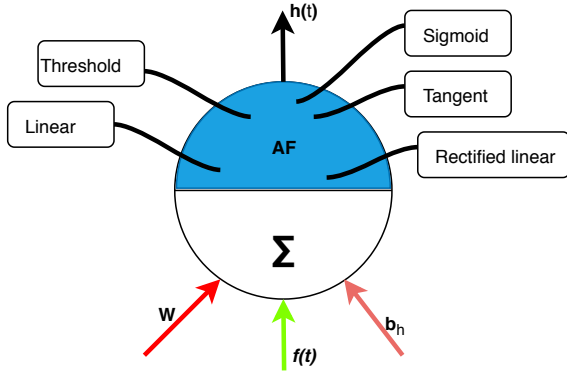


FIGURE 2. Activation functions for a neural network designed to introduce non-linearity.

which means the output is not simply a constant scaling of the input (i.e. the rate of change is not constant across all independent variables). Hence enabling it to solve complex problems. Common activation functions include linear, rectified linear hyperbolic tangent, and leaky rectified linear (see Fig. 2). In this work, we use the sigmoid function to introduce nonlinearity. The sigmoid function is defined as:

$$S(x) = \frac{1}{1 + e^x}. \quad (13)$$

Substituting equations (11) and (21) into equation (13), we get the following equation:

$$\mathbf{h}(t) = S(\mathbf{z}_h(t)) = S(\mathbf{w}\mathbf{x}(t) + \mathbf{F}\mathbf{y}(t-1) + \mathbf{b}_h). \quad (14)$$

$\mathbf{z}_h(t)$ is of size $N_h \times 1$, therefore, equation (14) is executed in an element wise operation. That is, $S(\mathbf{z}_h) = [S(z_1), \dots, S(z_{N_h})]^T$. A matrix \mathbf{V} of weights with dimensions $N_o \times N_h$ exists between the hidden and the output layer. The input to the output layer therefore becomes, $\mathbf{z}_o(t) = \mathbf{V}\mathbf{h}(t) + \mathbf{b}_y$ where \mathbf{b}_y is the bias matrix for the output layer of size $N_o \times 1$. The output equation is derived as:

$$\mathbf{y}(t) = S(\mathbf{z}_o(t)) = S(\mathbf{V}\mathbf{h}(t) + \mathbf{b}_y). \quad (15)$$

However, RNNs suffers from vanishing and exploding gradient problems. The vanishing gradient problem is experienced during back-propagation. This is where the partial derivative of the loss function with respect to the current weight progressively diminishes during back-propagation and hence has no effect on the weights when performing gradient descent. On the other hand, the exploding gradient is experienced when large gradient errors accumulate causing large updates on the network weights during training. LSTM (long short term memory) networks are special types of RNNs that use gates to overcome the problems experienced by conventional RNNs. The gates in LSTM networks include input, output, and forget gates, they facilitate better control of gradient flow and prevention of long-range dependencies. In this work, we utilize LSTMs as shown in Fig. 4. Equations that

describe the functionality of an LSTM unit cell are given below.

$$i_t = \sigma(W^{(i)}x_t + U^{(i)}h_{t-1}), \quad (16)$$

$$f_t = \sigma(W^{(f)}x_t + U^{(f)}h_{t-1}), \quad (17)$$

$$o_t = \sigma(W^{(o)}x_t + U^{(o)}h_{t-1}), \quad (18)$$

$$\hat{c}_t = \tanh(W^{(c)}x_t + U^{(c)}h_{t-1}), \quad (19)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \hat{c}_t, \quad (20)$$

$$h_t = o_t \circ \tanh(c_t). \quad (21)$$

\circ denotes element-wise multiplication. W denotes the recurrent connection between the previous and the current hidden layers. U is the weight matrix between the current input and the hidden layer. \hat{C} is calculated based on the current input and the previous hidden state and is a hidden state candidate. C is the internal memory of the cell, which is the sum of *previous_memory* \circ *forget_gate* and *newly_computed_hidden_state* \circ *input_gate*.

