Received: May 3, 2022. Revised: June 9, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Development of CNN-LSTM Hybrid Deep Learning Network for the Joint Detection of Non-Orthogonal Multiple Access Signals in 5G Uplink Receivers

Ali Hilal Ali[1]* **Raed S.H. AL-Musawi[2]** **Kadhum Al-Majdi[3]**

[1] *Department of Electronics and Communication Engineering, University of Kufa, Iraq*
[2] *Department of Electrical Engineering, University of Babylon, Iraq*
[3] *Ashur university college, Baghdad, Iraq*
* Corresponding author's Email: alih.alathari@uokufa.edu.iq

Abstract: This paper presents a novel approach for signal detection in non-orthogonal multiple access (NOMA) uplink receivers. We propose converting incoming packets into a stream of 2D image-like vectors. Thereby, converting a signal detection problem into a video classification problem. Our approach is true end-to-end learning where no manual feature engineering and/or pre-processing is required. Detection is done blindly in a joint fashion with no explicit channel estimation and equalization steps required. Successive interference cancellation (SIC) is the default method for detecting NOMA packets, but it requires perfect channel estimation to be maintained. Deep learning approaches have shown great promise. However, they suffer from overfitting and/or poor performance. We show how to improve performance by a better training dataset generation procedure and hyperparameter optimization along with the use of CNN as a feature filter. The CNN-LSTM hybrid network has registered a training and testing accuracies of 88.61 and 85.36 which are higher than the state-of-the-art approaches and its symbol error rate (SER) vs signal-to-noise ratio (SNR) performance is higher by about 9dB than the LSTM approach, maximum likelihood and other standard SIC based approaches like the minimum mean square error (MMSE) and least square (LS). This suggests deep learning-based receivers as strong candidates for the upcoming generations of wireless communication systems.

Keywords: Deep learning, Non-orthogonal multiple access, SIC, Wireless communication, 5G, CNN LSTM hybrid, Hyperparameter optimization.

1. Introduction

The next-generation wireless communication system is expected to support varieties of services and a numerous number of users. To serve such a big number of users, resources, such as spectrum, need to be shared. Traditionally, this has been provided via orthogonal frequency division multiple access (OFDMA). However, OFDMA is not efficient in terms of spectral usage and support for varieties of services and traffic such as IoT devices [1, 2].

To share resources between users of a communication channel with the OFDMA technique, each user is assigned some orthogonal sub-channel and guard intervals are inserted between these sub-channels. As the number of users grows tremendously, these guard intervals amount significantly and result in poor spectral efficiency [3]. In addition, the trade-off between energy efficiency and data rate should be ameliorated particularly in uplink scenarios where battery saving of mobile devices is of the essence [4].

One suitable technique for the next-generation wireless communication that can outperform OFDMA in terms of spectral and power efficiency and support huge varieties of users and services is the so-called non-orthogonal multiple access (NOMA) [5-7]. NOMA does not rely on the principle of orthogonality to separate users from each other, and users have access to all subcarriers. A NOMA transmission consists of packets all users superimposed to form a single signal or stream [8].

Hence, a more sophisticated algorithm is required to separate user data at the receiver. Successive interface cancelation (SIC) is one of the proven detection algorithms for NOMA. It starts by sorting incoming packets in terms of their channel condition or quality of service requirements, next the user with the strongest/highest requirements is detected while assuming all other users as noise, then the detected signal is subtracted from the combined signal where the next user is detected and so on [9, 10].

However, SIC-based techniques require perfect tracking of channel state by some channel estimation mechanism. High mobility networks can cause errors in channel estimation and could result in a loss of dependability and robustness in particular for vehicle users (VUs) [11]. To mitigate the issue of the rapid changing channel state, schemes for resource allocation have been suggested although it is largely still a hard problem [11]. Since user signals are detected successively, errors in detecting one user signal would result in errors in detecting all the remaining user packets. A more recent improvement to NOMA is cooperative-NOMA (C-NOMA). In C-NOMA, users near the base station act as a relay for those further away [12]. While C-NOMA improves rates significantly, it has been shown that imperfect channel state information would significantly impact outage probability and ergodic rates [13]. Additionally, it has also been proven that the error propagation rate is dominated by the users with better channel conditions which could degrade diversity [14], one of the promised features of 5G.

Since this paper aims to come up with a system where link performance is not jeopardized by errors in channel estimation or SIC sequential detection nature, we turned out to schemes where detection is performed in a single-shot fashion such as the case with receivers built with deep learning techniques.

Deep learning is a subset of machine learning where the mapping from input to output is learned from a massive number of input-output example pairs. Deep learning is attractive because it can be used to map the complex relationship between input and output and its performance continue to improve with the addition of more examples [15]. In addition, deep learning can be used in end-to-end learning scenarios where little to no pre-processing and/or feature engineering are required for the successful mapping of input-output data pairs [16].

