

# Channel state information estimation for 5G wireless communication systems: recurrent neural networks approach

Mohamed Hassan Essai Ali<sup>1</sup> and Ibrahim B.M. Taha<sup>2</sup>

<sup>1</sup> Department of Electrical Engineering, Faculty of Engineering, Al-Azhar University, Qena, Qena, Egypt

<sup>2</sup> Department of Electrical Engineering, College of Engineering, Taif University, Taif, Saudi Arabia

## ABSTRACT

In this study, a deep learning bidirectional long short-term memory (BiLSTM) recurrent neural network-based channel state information estimator is proposed for 5G orthogonal frequency-division multiplexing systems. The proposed estimator is a pilot-dependent estimator and follows the online learning approach in the training phase and the offline approach in the practical implementation phase. The estimator does not deal with complete *a priori* certainty for channels' statistics and attains superior performance in the presence of a limited number of pilots. A comparative study is conducted using three classification layers that use loss functions: mean absolute error, cross entropy function for  $k$ th mutually exclusive classes and sum of squared of the errors. The Adam, RMSProp, SGdm, and Adadelat optimisation algorithms are used to evaluate the performance of the proposed estimator using each classification layer. In terms of symbol error rate and accuracy metrics, the proposed estimator outperforms long short-term memory (LSTM) neural network-based channel state information, least squares and minimum mean square error estimators under different simulation conditions. The computational and training time complexities for deep learning BiLSTM- and LSTM-based estimators are provided. Given that the proposed estimator relies on the deep learning neural network approach, where it can analyse massive data, recognise statistical dependencies and characteristics, develop relationships between features and generalise the accrued knowledge for new datasets that it has not seen before, the approach is promising for any 5G and beyond communication system.

Submitted 3 May 2021  
Accepted 29 July 2021  
Published 26 August 2021

Corresponding author  
Mohamed Hassan Essai Ali,  
mhessai@azhar.edu.eg

Academic editor  
Yilun Shang

Additional Information and  
Declarations can be found on  
page 21

DOI 10.7717/peerj-cs.682

© Copyright  
2021 Essai Ali and Taha

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

**Subjects** Artificial Intelligence, Computer Networks and Communications

**Keywords** BiLSTM, Channel state information estimator, Deep learning neural networks, Loss functions

## INTRODUCTION

5G wireless communication is the most active area of technology development and a rapidly growing branch of the wider field of communication systems. Wireless communication has made various possible services ranging from voice to multimedia.

The physical characteristics of the wireless communication channel and many unknown surrounding effects result in imperfections in the transmitted signals. For example, the transmitted signals experience reflections, diffractions, and scattering, which produce multipath signals with different delays, phase shift, attenuation, and distortion arriving at

the receiving end; hence, they adversely affect the recovered signals (Oyerinde & Mneney, 2012).

*A priori* information on the physical characteristics of the channel provided by pilots is one of the significant factors that determine the efficiency of channel state information estimators (CSIEs). For instance, if not *a priori* information is available (no or insufficient pilots), channel estimation is useless; finding what you do not know is impossible. When complete information on the transmission channel is available, CSIEs are no longer needed. Thus, *a priori* uncertainty exists for communication channel statistics. However, the classical theory of detection, recognition, and estimation of signals deals with complete priory certainty for channel statistics, and it is an unreliable and unpractical assumption (Bogdanovich, Vostretsov & Electronics, 2009).

In the classic case, uncertainty is related to useful signals. In detection problems, the unknown is the fact of a signal existence. In recognition problems, the unknown is the type of signal being received at the current moment. In estimation problems, the unknown is the amplitude of the measured signal or one of its parameters. The rest of the components of the signal-noise environment in classical theory are regarded as *a priori* certain (known) as follows: the known is the statistical description of the noise, the known is the values of the unmeasured parameters of the signal and the known is the physical characteristics of the wireless communication channel. In such conditions, the classical theory allows the synthesis of optimal estimation algorithms, but the structure and quality coefficients of the algorithms depend on the values of the parameters of the signal-noise environment. If the values of the parameters describing the signal-noise environment are slightly different from the parameters for which the optimal algorithm is built, then the quality coefficients will become substantially poor, making the algorithm useless in several cases (Bogdanovich, Vostretsov & Electronics, 2009; O'Shea, Karra & Clancy, 2017). The most frequently used CSIEs are derived from signal and channel statistical models by employing techniques, such as maximum likelihood (ML), least squares (LS), and minimum mean squared error (MMSE) optimisation metrics (Kim, 2015).

One of the major concerns in the optimum performance of wireless communication systems is providing accurate channel state information (CSI) at the receiver end of the systems to detect the transmitted signal coherently. If CSI is unavailable at the receiver end, then the transmitted signal can only be demodulated and detected by a noncoherent technique, such as differential demodulation. However, using a noncoherent detection method occurs at the expense of a loss of signal-to-noise ratio of about 3–4 dB compared with using a coherent detection technique. To eliminate such losses, researchers have focused on the development of channel estimation techniques to provide perfect detection of transmitted information in wireless communication systems using the Orthogonal Frequency-Division Multiplexing (OFDM) modulation scheme (Oyerinde & Mneney, 2012).

The use of deep learning neural networks (DLNNs) is the state-of-the-art approach in the field of wireless communication. The amazing learning capabilities of DLNNs from training data sets and the tremendous progress of graphical processing units (GPUs), which are considered the most powerful tools for training DLNNs, have motivated its

usage for different wireless communication issues, such as modulation recognition (Zhou, Liu & Gravelle, 2020; Karra, Kuzdeba & Petersen, 2017) and channel state estimation and detection (Essai Ali, 2021; Joo et al., 2019; Kang, Chun & Kim, 2020; Ma, Ye & Li, 2018; Ponnaluru & Penke, 2020; Yang et al., 2019; Ye, Li & Juang, 2018). According to Karra, Kuzdeba & Petersen (2017), Kim (2015), Oyerinde & Mneney (2012), Zhou, Liu & Gravelle (2020) and Ma, Ye & Li (2018), all proposed deep learning-based CSIEs have better performance compared with the examined traditional channel ones, such as LS and MMSE estimators.

Recently, numerous long short-term memory (LSTM)- and BiLSTM-based applications have been introduced for prognostic and health management (Zhao et al., 2020), artificial intelligence-based translation systems (Wu et al., 2016; Ong, 2017) and other areas. For channel state information estimation in 5G-OFDM wireless communication systems, many deep learning approaches, such as convolutional neural network (CNN), recurrent neural network (RNN) (e.g., LSTM and BiLSTM NNs) and hybrid (CNN and RNN) neural networks have been used (Essai Ali, 2021; Liao et al., 2019; Luo et al., 2018; Ponnaluru & Penke, 2020; Yang et al., 2019; Yang et al., 2019; Ye, Li & Juang, 2018).

In Liao et al. (2019), a deep learning-based CSIE was proposed by using CNN and BiLSTM-NN for the extraction of the feature vectors of the channel response and channel estimation, respectively. The aim was to improve the channel state information estimation performance at the downlink, which is caused by the fast time-varying and varying channel statistical characteristics in high-speed mobility scenarios. In Luo et al. (2018), an online-trained CSIE that is an integration of CNN and LSTM-NN was proposed. The authors also developed an offline–online training technique that applies to 5G wireless communication systems. In Ye, Li & Juang (2018), a joint channel estimator and detector that is based on feedforward DLNNs for frequency selective channel (OFDM) systems was introduced. The proposed algorithm was found to be superior to the traditional MMSE estimation method when unknown surrounding effects of communication systems are considered. In Yang et al. (2019), an online estimator was developed by adopting feedforward DLNNs for doubly selective channels. The proposed estimator was considered superior to the traditional LMMSE estimator in all investigated scenarios. In Ponnaluru & Penke (2020), a one-dimensional CNN (1D-CNN) deep learning estimator was proposed. Under various modulation scenarios and in terms of MSE and BER metrics, the authors compared the performance of the proposed estimator with that of feedforward neural networks (FFNN), MMSE and LS estimators. 1D-CNN outperformed LS, MMSE and FFNN estimators. In Essai Ali (2021), an online pilot-assisted estimator model for OFDM wireless communication systems was developed by using LSTM NN. The conducted comparative study showed the superior performance of the proposed estimator in comparison with LS and MMSE estimators under limited pilots and *a priori* uncertainty of channel statistics. Sarwar, Shah & Zafar (2020) used the genetic algorithm-optimised artificial neural network to build a CSIE. The proposed estimator was dedicated for space–time block-coding MIMO-OFDM communication systems. The proposed estimator outperformed LS and MMSE estimators in terms of BER at high SNRs, but it achieved approximately the same performance as LS and MMSE estimators at low SNRs. Senol, Tahir & Özmen (2021)

proposed a CSIE for OFDM systems by using ANN under the condition of sparse multipath channels. The proposed estimator achieved a comparable SER performance as matching pursuit- and orthogonal matching pursuit-based estimators at a lower computational complexity than that of the examined estimators. [Le Ha et al. \(2021\)](#) proposed a CSIE that uses deep learning and LS estimator and utilizes the multiple-input multiple-output system for 5G-OFDM. The proposed estimator minimizes the MSE loss function between the LS-based channel estimation and the actual channel. The proposed estimator outperformed LS and LMMSE estimators in terms of BER and MSE metrics.

