Long-Short term Memory based Channel Prediction for SISO System

Draksham Madhubabu, Arpita Thakre
Department of Electronics and Communication Engineering
Amrita School of Engineering, Bengaluru,
Amrita Vishwa Vidyapeetham, India
madhubabu926@gmail.com, arpita.thakre@gmail.com

Learning the characteristics of a wireless channel is one of the most fundamental and challenging issues in the wireless communication. Many conventional signal processing techniques have been developed so far to estimate the channel state information. In this paper, we present a novel approach based on deep learning using long short term memory (LSTM) neural network to predict the fast-varying Rayleigh fading channel. Having known application of LSTM neural network in the field of time-series prediction problems, we have applied LSTM to predict the future state of the channel by providing the past channel state values. The predicted values of channel is used to recover the transmitted symbol from noisy signal received at the receiver end. We have optimized the configuration of LSTM network by correctly tuning the hyper-parameters so that we get a correct prediction of channel which is in turn used for reconstruction of signal at receiver end. The simulated Bit Error Rate vs. Signal-to-Noise ratio plot proves that LSTM based channel predictor works better than currently used interpolation based channel predictor.

Keywords— Deep Learning, Rayleigh fading channel, LSTM Neural Network, Wireless Communication, SISO system.

I. INTRODUCTION

Communication is a rich field consisting of domain knowledge regarding modelling different types of channels, signaling and detection techniques and different modulation schemes [1], [2] that ensure reliability of data transfer between transmitter and receiver [3]. Further to increase the efficiency in the performance of the wireless technologies. We need to be docile in adapting to the latest technologies, such as data-driven approaches by using machine learning and deep learning algorithms. There were different multiple access techniques like frequency division multiple access (FDMA) across frequency, time division multiple access (TDMA) across time and code division multiple access [4] schemes across different pseudo codes helped to increase the quality of service(QoS) and showed significant performance improvement but we cannot go any further with them. Now a days the data centric approach had gathered attention with huge success attained in fields as such realization of vision through computers, processing of human language and image classification [5], [6]. For a reliable data transmission to happen we need to model the channel correctly but accurate modelling of channel mathematically is almost impossible so we go for a probabilistic based approach. The core signal processing algorithms used in communications fails to improve in performance [7]. The deep learning based algorithms in the field of wireless communication [8] was used because it has the power to improve the performance in areas where it is impossible to generate mathematical models but by considering the deep learning model as a black-box we can tune hyper-parameters so that we can control the output of the system [9].

The next wave of wireless technology suffers with the problem of network latency as the number of connections over a network increases drastically in number the scope of deep learning is such that it has to automatically get tuned to the available network and increase or decrease power with respect to the corresponding receiver. In recent the machine learning which is a root class of deep learning had used in the field of communications for tasks such as recognition of modulation scheme, channel encoding techniques and decoding techniques, in estimation of channel and channel equalization [10], [11], [12], [13].

In the recent years the computing power has been drastically increased and the algorithms have been developed so far and fast that uses the data. In physical layer research, by considering the applications of deep learning we are not only trying to improve the performance of individual blocks but also we try to increase the performance of entire communication system by considering transmitter, channel and receiver as a deep learning model such type of arrangement is called as auto-encoder[14]. The channel estimation using least squares (LS) method and minimum mean square error (MMSE) method estimators have been utilized in many occasions [15]. Of that least squares estimator does not require any previous channel values. The MMSE estimator works better by considering the second order statistics of the channel. In a paper related to orthogonal frequency division multiple access (OFDM) [16] the application of deep learning which mainly focus on feed forward neural network have been studied it presents initial results in learning methods to deal with wireless channels and the removal of cyclic prefix in OFDM and usage of limited pilots and non-linear noise distortions has been discussed [17].

