Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Discrete Memoryless Channel Capacity Estimation using Deep Learning and 256QAM

Thanka Geetha T

vijaytgee@gmail.com

DMI Engineering College

Judson D

SRM Valliammai Engineering College

Research Article

Keywords: DMC, Channel Capacity, Deep learning, 256QAM, Deep neural network, Mutual information, Signal to noise ratio, Blahut Arimoto algorithm.

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9072677/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Discrete Memoryless Channel Capacity Estimation using Deep Learning and 256QAM

Mrs. Thanka Geetha T[1*], Dr. Judson D[2]

[1]Research Scholar, Assistant Professor, Department of Electronics and Communications Engineering, DMI Engineering College, Kanyakumari, Tamilnadu, India.

[2]Professor, Department of Electronics and Communications Engineering, SRM Valliammai Engineering College, Chennai, Tamilnadu, India.

[*]Corresponding Author: vijaytgee@gmail.com

Abstract

Channel capacity should be accurately estimated to design effective communication systems particularly in high-order modulation settings. In this paper, a deep learning-trained structure of the capacity estimation of discrete memoryless channels (DMCs) in the context of 256-Quadrature Amplitude Modulation (256QAM) is introduced. The presented approach is built upon Deep neural networks (DNNs) to train capacity directly based on input-output information, which is contrary to the classical methods like the Blahut-Arimoto algorithm which are both computationally intensive and depend on channel complexity. The simulation findings indicate that the CNN model has the capacity estimates that are close to the theoretical ones with low root mean square error (RMSE) and mean absolute error (MAE) at different signal-to-noise ratios (SNRs). The DNN-based capacity estimator takes 0.043 s (76% fast) to execute a test sample, which is much lower than the average execution time of the Blahut Arimoto algorithm of 0.184 s per SNR point, showing that the proposed approach is computationally efficient. The model is also more scalable and converges faster and hence can be used in real time applications. The article fits into the intersectional area of information theory and artificial intelligence, providing a viable and scalable capacity estimation methodology in next-generation wireless networks.

Keywords

DMC, Channel Capacity, Deep learning, 256QAM, Deep neural network, Mutual information, Signal to noise ratio, Blahut Arimoto algorithm.

1. Introduction

The key to the contemporary information systems is reliable communication and channel capacity has been one of the foundations of information theory research. Claude Shannon came up with the concept of channel capacity that gives the highest rate at which information can be transmitted to a communication channel however with probability of error that is arbitrarily low. One of the most basic

abstractions is the so-called Discrete Memoryless Channel (DMC) which helps to model the behavior of channels with a large number of different conditions and in which the behavior of channel is independent of successive transmissions. The capacity of such channels is not a mere theoretical exercise but it is also practically required when developing effective coding and modulation schemes.

The conventional methods of capacity estimation are that of analytical derivations or iterative algorithms, including the Blahut-Arimoto algorithm, which is mathematically accurate, but computationally demanding with complex channel structures[1.2]. In addition, the methods can have difficulties adapting when the channel statistics underlying are known, or dynamically varying. As the data-driven approaches rapidly develop, machine learning. and deep learning methods The methods have become a strong alternative, with the capacity to acquire complex patterns directly out of measuring behaviors with no modeling assumptions.

The concept of deep learning, and especially neural networks, have been shown to be extremely successful in a variety of fields like computer vision, natural language processing, and signal processing due to their ability to be used to estimate highly nonlinear functions. Deep learning offers a potential opportunity to estimate channel capacity when applied to communication systems, which makes use of large sets of input and output couples. This factual viewpoint helps the model to seed concealed dependencies and statistical variations, which may otherwise be hard to analytically characterize[3,4].

In recent years, studies have started to investigate the interface between information theory and deep learning, and these studies have shown the potential to improve the classical communication paradigms. With the incorporation of neural networks in the estimation, one can have an opportunity to attain a faster convergence, better accuracy, and flexibility to different channel conditions[5,6]. These innovations especially apply in the generation of next-generation wireless networks when heterogeneous environments require powerful and expandable solutions. These iterative methods are mathematically sound, but computationally infeasible when the alphabet of symbols to be used is too large or when the channel varies dynamically. This is specifically noticeable in those schemes that use 256-Quadrature Amplitude Modulation (256QAM) an advanced modulation format that facilitates tight constellation packing and high spectral efficiency but very vulnerable to noise and channel effects. Making proper capacity estimation in these cases is also important but challenging though in conventional means. [7].

Recent progress has attempted to add on the efficiency of capacity computation. The idea of vector flows, suggested by Beretta and Pelillo in the calculation of DMC capacity, provides a dynamical systems

approach that is complementary to the classical optimization-based approaches [1]. The articles emphasize that it is necessary to reconsider classical algorithms using modern tools of computation.

Parallel to these advances, the emergence of the so-called deep learning in communication systems gave rise to the new opportunities in the channel estimation and analysis of capacity. [10]. An overview of deep learning technologies that have been applied to physical-layer channel estimation with a focus on their versatility and resiliency to the wireless world. On the same note, [11] deep architecture channel estimation and beamforming in massive MIMO systems, and shows a huge improvement in performance compared to the traditional methods. Such researches highlight the possibility of the neural network to represent complicated statistical associations across communication media.

There is also the emergence of direct uses of deep learning to estimate channel capacity. [3]. A message importance based neural mutual information estimator is a model that allows learning capacity based on data, without necessarily modeling the channel. [12]. This concept to the case of multi-users demonstrates that neural networks can estimate the capacity region of memoryless multiple access channels. Collectively, these contributions demonstrate the increased synergy amid information theory and machine learning.

Although these developments have been made, few studies have paid special attention to the problem of estimating DMC capacity with deep learning frameworks. The available literature is either focused on general channel estimation or focused on multi-user extensions, thus a gap in the literature on single-user DMCs. Such a gap is the reason behind the current work, which will prove the efficiency of deep learning in estimating DMC capacity, to bridge the gap between theoretical frameworks and practical applications. We are concerned with the issue of the capacity estimation of discrete memoryless channels through deep learning models in this work. This is aimed at closing the divide between theory and practice by showing how neural architectures can be used to effectively estimate channel capacity. In addition to having a computational advantage, this could point to the changing nature of artificial intelligence in the reformulation of traditional communication theory.

Research Gap and Contribution

Proper estimation of channel capacity has been a classic in communication theory. Classical algorithms like BlahutArimoto algorithm can give the exact solution to discrete memoryless channels (DMCs), however, they are computationally intensive because they are iterative algorithms, particularly when used in high-order modulation schemes, e.g., 256Quadrature Amplitude Modulation (256QAM). In addition

to achieving substantial improvements in spectral efficiency, 256QAM has a dense constellation form and is susceptible to noise, which proves challenging to the conventional methods of analysis.

