JIMMA UNIVERSITY

JIMMA INSTITUTE OF TECHNOLOGY

FACULTY OF ELECTRICAL AND COMPUTER ENGINEERING

Performance Analysis of Deep Learning Detector for MIMO Cooperative Relay Communications with Imperfect CSI

By: NEWAY TESHOME

Advisor: DR. KINDE ANLAY (*Ass. professor*)

Co-Advisor: SOFIA ALI(*Msc*)

A thesis submitted to School of graduate studies, Jimma University in fulfillment of the requirements for Masters of Science

in the field of

Communication Engineering

March 16, 2022

Jimma, Ethiopia

JIMMA UNIVERSITY

JIMMA INSTITUTE OF TECHNOLOGY

FACULTY OF ELECTRICAL AND COMPUTER ENGINEERING

MASTERS THESIS ON

Performance Analysis of Deep Learning Detector for MIMO Cooperative Relay Communications with Imperfect CSI

By: NEWAY TESHOME

Advisor: DR. KINDE ANLAY (*Ass. professor*)

Co-Advisor: SOFIA ALI(*Msc*)

A thesis submitted to School of graduate studies, Jimma University in fulfillment of the requirements for Masters of Science

in the field of Communication Engineering

March 16, 2022

Jimma, Ethiopia

Declaration

I declare this thesis with the title of *" Performance Analysis of Deep Learning Detector for MIMO Cooperative Relay Communications with Imperfect CSI "* as my own work except where explicitly stated otherwise in the text and I assure it with my signature.

RESEARCH THESIS SUBMITTED BY:

NEWAY TESHOME GEMECHU	*Signed:* ________	*Date:* __ / __ / ____

APPROVED BY ADVISORS:

Main Advisor:	*Signed:*	*Date:*
DR. KINDE ANLAY (*Ass. professor*)	________	__ / __ / __
Co-Advisor:	*Signed:*	*Date:*
MS. SOFIA ALI	________	__ / __ / ____

APPROVED BY THE BOARD OF EXAMINERS:

Chairperson:	*Signed:* ________	*Date:* __ / __ / ____
Internal Examiner:	*Signed:* ________	*Date:* __ / __ / ____
External Examiner:	*Signed:* ________	*Date:* __ / __ / ____

Abstract

Cooperative communication is one of the promising approaches for achieving high data rates and efficient bandwidth utilization, but introducing relay nodes in the architecture brings a challenge in physical layer security. Scholars propose different approaches like a secure beamforming model and a combination of beamforming and jamming using artificial noise to overcome this challenge. The channel state information (CSI) of the eavesdropper and the legitimate user is necessary for the secrecy of the transmission, but in reality, the eavesdropper is always passive, and the channel state information is difficult to obtain, and the channel state information of the legitimate user is outdated. This thesis proposes a secure multiple input multiple output (MIMO) communication system to overcome security threats during cooperation with the relay node. A zero-forcing algorithm is used to secure leakage to the eavesdropping relay node by transmitting on null space using the beamforming technique. The deep convolutional neural network (DCNN) is trained with the imperfect version channel state information to produce the perfect channel state information then the input bit is recovered using a maximum likelihood detector. The Simulation was done for different performance factor parameters like imperfect correlation factor, doppler frequency, and the number of antennas to show the BER performance of the system. The results show that the deep convolutional neural network detector has a gain performance about 2dB in higher correlation factor and about 10.5dB in lowest imperfect correlation factor than the standard maximum likelihood detector.

Keywords: Deep learning, Deep CNN, Cooperative relay, AF protocol, DF protocol, MIMO communication, imperfect CSI, channel estimation, physical layer security .

Acknowledgements

First and foremost, thank you to Almighty God, and I'd like to thank my teacher Dr. Kinde Anlay (Ass. Professor) for teaching the majority of the course that is related to my thesis and advising me. And many thanks to my co-adviser Sofia Ali for her guidance and assistance with this project. In addition, I'd like to thank my employer and sponsor South Radio and Television Agency, as well as my family and friends.

Contents

List of Figures

List of Tables

List of Abbreviations

AI	**A**rtificial **I**ntelligence
AdaDelta	Adaptive Delta
AdaGrad	Adaptive gradient
Adam	Adaptive Moment Estimation
AF	Amplify-and-forward
ASK	amplitude shift keying
BER	bit error rate
BS	Base station
CC	Coded Cooperation
CSI	Channel state information
DCNN	Deep convolutional neural network
DF	Decode-and-forward
DFNN	deep feedforward neural network
FSK	frequency shift keying
LTE-A	Long-Term Evolution-Advanced
MGF	moment generating function
MIMO	Multiple-Input and Multiple-Output
ML	Maximum likelihood
MMSE	Minimum mean squire error
PDF	probability density function
PSAM	pilot symbol assisted modulation
PSK	phase-shift keying
QAM	Quadrature amplitude modulation
QSK	Quadrature phase-shift keying
RMSProp	Root Mean Square Propagation
RN	Relay node
SER	symbol error rate

SNR	Signal to noise ratio
SOP	closed-form secrecy outage probability
SVM	support vector machines
UE	user equipment
ZF	Zero force
α	channel attenuation
λ	wave length
d	path distance
Tc	coherence time
fd	Doppler spread
Ts	symbol duration
Jo(.)	zero order Bessel function
fc	coherence bandwidth
τ**max**	maximum delay spread
Nt	Number of transmitting antenna
Nr	Number of receiving antenna
H	channel fading matrix
y	received signal
x	input signal
$\hat{x}$	recovered signal

Chapter 1. Introduction

1.1 Introduction

With globalization, present-day wireless networks are facing high traffic demands. Different improvement techniques, such as optimizing, changing the architecture, or combining different networks, are used to meet those needs. Cooperative communication is one of the promising approaches for achieving high data rates as well as efficient utilization of the bandwidth. Cooperative communication uses a relay node (RN), to provide coverage in the holes within the Long-Term Evolution-Advanced (LTE-A) cellular networks [1]. Similarly, relaying techniques are used in 5G mm-wave communication to overcome various challenges such as link blockage, back-haul connectivity, and path loss [2]. But this combination of the network architecture brings a challenge in physical layer security [3]. In a cooperative network, security constraints and measurement must be taken into account for reliable and efficient end-to-end communication. In some scenarios, the eavesdropping link node may have a good channel quality for attracting in the selection process of the RN and aims to acquire the information during the communication process [4]. There are different techniques which overcome the different security problems in the network. One is directly facing the eavesdropper characteristics to achieve the higher secrecy capacity by artificial noise during the transmission of data to confuse the eavesdropper elements in the network [5] or by combining the optimal relay selection method with artificial noise increase secrecy capacity [6] and also by selecting the best relay, based on full and statistical eavesdropping CSI to derive closed-form secrecy outage probability (SOP) [7]. One of the key points for secure cooperative communication is the CSI between the relay and the legitimate user. To acquire the CSI most the time it uses pilot-base estimation method like maximum likelihood, least square and minimum mean square error but the performance of the algorithms degrades due to imperfection factors. The performance degradation can be optimized with the aid of mathematical models and expert knowledge, which heavily relies on the channel model and estimation theory. But this mathematical model cannot cope with an excessively complex scenario.Especially when the channel state information is dynamically changing over time. Currently deep learning-based detectors shows a greater performance in communication field [8]. Motivated by that we proposed deep learning-based detector for MIMO cooperative relay network with imperfect CSI.

1.2 Statement of the problem

Cooperative communication uses a relay node which is used to overcome different challenges of link blockage, backhaul connectivity, and path loss, etc. In some scenarios when the legitimate relay transmits a secure message to the receiver user, the eavesdropper relay node and eavesdropper user may intercept the message. Various researchers optimize the security rate by addressing the eavesdropper effect and/or optimizing channel estimation algorithms. maximum likelihood detector has greater performance than other with perfect CSI but the CSI obtained by the receiver is imperfect due to the time difference between channel estimation and data transmission. The performance of maximum likelihood detector will degrade due to imperfection factors.research shows, if the channel estimation is supported with deep learning the performance can be enhanced.So We are Motivated to study the performance of DCNN type detector for MIMO cooperative relay communication.

1.3 The objective of the research

1.3.1 General Objective

The objective of the study is to Secure MIMO Communications with imperfect CSI for cooperative relay Networks using Deep Learning.

1.3.2 Specific Objective

The specific objectives of the study include:

- To develop a MIMO cooperative relay communication that avoids information leakage to the eavesdropper node.
- To estimate the perfect CSI between the relay and the legitimate receiver using deep convolutional neural network which increases the BER performance of the receiver.
- To enhance the overall performance of cooperative relay communication

1.4 Methodology

In order to achieve the objectives described above, the following techniques has been used. First of all, reviewed the previous works related to this work. This include study of MIMO communication, cooperative relay communication, and application of Deep learning in communication.
Secondly, asses starting from the reason for requiring of new channel estimation method up to identifying the proposed candidate channel estimation method. Then the candidate channel estimation method formats will be reviewed in order to understand their operation principles and mathematical formulations.
The next step is the system development, which is common frame work for the comparison.

1.4.1 Developing system setup

A communication scenario will be setup in which the CSI is outdated at the receiver. The mathematical model for calculating the received signal will be outlined which helps to evaluate the BER performance and identify the input parameters for the simulation.

1.4.2 Developing and training proposed DCNN

Python simulation software is implemented to build model of deep learning estimation method and conventional maximum likelihood estimator in order to study their respective bit error rate for different performance factor parameters. The DCNN model first trained using the imperfect version of CSI to produce perfect CSI for different performance factor parameters.

1.4.3 Data preparation

The data set is a channel fading matrix between the sender, relay and the destination. To acquire training and testing data, continuous-time channel responses are sampled, adhering to the assumption of the selected communication scenario. The channel fading matrix will be reshaped as a column vector to form a one-dimensional array and the real and imaginary part of the channel fading matrix will be separated and form RIRI in preprocessing of data set. The generated data set will be Split into training, validating, and test sets to fed to the DCNN model. To generate the data a python software with tensor-flow framework will be used.

1.4.4 Result analysis

By plotting BER graphs we can make a reasonable comparison between the both estimator, which are useful in showing how the transmitted signal distorted by imperfect CSI. Finally, perform result analysis and interpretation

1.5 Significance of the Study

This study will decrease the security threats between a legitimate user and relay node which is occurred due to the open nature of the cooperative relay network. Because this study considers imperfection of the channel, the performance degradation due to imperfect CSI will be improved and will have a significant role to enhance the overall performance of Wireless cooperative relay communication capacity.
The contribution of this thesis is

- It uses the application of deep learning specifically DCNN estimator to enhance the BER performance of both AF and DF cooperative relay network with imperfect channel condition.this include preparation of data and training of the DCNN network for both protocol.

- The system considers the eavesdropper relay nodes security threat during the communication between legitimated relay and receiver.

1.6 Scope and limitation of the study

This thesis focuses on imperfect channel estimation for a MIMO DF-based cooperative relay network, which is accomplished by training the DCNN with an imperfect channel fading matrix between the relay and the source and destination independently, and then detecting the information symbol using a maximum likelihood detector. The addition of DCNN will increase the computational complexity of the system, but the performance degradation caused by the imperfect factor will be reduced.

1.7 Organization of the Thesis

This thesis work contains six chapters. The first chapter is the introduction part which contains a motivational overview, statement of the problem, objective, methodology, scope, and significance of the thesis. The second chapter discusses technical backgroung which contain overview of MIMO communication, cooperative relay communication and the application of deep neural networks .The third chapter is about literature review. Chapter four deals with proposed system, it contains the system model, DCNN architecture, and detection method. Chapter five is about simulation results and discussions and the last chapter is conclusion and recommendation.

Chapter 2. Technical background

2.1 Overview of MIMO Wireless communication system

MIMO communication refers to a link for which the transmitting end as well as the receiving end is equipped with multiple antenna elements. To transmit digital information through wireless channel needs certain procedures.

2.1.1 Modulation technique

Transmitting digital information through channels needs modulation. Modulation is the process of changing amplitude (amplitude shift keying (Ask)), frequency (frequency shift keying (FSK)) or phase (phase-shift keying (PSK)), or combination of them (quadrature amplitude modulation (QAM)) of the analog signal with respect to the digital information bit or symbol. This modulation techniques use two orthogonal sinusoidal signals which makes the modulated signal complex signal. Using these two orthogonal signals we can make a different constellation which is known as M-array. where M indicates the constellation size. PSK is an angle-modulated, constant-amplitude digital modulation technique.It is an M-array digital modulation scheme with M=2,4,8 called binary phase-shift keying (BPSK), quadrature phase-shift keying (QPSK), 8 phase-shift keying (8PSK), and so on respectively. In choosing modulation scheme we need to know the tradeoff between the BER performance and data rate performance. As the modulation order increases the data rate will be increase but the the BER performance will decrease so, for this thesis, We employ a QPSK modulation scheme in which the binary input data is divided into two-bit groups known as di-bits. Each di-bit code generates one of the four possible output phases $(+45, +135, -45, and - 135)$ but the system can work for any modulation scheme. [9]

2.1.2 Wireless Channel Model

MIMO systems are wireless transmission schemes that operate in the absence of direct line of sight and rely on multi-path propagation [10].