The authors in [17, 18] have proposed a semi joint detector of users in cooperative-NOMA (C-NOMA). However, their approach lacks vital details on the dataset generation and the accuracy of the training and testing datasets. These are vital because if the accuracy of testing is significantly lower than that of the training then the deep learning network is overfitting. Overfitting is when a DL network fails to generalize from the training dataset to the testing dataset which is kept away during training to assess the performance of the network to data that it has not encountered before. Analysis of the work in [19] shows that the proposed LSTM network overfits by a significant margin [20]. Although overfitting is studied in [20] and an enhanced version is proposed, its total testing accuracy of 74 % versus training accuracy of about 90 % still indicates overfitting. Authors in [21] have proposed converting incoming packets to 3-D tensors where the in-phase (I) and quadrature (Q) components of the received signal are treated as two image channels and a CNN followed by an LSTM network is used for detection. The resulting tensor would be of dimension ($m\times2\times2$) where $2m$ is the size of the received signal. Such tensor would be too shallow for the CNN and heavy padding of zeros should be added to keep dimensions from shrinking down between successive CNN layers. Since padding conveys no information, overfitting is inevitable. Although their approach is somewhat similar to ours, we will show how our modelling and construction of an LSTM-CNN will result in a far better performance. In addition, robustness of their approach was not demonstrated and the issue of generating dataset, training and bias/overfitting were never reported.

In this paper, we will propose a CNN-LSTM hybrid and we will show how this structure can improve performance and overcome overfitting. We will also discuss dataset generation and hyperparameter optimization in detail. The main contributions of this paper are:

- We propose a video classification approach for detecting communication signals. Each video frame is treated independently and is formed by converting incoming packets into 2D image-like-vectors.
- True end-to-end learning where there are no sub-systems, pre-processing, or manual feature extraction. A single deep learning network is the entire receiver. In addition, it does not require channel estimation and equalization.
- We study the effect of bias/overfitting to improve generalization capabilities. This is seldom done in the literature.
- We perform hyperparameter optimization using the black box method. In addition, we propose a better dataset generation procedure to improve accuracy and reduce overfitting.

Received: May 3, 2022. Revised: June 9, 2022.

The rest of the paper is divided into seven sections. Section 1 presents the theoretical modelling of the NOMA uplink channel and receiver. Section 2 discusses channel estimation in the standard SIC receivers such as MMSE and LS. Section 3 shows the reasoning for choosing the CNN-LSTM hybrid model. Section 4 will show how we have generated the training dataset. Section 5 gives the details of the deep learning model and its layers. Section 6 presents the results of the hyperparameters optimization and finally, section 7 discusses the SER vs SNR results.

2. System and channel modelling

This paper aims to develop an end-to-end deep learning network. Much like similar research in this area, the performance of the proposed network will be demonstrated by a two users receiver in an uplink NOMA scenario. By end-to-end, we mean that the deep learning network should perform channel estimation, equalization and detection implicitly and there would be no manual feature engineering and/or heavy pre-processing required before inputting the received signal to the DL network. The detection of the symbols transmitted by users 1 and 2 are done jointly in a single-shot fashion as is the case in [19, 20].

For a fair comparison with all similar work in the literature, we will follow the same assumptions and limitations in the modelling of the uplink NOMA channel. Specifically, two user equipment (UE), each with a single antenna, are connected to a base station (BS) in a typical micro-cell as shown in Fig. 1. These two users are classified as the intra-cell user and edge-user or the near and far user according to their fading channel coefficients [17]. In other words, we will model users with strong and weak power. The BS is modelled as a typical NOMA receiver which works by superimposing received users' signals on each other.

UE transmission comprises a 64-subcarrier OFDM data packet and two pilot packets in a total of three packets scenario. For comparison with the SIC-based receiver, we will assume perfect CSI at the BS. This is because if our DL network cannot beat SIC with perfect CSI then there is no feasible gain in preferring DL based receivers over SIC ones.

Referring to Fig. 1, assume the transmitted signal by a UE is $\mathrm{x}_i(t)$ where (x) is the time domain signal and the subscript (i) represent UE number. Each user will be transmitting a power of $\mathrm{P}_i(t)$ along a multi-fading channel given by the coefficient $\hbar_i(t)$ for each user (i). Mainly, a Rayleigh channel will be considered with additive white Gaussian noise having a variance of σ^2 : $\mathbb{N}(t) \sim N(0, \sigma^2)$, the superimposed received signal $\mathrm{Y}(t)$ at the BS for (N) users is given in Eq. (1).

$$\mathrm{Y}(t) = \sum_{i=1}^{\mathrm{N}} \sqrt{\mathrm{P_i}(t)}\hbar_i(t)\mathrm{x}_i(t) + \mathbb{N}(t) \tag{1}$$

The multi-fading channel $\hbar_i(t)$ is given in Eq. (2) which shows its discrete Fourier transform

$$\hbar_i(t) = \sum_{\lambda=1}^{\epsilon} g_{i,\lambda}\, \sigma\big(t - \tau_{i,\epsilon}\big) \tag{2}$$

where $g_{i,\lambda}$ is the complex channel gain along path λ of user (i) for a total delay of ϵ and $\sigma\big(t - \tau_{i,\epsilon}\big)$ is