In this study, a BiLSTM DLNN-based CSIE for OFDM wireless communication systems is proposed and implemented. To the best of the authors' knowledge, this work is the first to use the BiLSTM network as a CSIE without integration with CNN. The proposed estimator does not need any prior knowledge of the communication channel statistics and powerfully works at limited pilots (under the condition of less CSI). The proposed BiLSTM-based CSIE is a data-driven estimator, so it can analyse, recognise and understand the statistical characteristics of wireless channels suffering from many known interferences such as adjacent channel, inter symbol, inter user, inter cell, co-channel and electromagnetic interferences and unknown ones ([Jeya et al., 2019](#); [Sheikh, 2004](#)). Although an impressively wide range of configurations can be found for almost every aspect of deep neural networks, the choice of loss function is underrepresented when addressing communication problems, and most studies and applications simply use the 'log' loss function ([Janocha & Czarnecki, 2017](#)). In this study two customised loss functions known as mean absolute error (MAE), and sum of squared errors (SSE) are proposed to obtain the most reliable and robust estimator under unknown channel statistical characteristics and limited pilot numbers.

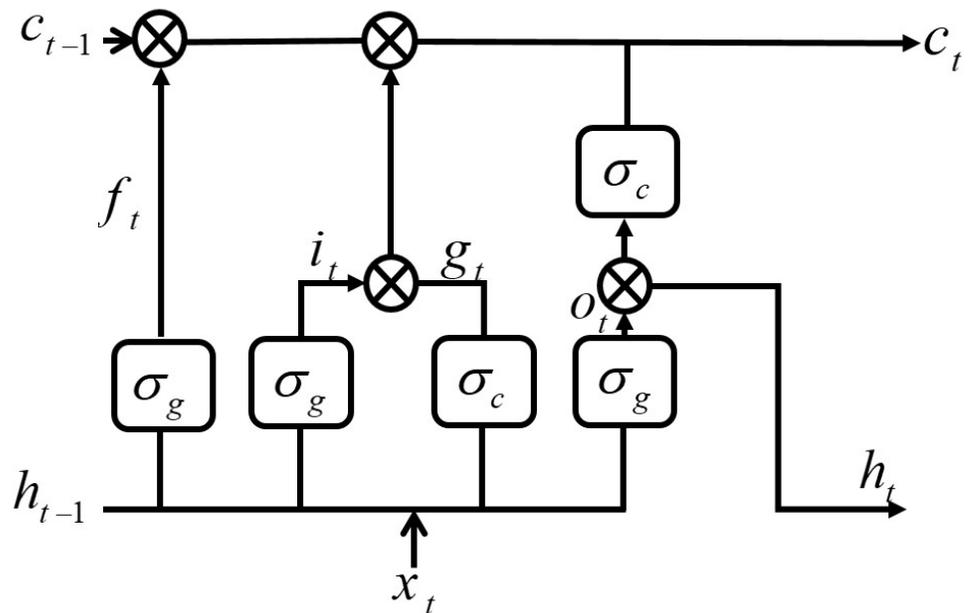
The performance of the proposed BiLSTM-based estimator is compared with the performance of the most frequently used LS and MMSE channel state estimators. The obtained results show that the BiLSTM-based estimator attains a comparable performance as the MMSE estimator and outperforms LS and MMSE estimators at large and small numbers of pilots, respectively. In addition, the proposed estimator improves the transmission data rate of OFDM wireless communication systems because it exhibits optimal performance compared with the examined estimators at a small number of pilots.

The rest of this paper is organised as follows. The DLNN-based CSIE is presented in Section II. The standard OFDM system and the proposed deep learning BiLSTM NN-based CSIE are presented in Section III. The simulation results are given in Section IV. The conclusions and future work directions are provided in Section V.

## DLNN-BASED CSIE

In this section, a deep learning BiLSTM NN for channel state information estimation is presented. The BiLSTM network is another version of LSTM neural networks, which are recurrent neural networks (RNN) that can learn the long-term dependencies between the time steps of input data ([Hochreiter & Schmidhuber, 1997](#); [Luo et al., 2018](#); [Zhao et al., 2020](#)).

The BiLSTM architecture mainly consists of two separate LSTM-NNs and has two propagation directions (forward and backward). The LSTM NN structure consists of



**Figure 1** Long short-term memory (LSTM) cell.

Full-size DOI: 10.7717/peerjcs.682/fig-1

input, output and forget gates and a memory cell. The forget and input gates enable the LSTM NN to effectively store long-term memory. Figure 1 shows the main construction of the LSTM cell (Hochreiter & Schmidhuber, 1997). The forget gate enables LSTM NN to remove the undesired information by currently used input  $x_t$  and cell output  $h_t$  of the last process. The input gate finds the information that will be used with the previous LSTM cell state  $c_{t-1}$  to obtain a new cell state  $c_t$  based on the current cell input  $x_t$  and the previous cell output  $h_{t-1}$ . Using the forget and input gates, LSTM can decide which information is abandoned and which is retained.

The output gate finds current cell output  $h_t$  by using the previous cell output  $h_{t-1}$  at current cell state  $c_t$  and input  $x_t$ . The mathematical model of the LSTMNN structure can be described through Eqs. (1)–(6).

$$i_t = \sigma_g(w_i x_t + R_i h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma_g(w_f x_t + R_f h_{t-1} + b_f) \quad (2)$$

$$g_t = \sigma_c(w_g x_t + R_g h_{t-1} + b_g) \quad (3)$$

$$o_t = \sigma_g(w_o x_t + R_o h_{t-1} + b_o) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (5)$$

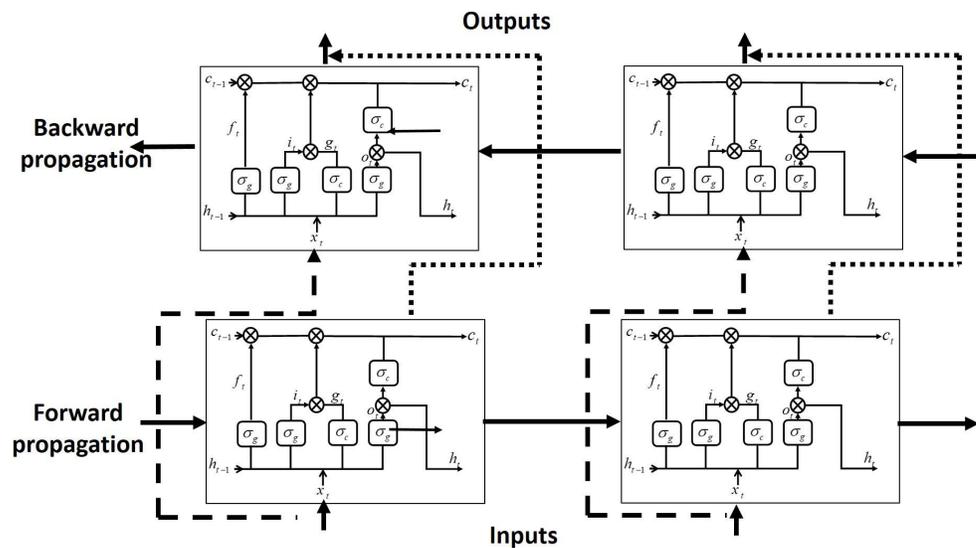


Figure 2 BiLSTM-NN architecture.

Full-size DOI: 10.7717/peerjcs.682/fig-2

$$h_t = o_t \odot \sigma_c(c_t) \quad (6)$$

where  $i, f, g, o, \sigma_c, \sigma_g$  and  $\odot$  denote the input gate, forget gate, cell candidate, output gate, state activation function (hyperbolic tangent function ( $\tanh$ ), gate activation function (sigmoid function) and Hadamard product (element-wise multiplication of vectors), respectively.  $\mathbf{W} = [w_i w_f w_g w_o]^T$ ,  $\mathbf{R} = [R_i R_f R_g R_o]^T$  and  $\mathbf{b} = [b_i b_f b_g b_o]^T$  are input weights, recurrent weights and bias, respectively.

LSTM DNN, only analyses the impact of the previous sequence in the present, disregarding information later on and failing to reach optimal performance. On the other hand BiLSTM connects the LSTM unit's output bidirectionally (forward and backward propagation directions) and capture bidirectional signals dependencies, increasing the overall model's performance.