We, in this paper have opted for a single input single output system [18](SISO-system), modeled the channel as a fast fading channel with Rayleigh distribution and a moving receiver at 500kmph scenario is considered. Then at the receiver end we have built a LSTM neural network and trained it online using the data from the zero forcing estimation block after adding additive white Gaussian noise (AWGN). We optimized the data over the loss function between the true values of the channel and the values that were corrupted due to noise. The predicted values have been used for channel equalization and successful reconstruction of the signal at receiver end. We have fine-tuned the hyper-parameters at a certain value of SNR and has been implemented for SNR of certain range. Finally, after the reconstruction of the transmitted bits the Bit Error Rate over different values of Signal to noise ratio has been plotted.

we will be describing about the wireless system model with the block diagram in Section II. In Section III we will be discussing about the Deep Learning basics and Deep Learning Architecture being used and how we have obtained the data and the corresponding performance metrics used in obtaining the results. That was followed by a discussion on the simulation results obtained in Section IV, Section V deals about conclusion of the topic and its future scope of the work done.

II. WIRELESS SYSTEM MODEL

We have assumed the problem of channel prediction as a supervised learning problem. Out of all the channel characteristics like attenuation, scattering, diversity, reflection and refraction we have been considering channel fading as the only phenomenon to tackle. The channel we have considered here is a Rayleigh fading channel. As *R* being its corresponding random variable and its probability density function is given by

$$f_R(r) = \frac{r}{\sigma^2} e^{\frac{-r}{2\sigma^2}}, \quad r \geq 0 \qquad (1)$$

we have been considering the case of moving receiver which introduces a Doppler shift in the receiving signal. We are considering the speed of the receiver as 500kmph in this problem. Here we have considered V as velocity and λ as wavelength and θ as the angle between the receiver and the incoming signal. The value of Doppler shift (ΔØ) affect the value of frequency of the received signal the change in frequency is calculated using Equation 2 and 3.

$$\Delta\text{Ø} = \frac{2\pi\Delta l}{\lambda} = \frac{2\pi v\Delta t}{\lambda}\cos\theta \qquad (2)$$

$$f_d = \frac{1}{2\pi}\frac{\Delta\text{Ø}}{\Delta t} = \frac{v}{\lambda}\cos\theta \qquad (3)$$

The problem we have taken is a complex optimization problem that address the time-series relationship in the data and we have modelled it as a supervised regression problem. The baseband equivalent of the bandpass SISO (single input single output) system for a faded channel of probability distribution as Rayleigh in the presence of noise will be

$$\mathbf{y} = \mathbf{h}\,\mathbf{x} + \mathbf{n} \qquad (4)$$

Where y is the received complex baseband signal and h is a Rayleigh faded channel |h| is the magnitude of the fading and x is the input modulated signal to be transmitted and n is the Additive White Guassian Noise (AWGN). Where noise follows Guassian distribution. While solving the problem we have split the equations into real and imaginary parts shown in Equation 5 and 6.

$$y_{real} = h_{real} * x_{real} + n_{real} \qquad (5)$$

$$y_{img} = h_{img} * x_{img} + n_{img} \qquad (6)$$

From Fig.1, we have generated the 'N' bits where N □ (0, 1) and that bits has been sent to modulator block where the QPSK modulation happens on the input bits and the modulator block output symbols which are of complex in nature and they are of size N/2. Then those symbols are transmitted through wireless channel which in our case is a Rayleigh faded channel and the output of the channel is received by the receiver and adds an Additive White Gaussian Noise (AWGN) at the beginning of the receiver after the addition of noise we will consider the signal as y. 'y' is complex signal consisting of real part and imaginary part and The first 150 symbols were given as input to the zero-forcing estimator the output 150 symbols of the estimator are separated into real and imaginary values and are given to two LSTM blocks separately the LSTM block takes the input as first 150 values of h and predicts the next 600 values of h in both real and imaginary part. We call it as h' and then we collect h' and proceeds for channel equalization with the obtained values of h' and then after the channel equalization we will be demodulating the signal to reconstruct the originally transmitted information.