The latest developments in deep learning have demonstrated the potential to overcome these weaknesses. Neural networks have been effectively used in channel estimation, detection and in mutual information approximation. The literature is however missing a single-user DMC capacity estimation field where most of the literature is either devoted to general channel estimation or multi-user scenarios.

The paper can fill this gap in the following ways:

New Framework: Our suggestion is a deep learning based model that is specifically designed to approximate the capacity of discrete memoryless channels with 256QAM modulation to bridge the gap between theoretical frameworks and data based approaches.

Computational Efficiency: The framework minimizes the use of iterative methods of analysis, converges more quickly and yet remains accurate.

Robustness and Adaptability: The model can learn and adapt to different channel conditions and levels of noise through direct input-output information, thus it is appropriate in the real-world wireless setting.

Bringing Theory and Practice Together: The paper shows that artificial intelligence can be used to complement classical information theory to provide scalable solutions to the next-generation communication system.

2. System Model

2.1 Discrete Memoryless Channel (DMC):

A discrete memoryless channel is characterized by a finite input symbols, often denoted as $X=\{X_1,X_2,\ldots..,X_i\}$, a finite output Symbols denoted as $Y=\{Y_1,Y_2,\ldots\ldots,Y_j\}$, and a set of conditional probabilities $P(Y_j/X_i)$. Mathematically, for a sequence of input and output symbols it is represented as,

$$P(Y_1, Y_2, \ldots., Y_n \mid X_1, X_2, \ldots\ldots\ldots, X_n) = \prod_{i=1}^{n} P(Y_i \mid X_i)$$



Fig:1 Discrete Memoryless Channel

2.1.1. Capacity Estimation in a Discrete Memoryless Channel

Capacity of discrete memoryless channel (DMC) is the highest rate at which information can be sent across a channel with low probability error. This is the ability to achieve the largest mutual information between the channel input and output, all subject to the input probability distribution in all possible combinations. It can be mathematically put as,

$$C = \max_{P(x)} I\,(X;Y)$$

2.1.2. Mutual Information

Mutual information $\big(I(X;Y)\big)$ quantifies the amount of information that the output (Y) reveals about the input (X). It's calculated using the input probability distribution (P(X)), output probability distribution (P(Y)), and the conditional probabilities of the channel $\big(P(\,Y \mid X\,)\big)$. Mathematically it is expressed as,

$$I(X;Y) = H(Y) - H(\,Y \mid X\,)$$

Where H(Y) is the entropy of the output and $H(\,Y \mid X\,)$ is the conditional entropy of the output given the input.

2.2.Classical Algorithms to Estimate the capacity of DMC

2.2.1. Blahut – Arimoto Algorithm(BAA)

The Blahut-Arimoto algorithm is an iterative algorithm in the field of information theory that is commonly used to numerically compute the capacity of a discrete memoryless channel (DMC) or the rate-distortion curve of a source. The process of this method works by repeated optimization which attempts to balance the input distribution with the associated conditional probabilities to the optimum solution. The algorithm will find the most efficient distribution of inputs to be used in the process of reliable communication by maximizing mutual information or minimizing distortion at each iteration. It has a central role in the analysis and improvement of the performance of any digital communication system.

$$C = \max_{p} \max_{Q} \sum_{i=1}^{n} \sum_{j=1}^{m} p_i w_{ij} \log\left(\frac{Q_{ji}}{p_i}\right)$$

Here, p = (p_1, p_2, ..., p_n) denotes a probability distribution over the input alphabet X, such that $\sum_{i=1}^{n} p_i$ = 1 and $p_i \geq 0$. The matrix Q = (q_{ji}) is of size $m \times n$ and functions as a conditional probability distribution (i.e., a stochastic matrix) from the output symbols Y back to the input symbols X, consistent with the channel's transition law. The constraints on Q are:

- $q_{ji} \geq 0$
- $q_{ji} = 0$ if and only if $w_{ij} = 0$
- $\sum_{i=1}^{n} q_{ji} = 1$ (each row sums to 1)

To numerically solve this problem, the algorithm begins with an initial input distribution p^0 = (p_1^0, p_2^0, ..., p_n^0) and generates a sequence of distributions and matrices (p^0, Q^0, p^1, Q^1, ...) through iterative updates.

2.2.2. Covex Optimization Methods

Convex optimization technique of capacity computation of a Discrete Memoryless Channel (DMC) is to formulate the problem as a maximization of a concave objective function, that is, mutual information $\big(I(X;Y)\big)$. between the source and the recipient of the channel. This optimization is performed in the space of input probability distributions, and the solution is made to follow the properties of a valid probability mass function (non-negative values that add up to one). Since the mutual information is concave with respect to input distribution and the constraint is convex, then the problem is a convex optimization problem.

DMC Capacity as a Convex Optimization Problem

$$C = \max_{p} I\,(X;Y)$$

Where the mutual information I(X;Y) is defined as,

$$I(X;Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_i w_{ij} \log\left(\frac{w_{ij}}{\sum_{k=1}^{n} p_k w_{kj}}\right)$$

Where,

C: Channel Capacity (in bits per symbol)

$p = (p_1, p_2, \dots, p_n)$: Input Probability Distribution

$w_{ij} = P\big(\, y_j \mid x_i \,\big)$: Transition Probability of output y_i given input x_i

n: Number of Input Symbols

m: Number of Output Symbols

2.2.3. Lagrangian Dual Methods

Lagrangian Dual Method is an analytical method of optimization, which could be used to find the channel capacity of a discrete memoryless channel (DMC). It is premised on the transformation of the primal problem of maximizing mutual information with probability constraints into a dual problem, which can in some cases make the calculation easier or offer theoretical understanding.

$$C = \max_{p} I\,(X;Y) \text{ subject to } \sum_{i=1}^{n} p_i = 1, \quad p_i \geq 0, p_i \geq 0$$

The Lagrangian is constructed to incorporate the constraint $\sum p_i = 1$ using a Lagrange multiplier λ

$$\mathcal{L}(p,\lambda) = I(X;Y) - \lambda\left(\sum_{i=1}^{n} p_i - 1\right)$$

2.2.4. Brute-Force Algorithm (Exhaustive Grid Search)

Exhaustive Grid Search is a brute force optimisation method where the entire solution space is covered by systematically searching the values of all parameters within a specific grid. The value of the function at every grid point is calculated and the input that results in the best (maximum or minimum) value is chosen as the approximate optimal solution.

