People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

University M'Hamed BOUGARA – Boumerdes

Institute of Electrical and Electronics Engineering

Department of Electronics

Final Year Project Report Presented in Partial Fulfilment of

The requirement for the degree of

Master

In **Electrical and Electronics Engineering**

Option: **Telecommunication Engineering**

Title:

Antenna Selection in Massive MIMO Using Machine Learning

Presented by:

Rahma CHERIGUI

Sarah BOUAZABIA

Supervisor:

Dr. Elhocine BOUTELLAA

Co-supervisor:

Dr. Fatima Zohra BOUCHIBANE

Dr. Hakim TAYAKOUT

Registration Number: 2024

To our beloved parents and siblings, for their endless encouragement and support . . .

And our friends, Nouha, Katia, Ibtihel, Rania and Houria

Acknowledgements

We are given, by these few lines, the opportunity to express our deepest appreciation and gratitude to the people who contributed to the successful completion of this project.

First and foremost, we extend our heartfelt thanks to our supervisor, Mr. Elhocine BOUTELAA and co-supervisor, Mrs. Fatima Zohra BOUCHIBANE, whose guidance, expertise, time, insightful feedback, encouragement and unwavering help have been indispensable and invaluable.

We would like to extend our thanks to the entire staff of CDTA, for their kindness and high level of professionalism in ensuring the health and safety of everyone involved.

We are indebted to all the teachers at the institute for their tireless dedication and genuine care for all of us students, which have significantly helped us empower our knowledge and consistently strive for excellence.

Finally, we are immensely grateful to our parents, whose love, patience and support have accompanied us throughout every step of this academic journey, helping us overcome every challenges. We cannot thank them enough for all what they do and sacrifice for us.

Thank you all for making this endeavour possible.

Abstract

In massive MIMO (Multiple Input Multiple Output) systems the overall performance (bit/s/Hz/cell) is significantly improved by equipping the base stations with arrays of a hundred antennas; which becomes one of its most significant challenges; economically and technically due to the high power consumption. To solve this, Antenna selection (AS) is increasingly gaining more interest, as it strategically reduces the hardware complexity while maximizing efficiency and throughput by selecting a specific subset of antennas to activate in each transmission slot. In this report, we examine the application of multi-label learning (MLL) based algorithms in AS, such as problem transformation methods, including first order binary relevance; and high order chain classification. Additionally, we investigate the Deep neural networks (DNN) based algorithms, namely Multi-Label Convolutional Neural Networks (MLCNN) and Multi-Layer Perceptron (MLP) classifier, and multi-View based algorithm. These proposed methods are rigorously evaluated based on their maximum capacity, performance and the computation time across various scenarios. Our work concludes that, in comparison with the convex relaxation based method, the Multi-view MLL achieves comparable results.

Contents

List of Figures

List of Tables

Acronyms

5G	*Fifth-generation wireless*
ABF	*Analog Beamforming*
ADC	*Analog Digital Converter*
AI	*Artificial Intelligence*
ANN	*Artificial Neural Network*
AWGN	*Additive white Gaussian noise*
Bps	*bits per second*
BS	*Base Station*
BW	*Bandwidth*
CC	*Classifier Chain*
CCA	*Canonical Correlation Analysis*
CNN	*Convolutional Neural Network*
CSI	*Channel State Information*
CSI-RS	*Channel State Information Reference Signal*
CV	*Calibrated Classifier*
DAC	*Digital Analog Converter*
dB	*Decibels*
DBF	*Digital Beamforming*
DCCA	*Deep Canonical Correlation Analysis*
DIS	*distributed scenario*
DL	*Deep Learning*
DNN	*Deep Neural Network*
ESA	*Exhaustive Search Algorithm*

FC	*Fully Connected*
FDD	*Frequency Division Duplex*
HBF	*Hybrid Beamforming*
KPI	*Key Performance Indicator*
KNN	*K-Nearest Neighbour*
LAN	*Local Area Network*
LOS	*Line Of Sight*
LTE	*Long Term Evolution*
LSTM	*Long Short-Term Memory*
M-MIMO	*Massive MIMO*
MIMO	*Multiple Input Multiple Output / Multiple-Input Multiple-Output*
MKL	*Multiple Kernel Learning*
MLCNN	*Multi-Label CNN*
MLP	*Multilayer Perceptron*
MV-MLL	*Multi-View Multi-Label Learning*
MVL	*Multi-View Learning*
MS	*Mobile Station*
MU-MIMO	*Multi-User MIMO / Multiuser MIMO*
OFDM	*Orthogonal frequency-division multiplexing*
PSO	*Particle Swarm Optimization*
RAM	*Random-Access Memory*
RBF	*Radial Basis Function*
RELU	*Rectifier Linear Unit*
RF	*Radio Frequency*
SE	*Spectral Efficiency*

SGD	*Stochastic Gradient Descent*
SNR	*Signal-to-Noise Ratio*
SU	*Single User*
SU-MIMO	*Single-User MIMO*
TDD	*Time Division Duplex*
TTI	*Transmission Time Interval*
UE	*User Equipment*
ULA	*Uniform Linear Array*
URA	*Uniform Rectangular Array*
WI-FI	*Wireless Fidelity*
mmW	*millimeter Wave*

Introduction

The digital age and new wireless communication are undergoing a very substantial increase in the mobile data traffic due to a number of emerging applications, which include machine-to-machine communications and video streaming…etc in response, the telecommunications industry is continuously innovating and evolving, to meet the escalating demands of the digital age with technologies.

Among the proposed new technologies capable of meeting these requirements is Massive MIMO (M-MIMO) also called Large-Scale Antenna Systems; a widely discussed and pivotal technology in the realm of wireless communication that stands as a beacon of innovation, offering immense potential to revolutionize wireless networks. M-MIMO employs a huge number of antennas at the base station while using a simple linear processing. It holds the promise of serving multiple users simultaneously in the same time-frequency resource with improved quality, capacity and without severe inter-user interference. However, amidst its advantages, challenges such as hardware complexity, cost and power consumption due to the number of Radio Frequency (RF) chains causes a significant issue that we aim to address within this work.

Numerous techniques have been proposed to improve energy efficiency and capacity, one of which is Antenna Selection (AS). This method entails selecting specific subsets of antenna elements from the whole array within the base station (BS) system to attain optimal capacity while economizing energy.

Our approach involves harnessing the power of a machine learning trained models for multi-label classification to select an optimal subset of antennas at the BS which offers the perfect balance between capacity enhancement and power efficiency.

The remainder of this report is outlined as follows:

- Chapter 1: elucidates the fundamental concepts of MIMO (Multiple Input Multiple Output) systems, examines the benefits and drawbacks of Massive MIMO technology, in addition to the antenna selection in Massive MIMO systems, and explores its various techniques and related work

- Chapter 2: examines machine and deep learning systems and further explores multi-label classification in the context of antenna selection for Massive MIMO systems.
- Chapter 3: presents the obtained results and engage in a comprehensive discussion of their implications and significance.

The report is finally concluded by summarizing the main findings and providing perspectives for future research endeavours.

Chapter 1 : Massive MIMO and Antenna Selection

1.1 Introduction

This chapter presents the groundwork by introducing the technology of Multiple-Input Multiple-Output (MIMO) wireless communication and its fundamental concepts. Subsequently, it presents Massive MIMO, elucidating its architecture, benefits, and drawbacks, where lays a critical issue concerning the energy consumption due to the deployment of many antenna elements in Massive MIMO systems. Additionally, Antenna Selection (AS) topic will be explored in this chapter, elaborating on several algorithms and techniques, aiming for the maximization of capacity considering both conventional MIMO and Massive MIMO.

1.2 MIMO systems

MIMO technology has gained significant attention for the past decades, representing a pivotal role in advancing wireless communication like Wi-Fi, LTE, and 5G, for it has Remarkably improved network capacity, reliability and spectral efficiency by employing multiple antennas at both transmitter and receiver ends, which enables the transmission of multiple data streams. Moreover, it mitigates interference and exploits spatial diversity making it particularly vital for applications such as wireless LANs and cellular telephony, where a single base station must communicate with numerous users concurrently.

A MIMO system refers to a system where multiple antennas are deployed at one or both ends of the communication link as shown Figure 1-1.

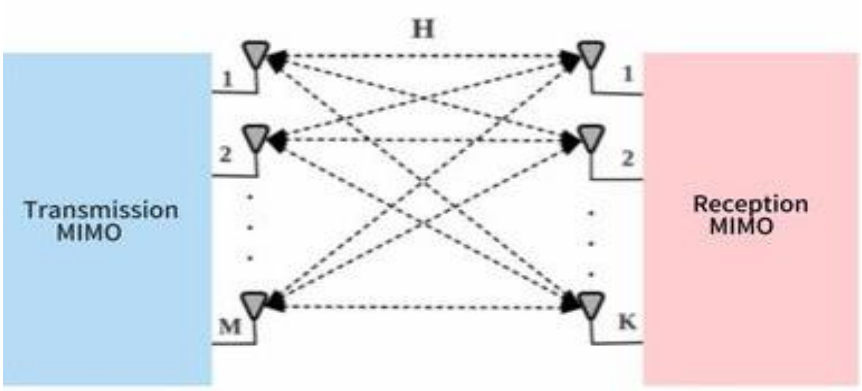


Figure 1-1: MIMO System.

Considering having a system with M antennas on the Base Station (BS) and K receiver antennas. The received signal can be given as [1]

$$Y(t) = H.x(t) + n(t) \tag{1.1}$$

where y(t) represents the K × 1 received vector sampled at time t, and x(t) represents the M × 1 vector transmitted by the antennas, n(t) is the K × 1 Additive white Gaussian noise (AWGN) vector at the receiver with zero mean and variance of σ^2, **H** is the K × M channel matrix, whose ij-th element is the scalar channel between the i-th receive and j-th transmit antenna.

$$H = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1M} \\ h_{21} & h_{22} & \cdots & h_{2M} \\ \vdots & \ddots & \ddots & \vdots \\ h_{K1} & h_{K2} & \cdots & h_{KM} \end{bmatrix} \tag{1.2}$$

The general formula of capacity of a MIMO system is given by [2]:

$$C = log_2[\ \det(I_K + (\rho / M).\ HH^{\dagger})]\ \ \text{bps} \tag{1.3}$$

where I_K is the K× K identity matrix, ρ is the mean Signal to Noise Ratio (SNR), and the superscript † denotes the Hermitian transpose.

1.3 Basic Concepts of MIMO Technology

MIMO technology can be divided into three categories namely: point-to-point MIMO (P2P MIMO), multiuser MIMO (MU-MIMO), and massive MIMO (M-MIMO).

1.3.1 Point-to Point MIMO

Point-to-point MIMO (P2P MIMO) represents the simplest MIMO system, where the BS is equipped with one array that transmits to a terminal with a receiving antenna array as shown in Figure 1-2. the users are accommodated in disjoint time/frequency blocks via a combination of time division and frequency division multiplexing [3]. P2P MIMO is commonly used in wireless communication scenarios where high data rates, reliability, and coverage are essential, such as backhaul links between base stations, wireless broadband access, and long-range communication links.

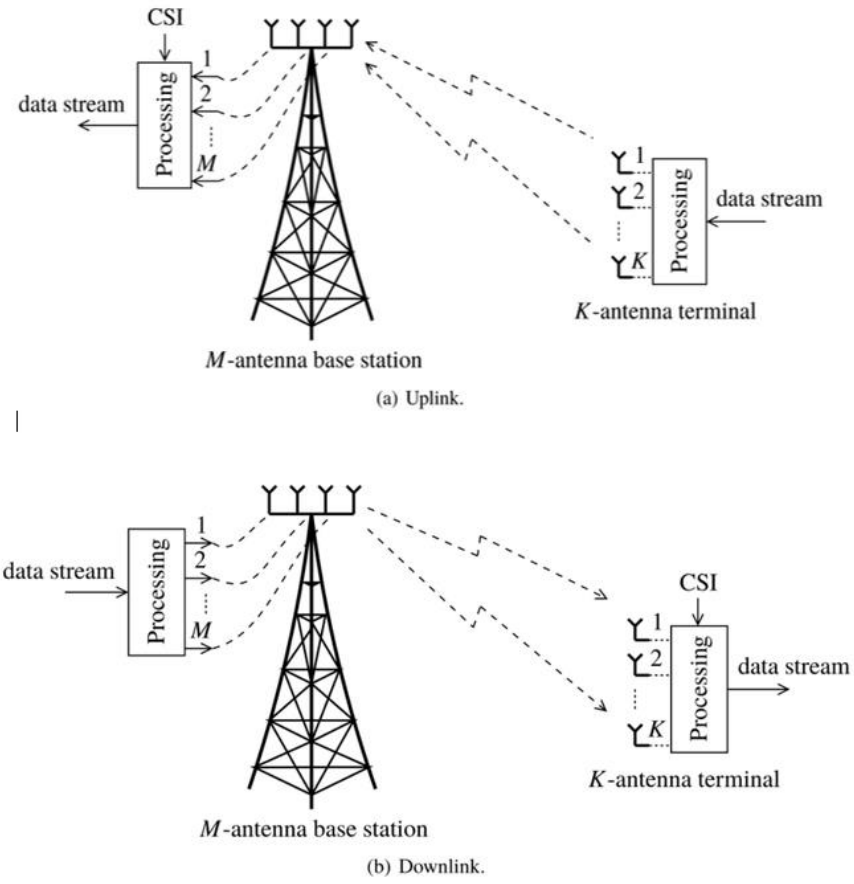


Figure 1-2 Downlink and Uplink Point-to-Point MIMO [4].

The point-to-point MIMO system capacity in both the uplink and downlink system, is expressed in bits/s/Hz the same as (1.3).

1.3.2 Multi-User MIMO

Multi-User MIMO (MU-MIMO) serves multiple users simultaneously by exploiting spatial multiplexing and interference management techniques. Multiple users communicate simultaneously with the BS using the same time-frequency resources where BS handles most of the processing, in both uplink and downlink scenarios.

A MU-MIMO system can be derived from the point-to-point MIMO setup by dividing the K-antenna terminal model into several independent and less complex terminals separated by many wavelengths as shown in Figure 1-3 [3]. The main difference between them is represented in the table 1-1.