The first operation done in an LSTM layer is to decide whether the information is kept or discarded. This operation is done in the forget gate. A number between 0 and 1 is produced, where 1 means keep the information, while 0 means completely forget. The next step is to decide which new information will be kept in the cell state. This process is done in two parts, first, a sigmoid function called the input gate determines which values will be updated, then a tanh layer determines candidate cell states to be added. Next, the previous cell state C_{t-1} needs to be updated, that is $C_t = f_t \times C_{t-1} + i_t \times \hat{C}_t$. where f_t makes the old state forget information discarded in the forget gate. Finally, the output of the cell is modeled by the output gate which is a sigmoid layer that determines which part of the cell state will be output. The cell state is passed through a layer of tanh which forces the cell state to be between -1 and 1, then the output of the output gate is multiplied by the output of the tanh layer to get the output of the cell.

The use of the forget gate in an LSTM cell helps it to decide which information needs to be discarded and which information is to be used to update the models parameters at each time step. Hence, this helps the model prevent against vanishing and exploding gradient. To understand this better, lets say that the error vanishing gradient ∂E with respect to some weight W at some time step $k < T$ is:

$$\sum_{t=1}^k \frac{\partial E_t}{\partial W} \rightarrow 0, \quad (22)$$

then for the gradient not to vanish, a suitable parameter is found for the next time step such that:

$$\frac{\partial E_{k+1}}{\partial W} \rightarrow 0. \quad (23)$$

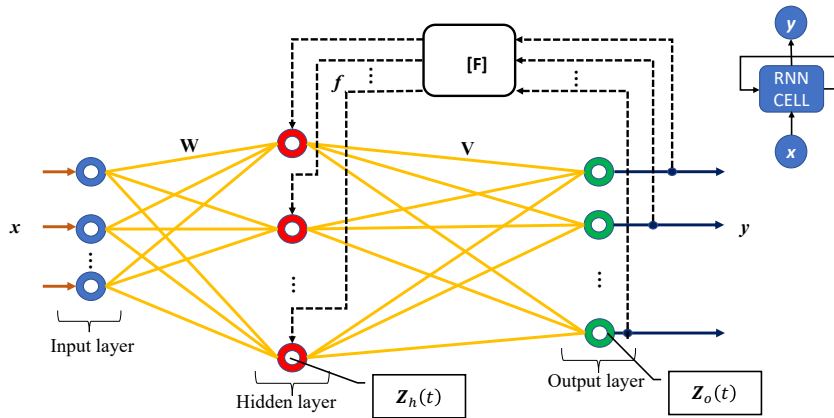


FIGURE 3. Conventional recurrent neural network system.

The presence of the activation vector in the forget gate allows it to find such a parameter.

The data set used is divided into two, training and prediction data sets. The training data set is used to train the model through backpropagation (BP). The model takes in the input x and the preferred output y then calculates the cost $C = ||y - x||^2$. The error is then propagated back through the network, causing the weights to iteratively adjust until convergence is obtained and a minimum cost is achieved. To commence prediction using LSTMs, the initialization process is described below.

1) LSTM INITIALIZATION AND PREDICTION PROCEDURE

- 1) Randomly select W, V, b_h, b_y .
- 2) Define sequence length. This is the number of historical data considered when performing the prediction.
- 3) Set feature period predict. This defines the time units predicted into the future.
- 4) Set epoch. This is the number of times that entire data points are passed through the network.
- 5) Set batch size. Number of grouped samples from the total samples processed at a time.
- 6) Preprocess the data. Data is scaled and cleaned, i.e. NaNs and empty rows are removed.
- 7) Split data into validation and train data sets.
- 8) Create an LSTM model.