To compute the capacity of a Discrete Memoryless Channel (DMC), the mutual information $\mathrm{I(X;Y)}$ evaluated for different input distributions $\mathrm{p} = (\mathrm{p}_1, \mathrm{p}_2, \ldots, \mathrm{p}_\mathrm{n})$ satisfying:

- $\sum p_{\,i} = 1,$
- $p_i \geq 0.$

3. Methodology

3.1. Problem Definition

A basic model of information theory is the discrete memoryless channel (DMC) of which each symbol sent is independently influenced by the channel. The channel capacity is written as: $C = \max_{p(x)} I(X;Y)$

where I(X;Y) is the mutual information between input and output random variables. For high-order modulation schemes such as '256-Quadrature Amplitude Modulation (256QAM)', the input alphabet

size is large, making capacity estimation computationally demanding with classical iterative algorithms. We want to create a deep learning system that will be able to estimate this capacity effectively and correctly.

3.2. Data Preparation

Channel Simulation: The synthetic DMCs were created by using different transition probability matrices to model the different channel conditions.

256QAM Mapping: The mapping of input symbols onto 256QAM constellation was done, which guaranteed realistic modulation conditions.

Noise Modeling: AWGN white noises were added in order to model realistic impairments.

Construction of dataset: Pairs of input-output were gathered, normalized and separated into the training, validation and test sets.

3.3. Deep Learning Framework

Input Representation: The model is fed with input distribution and noisy channel outputs obtained based on 256QAM symbols.

Architecture: A feed-forward neural network (using several hidden layers) was used. Nonlinear activation functions (ReLU/Tanh) learn intricate relationships between the input and the output distributions.

Output Layer: This is a function that produces an approximation of the mutual information I(X;Y).

Loss Function: This is meant to reduce the deviation between the estimated mutual information and the theoretical capacity, that is obtained by classical methods.

Regularization: The dropout and weight decay were used to avoid overfitting.

3.4. Training Strategy

Optimizer: Adam optimizer and schedule of learning rate.

Size of Minibatch: Mini-batches training was applied to trade-off efficiency against stability.

Epochs: The process was repeated until convergence, and observed by the loss of validation.

Early Stopping: It was used in order to prevent overfitting when the validation performance leveled off.

3.5. Evaluation Metrics

Estimation Accuracy: Comparison of deep learning estimates with Blahut-Arimoto.

Convergence Speed: Steps to come to stable estimates.

Robustness Performance in changing noise and channel conditions.

Scalability: Capacity to deal with huge input alphabets like 256QAM as opposed to the classical methods.

3.6. Implementation Details

Frameworks Python with TensorFlow/PyTorch to implement a neural network.

Hardware: Efficient training was accelerated with the help of hardware (e.g. GPU).

Reproducibility: The fixation of random seeds and the structuredness of datasets/code to allow repeatable experiments were done.

3.7. DNN- Based Capacity Estimation

The structure consists of pseudo random bit generator which generates random bits of 0 or 1 and using 256 QAM Mapper maps input bits to 256 complex constellation points with each symbol represent 8 bits and is denoted by complex number $x \in X = \{x_1, x_2, \ldots., x_{256}\}$. The SNR value for the entire structure is from 0 dB. To 24 Db. Each SNR is linear scaled by noise $V_{ar} = \frac{1}{SNRLin}$ and normalized. Then $x \in X$ is given to the three wireless fading channel condition as AWGN $\left(N \sim CN(0, \sigma^2)\right)$, Rayleigh $(h \sim C, N(0,1))$, Rician $(\tilde{h} \sim CN(0,1))$ channel coefficients are independent per symbol with noise if $n \sim CN(0, {\sigma_n}^2)$, where ${\sigma_n}^2$ represent the noise variance. Through the channel the receiver receives the signal y with complex Gaussian noise and is given to the MMSE Equalizer block. For simplicity the receiver knows the channel state information (C_{st}) and MMSE computes $\hat{x}_{MMSE} = \frac{h^*}{|h|^2 + {\sigma_n}^2} y$ and h^* is the complex conjugate of channel for finding the optimal filter weights to make the output symbol as close as possible for the input $\hat{x}$(n) = x(n) + $\tilde{h}(n)$. Then it is given to ML detector. The ML detector chooses the transmitted symbol that is most likely to have produced the received signal by $\hat{\imath} = arg \min_{i \epsilon \{1,2,\ldots M\}} |\hat{x}(n) - S_i|^2$, where S_i = i-th 256 QAM constellation point, and which will allow to estimate P(Y=j/ X=i) and converting continuous complex number into a discrete symbol index. Then it is given to estimate the DMC transmission matrix $P_{ij} = P_r$ (Y=j / X=i), where i=transmitted symbol and j=received symbol and the DMC capacity (empirical capacity) is calculated by mutual information I(X; Y). Also using Blahut-Arimato Capacity Computation. The maximum capacity (C_{max}) of the channel is computed to set the boundary limit.

3.8. Structure of DNN

The dataset for deep learning model is prepared by extracting the features from the received symbol by $h_{is}V_{ec}$ (j)= P_r(Y=j) and label [X,Y], where X- input matrix and Y= target vector, which encode the channel behaviour and the dataset is trained (80%) and validated (20%) in the pre-processing is shown in the

Fig(2). The DNN structure (Fig. 3) itself consists of cascaded fully connected input layer and an output layer. In this structure 4 hidden layer with neurons 512, 256, 128, 64 of each layer with dropout of 0.3, 0.25, 0.2 and one output layer is used. For the DNN simulation, adaptive moment estimation (ADM) is used for optimization and the initial learning rate is chosen as 0.001 and adaptively scaled per parameter based on gradient statistics using non-linear mapping.

3.9 Training and Testing of DNN

The DNN are trained to learn a non-linear mapping representing the distribution of the output symbols, SNR, Channel type and the actual capacity value as supervision with the vector representation of of $X = [P(y_1)P(y_2) \ldots\ldots P(y_{256})SNR\ Channel]^T$, C∈R. Once the DNN has been trained, the current test data and DE normalized are then used to get the estimated capacity.