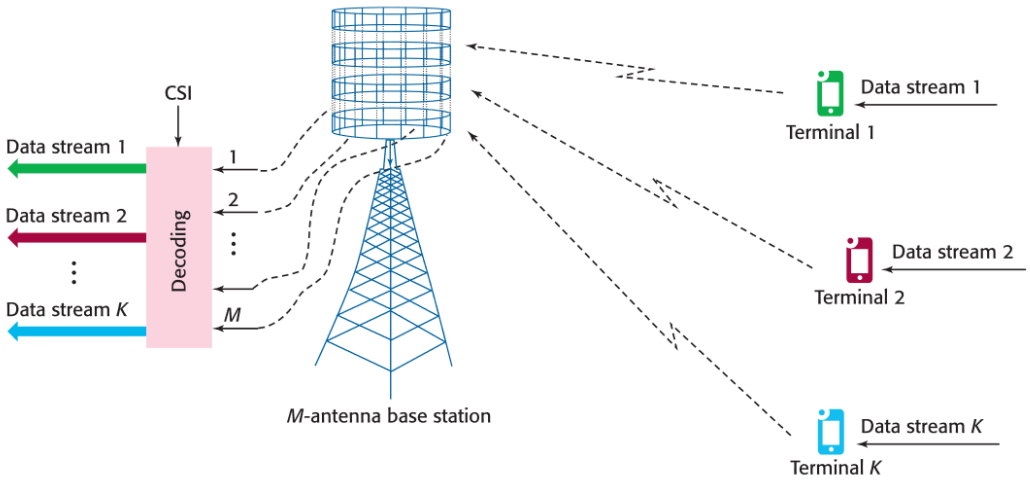


Figure 1-3: Uplink operation of a Multi-user link [3].

the Shannon sum-capacity for uplink Multi-User MIMO is identical to that of uplink Point-to-Point MIMO [3]:

$$C_{Up} = log_2[\ \det(I_K + (\rho / M).\ H_u H_u^{\dagger}\ )]\quad \text{bps} \tag{1.4}$$

However, the formula for downlink Shannon sum-capacity requires the solution of a convex optimization problem,

$$C_{down} = sup_a \left\{ log_2\ det\left((I_M + (\ \rho\ D_a H_d H_d^{\dagger}) \right) \right\} \text{bps}\ ;\ \ a >= 0, 1^T a = 1 \tag{1.5}$$

where D_a is a diagonal matrix, whose diagonal elements comprise the M × 1 vector, a, and 1 represents the M × 1 vector of ones. Crucially, this capacity is predicated on both ends of the link knowing the downlink channels.

Features	SU-MIMO	MU-MIMO
Service	One user at a time	Multiple users simultaneously
Complexity	Relatively simpler implementation	Requires sophisticated signal processing and coordination
Spectral efficiency	Limited by the channel conditions of a single user	Improved spectral efficiency
Spatial streams	Multiple spatial streams dedicated to a single user	Independent data streams to different users simultaneously

Table 1-1: Comparison between SU-MIMO and MU-MIMO.

1.4 Massive MIMO

The demand for communication reliability, wireless throughput, and user density is on a perpetual rise; thus, future wireless communication requires developed technologies capable of serving numerous users simultaneously with exceptionally high throughput. One such technology that addresses these challenges is Massive MIMO.

The massive MIMO concept was first mentioned in the seminal paper [5] by Thomas Marzetta, defined as a MU-MIMO system with M antennas and K users per BS. The system is characterized by the vast provisioning of antenna elements at base stations that serves multiple users' antennas simultaneously, operating in time-division duplexing (TDD) mode using linear Uplink (UL) and Downlink (DL) processing.

In a massive MIMO system, the number of transmit antennas can be as large as hundreds. such criteria introduce many benefits, such as capacity, multiplexing, diversity, and energy efficiency. However, it faces challenges like: power consumption and hardware complexity [6].

1.4.1 Massive MIMO Characteristics

The theoretical mathematical aspect of massive MIMO is based on the law of large numbers and the theorem of Lindeberg-Levy. Let h_i and h_j be two mutually independent 1 × M channel vectors whose elements are zero-mean random variables with σ_i^2, σ_j^2 their corresponding

variances. For a Single user (SU) case, channel hardening can be mathematically illustrated from the capacity equation as [7]:

$$\frac{1}{M}(h_i h_i^H) \xrightarrow{a.s.} \sigma_i^2 \ \ as \ \ M \to \infty \tag{1.6}$$

And favourable propagation for two users:

$$\frac{1}{M}(h_i h_i^H) \xrightarrow{a.s.} 0 \ \ as \ \ M \to \infty \tag{1.7}$$

where $\xrightarrow{a.s.}$ denotes almost sure convergence. Eq. 1.7 shows that the two vectors become orthogonal as the number M increases.

1.4.1.1 Channel hardening

Channel hardening involves minimizing the effects of channel fading and interference for a secure transmission. It might involve techniques such as beamforming, diversity schemes, and error correction coding. Given a user channel vector, the average received gain at user k when M antennas are transmitting converges to a deterministic value denoted β_k, the large-scale coefficient. Mathematically, it can be written for user k as [7]:

$$\frac{\|\boldsymbol{h_k}\|^2}{\boldsymbol{M}} \xrightarrow{a.s.} \beta_k \ \ as \ \ M \to \infty \tag{1.8}$$

1.4.1.2 Favourable propagation

It represents the circumstances that are conducive to the successful transmission of signals. Where the ultimate favourite propagation condition occurs for two users, channel vectors become pairwise orthogonal, when $\mathbf{h_i h_j^H = 0}$. Ultimately, the signals can be separated. However, this condition is very hard to satisfy.

1.4.1.3 Time Division Duplex (TDD)

In Time Division Duplex (TDD) mode, the same frequency band is used with different time slots for Uplink (UL) and Downlink (DL): In UL transmission, the channel is estimated based on the received K orthogonal pilot sequences signals sent by k users. Whereas, in the DL transmission, to make sure each user recovers its own data, the Base Station (BS) needs Channel State Information (CSI) to precode the transmitted signals.

In Frequency Division Duplex (FDD) mode, UL and DL use two different frequency bands to estimate the corresponding channel: In UL transmission, the process is same as in TDD. Whereas, in DL transmission, BS antenna elements send orthogonal pilot signals to K users, which then estimate the channel from the received signals relay this information back to the base station via a control channel, the achievable rates in bits/second [4]:

In TDD: $$D_{TDD} = \frac{B}{2}\, log_2\left(1 + \frac{P_R}{BN_0}\right) \quad (1.9)$$

In FDD: $$D_{FDD} = \frac{B}{2}\, log_2\left(1 + \frac{P_R}{(B/2)N_0}\right) \quad (1.10)$$

Where B represent the Bandwidth (BW). P_R denotes the power received by the user terminal, and N_0 is the noise spectral density for the downlink scenario.

TDD protocol is used in Massive MIMO, because the throughput required for channel estimation does not depend on the number of BS antennas M Unlike FDD protocol.

1.4.2 Massive MIMO system

We consider a Massive MIMO system (see Figure 1-4) with M transmit antennas and K receive antennas such that $\mathbf{M \gg K}$, where the input-output relationship is given by:

$$y(t) = H.x(t) + n(t) \quad (1.11)$$

$\mathbf{y} \in \mathbb{C}^{\mathbf{Kx1}}$ is the signal received by the M antennas, $\mathbf{x} \in \mathbb{C}^{\mathbf{M\ x\ 1}}$ is the transmitted signal vector, $\mathbf{H} \in \mathbb{C}^{\mathbf{K\ x\ M}}$ is the channel matrix composed of several independent and identically distributed wireless channels between the transmitter and the with zero mean and unit variance and $\mathbf{n} \in \mathbb{C}^{\mathbf{Kx1}}$ is the AWGN noise vector at the receiver with zero mean and variance of $\boldsymbol{\sigma^2_n}$ [8].

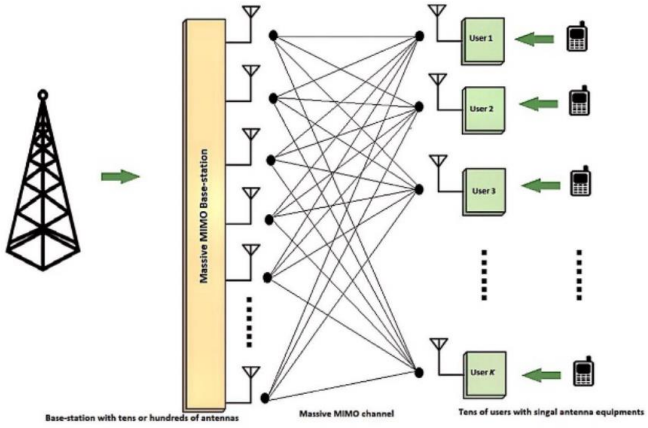


Figure 1-4: Massive MIMO architecture [9].

1.4.3 Why Massive MIMO?

Massive MIMO systems offer a host of unparalleled benefits for wireless communication (see Figure 1-5). By harnessing a multitude of antennas, these systems enhance capacity and throughput, maximize data rates while optimizing spectral utilization and ensure robust connectivity across diverse environments, allowing simultaneous servicing of multiple users with advanced signal processing capabilities. Moreover, massive MIMO utilizes sophisticated beamforming techniques systems to extend coverage and range through, diminishing interference and ensuring robust connectivity and optimal user experience. The major benefits of massive MIMO are:

- **Improved Spectral Efficiency**: By utilizing spatial multiplexing techniques in cellular networks, Massive MIMO systems achieve higher spectral efficiency, enabling more data to be transmitted over the available bandwidth (BW) [9].

- **High Throughput**: System throughput is a key parameter for performance evaluation, defined as the sum of data rates delivered to all users in a given cell and measured in bits per second (bits/s or bps), it is directly related to the BW and spectral efficiency (SE) as shown below:

$$\text{Throughput} = \text{Bandwidth (Hz)} \times \text{Spectral efficiency (bits/s/Hz)}$$

Massive MIMO can increase the system capacity 10 times or more and simultaneously improve the radiated energy efficiency [10], due to the aggressive spatial multiplexing that is

transmitting multiple data streams simultaneously over the same frequency channel, utilizing multiple antennas at both ends.

- **Scalability**: Massive MIMO is a Scalable Technology that accommodates a large number of users and devices, making them suitable for deployment in various scenarios. It departs from Shannon theoretic practice. First, the BS learns the channels via uplink training, under TDD operation. Second, the number of BS antennas can be made as large as desired with no increase in the channel estimation overhead since In a TDD system the time required to acquire CSI is independent of the number of base station antennas. Third, the signal processing at each user is very simple and does not depend on other users' existence. Therefore, as the number of the base station antennas increases, linear precoding and decoding performance can approach the Shannon limit [3].
- **Linear processing:** in a massive MIMO system, the large number of base station antennas over the number of users yields a favourable propagation when the channel matrix between the BS antenna array and the users is nearly orthogonal. Hence, a simple signal processing (linear combing schemes in the uplink and linear precoding schemes in the downlink) is preferable to remove the effect of interference, fading and noise [10]**.**

Furthermore, massive MIMO reduces the constraints on accuracy and linearity of each individual amplifier and Radio Frequency (RF) chain by relying on the law of large numbers and beamforming to make sure that noise, and hardware imperfections average out.

Figure 1-5: 5G requirements and benefits of M-MIMO [11].

1.4.4 Disadvantages of massive MIMO

Massive MIMO represents the key technology to this demand as it grants high spectrum efficiency and gain by deploying a large number of antennas at the BS; however, it faces the following issues:

- **Power consumption and cost:** Due to enormous antenna elements, a big amount of power is needed since the BS requires RF and Analog-to-Digital Converter (ADC) chains, which are expensive, and despite the hybrid combination to reduce RF chains that may decrease the hardware overhead cost, the price may rise as the propagation paths channel is larger than the mobile station (MS) antennas [9].
- **Pilot contamination:** Typically, the maximum number of orthogonal pilot sequences in a 1 ms coherence interval is estimated to be about 200 (the available supply of orthogonal pilot sequences is easily exhausted) [4]. Hence, the reusage of pilots from one cell yields the consequence referred to as pilot contamination, illustrated in Figure 1-6). This phenomenon affects the channel estimation accuracy and cannot be eliminated by increasing the number of antenna elements at the BS.

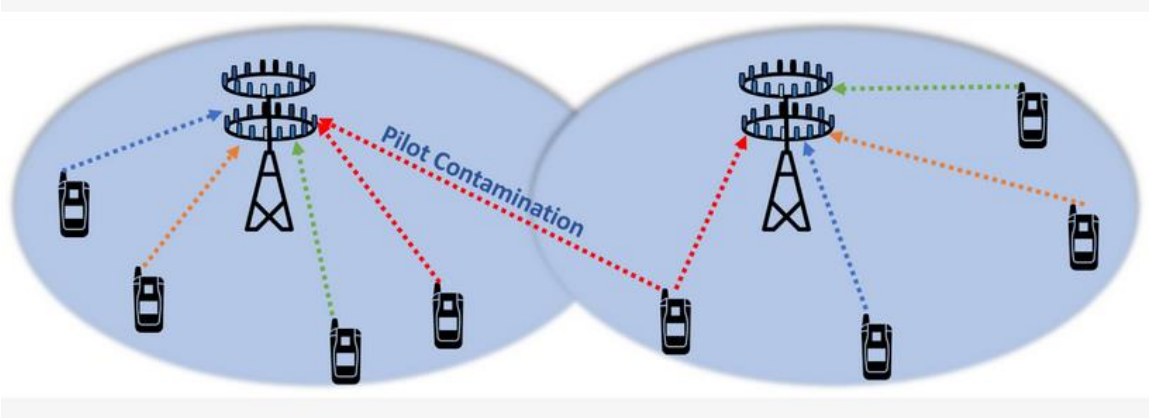


Figure 1-6: Illustration of pilot contamination in M-MIMO [12].

Moreover, Determining CSI between each transmit and receive antenna uses a considerable number of spectral resources [10].

1.4.5 Massive MIMO architecture

Beamforming is a signal processing technique used in sensor arrays for directional signal transmission or reception. It is applied to various technologies, including massive MIMO setups. The Three popular architectures of beamforming used in massive MIMO setups are illustrated in Figures 1-(7 and 8) as a, b, and c for analog, digital, and hybrid beamforming respectively, each with its unique applications and advantages.

1.4.5.1 Digital Beamforming

Digital Beamforming (DBF) is referred to as MU-MIMO in LTE/5G. It is already used in transmission modes 7, 8 and 9 in LTE Advanced Pro [8]. In this architecture, each radiating element in the antenna array is connected to its RF chain and signal pre-Processing circuit, including phase and amplitude adjustments for beamforming, is performed at the base station before transmission. Despite the usage of RF chains for each element, this setup allows for the simultaneous formation of multiple beams from the same set of antenna elements, providing full flexibility and adaptability to the system.

1.4.5.2 Analog beamforming

Unlike digital beamforming, which performs signal processing in the baseband domain, analog beamforming (ABF) is a technique that manipulates the signals directly in the RF domain, same signal is fed to each physical antenna element and the signal phases are adjusted in the RF domain using analog phase-shifters after digital-to-analog conversion (DAC) for the single stream user [7]. ABF grants advantages such as simplicity and lower power consumption compared to digital beamforming. However, it cannot dynamically adjust beamforming parameters in response to changing channel conditions.