The forward and backward propagation directions of BiLSTM are transmitted at the same time to the output unit. Therefore, old and future information can be captured, as shown in Fig. 2. At any time  $t$ , the input is fed to forward LSTM and backward LSTM networks. The final output of BiLSTM-NN can be expressed as follows:

$$h_t = \vec{h}_t \odot \vec{h}_t, \quad (7)$$

where  $\vec{h}_t$  and  $\vec{h}_t$  are forward and backward outputs of BiLSTM-NN, respectively. The operation of BiLSTM in the proposed estimator can be described briefly by the following algorithm:

- Input: sequence represents transmitted signal (original signal + channel model)
- Output: Prediction matrix of the extracted features of the input sequence
- Step 1: The forward LSTM layer receives the transmitted signal vectors from  $X$ .

```
for  $i \in \text{length}(X)$  do  
  send  $X_i$  to BiLSTM Layer  
end for
```

Step 2: Eqs. (1)–(6) are used to update the state of the LSTM cell.

Step 3: The backward LSTM layer receives the signal vectors from  $X$ , and the two previous steps are repeated.

Step 4: A hidden state sequence vector is created by splicing the forward and backward sequences of hidden layers.

Step 5: A hidden state sequence vector is sent into a full connection layer and the prediction matrix is obtained

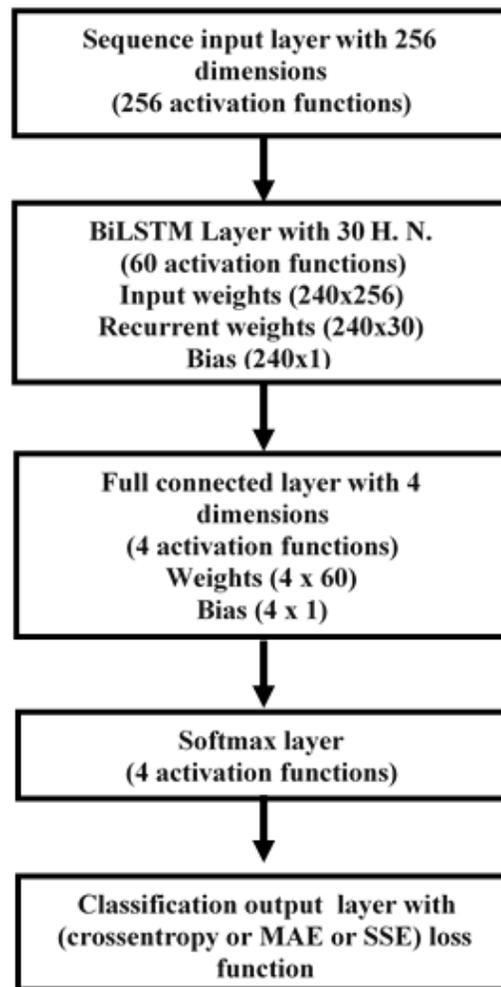
Step 6: Return the prediction matrix.

To build the DL BiLSTM NN-based CSIE, an array is created with the following five layers: sequence input, BiLSTM, fully connected, softmax and output classification. The input size was set to 256. The BiLSTM layer consists of 30 hidden units and shows the sequence's last element. Four classes are specified by considering the size 4 fully connected (FC) layer, followed by a softmax layer and ended by a classification layer. Figure 3 illustrates the structure of the proposed estimator (Essai Ali, 2021; Ye, Li & Juang, 2018).

As the proposed BiLSTM-based CSIE is built, the weights and biases of the proposed estimator are optimised (tuned) using the desired optimisation algorithm. The optimisation algorithm trains the proposed estimator by using one of three loss functions, namely, cross entropy function for  $k^{\text{th}}$  mutually exclusive classes (crossentropyex), mean absolute error (MAE), and sum of squared errors (SSE). The loss function estimates the loss between the expected and actual outcome. During the learning process, optimisation algorithms try to minimise the available loss function to the desired error goal by optimising the DLNN weights and biases iteratively at each training epoch. Figure 4 illustrates the training processes of the proposed estimator. Selecting a loss function is one of the essential and challenging tasks in deep learning. Also, investigating the efficiency of the training process using different optimization algorithms such as Adaptive Moment Estimation (Adam), Root Mean Square Propagation (RMSProp), Stochastic Gradient Descent with momentum (SGdm) (Dogo et al., 2018), and an adaptive learning rate method (Adadelta) (Zeiler, 2012). The proposed estimator is trained using above-mentioned three different loss functions and optimization algorithms to obtain the most optimal BiLSTM-based estimator for wireless communication systems with low prior information (limited pilots) for signal-noise environments.

## DL BILSTM NN-BASED CSIE FOR 5G-OFDM WIRELESS COMMUNICATION SYSTEMS

The standard OFDM wireless communication system and an offline DL of the proposed CSIE are presented in the following subsections.

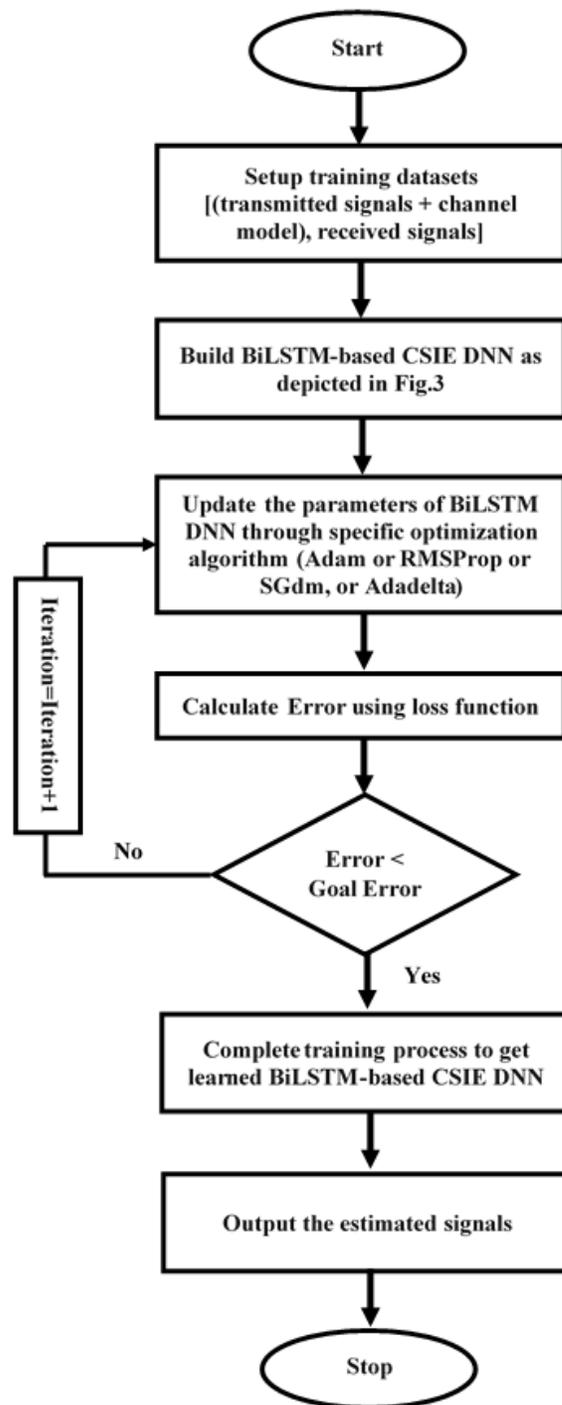


**Figure 3** Structure of the DL BiLSTM NN for the BiLSTM estimator.

Full-size  DOI: [10.7717/peerjcs.682/fig-3](https://doi.org/10.7717/peerjcs.682/fig-3)

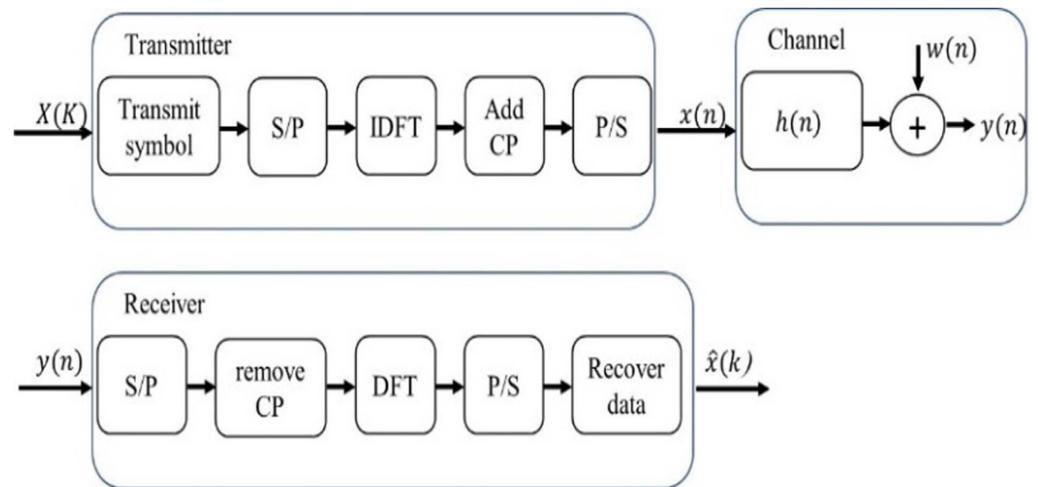
## OFDM SYSTEM MODEL

In accordance with *Essai Ali (2021)* and *Ye, Li & Juang (2018)*, *Fig. 5* clearly illustrates the structure of the traditional OFDM communication system. On the transmitter side, a serial-to-parallel (S/P) converter is used to convert the transmitted symbols with pilot signals into parallel data streams. Then, inverse discrete Fourier transform (IDFT) is applied to convert the signal into the time domain. A cyclic prefix (CP) must be added to



**Figure 4** Offline training of the BiLSTM-NN-based CSI estimator.

Full-size  DOI: [10.7717/peerjcs.682/fig-4](https://doi.org/10.7717/peerjcs.682/fig-4)



**Figure 5** Conventional OFDM system.

Full-size DOI: [10.7717/peerjcs.682/fig-5](https://doi.org/10.7717/peerjcs.682/fig-5)

alleviate the effects of inter-symbol interference. The length of the CP must be longer than the maximum spreading delay of the channel.