Multi-path propagation

Properties of multipath propagation include amplitude fade and phase variations, time and power delay spread information, angle of entry and exit, Doppler shift effect, and the amount of multipath components. The component of the received signal through different paths due to environmental effects like reflection, diffraction, and scattering implies add constructively so that the received signal is large or they add destructively, resulting in a very small or practically zero signal. Mathematically written as (2.1) where α is channel attenuation, λ is wave length and $\mathbf{d}$ is path distance.

$$g = \sum_{i=1}^{L} \sqrt{\alpha_i} e^{-j2\frac{(d_i - d)}{\lambda}} \tag{2.1}$$

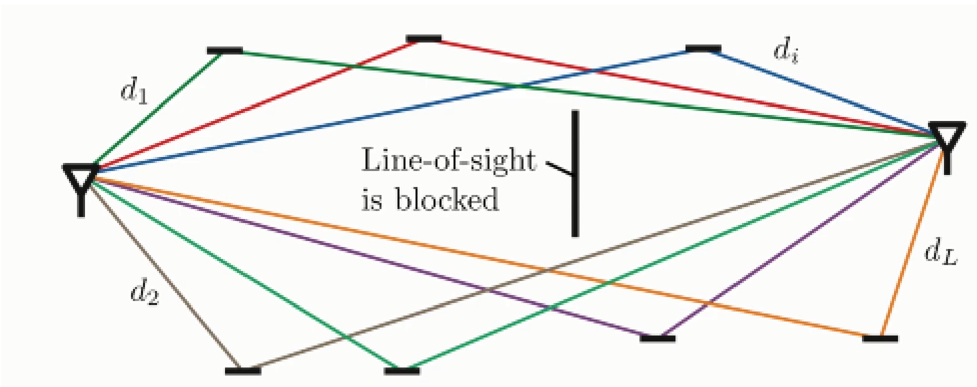


FIGURE 2.1: Non line of sight communication.

There are certain characteristics to consider while we model fading channels.

Slow and Fast Fading

The slow and fast fading scenario is related to the coherence time ***Tc***, which measures the period after which the correlation function of two samples of the channel response taken at the same frequency but at different time instants falls below a predefined value. The channel coherence time is also related to the channel Doppler spread ***fd*** which occurs due to the relative speed of the elements in the communication system.

$$T_c = 1/f_d \tag{2.2}$$

In slow fading, the symbol duration is significantly shorter than the channel coherence time, implying that the channel remains constant over an entire symbol period.

In fast fading the symbol duration is higher than the coherence time of the channel so, the communication will experience different channel fading coefficients. The detection decisions in a fast-fading scenario are based on the received signal with different symbol times. As a matter of fact, proper correlation models must be used to give an explanation for the fading channels' variation. This is accomplished through the use of a variety of correlation models, the majority of which are determined by the propagation environment and the underlying communication scenario. Because this thesis is concerned with the land mobile environment, the correlation coefficient factor between adjust sample is modeled as (2.3)[11].

$$\rho = J_0(2f_dTs) \tag{2.3}$$

where ***fd*** is doppler frequency, ***Ts*** is symbol duration and ***Jo(.)*** is zero order Bessel function

Frequency Flat and Frequency Selective Fading Channel

channels coherence bandwidth ***fc*** is defined as the frequency bandwidth over which the correlation function of two samples of the channel response is taken at the same time but different frequencies fall below a suitable value. The relationship between the coherence bandwidth and the maximum delay spread τ_{max} is given as:

$$f_c = 1/\tau_{max} \tag{2.4}$$

If the signal bandwidth is significantly smaller than the coherence bandwidth of the channel ***fc***, all the frequency components of the transmitted signals are passed through the same channel. This is called flat fading. In other cases, If the signal bandwidth exceeds the coherence bandwidth of the channel, the transmitted signal is modified with different amplitude gains and phase shifts; this is known as frequency selective fading. The fading characteristics is simply summarized as figure 2.2

Various models are used to describe the statistical behavior of multi-path fading, depending on the nature of the radio propagation environment.

Rayleigh, Rician and Nakagami Fading Channel

The wireless channel has a multiplicative effect on the transmitted signals in frequency flat fading channels, where the multiplicative term is a complex Gaussian random variable. If the mean of the channel coefficient is zero, the channel is considered Rayleigh fading because the absolute value of the channel gain is a Rayleigh

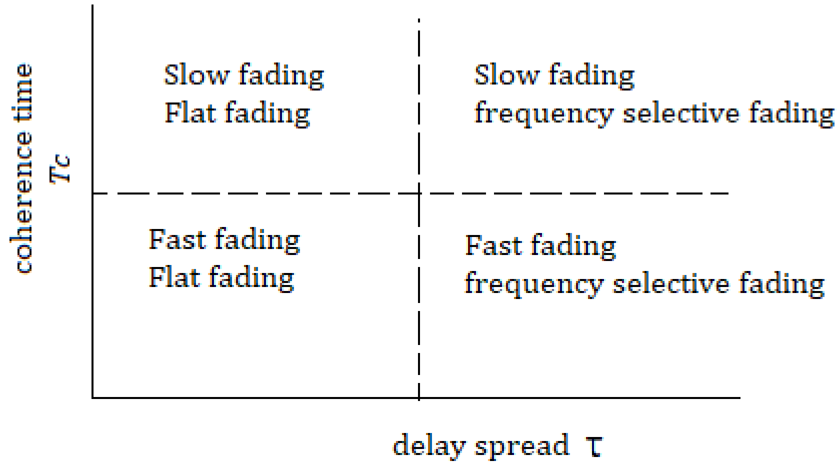


FIGURE 2.2: characteristics of fading channel delay spread vs coherence time.

random variable [12].If the mean of the channel gain is non-zero, its absolute value is Rician distributed, and the channel is said to be Rician fading. Another popular fading channel model is Nakagami fading, which is based on experimental observations rather than theoretical models like the ones used to develop Rayleigh and Rician models.[13].

2.1.3 MIMO system

MIMO system considers an antenna array with $N_t \times N_r$ transmitting and receiving elements. For the given j^{th} transmitting element and the i^{th} receiving elements, the channel impulse response between them is named as $h_{i,j}(\tau,t)$. The MIMO channel can then be described by the $N_t \times N_r$ size of $H(\tau,t)$ matrix.

$$H(\tau,t) = \begin{bmatrix} h_{11}(\tau,t) & h_{12}(\tau,t) & \dots & h_{1Nt}(\tau,t) \\ h_{21}(\tau,t) & h_{22}(\tau,t) & \dots & h_{2Nt}(\tau,t) \\ \vdots & \vdots & \ddots & \vdots \\ h_{Nr1}(\tau,t) & h_{Nr2}(\tau,t) & \dots & h_{NrNt}(\tau,t) \end{bmatrix} \tag{2.5}$$

The matrix elements are complex numbers that correspond to the propagation loss and phase shift introduced by the wireless channel to the signal arriving at the receiver with delay τ

The input output relationship can be described as

$$y = Hx + n \tag{2.6}$$

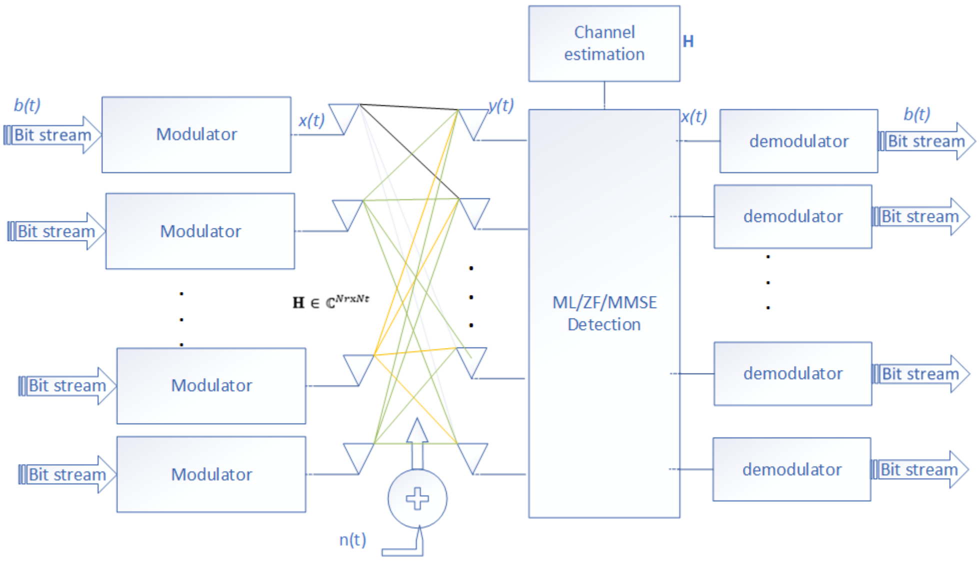


FIGURE 2.3: Block Diagram of M IMO wireless network[14] .

where $y \in \mathbb{C}^{Nr}$ received signal, $x \in X^{Nt}$ the input signal and H is $\mathbb{C}^{Nr \times Nt}$ channel fading matrix, and $n \sim \mathbb{C}N(0, I)$ the received noise.
The detection process of the transmitted symbol $\mathbf{x}$ is depending on the estimated channel. The estimation process done either by blind estimation method like Bussang algorithm and sub-space based which is carried out by evaluating the statistical information of the channel and particular properties of the transmitted signals. This blind channel estimation has no overhead loss and it is only suitable for slowly time varying channels [15]. Or pilot-base estimation method like least square and minimum mean square error,[16]–[18]. In training-based channel estimation algorithms, the transmitter will transmit training symbols or pilot tones that are known to the receiver, then the receiver will use the channel state information to detect the message signal.
The detection for the transmitted vector $\hat{x}$, based on its knowledge of the channel matrix H, x, and the observation y will be calculated using different algorithm as seen equations 2.7, 2.8, and 2.9

Maximum likelihood detection (MLD)

MLD is the optimum in terms of minimizing the overall error probability because the minimization is with all possible transmitted vectors [19].

$$\hat{x} = \arg \min_{x \in X^{Nt}} \|y - \hat{H}x\|^2 \tag{2.7}$$

where X^{Nt} is all possible constellation set, $\hat{x}$ is the recovered signal at receiver, $\hat{H}$

is estimated channel fading matrix using training pilot symbols. due to computing the function for all possible constellation set of potential value of $\boldsymbol{x}$, MLD has higher complexity than zero force and Minimum mean squire error.