Despite ABF being the best compromise between coverage and power/cost constraints, it is not adequate for massive MIMO scenarios with a large number of receivers. Nevertheless, it can be used where link reliability and high data rates are a must and with mmW for potential wireless backhaul design applications [13].

1.4.5.3 Hybrid beamforming

For a balance between the simplicity and power/cost constraints of analog beamforming and the flexibility and adaptability of digital beamforming, hybrid beamforming (HBF) was first introduced and analysed in the mid-2000s [10]. In this setup, both ABF and DBF are combined: the array of antenna elements of BS is divided into sub-arrays, each connected to a number of RF chains, where ABF is applied to adjust the phase and amplitude of signals before they are combined across the sub-arrays. After signals have been combined from the different sub-arrays, DBF is performed at the baseband level adjusting the signals to further refine the beamforming process.

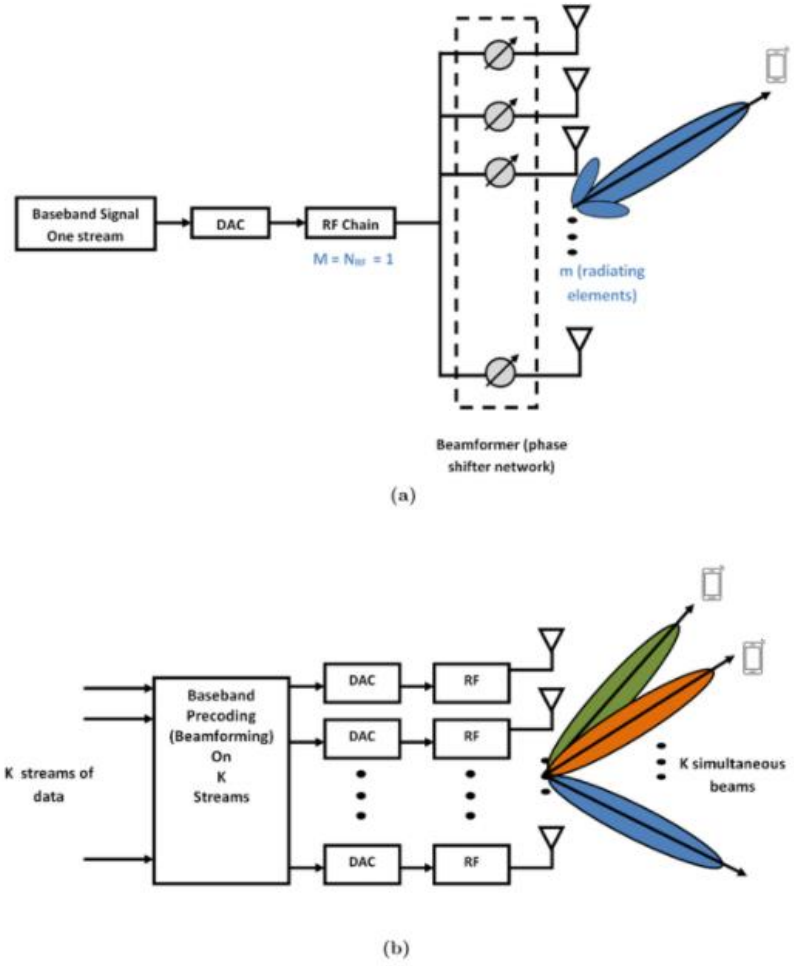


Figure 1-7: (a) Tx Analog Beamformer and (b) Tx Full Digital Beamformer [7].

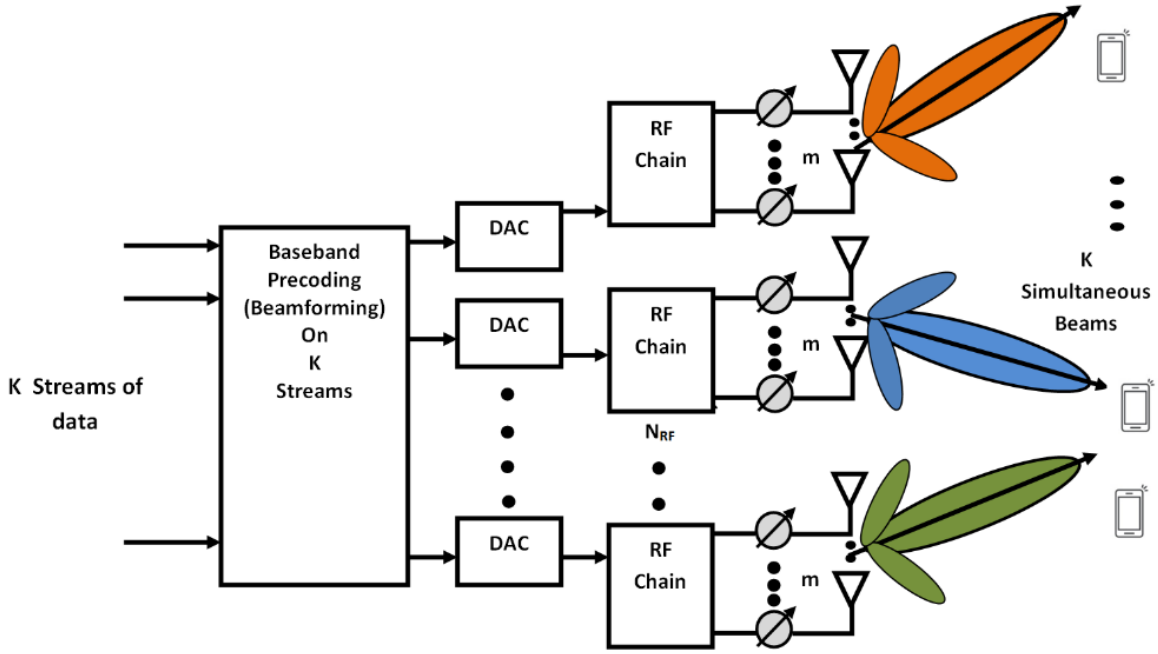


Figure 1-8: Example of a Hybrid Beamforming Architecture [7].

1.5 Antenna Selection

To transmit a signal from an antenna element at the BS, the element must be attached to an RF chain, that incorporates all analog components past the transmitting antennas, including power amplifiers, mixers, phase shifters, and ADCs/DACs (Analog-to-Digital Converters/Digital-to-Analog Converters). Hence, the total number of RF chains is equal to the number of transmitting antenna elements M at the BS, which represent the main power consumption and hardware cost of a cellular network. Antenna selection emerges as one of the most promising techniques. It involves identifying the most optimal subset of antennas for activation for further processing, while efficiently discarding the remainders.

1.6 Antenna selection in M-MIMO

As shown in Figure 1-9, there are M transmit antennas at the base station with a number of L_T RF chains ($L_T < M$), and K receive antennas at the receiver side with L_R RF chain ($L_R < K$). We denote the overall M×K channel matrix by H, and the $L_T \times L_R$ channel matrix representing the selected antennas by H.

The object of AS is to select the best L_R out of K antennas at the receive side and the best L_T out of M antennas at the transmission side so that the resulting system capacity is maximized. Following this selection, the transmitter experiences a significant reduction in complexity [14].

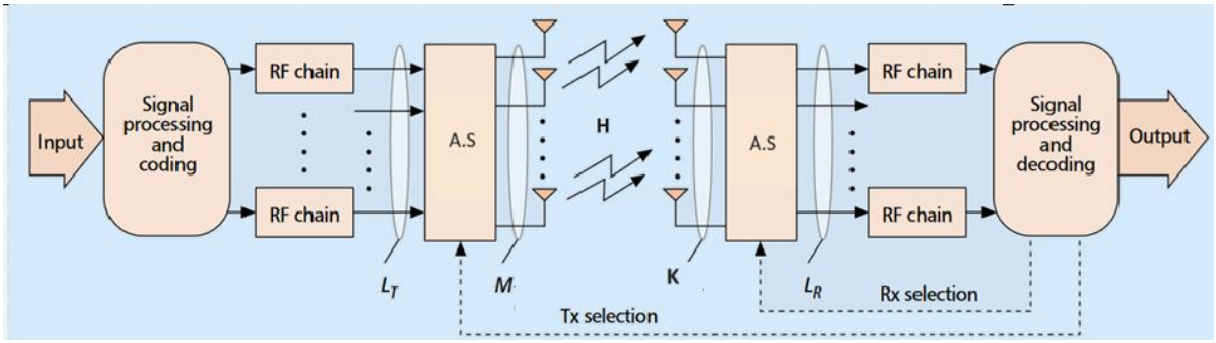


Figure 1-9: Antenna selection system in M-MIMO [14].

1.7 Antenna selection approaches

Over the span of decades, several antenna selection criteria and algorithms have been introduced for conventional small-scale MIMO systems. For instance, many research delved into error-rate-oriented selection criteria coupled with specific selection algorithms tailored for practical receivers. And many others focused on capacity-oriented selection criteria, including greedy search methods, convex optimization techniques, and dominant-submatrix searches [15]. However, many derivations in Antenna selection in MIMO (AS-MIMO) systems cannot be directly applied in massive AS-MIMO systems as deploying a large number of antennas will result in high computational complexity. To address this issue, some of these methods have been Adjusted and expanded for massive MIMO systems.

1.7.1 Exhaustive Search Method

Exhaustive search method can be used to achieve the optimal antenna allocation. In particular, this technique checks all possible subsets of antennas, and selects the subset that satisfies the best performance metric. However, a significant drawback of this approach derives from its high computational complexity. note that the number of potential subsets is $\sum_{m=1}^{L_R} C_m^{N_T}$ which entails very high complexity if N_T and/or L_R are large as in M-MIMO [16].

Example:

For simplicity, we assume all channels experience the same SNR. Assuming a base station with N_T =4 and the selected antennas is L_T=2 antennas for transmission. The possible combinations are: (1,2), (1,3), (1,4), (2,3), (2,4), (3,4).

For each combination, we calculate the capacity using the capacity equation (1.3) from the first chapter: the capacity for combination (1,2) might be 10 bps/Hz, for combination (1,3) is 12 bps/Hz, and so on. Upon evaluating all combinations, we find that combination (1,3) yields the highest capacity of 12 bps/Hz. Consequently, the algorithm opts for the (1,3) subset to activate, resulting in a capacity of 12 bps/Hz.

1.7.2 Related work

A bio-inspired algorithm was employed in [17], where the author tackled the challenge of maximizing capacity in wireless MIMO systems using Particle Swarm Optimization (PSO)[1],where the authors defined the “particle” as a set of antennas and the fitness function as the capacity achieved by the antenna subset. The work results showed better balance between capacity performance and computational complexity of PSO when compared with optimal Exhaustive Search Algorithm (ESA) method, and more effectiveness as the number of iterations increases leading to optimized MIMO system performance.

[1] PSO mimics the behaviour of organisms, like bird flocks, to seek optimal solutions through iterative updates of particle positions and velocities, influenced by local and global best solutions.

Heuristic Optimization algorithms can also be deployed to solve AS problem as discussed in [18]. However, a significant number of iterative calculations is required due to the large-scale antenna array, making them computationally impractical in real-world applications.

Gorokhov [19] proposed an algorithm for fast antenna subset selection, that iteratively removes the antenna with the lowest contribution to system capacity continuously until the desired number of antennas is reached, though the algorithm is complex. In contrast, a faster near-optimal antenna selection algorithm was introduced in [20], that begins with an empty set unlike the previous algorithm, and adds the antenna with the highest capacity contribution in each step. This approach [20] achieves nearly the same capacity as both the optimal method and Gorokhov's algorithm, but with significantly lower computational complexity.

While these traditional methods and heuristic approaches provide a foundation, Deep Learning (DL), a new Artificial Intelligence (AI) technique that has shown its powerful use in research domains such as image processing and speech processing, overcomes their limitations and enhances the performance and efficiency of antenna selection in massive MIMO systems. DL methods excel in addressing complex tasks such as multi-label classification, which is a challenging task in machine learning as it requires predicting multiple label categories for each input instance, which is practical yet difficult.

It is noteworthy to mention that multi-label classification is different from the standard multi-class recognition, as it predicts more than one label per instance (see Figure 1-10). And thus, different approaches have been adapted to solve multi label classification problem as it became more required by modern applications.

One of the most well-known strategies is Binary Relevance (BR), this method treats each label as an independent binary classification problem, enabling separate handling of each task [21]. The author in [22] introduced AS in single-user and multi-user MIMO systems by approaching it as a series of binary classification tasks, in which he employed two conventional machine learning algorithms; support vector machine (SVM) and k-nearest neighbours (KNN); to detect the selected antennas. However, these methods required a large number of binary classifiers, resulting in heightened computational complexity.

Furthermore, there have been efforts to utilize deep learning techniques in the context of a AS in massive MIMO systems. In [23], A pair of twin convolutional neural networks (CNNs) that

are trained using a substantial amount of labelled data was introduced. These networks are designed to sequentially carry out receive antenna selection and hybrid beamformer design in a single-user (massive) MIMO system, which is distinct from the configuration of our own system. Convolution layers are commonly employed in image processing tasks to capture the relationship between neighbouring pixels. CNN-based antenna selection algorithm was also introduced in [24]. However, the application of convolution in the context of wireless channels lacks a clear understanding of its physical implications [25].

Figure 1-10: The three types of classification.

1.8 Conclusion

In this chapter, we explored massive MIMO systems, along with its advantages, disadvantages and characteristics. We discussed antenna selection and its various methods that have been employed and studied in previous works on MIMO systems. These techniques include traditional approaches as well as machine and deep learning-based techniques. From our review, we conclude that deep learning-based methods demonstrate superior performance, lower computation complexity and are particularly well-suited for massive MIMO systems, where the number of transmitting antennas BS is significantly increased.

Chapter 2 : Multi-Label classification based AS

2.1 Introduction

This chapter provides a comprehensive analysis of different implementations of Multi-Label Learning (MLL) based algorithms in Antenna Selection (AS). We first review the concept of Multi-View Learning (MVL) in the scope of deep learning, and its methods. Then discuss the multi-View MLL model; combining both Deep Canonical Correlation Analysis (DCCA) and Auto-Encoder (AE) in a single network structure [25]; first order binary relevance, high order chain classifier, MLCNN, and finally MLP classifier.