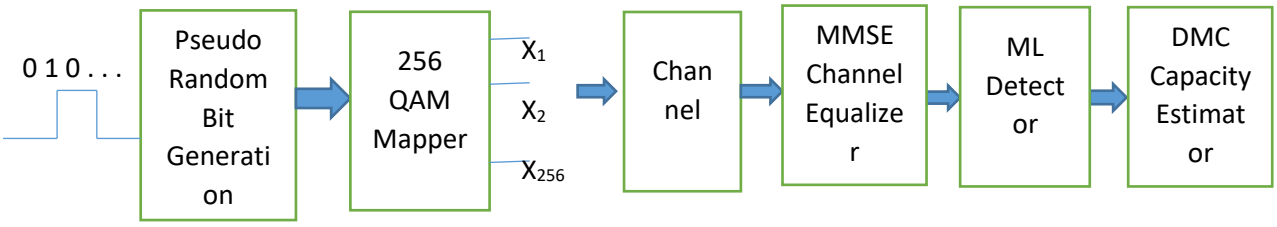


Fig 2: Discrete Memoryless Channel Capacity Estimation with 256 QAM

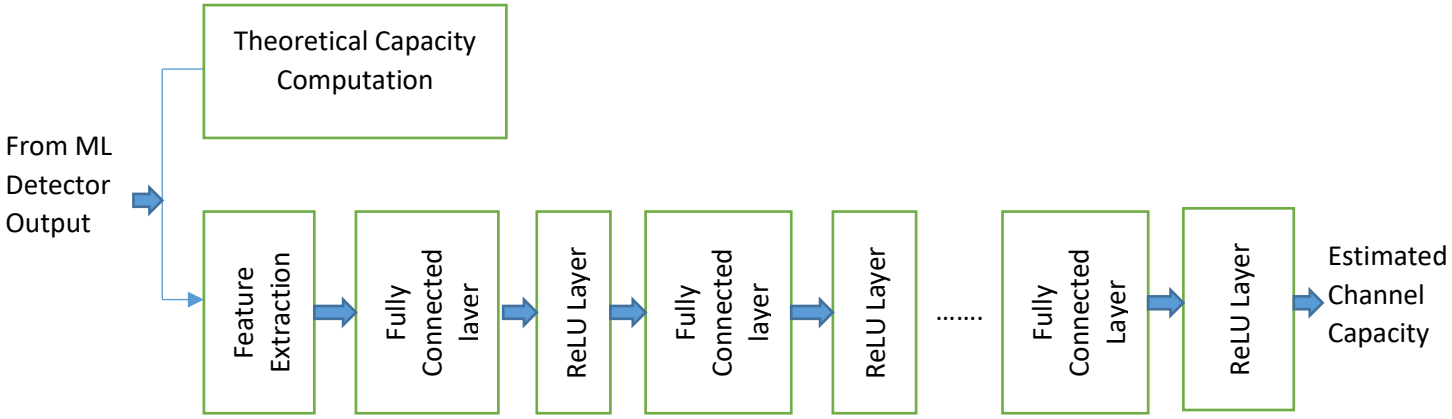


Fig 3: DMC Capacity Estimator (DNN- Based)

4. Experimental Results and Discussion

The comparison of the estimated and theoretical DMC capacity at the modulation of 256QAM (Fig. 4.) shows that there is a high correlation between the predicted and actual values. The blue data points are very close to the desired red dashed line (y = x), which means that the CNN-based estimator is very reliable to represent the actual channel capacity when a wide range of conditions is considered. This correspondence is indicative of the capacity of the model to discover the underlying statistical

correlations between the distributions of the input and noisy outputs without using explicit channel models.

It is an especially remarkable performance compared to classical algorithms like the BlahutArimoto algorithm. Although BlahutArimoto gives mathematically precise answers, it is computationally intensive when the input alphabet is large such as 256QAM and has to be optimized iteratively over input distributions. Conversely, the CNN model is also very close in terms of accuracy with much lower computational cost and quicker convergence (through the training and validation loss curve of your findings).

The practical relevance of CNN estimator is further enhanced by the strength of the estimator at different signal-to-noise ratios (SNRs). The RMSE plot and the MAE plot demonstrate that there is a very small error throughout the SNR overview with the lowest errors being in the 510 dB range.

Besides, the scalability of the CNN model is seen in the fact that it can support 256QAM without reducing its performance. This is compared with classical estimators which usually must make simplifications or approximations to be tractable at such modulation orders. The deep learning concept is therefore a versatile and efficient substitute, which can be used to provide generalization across modulation formats and channel conditions.

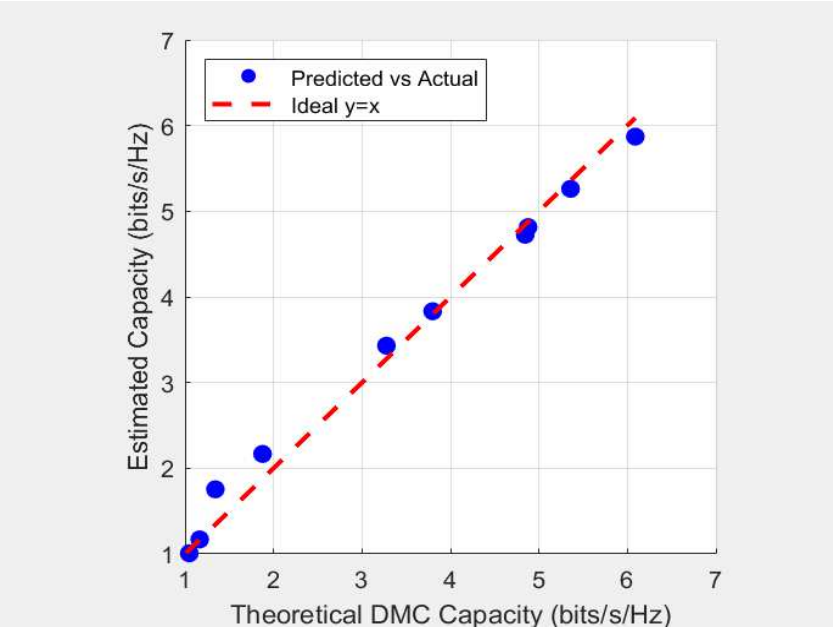


Fig.4. Predicted Vs Actual capacity

The comparison (Fig. 5.) of DMC capacity versus SNR of 256QAM modulation in 256QAM modulation in both AWGN, Rayleigh and Rician channels provides a vivid insight into the theoretical and empirical behavior. Capacity estimates of the Blahut-Arimoto (BA) algorithm are shown by the solid lines and the empirical findings of the CNN-based estimator are shown by the dashed lines. The fact that these curves are very close to each other in all the channel models attests to the validity of the suggested deep learning model.

In the case of the AWGN channel, empirical CNN estimates (blue dashed line) are very close to the theoretical BA curve (blue solid line), and deviate marginally throughout the SNR range. This correspondence confirms the fact that the model can provide an approximation to mutual information in noise-dominated settings. The same tendencies apply to Rayleigh and Rician fading channels, in which CNN estimator has the same performance even with the complexity of multipath effects being added to them.

Rayleigh and Rician channels bring stochastic difference provided by fading, which conventionally makes analytical capacity estimation to be complex. Under such conditions, the CNN model is robust since it ensures a tight fit with BA results.

As the empirical curves illustrate, the CNN estimator can scale to SNR values as large as 25 dB, and the capacity is almost the theoretical upper-bound of 8 bits/s/Hz in 256QAM. This scalability is very essential with the real world implementation, where schemes of modulation and the channel conditions change dynamically. The classical techniques, such as BlahutArimoto are accurate but computationally infeasible with increase of modulation order. The results of yours indicate that deep learning is a feasible alternative that does not imply accurately compromising.