VI. mMIMO CSI PREDICTION PROCESS USING RNNs

The data set h is divided into two, training (60%) and testing/prediction (40%) data sets. The training data set is used to train the model through backpropagation (BP). Referring to Fig. 4, the model takes in the input $x = h(t)$ and the preferred output $y = h(t+D)$, where D is the steps to be predicted into the future. The model then calculates the cost C , using mean squared error (MSE):

$$C = \frac{1}{T} \sum_{t=1}^T ||\hat{h}(t+D) - h(t+D)||^2, \tag{24}$$

T is the total number of channel samples. The error is then propagated back through the network, causing the weights to iteratively adjust through gradient descent until convergence is obtained and a minimum cost is achieved. Initial parameters such as C_{t-1} and h_{t-1} are generally initialised with zero values, while the weights are randomly initialized. The testing set is then used to measure the accuracy of the network. A 2-layer mMIMO RNN-based model predictor utilizing LSTM was trained using 200,000 samples of h from 128 antenna generated with Rayleigh distribution. In this simulation, Python 3.0 was used and the properties of Keras used as a deep learning framework were version 2.3.1, Keras-Applications 1.0.8, and Keras-Preprocessing 1.1.0. A summary table of the utilized network indicating the number of parameters and the network topology is depicted in Table 2. Adam optimizer was employed with an initial learning rate of 0.01 and a batch size of 256. Moreover, we use dropout regularization to prevent overfitting.

TABLE 2. Proposed LSTM network properties

Layer (type)	Output Shape	Param No.
lstm 10 (LSTM)	(None, 128, 32)	4480
dropout 1 (Dropout)	(None, 128, 32)	0
lstm 11 (LSTM)	(None, 16)	3136
dropout 2 (Dropout)	(None, 16)	0
dense 3 (Dense)	(None, 64)	1088
Total params :	8,704	
Trainable params :	8,704	
Non-trainable params :	0	

VII. RESULTS

Fig. 5 indicates characteristics of the mMIMO channel, where a and b are real and imaginary parts of the channel respectively. As it can be observed, the channel is near to a random signal due to the random nature of the environment. However, we can also observe that the data is a time series data with some repetition. Fig. 6 shows

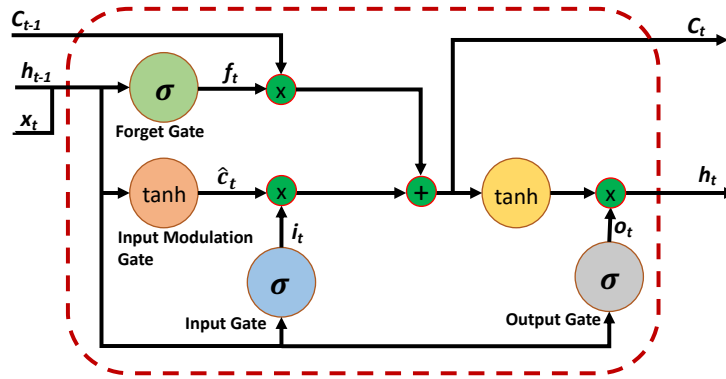


FIGURE 4. Long Short Term Memory (LSTM) unit cell.

historical and true future real part of the channel data used to train the model. Similarly, the imaginary part with the same historical and future window is fed into the model for training. Fig. 7 shows the validation and

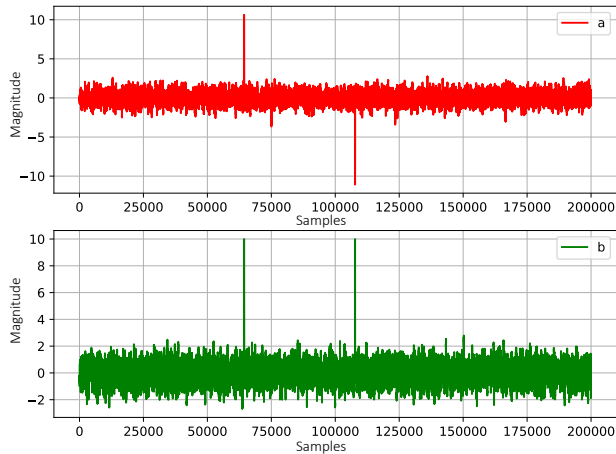


FIGURE 5. Real and imaginary parts of the channel (a is real, b is imaginary).