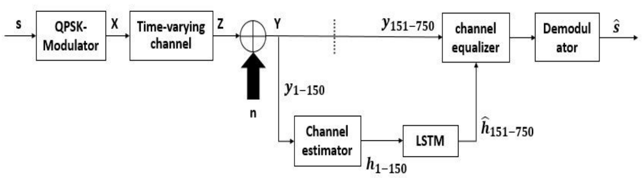

Fig.1: Signal-flow Diagram of the SISO system using LSTM neural network as channel predictor.

III. DATASET AND DEEP LEARNING ARCHITECTURE

A. Dataset

The dataset we used in this problem is a custom dataset it can be obtained by simulating the wireless communication system under specific channel conditions and the corresponding AWGN channel the data collected was the data before the noise being added to the wireless system and the data after the noise being added those two were collected and the data collected after the addition of noise is given as the input for the model to learn.

B. Deep Learning basics

Neural networks, an extension of machine learning has its own significance. Unlike, machine learning there is no need for doing feature extraction in neural networks the feature engineering is automatically done. The neural networks ideally consists of three layers. The first one is input layer, next one is hidden layer and the last one is output layer. Any neural network layer basically made up of nodes. Every node in one layer makes connection with every other node of the next layer by a weight and a bias at every node this forms the basic structure of feed-forward neural network.

Usually, the input and output nodes are common and the size of the network depends on the count of hidden layers. A neural network with one hidden layer can equally approximate any mathematical function. The Weights in the feed forward neural network are updated by an algorithm called back propagation as it updates the weights by minimizing the distance between the actual values and obtained values. The process of continuous update of weights continue until a global minima is attained on the error surface curve. The update of weights usually happen by a factor called learning rate. Generally, the output of a feed forward neural network is given by Equation 7.

C. Deep Learning Architeture

But, for our problem we need a neural network that

$$\mathbf{O} = \varnothing\,(\textstyle\sum_{i=1}^{m} W_i . X_i + b_i) \qquad (7)$$

works a neural network that work well for time-series data. So, we have shifted our attention towards RNN (recurrent neural network). It has a same function similar to that of a feed forward neural network except for that each node in RNN consists of a feed-back which makes them efficient to deal with time-series data. RNN which is similar to feed forward net after unwrapping it in time. After unwrapping the RNN it appears like this

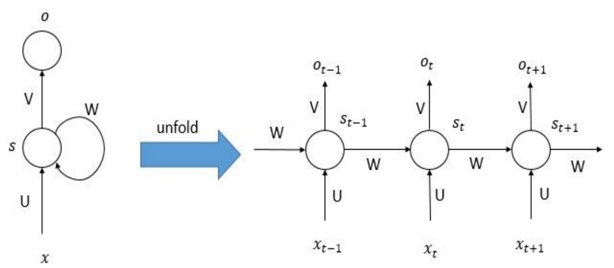

Fig.2: RNN neural network unwrapping in time.

RNN have a problem with length of time-series as the length of time-series goes on increasing the RNN suffers from a problem of disappearing gradient. The problem of disappearing gradient can be resolved by introducing a memory into RNN. This typical structure is named as LSTM (long short term memory) neural network which typically looks like this