Zero forcing (ZF) at the receiver

To reduce the complexity of MLD linear receiver like zero forcing is introduced [19]. It involves design a matrix which is the inverse of channel fading matrix H^{-1} and multiply with the received signal. If the Matrix **H** is well conditioned, we get a good bit error rate but if the channel fading matrix **H** is ill conditioned which means if **H** is closed to zero the inverse will be closed to infinity which causes the noise to amplify

$$\hat{x} = H^{-1}(Hx + n) = x + H^{-1} \tag{2.8}$$

Minimum mean squire error (MMSE)

To maintain the ill-conditioning of the matrix Hin order to reduce the sensitivity of linear receivers, regularization term will be added

$$\hat{x} = H^H(H^H H + \lambda I)^{-1} y \tag{2.9}$$

Where, λ is regularization weight and I is identity matrix with the same size of H. Since it minimizes the mean squared error in the estimate of $\boldsymbol{x}$, it is called linear minimum mean squire error detector (LMMSE)[20],[21]

Neural Network based detector

This method estimates the unknown channel response at non-pilot subcarriers by leveraging knowledge of pilot channel properties. This estimator learns to adapt to channel variations before estimating channel frequency response. This method is less complex and high quality than conventional methods such as least Square (LS), Minimum Mean Square Error (MMSE).[22]

2.2 Overview on Deep Learning

Deep learning is a subset of machine learning. It is based on the idea that systems can learn from data, identify patterns and make decisions. The leaning process can be categorized in to supervised, unsupervised and reinforcement.
In supervised learning a labeled data is used to train the network and develop a function that govern the input/output relationship then it will predict the value of the label for an input data that is not in the training set. The prediction is classification, if the label is discrete and regression, if the label is continuous. For this thesis we use unsupervised learning to predict the CSI of the system. Where the system learn the association between the input data. Reinforcement learning based on rewarding and punishing depending to the desired.
Deep learning has its origins in early work that tried to model networks of neurons in the brain with computational circuits. For this reason, the networks trained by deep learning methods are often called neural networks [23]. A single neuron in deep learning is constructed as figure 2.4 with input **x**, weight **w**, bias **b**, activation function $f(.)$

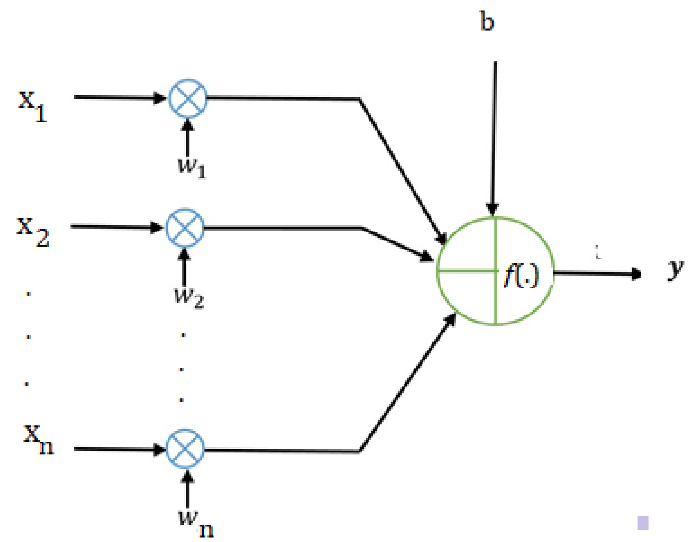


FIGURE 2.4: Structure of single neuron in deep learning.

Mathematical represented as (2.10)

$$y = f((x_1w_1 + x_2w_2 + x_nw_n) + b) \tag{2.10}$$

The activation function was included to give the function non-linear behavior, allowing the network to learn more complex things. By creating the corresponding output, the activation function determines whether a neuron will respond or not for a particular input. The most commonly used activation functions in deep neural networks are summarized on table 2.1.

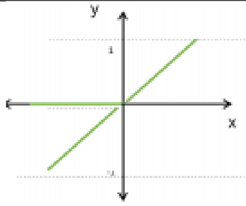
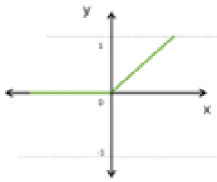
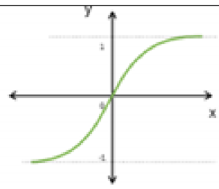
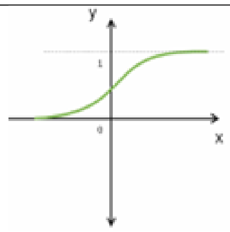
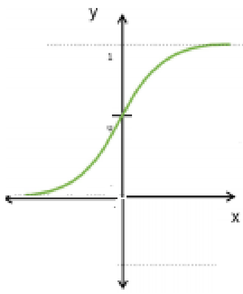
Activation Function	**Mathematical Model**	**Range and graph**
Linear	$f(x) = ax$	$(-\infty, \infty)$ 
ReLu	$f(x) = max(0, x)$	$[0, \infty)$ 
tanh	$f(x) = tanh(x)$	(-1, 1) 
Sigmoid	$f(x) = \frac{1}{1+e^{-x}}$	(0, 1) 
Softmax	$f(x) = \frac{e^{x_i}}{\sum_{j=1}^{k} e^{x_j}}$	(0, 1) 

TABLE 2.1: List of some activation functions and their characteristics for deep neural networks

Deep neural network is constructed from multiple neurons and layers. It has the input layer, hidden layers (one or more than one depending on the depth of the information to be extracted) and the output layer. If the connection of the neurons and propagation of the signals are only in forward direction it is categorized as feed forward neural network (FFNN). If it has feedback to previous neuron the network is categorized as Recurrent neural network. A FFNN type Convolutional neural network is used in this thesis.

2.3 Convolutional neural network

Convolutional Neural Network (CNN), also known as ConvNet, is a form of Artificial Neural Network (ANN) with a deep feed-forward architecture and incredible generalization power. It can learn highly abstracted aspects of objects, particularly spatial data, and recognize them more effectively. A deep CNN model is made up of a finite number of processing layers that can learn many levels of abstraction from incoming data. The higher-level features (with lower abstraction) are learned and extracted by the initiatory layers, while the lower-level characteristics are learned and extracted by the deeper layers (with higher abstraction). The basic conceptual model of CNN was shown in figure 2.5.

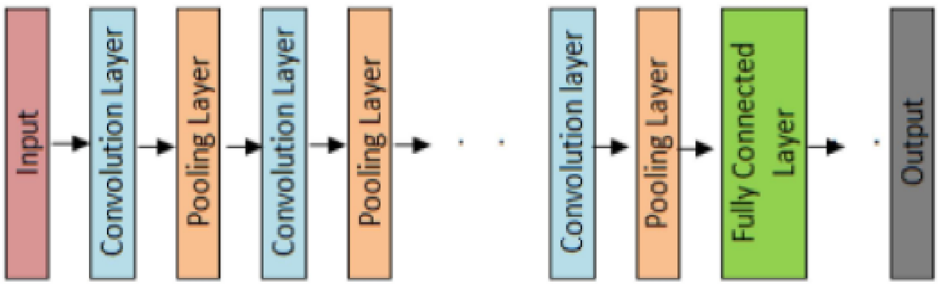


FIGURE 2.5: Conceptual model of CNN [24].

One of the essential building elements of a convolutional neural network is the convolution process. The parameters of the convolutional layers are made up of a collection of learnable filters (kernels). The pooling layers are used to sub-sample the feature maps, which means they shrink the larger feature maps to smaller feature maps.While decreasing the feature maps, the most prominent features in each pool step are always preserved.

Convolution operation, is the process of taking the kernel with filter size and slide it over all data set horizontally as well as vertically with given stride and by multiplying the corresponding values of the kernel and the input data set and we sum up all values to generate one scalar value in the output feature map. This process continues until the kernel can no longer slide further. The convolution process will eliminate the border of the data set to overcome this problem padding is used to give border size information of the input data more importance. The padding is also used to increase the input data size, as a result the output feature map size also gets increased.

The feature map size after convolution operation can be calculated as 2.11 for height and 2.12 for width of feature map size.

$$\hat{h} = \frac{h - f + p}{s} + 1 \tag{2.11}$$

$$\hat{w} = \frac{w - f + p}{s} + 1 \tag{2.12}$$

Where $\hat{h}$ and $\hat{w}$ is the height and width of the output feature map, h and w is the height and width of the intput data, f is the filter size, p is the padding and s is stride of convolution operation.

We use a Loss Function to calculate the prediction error created by the CNN model over the training data at the output layer. This prediction error indicates how far the network's prediction is off from the actual output.If your prediction is completely wrong, your loss function will produce a higher number. If they are good enough, it will generate a lower number. As you change parts of your algorithm to try to improve your model, the loss function tells you where you're going. Some of the most used loss functions are

2.4 Loss functions

Cross-Entropy: also called log loss function is widely used to measure the performance of the CNN model, whose output is a binary number (0,1) mathematically expressed as

$$H(P,y) = -\sum_{i=1}^{N}(P_i)log_2(y_i) \tag{2.13}$$

where P is the probability for each output category and y denotes the desired output.

Hinge loss function: is utilized in maximum margin classification problems, particularly for support vector machines (SVMs). The optimizer is attempting to maximize the margin between two target classes in this case.

$$H(\hat{y},y) = \sum_{i=1}^{N} max(0, m(2y_i 1)\hat{y}_i \tag{2.14}$$

Where m is the margin, $\hat{y}_i$ denotes the predicted output and y_i denotes the desired output.

Mean squared error(MSE): is commonly used in regression problems. The mean squared error between the predicted output $\hat{y}$ and the actual output y

$$H(\hat{y}, y) = \frac{1}{2N} \sum_{i=1}^{N} (\hat{y}_i y_i)^2 \tag{2.15}$$

where N is number of neuron in output layer

2.5 Optimizer selection

The model parameters are adjusted constantly during each training epoch to reduce error, and the model iteratively searches for the locally optimal solution in each training session. The learning rate is the size of parameter updating steps, and a training epoch is a complete iteration of parameter update that contains the entire training data set. There exist a number of variants of the gradient-based learning algorithm, out of them the most widely used are:

2.5.1 Batch Gradient Descent

The gradient is calculated throughout the entire training set and then used to update the parameters. The CNN model creates a more stable gradient when using Batch gradient descent, and it also converges faster for small data sets. However, as the training data set grows larger, convergent time increases, and the solution may converge in a locally optimal state. [25].

2.5.2 Stochastic Gradient Descent

Each training sample's parameters are changed independently [26]. It is more faster and memory economical when dealing with huge training data sets. However, because of the frequent updates, it takes highly noisy steps towards the solution, causing the convergence behavior to be quite unstable.

2.5.3 Mini Batch Gradient Descent

Separate the training examples into non-overlapping mini-batches and process them separately. It was more memory efficient, faster to compute, and had a more steady convergence. The main disadvantage of the Gradient-Based learning algorithm is that it easily stuck in a local minimum instead of a global minimum

2.5.4 Momentum

Improves both training speed and accuracy by adding the gradient calculated at the previous training step weighted by a parameter called the momentum factor.

$$\Delta w_{ij}^{t} = (\eta * \frac{dE}{dw_{ij}}) + (\lambda * \Delta w_{ij}^{t-1}) \tag{2.16}$$

WhereΔw_{ij}^{t} is current weight, Δw_{ij}^{t-1} previous weight, is the learning rate and is the momentum factor.

2.5.5 AdaGrad

Adaptive learning rate method updates each network parameter differently, based on their significance for the problem. perform larger updates for infrequent and smaller updates for frequent parameters.

2.5.6 AdaDelta

AdaDelta can be imagined as the extension of AdaGrad. The problem with AdaGrad is that, if we train the network with many large training epochs (t), then the sum of the square of all the past gradients becomes large, as a result, it almost vanishes the learning rate. AdaDelta method divides the learning rate of each parameter with the sum of the square of some past gradients (instead of using all the past gradients) for each parameter in each training epoch.

2.5.7 RMSProp

Root Mean Square Propagation (RMSProp) is also designed to solve the Adagrads radically diminishing learning rates problem

2.5.8 Adaptive Moment Estimation (Adam)

Adaptive Moment Estimation (Adam)[27] is another learning strategy, which calculates adaptive learning rate for each parameter in the network and it combines the advantages of both Momentum and RMSprop by maintaining both the exponential moving average of the gradients and the exponential moving average of the squared gradients. Adam is more memory efficient than others and also needs less computational power. In this thesis, we use Adam optimizer.

$$\Delta w_{ij}^{t} = \Delta w_{ij}^{t-1} - \frac{\eta}{\sqrt{\widehat{E[\delta^2]}^{t} + \epsilon}} * \widehat{E[\delta]}^{t} \tag{2.17}$$

Where $\widehat{E[\delta]}^{t}$ is the estimate of the first moment (the mean) and $\widehat{E[\delta^2]}^{t}$ is the estimate of the second moment (the uncentered variance) of the gradients.

2.6 Cooperative communications

The use of diversity technology highly improves the performance of wireless communications is improved using diversity gain technique. The transmission of the signals through multiple fading path exploit diversity in different channel dimensions, such as time, frequency, and space, and hence achieve diversity gains. The principle is similar to that of achieving spatial diversity gains in MIMO systems.
Cooperative communications allow nodes or terminals in a communication network to collaborate in the transmission of information, allowing for more effective use of communication resources. It's a technology that could be useful in future communication systems.
Structures of cooperative relaying technique has three types which are coordinated multi-point transmission (CoMP) which coordinate their transmissions in the downlink and jointly process the received signals in the uplink, fixed relay, and mobile relay. Fixed and mobile relay may have sigle relay or multiple relay models.

A relay system has three components: a source (S), a relay node (RN), and a destination (D). The RNs receive the data from the sources first. Each RN then applies a protocol to the data it receives and sends it to the destination nodes. The destinations then decode the data from their relevant sources using the received signal from the RNs.