2.2 System model

We consider transmit AS in the downlink of a single-cell multi-user large-scale MIMO-OFDM (Orthogonal frequency-division multiplexing) system with L subcarriers. The BS is equipped with N = 64 transmitting antennas and Ns RF chains that simultaneously serve k = 4 single-antenna mobile users. Considering an ideal scenario with perfect CSI for all antennas at the BS, denoted by $H \in \mathbb{C}^{KxN}$ Channel normalization is performed such that the elements of H have unit energy, averaged over L = 100 subcarriers. Both the transmit power and the number of selected antennas is assumed to be fixed. The objective of this study is to transmit signals over Ns antennas that maximize the key performance indicator (KPI), which is the channel capacity given by:

$$C = log_2[det(I_K + (\rho / N_s).H_s H_s^{\dagger})] \quad \text{bps} \qquad (2.1)$$

where I_K is the K× K identity matrix, ρ is the mean SNR, and the superscript † denotes the Hermitian transpose. from the complete MIMO matrix H, Ns columns are selected by using a diagonal matrix Δ of size N × K with binary diagonal elements. Each binary element indicates whether the i-th antenna is selected, ensuring that the sum of the diagonal elements is equals to Ns, $\sum_{i=1}^{N} \Delta i$ =**Ns.**

$$\Delta = \begin{cases} 1 & ,selected \\ 0 & ,otherwise \end{cases} \quad (2.2)$$

Equation (3.1) can be rewritten as:

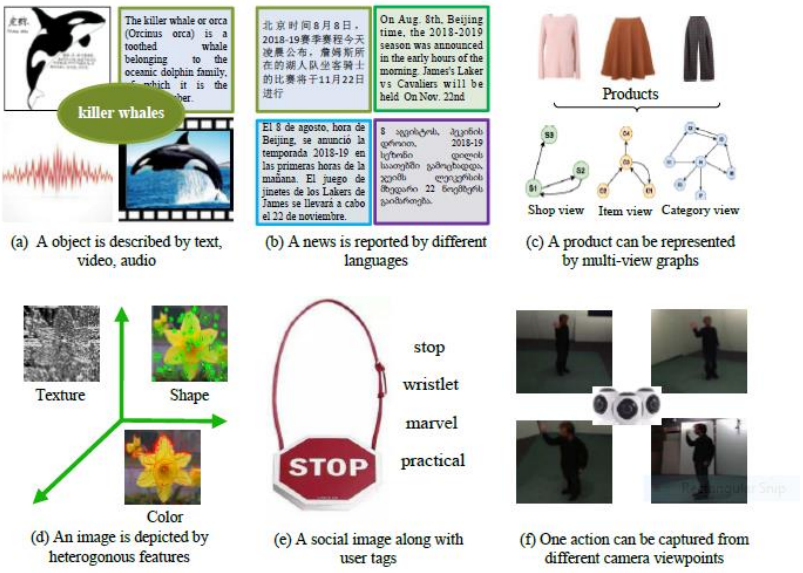
$$C = log_2[\det(I_K + \left(\rho / N_s\right).\ H\Delta H^{\dagger})] \text{ bps} \quad (2.3)$$

Thus, the optimal Δ can be found by:

$$\Delta_{opt} = arg\ max\Delta(log_2[\det(I_K + \left(\rho / N_s\right).\ H\Delta H^{\dagger})]) \text{ bps} \quad (2.4)$$

2.3 Multi-view overview

The objective MVL is to uncover the common feature spaces or shared patterns by merging different sets of features or data sources [26]. Over the past few decades, MVL has experienced a notable surge in popularity within the realms of machine learning and computer vision, as evidenced by various studies [27, 28, 29, 30, 31]. This surge has led to the development of several promising algorithms, including the co-training mechanism [32], subspace learning methods [33], and multiple kernel learning (MKL) [34]. Among the various MVL approaches, one of the most widely adopted strategies involves mapping multi-view data (example in Figure 2-1) into a shared feature space that maximizes the consensus among multiple perspectives [35, 36, 37, 38, 39]; known as canonical correlation analysis (CCA) [27], which is a statistical technique utilized to explore the linear relationships between two sets of feature vectors. Moreover, many variations of CCA have been developed to learn about a common feature space with fewer dimensions for multiple modalities or viewpoints, covering a diverse set of uses, such as deep CCA [40]. However, CCA is a conventional learning technique. In this chapter we will discuss the multi-view extension works of conventional learning methods.



(a) A object is described by text, video, audio

(b) A news is reported by different languages

(c) A product can be represented by multi-view graphs

(d) An image is depicted by heterogonous features

(e) A social image along with user tags

(f) One action can be captured from different camera viewpoints

Figure 2-1: Examples of multi-view data [40].

2.4 Deep Multi-view Canonical Correlation Analysis

In this section, we first introduce the conventional CCA technique, followed by the Deep Canonical Correlation Analysis with Autoencoders (DCCAE).

2.4.1 Canonical Correlation Analysis

Canonical Correlation Analysis CCA is a widely used and well-known technique that aims to uncover the connections between two given variables by converting them into a shared subspace. The main goal of CCA is to find a linear transformation that brings the variables together in a unified space, where their correlations are both optimized and maximized. Its performance has shown great potential in multiple tasks, including classification [41, 42].

2.4.2 Deep Canonical Correlation Analysis

The traditional CCA is constrained in its ability to uncover only the linear correlation between multiple views of data; thus, various nonlinear extensions of CCA have been proposed in the literature, such as DCCA, that was introduced by Andrew et al in [36]. DCCA aims to capture nonlinear associations among various perspectives by integrating deep neural networks (DNNs) and CCA techniques.

As illustrated in Figure 2-2, Two separate neural networks labelled as $\boldsymbol{f}$ and $\boldsymbol{g}$, are employed in deep **CCA** to optimize the parameters $\boldsymbol{\alpha_f}$ and $\boldsymbol{\alpha_g}$ of these networks. The main goal is to efficiently preserve the canonical correlation between the outputs generated by **f** and **g** at their highest level. These parameters are then trained using back-propagation to optimize the following objective function:

$$(\theta_f^* \theta_g^*) = arg \max_{\theta_f \theta_g} corr\big(f(X_1; \theta_f), g(X_2; \theta_g) \big) \tag{2.5}$$

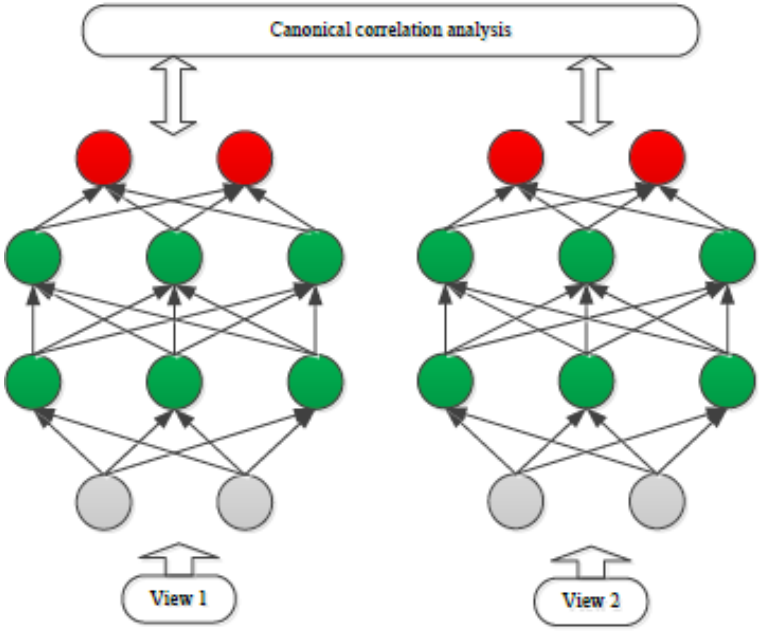


Figure 2-2: The framework of (DCCA) [40].

2.4.3 Deep Canonical Correlated Auto-encoders

Wang et al. [43] introduced a novel deep learning model known as DCCAE, building upon the foundation of deep CCA and reconstruction-based methodologies. This model serves as an extension of the deep CCA framework, offering a new perspective on data representation and correlation learning.

The DCCAE architecture consists of a pair of auto-encoders, as depicted in Figure 2-3. These auto-encoders are designed to optimize the compressed features' canonical relationship and the reconstruction errors simultaneously.

DCCAE's objective function is defined as follows:

$$\mathcal{L}\text{DCCAE} = \max_{\theta_f,\theta_g,\theta_p,\theta_q,U,V} -\frac{1}{n}\text{tr}(U^T f(X_1)g(X_2)^T V)$$

$$+\frac{\lambda}{n}\sum_{i=1}^{n}\left(\left\|x_1^i - p\left(f(x_1^i)\right)\right\|^2 + \left\|x_2^i - p\left(f(x_2^i)\right)\right\|^2\right) \quad (2.7)$$

$$s.t.,\ U^T\left(\frac{1}{n} f(X_1)\, f(X_1)^T + r_{X_1} I\right) = I \quad (2.8)$$

$$V^T\left(\frac{1}{n} f(X_2)\, f(X_2)^T + r_{X_2} I\right) = \text{I} \quad (2.9)$$

$$u_i^T f(X_1)g(X_2)^T v_j = 0 \quad ,\forall ij \quad (2.10)$$

Where the variable ***I*** represents the identity matrix, ***U* = {u_1, u_2, ..., u_L}** and ***V* = {v_1, v_2, ..., v_L}** are the CCA directions that transform the characteristics of auto-encoders into a condensed space with ***L*** units. The terms (rX1, rX2) denote the regularization terms used [40].

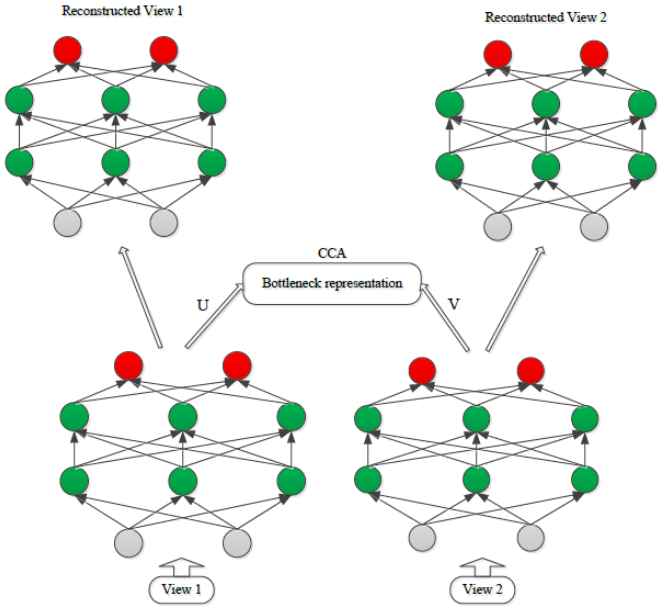


Figure 2-3: The framework of (DCCAE) [40].

2.5 Multi-Label Learning Based AS

In the given AS problem, the complete channel matrix H is regarded as the input instance that is received at every Transmission Time Interval (TTI). The AS vector α is considered as the

label vector consisting of N interdependent labels. Consequently, the AS in massive MIMO systems can be converted into an MLL problem, which focuses on constructing a classifier to forecast α based on H. Starting with the Multi-View Multi-Label Learning (MV-MLL), followed by the neural network-based approaches, namely the Multi-Layer Perceptron (MLP) classifier and Multi-Label Convolutional Neural Networks (MLCNN). Finally, we discuss traditional approaches such as Binary Relevance (BR) and Classifier Chains (CC), known as problem transformation methods, using Three base estimators: SVM, KNN, and Random Forest classifier.

2.5.1 Multi-view MLL

As discussed earlier, MVL Incorporates diverse representations of data to improve generalization. In the case of AS, labels are treated as the second view only in the training phase. To address this, we present a DNN-based framework, integrating the architectures of DCCAE for multi-label classification.

2.5.1.1 Model architecture

As illustrated in Figure 2-4, our Multi-view MLL combines two DNN models (i.e., DCCA and autoencoder), composed of three main components: an encoding network $\boldsymbol{F_e}$, a decoding network $\boldsymbol{F_d}$, and a feature mapping network $\boldsymbol{F_x}$. Each of these components is structured with three fully connected (FC) layers, each with its own specific activation functions; including leaky ReLU, and sigmoid activation functions.

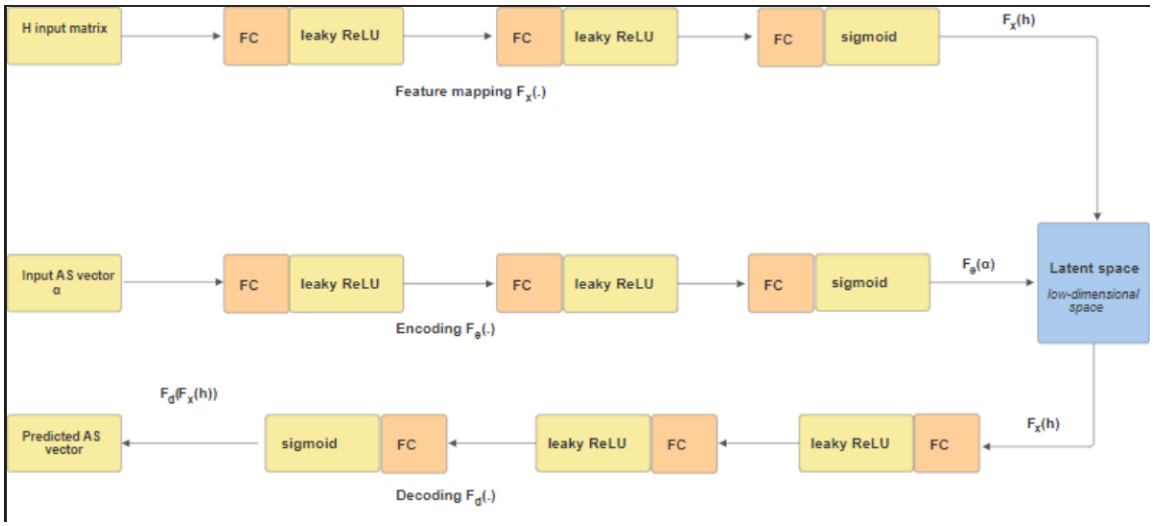


Figure 2-4: Multi-view MLL (DCCAE) architecture.

Where, the objective function of the respective model is as follows:

$$\psi = \min_{F_x F_e F_d} \phi(F_x, F_e) + \alpha\Gamma(F_e, F_d) \tag{2.11}$$

The losses at the latent space and output of Multi-View MLL are represented by $\boldsymbol{\phi(F_x, F_e)}$ and $\boldsymbol{\Gamma(F_e, F_d)}$ respectively, with the parameter $\boldsymbol{\alpha}$ serving as a balance between these two types of loss functions.

After training, the test $\boldsymbol{h}$ channel matrix is transformed into the derived latent space by $\boldsymbol{F_x}$, followed by the decoding mapping of $\boldsymbol{F_x}$ to predict AS vector (i.e., $\boldsymbol{y = F_d(F_x(h))}$); where $\boldsymbol{h}$ is a $\boldsymbol{MN \times 1}$ real vector which is obtained by flattening the channel matrix $\boldsymbol{H}$.