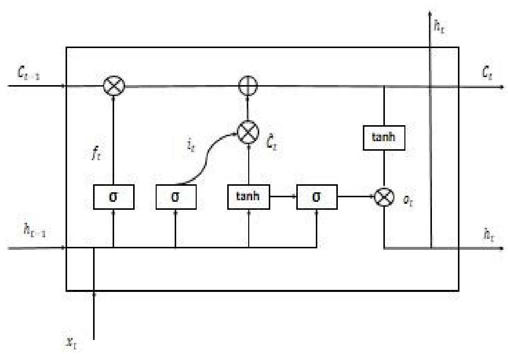

Fig. 3: LSTM memory unit structure

As in Fig.3, The LSTM memory unit consists of a three gates. These gates are responsible for the flow-control and storage of information. These functionality of these gates are used in naming them. The memory cell is responsible for the storage of current cell state which is to be applied for the future time-step. The input gate denoted by **i** responsible for the flow of fresh information into the memory cell. The forget gate which is denoted by **f** is responsible for the amount of information to be forget in the current state. The output gate which is taken as **o** is responsible for the control of amount of information to be flown into the rest of the network. The last layer in the LSTM neural network is connected to a fully connected neural network. The equations from 8 to 13 describe the functionality of LSTM.

$$i_t = \sigma(W_x^i . x_t + W_h^i . h_{t-1}) \qquad (8)$$

$$f_t = \sigma(W_x^f . x_t + W_h^f . h_{t-1}) \qquad (9)$$

$$o_t = \sigma(W_x^o . x_t + W_h^o . h_{t-1}) \qquad (10)$$

$$\tilde{C}_t = \tanh(W_x^c . x_t + W_h^c . h_{t-1}) \qquad (11)$$

$$C_t = i_t \otimes \tilde{C}_t + f_t \otimes C_{t-1} \qquad (12)$$

$$h_t = o_t \otimes \tanh(C_t) \qquad (13)$$

D. Performance metrics

The predicted values of the channel are evaluated using a metric called root mean square error (RMSE) the error function upon which the weights are optimized. We will be dealing with the real component values and imaginary component values of the channel separately using two LSTM networks which are tuned differently and the result is combined to form a complex value prediction. The RMSE values are computed using a mathematical equation given by

$$\mathbf{RMSE} = \sqrt{\tfrac{1}{N}\textstyle\sum_{i=1}^{N}(y_i^{pred} - y_i^{actual})} \qquad (14)$$

the problem we are solving is a regression problem so we have taken the cost function as root mean square error. There are different cost functions for different types of problems we need to choose the cost function or loss function according to the problem in hand.

IV. Results and Discussion

Experiment was carried in a simulation based environment using MATLAB 2018b and to implement neural networks in MATLAB we used neural network toolbox. A SISO wireless system was simulated and the channel was modeled as a Rayleigh fading channel and we have considered a moving receiver scenario which is offering a Doppler shift of 0.0001534 as we are modeling the channel as a fast fading channel and we have been using a predefined function to simulate Rayleigh fading channel in MATLAB with carrier frequency 2.5GHz and signal bandwidth of 5MHz and with a sampling rate of twice the signal bandwidth and we are considering Jakes spectrum. The modulation scheme used at the transmitter side was QPSK-modulation and its counterpart QPSK demodulator was used at the receiver end.

Firstly, we have generated 1500 bits and then we have passed onto a QPSK modulator so that we may get 750 symbols which are complex numbers are obtained as the output of the modulator block and those symbols were transmitted through the channel with specifications as mentioned above and at the beginning of receiver we are adding a AWGN noise which is a byproduct due to receiver components and that is termed as ‘y’. Now y which is given as input to zero forcing channel estimator and the output of that is of dimension 750x1 is separated as 150 units for training is termed as h and the rest of the values are predicted using the trained neural network model that predicted values is denoted by h’. Here the neural network model consists of 2 LSTM layers with a fully-connected layer in the end and in the LSTM layer the activation function used at the gate level was sigmoid activation function and at the cell-state level we use tanh activation function. We use two separate models that were tuned separately one to predict the real values of h’ and another to predict the imaginary values of h’. For the model which predicts the real values we have taken 50 hidden units that is 50 nodes and the we have considered certain parameters like Max.epochs 200 and Mini Batch size as 64 and an Initial Learning rate as 0.05 and Learning rate drop period as 125. And for the other model which predicts the imaginary values of h uses 50 hidden units and Max.epochs of 250 and a Mini Batch size of 64 and with an Initial rate of learning as 0.05 and a Rate of Learning drop period as 125. Then the predicted values are combined to form a complex value and this is used for channel equalization and after that we give the output of this block as an input to the demodulator block and after the demodulation we will get the transmitted bits while some are corrupted due to channel impairments and the receiver noise. The hyper-parameters that were considered separately for real and imaginary were indicated in Table. I and Table. II

Table. I: Hyper parameters tuned for real channel value prediction.