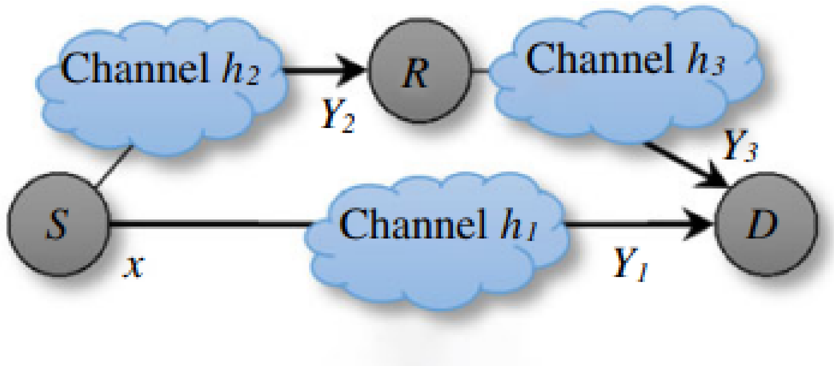


FIGURE 2.6: Block diagram of cooperative relay communication

Some of the basic cooperation protocols are:

2.6.1 Amplify-and-forward (AF)

Each RN basically scales the received signal to fit its transmit power constraints and sends the scaled signal to the next transmission slot. Laneman et al. analyzed a simple cooperative signaling method[28]. The fundamental disadvantage of the AF relaying protocol is that noise is amplified in the cooperative network, resulting in inter-symbol interference (ISI) between the source and destination channels[29]. Because the AF protocol includes less processing at the RN, it has a low computational complexity and thus a cheap cost when compared to other protocols[2]. Furthermore, the time it takes to send the information to D is short [30]. The AF is the finest solution for a quick communication application.

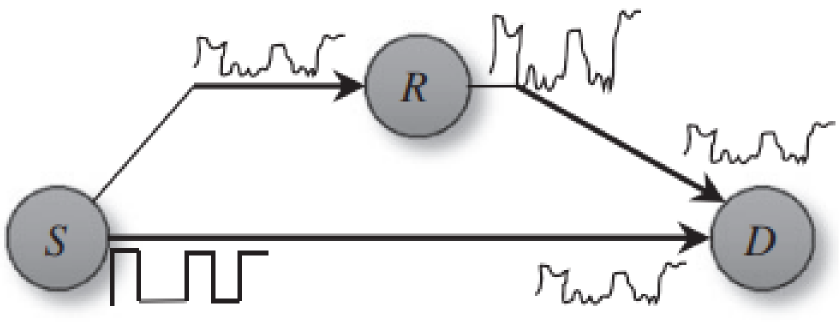


FIGURE 2.7: block diagram of an amplify-and-forward cooperation protocol

2.6.2 Decode-and-forward (DF)

Each RN decodes the source message from the signal it receives, re-encodes it into a new codeword, and broadcasts it in the next timeăslot. This helps to avoid noise amplification along the message signal, lowering the risk of ISI and reducing the likelihood of interference in the cooperative communication network. The fundamental drawback of this protocol is the processing delay, which required additional time from RN to demodulate, decode, modulate, and encode the incoming signal [31].

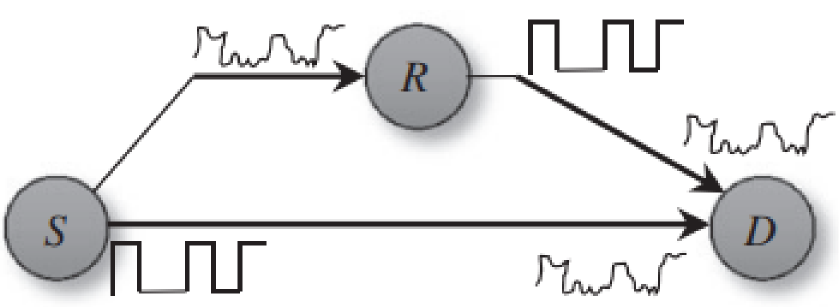


FIGURE 2.8: block diagram of Decode-and-forward cooperation protocol

2.6.3 Coded Cooperation (CC)

When repetition codes are used, the same codeword is sent twice either by the source or the relay, this will reduce bandwidth efficiency by half. CC basically incorporates cooperation into channel coding. Different chunks of the same message are conveyed in the two phases of coded collaboration schemes [32]. In particular, the source message is encoded in the first component of the codeword sent by the source, and incremental redundancy can be sent by the relay in the second portion of the codeword. Even though the time it takes to execute this operation causes a delay, the communication system's accuracy and dependability are much improved, and interference is decreased.

One protocol may outperform the other in terms of system capacity or diversity, depending on the network topology and the strength of the backhaul link between the source and the RN. In general, DF-based cooperation schemes are more advantageous for systems with decent backhaul links, whereas AF-based cooperation schemes are more advantageous for systems with relatively poor backhaul links[33].

2.6.4 Selective Detect-and-Forward

This protocol checks if the detected signal from source have an error due to channel noise then relay detects the source transmission; if the detection is error free it will be forwarded to the destination. To detect the source transmission correctly, it uses cyclic redundant check (CRC) error detection mechanism. This kind of protocol eliminates the problem of error propagation.

Chapter 3. Literature review

The performance of cooperative communication depends on the relay selection technique [34]. One of the efficient approaches is stablishing communication channel between RN and legitimate user by finding the channel condition using training bit [35]. Due to the variation of the channel the selected RN might not remain optimal. The performance of the system will degrade because of outdated CSI [36]. In the most cases, the obtained CSI through training was imperfect during the transmission time [37].

[38] proposed smart relay selection systems to overcome the performance degradation problem when spatial information channels are correlated in a MIMO environment. The eigenvalue properties-based relay selection method reduces the processing complexity of user equipment (UE) at the receiving end. However, especially in the case of frequency-flat fading channels, efficiency and performance are not achieved. [39].

Security is the most essential factor in general in wireless communication. Specifically in cooperative communication, the process of selecting the best relay during the transmission of data is highly susceptible to malicious attacks that generate a lot of security threads [40]. In some scenarios, the eavesdropping link node may have a good channel quality to acquire the information during the communication process [4].

To overcome this security threads scholars proposed different techniques. Some of them are uses a technique to enhance physical layer security by facing the effect of eavesdroppers.

In [3], the author designs a secure beamforming model (SBM) based on machine learning to modulate the signals on the relay. The input of SBM was signals received by the relay, the outputs of the network pass through the legitimate and the eavesdroppers channel respectively, and finally, arrive at the receiver or eavesdropper. After the signals are received, the signals will pass through the SBM network again and then the SBM network can output the plaintext. Through iterative learning, the SBM network can learn the statistical characteristics of the legitimate and the eavesdroppers channel, so that the legitimate user can decrypt the signals and the eavesdropper cannot decrypt the signals transmitted by the relay. However, the proposed network requires that the CSI of legitimate users and eavesdroppers are known. But in reality, the eavesdropper is always passive and the CSI is difficult to obtain.

In [41] a combination of jamming and beamforming is used to enhance overall information security. A unique interference node is created in the first time slot to transmit jamming signals, so that the eavesdropper receives both the information-bearing signal from the source node and the jamming signal from the interfering node in the first time slot. The SINR for eavesdropping has decreased. The relay broadcasts the relayed signal in the second time slot, along with fake noise projected over the null space of the legitimate channel to boost secrecy even more. However, sending the jamming signal through a specialized interfering node may cause interference with other relays or legitimate receivers.

In [42], secrecy enhancement for a three-timeslot two-way AF relaying scheme is investigated. Instead of using a dedicated jamming node, it is proposed that in the first two timeslots designated for legitimate node transmissions, the legitimate user that is not transmitting information-bearing signals acts as jamming to interfere with eavesdropping thereby reducing system complexity and delay. In the final time slot, when the relay amplifies and forwards the information received in the first two time slots, the bidirectional signals are processed separately using two beamforming matrices and optimizes each beamforming matrix based on knowledge of the CSI. The jamming signals are projected onto the null space of the legitimate channels. The sum secrecy rate of the legitimate users can be maximized through joint optimization.

In [6] an Optimal Relay Selection for Secure Cooperative Communications with an Adaptive Eavesdropper was proposed in which it derives closed-form secrecy outage probability expressions for the optimal relay selection schemes in the full and statistical eavesdroppers CSI cases and derives approximate secrecy outage probability expression for the optimal relay selection scheme in the partial eavesdroppers CSI case. In any case, the CSI of the eavesdropper and legitimate user is required for optimal relay selection and transmission secrecy, but in practice, the eavesdropper is always passive and the CSI is difficult to obtain; additionally, the CSI of legitimate users is outdated due to the time difference between channel estimation and packet transmission instant. The security risks can be reduced by improving the legitimate user's channel estimation performance.

[43] dealt with the problem of performance degradation of cooperative communication systems due to imperfect CSI. It uses a pilot symbol assisted modulation (PSAM)-based LMMSE scheme for the channel estimation and by deriving probability density function (PDF) and the moment generating function (MGF) of the

instantaneous SNR at the destination terminal, statistical quantities are applied to developed an accurate SER formula. But the performance of the algorithm decreases as the imperfection factor increases.

The deep learning assisted channel estimation is outperformed the conventional estimators for MIMO communications.

In [44] the authors argue for applying conventional neural network (CNN) to extract CSI pattern and present a CNN-RNN architecture for CSI aging. [45] build a decision-directed estimation with deep feed forward neural network-based channel prediction for MIMO transmission .

[46] Shows all proposed deep learning-based channel estimation models outperformed the conventional methods, even when channel imperfections were present. even if the performance of bi-LSTM model outperform in comparison to the FDNN and CNN models, it is more sensitive to Doppler frequency. The Doppler frequency has more serious consequences on the bi-LSTM model since it exploits the time-varying features of channels. But the complexity of bi-LSTM model is higher than both FDNN and DCNN is high.

In [47] a DCNN type detector is proposed as a MIMO communication with imperfect channel state information for the Internet of Things where the DCNN is trained offline and then used online to increase the BER of wireless systems by refining the imperfect CSI. Simulation results suggest that the DCNN outperform compared with the classical maximum likelihood detector (MLD). The network is not trained for cooperative relay communication.

The use of deep learning-based algorithms to overcome the imperfect CSI for MIMO communication systems is a promising approach. Motivated by that we extend the application of DCNN for a Cooperative relay network that considers the security issue of the system.

Chapter 4. Proposed system

4.1 System model

The system considers both AF and DF-based cooperative communication protocol which consists of the sender, receiver, and relay nodes. Each is equipped with N number of antennas and the number of antennas at a relay node is greater than the number of antennas of the sender, and the receiver. All nodes operate in half-duplex mode which means the relay node and sender or relay node and receiver are communicating one way only at a time to avoid inter symbol interference. Considering no direct link available, the sender and the receiver need to communicate with each other with the assistance of a relay. The channel fading matrix between the relay node and source, destination and relay node, and relay node1 and 2 is denoted by H_{SR1}, H_{DR1}, and H_{R1R2}, respectively. Two-time slots are used to transmit one data symbol. In the first time slot, the source terminal communicates with the relay. In the second time slot, the relay terminal communicates with the destination terminal. The channel is modeled as independent identically distributed Rayleigh fading and, fading characteristics are considered as time-correlated, fast and, flat fading. The correlation function is modeled as (2.3) so the channel fading matrix of adjacent samples will be calculated as

$$H(n) = \rho H(n-1) + \sqrt{1-\rho^2}N(n) \tag{4.1}$$

Where **n** is the sample time and **N** is the received noise the same size as **H**

One data symbol is transmitted over two time slots. The source terminal connects with the relay in the first time slot. The relay terminal connects with the destination terminal in the second time slot. During communication between the relay node one (**R1**) and the source or destination, relay node two (**R2**) will intercept the information due to the open nature of the cooperative network. To avoid this threat the **R1** will zero forcing the channel fading matrix between **R1** and **R2** (H_{R1R2}). Which is, first the CSI of H_{R1R2} will be estimated using pilot symbol then by applying eigenvalue decomposition it will produce null-space eigenvectors.

$$(A, V) = Eig(H_{R1R2}^*, H_{R1R2}), \tag{4.2}$$

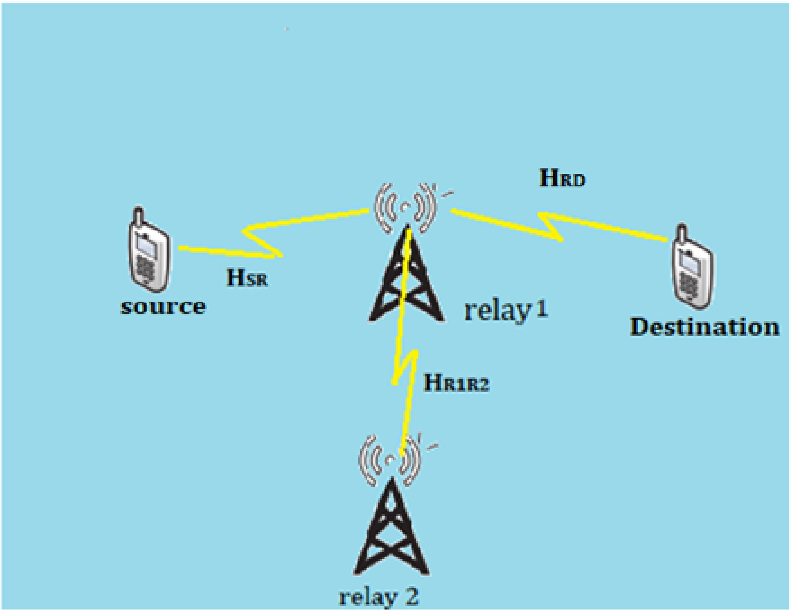


FIGURE 4.1: System Diagram of cooperative relay network with no direct link between source and destination

where $Eig()$ is the eigenvalue decomposition function, A ,V are the eigenvalues and eigenvectors respectively. Then **R1** will transmit the signal to the receiver using beamforming matrix which the information lies on the null-space of H_{R1R2}. Therefore, the received signal at R2 will be zero forced.