The multipath channel of a sample space defined by complex random variables  $\{h(n)\}_{n=0}^{N-1}$  is considered. Then, the received signal can be evaluated as follows:

$$y(n) = x(n) \oplus h(n) + w(n), \quad (8)$$

where  $\oplus x(n)$  is the input signal,  $\oplus$  is circular convolution,  $w(n)$  is additive white Gaussian noise (AWGN) and  $y(n)$  is the output signal.

The received signal in the frequency domain can be defined as

$$Y(k) = X(k)H(k) + W(k), \quad (9)$$

where the discrete Fourier transformations (DFT) of  $x(n)$ ,  $h(n)$ ,  $y(n)$  and  $w(n)$  are  $X(k)$ ,  $H(k)$ ,  $Y(k)$  and  $W(k)$ , respectively. These discrete Fourier transformations are estimated after removing CP.

The OFDM frame includes the pilot symbols of the 1st OFDM block and the transmitted data of the next OFDM blocks. The channel can be considered stationary during a certain frame, but it can change between different frames. The proposed DL BiLSTM NN-based CSIE receives the arrived data at its input terminal and extracts the transmitted data at its output terminal (*Essai Ali, 2021; Ye, Li & Juang, 2018*).

## OFFLINE DL OF THE DL BiLSTM NN-BASED CSIE

DLNN utilisation is the state-of-the-art approach in the field of wireless communication, but DLNNs have high computational complexity and long training time. GPUs are the most powerful tools used for training DLNNs (*Sharma, Vinutha & Moharir, 2016*). Training should be done offline due to the long training time of the proposed CSIE and the large number of BiLSTM-NN's parameters, such as biases and weights, that should be

tuned during training. The trained CSIE is then used in online implementation to extract the transmitted data (Ye, Li & Juang, 2018; Essai Ali, 2021).

In offline training, the learning dataset is randomly generated for one subcarrier. The transmitting end sends OFDM frames to the receiving end through the adopted (simulated) channel, where each frame consists of single OFDM pilot symbol and a single OFDM data symbol. The received OFDM signal is extracted based on OFDM frames that are subjected to different channel imperfections.

All classical estimators rely highly on tractable mathematical channel models, which are assumed to be linear, stationary and follow Gaussian statistics. However, practical wireless communication systems have other imperfections and unknown surrounding effects that cannot be tackled well by accurate channel models; therefore, researchers have developed various channel models that effectively characterise practical channel statistics. By using these channel models, reliable and practical training datasets can be obtained by modelling (Bogdanovich, Vostretsov & Electronics, 2009; Essai Ali, 2021; 2019).

In this study, the 3GPP TR38.901-5G channel model developed by (2019) is used to simulate the behaviour of a practical wireless channel that can degrade the performance of CSIEs and hence, the overall communication system's performance.

The proposed estimator is trained *via* the algorithm, which updates the weights and biases by minimising a specific loss function. Simply, a loss function is defined as the difference between the estimator's responses and the original transmitted data. The loss function can be represented by several functions. MATLAB/neural network toolbox allows the user to choose a loss function amongst its available list that contains crossentropyex, MSE, sigmoid and softmax. In this study, another two custom loss functions (MAE and SSE) are created. The performance of the proposed estimator when using three loss functions (*i.e.*, MAE, crossentropyex and SSE) is investigated. The loss functions can be expressed as follows:

$$\text{crossentropyex} = - \sum_{i=1}^N \sum_{j=1}^c X_{ij}(k) \log(\hat{X}_{ij}(k)), \quad (10)$$

$$\text{MAE} = \frac{\sum_{i=1}^N \sum_{j=1}^c |X_{ij}(k) - \hat{X}_{ij}(k)|}{N}, \quad (11)$$

$$\text{SSE} = \sum_{i=1}^N \sum_{j=1}^c (X_{ij}(k) - \hat{X}_{ij}(k))^2, \quad (12)$$

where  $N$  is the sample number,  $c$  is the class number,  $X_{ij}$  is the  $i$ th transmitted data sample for the  $j$ th class and  $\hat{X}_{ij}$  is the DL BiLSTM-based CSIE response for sample  $i$  for class  $j$ .

Figure 4 illustrates the offline training processes to obtain a learned CSIE based on BiLSTM-NN.

**Table 1** BiLSTM- and LSTM-NN structure parameters and training process options.

| Parameter            | Value                    |
|----------------------|--------------------------|
| Input Size           | 256                      |
| BiLSTM Layer Size    | 30 hidden neurons        |
| LSTM Layer Size      | 30 hidden neurons        |
| FC Layer Size        | 4                        |
| Loss Functions       | Crossentropyex, MAE, SSE |
| Mini Batch Size      | 1000                     |
| Epochs Number        | 1000                     |
| Learning Algorithm   | Adam                     |
| Training Data Size   | 8000 - OFDM frame        |
| Validation Data Size | 2000 - OFDM frame        |
| Test Data Size       | 10000 - OFDM frame       |

**Table 2** OFDM system and channel parameters.

| Parameter         | Value       |
|-------------------|-------------|
| Modulation Mode   | QPSK        |
| Carrier Frequency | 2.6 GHz     |
| Paths Number      | 24          |
| CP Length         | 16          |
| Subcarrier Number | 64          |
| Pilot Number      | 64, 8 and 4 |

## SIMULATION RESULTS

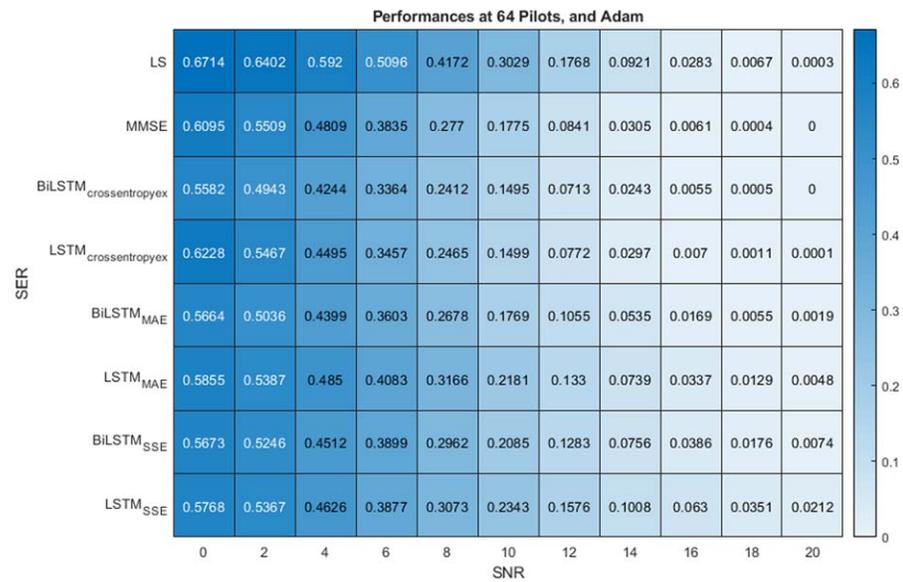
### Studying the performance of the proposed, LS and MMSE estimators by using different pilots and loss functions

Several simulation experiments are performed to evaluate the performance of the proposed estimator. In terms of symbol error rate (SER) performance analysis, the SER performance of the proposed estimator under various SNRs is compared with that of the LSTM NN-based CSIE (*Essai Ali, 2021*), the well-known LS estimator and the MMSE estimator, which is an optimal estimator but requires channel statistical information. *A priori* uncertainty of the used channel model statistics is assumed and considered for all conducted experiments.

Moreover, the Adam optimisation algorithm is used to train the proposed estimator whilst using different loss functions to obtain the most robust version of the proposed CSIE. The proposed model is implemented in 2019b MATLAB/software.

**Table 1** lists the parameters of BiLSTM-NN and LSTM-NN architectures and their related training options. These parameters are identified by a trial-and-error approach. **Table 2** lists the parameters of the OFDM system model and the channel model.