Number of hidden units	50
Max. epochs	200
optimizer	Adam
Initial learning rate	0.05
Learning rate drop of period	125
Learning rate drop of factor	0.2
Mini Batch size	64

Table. II: Hyper parameters tuned for imaginary channel value prediction.

Number of hidden units	50
Max. epochs	250
Optimizer	Adam
Initial learning rate	0.05
Learning rate drop of period	125
Learning rate drop of factor	0.2
Mini Batch size	64

Fig.4, demonstrates the predicted real values of the channel at 25dB SNR. Fig.5, demonstrates the predicted

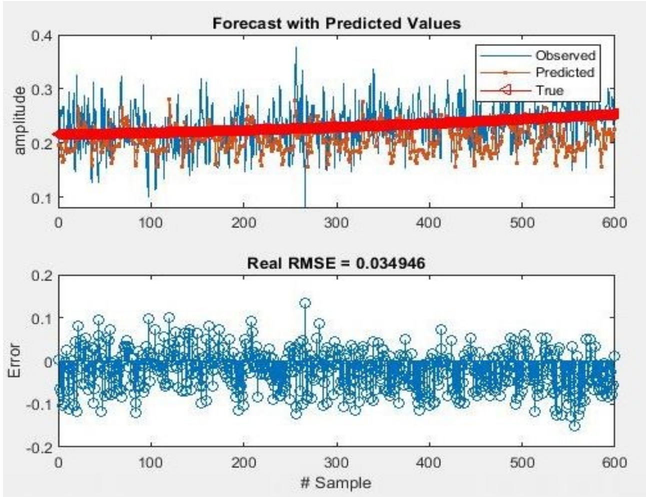

Fig. 4: predicted real values of LSTM block and Real RMSE value for sample.

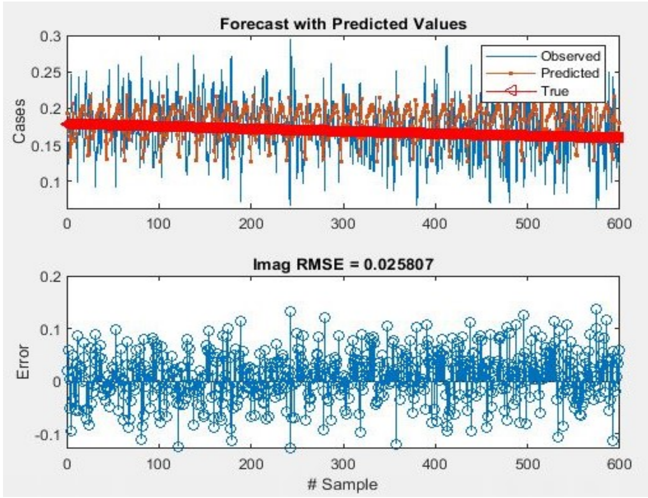

Fig. 5: predicted imaginary values of LSTM block and Imaginary RMSE value for sample.

