UQO
UNIVERSITÉ
DU QUÉBEC
EN OUTAOUAIS

Deep Learning Modeling of channel propagation in indoor environments

Par

Khaled Kedjar

Thèse présentée au

Département d'informatique et ingénierie

Pour l'obtention du grade de

Philosophiae Doctor (Ph.D.)

en Sciences et technologies de l'information

Membres du jury : *Président* *Prof. Michael Korwin-Pawlowski, UQO*
  Examinateur externe *Dr. Zhihong Hunter Hong, CRC- Ottawa*
  Examinateur interne *Prof. Ana-Maria Cretu, UQO*
  Directeur de thèse *Prof. Larbi Talbi, UQO*
  Codirecteur *Prof. Nedil Mourad, UQAT*

This thesis is dedicated to

Yema dh Vava

Lovely wife

Brothers and sister

As Roma & Toti's family

And

The memory of Algerian martyrs

ABSTRACT

Recently the rising internet of things technology has pushed the wireless communication systems to require more channel capacity and coverage, resulting in using high frequency bands. Moreover, deep learning algorithms can be very useful to explore the huge amount of the collected data during the experimental measurement and it is observed that they are suitable to non-linear problems. Therefore, it brings a new model as a solution for channel characterization and modeling in any complex environment. The first contributions of this PhD thesis consist in developing an efficient channel characterization and modeling based on deep learning algorithms in LOS (Line-of-Sight) scenario. The learning and the validation process are performed using measurements from only one environment, enabling robust model learning and prediction results. Then, the model efficiency is analyzed and validated using measurements from different environments which are not present during the learning process. Finally, the channel characterization is made with the predicted and the measured ones. The proposed model achieved a highly accurate channel frequency response prediction within different environments without any prior information. The model achieves lower Root Mean Square Error (RMSE) up to 2% compared to the latest proposed models in the literature. Hence, an efficient modeling tool is provided for the future wireless communication design in complex confined environment in LOS scenarios.

The second contribution of this PhD consists in developing an efficient model is developed for 2 × 2 Wireless Body Area Network Multiple-Input-Multiple-Output (WBAN-MIMO) channel, based on deep learning algorithms. The model is composed of three deep learning algorithms which is known as Stacked Model (SM). The SM predicts simultaneously the channel matrix H in underground mine, identifies the positions of the collected data in both Line of Sight (LoS) and non-Line of the Sight (NLoS) scenarios. The model is trained and evaluated using the magnitude and phase data collected in an underground mine environment within a frequency range of 2.3 GHz - 2.5 GHz. These measurements, conducted with different antenna configurations in LoS and NLoS scenarios, constitute an input to the model. The latest predict the channel matrix H with the position and identify whether the channel is in LoS or NLoS. Finally, the path loss and the channel impulse response models are compared with the measurements-based ones. The modeled channel prediction

exhibited lower Root Mean Square Error (RMSE) for channel prediction, and high classification accuracy for LoS-NLoS and position identification, respectively. The model offers a powerful solution for future wireless designs in terms of channel prediction in underground mine environments.

The methodologies have demonstrated their uses in channel prediction whether Single Input and Single Output (SISO) or MIMO in different environments, considering the variety of polarization, where the same discussions could be held as those measured.

This research will serve as a useful reference for the development of wireless communication systems in any complex area for their applications.

ACKNOWLEDGMENT

LIST OF ACRONYMS AND ABBREVIATIONS

CFR	Channel Frequency Response
CIR	Channel Impulse Response
Cir	Circular
CP	Co-positioned
CSI	Channel State Information
dB	Decibel (ratio in log scale)
dBm	Decibel relative to 1 milliwatt
DL	Deep Learning
FFT	Fast Fourier Transform
EFs	Environment Features
LC	Learning Curve
Lin	Linear
LNA	Low Noise Amplifier
LSTM	Long short term with memory
LOS	Line Of the Sight
LR	Linear Regression
IOT	Internet of things
K	Rician factor

MIMO	Multiple Input Multiple Output
MSE	Mean Square Error
MPCs	Multi-Path components
MPL	Measured Path Loss
NLOS	Non-Line-Of-Sight
NMSE	Normalized Mean Square Error
OFDM	Othogonal Frequency Division Multiplex
PPL	Predicted Path Loss
PDP	Power Delay Profile
PL	Path Loss
RMS	Root Mean Square
RMSE	Root Mean Square Error
RP	Radiation Pattern
Rx	Receiver
SISO	Single Input Single Output
SM	Stacked Model
SLP	Supervised Learning Problem
SNR	Signal-to-Noise Ratio

Tx   Transmitter

VNA   Vector Network Analyzer

WBAN   Wireless Body Area Network

LIST OF TABLES

LIST OF FIGURES

CONTENT

RÉSUMÉ DE LA THÈSE

La caractérisation et la modélisation du canal de propagation sont d'une grande importance pour la conception des communications sans fil, l'analyse de la qualité de la communication et la simulation de la performance du réseau. Dans la littérature, le canal de propagation est défini comme le milieu où les ondes électromagnétiques voyagent entre les antennes, et sont soumises à divers phénomènes physiques qui affectent la qualité de la transmission. Ces phénomènes sont catégorisés par les réflexions, les diffractions et les diffusions, causées par la nature des surfaces et la densité des obstacles dans le site spécifique. Ils provoquent donc des interférences et des distorsions de l'information dans le canal. Plusieurs approches ont été utilisées pour exploiter les ressources des réseaux de propagation, par exemple les techniques seule entrée et une seule sortie (*Single-Input Single-Output* ou bien *SISO*) et multiple entrée et une multiple sortie (*Multiple-Input-Multiple-Output* ou bien *MIMO*). Au cours de la dernière décennie, la communauté des chercheurs a multiplié les efforts pour étudier la propagation des canaux sans fil dans les environnements intérieurs, les bureaux, les bâtiments et les mines souterraines. De plus, l'effet du corps est également considéré dans plusieurs études comme un paramètre important, en particulier pour les futurs systèmes de communication dans les mines souterraines. Par conséquent, les canaux des réseaux corporels sans fil (*Wireless body Area Network ou WBAN*) ont été étudiés dans les mines souterraines afin de comprendre le mécanisme de propagation résultant de l'effet du corps dans des scénarios *SISO* et *MIMO*, en considérant les situations (*Line of the Sight* ou bien *LoS*) et (*No - Line of the Sight* ou bien *NLoS*). Par conséquent, la nécessité croissante d'améliorer les conditions d'exploitation minière en mettant en œuvre des réseaux de communication souterrains efficaces a suscité diverses recherches sur les sites miniers souterrains. Ces études visent à remplacer les communications filaires, dont la maintenance est coûteuse et l'évolutivité limitée, par une liaison sans fil fiable et sûre. Ces études ont été développées dans le but de développer une meilleure compréhension et des solutions pour l'environnement complexe des mines souterraines, qui est caractérisé par des surfaces rugueuses et aléatoires. Les chercheurs ont travaillé sur différentes configurations de systèmes d'antennes telles que MIMO, SISO. De plus, des scénarios de LoS et de NLoS ont été utilisés pour stimuler le comportement réel du signal à bande étroite ou à large bande dans n'importe quelle situation. Bien que ces algorithmes permettent une bonne compréhension des canaux sans fil, ils restent limités à un seul environnement, et il est difficile

d'appliquer ou de conclure les mêmes résultats dans des environnements différents. En conséquence, ces modèles de canaux de propagation ne seront pas suffisants pour concevoir le système d'information. En conséquence, ces modèles de canaux de propagation ne seront pas suffisants pour concevoir le système sans fil dans n'importe quel environnement en raison de la croissance rapide des réseaux et des installations pour les canaux de propagation sans fil qui comportent des normes plus élevées pour une efficacité spectrale élevée.

De nos jours, l'Intelligence Artificielle (*IA*) fait son apparition dans notre vie quotidienne. Nous ne pouvons pas regarder un film ou une émission de télévision sans qu'un robot *IA* n'apprenne le schéma de notre routine quotidienne pour finalement nous faire des suggestions le lendemain, c'est fascinant. Cependant, l'arrière-plan de ces applications est basé sur des algorithmes, à savoir des algorithmes d'apprentissage automatique (*Machine Learning* ou bien *ML*) qui ont été utilisés dans plusieurs applications, telles que la reconnaissance faciale, la détection d'objets, etc. En fait, ces applications collectent plus de données afin de performer et de bien prédire le choix du film que nous voulons regarder. En outre, il existe deux types de problèmes de prédiction : les problèmes de classification, où la cible de sortie est une étiquette de classe discrète, et les problèmes de régression, où la cible est une quantité continue. Par conséquent, comme nous vivons également à l'époque de la technologie de l'Internet des objets (*Internet of things* ou bien *IOT)*, le domaine de l'IA est très utile pour apporter une nouvelle solution dans la vie quotidienne. Cela crée de nouvelles technologies massives où plusieurs objets doivent être connectés en même temps entre eux et aussi avec le serveur hôte. Par conséquent, le besoin de capacité de canal et de couverture devient crucial pour les futurs systèmes de communication sans fil dans les environnements intérieurs. L'algorithme d'apprentissage profond (*Deep Learning* ou bien *DL*) est un sous-domaine de la ML qui structure les algorithmes en plusieurs couches pour créer un réseau neuronal artificiel capable d'apprendre et de prendre des décisions intelligentes par lui-même.