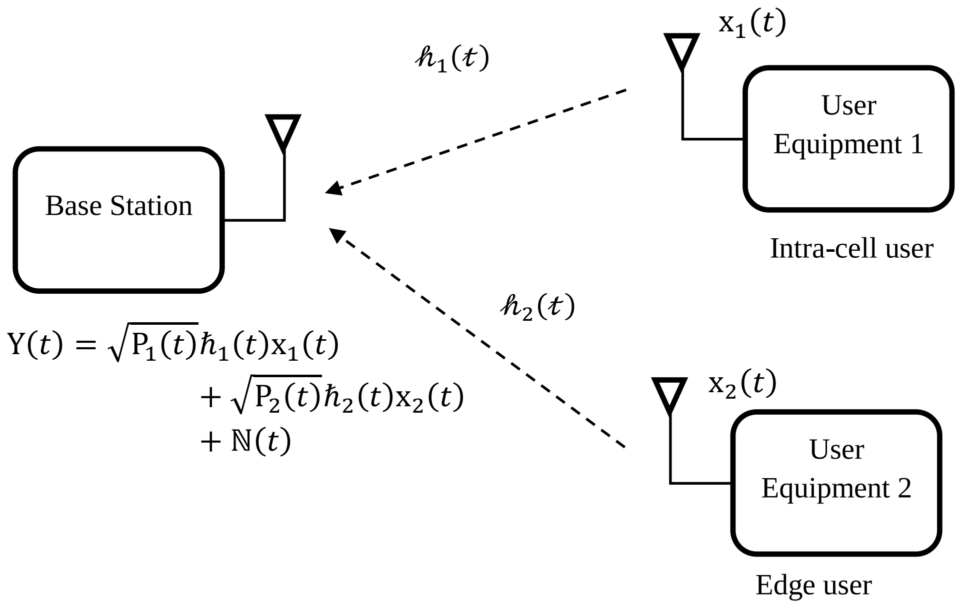


Figure. 1 Two user equipment and a base station operating as NOMA receiver

 DOI: 10.22266/ijies2022.0831.43

the impulse function delayed by a factor of $\tau_{i,\epsilon}$. We will assume a Rayleigh channel fading with a total delay of 20, also referred to as cyclic prefix (CP). SIC decodes signal subsequently starting with the highest power user, the power to Interference plus noise ratio (SINR) for user (i) where i≠1 is given by [22]:

$$SINR_l = \frac{\kappa \mathcal{E}_l |\hbar_l|^2}{\kappa \sum_{i=1}^{N} \mathcal{E}_i |\hbar_i|^2 + 1} \tag{3}$$

where $\hbar_l$ is the multi-fading channel of user (l), $\mathcal{E}_i$ is the power coefficient associated with user (i) constrained by:

$$\sum_{i=1}^{N} \mathcal{E}_i = 1 \tag{4}$$

where N is the total number of users, κ is the signal to noise ratio given by:

$$\kappa = \frac{\mathrm{P}}{\sigma^2} \tag{5}$$

where σ is the noise variance, P is the total power of all users. For user 1, SINR can be written as [22]:

$$SINR_1 = \mathcal{E}_1 \kappa |\hbar_l|^2 \tag{6}$$

and finally, the sum rate for all users is [22]:

$$\mathcal{R}_{sum} = \log_2\left(1 + \kappa \sum_{i=1}^{N} \mathcal{E}_i |\hbar_i|^2\right) \tag{7}$$

We will elaborate further on how the OFDM packets of each user are generated in the next section where the deep learning model is discussed along with the generation of the training dataset.

3. Channel estimation and detection

Before the SIC algorithm can be applied, the channel must be tracked and estimated. Loss of channel state information (CSI) could result in degradation of performance and robustness [23]. In this paper, we will consider two widely used estimators for SIC: Least squares (LS) and minimum mean square error (MMSE) estimators. These will be used as a benchmark for comparing our deep learning model.

Fig. 2 shows a typical OFDM baseband system [24] and assuming the channel has finite impulse length, the received signal (Y) can be modelled by an N-point discrete Fourier transform [25]

$$Y = DFT_N\left(IDFT_N(x) \otimes \frac{\hbar}{\sqrt{\mathbb{N}}} + \tilde{n}\right) \tag{8}$$

where DFT_N is an N-points discrete Fourier transform, $IDFT_N$ is the N-points inverse discrete Fourier transform of signal (x), the symbol $\otimes$ stands for cyclic convolution operation and $\tilde{n}$ is a vector of independent and identically distributed complex Gaussian noise. Eq. (1) and (8) can be re-written in vector format as follows [25]:

$$Y = X\mathrm{F}\hbar + \tilde{n} \tag{9}$$

where X is a diagonal matrix of $[x_0, x_2, \ldots x_{N-1}]^T$ and F is the discrete Fourier transform matrix given by [25]:

$$\mathrm{F} = \left|\begin{pmatrix} \mathcal{M}_N^{00} & \cdots & \mathcal{M}_N^{0(N-1)} \\ \vdots & \ddots & \vdots \\ \mathcal{M}_N^{(N-1)0} & \cdots & \mathcal{M}_N^{(N-1)(N-1)} \end{pmatrix}\right| \tag{10}$$

with

$$\mathcal{M}_{\mathrm{N}}^{nl} = \frac{1}{\sqrt{N}} e^{-i2\pi\frac{nl}{N}} \tag{11}$$

where $l = 0, 1, 2, \ldots, N-1$ and $n = [n_0, n_1, \ldots n_{N-1}]^T = DFT_N(\tilde{n})$.