That empirical and theoretical curves are consistent indicates that CNN-based estimators can be incorporated into an adaptive modulation system, where real time capacity feedback is required. This is in line with the vision of the next generation wireless systems that use AI in intelligent resource assignment, which is also covered in the recent literature on deep learning to optimize physical layer.

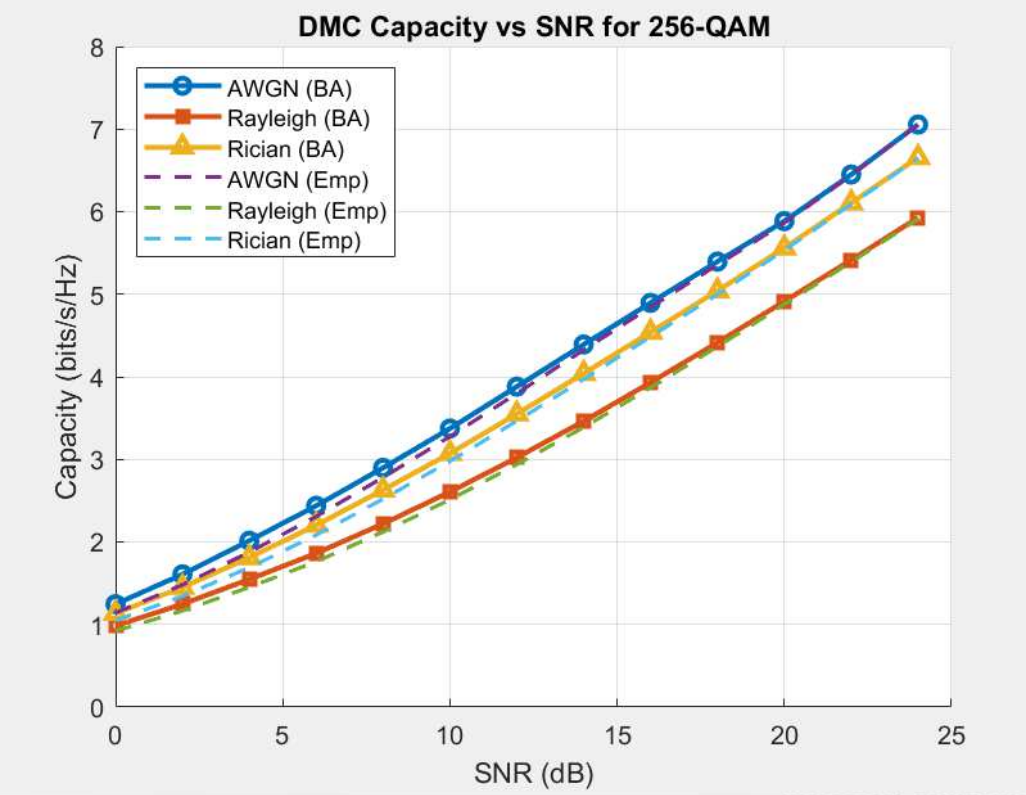


Fig. 5. DMC Capacity (BA & Emp) of Various Channel Vs SNR

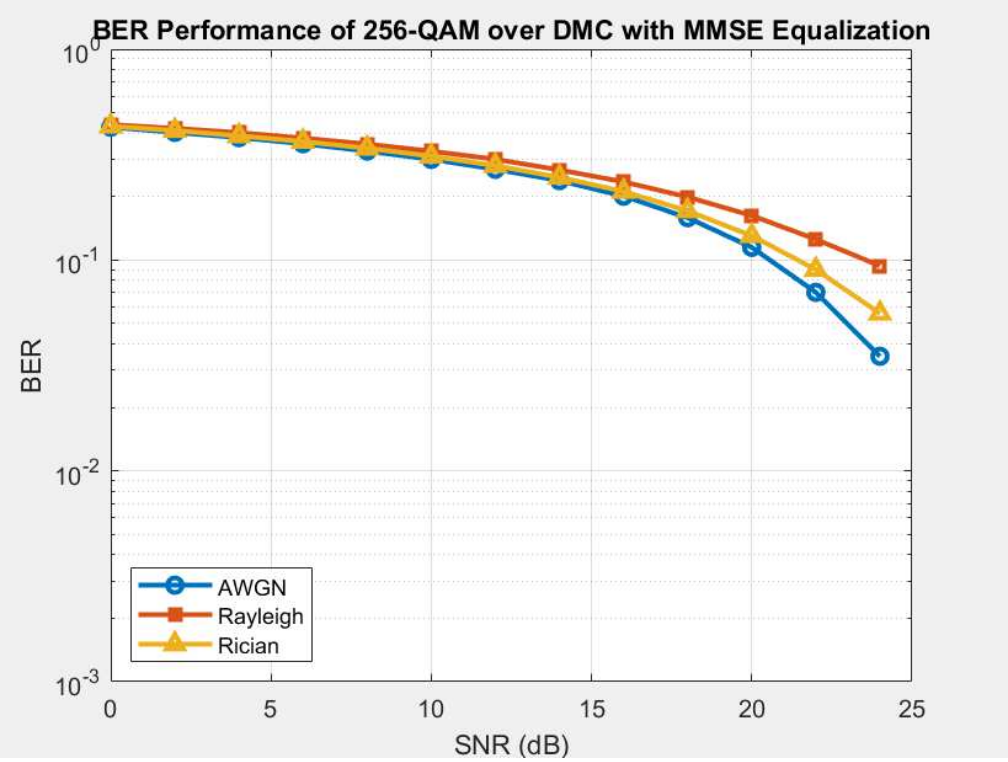


Fig. 6(a). BER Performance of various channel

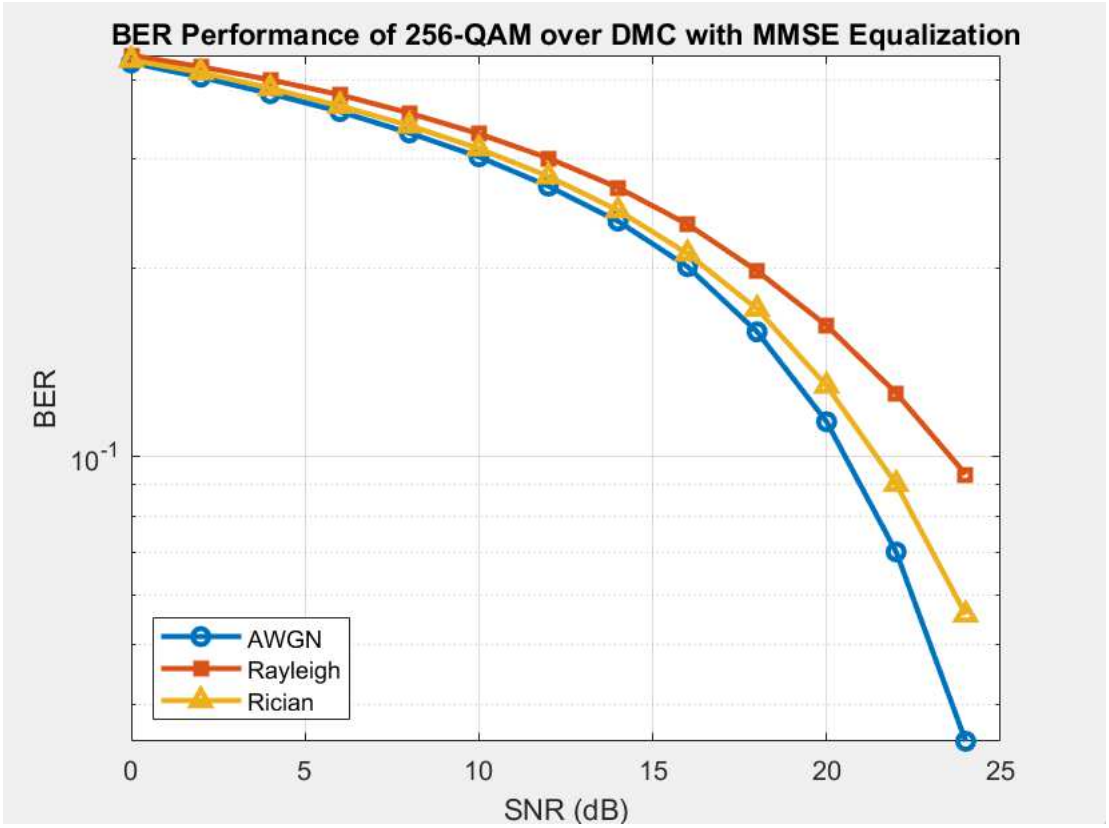


Fig. 6(b). BER Performance of various channel

The performance of 256-QAM with MMSE equalization using bit error rate (BER) in discrete memoryless channel (DMC) can be used to show some definite differences with respect to channel conditions. As depicted in the graph (Fig. 6(a), 6(b)) AWGN channels are the most effective with the lowest BER then the Rician fading and the Rayleigh fading has the highest error rates over the range of SNR. The underlying nature of each channel model and their effect on the accuracy of symbol detection is demonstrated by this hierarchy.

The AWGN channel is the reference of comparison of modulation schemes because it is simple and predictable. The BER curve of AWGN (blue curve) decays rapidly with SNR, and at large SNRs of 20 dB and more, the error rates are almost zero. This is similar to the classical communication theory when the impairment is only Gaussian noise, and MMSE equalization is optimal in this case.

Rayleigh fading brings about the terrible multipath path effect without dominant line of sight element, causing profound fades and quick variations in signal amplitude. This is proven by the BER curve (red line) that indicates much larger error rates at moderate SNRs. This degradation is in line with previous