The examined estimators' performance is evaluated at different pilot numbers of 4, 8 and 64 as well as crossentropyex, MAE and SSE loss functions. The Adam optimisation algorithm is used for all simulation experiments.



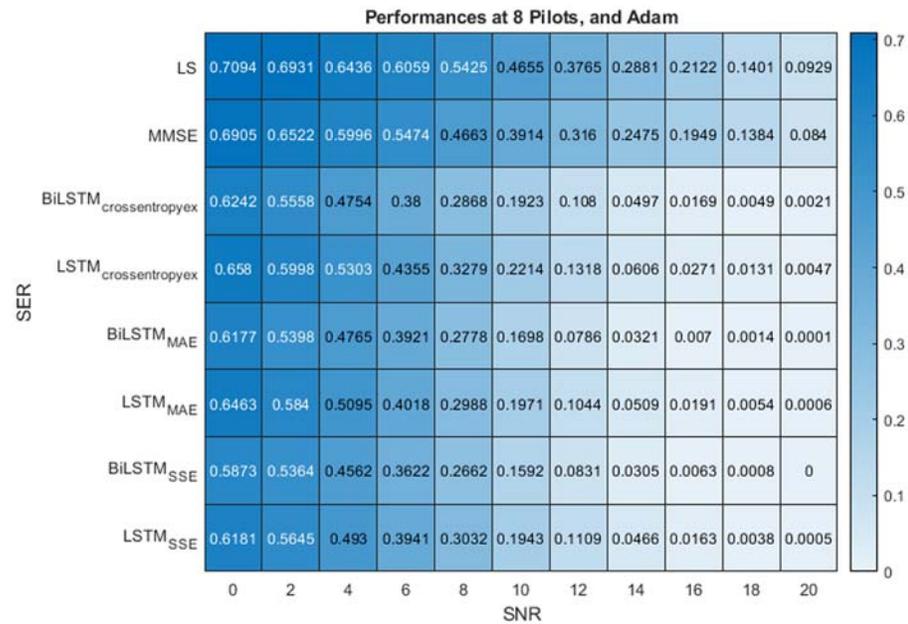
**Figure 6** SER comparison of LS, MMSE, BiLSTM and LSTM estimators using 64 pilots, the Adam learning algorithm and crossentropyex, MAE and SSE loss functions.

Full-size DOI: [10.7717/peerjcs.682/fig-6](https://doi.org/10.7717/peerjcs.682/fig-6)

With a sufficiently large number of pilots (64) and the use of the crossentropyex loss function, the proposed BiLSTM<sub>crossentropyex</sub> estimator outperforms LSTM<sub>crossentropyex</sub>, LS and MMSE estimators over the entire SNR range, as shown in Fig. 6. At the use of the MAE loss function, the BiLSTM<sub>MAE</sub> estimator outperforms the LS estimator over the SNR range [0–18 dB], but LSTM<sub>MAE</sub> outperforms it over the SNR range [0–14 dB]. In addition, the BiLSTM<sub>MAE</sub> and LSTM<sub>MAE</sub> estimators are at par with the MMSE estimator over the SNR ranges [0–10 dB] and [0–4 dB], respectively. Beyond these SNR ranges, the MMSE estimator outperforms BiLSTM<sub>MAE</sub> and LSTM<sub>MAE</sub> estimators. BiLSTM<sub>MAE</sub> outperforms LSTM<sub>MAE</sub> starting from 0 dB to 20 dB.

At the use of the SSE loss function, Fig. 6 shows that the BiLSTM<sub>SSE</sub> and LSTM<sub>SSE</sub> estimators achieve approximately the same performance as the MMSE estimator over a low SNR range [0–6 dB]. MMSE outperforms the BiLSTM<sub>SSE</sub> and LSTM<sub>SSE</sub> estimators starting from 8 dB, and the LS estimator outperforms BiLSTM<sub>SSE</sub> starting from 16 dB and LSTM<sub>SSE</sub> starting from 14 dB. BiLSTM<sub>SSE</sub> outperforms LSTM<sub>SSE</sub> starting from 10 dB to 20 dB. LS provides poor performance compared with MMSE because it does not use prior information about channel statistics in the estimation process. MMSE exhibits superior performance, especially with sufficient pilot numbers, because it uses second-order channel statistics. Concisely, MMSE and the proposed BiLSTM<sub>crossentropyex</sub> attain close SER performance with respect to all SNRs. Furthermore, at low SNR (0–6 dB), BiLSTM<sub>(crossentropyex, MAE, and SSE)</sub>, LSTM<sub>(crossentropyex, MAE, and SSE)</sub> and MMSE attain approximately the same performance.

Figure 7 present the performance comparison of LS, MMSE, BiLSTM and LSTM-based estimators using the Adam optimisation algorithm and the different (crossentropyex, MAE and SSE) loss functions at 8 pilots. Figure 7 shows



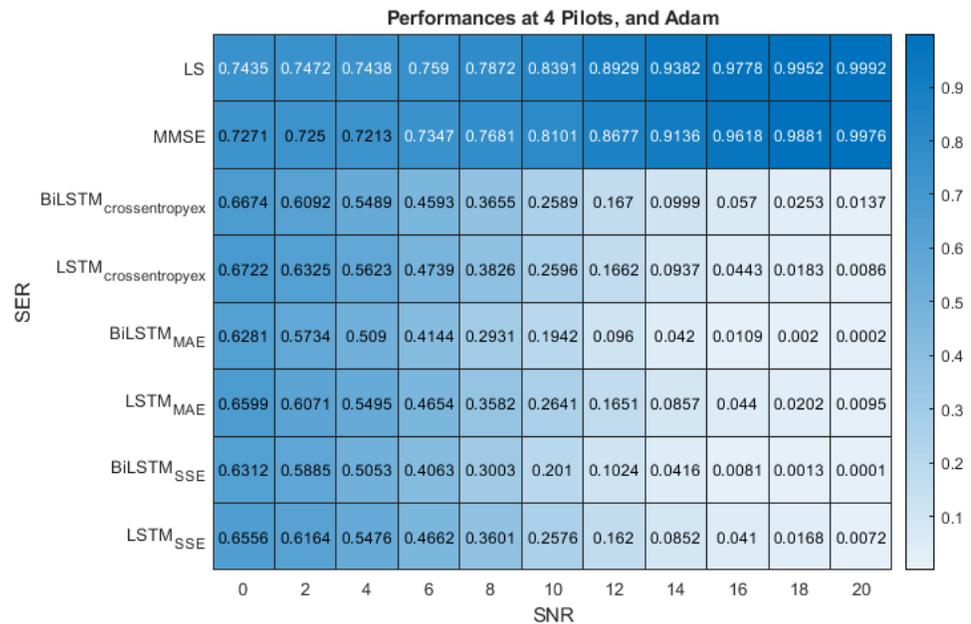
**Figure 7** SER performance comparison of LS, MMSE, BiLSTM, and LSTM estimators using 8 pilots, the Adam learning algorithm and crossentropyex, MAE and SSE loss functions.

Full-size DOI: [10.7717/peerjcs.682/fig-7](https://doi.org/10.7717/peerjcs.682/fig-7)

that the proposed BiLSTM<sub>(crossentropyex, or MAE or SSE)</sub> estimators outperform the LSTM<sub>(crossentropyex, or MAE or SSE)</sub> estimators and the traditional estimators over the examined SNR range. At a low SNR (0–7 dB), the proposed BiLSTM<sub>(crossentropyex, or MAE or SSE)</sub> estimators exhibit semi-identical performance. Furthermore, the proposed BiLSTM<sub>SSE</sub> estimator trained by minimising the SSE loss function outperforms the BiLSTM<sub>crossentropyex</sub> estimator trained by minimising the crossentropyex loss function starting from 0 dB; also it outperforms BiLSTM<sub>MAE</sub>, which is trained by minimising the MAE loss function starting from 14 dB. Concisely at 8 pilots BiLSTM<sub>SSE</sub> estimator achieved the most minimum SER.

Figure 8 show the performance comparison of the LS, MMSE, BiLSTM<sub>(crossentropyex, or MAE or SSE)</sub> and LSTM<sub>(crossentropyex, or MAE or SSE)</sub> estimators at four pilots. Figure 8 shows the superiority of the proposed BiLSTM<sub>(crossentropyex, or MAE or SSE)</sub> estimators in comparison with the traditional estimators, which have lost their workability starting from 0 dB. It also shows the superiority of the proposed estimator BiLSTM<sub>(MAE or SSE)</sub> over LSTM<sub>(MAE or SSE)</sub>. LSTM<sub>(crossentropyex)</sub> exhibits a competitive performance as BiLSTM<sub>(crossentropyex)</sub> starting from 0 dB to 12 dB, and LSTM<sub>(crossentropyex)</sub> outperforms BiLSTM<sub>(crossentropyex)</sub> starting from 14 dB. At very low SNRs (0–3 dB), the proposed BiLSTM<sub>(crossentropyex, or MAE or SSE)</sub> estimators have the same performance. The proposed BiLSTM<sub>SSE</sub> estimator outperforms the BiLSTM<sub>crossentropyex</sub> estimator starting from 4 dB, and it exhibits an identical performance as the BiLSTM<sub>MAE</sub> estimator until 14 dB and outperforms it in the rest of the SNR examination range.