It can be shown that the MMSE estimator of $\hbar$ is given by [26]:

$$\grave{\hbar}_{MMSE} = \mathbb{C}_{\hbar y}\mathbb{C}_{yy}^{-1} \tag{12}$$

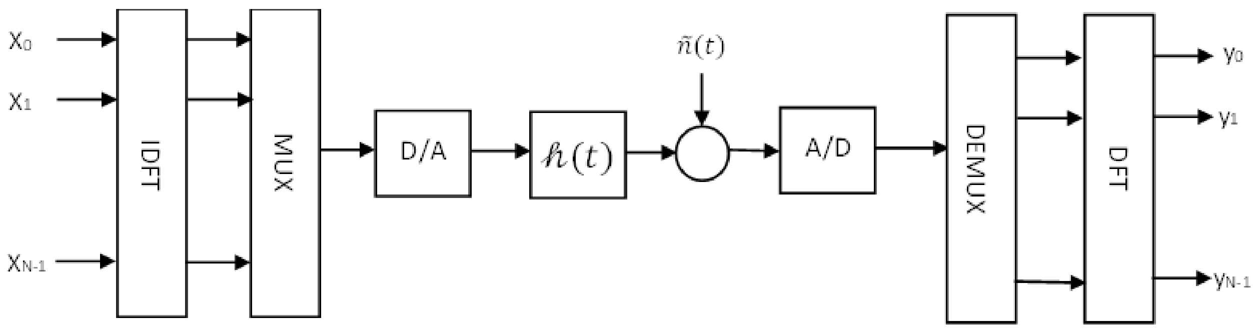


Figure 2. Typical OFDM baseband system

where $\mathbb{C}_{yy}$ and $\mathbb{C}_{\hbar y}$ are the auto and cross covariance matrices between $\hbar$ and y given by:

$$\mathbb{C}_{\hbar y} = \mathbb{C}_{\hbar\hbar} F^H X^H \quad (13)$$

$$\mathbb{C}_{yy} = X\mathrm{F}\mathbb{C}_{\hbar\hbar}\mathrm{F}^H X^H + \sigma^2 I_N \quad (14)$$

where X is a diagonal matrix representation of the vector (x) and I_N is the identity matrix. Similarly, the least square estimator of $\hbar$ becomes [25]:

$$\grave{\hbar}_{LS} = \mathrm{F}\mathbb{Q}_{LS} F^H X^H y \quad (15)$$

with

$$\mathbb{Q}_{LS} = (F^H X^H F X)^{-1} \quad (16)$$

Along with these two estimators, we also consider the maximum likelihood (ML) detector because each of the three methods has its advantages and drawback. Nonetheless, they provide typical benchmark performance to compare performance with.

4. Deep learning model selection

This section will explain the details of the deep learning model that has been selected and the reasoning behind our assumptions and choices. Firstly, the main goal of this research is to develop an end-to-end deep learning network to detect user equipment transmission in a NOMA uplink scenario. By end-to-end, we mean that the model does not need pre-processing and/or feature engineering albeit it can be done implicitly within the model itself. In addition, channel estimation and equalization should be done explicitly as well. Hence, the input to the model is the raw received superimposed signals of users 1 & 2 and the output is a joint label representing the symbols transmitted by these users. This is a classification problem.

While an LSTM implementation may seem like the rational choice for the deep learning model as the case with previous research discussed in the previous section, an LSTM model has some drawbacks. Firstly, for the LSTM network to work, the received packets should be converted to a single sequence. This would result in a long-term dependency between the bits which are inserted earlier in the sequence with those at the end. Long-term dependency is a major issue in sequence models because it can lead to exploding/vanishing gradient during training [27]. Secondly, proposed LSTM models in the literature do not generalize well from the training data [20]. This issue is often called overfitting or high variance.

Thus, we propose to combine the best of both approaches in a CNN-LSTM hybrid model. The CNN acts as a feature filter and the LSTM works as a classifier for the sequence of features coming from the CNN part. This approach is often used to classify videos by considering them as a sequence of independent frames [28]. We will present the training set generation in the next section before discussing our proposed CNN-LSTM hybrid model.

5. Training dataset generation

To keep fair and meaningful with current research which also proposed deep learning for NOMA uplink receivers, we will consider the same scenario as in [19-21]. Referring to Fig. 3, consider two user equipment OFDM system with 64-subcarriers. Data from each user is baseband modulated with QPSK before going into the OFDM system. In addition, two fixed pilot sequences are to form the OFDM packet which is now consisting of three symbols: one data and two pilot symbols. Next, the inverse discrete Fourier transform is applied to the OFDM packet, and a cyclic prefix is added to reduce inter-symbol interference. Finally, AWGN is added to the transmitted signal. We have assumed a Rayleigh channel with a total delay of 20 and a total transmitted power of 1.

At the receiver, the superposition of user equipment 1 and 2 is collected and DFT is performed. Since our OFDM has 64 subcarriers and there are 3 symbols per packet and including both the real and imaginary parts out of the DFT step, there will be a 64×3×2 feature vector per packet. The feature vector is converted to a 2-D vector to form an image-like structure through a simple vector reshaping operation from a 1×384 1-D vector to a 64×6 2-D image-like vector, as shown in Fig. 3.

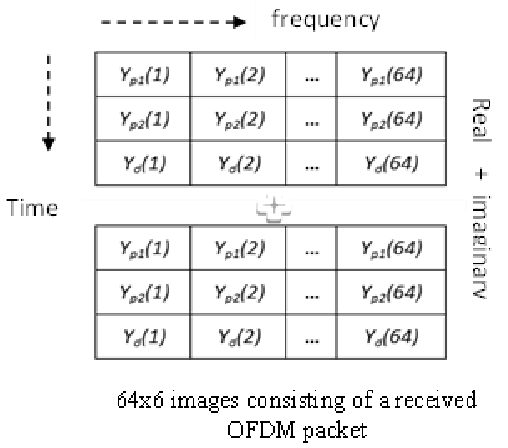


Figure. 3 Forming a 2-D image-like structure from the received OFDM packets

Received: May 3, 2022. Revised: June 9, 2022.