DNNs based DCCA performs joint feature and label embedding by transforming both h and α into the same latent space. (i.e. common low-dimensional space, where the projected representations are highly linearly correlated) [44, 36].

In order to calculate $\boldsymbol{\Phi(F_x, F_e)}$ in equation (2.11), we employ the concept introduced by [45] and rewrite the objective function based on correlation as follows:

$$\min_{F_x F_e} \|F_x(h) - F_e(\alpha)\|_F^2 \tag{2.12}$$

$$s.t. \quad F_x(h)F_x(h)^T = F_e(\alpha)F_e(\alpha)^T = I \tag{2.13}$$

the transformed feature and label data in the derived latent space **L** are represented by $F_x(h)$ and $F_e(\alpha)$ respectively. Additionally, the identity matrix $I \in \mathbb{R}^{l \times l}$, where I represents the dimension of the latent space L used, As demonstrated in [31].

DNNs are employed to solve $F_x(h)$ and $F_e(\alpha)$ in equation 2.12 with the aim of establishing a robust relationship between feature and label data. This enforces a cohesive and integrated association between the two. Furthermore, the autoencoder component of Multi-view MLL recovers the AS vector preserving cross-label dependency. The loss function is given by [46]:

$$\Gamma(F_e, F_d) = \sum_{i=1}^{N} E_i \tag{2.14}$$

$$E_i = \frac{1}{|\alpha_i^1||\alpha_i^0|} \sum_{(p,q)\in\alpha_i^1\times\alpha_i^0} exp\left(-\left(F_d\left(F_e(h_i)\right)^p - F_d(F_e(h_i))^q\right)\right) \tag{2.15}$$

Where $\alpha_i^1\{\alpha_p(i) = 1\}$ and $\alpha_i^0\{\alpha_q(i) = 0\}$ respectively denote the sets of selected and unselected antennas for the i-th sample.

$$\left(F_d\left(F_e(h_i)\right)^p - F_d(F_e(h_i))^q\right) \tag{2.16}$$

measures the distance between the network's output consisting of the ***p**-th* selected antenna and the ***q***-th unselected antenna**.** Therefore; by minimizing (2.15) we maximize the prediction of positive labels.

2.5.1.2 Optimization

To train the model we need to solve the optimization problem of (2.11), in which the loss terms $\Phi(F_x, F_e)$ and $\Gamma(F_e, F_d)$ are computed in the latent space and at the output of the model, respectively.

The technique of gradient descent is employed to update the network parameters for each loss term. The gradient of $\Phi(F_x,F_e)$ is responsible for updating the feature mapping F_x and encoding F_e, while the gradient of $\Gamma(F_e,F_d)$ updates both the encoding Fe and decoding functions F_d. To compute the gradient term of $\Phi(F_x,F_e)$, equation (2.12) is reformulated with the assistance of Lagrange multipliers [25].

$$\phi(F_x, F_e) = Tr(C_1^T C_1) + \lambda Tr(C_2^T C_2 + C_3^T C_3) \tag{2.17}$$

Where;

$$C_1 = F_x(h) - F_e(\alpha) \tag{2.18}$$

$$C_2 = F_x(h)F_x(h)^T - \mathrm{I} \tag{2.19}$$

$$C_3 = F_e(\alpha)F_e(\alpha)^T - \mathrm{I} \tag{2.20}$$

We can derive then the gradient of Φ (F_x,F_e) with respect to F_e and F_x to as follows:

$$\frac{\partial \phi(F_x,F_e)}{\partial F_x(h)} = 2C_1 + 4\lambda F_x(h)C_2 \tag{2.21}$$

$$\frac{\partial \phi(F_x,F_e)}{\partial F_e(\alpha)} = 2C_1 + 4\lambda F_e(\alpha)C_3 \tag{2.22}$$

Next, the gradient of Γ(F_e,F_d) can be calculated using eq (2.15) with respect to each $F_d(F_e(x_i))^j$ [25]. To simplify the calculations, we let:

$$c_i^j = F_d(F_e(x_i))^j \tag{2.23}$$

The gradient is as follows [25]:

$$\frac{\partial \Gamma(F_e,F_d)}{\partial c_i^j} = \sum_{i=1}^{N} \frac{\partial E_i}{\partial c_i^j} \tag{2.24}$$

$$\frac{\partial E_i}{\partial c_i^j} = \begin{cases} -\frac{1}{|\alpha_i^1||\alpha_i^0|} \sum_{q\in\alpha_i^0} \exp\left(-\left(c_i^j - c_i^q\right)\right), if\ j \in\ \alpha_i^1 \\ \frac{1}{|\alpha_i^1||\alpha_i^0|} \sum_{p\in\alpha_i^1} \exp\left(-\left(c_i^p - c_i^j\right)\right), if\ j \in\ \alpha_i^0 \end{cases} \tag{2.25}$$

Where α_i^1represent the sets of positive labels in α_i; α_i^0 represents the set of negative labels in α_i.

Based on these calculations, it is possible to train our model using gradient descent. The algorithm for this process is outlined in Algorithm 1. Upon the successful completion of training the model, making predictions on a test input $\underline{h}$ can be straightforwardly accomplished by rounding

$$y = F_d\left(F_x(h)\right) \tag{2.26}$$

s.t, $\underline{y}$ and $\underline{h}$ are 2D vectors.

Algorithm 1: Multi-View MLL

Input: Feature matrix H, label matrix α, and l dimension of latent space

Output: F_x , F_e , and F_d

Randomly initialize F_x,F_e,F_d

Repeat

- Randomly Select a batch of data S{h}, S{α}
- Define the loss function by (2.11)
- Perform gradient descent on F_d by (2.25)
- Perform gradient descent on F_x by (2.21)
- Perform gradient descent on F_e by (2.25) and (2.22)

Until Converge

2.5.1.3 hyperparameters

The performance of the proposed DNN-based framework integrating DCCAE for multi-label classification is influenced by several hyperparameters. They are illustrated below in Table 2-1.

Hyperparameters	
Number of neurons in each layer	512
Latent embedding dimension	51
Batch size	500
λ in (2.21) & (2.22)	0.5
α in (2.11)	2
Initial learning rate	10e-4
Decay	2%
Momentum	0.99

Table 2-1: Hyperparameters for DCCAE Framework.

2.5.2 Neural network-based approaches

2.5.2.1 Multilayer Perceptron classifier

The multilayer perceptron (MLP) is a type of deep feedforward neural network (DFN) that maps the inputs to the outputs. It contains multiple fully connected layers (FCs), the Neurons within these layers use nonlinear activation functions, enabling the network to capture complex patterns in the data which makes MLPs powerful for tasks like classification, regression, and pattern recognition [47].

MLP is a finite acyclic graph (i.e. information flow is unidirectional, from the input layer through the hidden layers to the output layer, without any feedback loops) [48]. The nodes are neurons with logistic activation as shown in Figure 2-5 where:

neurons of i-th layer serve as input features for neurons of i + 1th layer and every complex function can be calculated combining many neurons

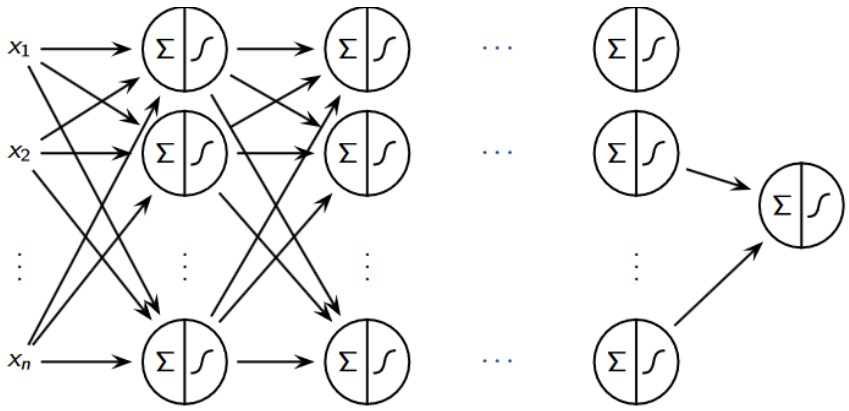


Figure 2-5: General model architecture of MLP [48].

2.5.2.1.1 MLP Architecture

The model consists of three dense layers with a Rectifier Linear Unit (ReLU) as shown in Table 2-2, each followed by a dropout rate and an L2 regularization for a performance of models by reducing overfitting. The output layer uses a sigmoid activation with a number of units equal to the number of output classes which is equal to 64. For training and optimization, the model utilizes the Adam optimizer with an initial learning rate of 0.0001 and binary cross-entropy loss.

Layer Type	Configuration Details
Dense Layer 1	Units: 512, Activation: ReLU, Kernel Regularizer: L2 (0.01), Dropout (Rate).
Dense Layer 2	Units: 128, Activation: ReLU, Kernel Regularizer: L2 (0.01), Dropout (Rate).
Dense Layer 3	Units: 128, Activation: ReLU, Kernel Regularizer: L2 (0.01), Dropout (Rate).
Output Layer	Units: 64, Activation: Sigmoid, Kernel Regularizer: L2 (0.01).

Table 2-2: MLP Architecture.

2.5.2.2 Multi-Label Convolutional Neural Networks

Convolutional Neural Networks are similar to conventional Artificial Neural Networks (ANNs) as they comprise of neurons that enhance their performance through learning. These networks carry out tasks such as scalar products followed by non-linear functions. Inspired by the visual cortex of animals, CNNs are primarily employed in image recognition, leading to quicker learning and reduced error rates [49]. And unlike the MLP model, that only consists of a classification part, CNN model has two distinct parts:

- Convolutional part: the objective is to extract distinctive features from images by passing them through various filters, resulting in convolution maps which then are combined into a feature vector, known as the CNN code.
- Classification part: The CNN code from the convolutional part is fed into fully connected layers (FC); that are positioned at the end of the CNN architecture and are fully connected to all output neurons from the convolutional layers; forming a Multi-Layer Perceptron, that combine the features to classify the image.

A modified version known as the multi-label CNN (MLCNN) is designed to handle tasks where each input can have multiple labels simultaneously like in our case. This enables the network to make more intricate and nuanced predictions compared to single-label classification.

2.5.2.2.1 MLCNN Architecture

Our MLCNN model consists of 4 main layers as shown in Figure 2-6: an input layer, two convolutional layers, fully connected layers, and an output layer. A ReLU activation function defined as f(x)= max(0, x) was utilized in the layers to introduce non-linearity, in addition to the Sigmoid activation function at the output layer to produce the final classification output that is a 64 x 1 vector of ones and zeros representing selected and unselected antennas respectively.

In terms of training and optimization, the model employs Stochastic Gradient Descent (SGD) with an initial learning rate of 0.0001 and binary cross-entropy as the loss function. Dropout ratios of 0.5 and 0.7 and L2 regularization are used to improve the generalization performance of models by reducing overfitting.

Layer Type	**Configuration Details**
Input Layer	Input shape: (None, 1, NR, NT)
Convolutional Layer 1	Filters: 32, Kernel Size: 4x4, Strides: 1, Padding: 'same', Activation: ReLU, Kernel Regularizer: L2 (0.01).
Convolutional Layer 2	Filters: 32, Kernel Size: 5x5, Strides: 1, Padding: 'same', Activation: ReLU, Kernel Regularizer: L2 (0.01).
Fully Connected Layers	Flatten function **Dense layer 1**: Units: 512, Activation: ReLU, Kernel Regularizer: L2 (0.01), Dropout (ratio). **Dense layer 2:** Units: 218, Activation: ReLU, Kernel Regularizer: L2 (0.01), Dropout (ratio). **Dense layer 3:** Units: 256, Activation: ReLU, Kernel Regularizer: L2 (0.01), Dropout (ratio).
Output Layer	Units: 64, Activation: Sigmoid.

Table 2-3: MLCNN Architecture.

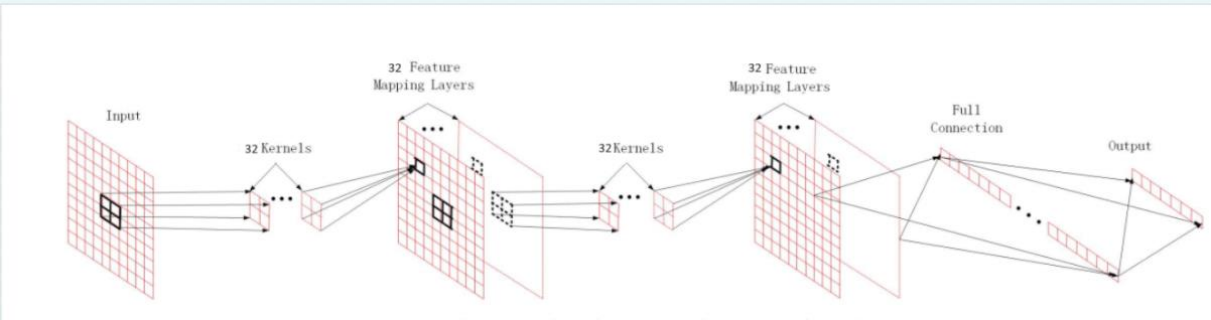


Figure 2-6: MLCNN architecture [50].

Remark: the proposed MLCNN model does not employ max pooling in the convolution layers like CNN models as shown in Table 2-3, and that is because the main functions of the pooling layer employed in conventional CNN models are down sampling, dimensionality reduction, redundant information removal, compression features and overfitting reduction [51].

2.5.3 Problem transformation methods

Problem transformation methods aim to simplify the multi-label classification problem by converting it into single-label classification or regression problems. This makes it easier to apply existing algorithms and models designed for single-label tasks [50].

2.5.3.1 First-order approaches Binary Relevance

Binary Relevance (BR) is by far the simplest approach: The fundamental concept behind this algorithm involves dividing the multi-label learning problem into q_i distinct binary classification problems for each p_i label. Each binary classification problem is specifically linked to one of the potential labels within the label space as shown in Figure 2-7 [51].

For input h, predict independently:

$$\alpha_j = C_j\,(h) \tag{2.27}$$

$$\equiv \arg\max_{\alpha_{j\in\{0,1\}}} p\left(\alpha_j \middle| h\right) \tag{2.28}$$

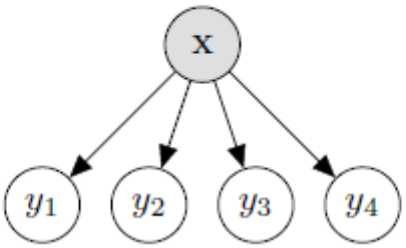


Figure 2-7: Binary relevance [52].

To implement this approach, SVM was utilized as a binary classifier base estimator. SVM offers robust performance and flexibility in handling complex datasets. The Radial Basis Function (RBF) kernel function was employed in conjunction with grid search technique to optimize the gamma and cost parameters, enhancing the model's performance.