4.1.1 DF based cooperative protocol

For DF based protocol the signal from the source will be decoded and re encoded in relay node then transmitted to the destination. During the communication between the source and R1 in first time slot the pilot signal will be sent with the beamforming matrix to estimate the channel.Then the source will transmit the symbol with the knowledge of beamforming matrix. The received signal at R1 will be calculated as

$$y_{R1} = \sqrt{P_S} H_{SR1} B x + N_{SR1} \tag{4.3}$$

where P_S is the normalized transmission power of source terminal, B is beamforming matrix, $x \in CN(0, I)$ is the original message with size of B and,$N_{SR1} \in CN(0, \sigma^2 I)$ complex additive Gaussian white noise vector with zero mean and covariance of σ^2 between source and R1.

During the communication between R1 and the destination in the second time slot first the symbol will be detected using maximum likelihood detector as 4.7 then re-encoded and transmitted to the destination. The received signal at the destination terminal will be calculated as

$$y_D = \sqrt{P_{R1}} H_{DR1} B x + N_{DR1} \tag{4.4}$$

Where, P_{R1} is the transmission power of R1, B is beamforming matrix, $x \in CN(0, I)$ is the re-encoded original message with size of B and,$N_{DR1} \in CN(0, \sigma^2 I)$ complex additive Gaussian white noise vector with zero mean and covariance of σ^2 between R1 and destination.

The assumption of perfect CSI from the source to the relay and from the relay to the destination terminal will reduce system performance. In actuality, the CSI is never completely understood by the source to relay and relay to destination terminals. Imperfect CSI might occur as a result of a faulty channel estimating technique or as a result of channel fluctuations after it has been accurately measured. The flaw in our scenario is caused by a time delay between the estimation and the packet transmission instant. The imperfect equation from source to R1 and R1 to destination is modeled as 4.5 and 4.6 respectively.

$$\hat{H}_{SR1} = \sqrt{\zeta} H_{SR1} + \sqrt{1-\zeta} N_{SR1}, \tag{4.5}$$

$$\hat{H}_{DR1} = \sqrt{\zeta} H_{DR1} + \sqrt{1-\zeta} N_{DR1}, \tag{4.6}$$

Where, ζ is the correlation factor of the imperfect version of the channel fading matrix. Therefore, the detected original message at R1 from source and at destination from R1 using standard maximum likelihood detector is calculated as equation 4.7 and 4.8 respectively

$$\hat{x_{R1}} = \arg \min_{x \in X^{Nt}} \|y_{R1} - \hat{H}_{SR1} B x\|^2 \tag{4.7}$$

$$\hat{x_D} = \arg \min_{x \in X^{Nt}} \|y_D - \hat{H}_{DR1} B x\|^2 \tag{4.8}$$

where y_{R1} is the received signal at R1 terminal in first time slot and y_D is the received signal at R1 terminal in second time slot.

4.1.2 AF based cooperative protocol

For AF based protocol the signal will be sent in first time slot to the relay then the relay will first normalize the received signal to ensure the unity of average energy. Then, the normalized signal will be amplified and forwarded to the destination terminal during the second time slot. So the received signal at relay node and destination is given by 4.9 and 4.10 respectively

$$y_{R1} = \sqrt{P_S} H_{SR1} B x + N_{SR1} \tag{4.9}$$

$$y_D = \beta H_{SR1} H_{DR1} B x + N_{DR1} \tag{4.10}$$

Where, P_S and P_{R1} are the transmission power of source terminal and R1, B is beamforming matrix, $x \in CN(0, I)$ is the original message with size of B, $N_{DR1} \in CN(0, \sigma^2 I)$ complex additive Gaussian white noise vector with zero mean and covariance of σ^2 between R1 and destination and amplification factor β is given by equation .

$$\beta = \sqrt{\frac{P_{R1}}{\sigma_h^2 P_S + \sigma_n^2}} \tag{4.11}$$

where, σ_h^2 and σ_n^2 are covariance of channel fading matrix between source and R1 and, channel noise between R1 and destination.
Again the assumption of perfect CSI at destination terminal will degrade the system performance. The imperfect or outdated channel fading equation is modeled as

$$\hat{H}_{SR1} \hat{H}_{DR1} = \sqrt{\zeta} H_{SR1} H_{DR1} + \sqrt{1-\zeta} N_{DR1}, \tag{4.12}$$

Where, ζ is the correlation factor of the imperfect version of the channel fading matrix.

But the imperfect CSI will decrease the performance and security of the system. To solve this problem, we propose a deep convolutional neural network estimator. The deep learning networks can effectively capture the correlation features of the training data set.

4.2 Deep CNN architecture

Some well-known CNN models, such as VGG [48] and ResNet [49], enhance detection probability by increasing model depth. Specifically, the number of learnable parameters for VGG-19 is up to 144M, VGG-16 is 134.7, ResNet-18 11.4 M, ResNet-34 21.5 M, ResNet-50 23.9 M, ResNet-101 42.8 M. We must strike a balance between complexity and performance when developing the DCNN's architecture. Although a fully connected DNN performs better, the computing complexity of a fully connected DNN is proportional to the square of the number of nodes. In addition, the training period and data set are both very large. As a result, we adopt simplified classical DCNN models from[47] to reduce computational complexity as shown figure 4.2 and number of learnable parameters are summarized in table 4.1.

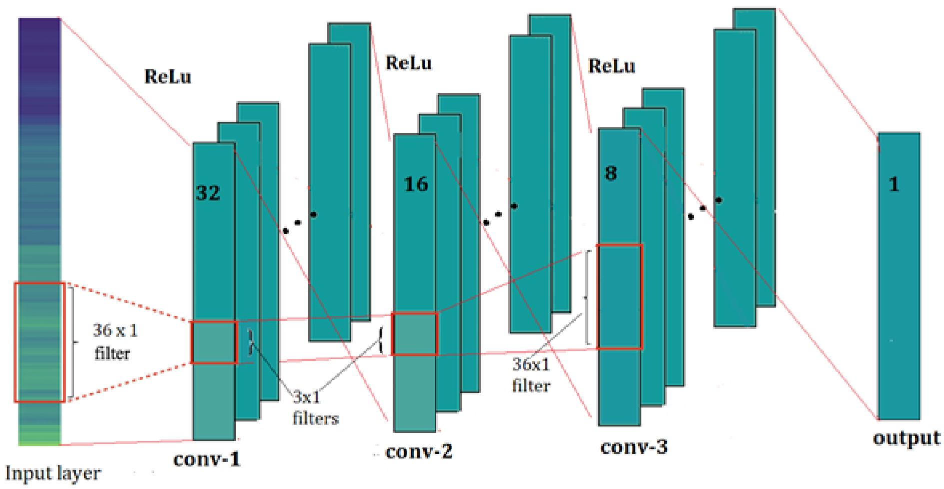


FIGURE 4.2: A four layer classical DCNN architecture

As shown in the figure 4.2 the DCNN model has 4 one dimensional layers excluding the input layer. The feature map extracted from input, Conv-1, Conv-2, and Conv-3 are 32,16, 8, and 1 and the filter length are 36, 3, 3, and 36 respectively each convolutional layer is followed by the ReLu activation function except the last one. The last one is followed by a SoftMax activation function for optimization purpose. The number of layers or depth of the network is chosen with the trade-off between complexity and performance in mind, but excessive depth reduces accuracy.

Also Setting the kernel size is always a tradeoff between speed and accuracy. The smeller Kernel size has better accuracy with lower execution speed and the larger kernel size has less accuracy with better execution speed but at some level there is an accuracy saturation, and computations grow up quadratically. A common choice is to keep the kernel size at lower. The first convolutional layer is often kept larger. Its size is less important as there is only one first layer, and it has fewer input channels.

In our model the first and last layer has lager kernel size which helps the network to learn general characteristics of the input data but the middle layers has lower kernel size which help the network to learn deeper level of the data set.
The input size of the network is a batch of $N_r \times N_t$. The batch size depends on the coherence time to transmit packet data. Similarly the hidden and output layers have the same size as input to keep information loss. To make sure the input and output have the same size we apply padding and the convolution process slides with single stride.

The number of learnable parameters L_p is calculated as

$$L_p = \sum_{i=1}^{L_n} Input * filtersize * numberoffilters + bias \tag{4.13}$$

TABLE 4.1: List of learnable parameters and computational complexity

Layers	Input * filter size * number of filters + bias	Number of learnable Parameters
Input	Packet size* packet per batch	-
Conv-1	1*36*1*32+32	1,184
Conv-2	32*3*1*16+16	1,552
Conv-3	16*3*1*8+8	392
Output	8*36*1*1+1	289
Total		**3,417**

4.3 Training process of the DCNN

The aim of the training is to predict the actual output (perfect CSI) from imperfect version of CSI. That means it will predict the perfect CSI for a given imperfect CSI. When we say imperfect CSI the previously estimated channel is changed due to different factors through time. Some of the factors are doppler shift which occurs due to movement of communication nodes, channel noise and etc. The DCNN network is expected to learn the changing pattern of affected CSI through time by adjusting the weight and bias of the network for given training data set.

Figure 4.3 shows the training process of the DCNN detector in which a batch of channel fading matrix $\hat{H}_{RD}$ with the size of $N_r \times N_t$ is reshaped as a column vector to form a one dimensional array. The batch size is equal to the coherence transmission time which is the time taken to transmit a group of packet data. Then the complex data of $\hat{H}_{RD}$ will be separated into real and imaginary to be treated as two real channel because the CNN model can only process real data [50]. The DCNN

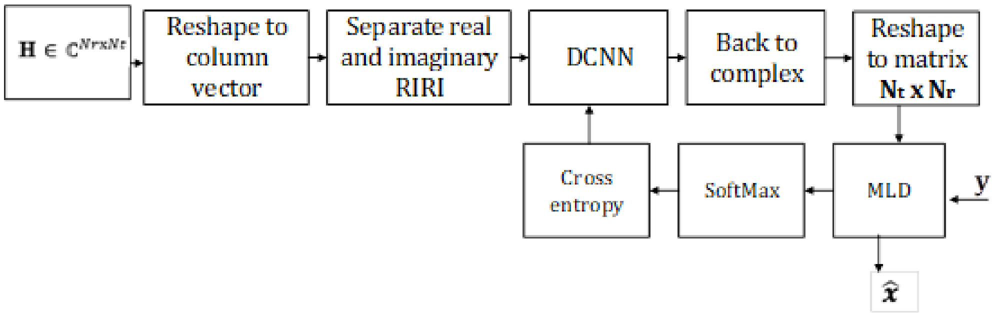


FIGURE 4.3: Block diagram of the DCNN training procedure

will convolve the channel fading with initialized filters weight for first time and then update the filter weight through training. The convolution will be held throughout the layer with respected filter size. Then after passed through the DCNN the data will be re arranged back to complex and reshaped to $N_r \times N_t$ matrix. The expected output using maximum likelihood detector will be calculated as 4.14

$$\hat{x}_i = exp\{-|y - \hat{H}_{RD}Bx_i|^2\}, x_i \in X^{Nt} \tag{4.14}$$

In the training process, the weights and biases of the DCNN will be updated by minimizing the loss function.So the normalized likelihood probability of each output class can be obtained by the SoftMax activation function as 4.15. From likelihood detection for all possible constallation set, if the probabilllaity $Pr_i = 1$ the message is correctly decoded otherwise the probability will be 0.