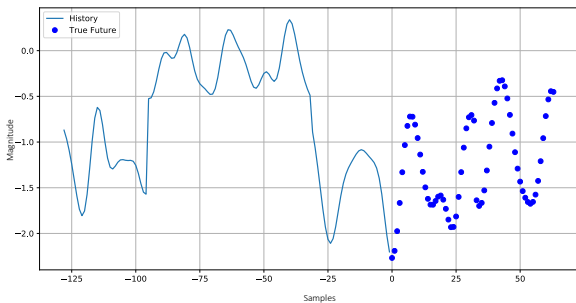


FIGURE 6. History and future true channel data for massive MIMO.

training loss after 10 epochs with a batch of 500. It can be observed that both training loss and validation loss gradually converge. This indicates that the model is neither over-fitting nor under-fitting due to the use of dropout layers in the model as it can be observed from Table 2. In this case, the accuracy of the model can be

improved by increasing the number of epoch since the losses are still on a downwards trend. Moreover, Fig. 8 and

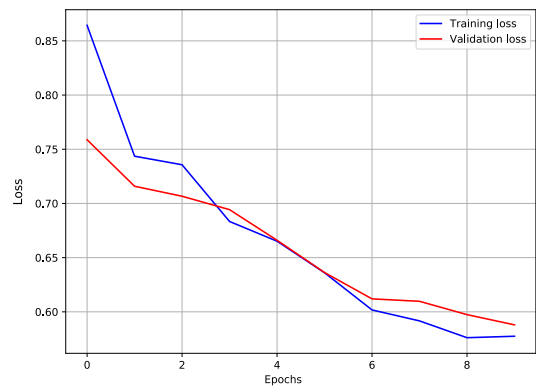


FIGURE 7. mMIMO training and validation loss.

9 shows the prediction of the real and imaginary channel parts respectively of the channel. The results show that the model can estimate the channel's characteristics with minimum complexity compared to other methods mentioned in section IV. It can also be observed from the figure that the accuracy of the predictor is not 100%. To improve the accuracy of the predictor, training data and iterations can be increased as well as LSTM layers or using different regularizers.

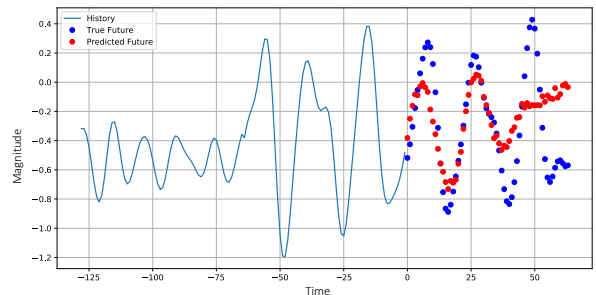


FIGURE 8. Real part prediction of the massive MIMO channel.

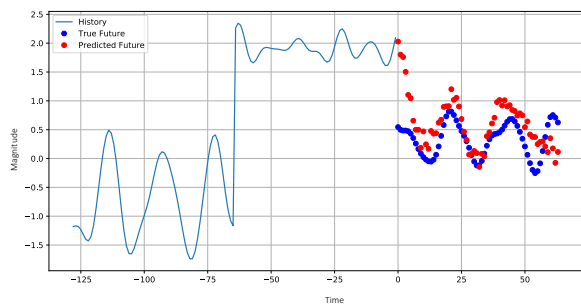


FIGURE 9. Imaginary part prediction of the massive MIMO channel.