Pilot sequences are shown with subscript (p) and data sequences with subscript (d). The real and imaginary parts are stacked together to form the required dimensions.

Since QPSK is considered, data symbols consist of two bits only which could have one of four possible values. Hence, if the symbols transmitted by the two users are jointly formed there could be one out of 16 possible cases for the transmitted bits from both users. Hence, the label is 1 to 16 integers.

The training dataset is collected by generating random data sequences at the transmitter with E_s/N_o values ranging from 5 to 40dB with a step of 7dB. At each E_s/N_o value, 1000 packets are transmitted, received, converted to images and stored along with their corresponding label. Hence, the total number of samples in the dataset is 96000. Sampling uniformly at all extremes of E_s/N_o would result in lower overfitting [20]. It has been further divided into training, validation and testing subsets at ratios of 90 %, 5 % and 5 % respectively.

6. The CNN-LSTM hybrid architecture

The core joint detector for the NOMA uplink receiver in this paper is the CNN-LSTM hybrid. It was built using MATLAB deep learning toolbox and was built from the ground up. The weights of the CNN and LSTM layers were trained at the same time rather than separately or using via transfer learning as the case in [29]. This will allow us to efficiently optimize the number of layers, filters, and LSTM unit numbers which would not be possible with vanilla CNN networks such as VGG16, GoogleNet, AlexNet…etc.

Fig. 4 shows the CNN-LSTM hybrid layers used during training. The input is a 64×6 OFDM packet formed as an image like a 2-D vector. The input layers are MATLAB toolbox sequence input layer that is used to input a sequence of vector data, or images, to a deep learning network. The folding layer converts these sequences into images which are required by the CNN layers. These layers have no tuneable parameters. Each CNN layer has a number of filters of size (n,m) which are tuneable. We will consider finding the optimal number of CNN layers, number of filters and their sizes in the next section. CNN layers are feature filters where each image is converted into a feature vector for classification usually by a fully connected neural network. Once the feature vector is obtained, it is converted to sequence format and fed to LSTM layers. Each LSTM layer has a tuneable number of units. LSTM is used to transform the input sequence of data to a domain where the classes could be somehow separable and therefore much easier to classify. The LSTM sequence at the output is rolled into a 1-D vector and fed to a fully connected (FC) neural network. The softmax layer calculates the probability of associating each inputted image with a label. Finally, the classification layer picks the label with the highest probability. The arrow between the folding and unfolding layers indicates sharing of parameters, more specifically, the batch size used in training.

7. Hyperparameter optimization results

In deep learning, a hyperparameter is a parameter that can be tuned during training to maximize the performance of the network [30]. Although there are many algorithms for automating the process of finding the best values for the hyperparameter vector, we opted for a simple grid search algorithm due to the discrete nature of the hyperparameters to be optimized and their limited range. It works by specifying values for each hyperparameter and letting the network cycle through these values while registering performance.

The hyperparameters that have been considered are number of CNN layers, number of filters in each layer, filter size, and number of LSTM units. Our approach is to only vary one hyperparameter while keeping all the others fixed. Once the best value is registered, we update its value and vary the next one

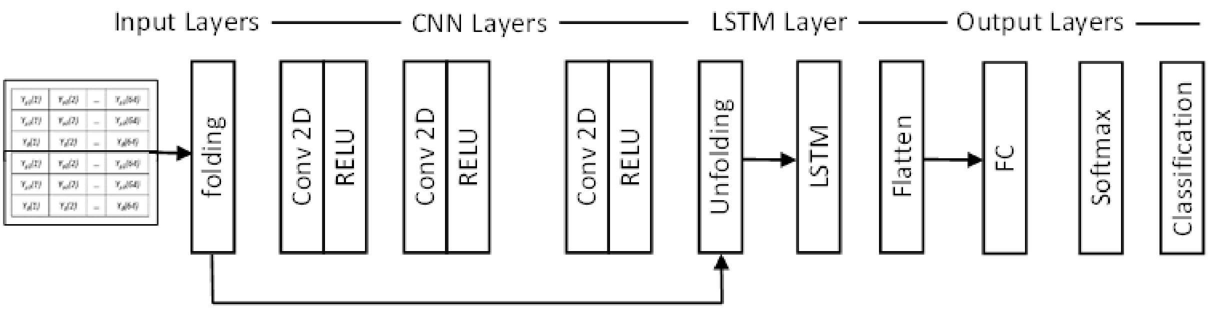


Figure. 4 The CNN-LSTM hybrid deep learning showing images as input and how CNN and LSTM is interconnected via folding and unfolding

Received: May 3, 2022. Revised: June 9, 2022.

Table 1. Results of performing hyperparameter optimization.