Figures 6–8 emphasise the robustness of the BiLSTM-based estimators against the limited number of pilots, low SNR, and under the condition of *a priori* uncertainty of



**Figure 8** SER performance comparison of LS, MMSE, BiLSTM, and LSTM estimators using 4 pilots, the Adam learning algorithm and crossentropyex, MAE and SSE loss functions.

Full-size DOI: [10.7717/peerjcs.682/fig-8](https://doi.org/10.7717/peerjcs.682/fig-8)

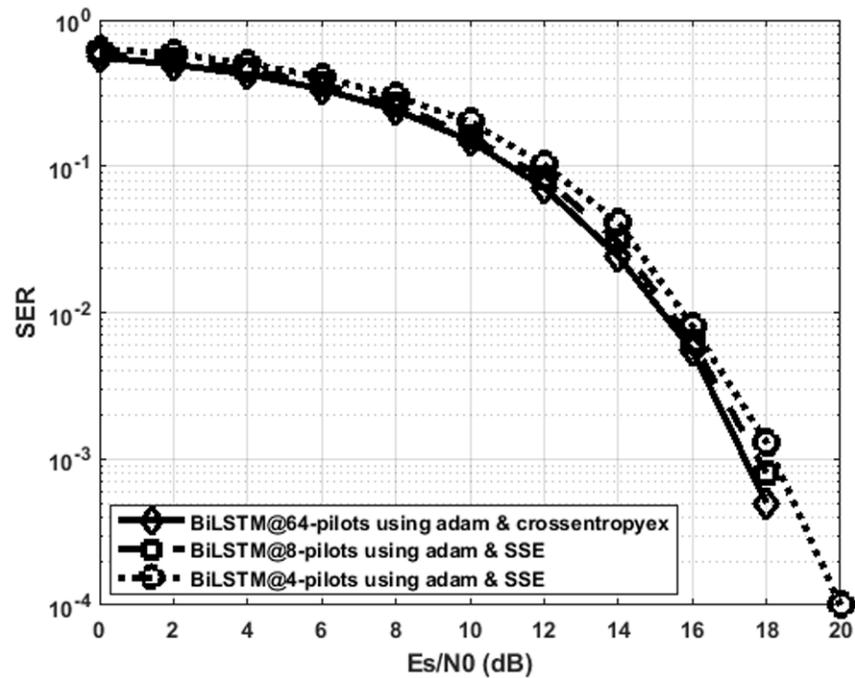
channel statistics. They demonstrate the importance of testing various loss functions in the deep learning process to obtain the most optimal architecture of any proposed estimator.

Figure 9 indicates that the proposed BiLSTM<sub>crossentropyex</sub>, BiLSTM<sub>SSE</sub> and BiLSTM<sub>SSE</sub> estimators have close SER performance at 64, eight and four pilots, respectively. The performance of BiLSTM<sub>SSE</sub> at eight pilots coincides with the performance of BiLSTM<sub>crossentropyex</sub> at 64 pilots. Therefore, using the proposed estimators with few pilots is recommended for 5G OFDM wireless communication systems to attain a significant improvement in their transmission data rate. Given that the proposed estimator adopts a training data set-driven approach, it is robust to *a priori* uncertainty for channel statistics.

### Loss curves

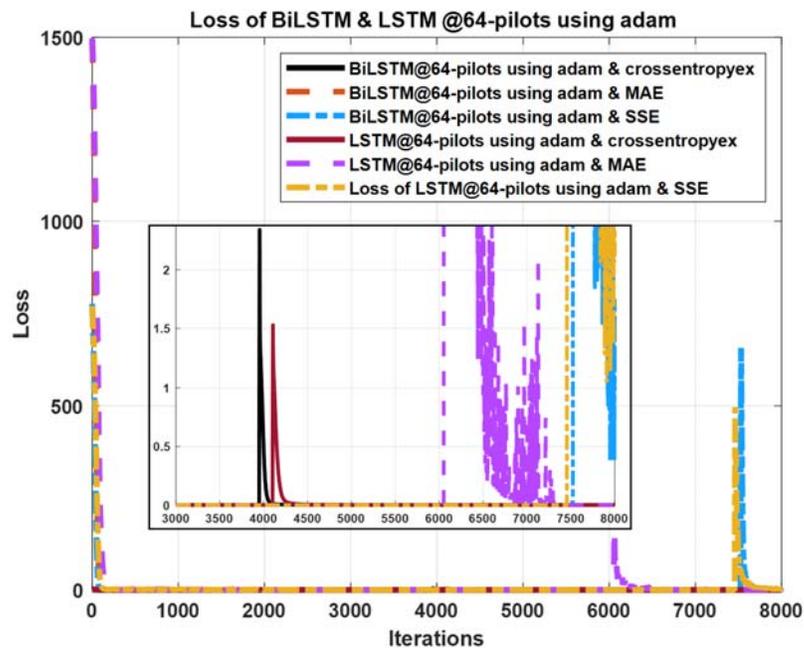
The quality of the DLNNs' training process can be monitored efficiently by exploring the training loss curves. These loss curves provide information on how the training process goes, and the user can decide whether to let the training process continue or stop.

Figures 10–12 show the loss curves of the DLNN-based estimators (BiLSTM and LSTM) at pilot numbers = 64, eight and four and with the three examined loss functions (crossentropyex, MAE and SSE). The curves emphasise and verify the obtained results in Figs. 6, 7, and 8. For example, the sub-curves in Fig. 10 for BiLSTM<sub>crossentropyex</sub> and LSTM<sub>crossentropyex</sub> estimators emphasise their superiority over the other estimators. This superiority can be seen clearly from Fig. 6. Moreover, the training loss curves in Figs. 11 and 12 emphasise the obtained SER performance in Figs. 7 and 8, respectively, of each examined DLNN-based CSIE. For more details, good zooming, and analysis of the presented loss curves, they can be downloaded from this link ([shorturl.at/lqxGQ](http://shorturl.at/lqxGQ)).



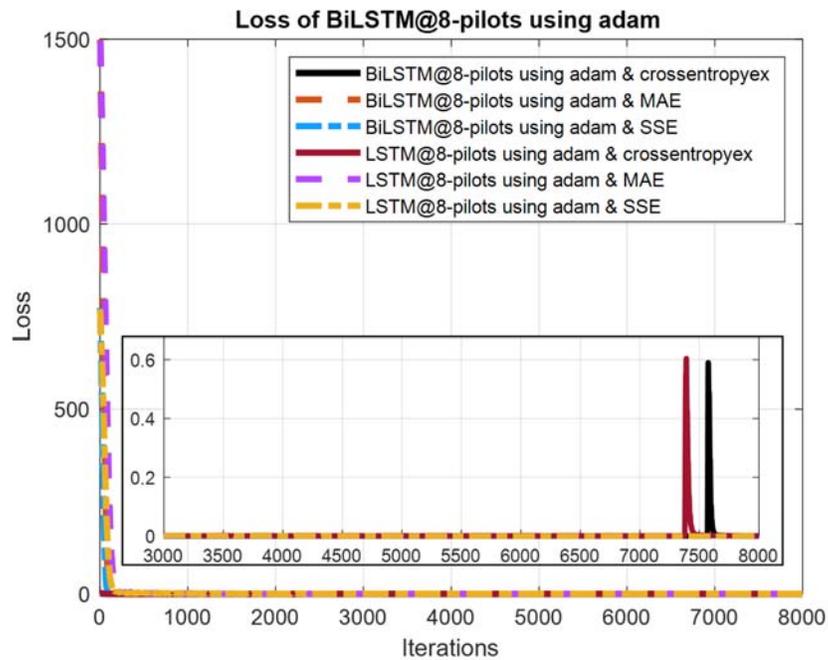
**Figure 9** SER performance comparison of the best DL BiLSTM-based CSIEs using various pilots and loss functions.