Cette thèse étudie l'application de l'algorithme *DL* pour modéliser le canal intérieur dans les couloirs et les environnements souterrains dans des configurations *SISO* et *MIMO*. Il s'agit d'un travail nécessaire visant à concevoir d'abord un modèle basé sur les données pour prédire le canal dans tout environnement complexe dans un scénario *SISO*. Ensuite, un modèle empilé (*Stacked model* ou bien *SM*) est conçu pour le canal *WBAN-MIMO* afin de prédire, simultanément, le sous-

canal de la matrice *H*, la position de l'endroit où les données sont collectées (classificateur de position) et s'il s'agit d'une situation *LoS* ou *NLoS* (détection LoS-NLoS).

CHAPTER 1   INTRODUCTION

The propagation channel characterization and modeling are of great importance for wireless communication design, analyzing the communication quality and simulating the network performance. In the literature [1], the propagation channel is defined as the medium where the electromagnetic travel between antennas, and are subject to various physical phenomena that affect the transmission quality. In indoor environment, these effects known as propagation mechanisms. They are categorized by reflections, diffractions, and scattering, caused by the nature of the surfaces and the density of the obstacles within the site specific. Therefore, it causes interferences and distortions of the information within the channel. Several approaches have been used, leveraging the resources from propagation networks, e.g., Single-Input-Single -Output (SISO), Multiple-Input-Multiple-Output (MIMO) techniques [2-6]. In other studies, deterministic algorithms are used to describe the signal behaviour at a given environment and provides an appropriate coverage criterion in that environment. The determinist models are based on electromagnetic wave propagation theory [7,8]. Their application necessitates a thorough understanding of the propagation environment and allows precise and accurate signal propagation predictions in the channel corresponding to the considered environment [7,9]. Ray Tracing (RT) is used in the design of propagation environments and wireless networks [8,10-13]. It is among the most deterministic and efficient techniques. In fact, the RT technique is complex, it allows calculating the received power and to perform statistical operations on the received signal strength (RSS). The RT models are designed with a different approach depending on the surface roughness of the environment. Several research reported [9,14-16] on channel propagation modeling and characterization inside confined environment with smooth surfaces. From an electromagnetic perspective, there are no smooth surfaces. The surface roughness depends on the relationship between the standard deviation of the roughness high and the wavelength [17-19]. Indoor mm-wave channel modeling, on the other hand, has been extensively researched in the literature to forecast the propagation characteristics in diverse places. The majority of known research publications imply that deterministic modeling of mm-waves channels is problematic. In the literature and standardization domain, Saleh-Valenzuela [20] and Shoji [21] models are among the most often used models. In fact, the channel measurement may be performed to model a specific site, depending on whether the measured signal is narrowband or wideband [1,22,23]. Thus,

narrowband measurements often correlated with flat fading frequency, where the wideband measurements are specifically used to describe the multipath fading phenomena [1]. In this area, the Channel Impulse Response (CIR) is generally measured either in the time domain [1,24] or obtained from using Inverse Fourier (IFFT) of the measured Channel Frequency Response (CFR) for large bandwidth $[\,f_0 - f_n]$ [5,25]. On the other hand, Continuous Wave (CW) channel measurements are used to validate experimentally deterministic models. Generally, the CW measurements are performed by moving continuously the receiver from the transmitter to assess the power decay, and by the effect of propagation mechanisms such as reflections, scattering, and diffractions. Therefore, the collected data aim to validate some analytical models experimentally including ray tracing and edge diffractions model [8,9,26-28].

In the past decades, the research community has multiplied efforts in studying wireless channel propagation in indoor environment, offices, buildings, and underground mines [5,6,29-34]. Moreover, the body effect is also considered in several studies as an important parameter [35]. Particularly, for future communications systems in underground mine environment. As results, Wireless Body Area Network (WBAN) channels have been studied in underground mine environments to understand the propagation mechanism resulting by the body effect [6,29-32,35] within SISO and MIMO scenarios, considering LoS and NLoS situations. However, a growing need to enhance mining conditions by implementing effective underground communication networks prompted various investigations in underground mine sites. The study's aim to replace the wired communications, which is costly to maintain and has limited scalability [36], by a reliable and safe wireless link. These studies were developed in the purpose to develop a better understanding and bring solutions for the underground mine complex environment. This is due to the roughness and the randomness of the underground mine surfaces. The researchers worked under different antenna systems configurations such as MIMO, SISO [5,20,23,37-42]. Line of Sight (LoS) and Non-LoS scenarios [6,29-32,35] were used to stimulate the real behaviour of the narrowband or broadband signal within any situation. Although, these characterizations provide a good understanding of the wireless channel, it remains limited to only one environment or a part of the underground mine environment. This later, makes it difficult to apply or conclude same results in different environments. As results, these propagation channel models will not be sufficient to design the wireless system within all the underground mine environments due to the

fast-growing networks and facilities for wireless propagation channel which carry higher standards for high spectral efficiency.

Nowadays, the Artificial Intelligence (AI) is emerging in our daily life. It's interesting that we can't watch a movie or a TV show without an AI robot learning the pattern of our daily routine and then offering suggestions the next day. However, the back end of these applications is based on algorithms, namely Machine Learning algorithms (ML) which were used in several applications, such as facial recognition, object detection, etc. As a matter of fact, these applications collect more data in order to perform and effectively predict the choice of the movie that we want to watch. Moreover, two types of prediction problems exist, classification problems where the output target is a discrete class labels and regression problem where the target is continuous quantity. Therefore, as we also live-in time of the Internet of Objects technology (IOTs). However, the dataflow through IoTs makes the AI field is very helpful to bring a new solution in daily life. This creates massive new technologies where several objects have to be connected at the same time between them and also with the host server. Consequently, the need of channel capacity and coverage becomes crucial for future wireless communications systems in indoor environments. Due to the amount of the dataflow, the use of the Deep learning (DL) algorithms are helpful to the capability to create a model that can learn and predict the wireless communications channels.

This PhD thesis studies the application of DL algorithms to model the indoor channel within corridors and underground environments in SISO and MIMO configurations. It is a necessary work aimed to design for first a data-driven model to predict the channel within any complex environment in SISO scenarios. Then, a Stacked Model (SM) is designed for WBAN-MIMO channel to predict, simultaneously, the subchannel of the matrix H, the position of where the data is collected (position classifier) and whether it is in LoS or NLoS situation (LoS-NLoS detection).

1.1 Motivation

The channel characterization based both on theoretical approaches and real measurements are going to become critical in the future. Due to the availability and removal of equipment from one site to another, experimental measurements are usually difficult to perform. The simulation method will not be sufficient to deal with the reality of the future communication system caused by the difficulty to simulate and model the channel within any complex environments and different

frequency range. In particular, the channel characterisation depends on highly experimental site-specific measurements. As results of these measurements campaign, numerous amounts of the data are collected, where different complex environments were used under different antenna configuration (MIMO - SISO) and LOS-NLOS scenarios. These collected data can be used to develop a suitable technique which improves the design of the wireless channel propagation system within any complex environments in different scenarios. As literature point of view addresses, the propagation channel is considered as a non-linear problem [1,25,43]. Therefore, to design wireless communications systems, more experimental data is needed and continuously looking forward to improving the transmission link by resolving a non-linear problem. As results, DL algorithms can be very useful to explore the huge amount of the collected data. It is observed that it is suitable to non-linear problems, to bring a data driven model as a solution for channel characterization and model in any complex environment. Moreover, DL algorithms can also be used to enhance the radio-localization in any environments to bring more safety for miners in case of underground mine hazards [6].

1.2   Research Problems and Objectives

Freshly applied to channel propagation field, the Deep Learning applications were seen in different perspectives [44-57]. However, it is always a problem when it comes to understanding the way to develop a model which can be helpful for now and for future communications. Even though, researchers made it clear that there is a large possibility to create a model which can reproduce and predict same CFR and CIR, with great prediction accuracy, many uncertainties are questioned which bring doubt on the performance of the proposed models in the literature. The uncertainties are that these modeling techniques tend to update once new data are available, the input data processing is not well demonstrated, and the training process used in all environments. Therefore, it increases the complexity of the DL algorithms in terms of the computational simulation and the loss on the generalization efficiency. However, the need of a data driven model which can predict with high accuracy the channel in any complex environment become crucial. This can be designed to predict any new experimental data collected in new environment. Moreover, as *Moulay, et al* [35] have specified, a wireless system needs to provide a reliable service to ensure a secure environment work in underground mine, e.g., the localization of the miners for prompt intervention in any emergency event, as well as collision avoidance systems that safeguard the miners from

passing mining machinery. To overcome, these limitations, a DL algorithm combined with strong data processing algorithms for the input data is designed specifically to be able to predict in SISO and MIMO scenarios, also considering the localization problem, which is crucial in underground mine environment.