Run No.	No. of CNN layers	No. of filters	Filter size	No. of LSTM Units	Training accuracy (%)	Validation accuracy (%)
1	1	4	(3,3)	64	83.69	80.98
2	1	8	(3,3)	64	86.02	83.83
3	1	16	(3,3)	64	87.65	85.83
4	1	32	(3,3)	64	88.41	84.97
5	1	64	(3,3)	64	87.03	82.95
6	1	16	(5,5)	64	88.03	85.34
7	**1**	**32**	**(5,5)**	**64**	**88.61**	**85.36**
8	1	16	(7,7)	64	88.22	84.81
9	1	32	(5,5)	32	84.94	82.37
10	1	32	(7,7)	128	89.18	85.42
11	2	32	(5,5)	64	86.72	84.73

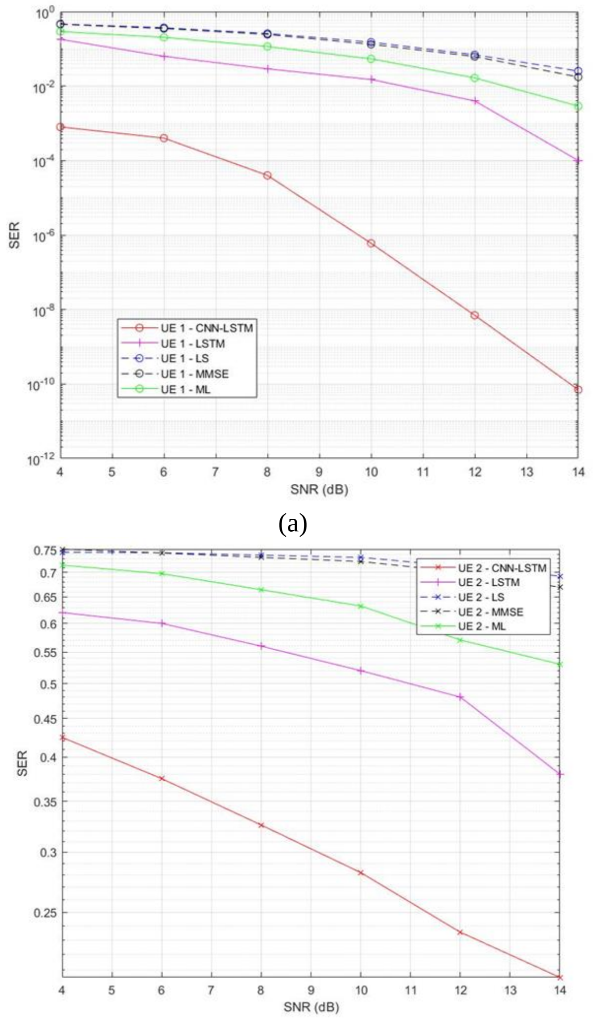
and so on. Table 1 shows the result of the hyperparameter optimization where the best values are highlighted in bold.

In the beginning, we considered a single CNN and LSTM layers and started varying the number of filters and their sizes. There has not been a considerable increase in accuracy when filter size is raised beyond (5, 5). Similarly, LSTM units of 128 have registered the best training accuracy but at the cost of a slight increase in overfitting due to the difference between training and validation accuracy. Increasing the number of CNN layers did not have much of an impact on the results.

It seems that the crucial step in coming up with a good deep learning model for the task of this paper is the CNN filter that transforms raw data vector into a feature vector. Once that is done, the job of the deeper layers, i.e., the LSTM and fully connected layers, becomes much easier. LSTM units of 64 rival the number of OFDM subcarriers and seem like a logical number. We have concluded that a single CNN and LSTM layers are the best choices with details as shown in run 7 of table number 1.

8. Symbol error rate results

Once the hyperparameter optimization step is done as described in the previous section, we tested the resulting network to obtain the symbol error rate (SER) versus signal to noise ratio (SNR) curve. For that, we have followed the same two-users uplink scenario described in the previous sections and in [19-21] for a fair comparison. We have compared our results with standard SIC MMSE, LS-based receivers and the Maximum Likelihood method. We will also compare our approach to the recently proposed LSTM and CNN-LSTM deep learning receivers of references [20, 21].



(a)

(b)

Figure. 5 SER vs SNR curves of the CNN-LSTM approach vs LSTM, SIC MMSE, LS and ML for: (a) user equipment 1 and (b) user equipment 2

Fig. 5 shows two symbol error rate (SER) vs. signal-to-noise ratio (SNR) curves: (a) one for user equipment 1, labelled as UE1, and (b) for user equipment 2, labelled UE2. Our approach is shown in red and labelled "CNN-LSTM" and the LSTM network of reference [20] is shown as "LSTM" in magenta. Other standard methods are shown with their corresponding abbreviations.

We noticed that our approach is about 9dB ahead of the LSTM approach even when SNR is as low as 4dB. This result is consistent in both UE1 and UE2. Comparison with reference [21] was not

Received: May 3, 2022. Revised: June 9, 2022.

possible at CP of 20 due to lack of data but will be considered at CP of 16. Although the LSTM and CNN-LSTM methods lie ahead of the SIC based receivers which proves the performance power of the deep learning approach, the effect of overfitting can be noticed once the scenario is slightly changed. If the network could not cope with changes, then it indicates an inability to generalize from learnt examples. For that, we have changed the cyclic prefix (CP) value from 20 to 16. CP should be chosen larger than the impulse response time of the channel to reduce interference. Reducing its value would increase interference but allows us to assess the robustness of the deep learning approach to these changes. Fig. 6 shows the effect of reducing CP to 16 for user equipment 1 and 2. It can be clearly shown that the LSTM approach has lost its performance benefits over MMSE, and LS as shown in Fig. 5. However, our approach is still ahead of all other receiver types which proves the robustness of our approach.