Furthermore, the one-vs-rest strategy was used for the construction of multiple binary classifiers, each trained to predict the presence or absence of a specific label. This strategy provides versatility in handling diverse label sets and contributes to the model's overall accuracy. Additionally, the Calibrated Classifier (CV) was adopted to ensure reliable probability estimates and further enhance the accuracy of the model's predictions.

2.5.3.2 High-order approach Classifier Chains

Chain Classifiers (CC) involve converting the multi-label learning problem into a series of binary classification problems. Each subsequent binary classifier in the sequence is constructed based on the predictions made by the preceding classifiers, example in Figure 2-8) [53].

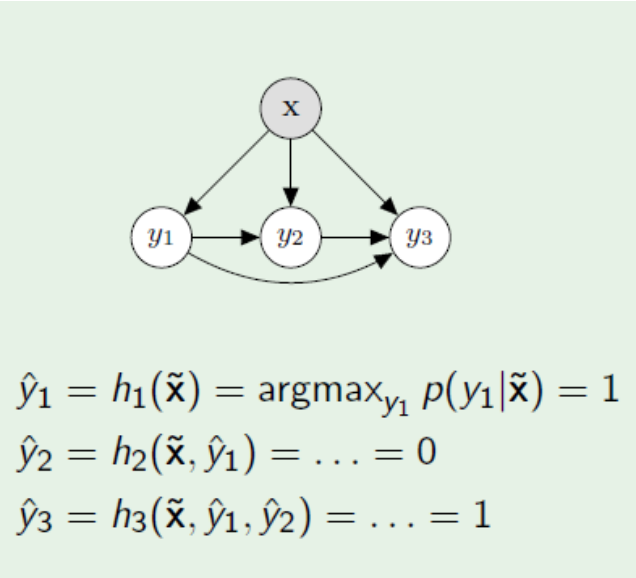


Figure 2-8: Chain classifier [52].

In this approach, three base estimators were employed: the K-Nearest Neighbours (KNN), which relies on proximity in feature space for classification; Random Forest, known for its ensemble-based approach and robustness to noise and overfitting SVM as well, famous for its ability to delineate complex decision boundaries. This diverse ensemble of classifiers offers versatility and resilience to various data distributions and complexities.

For the parameters tuning, a Calibrated Classier was leveraged along with the grid search technique, exploring the hyperparameter space to identify the optimal configuration along with. For instance, grid search optimizes parameters such as 'n_neighbors' for KNN and 'n_estimators' and 'max_depth' for Random Forest, to improve the models' predictive performance.

2.6 Conclusion

In our study of antenna selection, we have leveraged a combination of Deep Canonical-Correlation analysis and Autoencoder DCCAE architectures within a DNN framework, which enables end-to-end learning and prediction with the ability to exploit label dependency. We have exploited the potential of neural network architectures such as MLCNN and MLP to capture complex data relationships and patterns. Furthermore, our methodology incorporated advanced techniques like ClassifierChain, hyperparameter tuning via grid search, and calibrated probability estimation to refine model performance, using various base estimators.

The outcomes of our endeavours will be discussed in detail in the subsequent chapter, culminating in the identification of the best model for antenna selection in communication systems.

Chapter 3 : Results and discussion

3.1 Introduction

This chapter presents the implementation of different multi-label Learning techniques, previously studied, for antenna selection in massive MIMO systems; using indoor channel measurements with three different antenna array configurations. We evaluate their performance in terms of average precision and Hamming loss, demonstrating their capability to identify the optimal subset of antennas that maximize system capacity. Additionally, we assess the computational complexity of each approach.

3.2 Dataset generation

In this work, we used a dataset collected in a realistic indoor setting using the KU Leuven Massive MIMO testbed and a CNC XY table. The dataset includes 252,004 CSI samples taken from various user locations within an office environment, where four single-antenna user devices were placed on CNC XY tables, allowing their antennas to move along a predefined path.

This path followed a zigzag pattern on a grid with 5mm steps, covering an area of 1.25m by 1.25m. The BS had 64 antennas capable of transmitting and receiving simultaneously and was set to a centre frequency of 2.61 GHz. Three antenna array configurations were used: an 8x8 Uniform Rectangular Array (URA), a 64-antenna Uniform Linear Array (ULA), and a distributed scenario (DIS) where antennas were grouped in pairs of eight and distributed throughout the room. This dataset is available for download from [54].

The transmit AS problem has been formulated as a convex optimization problem and solved with numerical methods at CDTA, using the MATLAB based tool for convex optimization called CVX [55]. We then used the conducted results as the optimal benchmarks and compared them with our work findings.

All our simulations were implemented in Python. They were conducted in a virtual environment running on i5-8350U CPU with 16 GB RAM. All models were trained and tested on the same dataset for a fair comparison, with train, test and validation data of size 6000, 2000 and 2000 respectively.

3.3 Evaluation Metrics

In a traditional supervised learning, conventional metrics such as accuracy, and F-measure are used to evaluate the model's performance in single-label classification. However, in MLL, to gain a full understanding of the model's performance, specific MLL evaluation metrics should be used.

In this work, Example-based metrics are used to assess the model's performance, namely Hamming loss, and Average precision. These metrics evaluate the performance on each test sample individually, then return the mean value across the entire test set.

Before defining the metrics, it's essential to identify the types of predictions a classifier can make:

- True Positive (TP): Label: 1, Prediction: 1
- False Positive (FP): Label: 0, Prediction: 1
- False Negative (FN): Label: 1, Prediction: 0
- True Negative (TN): Label: 0, Prediction: 0.

3.3.1 Hamming Loss

Hamming loss represents the fraction of labels that are incorrectly predicted (Hamming Distance between predictions and labels), and is calculated as follows [56]:

$$Hloss = \frac{1}{p}\sum_{i=1}^{p}|h(x_i)\Delta Y_i| \tag{3.1}$$

Where, Δ stands for the symmetric difference between the two sets: h and Y that represents the channel matrix and the labels vector of size p, respectively.

3.3.2 Average precision:

The average precision evaluates the average fraction of relevant labels ranked higher than the irrelevant ones for each instance. Its equation is represented as [56]:

$$Avgprec(f) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{|Y_i|}\sum_{y\in Y_i}\frac{|\{y'|rank_f(x,y')\leq rank_f(x_i,y),y'\in Y_i\}|}{rank_f(x_i,y)} \quad (3.2)$$

3.4 Classification performance evaluation

The histogram in Figure 3-1, compares the models' performance across the three configurations: Distributed, Linear and Rectangular for the case of Ns=44 and SNR=0 Db.

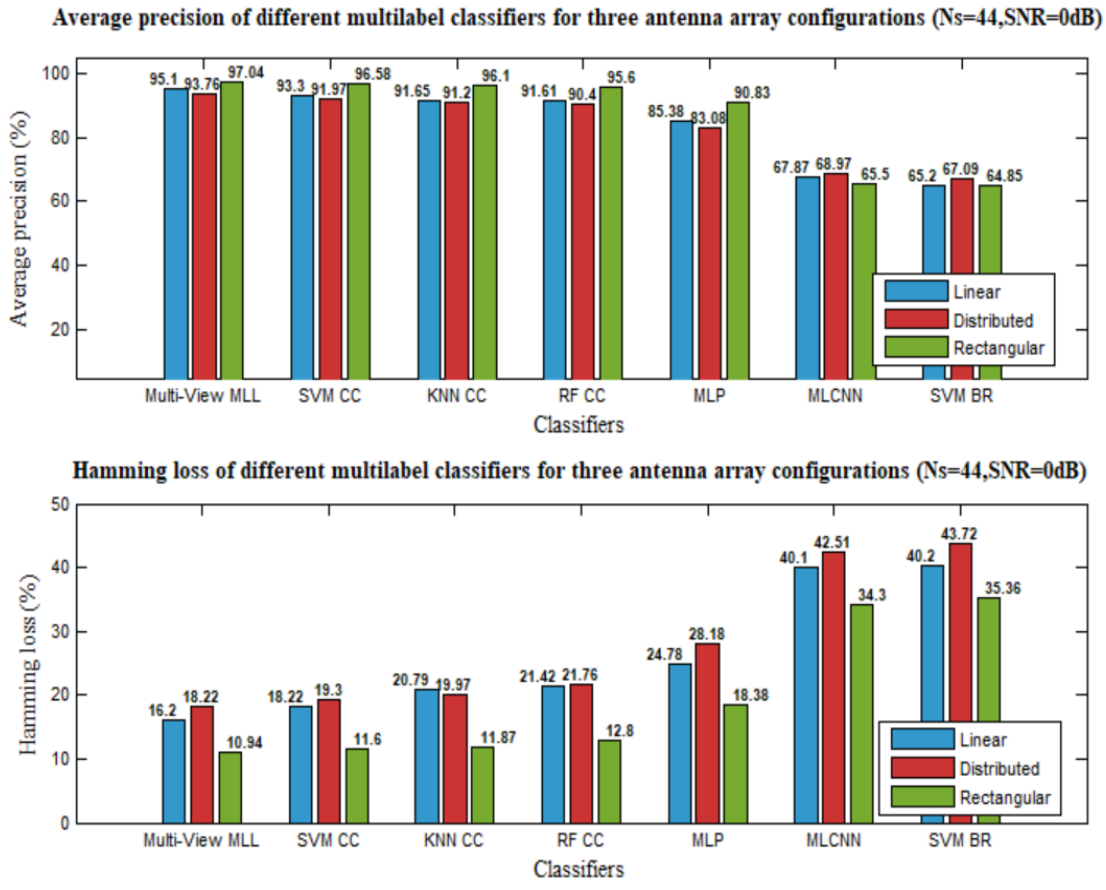


Figure 3-1: Performance evaluation of different multilabel classifiers for three antenna array configurations.

It is observed that the models that lack the ability to exploit label co-occurrence information such as, MLCNN, MLP, and SVM BR, fail to have a satisfactory result. However, the models that exploit label dependencies, namely: multi-View MLL, SVM CC, KNN CC, and Random Forest CC; multi-View MLL had the best performance results due to its low complexity and ability to perform joint feature and label embedding while simultaneously exploit cross-label dependencies with the label-correlation sensitive loss function introduced at the output of the DCCAE.

Despite chain classifiers' ability to capture label correlations, errors may propagate down the chain, which explains the difference in performance compared to DCCAE based model.

3.5 The antenna subset size impact on the capacity

To evaluate the performance of the different MLL approaches relative to the size of the selected antenna subset, we compute the system capacity for various Ns values across the three different array topologies at SNR=0dB. As shown in Figures 3-(2,3 and 4), Capacity increases with the size of antenna subset. This is expected due to the increase of spatial degree of freedom for all schemes.

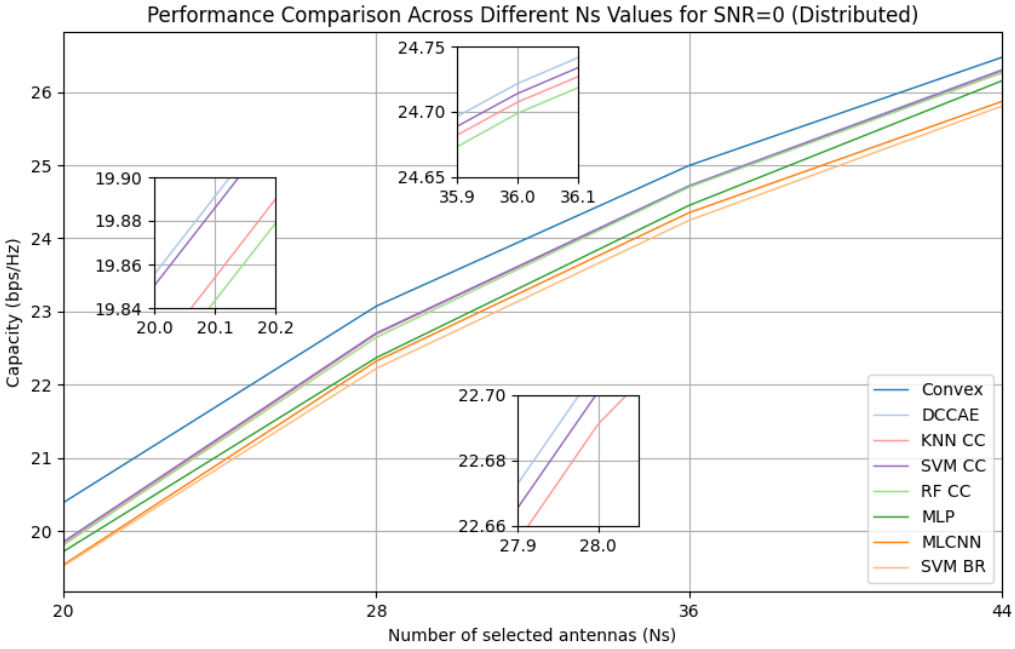


Figure 3-2: Performance comparison Across Different Ns values for SNR=0 (Distributed).

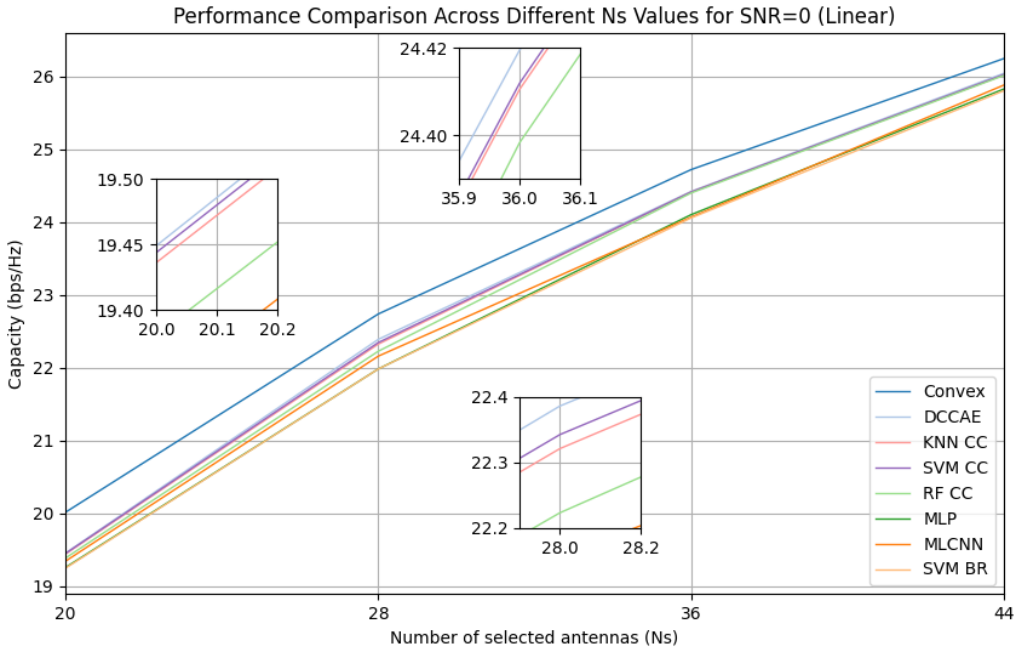


Figure 3-3: Performance comparison Across Different Ns values for SNR=0 (Linear).