The implementation of the DL algorithms requires details and strong knowledge of the channel characterization and modeling. As results, the DL model establishment allows reducing the cost to build an efficient wireless systems, by its accuracy to predict the channel. Specifically in underground mine environment, where the mine geometry is characterized by the randomness. In fact, deterministic approach requires high complexity computations to implement it and a stochastic model is not really reliable enough to provide more accuracy in that environment.

1.3   Thesis Contributions

To the best of the authors "knowledge," no such algorithm was applied to model the channel frequency response within any environment. The novelty of this thesis consists on SISO and MIMO Channel prediction models. The main contributions are

1.3.1   In the case of SISO channel systems:

In this case, the SISO channel is proposed in chapter 4. The novelty of this study consists of:

First, a novel CFR model is proposed for different environments such as smooth, rough corridors and underground mine environments. The Line-of-Sight Single Input Single Output (LOS-SISO) scenario has been carried out for different antenna configurations and frequency bands. Second, a learning framework is designed based on LSTM combined with linear networks. The learning curve is performed in only one environment, which is crucial in the deep learning field to demonstrate the ability of the model learned from the collected CFR. The model is evaluated for CFR prediction in different environments, considering different antenna radiation patterns. Hence, three different antenna combinations were used:

1) Directional transmitter versus Directional receiver antennas (D vs. D).

2) Directional transmitter versus Patch receiver antennas (D vs. Prx) for the corridors environments.

3) Patch antennas used in the underground mine environment (P vs. P).

1.3.2 In the case of WBAN-MIMO channel systems:

In this case, the WBAM-MIMO channel is proposed in chapter 5. A stacked model (SM) based on LSTMs deep learning networks is designed. The model is validated with extensive WBAN-MIMO channel measurements published by *Moulay et al* [35]. The SM intent to bring a new concept for the wireless communication system in underground mines environments. Moreover, the SM composed of three parallel models, where only one input data is used to predict simultaneously the transmitter position, LOS-NLOS scenario and the channel matrix (H). The collected magnitude and phase of the Channel matrix were used in the SM to empower the model prediction in terms of accuracy. Therefore, the SM aim to predict the channel magnitude and phase for the channel matrix (H), simultaneously. Then, it uses this information to predict the positions of the received H and whether it is in LOS or NLOS scenario.

1.4 Thesis Structure

Six chapters contribute to this study. In the first chapter, the introduction, the motivation of the work, the problems and the engineering contributions of this thesis are outlined. The following chapters deal with the following topics in detail:

Chapter 2 explains the channel characterization and sounding

Chapter 3 provides the applied Deep Learning algorithms in times series data

Chapter 4 presents the channel prediction model for SISO systems within LOS scenarios. The channel characterisation was made to compare the measured and predicted channel frequency response within rough, smooth indoor and underground mine environments using different antenna configurations.

Chapter 5 presents the channel prediction model for WBAN -MIMO 2x2 systems within different antenna setups. SM framework is designed to predict the channel matrix H, the position and LOS-NLOS identification. Considering different antenna's setups, such as linearly polarized (Lin) patch antennas MIMO system, circularly polarized patch antennas MIMO system, co-polarized (CP) and 90 degrees rotated (90 deg) configurations.

Chapter 6 presents conclusions and discusses the open research problems

1.5   Publications

1. K. Khaled and L. Talbi, "Case study of radio coverage in complex indoor environments for 5G communications," 2019 IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE), 2019, pp. 105-110, doi: 10.1109/WiSEE.2019.8920388.

2. Kedjar, K., Talbi, L., Nedil, M.: Efficient indoor propagation channel prediction based on deep learning approach. IET Microw. Antennas Propag. 1– 17 (2021). https://doi.org/10.1049/mia2.12183

3. K.Kedjar, M. ELazhari L.Talbi and M.Nedil "Deep Learning Modeling of a WBAN-MIMO Channel in Underground Mine" to be submitted in IEEE Access journal

1.6   References

[1]   T. S. Rappaport, *Wireless communications: principles and practice*. prentice hall PTR New Jersey, 1996.

[2]   K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Second order statistics of NLOS indoor MIMO channels based on 5.2 GHz measurements," in *GLOBECOM'01. IEEE Global Telecommunications Conference (Cat. No. 01CH37270)*, 2001, vol. 1: IEEE, pp. 156-160.

[3]   R. S. Blum, "MIMO with limited feedback of channel state information," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).* 2003, vol. 4: IEEE, pp. IV-89.

[4]   H. K. Bizaki, *MIMO systems: theory and applications*. BoD–Books on Demand, 2011.

[5]   I. B. Mabrouk, L. Talbi, M. Nedil, and K. Hettak, "MIMO-UWB channel characterization within an underground mine gallery," *IEEE Transactions on Antennas and Propagation,* vol. 60, no. 10, pp. 4866-4874, 2012.

[6]   M. El Hassan El Azhari, L. Talbi, L. Arabi, M. Nedil, M. L. Seddiki, and N. Kandil, "Channel Characterization of Circularly Polarized Antenna MIMO System in an Underground Mine," *Progress In Electromagnetics Research,* vol. 67, pp. 9-19, 2018.

[7]   V. A. Fono, "Étude de la propagation radio en environnement doté de surfaces irrégulières à profil périodique," Université du Québec en Outaouais, 2018.

[8]   V. A. Fono, L. Talbi, O. A. Safia, M. Nedil, and K. Hettak, "Deterministic Modeling of Indoor Stairwells Propagation Channel," *IEEE Antennas and Wireless Propagation Letters,* vol. 19, no. 2, pp. 327-331, 2019.

[9]   M. Ghaddar, I. B. Mabrouk, M. Nedil, K. Hettak, and L. Talbi, "Deterministic modeling of 5G millimeter-wave communication in an underground mine tunnel," *IEEE Access,* vol. 7, pp. 116519-116528, 2019.

[10]   H. R. Anderson, "A ray-tracing propagation model for digital broadcast systems in urban areas," *IEEE Transactions on Broadcasting,* vol. 39, no. 3, pp. 309-317, 1993.

[11]   R. P. Torres, L. Valle, and M. Domingo, "Computer tool to analyze radio channel in arbitrary enclosed spaces using ray tracing," in *VTC'98. 48th IEEE Vehicular Technology Conference. Pathway to Global Wireless Revolution (Cat. No. 98CH36151)*, 1998, vol. 1: IEEE, pp. 581-585.

[12]   K. Remley, A. Weisshaar, and H. Anderson, "Full-wave study on the accuracy of ray-tracing in multipath environments," in *RAWCON 99. 1999 IEEE Radio and Wireless Conference (Cat. No. 99EX292)*, 1999: IEEE, pp. 203-206.

[13]   P. Međeđović, M. Veletić, and Ž. Blagojević, "Wireless insite software verification via analysis and comparison of simulation and measurement results," in *2012 Proceedings of the 35th International Convention MIPRO*, 2012: IEEE, pp. 776-781.

[14]   S. Y. Lim, A. K. Awelemdy, Z. Yun, and M. F. Iskander, "Experimental study of propagation characteristics in an open-trench drain," *IEEE Antennas and Wireless Propagation Letters,* vol. 15, pp. 60-63, 2015.

[15]   A. Hrovat, G. Kandus, and T. Javornik, "A survey of radio propagation modeling for tunnels," *IEEE Communications Surveys & Tutorials,* vol. 16, no. 2, pp. 658-669, 2013.

[16]   S. Helhel, "Comparison of 900 and 1800 MHz indoor propagation deterioration," *IEEE transactions on antennas and propagation,* vol. 54, no. 12, pp. 3921-3924, 2006.

[17]   D. Didascalou and W. Wiesbeck, "Ray-optical wave propagation modeling in arbitrarily shaped tunnels," *Dissertation, Institut für Höchstfrquenztechnik und Elek-tronik, Universität Karslruhe,* 2000.

[18]   N. PINEL and C. Bourlier, "Study of asymptotic models of electromagnetic wave scattering from natural interfaces–Application to a sea covered in oil–."

[19]   L. Tsang, J. A. Kong, K.-H. Ding, and C. O. Ao, *Scattering of electromagnetic waves: numerical simulations.* John Wiley & Sons, 2004.

[20]   A. A. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE Journal on selected areas in communications,* vol. 5, no. 2, pp. 128-137, 1987.