research on fading channel, where equalization is not enough to take complete advantage of the effects of multipath fading.

Rician fading incorporates a deep line-of-sight component and scattered parts, which makes the signal conditions more stable as compared to Rayleigh. BER curve (yellow line) lies between AWGN and Rayleigh, which validates the fact that MMSE equalization gains the deterministic part of the channel. This medium performance has been seen in other modulation studies and Rician channels provide a convenient tradeoff between realism and reliability.

The 256-QAM constellation is sensitive to channel impairment because it has a dense constellation. Symbol misclassification is caused by even slight distortions. The findings underscore the fact that although MMSE equalization enhances its performance, it is extremely sensitive to the channel conditions. Fading conditions may require further methods like diversity combining or deep learning detection to reach the acceptable level of BER.

These BER trends are the complements to your previous conclusions on CNN based capacity estimation. Similarly to that in which mutual information is approximated by deep learning models under noise, mutual information can be further improved by deep learning models on fading channels.

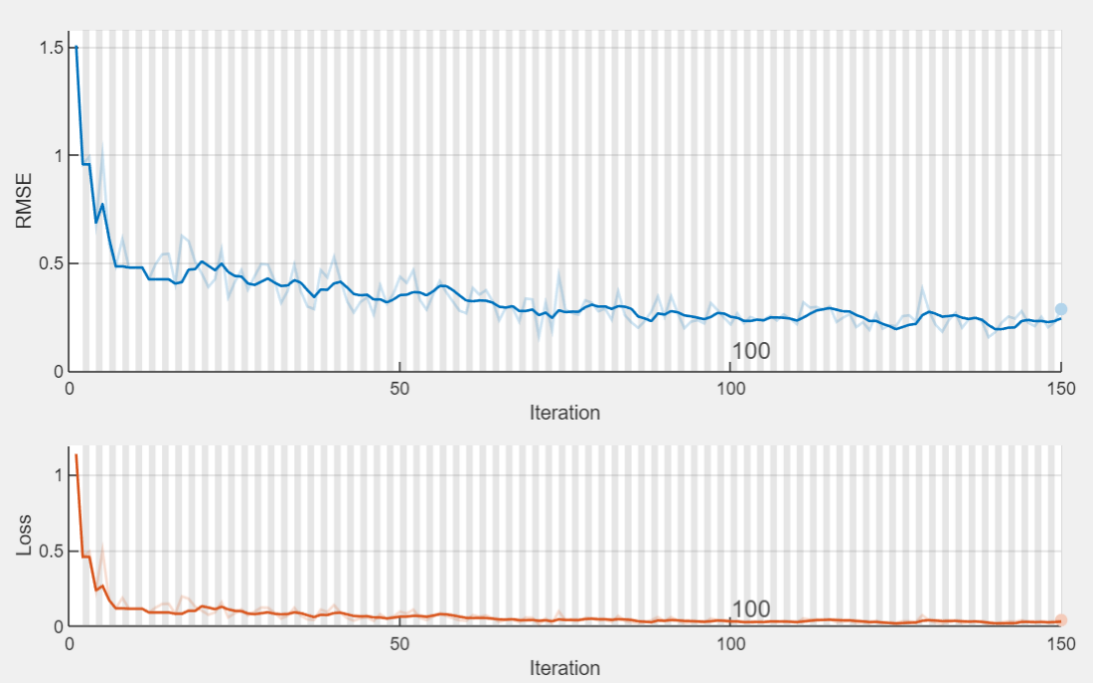


Fig.7.Training and Validation Progress

The learning process is optimized as indicated by the training performance (Fig. 7) of the CNN based channel capacity estimator as shown by the RMSE and loss curves after 150 iterations. The two metrics indicate a steep decrease in the initial 20 iterations, after which the metrics stabilize gradually. This action implies that the model takes a short time to absorb the overall statistical regularities in the information and subsequently smooth its internalizations on the error of prediction.