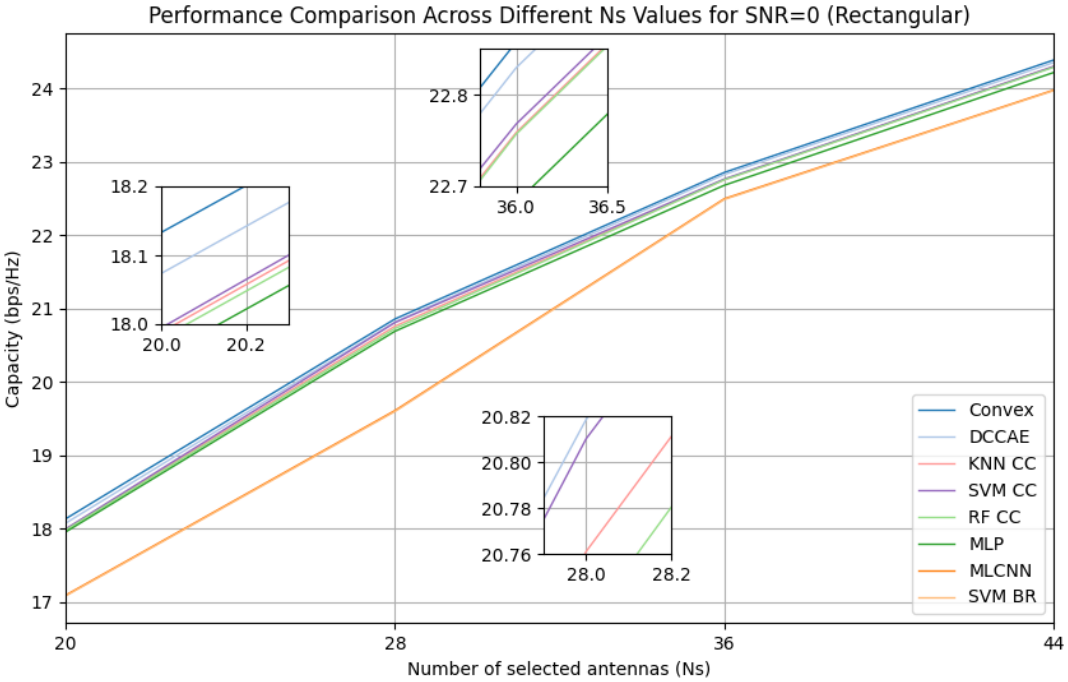


Figure 3-4: Performance comparison Across Different Ns values for SNR=0 (Rectangular).

Noticeably, the DCCAE based approach showcased an overall better performance across all the cases, since integrating DCCA ensures maximizing features-labels correlations while simultaneously predicting labels; achieving a satisfactory labels recovery thanks to the integration of the autoencoder.

Chain classifiers proved to outperform BR and MLP classifier since they can effectively exploit label correlations. Whereas both BR and MLP treat each label independently.

Both DCCAE based approach and chain classifiers explicitly exploit label dependencies, either using custom loss function (like DCCAE) or using chaining property like chain classifiers (i.e. The first classifier makes a prediction of the first label; subsequent classifiers make use of predictions from the preceding ones), unlike CNN based model, it does not directly/ explicitly exploit label dependencies explaining the performance gap between them.(i.e. binary cross-entropy used in CNN fails to capture dependencies among labels).

3.6 Evaluation of MLL based Classifiers' Performance in Terms of Capacity

We evaluate the performance of different MLL classifiers using channel capacity as the metric. This is calculated by averaging the results from 2000 individual capacity values, each corresponding to different CSI instances. For each instance, the capacity is computed independently by identifying the antenna subset using each classifier. We compare the obtained results to those of the convex relaxation-based method, with the SNR set in the range of 0 to 10 dB.

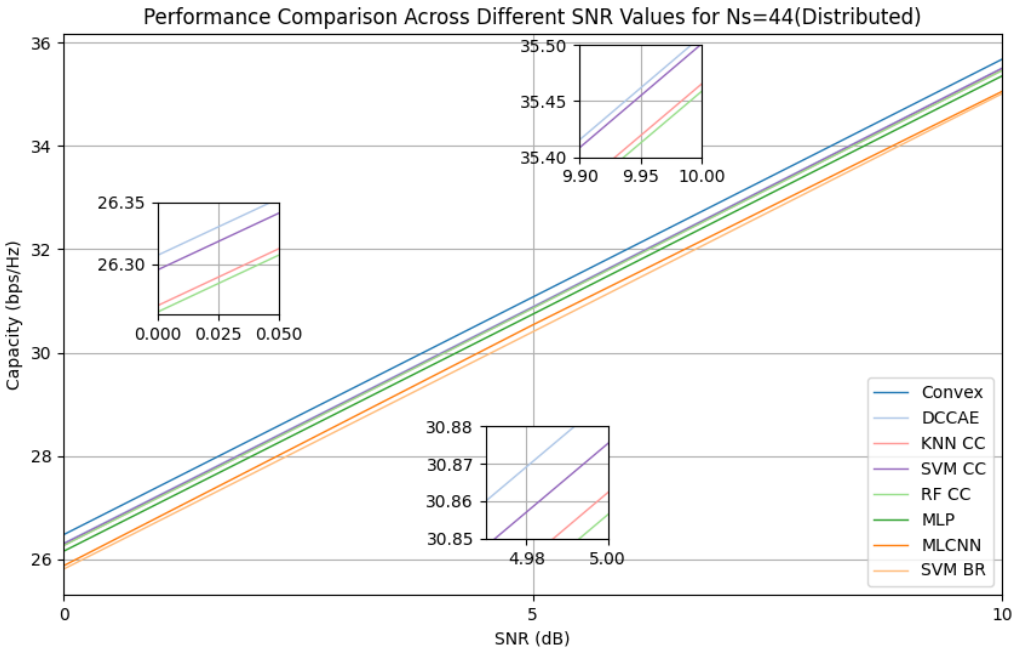


Figure 3-5: Performance Comparison Across Different Values for Ns=44 (Distributed).

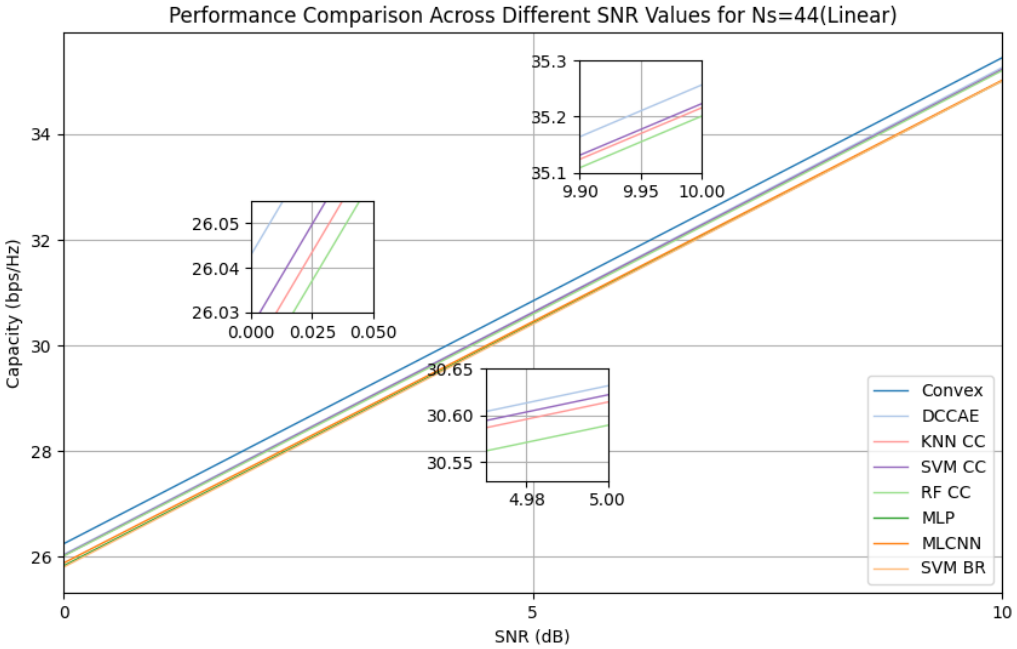


Figure 3-6 Performance Comparison Across Different Values for Ns=44 (Linear).

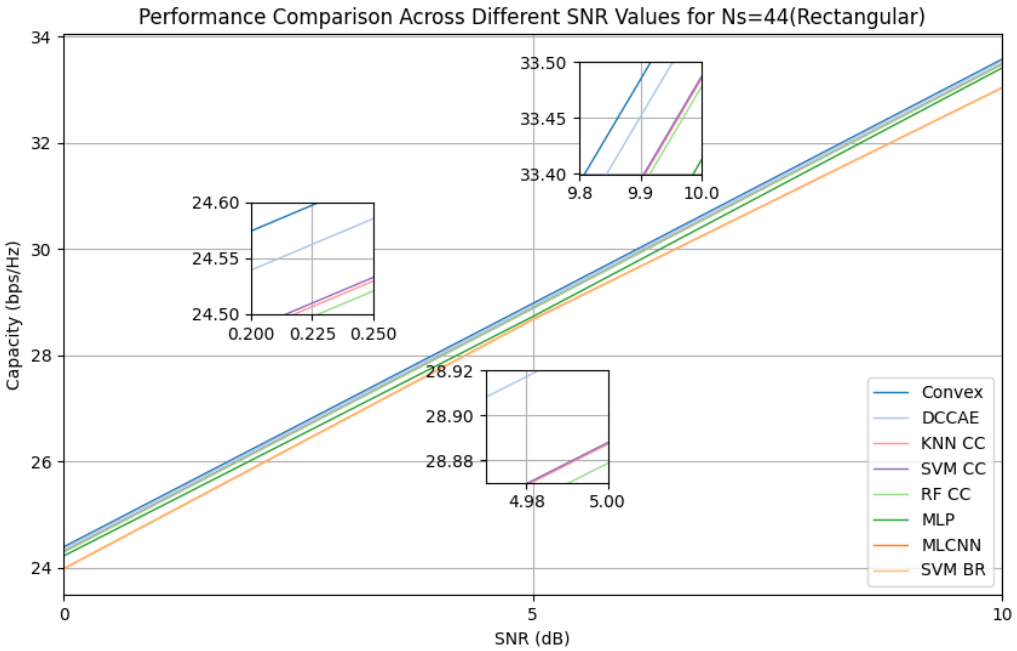


Figure 3-7: Performance Comparison Across Different Values for Ns=44 (Rectangular).

Results depicted in Figures 3-(5,6 and 7) illustrate the capacity graphs of the case of Ns=44 antennas across three array topologies: Linear, Distributed, and Rectangular. It can be seen that for the varying values of SNR from 0 to 10 dB, there is a slight difference between the three configurations and the DCCAE-based approach demonstrated superior performance across all cases. Additionally, chain classifiers (SVM CC and RF CC) outperformed both BR and MLP and MLCNN classifiers.

3.7 Computational complexity

One of the characteristics that reflects the model's complexity is computational time, it is evident that in real life application, elapsed time plays a vital role in determining a model's practicality. Moreover, Antenna Selection is a real-time task requiring fast prediction that lie within the channel coherence time interval [25].

Ns=44	SNR 0	SNR 5	SNR 10
DCCAE	0.1624	0.1605	0.1743
KNN CC	12.0474	12.0773	12.1480
RF CC	12.0470	12.0665	12.136
SVM CC	33.2187	29.9430	29.9407
MLP	0.3100	0.3100	0.3300
MLCNN	0.4100	0.3500	0.3400
SVM BR	46.9205	42.0679	42.4178

Table 3-1: Inference time in the distributed configuration array.

Ns=44	SNR 0	SNR 5	SNR 10=
DCCAE	0.1597	0.1670	0.1683
KNN CC	18.2705	16.4433	16.5751
RF CC	12.0289	9.8039	10.0291
SVM CC	26.3945	27.0207	26.7579
MLP	0.1600	0.1650	0.1601
MLCNN	0,3300	0.3200	0.3500
SVM BR	46.6494	38.8872	38.8917

Table 3-2: Inference time in the linear configuration array.

Ns=44	SNR 0	SNR 5	SNR 10
DCCAE	0,1699	0.1667	0.1560
KNN CC	14.6173	13.9033	15.0775
RF CC	8.7044	8.4613	8.0776
SVM CC	12.8950	12.5829	12.7150
MLP	0.1400	0.1401	0.1420
MLCNN	0.7302	0.5100	0.7110
SVM BR	20.3299	20.4463	20.05395

Table 3-3: Inference time in the rectangular configuration array.

In Tables 3-(1,2 and 3), computational time is compared for each case, all models need to be trained first then tested using unseen data. We notice that Neural network-based methods take

few minutes to converge, while the classifiers require a prolonged period of time; more than 50 minutes for SVM BR.

Despite the fact that the performance gap between DCCAE based model and the chain classifiers are relatively close, it is clear that these classifiers suffer from a high computational complexity, as for multi-View MLL DCCAE, it takes around 0.16 ms to predict the labels vector, meanwhile it takes the classifiers a minimum of 14 ms.

Therefore, chain classifiers are not a practical solution for AS due to its high computational time, on the other hand, multi-View MLL network offers a substantial acceleration, combined with its low complexity architecture, positioning it as a viable option for AS.

Moreover, we can see that multi-label learning is generally a difficult task to solve when the size of the label space becomes large i.e. As the number of class labels increases, the number of possible label sets grows exponentially. For example, in a label space with 64 class labels (q = 64), there would be over one million potential label sets (2^64). The key to overcome this challenge, is to improve the learning process by utilizing correlations or dependencies among labels [55]. For instance, if an image is tagged with labels such as pyramids and sands, there is a high likelihood that it will also be labelled as Egypt. Conversely, a document that pertains to entertainment is improbable to be categorized as science. In the context of wireless communication, labels are often correlated due to the underlying physical and operational dependencies between antennas; as an example, antennas that are closer to one another experience similar channel conditions such as interference pattern. hardware resources like RF chains are often shared among multiple antennas due to the overwhelming number of bs antennas and limited number of RF chains in Massive MIMO, and therefore selecting one antenna can affect the performance of others. Therefore, effectively exploiting label co-dependency is considered essential for the success of multi-label learning techniques.