$$x_i = \frac{\hat{x}_i}{\sum_{i=1}^{X^{Nt}} \hat{x}_i} \tag{4.15}$$

Then using cross-entropy the loss will be calculated as 4.16

$$H(Pr_i, \hat{x}_i) = -\sum_{i=1}^{x^{Nt}} (Pr_i) log_2(\hat{x}_i) \tag{4.16}$$

After calculating the loss, it will be back-propagate to the DCNN network to update the weights and biases. To reduce error, the model's weights and biases are continuously updated during each training epoch, and the model iteratively searches for the locally optimal solution in each training epoch using the Adam optimizer, which is calculated as 2.17

4.4 Computational complexity

The computational complexity includes the number of multiplications/divisions and summations/subtractions. It is known as the number of floating point operations (f lops). The complexity of summations/subtractions is ignored because these operations are much easier to implement in hardware; instead, the concern is more about the number of real valued multiplications than the number of summations. Two complex multiplications involve four real valued multiplications and two summations. Based on this notion, the computational complexity for traditional ML detection of equation 4.8 with the search of N_tM, where M is constellation size or modulation order, are the multiplication of h_jx has $4N_r$ real valued multiplication and the Euclidean norm $||y - \hat{H}_{DR1}x||^2$ has $2N_r$ multiplications. Hence, the computational complexity of ML detector is

$$C_{ML} = O(6N_rN_tM) \tag{4.17}$$

with L_n kernels of size k_n in the n^{th} convolution layer and a depth of d, the number of multiplications for the n^{th} convolution layer is $k_na_nL_{n-1}L_n$, where a_n is sizes of the n^{th} layer. The complexity of all convolution layers is $O(\sum_{n=1}^{d}(k_na_nL_{n-1}L_n)$. The input layer size is $packetsize(ps) \times packetperbatch(p/b)$where $packetsize = transmissiontimeperpacketlength \times N_t \times N_r$ Therefore, the over all DCNN detector computational complexity is

$$C_{DCNN} = O(ps \times p/b + \sum_{n=1}^{d}(k_na_nL_{n-1}L_n) + 6N_rN_tM) \tag{4.18}$$

Chapter 5. Simulation result and discussion

Here, the simulation results are provided to show DCNN type estimator is efficient than the traditional maximum likelihood estimator. The provided simulations consider the system performance factors like correlation factor for the imperfection of channel, doppler effect, and the number of antennas to evaluate the BER of the received signal.
To generate the training data, we use QPSK modulation with a packet bit length of 900, 20 data packets are grouped to form one batch of data set, the complex channel gains are described by the autocorrelation functions using ***fd*** with first-order Bessel function and the variances of the complex channel gains and the noise are normalized to unity. A single batch has a size of $N_t \times N_r \times transmissiontimeperbatch \times 20$.
In learning process the data will divided into training, validation and test data set. For this simulation the training data set is set to 10,000 batch and the validation data set is set 1000 then to evaluate the BER performance of the system 1000 batch test data is used. In training process parameter initialization and optimizer selection is most important part so we select popular Xavier initializer and Adam optimizer by setting learning rate to 0.005 and $\gamma_1 = 0.9$, $\gamma_2 = 0.999$ and ϵ is set to 10^{-7} in case the estimate of the second moment is zero it avoids dividing by zero. The simulation was done in Anaconda navigator environment with Jupyter notebook, TensorFlow framework and, python program.

5.1 Effect of imperfect correlation factor

Figure 5.1 shows the DCNN and standard maximum likelihood (with perfect and imperfect CSI) detectors BER performance verses SNR simulation result for MIMO DF cooperative relay network with correlation factor of 0.95, normalized doppler frequency of 0.1, $N_{R1} = 4$, $N_{R2} = 2$, and $N_D = 4$. As we observe from the graph the BER performance of DCNN detector have a gain of about 0.5 dB at lowest SNR and 2 dB at highest SNR for DF protocol and for correlation factor of 0.95 in AF protocol, it has a gain difference of about 0.2 dB at lowest SNR and 1.5 at highest SNR in comparison to standard maximum likelihood detector. The reason is the DCNN type detectors learns the imperfect pattern of adjacent samples and produce more accurate CSI which increase the BER performance of the DF cooperative relay network.

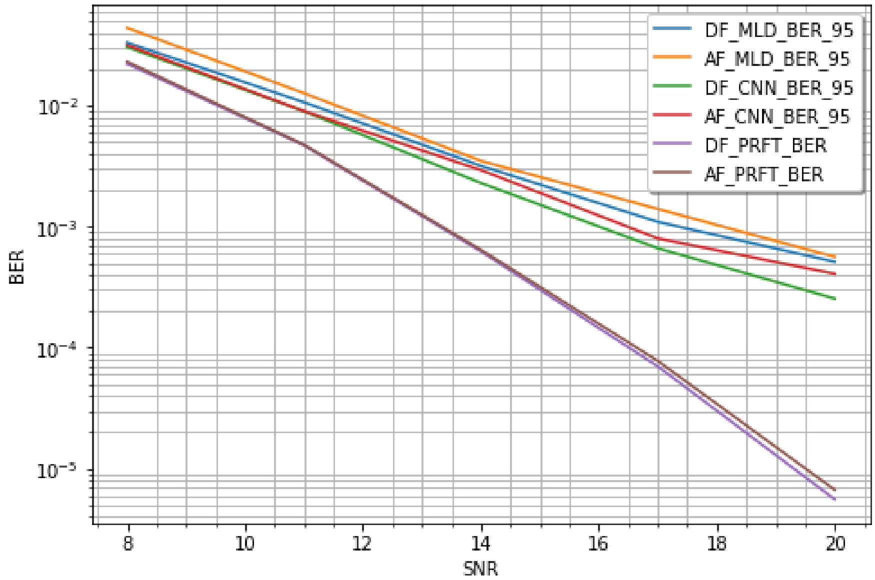


FIGURE 5.1: The BER performance comparison of the DCNN and standard maximum likelihood detector with correlation factor=0.95, $fd = 0.1$, $N_{R1} = 4$, $N_{R2} = 2$, $N_D = 4$

Figure 5.2, 5.3, 5.4, 5.5, and 5.6 shows the BER performance comparison of the DCNN and standard maximum likelihood detector with correlation factors of 0.9, 0.85, 0.8, 0.75, and 0.7 respectively. As the imperfection factor increases the performance difference of DCNN and standard ML detectors become increase. For instance, when the correlation factor is 0.9, 0.85, 0.8, 0.75, and 0.7 they have a gain difference of about 1dB, 1.5dB,2dB,2.75dB,and 5dB at lower SNR and about 4.5dB, 6.5dB, 8dB,9.5dB, and 10.5dB at higher SNR respectively for DF protocol and summarized on figure 5.7.

5.2 The effect of cooperation protocol

Figure 5.8 shows the BER performance comparison of the DF and AF protocol using DCNN detector with correlation factors of 0.9, normalized Doppler frequency of 0.1, NR1=4, NR2=2, and ND=4. As we observe from the graph the BER performance of the the DF protocol outperform than the AF protocol with a gain of about 0.5 dB at the lowest SNR and 2 dB at the highest SNR.

The DF protocol always outperform than the AF protocol. This is because of the presence of amplified noise in AF cooperative protocol during the transmission of the signal. Transmitting a symbol with presence of amplified noise has a negative

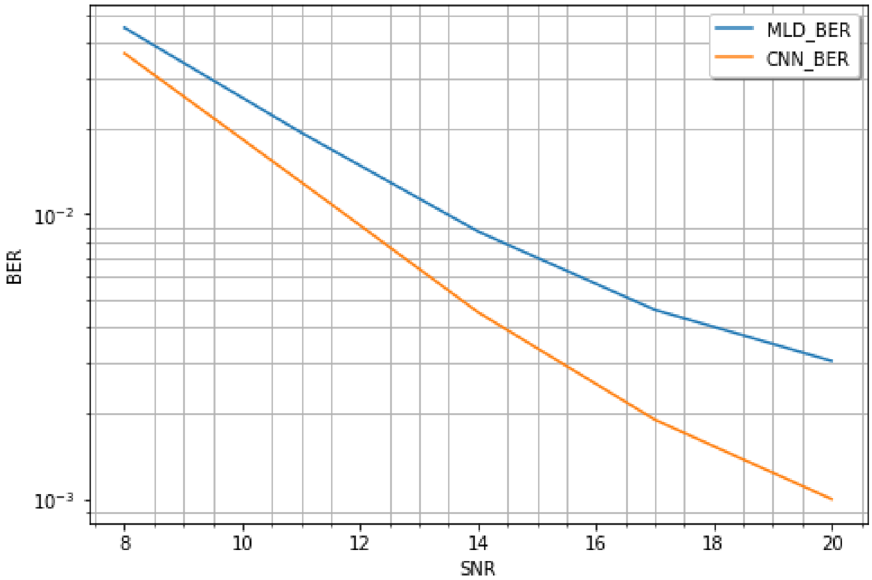


FIGURE 5.2: The BER performance comparison of the DCNN and standard maximum likelihood detector with correlation factor=0.90, $fd = 0.1$, $N_{R1} = 4$, $N_{R2} = 2$, $N_D = 4$

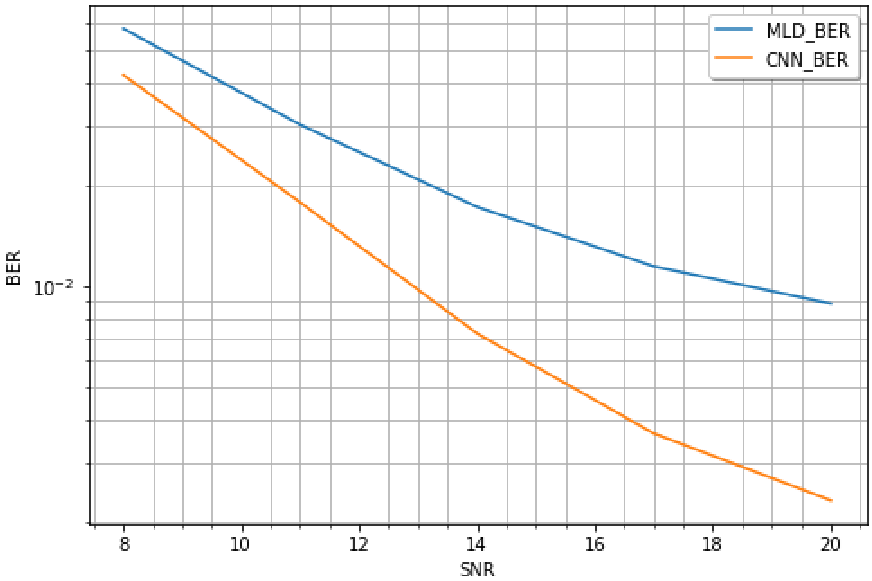


FIGURE 5.3: The BER performance comparison of the DCNN and standard maximum likelihood detector with correlation factor=0.85, $fd = 0.1$, $N_{R1} = 4$, $N_{R2} = 2$, $N_D = 4$

outcome on the quality of the signal received at the destination due to the inclusion of noise in the amplified signal.