[21]   Y. Shoji, H. Sawada, C.-S. Choi, and H. Ogawa, "A modified SV-model suitable for line-of-sight desktop usage of millimeter-wave WPAN systems," *IEEE Transactions on Antennas and Propagation,* vol. 57, no. 10, pp. 2940-2948, 2009.

[22]   K. Pahlavan, T. H. Probert, and M. E. Chase, "Trends in local wireless networks," *IEEE Communications Magazine,* vol. 33, no. 3, pp. 88-95, 1995.

[23]   H. L. Bertoni, *Radio propagation for modern wireless systems.* Pearson Education, 1999.

[24]   M. Otmani and M. Lecours, "Indoor radio impulse response measurements with polarization diversity," in *Proceedings of Vehicular Technology Conference-VTC*, 1996, vol. 1: IEEE, pp. 151-154.

[25]   X. Yin and X. Cheng, *Propagation channel characterization, parameter estimation, and modeling for wireless communications.* John Wiley & Sons, 2016.

[26] M. Celaya-Echarri, L. Azpilicueta, P. Lopez-Iturri, F. Falcone, M. G. Sanchez, and A. V. Alejos, "Validation of 3D simulation tool for radio channel modeling at 60 GHz: A meeting point for empirical and simulation-based models," *Measurement,* vol. 163, p. 108038, 2020.

[27] F. K. Schwering, E. J. Violette, and R. H. Espeland, "Millimeter-wave propagation in vegetation: Experiments and theory," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 26, no. 3, pp. 355-367, 1988.

[28] P. A. Tenerelli and C. W. Bostian, "Measurements of 28 GHz diffraction loss by building corners," in *Ninth IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (Cat. No. 98TH8361)*, 1998, vol. 3: IEEE, pp. 1166-1169.

[29] L. Arabi, M. Nedil, N. Kandil, M. L. Seddiki, M. El Azhari, and L. Talbi, "2.4 GHz radio-channel characterization of an underground mine using patch antennas," in *2017 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*, 2017: IEEE, pp. 1841-1842.

[30] M. El Hassan El Azhari, M. Nedil, I. B. Mabrouk, K. Ghanem, and L. Talbi, "Characterization of an off-body channel at 2.45 GHz in an underground mine environment," *Progress In Electromagnetics Research,* vol. 43, pp. 91-100, 2015.

[31] M. El Azhari, M. Nedil, I. B. Mabrouk, and L. Talbi, "Multipath effect on off-body channel parametters of a MIMO system using patch antennas inside a mine," in *2016 IEEE International Symposium on Antennas and Propagation (APSURSI)*, 2016: IEEE, pp. 1693-1694.

[32] M. El Azhari, M. Nedil, I. B. Mabrouk, and L. Talbi, "Path loss effect on off-body channel capacity of a MIMO system using patch antennas inside a mine," in *2016 IEEE International Symposium on Antennas and Propagation (APSURSI)*, 2016: IEEE, pp. 1697-1698.

[33] I. B. Mabrouk, L. Talbi, M. Nedil, Y. Coulibaly, and T. Denidni, "Effect of antenna directivity on performance of multiple input multiple output systems in an underground gold mine," *IET microwaves, antennas & propagation,* vol. 6, no. 5, pp. 555-561, 2012.

[34] I. B. Mabrouk, L. Talbi, and M. Nedil, "Performance evaluation of a MIMO system in underground mine gallery," *IEEE Antennas and Wireless Propagation Letters,* vol. 11, pp. 830-833, 2012.

[35] M. Elazhari, L. Talbi, and M. Nedil, "Body-to-Body Channel Characterization and Modeling Inside an Underground Mine," *IEEE Transactions on Antennas and Propagation,* 2020.

[36] A. Benzakour, S. Affes, C. Despins, and P.-M. Tardif, "Wideband measurements of channel characteristics at 2.4 and 5.8 GHz in underground mining environments," in *IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004*, 2004, vol. 5: IEEE, pp. 3595-3599.

[37] T. Rappaport and S. Seidel, "Multipath propagation models for in-building communications," in *1989 Fifth International Conference on Mobile Radio and Personal Communications*, 1989: IET, pp. 69-74.

[38] S.-C. Kim, H. L. Bertoni, and M. Stern, "Pulse propagation characteristics at 2.4 GHz inside buildings," *IEEE transactions on vehicular technology,* vol. 45, no. 3, pp. 579-592, 1996.

[39] S. R. Saunders and A. Aragï, *Antennas and propagation for wireless communication systems*. John Wiley & Sons, 2007.

[40] Y. Azar *et al.*, "28 GHz propagation measurements for outdoor cellular communications using steerable beam antennas in New York City," in *2013 IEEE international conference on communications (ICC)*, 2013: IEEE, pp. 5143-5147.

[41] X. Wu, Y. Zhang, C.-X. Wang, G. Goussetis, and M. M. Alwakeel, "28 GHz indoor channel measurements and modeling in laboratory environment using directional antennas," in *2015 9th European Conference on Antennas and Propagation (EuCAP)*, 2015: IEEE, pp. 1-5.

[42] A. Karttunen, K. Haneda, J. Järveläinen, and J. Putkonen, "Polarisation characteristics of propagation paths in indoor 70 GHz channels," in *2015 9th European Conference on Antennas and Propagation (EuCAP)*, 2015: IEEE, pp. 1-4.

[43] C. A. R. Fernandes, "Nonlinear MIMO communication systems: Channel estimation and information recovery using volterra models," Université de Nice Sophia Antipolis, 2009.

[44] H. Ngo, H. Fang, and H. Wang, "Deep Learning-based Adaptive Beamforming for mmWave Wireless Body Area Network," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*, 2020: IEEE, pp. 1-6.

[45] J. Wang, Y. Ding, S. Bian, Y. Peng, M. Liu, and G. Gui, "UL-CSI data driven deep learning for predicting DL-CSI in cellular FDD systems," *IEEE Access,* vol. 7, pp. 96105-96112, 2019.

[46] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5G wireless communications: A deep learning approach," *IEEE Transactions on Network Science and Engineering,* 2018.

[47] L. Bai *et al.*, "Predicting wireless mmWave massive MIMO channel characteristics using machine learning algorithms," *Wireless Communications and Mobile Computing,* vol. 2018, 2018.

[48] S. Li, Q. Wang, X. Liu, and J. Chen, "Low cost LSTM implementation based on stochastic computing for channel state information prediction," in *2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, 2018: IEEE, pp. 231-234.

[49] S. Navabi, C. Wang, O. Y. Bursalioglu, and H. Papadopoulos, "Predicting wireless channel features using neural networks," in *2018 IEEE international conference on communications (ICC)*, 2018: IEEE, pp. 1-6.

[50] J. Gante, G. Falcão, and L. Sousa, "Deep Learning Architectures for Accurate Millimeter Wave Positioning in 5G," *Neural Processing Letters,* vol. 51, no. 1, pp. 487-514, 2020.

[51] Y. Zhou, Y. Zhong, Z. Wei, T. Yin, and X. Chen, "An Improved Deep Learning Scheme for Solving 2D and 3D Inverse Scattering Problems," *IEEE Transactions on Antennas and Propagation,* 2020.

[52] W. Jiang and H. D. Schotten, "Deep learning for fading channel prediction," *IEEE Open Journal of the Communications Society,* vol. 1, pp. 320-332, 2020.

[53] C. Jiang, J. Shen, S. Chen, Y. Chen, D. Liu, and Y. Bo, "UWB NLOS/LOS classification using deep learning method," *IEEE Communications Letters,* vol. 24, no. 10, pp. 2226-2230, 2020.

[54] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," *arXiv preprint arXiv:1902.06435,* 2019.

[55] C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, "Indoor signal focusing with deep learning designed reconfigurable intelligent surfaces," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2019: IEEE, pp. 1-5.

[56] T. Imai, K. Kitao, and M. Inomata, "Radio propagation prediction model using convolutional neural networks by deep learning," in *2019 13th European Conference on Antennas and Propagation (EuCAP)*, 2019: IEEE, pp. 1-5.

[57] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Communications Surveys & Tutorials,* vol. 21, no. 3, pp. 2224-2287, 2019.