3.8 Conclusion

On the basis of our simulation results, along with the state of art studies and investigations, we can derive the following conclusions:

- The selection of the loss function is crucial for it directly influences to what extent the label-correlation is captured.
- The Multi-View MLL model achieved comparable capacity results to Convex optimization with reduced prediction time.
- High computational models are not practical in real world application.
- Due to the large output space, problem transformation methods require a large number of binary classifiers, which results in high computational complexity.
- Low complexity models are the best candidate for AS.

Conclusion

This work aims to explore the possibility of using Machine Learning for Antenna Selection to mitigate the complexity of massive MIMO system by reducing hardware costs, power consumption, and signal processing complexity. Several approaches were tested and their ability to accurately predict a subset of the antenna set was evaluated.

- **Is it Possible to Approach AS as a Multi-Label Classification Problem?**

As shown, AS in massive MIMO can be transformed into a MLL problem by taking the channel matrix as the input and antenna vector can be treated as the label vector. The investigation then primarily focuses on which MLL algorithm is a potential candidate for AS to achieve comparable results to the near optimal ones calculated using convex optimization.

In the case of AS, and unlike typical multi-label classification problems, the goal is to achieve the best communication performance instead of classification performance. That is, computation time, and capacity calculation were used in addition to average precision and Hamming loss for the purpose of performance evaluation.

- **What are the main MLL approaches?**

In this work, two main approaches were investigated: problem transformation methods, DNN based algorithms. The first is a machine learning-based approach and the latter is a deep learning-based approach. In problem transformation methods, the MLL problem is transformed into a single-label classification problem by dividing the MLL problem into N independent binary classifiers.

Three different base estimators were used, namely: SVM, KNN, and Random Forest.

In DNN based algorithms, three different architectures were investigated, CNN based-model, MLP classifier, and multi-View based model. CNN-based model employs convolutional layers to automatically extract spatial features from input data. On the other hand, MLP classifier Employs fully connected layers to learn and map input features to output labels. Meanwhile, the Multi-View based model integrates both DCCA and AE architectures in a

unified network, with label-correlation-sensitive function introduced at the output to ensure label-dependencies exploitation.

- **Which model yields the highest prediction performance?**

The multi-View based MLL model used consistently yielded higher prediction accuracies, faster prediction time, and achieved comparable capacity. The multi-View MLL integrates both DCCA and autoencoder architecture to perform joint feature and label embedding, while simultaneously extracting the interdependency among selected antennas. This low-complex architecture solves the large label space problem and achieved a trade-off between the computational complexity and the system performance,

Future work

In this report, the multi-View MLL was the best performing model. However, other aspects of the models' generalization ability have not been investigated, such as, topology mismatch. That is, when there is a mismatch in terms of the user number. Therefore, there is still a lot of potential future work to do to investigate more MLL based approaches and how they can be combined with multi-View MLL to further enhance the performance.

We note that there have been some recent advances to tackle multi-label classification problem:

Advanced Kernel-Based Learning Systems: is recently introduced and specifically designed for multi label classification problems, it ensures faster convergence, and better generalization performance. Future research should explore the application of these systems to AS [57].

Multi-label Classification with Reinforcement Learning: Reinforcement learning offers a robust approach to achieving optimal decision-making in numerous complex tasks that involve multi-dimensional data. Applications of deep RL method should be further explored to develop RL-based AS algorithms that learn optimal antenna subsets through continuous interaction with the environment [58].

References

[1] *molisch, a. f., win, m. z., choi, y. s., & winters, j. h. (2005). capacity of mimo systems with antenna selection. ieee transactions on wireless communications, 4(4), 1759-1772.*

[2] *Foschini, G. J., & Gans, M. J. (1998). On limits of wireless communications in a fading environment when using multiple antennas. Wireless personal communications, 6, 311-335.*

[3] *Marzetta, T. L. (2015). Massive MIMO: An Introduction. Bell Labs Technical Journal, 20, 11–2.*

[4] *Thomas L. Marzetta, Erik G. Larsson, Hong Yang, and Hien Quoc Ngo. Fundamentals of Massive MIMO. Cambridge University Press, 2016.*

[5] *T. L. Marzetta. Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas. IEEE Transactions on Wireless Communications, 9(11):3590–3600, November 2010.*

[6] *L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," IEEE J. Sel. Topics Sig. Proc., vol. 8, no. 5, pp. 742 - 758, May 2014.*

[7] *Challita, F. (2019). Massive MIMO channel characterization and propagation-based antenna selection strategies: application to 5G and industry 4.0 (Doctoral dissertation, Université de Lille).*

[8] *Tai, T.-H., Chung, W.-H., & Lee, T.-S. (2015). A Low Complexity Antenna Selection Algorithm for Energy Efficiency in Massive MIMO Systems. 2015 IEEE International Conference on Data Science and Data Intensive Systems.*

[9] *Singh, M. S. J., Saleh, W. S. W., Abed, A. T., & Fauzi, M. A. (2023). A review on massive MIMO antennas for 5G communication systems on challenges and limitations. Jurnal Kejuruteraan, 35(1), 95-103.*

[10] *Molisch, A. F., Ratnam, V. V., Han, S., Li, Z., Nguyen, S. L. H., Li, L., & Haneda, K. (2017). Hybrid beamforming for massive MIMO: A survey. IEEE Communications magazine, 55(9), 134-141.*

[11] *Ibrahim, S. K., Singh, M. J., Al-Bawri, S. S., Ibrahim, H. H., Islam, M. T., Islam, M. S., ... & Abdulkawi, W. M. (2023). Design, challenges and developments for 5G massive*

MIMO antenna systems at sub 6-GHz band: a review. Nanomaterials, 13(3), 520.

[12] *Chataut, R., & Akl, R. (2020). Massive MIMO systems for 5G and beyond networks—overview, recent trends, challenges, and future research direction. Sensors, 20(10), 2753.*

[13] *T. E. Bogale and L. B. Le. Massive MIMO and mmWave for 5G Wireless HetNet: Potential Benefits and Challenges. IEEE Vehicular Technology Magazine, 11(1):64–75, March 2016.*

[14] *Sanayei, S., & Nosratinia, A. (2004). Antenna Selection In Mimo Systems. Ieee Communications Magazine, 42(10), 68–73. Doi:10.1109/Mcom.2004.134126.*

[15] *Gao, Y., Vinck, H., & Kaiser, T. (2018). Massive MIMO Antenna Selection: Switching Architectures, Capacity Bounds, and Optimal Antenna Selection Algorithms. IEEE Transactions on Signal Processing, 66(5), 1346–1360.*

[16] *Le, N. P., Safaei, F., & Tran, L. C. (2016). Antenna Selection Strategies for MIMO-OFDM Wireless Systems: An Energy Efficiency Perspective. IEEE Transactions on Vehicular Technology, 65(4), 2048–2062.*

[17] *Yongqiang, H., Wentao, L., & Xiaohui, L. (2011). Particle Swarm Optimization for Antenna Selection in MIMO System. Wireless Personal Communications, 68(3), 1013–1029. doi:10.1007/s11277-011-0496-z.*

[18] F. Bouchibane, H. Tayakout, N. Ziane, F. Siahmed And S. Hebib, "Upgraded-Abc Algorithm For Antenna Selection In Energy Efficient Massive Mimo System," *2023 International Conference On Advances In Electronics, Control And Communication Systems (Icaeccs)*, Blida, Algeria, 2023, Pp. 1-5.

[19] *A. Gorokhov, "Antenna Selection Algorithms For Mea Transmission Systems," Proc. Ieee Icassp, Pp. 2857–2860, May 2002.*

[20] *Gharavi-Alkhansari, M., & Gershman, A. B. (2004). Fast Antenna Subset Selection In Mimo Systems. Ieee Transactions On Signal Processing, 52(2), 339–347.*

[21] *Tsoumakas, G., and Katakis, I. 2006. Multi-label classification: An overview. Dept. of Informatics, Aristotle University of Thessaloniki, Greece.*

[22] *J. Joung, "Machine learning-based antenna selection in wireless communications,"IEEE Commun. Lett., vol. 20, no. 11, pp. 2241–2244, Nov. 2016.*

[23] *A. M. Elbir and K. V. Mishra, "Joint antenna selection and hybrid beamformer design*

using unquantized and quantized deep learning networks," IEEE Trans. Wireless Commun., vol. 19, no. 3, pp. 1677–1688, Mar. 2020.

[24] Bouch*ibane, F. Z., Tayakout, H., & Boutellaa, E. (2023). A deep learning-based antenna selection approach in MIMO system. Telecommunication Systems, 84(1), 69-76.*

[25] *W. Yu, T. Wang and S. Wang, "Multi-Label Learning Based Antenna Selection in Massive MIMO Systems," in IEEE Transactions on Vehicular Technology, vol. 70, no. 7, pp. 7255-7260, July 2021.*

[26] *S. Sun, A survey of multi-view machine learning, Neural Computation Applications 23 (7-8) (2013) 2031–2038..*

[27] *J. R. KETTENRING, Canonical analysis of several sets of variables, Biometrika 58 (3) (1971) 433–451.*

[28] *Y. Zhang, J. Zhang, Z. Pan, D. Zhang, Multi-view dimensionality reduction via canonical random correlation analysis, Frontiers Computer Science 10 (5) (2016) 856–869.*

[29] *L. Sun, B. Ceran, J. Ye, A scalable two-stage approach for a class of dimensionality reduction techniques, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD), 2010, pp. 313–322.*

[30] *H. Avron, C. Boutsidis, S. Toledo, A. Zouzias, Ecient dimensionality reduction for canonical correlation analysis, in: Proceedings of the 30th International Conference on Machine Learning (ICML), Vol. 28, 2013, pp. 347–355.*

[31] *X. Zhang, X. Yang,W. Zhang, G. Li, H. Yu, Crowd emotion evaluation based on fuzzy inference of arousal and valence, Neurocomputing 2021.*

[32] *A. Kumar, H. D. III, A co-training approach for multi-view spectral clustering, in: Proceedings of the International Conference on Machine Learning (ICML), 2011, pp. 393–400.*

[33] *Z. Xue, J. Du, D. Du, S. Lyu, Deep low-rank subspace ensemble for multi-view clustering, Information Sciences 482 (2019) 210–227.*

[34] *F. R. Bach, M. I. Jordan, Kernel independent component analysis, Journal of Machine Learning Research (JMLR) 3 (2002) 1–48.*

[35] *G. Bhatt, P. Jha, B. Raman, Representation learning using step-based deep multi-modal autoencoders, Pattern Recognition (PR) 95 (2019) 12–23.*

[36] *G. Andrew, R. Arora, J. A. Bilmes, K. Livescu, Deep canonical correlation analysis, in: Proceedings of the International Conference on Machine Learning (ICML), 2013, pp. 1247–1255.*

[37] *Y. Ming, X. Meng, C. Fan, H. Yu, Deep learning for monocular depth estimation: A review, Neurocomputing Volume 438, 28 May 2021, Pages 14-33.*

[38] *W. Wang, R. Arora, K. Livescu, J. A. Bilmes, On deep multi-view representation learning, in: Proceedings of the International Conference on Machine Learning (ICML), 2015, pp. 1083–1092.*

[39] *P W. Fan, Y. Ma, H. Xu, X. Liu, J. Wang, Q. Li, J. Tang, Deep adversarial canonical correlation analysis, in: Proceedings of the SIAM International Conference on Data Mining (SDM), 2020, pp. 352–360.*

[40] *Xiaoqiang Yana , Shizhe Hua , Yiqiao Maoa , Yangdong Yea,∗ , Hui Yub. Deep Multi-view Learning Methods: A Review.*

[41] *T. Kim, S. Wong, R. Cipolla, Tensor canonical correlation analysis for action classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007.*

[42] *L. Sun, S. Ji, J. Ye, Canonical correlation analysis for multilabel classification: A least-squares formulation, extensions, and analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 33 (1) (2011) 194–200.*

[43] *T. X. Vu et al., "Machine learning based antenna selection and power allocation in multi-user MISO systems," in Proc. WiOPT, Avignon, France, Jun. 2019.*

[44] *G. Andrew et al., "Deep canonical correlation analysis," in Proc. ICML,Atlanta, GA, USA, Jun. 2013.*

[45] *Kettenring, J. R. 1971. Canonical analysis of several sets of variables. Biometrika 58(3):433–451.*

[46] *Zhang, M.-L., and Zhou, Z.-H. 2006. Multilabel neural networks with applications to functional genomics and text categorization. IEEE transactions on Knowledge and DataEngineering 18(10):1338–1351.*

[47] *Taud, H., & Mas, J. F. (2018). Multilayer perceptron (MLP). Geomatic approaches for modeling land change scenarios, 451-455.*

[48] *Riedmiller, M., & Lernen, A. (2014). Multi layer perceptron. Machine Learning Lab*

Special Lecture, University of Freiburg, 24.

[49] *O'shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458.*

[50] *An, W., Zhang, P., Xu, J., Luo, H., Huang, L., & Zhong, S. (2020). A novel machine learning aided antenna selection scheme for MIMO Internet of Things. Sensors, 20(8), 2250.*

[51] *Boureau, Y. L., Ponce, J., & LeCun, Y. (2010). A theoretical analysis of feature pooling in visual recognition. In Proceedings of the 27th international conference on machine learning (ICML-10) (pp. 111-118).*

[52] *Tsoumakas, Grigorios & Katakis, Ioannis. (2009). Multi-Label Classification: An Overview. International Journal of Data Warehousing and Mining. 3. 1-13. 10.4018/jdwm.2007070101.*

[53] *Zhang, Min-Ling & Zhou, Zhi-Hua. (2014). A Review On Multi-Label Learning Algorithms. Knowledge and Data Engineering, IEEE Transactions on. 26. 1819-1837. 10.1109/TKDE.2013.39.*

[54] *De Bast S, Pollin S, Ultra Dense Indoor MaMIMO CSI Dataset. IEEE Dataport; 2021. https://dx.doi.org/10.21227/nr6k-8r78.*

[55] *Grant M, Boyd S, CVX: Matlab Software for Disciplined Convex Programming, version 2.1; 2014.*

[56] *Zhang, M.-L., & Zhou, Z.-H. (2014). A Review on Multi-Label Learning Algorithms. IEEE Transactions on Knowledge and Data Engineering, 26(8), 1819–1837. doi:10.1109/tkde.2013.39.*

[57] *Mohammad Yekta Saidabad, Hiwa Hassanzadeh, Seyed Hossein Seyed Ebrahimi, Edris Khezri, Mohammad Reza Rahimi, Mohammad Trik,"An efficient approach for multi-label classification based on Advanced Kernel-Based Learning System, Intelligent Systems with Applications, "Volume 21,2024,200332, ISSN 2667-3053.*

[58] H. Lee, M. Girnyk, and J. Jeong, "Deep reinforcement learning approachto MIMO precoding problem: Optimality and robustness," arXiv preprint arXiv:2006.16646, Jun. 2020.