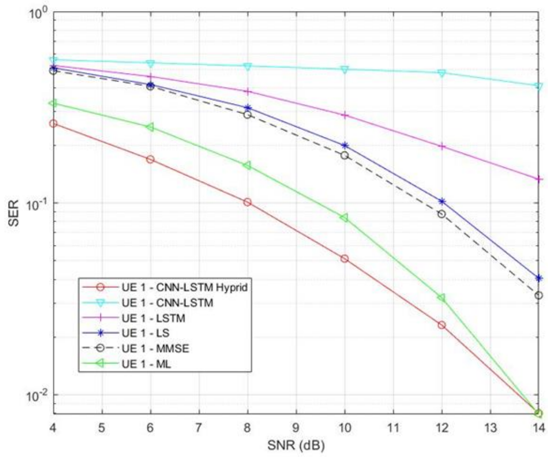
The CNN-LSTM network of reference [21] did not perform well at low SNR values as the case with our approach. This could be the result of training the network at a high SNR value of 50db. Thus, the DL network did not have enough training examples to generalize a good transformation relationship at low SNR values. There are also some other differences including the folding/unfolding layer which [21] did not discuss.

All in all, results show a great impact on the choice of the right receiver for the next generation technology. Firstly, deep learning is a great competitor to classical receivers such as SIC. Secondly, an end-to-end communication receiver designed blindly with deep learning is quite possible. Thirdly, the power of CNN as a feature filter had resulted in improving the performance of the deep learning receiver over other LSTM based receivers which did not utilize CNN. The SER vs SNR curves of our approach showed improvement by over one significant figure which strongly suggests CNN-LSTM receiver for future consideration in NOMA and other communication systems.

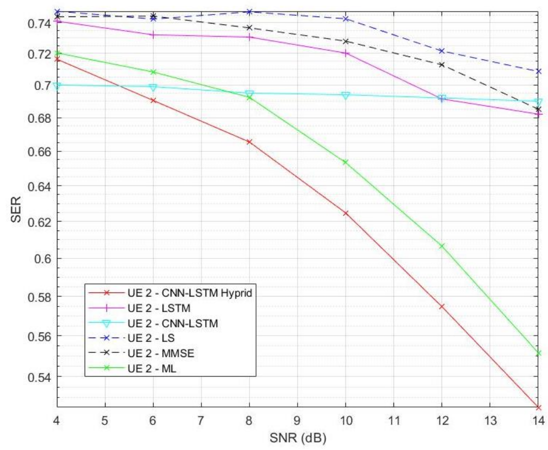
9. Conclusion

NOMA is a widely accepted promising technology for next-generation wireless technology. Its appeal comes from its spectral efficiency which would support a greater number of users. Its robustness, power and bandwidth management capabilities would result in a great variety of user types including IoTs.

The de facto method for detecting NOMA signals is the SIC technique. However, SIC requires perfect channel knowledge to maintain reliability and link performance. In addition, SIC works sequentially by detecting the user with the highest power and then the next in ascending order. If the previous user signal is not detected correctly, then the next ones are lost.



(a)



(b)

Figure 6. SER vs SNR when CP is 16 for CNN-LSTM of reference [21], CNN-LSTM Hybrid, LSTM, SIC MMSE, SIC LS and ML methods for: (a) user equipment 1 (b) user equipment 2

Deep learning is another recently proposed method. However, the deep learning approaches in the literature lack bias/variance analysis. Hence, their generalization capabilities are unknown. In addition, using a vanilla LSTM network may result in long-term dependencies which gives rise to vanishing/exploding gradient problems during training.

We have proposed an end-to-end deep learning network with both CNN and LSTM parts. We have shown how converting the received NOMA signal into a 2-D image-like vector and feeding that to the CNN part have improved performance and reduced overfitting compared to the LSTM approaches in the literature.

We have shown the superiority of our approach using a two-user uplink NOMA scenario and trained a CNN-LSTM network jointly and showed a procedure for the optimization of its hyperparameters. The resulting network showed an improvement in the SER vs SNR as compared to the LSTM method in the literature by at least one significant figure. The generalization power of our approach is shown by reducing the cyclic prefix and noticing its effect on SER. Results indicate significant robustness as compared to the LSTM method and other standard methods such as MMSE, LS and ML.

Conflicts of interest

There is no conflicting interest regarding this paper to the best of our knowledge.

Author contributions

We confirm that all authors have contributed to this work. The main author was responsible for doing the simulation and writing of the paper whereas the 2nd and 3rd authors were responsible for the modeling, theoretical work, and proofreading the manuscript.

References

[1] W. H. Chin, Z. Fan, and R. Haines, "Emerging technologies and research challenges for 5G wireless networks", *IEEE Wireless Communications*, Vol. 21, No. 2, pp. 106-112, 2014.

[2] M. K. Hasan, M. Shahjalal, M. M. Islam, M. M. Alam, M. F. Ahmed, and Y. M. Jang, "The role of deep learning in NOMA for 5G and beyond communications", In: *Proc. of 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2020.

[3] K. Selvam and K. Kumar, "Energy and spectrum efficiency trade-off of non-orthogonal multiple access (NOMA) over OFDMA for machine-to-machine communication", In: *Proc. of 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, 2019.

[4] A. Khalili and D. W. K. Ng, "Energy and spectral efficiency tradeoff in OFDMA networks via antenna selection strategy", In: *Proc. of 2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020.