CHAPTER 2   RADIO WAVE PROPAGATION CHANNEL

In general, any wireless communication system includes three parts: Transmitter (Tx), Receiver (Rx), and the wireless channel in between to connect them, as shown in Figure 2-1. The wireless channel, unlike the Tx and Rx, cannot be engineered to help the devices make a better trade-off between reliability and performance. However, consistent knowledge of the wireless propagation channel is the foundation of the design and analysis of any wireless communication system. This chapter will attempt to provide a unified and conceptually simple explanation of a morass of concepts around wireless channels.
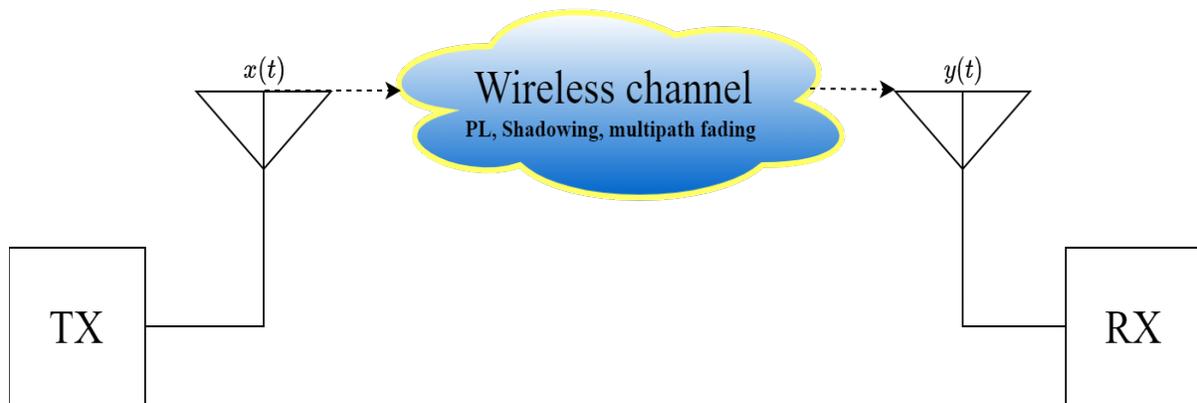
Figure 2-1 Wireless communications system

2.1 Wireless Propagation Channel Phenomena

The multiple radio channel path may be modeled as a linear time invariant impulse response h (t) filter if the channel exhibits small or no changes over time and linear time-varying h (t) if the distance d is varying with the time [3,4] . As illustrated in Figure 2-1. the received signal y(t) is the convolution of the transmitted signal x(t) and the channel impulse response $h(t, \tau)$, expressed as:

$$y(t) = x(t)*h(t, \tau)$$

(Chapter 2.1)

Where the variable t represents the time variation due to motion of the receiver, and τ represent the channel multipath delay for a fixed value of t. The received signal consists of a variety of attenuated, time delay replicas of the transmitted signal that have been shifted from phase to phase. The baseband impulse response model of a multipath mobile radio can be expressed as [1,4,5]:

$$h(t,\tau) = \sum_{n=1}^{N(t)} a_n(t,\tau)e^{-j\theta_n(t,\tau)}\,\delta\big(\tau\text{-}\tau_n(t)\big)$$

<div align="right">(Chapter 2.2)</div>

$a_n(t,\tau)$ and $\theta_n(t,\tau)$ are the amplitude and phase of the nth arriving multipath component, δ is the Kronecker delta function, and $\tau_n(t)$ is the excess delays. In fact, instead of the theoretical delta function, a small pulse is used for research. The propagation channel is described by its impulse response $h(t)$ which is the sum of time varying multipath impulse $N(t)$. t denotes the time of the various paths of the transmitted signal. Figure 2-2 shows an example of the impulse response and the frequency response for SISO-LOS channel within an indoor environment. The method acquiring to calculate the impulse response will be discussed in the channel sounding section. Due to the nature of wireless propagation, fading is inevitable in wireless channels. Fading refers to the time variation of the received signal power induced by shadowing or multipath fading. Generally speaking, fading can be categorized into large-scale fading, consisting of path loss and shadowing, and small-scale fading [1,6]. Therefore, in total we have three phenomena in wireless channels. Path loss (PL) and shadowing belong to large-scale fading category since they are dominant as the mobile station moves for several dozen wavelengths. The PL is the attenuation in the transmitted signal as it propagates from the TX to the RX and is observed over a distance, while shadowing is the slow variations obtained over distances due to large terrain features such as buildings and hills. Large scale fading is very important for a system such as cell coverage area, outages and handoffs which are influenced by these effects. In the opposite, small scale fading is a result of multipath propagation. Multipath fading refers to fast variation obtained over distance due to the constructive and destructive interference of the multiple signal paths between the Tx and Rx. These variations are observed over a distance. Small-scale fading plays an important role in determining the link-level performance according to bit error rates (the effect of inter symbol interference), average fade duration, and so on .
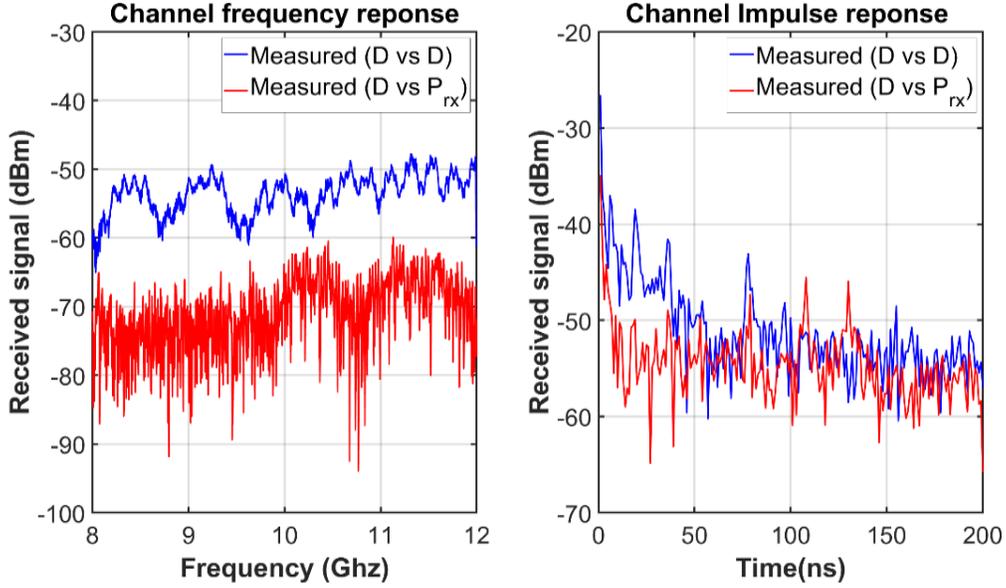
Figure 2-2 Channel impulse and frequency responses for two antenna configurations.

2.1.1 Large-scale Channel Characterization

The path loss (*PL*) is defined as the difference (in dB) between the effective transmitted power and the received power [1]. It is obtained from the path gains, by averaging over the frequencies, snapshots and the number of antennas, and is mathematically represented as follows [1,6,7]:

$$PL\big(d(z)\big) = -20 \log_{10} \frac{1}{N_s N_f} \sum_{j=1}^{N_s} \sum_{n=1}^{N_f} \left| H_j^z(n) \right| \tag{2.3}$$

Where $PL\big(d(z)\big)$ is the PL at the position of z, N_s and N_f are the number of the snapshot and frequency samples, respectively. $H_j^z(n)$ is the measured S parameters (S_{21}) for the position z [1,6-10]. Therefore, the path loss is modeled as a function of the distance between Tx and Rx as follows [1,8]

$$PL_{(d)} = PL_{db}(d_0) + 10.\beta \log_{10}\left(\frac{d}{d_0}\right) + X \tag{2.4}$$

Where $PL(d)$ is the mean path loss at the reference distance d_0, d is the distance where the path loss is calculated, β is the path loss exponent which determined by using linear regression analysis. X *(dB)* is a zero mean Gaussian variable. The model in equation (2.1), assumes the path loss is

constant over a given distance. However, the presence of obstacles within the environment leads to random variation of the received power. This effect is termed as shadowing. Several experiments showed that shadowing can be modeled as lognormal random variable [1,8]. It is known that shadows are distributed in a normal or Gaussian distribution in the log domain [11], and thus shadows can be modelled on as a log-normal distribution.