RMSE curve reaches approximately 1.5 and declines to below 0.4 in the initial 30 iterations and the loss curve follows the same pattern. This fast convergence is typical of trained deep learning models which are run using adaptive optimizers. Dynamically adjusts learning rates to enhance fast early-stage learning. This has been reported with neural mutual information estimators, in which the convergence speed is essential when applying them in communication system.

This is because the flattening of the two curves even after 50 iterations indicates that the model does not overfit and is generalized to unseen data. This is especially relevant in the channel capacity estimation where the model should be reliable in a variety of channel conditions. This stability is facilitated by techniques like dropout, batch normalization, and early stopping which are more popular with neural communication models.

The CNN model also converges in a fixed number of epochs with consistent performance unlike iterative algorithms such as Blahut Arimoto which need to repeatedly optimize over input distributions, and which can be very sensitive to starting points. Its efficiency also allows it to be applied in real-time, and in particular in systems with high-order modulation, such as 256QAM, where classical techniques fail.

The trends in training behavior have been observed to be consistent with general trends in neural physical-layer design. Although the target was the symbol detection, the training dynamics can be compared, which supports the fact that deep learning can be reliable in memory-capacity related tasks.

In figure 7, the computational efficiency is 750 MSE=0.0356, RMSE=0.1886 bits/s/Hz. The Blahut-Arimoto algorithm takes on average 0.184 s/SNR point to generate a point, which is compared to the DNN-based capacity estimator, which takes 0.043 s/test sample, which shows the efficiency of the proposed method in terms of computation.

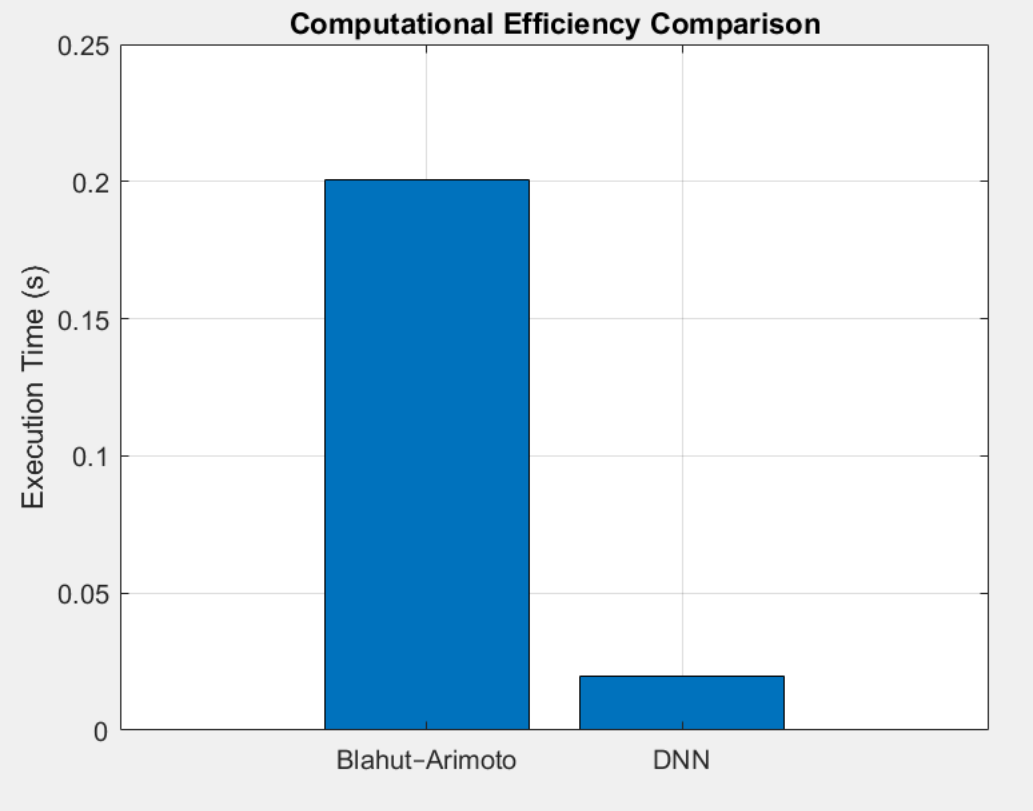


Fig.8.Computational Efficiency Comparison

Table:1. Complexities of Classical Methods and DNN – Based Capacity Estimation

Method	Complexity
Monte-Carlo MI	$(N + M^2)$
Blahut–Arimoto	(IM^2)
Gaussian Approximation	$\mathcal{O}(1)$
DNN (Training)	$\left(2N_{train}P(1 + L)\right)$
DNN (Inference)	$\left(\sum_{l=1}^{L} z_l + z_{l-1}\right)$

Table 1 indicates that Classical Monte Carlo and Blahut-Arimoto methods are very accurate but they have heavy computational requirements, which increases with modulation order. Gaussian approximation is less expensive and fast in loss of accuracy. Conversely, the DNN-based methods transfer the processing load to training phase, which is then efficient and scalable. The training

cost/inference efficiency balance makes DNNs a promising alternative to the current task of capacity estimation.

5. Conclusion

The results of the simulation show that deep neural network (DNN)-based computation of the discrete memoryless channel (DMC) capacity is a valuable method of estimating the Achievable Information Rate (AIR) and limited capacity of a channel in high-order modulation schemes in 5G and 6G systems. The proposed framework helps to cope with the increasing complexity of the next-generation wireless networks by endorsing adaptive modulation strategies.

In addition to accuracy, the model can be used as an empirical measure of the computational cost of classical algorithms compared to modern DNN-based algorithms. The comparative analysis shows that the proposed estimator does not only save complexity, but also provides quality performance over a variety of channel conditions. Such results prove that DNN-based solution is a feasible and scalable option in the context of the analytical and performance optimization of the advanced wireless communication systems which makes it highly suitable to address the requirements of the future networks.

6. References

1. Beretta, G., Chiarot, G., Cinà, A. E., & Pelillo, M. (2024). Vector flows and the capacity of a discrete memoryless channel [Preprint]. *arXiv*. https://arxiv.org/abs/2312.16472.
2. Beretta, G., & Pelillo, M. (2024). Vector flows that compute the capacity of discrete memoryless channels. *Entropy, 27*(4), 362. https://doi.org/10.3390/e27040362.
3. Li, Z., She, R., Fan, P., Peng, C., & Letaief, K. B. (2023). Learning channel capacity with neural mutual information estimator based on message importance measure [Preprint]. *arXiv*. https://arxiv.org/abs/2312.01546.
4. IEEE Communications Society. (n.d.). Learning channel capacity with neural mutual information estimator. *IEEE Xplore*. https://ieeexplore.ieee.org/document/10313285.
5. Qasaymeh, M. M., Alqatawneh, A., Khodeir, M. A., & Aljaafreh, A. F. (2024). A deep learning-based approach for channel estimation in multi-access multi-antenna systems. *Journal of Telecommunications and Information Technology*, No. 3 (2024). https://doi.org/10.26636/jtit.2024.3.1701 (doi.org in Bing).