Full-size [DOI: 10.7717/peerjcs.682/fig-9](https://doi.org/10.7717/peerjcs.682/fig-9)



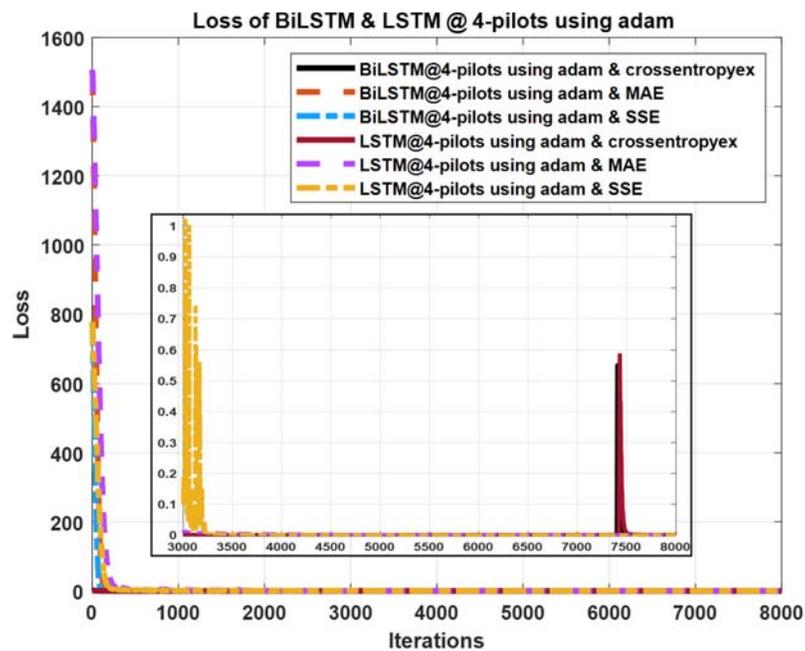
**Figure 10** Loss curves comparison of BiLSTM- and LSTM- based estimators using 64 pilots, the Adam learning algorithm and crossentropyx, MAE and SSE loss functions.

Full-size [DOI: 10.7717/peerjcs.682/fig-10](https://doi.org/10.7717/peerjcs.682/fig-10)



**Figure 11** Loss curves comparison of BiLSTM- and LSTM-based estimators using eight pilots, the Adam learning algorithm and crossentropy, MAE and SSE loss functions.

Full-size [DOI: 10.7717/peerjcs.682/fig-11](https://doi.org/10.7717/peerjcs.682/fig-11)



**Figure 12** Loss curves comparison of BiLSTM- and LSTM-based estimators using four pilots, the Adam learning algorithm and crossentropy, MAE and SSE loss functions.

Full-size [DOI: 10.7717/peerjcs.682/fig-12](https://doi.org/10.7717/peerjcs.682/fig-12)

**Table 3** Accuracy comparison of the examined estimators using 64 pilots.

| 64 pilots      |        |       |      |       |
|----------------|--------|-------|------|-------|
|                | BiLSTM | LSTM  | MMSE | LS    |
| Crossentropyex | 100    | 99.99 | 100  | 99.94 |
| SSE            | 99.23  | 97.88 | 100  | 99.96 |
| MAE            | 99.87  | 99.52 | 100  | 99.97 |

**Table 4** Accuracy comparison of the examined estimators using eight pilots.

| 8 pilots       |        |       |       |       |
|----------------|--------|-------|-------|-------|
|                | BiLSTM | LSTM  | MMSE  | LS    |
| Crossentropyex | 99.84  | 99.53 | 91.34 | 91.62 |
| SSE            | 100    | 99.95 | 91.60 | 91.49 |
| MAE            | 100    | 99.94 | 91.53 | 91.50 |

**Table 5** Accuracy comparison of the examined estimators using four pilots.

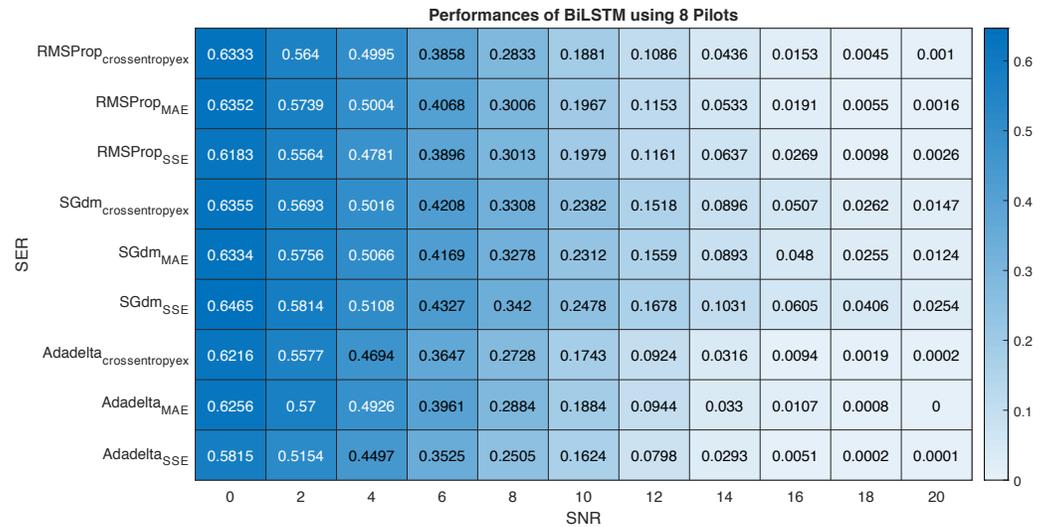
| 4 pilots       |        |       |      |      |
|----------------|--------|-------|------|------|
|                | BiLSTM | LSTM  | MMSE | LS   |
| Crossentropyex | 98.61  | 97.94 | 0.24 | 0.02 |
| SSE            | 100    | 99.28 | 0.24 | 0.09 |
| MAE            | 99.97  | 99.05 | 0.26 | 0.04 |

### Accuracy calculation

The accuracy of the proposed and other examined estimators is a measure of how the estimators recover transmitted data correctly. Accuracy can be defined as the number of correctly received symbols divided by the total number of transmitted symbols. The proposed estimator is trained in different conditions as indicated in the previous subsection, and we wish to investigate how well it performs in a new data set. Tables 3, 4 and 5 present the obtained accuracies for all examined estimators under all simulation conditions.

As illustrated in Tables 3 to 5, the proposed BiLSTM-based estimator attains accuracies from 98.61 to 100 under different pilots and loss functions. The other examined DL LSTM-based estimator has accuracies from 97.88 to 99.99 under the same examination conditions. The achieved accuracies indicate that the proposed estimator has robustly learned and emphasises the obtained SER performance in Fig. 9. The obtained results of MMSE and LS in Tables 1, 2 and 3 emphasise the presented SER performance in Figs. 6, 7 and 8, respectively, and show that as the pilot number decreases, the accuracy of the conventional estimators dramatically decreases.

The proposed BiLSTM- and LSTM-based estimators rely on DLNN approaches, where they can analyse huge data sets that may be collected from any plant, recognise the statistical dependencies and characteristics, devise the relationships between features and generalise the accrued knowledge for new data sets that they have not seen before. Thus, they are applicable to any 5G and beyond communication system.



**Figure 13** Performance comparison of BiLSTM-based estimator using eight pilots, the RMSProp, SGdm, and Adadelta optimisation algorithms and crossentropyex, MAE and SSE loss functions.

Full-size DOI: [10.7717/peerjcs.682/fig-13](https://doi.org/10.7717/peerjcs.682/fig-13)

### Impact of using different optimization algorithms on the proposed estimator performance

DL procedures benefit greatly from optimization methods. DNN training can be thought of as an optimisation issue that aims to discover a global optimum by applying gradient descent methods to obtain a robust training, and hence reliable prediction or classification models. Choosing the best optimization method for a particular scientific topic is a difficult task. Using the wrong optimization strategy during training can cause the DN to stay at the local minimum, which results in no training progress (*Dogo et al., 2018*). As a result, examination is required to evaluate the performance of various optimisers to get the optimal CSIE.

This section provides performance comparison experiments using RMSProp, SGdm, and Adadelta optimisation algorithms (*Soydancer & Intelligence, 2020*) for training the proposed BiLSTM-based CSIE at using 8-pilots, as illustrated in [Fig. 13](#). [Table 6](#) arranges the proposed BiLSTM CSIE estimators using different optimisation algorithms and loss functions from the highest performance to the lowest and their related accuracies.

It is clear from [Fig. 13](#) and [Table 6](#) that the trained BiLSTM-based CSIE using Adadelta optimisation algorithm and SSE loss function achieves the best SER performance and provides the highest accuracy with 100%. On the other hand, the same estimator achieves the lowest SER performance and provides accuracy with 97.46% using SGdm optimization algorithm and SSE loss function. This, in turn, shows the importance of studying the training process efficiency using different optimization algorithms in the case of using a specific loss function.

**Table 6** Performance comparison of different optimisation algorithms and its related accuracies.

| Order   | Optimisation algorithm <sub>Loss function</sub> | Accuracy |
|---------|---|----------|
| First   | Adadelta <sub>SSE</sub>                         | 100%     |
| Second  | Adadelta <sub>crossentropyex</sub>              | 99.99%   |
| Third   | Adadelta <sub>MAE</sub>                         | 99.98%   |
| Fourth  | RMSProp <sub>crossentropyex</sub>               | 99.90%   |
| Fifth   | RMSProp <sub>MAE</sub>                          | 99.84%   |
| Sixth   | RMSProp <sub>SSE</sub>                          | 99.74%   |
| Seventh | SGdm <sub>MAE</sub>                             | 98.76%   |
| Eighth  | SGdm <sub>crossentropyex</sub>                  | 98.53%   |
| Ninth   | SGdm <sub>SSE</sub>                             | 97.46%   |

## CONCLUSIONS AND FUTURE WORK

The proposed DL-BiLSTM-based CSIE is an online pilot-assisted estimator. It is robust against a limited number of pilots and exhibits superior performance compared with conventional estimators; it is also robust under the conditions of *a priori* uncertainty of communication channel statistics (non-Gaussian/stationary statistical channels) and demonstrates superior performance compared with conventional estimators and DL LSTM NN-based CSIEs.