VIII. CONCLUSION

Considering that mMIMO is a technology that has shown great potential in enhancing wireless communication in the future, this work has demonstrated that RNN-based CSI prediction is the ideal technology to boost the performance of mMIMO systems by lowering complexity and increasing accuracy. Therefore, we intent to further this work by perfecting mMIMO CSI prediction using different configurations of RNNs utilizing LSTMS or GRUs and improving the accuracy of the designed predictor.

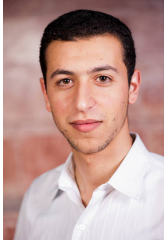
REFERENCES

- Jiang W, Schotten HD. "Neural Network-Based Fading Channel Prediction: A Comprehensive Overview." *IEEE Access* 7 (2019): 118112-118124.
- Luo C, Ji J, Wang Q, Chen X, Li P. "Channel state information prediction for 5G wireless communications: A deep learning approach." *IEEE Transactions on Network Science and Engineering* (2018).
- Zheng J, Rao BD. "Capacity analysis of MIMO systems using limited feedback transmit precoding schemes." *IEEE Transactions on Signal Processing* 56, no. 7 (2008): 2886-2901.
- Jiang W, Schotten HD. "Deep Learning for Fading Channel Prediction." *IEEE Open Journal of the Communications Society* 1 (2020): 320-332.
- Poncha LJ, Abdelhamid S, Alturjman S, Ever E, Al-Turjman F. "5G in a convergent Internet of Things era: An overview." In 2018 IEEE International Conference on Communications Workshops (ICC Workshops), pp. 1-6. IEEE, 2018.
- Al-Turjman F, Lemayian JP. "Intelligence, security, and vehicular sensor networks in internet of things (IoT)-enabled smart-cities: An overview." *Computers Electrical Engineering* 87 (2020): 106776.
- Lemayian JP, Al-Turjman F. "Intelligent IoT communication in smart environments: An overview." In *Artificial Intelligence in IoT*, pp. 207-221. Springer, Cham, 2019.
- Al-Turjman F, Lemayian JP, Alturjman S, Mostarda L. "Optimal Placement for 5G Drone-BS Using SA and GA." In *Drones in IoT-enabled Spaces*, pp. 43-58. CRC Press, 2019.
- Al-Turjman F, Lemayian JP, Alturjman S, Mostarda L. "Enhanced deployment strategy for the 5G drone-BS using artificial intelligence." *IEEE Access* 7 (2019): 75999-76008.
- LEMAYIAN JP, HAMAMREH JM. "Autonomous First Response Drone-Based Smart Rescue System for Critical Situation Management in Future Wireless Networks." (2020).
- Lemayian JP, Hamamreh JM. "First Responder Drones for Critical Situation Management." In 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 1-6. IEEE, 2019.
- Hamamreh JM, Furqan HM, Ali Z, Sidhu GA. "Enhancing the security performance of OSTBC using pre-equalicodization." In 2017 International Conference on Frontiers of Information Technology (FIT), pp. 294-298. IEEE, 2017.
- Love DJ, Heath RW, Lau VK, Gesbert D, Rao BD, Andrews M. "An overview of limited feedback in wireless communication systems." *IEEE Journal on selected areas in Communications* 26, no. 8 (2008): 1341-1365.
- Hamamreh JM, El_sallabi H, Abdallah M, Qaraqe K. "Advance in Adaptive Modulation for Fading Channels." (2013).
- Furqan HM, Hamamreh JM, Arslan H. "Secret key generation using channel quantization with SVD for reciprocal MIMO channels." In 2016 International Symposium on Wireless Communication Systems (ISWCS), pp. 597-602. IEEE, 2016.
- Hamamreh JM, Hajar A, Abewa M. "Orthogonal frequency division multiplexing with subcarrier power modulation for doubling the spectral efficiency of 6G and beyond networks." *Transactions on Emerging Telecommunications Technologies* 31, no. 4 (2020): e3921.
- Jaradat AM, Hamamreh JM, Arslan H. "OFDM With Hybrid Number and Index Modulation." *IEEE Access* 8 (2020): 55042-55053.
- Wang J, Ding Y, Bian S, Peng Y, Liu M, Gui G. "UL-CSI data driven deep learning for predicting DL-CSI in cellular FDD systems." *IEEE Access* 7 (2019): 96105-96112.
- Zhang R, Isola P, Efros AA. "Split-brain autoencoders: Unsupervised learning by cross-channel prediction." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1058-1067. 2017.
- Yaghjian A. "Approximate formulas for the far field and gain of open-ended rectangular waveguide." *IEEE Transactions on Antennas and Propagation* 32, no. 4 (1984): 378-384.
- Yaghjian A. "An overview of near-field antenna measurements." *IEEE Transactions on antennas and propagation* 34, no. 1 (1986): 30-45.
- Yang Y, Gao F, Li GY, Jian M. "Deep Learning-Based Downlink Channel Prediction for FDD Massive MIMO System." *IEEE Communications Letters* 23, no. 11 (2019): 1994-1998.
- Liao RF, Wen H, Wu J, Song H, Pan F, Dong L. "The Rayleigh fading channel prediction via deep learning." *Wireless Communications and Mobile Computing* 2018 (2018).
- Peng W, Zou M, Jiang T. "Channel prediction in time-varying massive MIMO environments." *IEEE Access* 5 (2017): 23938-23946.
- Zhou S, Giannakis GB. "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over Rayleigh MIMO channels?." *IEEE Transactions on Wireless Communications* 3, no. 4 (2004): 1285-1294.
- Oien GE, Holm H, Hole KJ. "Impact of channel prediction on adaptive coded modulation performance in Rayleigh fading." *IEEE Transactions on Vehicular Technology* 53, no. 3 (2004): 758-769.
- Wong IC, Evans BL, Evans. "Joint channel estimation and prediction for OFDM systems." In *GLOBECOM'05. IEEE Global Telecommunications Conference, 2005.*, vol. 4, pp. 5-pp. IEEE, 2005.
- Adeogun RO, Teal PD, Dmochowski PA. "Parametric channel prediction for narrowband mobile MIMO systems using spatio-temporal correlation analysis." In 2013 IEEE 78th Vehicular Technology Conference (VTC Fall), pp. 1-5. IEEE, 2013.
- Huang L, Long T, Mao E, So HC. "MMSE-based MDL method for accurate source number estimation." *IEEE Signal Processing Letters* 16, no. 9 (2009): 798-801.