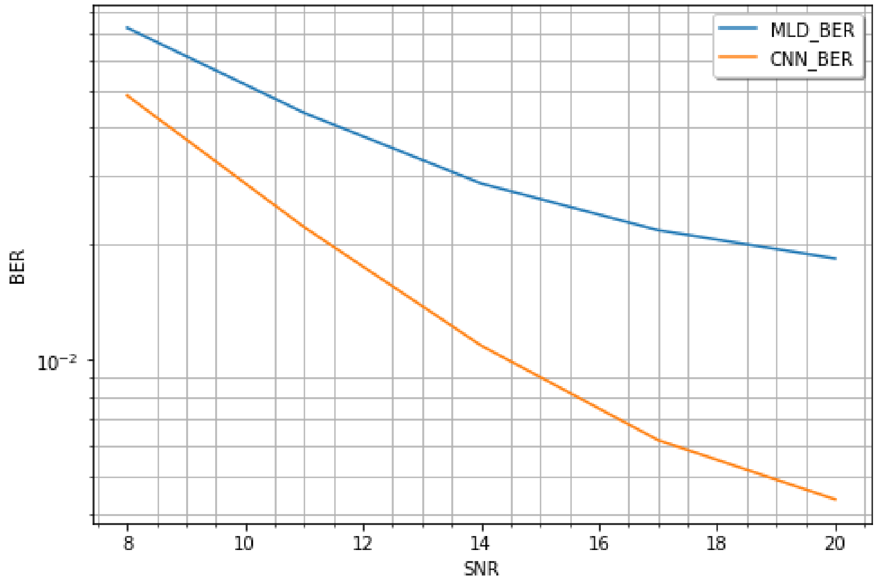


FIGURE 5.4: The BER performance comparison of the DCNN and standard maximum likelihood detector with correlation factor=0.8, $fd = 0.1$, $N_{R1} = 4$, $N_{R2} = 2$, $N_D = 4$

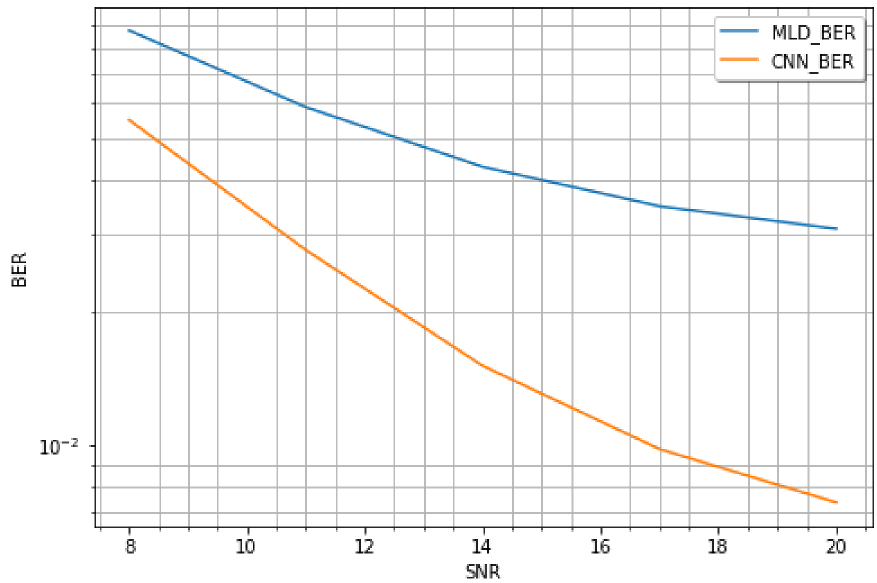


FIGURE 5.5: The BER performance comparison of the DCNN and standard maximum likelihood detector with correlation factor=0.75, $fd = 0.1$, $N_{R1} = 4$, $N_{R2} = 2$, $N_D = 4$

5.3 Effect of normalized doppler frequency

To show the effect of doppler frequency we can choose any value between $0 < f_d < 1$ but it is easy to analyze if we choose in such a way the first number is double of the second number. So, we chose the popular value in communication for normalized doppler frequency. Figure 5.9 shows the effect of normalized doppler frequency

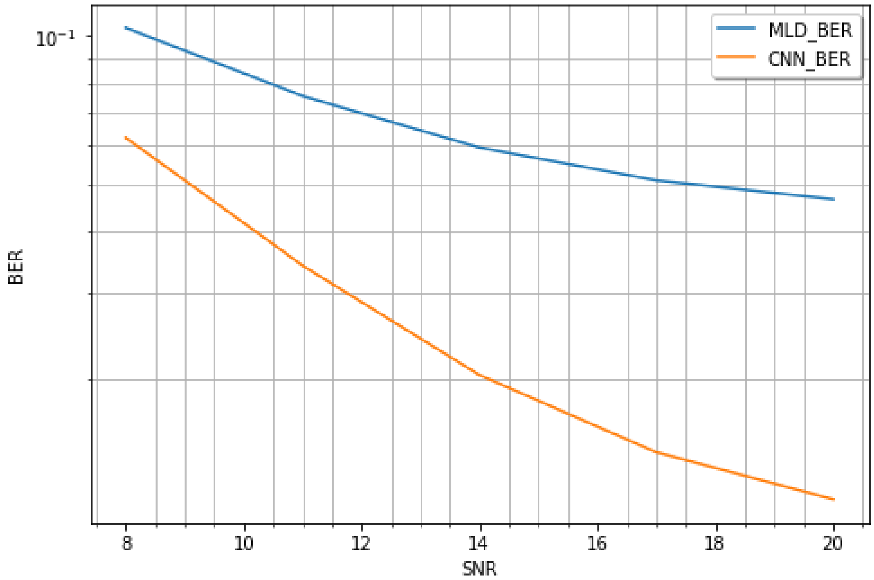


FIGURE 5.6: The BER performance comparison of the DCNN and standard maximum likelihood detector with correlation factor=0.7, $fd = 0.1$, $N_{R1} = 4$, $N_{R2} = 2$, $N_D = 4$

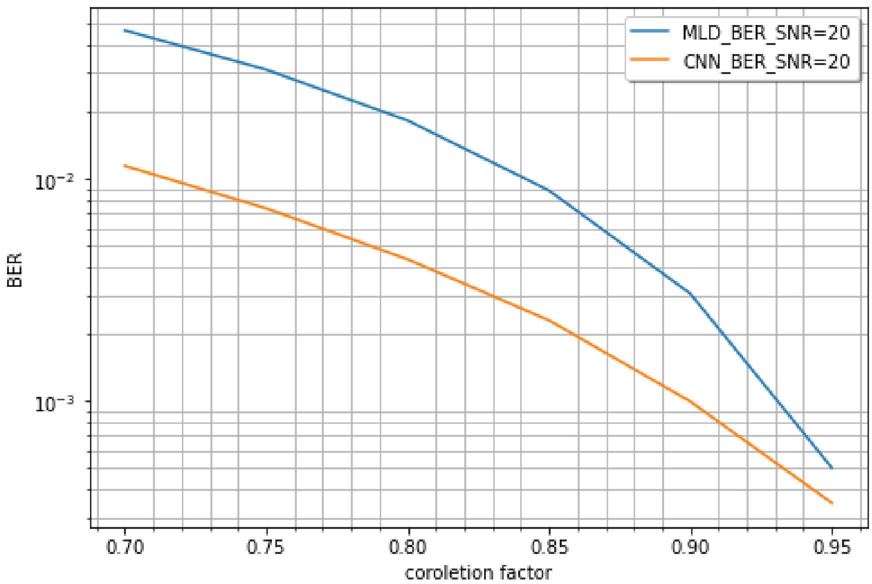


FIGURE 5.7: The effect of correlation factor on the BER performance DCNN and Standard maximum likelihood detectors using SNR=20, $fd = 0.1$, $N_{R1} = 4$, $N_{R2} = 2$, $N_D = 4$

with $fd = 0.1$ and $fd = 0.05$ on the BER performance of the DF cooperative communication system with a correlation factor of 0.9. As we see from the figure the DCNN detector BER performance is degraded to about 0.5dB at lower SNR and 3dB at higher SNR as the normalized Doppler frequency increases from 0.05 to 0.1. The reason is at the smallest normalized doppler frequency, the channel characteristics are changed slowly in which the symbol duration is significantly smaller than the

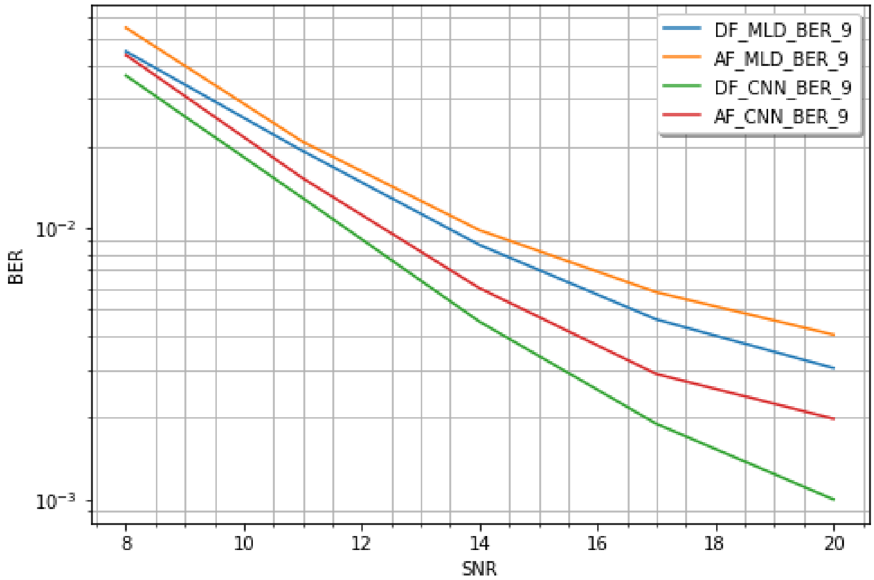


FIGURE 5.8: The effect of coopration protocol factor on the BER performance DCNN and Standard maximum likelihood detectors using SNR=20, $fd = 0.1$, $N_{R1} = 4$, $N_{R2} = 2$, $N_D = 4$

coherence time but at the highest normalized doppler frequency, the channel characteristics are changed for the transmission of single symbol duration. The doppler effect also affect the BER performance of maximum likelihood detector.

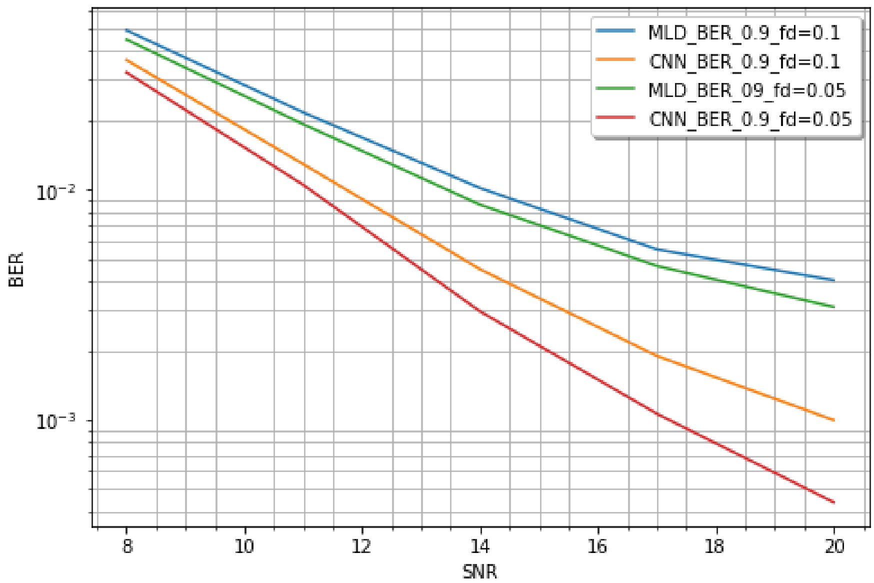


FIGURE 5.9: The effect of doppler frequency on the BER performance of the DCNN and Standard maximum likelihood detectors using, $fd = 0.1 and 0.05$, $N_{R1} = 4$, $N_{R2} = 2$, $N_D = 4$

5.4 Effect of number of antennas

Figure 5.10 shows the BER performance of the DCNN and standard ML detector with antenna configuration of N_{R1}, N_D and N_{R2} as 4, 4, and 2 for first and as 4, 2, and 1 for second respectively. As seen from figure the first configuration has better BER performance than the second. The reason for this is that as the number of antennas rises, the capacity increases linearly as a result of the multiplexing gain. Therefore for a fixed transmit power and bandwidth at high SNR, increasing the number of transmit and receive antennas results in an increase of the capacity and vise verse.

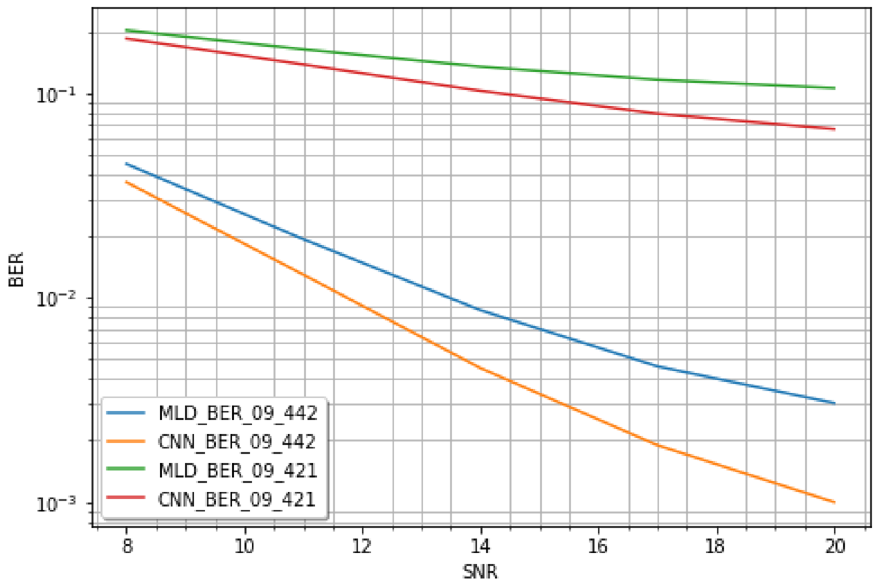


FIGURE 5.10: The effect of a number of antennas on the BER performance of the DCNN detector with different antenna configuration using, fd=0.1,corolation factor=0.9

5.5 Result comparison with related work

In [47] a deep learning detectors with and without accurate CSI are proposed to estimate and detect information bit considering the effect of imperfect CSI. The architecture of the DCNN model is the same as the model we use. The system setup is only considering base station and user but, in our setup, we consider eavesdropper relay node also we consider for both AF and DF cooperation protocol. By fine-tuning the optimizer parameter, the result we get has a gain difference of about 3dB.