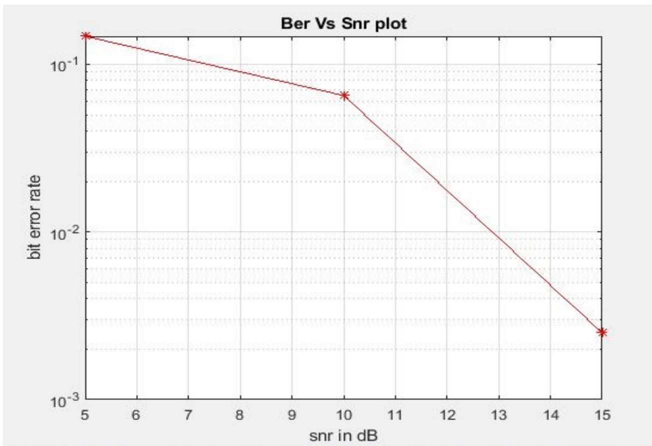

Fig. 6: BER versus SNR plot when N=1500.

imaginary values of the channel. Fig.6, indicates the Bit Error Rate (BER) versus Signal to Noise Ratio (SNR) curve for the number of bits equal to 1500. Now we will extend the concept further by extending the number of bits to 15000 and we will divide the entire 15000 bits into 10 blocks consisting of 1500 each. And then they are QPSK modulated. And these symbols are transmitted through the channel modeled and AWGN noise was added at the beginning of receiver end. Then that symbols were collected. The symbols obtained may be represented like Fig. 7.

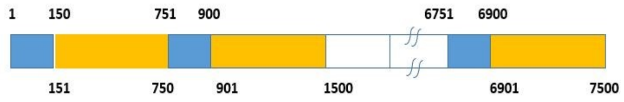

Fig .7: Representation of the symbols training and prediction.

Now the obtained symbols which have dimension of 7500x1 can be represented like the above shown figure. We consider 750 symbols as one set. Of that we will use 150 for training and the next 600 values will be predicted separately for real and imaginary values of the channel by using two different LSTM neural networks which have the hyper parameters as shown in Table. I and Table .II.

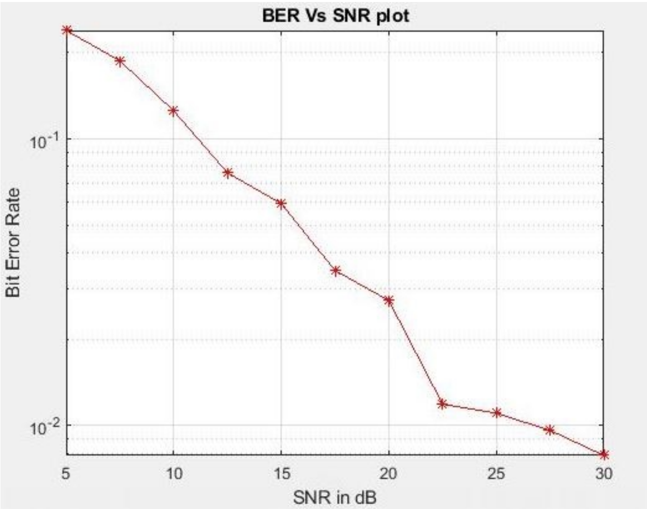

Fig. 8: BER versus SNR plot when N=15000.

We will train for every 150 symbols and predict the next 600 symbols and we will repeat this until 7500 symbols in sets of 750. The entire experiment is repeated for different values of SNR and the BER values were obtained. The BER Versus SNR plot is shown in Fig. 8. The activation function considered are for gate-level is sigmoid and cell-state level is tanh. The obtained Fig. 8, BER versus SNR plot indicates that LSTM are a good tools that can deal with time series data and can be used in application of wireless communication systems.

V. CONCLUSION AND FUTURE SCOPE

In this paper, we develop a system model that predicts the characteristics of a fast varying wireless fading channel in a moving/mobile receiver which is moving at a speed of 500kmph scenario by using a LSTM neural network and we have used that predicted channel values in order to reconstruct the message transmitted at the receiver end. For that we have tuned the hyper parameters of the LSTM neural network so that we will get a system performance better than the existing interpolation based systems. The experimental results suggested that LSTM's have the capability of predicting the time-series data i.e., getting insights from the historical data to predict the future state. we have modeled a SISO channel under the moving receiver case and trained the LSTM network online to predict the channel characteristics. LSTM's are helpful assets in building the next generation wireless technologies.