2.1.2 Small-scale Channel Characterization

Multipath fading comes as a result of small path length differences between rays coming from scatterers surrounding the transmitter and the receiver [1]. These variations contribute to major phase differences in the order of a few wavelengths (fewer than 10 wavelengths). Small movements of one of the antennas can then lead to drastic variation in the obtained signal envelope [12,13]. Their random phases and amplitudes give rise to rapid fluctuations in the received signal strength, thereby inducing small-scale fading and signal distortion. Small-scale fading is approximately superimposed on the constant large-scale fading, as seen in Figure 2-3. The small-scale fading caused by the multipath components is manifested by steep variations in the signal obtained by small intervals, by spontaneous frequency changes due to the Doppler effect on the various multipath signals, and time dispersion due to the various delays. The involvement of locals such as mountains and buildings often inhibit the line of sight (LoS) from Tx to Rx. Therefore, a non-LoS (NLoS) propagation path will appear between the Tx and Rx. The waves would then disperse by reflection, diffraction, and scattering. This refers to waves from different directions and different delays. At the receiver antenna, the waves merge to produce a composite received signal.
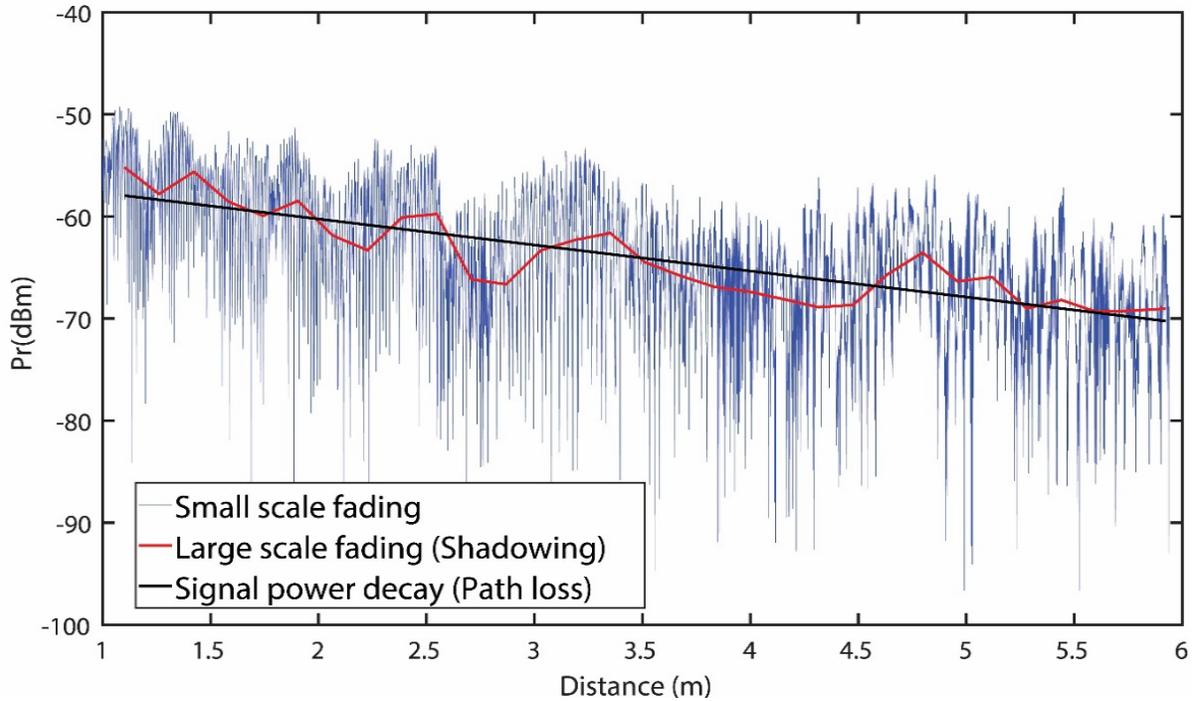
Figure 2-3 Small-scale and large-scale fading

2.1.2.1 Rician K-Factor.

The Rician k-factor describes the fading distribution of the power in presence of a dominant multipath component. The effect of a dominant signal arriving with many weaker multipath signals gives rise to the Rician distribution [1,8]. The Rician K-factor is expressed as follows [9,10]

$$K = \frac{P_D}{2P_R} \qquad (2.5)$$

Where P_D corresponds to the power of the LOS component and $2P_R$ is the power of the multipath component.

2.1.2.2 Time dispersion parameters

RMS delay is the time dispersive property of wideband multipath channel [1,8] and the coherence bandwidth is considered as the range of frequencies over which the channel can be considered flat. The RMS delay spread is equal to the square of the second central moment of Power Delay Profile (PDP) which is calculated using this formula [1,14]

$$PDP(t) = \|h(t)\|^2 \qquad (2.6)$$

16

Where $\| . \|$ and h(n) are the modulus operation and the impulse response, respectively. The signal power (P_t) of each multipath is shown against their respective propagation delays (τ)in a typical PDP plot. Figure 2-4 shows a sample PDP plot. It illustrates the received signal with various strengths as it travels across a multipath channel with various propagation delays (due to the environments) [1]. The RMS delay is calculated using the equation reported in [1]:

$$\tau_{RMS} = \sqrt{\overline{\tau^2} - \overline{\tau}^2} \tag{2.7}$$

Where $\overline{\tau^2}$represent the second moment of the PDP, $\overline{\tau}$ is the mean excess delay, and is expressed as follows [2]:

$$\overline{\tau} = \frac{\sum_k a_k^2 \, \tau_k}{\sum_k a_k^2} = \frac{\sum_k P(\tau_k) \, \tau_k}{\sum_k P(\tau_k)} \tag{2.8}$$

$P(\tau_k)$ is the relative amplitude of the Multipath components (MPCs) at the corresponding delay τ_k. The coherence bandwidth was evaluated for the 50% correlation using the RMS delay spread:

$$B_c \simeq \frac{1}{5\tau_{RMS}} \tag{2.9}$$
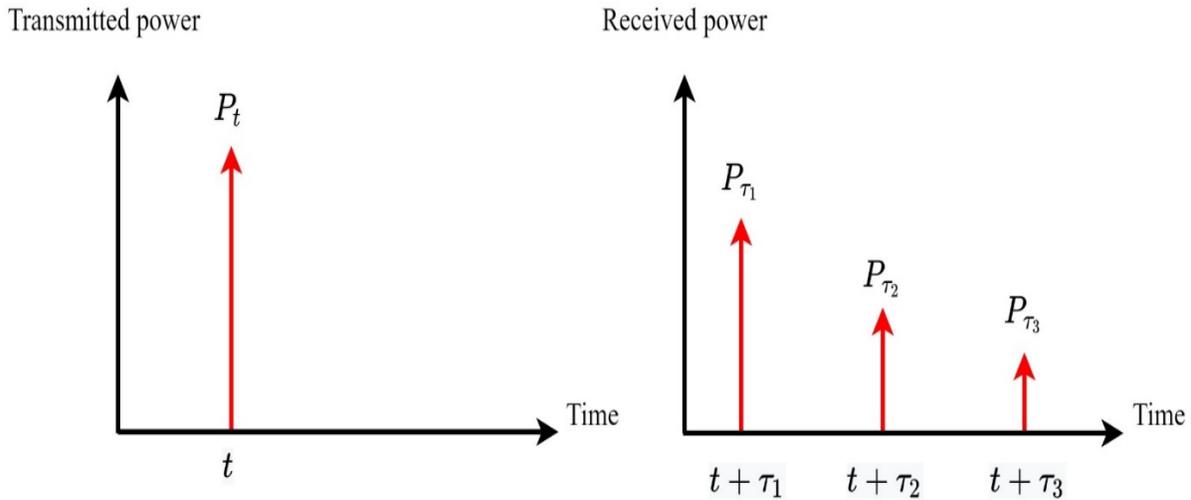
Figure 2-4 PDP for a multipath channel with 3 paths

2.2 Channel Sounding Techniques

Most research and studies on wireless communications modeling are based on the measurement campaign of the channel transmission characteristics. Channel sounding plays a crucial role in wireless engineering where the measurements of radio propagation channels provide understanding of and insights into the characteristics of radio channels in different environments. Channel investigation started at the end of the 1960s [15], as did channel sounding when researchers had to measure just the received field strength. Since then, the complexity of the systems and the required channel models has evolved considerably. New class of channel sounders has seen the light in research and industry to accommodate the transition towards wideband systems which require power delay profile measurement. In general, the impulse response of a channel can be approximated by equipment designed on the basis of one of the following two main measurement techniques in time and frequency domain.

In the time domain or the frequency domain, the properties of wideband channels can be measured to generate the time-varying channel impulses, $h(\tau,t)$ or transfer function $H(f,t)$, respectively. The results are technically similar in both fields and can be translated from one domain to the other via the Fourier transform (FT). The practicality of the steps is therefore very different.