Chapter 6. Conclusion and Recommendation

6.1 Conclusion

This paper studied the secure MIMO communication with imperfect CSI for cooperative relay network by zero forcing the equivalent channel fading matrix between legitimate RN and eavesdropping RN and then by estimating the equivalent channel fading matrix between RN and source and, RN and destination. The pattern of this imperfect version of the channel fading matrix is learned by the DCNN to produce accurate CSI then a maximum likelihood detector is used to extract the information. The Simulation was done for different performance factor parameters like imperfect correlation factor, doppler frequency, and the number of antennas to show the BER performance of the system. The results show that the DCNN detector has a higher gain performance than the standard maximum likelihood detector.

6.2 Recommendation

The following few promising future works are recommended by the author

- The security issue in cooperative communication can't only mitigated with help of accurete CSI of legitimate user it also affected by passive eavesdroppers this is also other interesting area to investigate.
- Even though the DCNN detector shows a great performance improvement than maximum likelihood detector, the bi-LSTM type deep learning can achieve more improvement by arranging the data set and training the network

References

[1] Omer Bulakci. “On backhauling of relay enhanced networks in LTE-advanced”. In: *arXiv preprint arXiv:1202.0212* (2012).

[2] Muhammad Asshad, Sajjad Ahmad Khan, Adnan Kavak, et al. “Cooperative communications using relay nodes for next-generation wireless networks with optimal selection techniques: A review”. In: *IEEJ Transactions on Electrical and Electronic Engineering* 14.5 (2019), pp. 658–669.

[3] Jintao Xing, Tiejun Lv, and Xuewei Zhang. “Cooperative relay based on machine learning for enhancing physical layer security”. In: *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE. 2019, pp. 1–6.

[4] Chensi Zhang, Jianhua Ge, Jing Li, et al. “Complexity-aware relay selection for 5G large-scale secure two-way relay systems”. In: *IEEE Transactions on Vehicular Technology* 66.6 (2016), pp. 5461–5465.

[5] Yulong Zou, Jia Zhu, Xianbin Wang, et al. “Improving physical-layer security in wireless communications using diversity techniques”. In: *IEEE Network* 29.1 (2015), pp. 42–48.

[6] Long Yang, Jian Chen, Hai Jiang, et al. “Optimal relay selection for secure cooperative communications with an adaptive eavesdropper”. In: *IEEE Transactions on Wireless Communications* 16.1 (2016), pp. 26–42.

[7] Nan Run Zhou, Xiao Rong Liang, Zhi Hong Zhou, et al. “Relay Selection Scheme for Amplify-and-Forward Cooperative Communication System with Artificial Noise”. In: *Sec. and Commun. Netw.* 9.11 (July 2016), pp. 13981404. ISSN: 1939-0114. DOI: 10.1002/sec.1425. URL: https://doi.org/10.1002/sec.1425.

[8] Linglong Dai, Ruicheng Jiao, Fumiyuki Adachi, et al. “Deep learning for wireless communications: An emerging interdisciplinary paradigm”. In: *IEEE Wireless Communications* 27.4 (2020), pp. 133–139.

[9] Marvin K Simon and Mohamed-Slim Alouini. *Digital communication over fading channels*. Vol. 95. John Wiley & Sons, 2005.

[10] Tommi Tuovinen, Nuutti Tervo, and Aarno Pärssinen. “Downlink output power requirements with an experimental-based indoor LOS/NLOS MIMO channel

models at 10 GHz to provide 10 Gbit/s". In: *2016 46th European Microwave Conference (EuMC)*. IEEE. 2016, pp. 505–508.

[11] LLOYDJ Mason. "Error probability evaluation for systems employing differential detection in a Rician fast fading environment and Gaussian noise". In: *IEEE transactions on communications* 35.1 (1987), pp. 39–46.

[12] William CY Lee. "Estimate of channel capacity in Rayleigh fading environment". In: *IEEE transactions on Vehicular Technology* 39.3 (1990), pp. 187–189.

[13] Tolga M Duman and Ali Ghrayeb. *Coding for MIMO communication systems*. John Wiley & Sons, 2008.

[14] Qian Chen, Shunqing Zhang, Shugong Xu, et al. "Efficient MIMO detection with imperfect channel knowledge-a deep learning approach". In: *2019 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE. 2019, pp. 1–6.

[15] N Sarmadi and S Shahbazpanahi. "andA. B. Greshman,Blind channel estimation in orthogonally coded MIMO-OFDM systems," in: *IEEE Trans. Signal Process* 57.6 (2009), pp. 2354–2364.

[16] Xu Zhu and Ross D Murch. "Performance analysis of maximum likelihood detection in a MIMO antenna system". In: *IEEE Transactions on Communications* 50.2 (2002), pp. 187–191.

[17] Kala Praveen Bagadi and Susmita Das. "MIMO-OFDM channel estimation using pilot carries". In: *International Journal of computer applications* 2.3 (2010), pp. 81–88.

[18] RS Ganesh, J Jayakumari, and IP Akhila. "Channel estimation analysis in MIMO-OFDM wireless systems". In: *2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies*. IEEE. 2011, pp. 399–403.

[19] Ahmadreza Hedayat and Aria Nosratinia. "Outage and diversity of linear receivers in flat-fading MIMO channels". In: *IEEE Transactions on Signal Processing* 55.12 (2007), pp. 5868–5873.

[20] Namshik Kim, Yusung Lee, and Hyuncheol Park. "Performance analysis of MIMO system with linear MMSE receiver". In: *IEEE Transactions on Wireless Communications* 7.11 (2008), pp. 4474–4478.

[21] Ping Li, Debashis Paul, Ravi Narasimhan, et al. "On the distribution of SINR for the MMSE MIMO receiver and performance analysis". In: *IEEE Transactions on Information Theory* 52.1 (2005), pp. 271–286.

[22] A Omri, R Bouallegue, R Hamila, et al. "Channel estimation for LTE uplink system by perceptron neural network". In: *International Journal of Wireless & Mobile Networks (IJWMN)* 2.3 (2010), pp. 155–165.

[23] Stuart Russell and Peter Norvig. "Artificial Intelligence: A Modern Approach, Global Edition 4th". In: *Foundations* 19 (2021), p. 23.

[24] Farhana Sultana, Abu Sufian, and Paramartha Dutta. "Advancements in image classification using convolutional neural network". In: *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*. IEEE. 2018, pp. 122–129.

[25] Sebastian Ruder. "An overview of gradient descent optimization algorithms". In: *arXiv preprint arXiv:1609.04747* (2016).

[26] Léon Bottou. "Large-scale machine learning with stochastic gradient descent". In: *Proceedings of COMPSTAT'2010*. Springer, 2010, pp. 177–186.

[27] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[28] J Nicholas Laneman, Gregory W Wornell, and David NC Tse. "An efficient protocol for realizing cooperative diversity in wireless networks". In: *Proceedings. 2001 IEEE International Symposium on Information Theory (IEEE Cat. No. 01CH37252)*. IEEE. 2001, p. 294.

[29] Krishna Srikanth Gomadam and Syed Ali Jafar. "Optimal relay functionality for SNR maximization in memoryless relay networks". In: *IEEE Journal on Selected Areas in Communications* 25.2 (2007), pp. 390–401.

[30] Zijian Mo, Weifeng Su, and John D Matyjas. "Amplify and forward relaying protocol design with optimum power and time allocation". In: *MILCOM 2016-2016 IEEE Military Communications Conference*. IEEE. 2016, pp. 412–417.

[31] Bin Zhao and Matthew C Valenti. "Distributed turbo coded diversity for relay channel". In: *Electronics letters* 39.10 (2003), pp. 786–787.

[32] Andrew Sendonaris, Elza Erkip, and Behnaam Aazhang. "User cooperation diversity. Part II. Implementation aspects and performance analysis". In: *IEEE Transactions on communications* 51.11 (2003), pp. 1939–1948.

[33] Qian Li, Rose Qingyang Hu, Yi Qian, et al. "Cooperative communications for wireless networks: techniques and applications in LTE-advanced systems". In: *IEEE Wireless Communications* 19.2 (2012), pp. 22–29.

[34] Ahmed S Ibrahim, Ahmed K Sadek, Weifeng Su, et al. "Cooperative communications with relay-selection: when to cooperate and whom to cooperate with?" In: *IEEE Transactions on wireless communications* 7.7 (2008), pp. 2814–2827.

[35] Shunya Iwata, Tomoaki Ohtsuki, and Pooi-Yuen Kam. "A lower bound on secrecy capacity for MIMO wiretap channel aided by a cooperative jammer with channel estimation error". In: *IEEE Access* 5 (2017), pp. 4636–4645.

[36] Jose Lopez Vicario, Albert Bel, Jose A Lopez-Salcedo, et al. "Opportunistic relay selection with outdated CSI: Outage probability and diversity analysis". In: *IEEE Transactions on Wireless Communications* 8.6 (2009), pp. 2872–2876.

[37] Ke Geng, Qiang Gao, Li Fei, et al. "Relay selection in cooperative communication systems over continuous time-varying fading channel". In: *Chinese Journal of Aeronautics* 30.1 (2017), pp. 391–398.

[38] Ivo Sousa, Maria Paula Queluz, and António Rodrigues. "A smart scheme for relay selection in cooperative wireless communication systems". In: *EURASIP Journal on Wireless Communications and Networking* 2013.1 (2013), pp. 1–13.

[39] J Nicholas Laneman, David NC Tse, and Gregory W Wornell. "Cooperative diversity in wireless networks: Efficient protocols and outage behavior". In: *IEEE Transactions on Information theory* 50.12 (2004), pp. 3062–3080.

[40] Emad S Hassan. *Security and data reliability in cooperative wireless networks*. CRC Press, 2018.

[41] Hui-Ming Wang, Miao Luo, Qinye Yin, et al. "Hybrid cooperative beamforming and jamming for physical-layer security of two-way relay networks". In: *IEEE Transactions on Information Forensics and Security* 8.12 (2013), pp. 2007–2020.

[42] Zonghao Ma, Yanhui Lu, Lingfeng Shen, et al. "Cooperative Jamming and relay beamforming design for physical layer secure two-way relaying". In: *2018 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*. IEEE. 2018, pp. 333–3336.

[43] Yi Wu and Matthias Patzold. "Performance analysis of cooperative communication systems with imperfect channel estimation". In: *2009 IEEE International Conference on Communications*. IEEE. 2009, pp. 1–6.

[44] Jide Yuan, Hien Quoc Ngo, and Michail Matthaiou. "Machine learning-based channel prediction in massive MIMO with channel aging". In: *IEEE Transactions on Wireless Communications* 19.5 (2020), pp. 2960–2973.

[45] Mehrtash Mehrabi, Mostafa Mohammadkarimi, Masoud Ardakani, et al. "Decision Directed Channel Estimation Based on Deep Neural Network k-Step Predictor for MIMO Communications in 5G". In: *IEEE Journal on Selected Areas in Communications* 37.11 (2019), pp. 2443–2456.

[46] Ha An Le, Trinh Van Chien, Tien Hoa Nguyen, et al. "Machine Learning-Based 5G-and-Beyond Channel Estimation for MIMO-OFDM Communication Systems". In: *Sensors* 21.14 (2021), p. 4861.

[47] Dan Deng, Xingwang Li, and Varun G Menon. "Learning based MIMO communications with imperfect channel state information for Internet of Things". In: *Multimedia Tools and Applications* (2021), pp. 1–12.

[48] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[49] Kaiming He, Xiangyu Zhang, Shaoqing Ren, et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[50] Timothy J OShea, Johnathan Corgan, and T Charles Clancy. "Convolutional radio modulation recognition networks". In: *International conference on engineering applications of neural networks*. Springer. 2016, pp. 213–226.