Two customized classification layers using the loss functions (MAE and SSE) are introduced. The proposed CSIE exhibits a consistent performance at large and small pilot numbers and superior performance at low SNRs, especially at limited pilots, compared with conventional estimators. It also achieves the highest accuracy amongst all examined estimators at 64, eight, and four pilots for all the used loss functions.

The proposed BiLSTM- and LSTM-based estimators have high prediction accuracies of 98.61% to 100% and 97.88% to 99.99%, respectively, when using crossentropyex, MAE, and SSE loss functions for 64, eight, and four pilots. The proposed BiLSTM using (Adam, and crossentropyex), BiLSTM using (Adam, MAE, and SSE; and Adadelta, and SSE), and BiLSTM using (Adam, and SSE), achieve the best SER performance and provide accuracies with 100% at 64, eight, and four pilots respectively. The proposed estimator is promising for 5G and beyond wireless communication systems.

For future work, authors suggest the following research plans:

1. Investigating the proposed estimator's performance and accuracy by using different cyclic prefix lengths and types.
2. Developing robust loss functions by using robust statistics estimators, such as Tukey, Cauchy, Huber and Welsh.
3. Investigating the performance of CNN-, gated recurrent unit (GRU)- and simple recurrent unit (SRU)-based CSIEs whilst using crossentropyex, MAE and SSE loss functions and for 64, eight, and four pilots.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This work was supported by the Taif University, Taif, Saudi Arabia, through the Taif University Researchers Supporting Project, under Grant TURSP-2020/61. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

Taif University, Taif, Saudi Arabia, through the Taif University Researchers: TURSP-2020/61.

### Competing Interests

The authors declare there are no competing interests.

### Author Contributions

- Mohamed Hassan Essai Ali conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Ibrahim B.M. Taha conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

### Data Availability

The following information was supplied regarding data availability:

Matlab code is available in the [Supplemental Files](#).

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.682#supplemental-information>.

## REFERENCES

- 3GPP. TR38.901. 2019.** Study on Channel Model for Frequencies from 0.5 to 100 GHz; 3GPP: Sophia Antipolis, France, 2019.
- Bogdanovich V, Vostretsov AG, Electronics. 2009.** Application of the invariance and robustness principles in the development of demodulation algorithms for wideband communications systems. *Journal of Communications Technology and Electronics* 54:1283–1291.
- Dogo E, Afolabi O, Nwulu N, Twala B, Aigbavboa C. 2018.** A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks. In: *2018 international conference on computational techniques, electronics and mechanical systems (CTEMS)*. Piscataway: IEEE, 92–99.

- Essai Ali MH. 2021.** Deep learning-based pilot-assisted channel state estimator for OFDM systems. *IET Communications* 15:257–264 DOI 10.1049/cmu2.12051.
- Hochreiter S, Schmidhuber J. 1997.** Long short-term memory. *Neural Computation* 9:1735–1780 DOI 10.1162/neco.1997.9.8.1735.
- Janocha K, Czarnecki WMJapa. 2017.** On loss functions for deep neural networks in classification. arXiv preprint. arXiv:1702.05659.
- Jeya R, Amutha B, Nikhilesh N, Immaculate RRJs. 2019.** Signal interferences in wireless communication-an overview. *Eurasian Journal of Analytical Chemistry* 2:3.
- Joo J, Park MC, Han DS, Pejovic VJIA. 2019.** Deep learning-based channel prediction in realistic vehicular communications. *IEEE Access* 7:27846–27858 DOI 10.1109/ACCESS.2019.2901710.
- Kang J-M, Chun C-J, Kim I-MJIA. 2020.** Deep learning based channel estimation for MIMO systems with received SNR feedback. *IEEE Access* 8:121162–121181 DOI 10.1109/ACCESS.2020.3006518.
- Karra K, Kuzdeba S, Petersen J. 2017.** Modulation recognition using hierarchical deep neural networks. In: *2017 IEEE international symposium on dynamic spectrum access networks (DySPAN)*. Piscataway, 1–3.
- Kim H. 2015.** *Wireless communications systems design*. Chichester: John Wiley & Sons.
- Le Ha A, Van Chien T, Nguyen TH, Choi W. 2021.** Deep learning-aided 5G channel estimation. In: *2021 15th international conference on ubiquitous information management and communication (IMCOM)*. Piscataway: IEEE, 1–7.
- Liao Y, Hua Y, Dai X, Yao H, Yang X. 2019.** ChanEstNet: a deep learning based channel estimation for high-speed scenarios. In: *ICC 2019-2019 IEEE international conference on communications (ICC)*. Piscataway: IEEE, 1–6.
- Luo C, Ji J, Wang Q, Chen X, Li PJIToNS, Engineering. 2018.** Channel state information prediction for 5G wireless communications: a deep learning approach. *IEEE Transactions on Network Science and Engineering* 7:227–236.
- Ma X, Ye H, Li Y. 2018.** Learning assisted estimation for time-varying channels. In: *2018 15th international symposium on wireless communication systems (ISWCS)*. 1–5.
- Ong T. 2017.** Facebook’s translations are now powered completely by AI. *The Verge*. Available at <https://www.theverge.com/2017/8/4/16093872/facebook-ai-translations-artificial-intelligence> (accessed on 9 August 2021).
- O’Shea T, Karra K, Clancy TC. 2017.** Learning approximate neural estimators for wireless channel state information. In: *2017 IEEE 27th international workshop on machine learning for signal processing (MLSP)*. Piscataway: IEEE, 1–7.
- Oyerinde OO, Mneney SH. 2012.** Review of channel estimation for wireless communication systems. *IETE Technical Review* 29:282–298 DOI 10.4103/0256-4602.101308.
- Ponnaluru S, Penke S. 2020.** Deep learning for estimating the channel in orthogonal frequency division multiplexing systems. *Journal of Ambient Intelligence and Humanized Computing* DOI 10.1007/s12652-020-02010-1.
- Sarwar A, Shah SM, Zafar I. 2020.** Channel estimation in space time block coded MIMO-OFDM system using genetically evolved artificial neural network. In: *2020*

- 17th international bhurban conference on applied sciences and technology (IBCAST)*. Piscataway: IEEE, 703–709.
- Senol H, Tahir ARB, Özmen AJTS. 2021.** Artificial neural network based estimation of sparse multipath channels in OFDM systems. *Telecommunication Systems* 77:1–10.
- Sharma R, Vinutha M, Moharir M. 2016.** Revolutionizing machine learning algorithms using gpus. In: *2016 international conference on computation system and information technology for sustainable solutions (CSITSS)*. Piscataway: IEEE, 318–323.
- Sheikh AUH. 2004.** Interference, distortion and noise. In: *Wireless communications: theory and techniques*. Boston: Springer US, 225–285.
- Soydaner DJIJoPR, Intelligence A. 2020.** A comparison of optimization algorithms for deep learning. *International Journal of Pattern Recognition and Artificial Intelligence* 34:2052013.
- Wu Y, Schuster M, Chen Z, Le QV, Norouzi M, Macherey W, Krikun M, Cao Y, Gao Q, Japa Macherey K. 2016.** Google’s neural machine translation system: bridging the gap between human and machine translation. arXiv preprint. [arXiv:09.08144](https://arxiv.org/abs/09.08144).
- Yang Y, Gao F, Ma X, Zhang S. 2019.** Deep learning-based channel estimation for doubly selective fading channels. *IEEE Access* 7:36579–36589  
[DOI 10.1109/ACCESS.2019.2901066](https://doi.org/10.1109/ACCESS.2019.2901066).
- Ye H, Li GY, Juang B. 2018.** Power of deep learning for channel estimation and signal detection in OFDM systems. *IEEE Wireless Communications Letters* 7:114–117  
[DOI 10.1109/LWC.2017.2757490](https://doi.org/10.1109/LWC.2017.2757490).
- Zeiler MDJapa. 2012.** Adadelata: an adaptive learning rate method. arXiv preprint. [arXiv:1212.5701](https://arxiv.org/abs/1212.5701).
- Zhao C, Huang X, Li Y, Yousaf Iqbal MJS. 2020.** A double-channel hybrid deep neural network based on CNN and BiLSTM for remaining useful life prediction. *Sensors* 20:7109.
- Zhou R, Liu F, Gravelle CW. 2020.** Deep learning for modulation recognition: a survey with a demonstration. *IEEE Access* 8:67366–67376  
[DOI 10.1109/ACCESS.2020.2986330](https://doi.org/10.1109/ACCESS.2020.2986330).