Time domain measurements obtain the channel impulse response presented in equation (2.2) by exciting the channel with short pulses or pseudo-noise sequences then recording the received signal with a sampling oscilloscope (Figure 2-5) [16,17]. This technique is conceptually simple but in practice the generation of ultra-narrow pulses is a challenging task in terms of analog to digital converters sampling rates and speed. Moreover, the requirements of RF and analog circuits are considered to be complex.
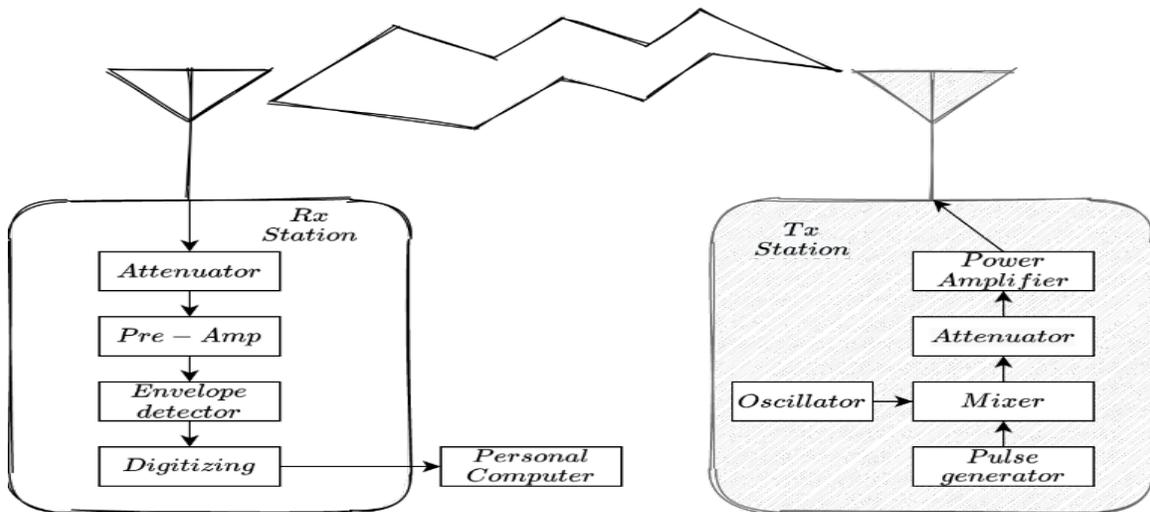
Figure 2-5 Example of pulse transmission system for wideband time-domain measurements [2]

On the other hand, frequency domain measurements usually use a chirp-like tone, or other multi-connect signals to sound the channel across a number of frequencies [18]. The vector network analyzer (VNA) makes this method easy to apply. VNAs essentially sound the channel by slowly balancing the interest level and calculating the transmission coefficient S_{21} between its two ports. The greatest downside of the VNA schemes is that, depending on the number of measurements, any measurement sweep takes a considerable amount of time [8]. Figure 2-6 presents the frequency channel domain sounder where a VNA is used to measure the channel impulse response. By stepping through individual frequencies, the sweeper scans a specific frequency band. The S-parameter test set transmits a known signal level at port 1 for each frequency and monitor the received signal at port 2. These signal levels allow the analyzer to determine the complex response S_{21} of the channel over the measured frequency. Then, the response is converted into the time domain by applying the inverse discrete Fourier transform which gives a limited channel impulse response [1,8].
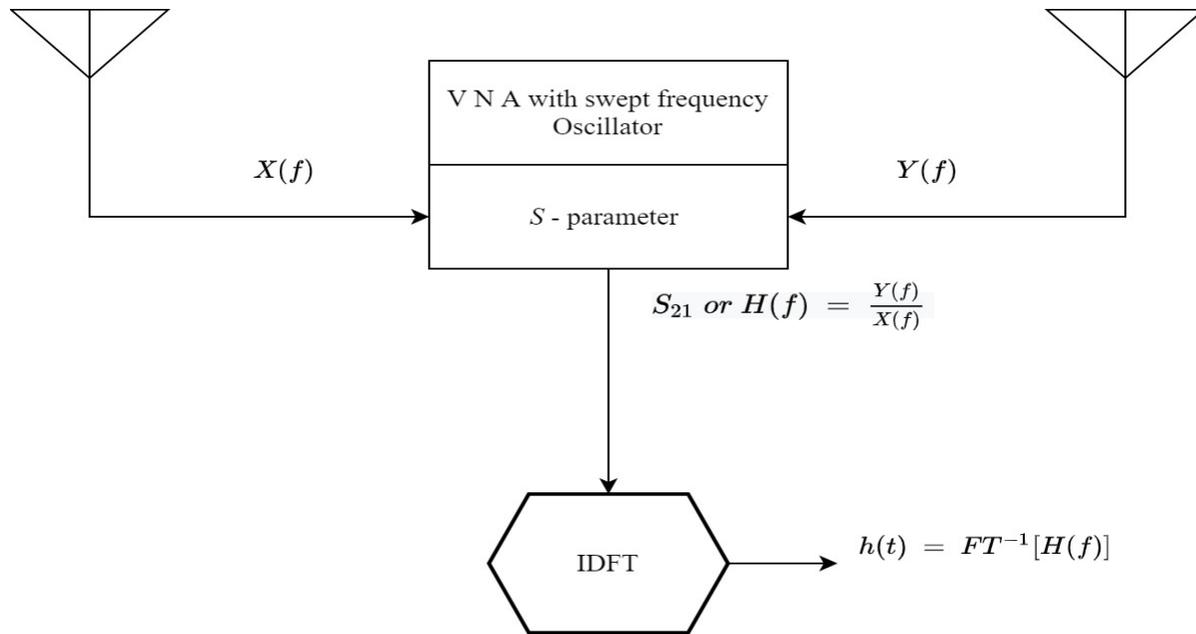
Figure 2-6 Frequency channel impulse response measurement system [1]

2.3 References

[1]   T. S. Rappaport, *Wireless communications: principles and practice*. prentice hall PTR New Jersey, 1996.

[2]   K. Pahlavan and A. H. Levesque, *Wireless information networks*. John Wiley & Sons, 2005.

[3]   J. G. Proakis, "Digital communications," *McGraw-Hill Book Company, 4th Edition,,* 2007.

[4]   M. Ghaddar, I. B. Mabrouk, M. Nedil, K. Hettak, and L. Talbi, "Deterministic modeling of 5G millimeter-wave communication in an underground mine tunnel," *IEEE Access,* vol. 7, pp. 116519-116528, 2019.

[5]   G. L. Turin, F. D. Clapp, T. L. Johnston, S. B. Fine, and D. Lavry, "A statistical model of urban multipath propagation," *IEEE Transactions on Vehicular Technology,* vol. 21, no. 1, pp. 1-9, 1972.

[6]   T. S. Rappaport, "Mobile radio propagation: Small scale fading and multipath," *Wireless Communications Principles and Practice, 2nd Ed. Prentice Hall,* 2001.

[7]   T. Rappaport and S. Seidel, "Multipath propagation models for in-building communications," in *1989 Fifth International Conference on Mobile Radio and Personal Communications*, 1989: IET, pp. 69-74.

[8] J. B. Andersen, T. S. Rappaport, and S. Yoshida, "Propagation measurements and models for wireless communications channels," *IEEE Communications Magazine,* vol. 33, no. 1, pp. 42-49, 1995.

[9] M. El Hassan El Azhari, L. Talbi, L. Arabi, M. Nedil, M. L. Seddiki, and N. Kandil, "Channel Characterization of Circularly Polarized Antenna MIMO System in an Underground Mine," *Progress In Electromagnetics Research,* vol. 67, pp. 9-19, 2018.

[10] I. B. Mabrouk, L. Talbi, M. Nedil, and K. Hettak, "MIMO-UWB channel characterization within an underground mine gallery," *IEEE Transactions on Antennas and Propagation,* vol. 60, no. 10, pp. 4866-4874, 2012.

[11] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 2002.

[12] J.-M. Molina-Garcia-Pardo, M. Lienard, A. Nasr, and P. Degauque, "Wideband analysis of large scale and small scale fading in tunnels," in *2008 8th International Conference on ITS Telecommunications*, 2008: IEEE, pp. 270-273.

[13] C. C. Chai and T. T. Tjhung, "A unified probability density function for small scale fading envelopes," in *2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, 2007: IEEE, pp. 1-5.

[14] M. Elazhari, L. Talbi, and M. Nedil, "Body-to-Body Channel Characterization and Modeling Inside an Underground Mine," *IEEE Transactions on Antennas and Propagation,* 2020.

[15] Y. Okumura, "Field strength and its variability in VHF and UHF land-mobile radio service," *Rev. Electr. Commun. Lab.,* vol. 16, pp. 825-873, 1968.

[16] D. Cassioli, M. Z. Win, and A. F. Molisch, "The ultra-wide bandwidth indoor channel: from statistical model to simulations," *IEEE Journal on selected areas in Communications,* vol. 20, no. 6, pp. 1247-1257, 2002.

[17] R.-M. Cramer, R. A. Scholtz, and M. Z. Win, "Evaluation of an ultra-wide-band propagation channel," *IEEE Transactions on Antennas and Propagation,* vol. 50, no. 5, pp. 561-570, 2002.

[18] J. Mar, Y.-R. Lin, and Y.-C. Yeh, "Ultra-wide bandwidth in-vehicle channel measurements using chirp pulse sounding signal," *IET science, measurement & technology,* vol. 3, no. 4, pp. 271-278, 2009.

CHAPTER 3   A REVIEW OF DEEP LEARNING ALGORITHMS FOR CHANNEL CHARACTERIZATION AND MODELING

As discussed previously, experimental measurements are performed to collect the data in both SISO and WBAN MIMO channels. To achieve a better understanding, some theories, and concepts of the Artificial neural networks (ANNs) are explained in detail. In research problems (chapter 1), it is mentioned that strong data processing and Deep Learning (DL) architectures are needed to develop a model to achieve high prediction accuracy and capable of providing a reliable channel information to design a wireless system.