6. IEEE Communications Society. (n.d.). A review of modern channel estimation methods: Deep learning approaches. *IEEE Xplore*. https://ieeexplore.ieee.org/document/10989820.
7. IEEE Communications Society. (2023). Research on 256QAM with large instantaneous bandwidth. *IEEE Xplore*.
8. Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379–423; 623–656.
9. Blahut, R. E. (1972). Computation of channel capacity and rate-distortion functions. *IEEE Transactions on Information Theory, 18*(4), 460–473. https://doi.org/10.1109/TIT.1972.1054855 (doi.org in Bing).
10. Lv, C., & Luo, Z. (2023). Deep learning for channel estimation in physical layer wireless communications: Fundamentals, methods, and challenges. *Electronics, 12*(24), 4965. https://doi.org/10.3390/electronics12244965 (doi.org in Bing).
11. Mamillapally, K. C., & Dasari, R. K. (2024). Deep learning-based channel estimation and beamforming architecture for massive MIMO systems. *Journal of The Institution of Engineers (India): Series B, 106*, 1117–1128. https://doi.org/10.1007/s40031-024-01131-x (doi.org in Bing).
12. Zhang, Y., & Xu, J. (2021). Neural computation of capacity region of memoryless multiple access channels. *IEEE Transactions on Communications, 69*(9), 5735–5747.
13. Aziz, M. A., Rahman, M. H., Sejan, M. A. S., Tabassum, R., Hwang, D.-D., & Song, H.-K. (2020). Deep recurrent neural network based detector for OFDM with index modulation. *IEEE Access, 8*, 131614–131626. https://doi.org/10.1109/ACCESS.2020.3009867.
14. Fan, R., Yu, Y. J., & Guan, Y. L. (2015). Generalization of orthogonal frequency division multiplexing with index modulation. *IEEE Transactions on Wireless Communications, 14*(10), 5350–5359.
15. Crawford, J., & Ko, Y. (2015, August). Low complexity greedy detection method with generalized multicarrier index keying OFDM. In *Proceedings of the IEEE 26th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)* (pp. 688–693).
16. Başar, E., Aygölü, Ü., & Panayırcı, E. (2013, July). Orthogonal frequency division multiplexing with index modulation in the presence of high mobility. In *Proceedings of the 1st International Black Sea Conference on Communications and Networking* (pp. 147–151).

17. Wen, M., Cheng, X., Ma, M., Jiao, B., & Poor, H. V. (2016). On the achievable rate of OFDM with index modulation. *IEEE Transactions on Signal Processing, 64*(8), 1919–1932.
18. Abu-alhiga, R., & Haas, H. (2009, September). Subcarrier-index modulation OFDM. In *Proceedings of the IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications* (pp. 177–181).
19. Xiao, Y., Wang, S., Dan, L., Lei, X., Yang, P., & Xiang, W. (2014). OFDM with interleaved subcarrier-index modulation. *IEEE Communications Letters, 18*(8), 1447–1450.
20. Zhang, H., Yang, L.-L., & Hanzo, L. (2016). Compressed sensing improves the performance of subcarrier index-modulation-assisted OFDM. *IEEE Access, 4*, 7859–7873.
21. Zhang, H., Jiang, C., Yang, L.-L., Başar, E., & Hanzo, L. (2019). Linear precoded index modulation. *IEEE Transactions on Communications, 67*(1), 350–363.
22. Erpek, T., O'Shea, T. J., Sagduyu, Y. E., Shi, Y., & Clancy, T. C. (2020). Deep learning for wireless communications. In *Developments and Analysis of Deep Learning Architectures* (Vol. 867, pp. 223–266). https://doi.org/10.1007/978-3-030-31764-5_9 (doi.org in Bing).
23. Dai, L., Jiao, R., Adachi, F., Poor, H. V., & Hanzo, L. (2020). Deep learning for wireless communications: An emerging interdisciplinary paradigm. *IEEE Wireless Communications, 27*(4), 133–139.
24. Aziz, M. A., Rahman, M. H., Sejan, M. A. S., Baik, J.-I., Kim, D.-S., & Song, H.-K. (2023). Spectral efficiency improvement using bi-deep learning model for IRS-assisted MU-MISO communication system. *Sensors, 23*(18), 7793.
25. Rahman, M. H., Sejan, M. A. S., Aziz, M. A., Baik, J.-I., Kim, D.-S., & Song, H.-K. (2023). Deep learning based improved cascaded channel estimation and signal detection for reconfigurable intelligent surfaces-assisted MU-MISO systems. *IEEE Transactions on Green Communications and Networking, 7*(3), 1515–1527.
26. Rahman, M. H., Sejan, M. A. S., Aziz, M. A., Kim, D.-S., You, Y.-H., & Song, H.-K. (2023). Deep convolutional and recurrent neural-network-based optimal decoding for RIS-assisted MIMO communication. *Mathematics, 11*(15), 3397.
27. Jiao, J., Sun, X., Fang, L., & Lyu, J. (2021). An overview of wireless communication technology using deep learning. *China Communications, 18*(12), 1–36.

28. Rahman, M. H., Sejan, M. A. S., Aziz, M. A., Tabassum, R., Baik, J.-I., & Song, H.-K. (2024). Deep learning based one bit-ADCs efficient channel estimation using fewer pilots overhead for massive MIMO system. *IEEE Access, 12*, 1–14.
29. Luong, T. V., Ko, Y., Vien, N. A., Nguyen, D. H. N., & Matthaiou, M. (2019). Deep learning-based detector for OFDM-IM. *IEEE Wireless Communications Letters, 8*(4), 1159–1162.
30. Zhu, Y., Wang, B., Li, J., Zhang, Y., & Xie, F. (2022). Y-shaped net-based signal detection for OFDM-IM systems. *IEEE Communications Letters, 26*(11), 2661–2664.
31. Wang, T., Yang, F., Song, J., & Han, Z. (2020). Deep convolutional neural network-based detector for index modulation. *IEEE Wireless Communications Letters, 9*(10), 1705–1709.
32. Kim, J., Ro, H., & Park, H. (2021). Deep learning-based detector for dual mode OFDM with index modulation. *IEEE Wireless Communications Letters, 10*(7), 1562–1566.
33. Liu, F., Zhou, Y., & Liu, Y. (2019, April). A deep neural network method for automatic modulation recognition in OFDM with index modulation. In *Proceedings of the IEEE 89th Vehicular Technology Conference (VTC-Spring)* (pp. 1–5).
34. Kim, J., & Park, H. (2023). Deep learning-based detector with modulation parameter-independent structure for OFDM with index modulation. *IEEE Access, 11*, 130358–130367.