JOEL P. LEMAYIAN received the B.Sc. degree in electrical and electronics engineering from Middle East Technical University Turkey, in 2017. He is presently pursuing the master's (M.Sc.) degree in electrical and computer engineering. He is currently with Antalya Bilim University, Turkey.

He has worked as a research assistant in both Middle East Technical University and Antalya Bilim University in IoT lab and Neuroscience lab respectively. He is an author of numerous journals, conference papers and book chapters. His research interests include UAVs, 5G Communication networks, Artificial Intelligence, Machine Learning, and the Internet of Things (IoT) applications.



JEHAD M. HAMAMREH received the B.Sc. degree in electrical and telecommunication engineering from An-Najah University, Nablus, in 2013, and the Ph.D. degree in electrical-electronics engineering and cyber systems from Istanbul Medipol University, Turkey, in 2018. He was a Researcher with the Department of Electrical and Computer Engineering, Texas A and M University at Qatar. He is currently an Assistant Professor with the Electrical and

Electronics Engineering Department, Antalya International (Bilim) University, Turkey.

His current research interests include wireless physical and MAC layers security, orthogonal frequency-division multiplexing multiple-input multiple-output systems, advanced waveforms design, multi-dimensional modulation techniques, IoT, 5G & 6G and orthogonal/non-orthogonal multiple access schemes for future wireless systems. He is a Regular Reviewer for various refereed journals as well as a TPC Member for several international conferences.

...