[5] R. Shankar, "Examination of a non-orthogonal multiple access scheme for next generation wireless networks", *The Journal of Defense Modeling and Simulation*, Vol. 19, No. 3, pp. 453-465, 2022.

[6] X. Yan, K. An, T. Liang, G. Zheng, Z. Ding, S. Chatzinotas, and Y. Liu, "The application of power-domain non-orthogonal multiple access in satellite communication networks", *IEEE Access*, Vol. 7, pp. 63531-63539, 2019.

[7] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey", *IEEE Communications Surveys & Tutorials*, Vol. 22, No. 3, pp. 1805-1838, 2020.

[8] R. C. Kizilirmak and H. K. Bizaki, "Non-orthogonal multiple access (NOMA) for 5G networks", *Towards 5G Wireless Networks-A Physical Layer Perspective*, Vol. 83, pp. 83-98, 2016.

[9] Z. Ding, R. Schober, and H. V. Poor, "Unveiling the importance of SIC in NOMA systems—Part 1: State of the art and recent findings", *IEEE Communications Letters*, Vol. 24, No. 11, pp. 2373-2377, 2020.

[10] Z. Ding, R. Schober, and H. V. Poor, "Unveiling the importance of SIC in NOMA systems—Part II: New results and future directions", *IEEE Communications Letters*, Vol. 24, No. 11, pp. 2378-2382, 2020.

[11] S. Guo and X. Zhou, "Robust resource allocation with imperfect channel estimation in NOMA-based heterogeneous vehicular networks", *IEEE Transactions on Communications*, Vol. 67, No. 3, pp. 2321-2332, 2018.

[12] M. Zeng, W. Hao, O. A. Dobre, and Z. Ding, "Cooperative NOMA: state of the art, key techniques, and open challenges", *IEEE Network*, Vol. 34, No. 5, pp. 205-211, 2020.

[13] V. Aswathi and A. Babu, "Full/half duplex cooperative NOMA under imperfect successive interference cancellation and channel state estimation errors", *IEEE Access*, Vol. 7, pp. 179961-179984, 2019.

[14] F. Kara and H. Kaya, "On the error performance of cooperative-NOMA with statistical CSIT", *IEEE Communications Letters*, Vol. 23, No. 1, pp. 128-131, 2018.

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT press, Cambridge, Massachusetts, M.A.2016.

[16] H. Ye, L. Liang, G. Y. Li, and B. H. Juang, "Deep learning-based end-to-end wireless communication systems with conditional GANs

as unknown channels", *IEEE Transactions on Wireless Communications*, Vol. 19, No. 5, pp. 3133-3143, 2020.

[17] A. Emir, F. Kara, H. Kaya, and H. Yanikomeroglu, "Deep learning empowered semi-blind joint detection in cooperative NOMA", *IEEE Access*, Vol. 9, pp. 61832-61852, 2021.

[18] A. Emir, F. Kara, H. Kaya, and X. Li, "Deep learning-based flexible joint channel estimation and signal detection of multi-user OFDM-NOMA", *Physical Communication*, Vol. 48, p. 101443, 2021.

[19] J. Thompson, "Deep learning for signal detection in non-orthogonal multiple access wireless systems", In: *Proc. of 2019 UK/China Emerging Technologies (UCET)*, 2019.

[20] A. H. Ali, R. S. A. Musawi, and K. A. Majdi, "Enhancements to the Deep Learning Signal Detection Model in Non-Orthogonal Multiple Access Receivers and Noisy Channels", *Journal of Telecommunications and the Digital Economy*, Vol. 10, No. 1, pp. 92-100, 2022.

[21] Y. Xie, K. C. Teh, and A. C. Kot, "Deep learning-based joint detection for OFDM-NOMA scheme", *IEEE Communications Letters*, Vol. 25, No. 8, pp. 2609-2613, 2021.

[22] M. Aldababsa, M. Toka, S. Gökçeli, G. K. Kurt, and O. Kucur, "A tutorial on nonorthogonal multiple access for 5G and beyond", *Wireless Communications and Mobile Computing*, Vol. 2018, 2018.

[23] K. Higuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access", *IEICE Transactions on Communications*, Vol. 98, No. 3, pp. 403-414, 2015.

[24] R. Prasad, *OFDM for Wireless Communications Systems*, Artech House, Boston, M.A.2004.

[25] J. J. V. D. Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems", In: *Proc. of 1995 IEEE 45th Vehicular Technology Conference. Countdown to the Wireless Twenty-First Century*, 1995.

[26] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*, Cambridge university press, Cambridge, Cambs.2010.

[27] J. Zhang, Y. Lin, Z. Song, and I. Dhillon. "Learning long term dependencies via fourier recurrent units", In: *Proc. of International Conference on Machine Learning*, 2018.

[28] A. Tasdelen and B. Sen, "A hybrid CNN-LSTM model for pre-miRNA classification", *Scientific Reports*, Vol. 11, No. 1, pp. 1-9, 2021.

[29] B. Barros, P. Lacerda, C. Albuquerque, and A. Conci, "Pulmonary covid-19: Learning spatiotemporal features combining CNN and LSTM networks for lung ultrasound video classification", *Sensors*, Vol. 21, No. 16, p. 5486, 2021.

[30] M. Feurer and F. Hutter, "Hyperparameter optimization", *Automated Machine Learning, The Springer Series on Challenges in Machine Learning*, Springer, Cham., 2019.