By considering this as a sub-channel in an OFDM system this concept with a little modification can be extended to multiple sub-channels in an OFDM system.

REFERENCES

[1] T. S. Rappaport, Wireless communications: Principles and practice, 2nd ed. Prentice Hall, 2002.

[2] R. M. Gagliardi and S. Karp, Optical communications, 2nd ed. Wiley, 1995.

[3] J. Proakis and M. Salehi, Digital Communications, 5th ed. McGraw-Hill Education, 2007.

[4] K. Akkarajitsakul, E. Hossain, D. Niyato and D. I. Kim, "Game Theoretic Approaches for Multiple Access in Wireless Networks: A Survey," in IEEE Communications Surveys & Tutorials, vol. 13, no. 3, pp. 372-395, Third Quarter 2011.

[5] Y. LeCun et al., “Generalization and network design strategies,” Connectionism in perspective, pp. 143–155, 1989.

[6] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in Proc. IEEE Int. Conf. Computer Vision, 2015, pp. 1026–1034.

[7] Tianqi Wang, Chao-Kai Wen, Hanqing Wang, Feifei Gao, Tao Jiang and Shi Jin, ”Deep Learning for Wireless Physical Layer: Opportunities and Challenges”, https://arxiv.org/pdf/1710.05312

[8] T. J. O’Shea and J. Hoydis, “An introduction to machine learning communications systems,” arXiv preprint arXiv:1702.00832, 2017.

[9] T. O’Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," in IEEE Transactions on Cognitive Communications and Networking, vol. 3, no. 4, pp. 563-575, Dec. 2017.doi: 10.1109/TCCN.2017.2758370

[10] E. E. Azzouz and A. K. Nandi, ”Modulation recognition using artificial neural networks”, in Proc. Automatic Modulation Recognition of Communication Signals, 1996, pp. 132176.

[11] J. Bruck and M. Blaum, ”Neural networks, error-correcting codes, and polynomials over the binary n-cube”, IEEE Trans. Inf. Theory, vol. 35, no. 5, pp. 976987, Sep. 1989.

[12] C. K. Wen, S. Jin, K. K. Wong, J. C. Chen, and P. Ting, ”Channel Estimation for Massive MIMO Using Gaussian-Mixture Bayesian Learning”, IEEE Trans. Wireless Commun., vol. 14, no. 3, pp. 13561368, Mar. 2015.

[13] J. Cid-Sueiro and A. R. Figueiras-Vidal, ”Digital equalization using modular neural networks: an overview”, in Proc. Signal Processing in Telecommunications. Springer, 1996, pp. 337345.

[14] T. J. OShea, K. Karra, and T. C. Clancy, ”Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention”, IEEE Int. Symp. Signal Process. Inform. Tech. (ISSPIT), pp. 223228, 2016.

[15] Y. G. Li, L. J. Cimini, and N. R. Sollenberger, “Robust channel estimation for OFDM systems with rapid dispersive fading channels,” IEEE Trans. Commun., vol. 46, no. 7, pp. 902–915, Jul. 1998.

[16] S. R. Hari, K. R. Kannan, N. Visagan, Latha and A. Thakre, "Performance of Dual mode OFDM-IM using Reduced Complexity Receiving Technique," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, pp. 2290-2294. doi: 10.1109/ICACCI.2018.8554370.

[17] H. Ye, G. Y. Li and B. Juang, "Power of Deep Learning for Channel Estimation and Signal Detection in OFDM Systems," in *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114-117, Feb. 2018.

[18] S. Ganesh, V. Sayee Sunder and A. Thakre, "Performance Improvement in Rayleigh Faded Channel using Deep Learning," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, pp. 1307-1312.