For this purpose, it is crucial to understand the background of DL to apply it to channel characterization. The important part in this model is the measurement CFR which is considered as an input of our model in both classification and regression cases. Then, considering the CFR data nature and the field of knowledge, a proper model is selected to predict and classify the input data. However, the model is considered as a specialised learning representation from the CFRs, and the algorithm as the learning process, as noticed below.

$$\text{Model} = \text{Algorithm}(\text{CFRs}) \tag{3.1}$$

Understanding the measured CFR data and how to gather it, is the best method to pick the algorithm for the DL model. Although, the propagation channel characterization is discussed previously, which gave us great understanding of the channel, it will not be enough for applying the DL. During the measurement campaign up to four thousand points were taken at each position for each measurement. These measurements were taken during the channel measurement for SISO and WBAN-MIMO cases. The channel measurements were done with the VNA by a frequency scan from f_0 to f_n, which is made within a temporal interval in each position. However, frequency time sweep known as sweep time is configured in to carry out a frequency scan. The time characteristic of the channel measurement made the data look like a sequence, where several channel frequency records were taken sequentially at each position. As time series data, the collected data were used in different architecture with Recurrent Neural Network (RNN), such as Long Short-Term memory (LSTM) and Gated Recurrent Unit (GRU). They are used for channel state information which considers the channel response as time series data [2-8]. Using this

concept, the choice of the type of the DL architecture is narrowed. Two primary DL topics, classification and regression problems, are explored to fulfil the goal stated in the introduction. Hence, this chapter is organized as follows: First, some specialized models for sequence processing are discussed. Second, the sequence modeling using LSTM and Encoder-decoder architectures are discussed. Finally, a classification and regression metrics are presented.

3.1 Neural Network for Sequence Modeling

ANNs are, as the name implies, artificial networks that seek to grossly replicate the decision mechanism in the biological (human or animal) central nervous system's nerve cell (neuron) networks, which is a highly parallel process. Over the last twenty years or so [1], the use of ANNs has expanded in a wide variety of industrial and service fields. ANNs are a robust, practical, and reliable method for estimating a target function. The ANN parallelism allows the use of highly parallel computer systems to train and evaluate ANNs. The backpropagation learning algorithm makes the model robust against errors and noise in the training data. Since there are various of artificial neural network structures for time series prediction, it is necessary to summarize their characteristics to better understand how the ANN algorithm operates and achieves high accuracy prediction.

3.1.1 Structure of Artificial Neural Networks

McCulloch and Pitts in 1943, implemented an artificial neural network system, which is analogous to biological neurons based on a complex structure Table1 . The ANNs system is thought to be a collection of identical structural elements called neurons that are linked together in a manner similar to the human nervous system's cells. It is also composed of a succession of layers connected in such a way that each neuron holds its input from the output of the preceding neuron. In this case, each neuron acts independently of the others so that the entire neurons form a compact structure. The information is stored in a distributed way in the network in form of synaptic coefficients. As shown in Figure 3-1, each neuron receives an input in vector form and then calculates a weighted sum of its inputs, as a result of which, the outcome is passed through the activation n function to give the prediction output.

Table 3-1 Artificial versus biological nomination

Artificial neuron	Biological neuron
Connection weight	Synapses
Output signal	Axons
Input signal	Dendrites
Activation function	Soma
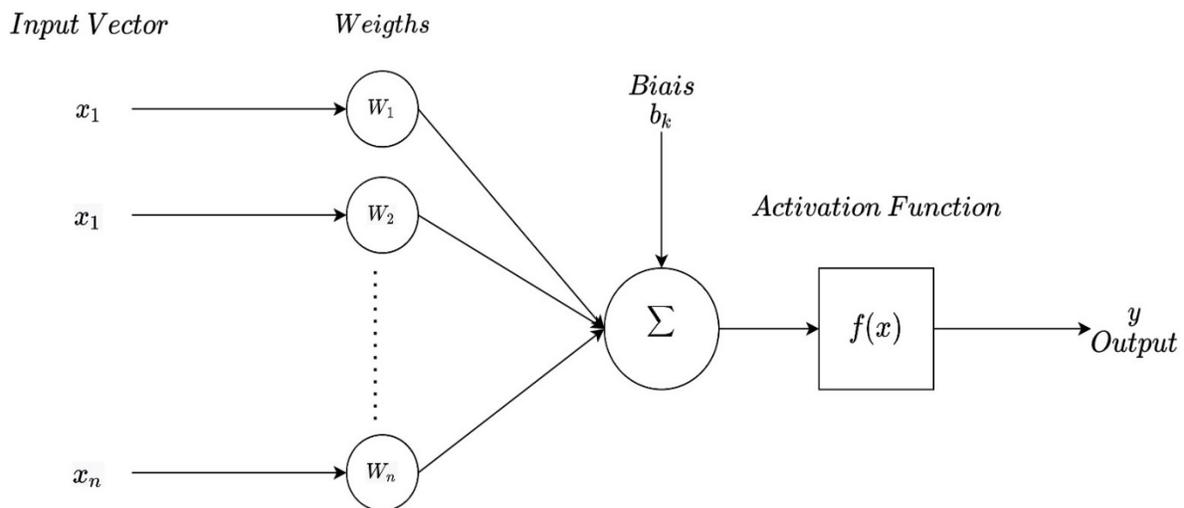
Figure 3-1 Perceptron's schematic input/output structure [1]

3.1.1.1 Activation function

An activation function is used to transform the product of the weighted sum into the output value. This conversion is carried out by measuring the state of the neuron by adding non-linearity in the action of the neuron. The bias b_k acts a threshold role, when the result of the weighted sum exceeds this threshold, the argument of the transfer function becomes positive or null; otherwise, it is considered negative. Finally, if the result of the weighted sum is

1- below the threshold, the neuron is considered non-active,

2- around the threshold, the neuron is in a transition phase, and

3- above the threshold, the neuron is considered active.

There are several types of transfer functions that can be used in ANNs. Frequently used activation functions are shown below:

3.1.1.1.1 The nonlinear sigmoid function

The non-linear sigmoid function is frequently used in ANNs, particularly in networks using the backpropagation algorithm [9]. Unlike the sigmoid function, the other functions give only a binary output which makes it more difficult to estimate optimal weights. The sigmoid function is defined by :

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.2}$$
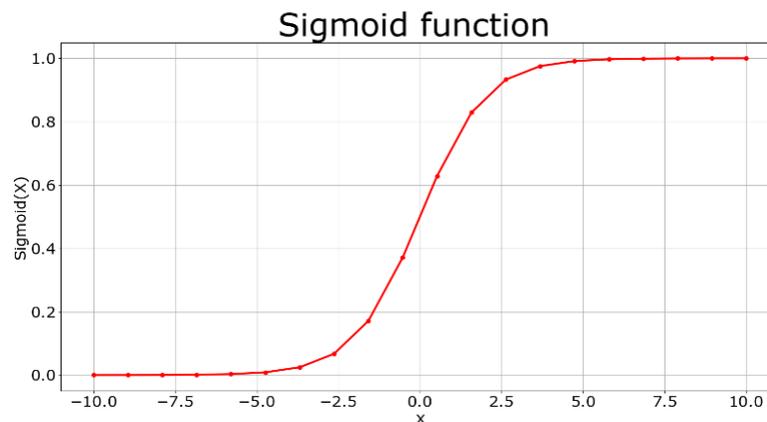
Figure 3-2 The sigmoid function

3.1.1.1.2 SoftMax function

The SoftMax function [10] takes a vector of K real values and reduces it to a vector of K real values that add up to one. The SoftMax turns the input values, which might be positive, negative, zero, or higher than one, into values between 0 and 1, allowing them to be understood as probabilities. If one of the inputs is tiny or negative, the SoftMax converts it to a small probability; if one of the inputs is high, it becomes a great probability; nonetheless, it will always be between 0 and 1. Soft-argmax, or multi-class logistic regression, is another name for the SoftMax function. This is because the SoftMax is a multi-class classification generalization of logistic regression,

and its formula is extremely similar to the sigmoid function used in logistic regression. The SoftMax formula is as follows [10-14]:

$$\sigma(x_i) = \frac{e^{x_i}}{\sum_j^k e^{x_j}} \qquad j = 0,1,....k \tag{3.3}$$

Where x_i represents the input vector element, and k is the number of classes in the multi-class classifier.

3.1.1.1.3 *Hyperbolic Tangent function (Tanh):*

The Tanh functionality is yet another potential function that can be used across layers, of a neural network as a nonlinear activation function [15]. In fact, it shares a few things in common with the role of sigmoid activation. Both seem quite equivalent. However, while the sigmoid function maps input values between 0 and 1, Tanh maps values between -1 and 1. The Tanh function is described as follows.

$$\sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3.4}$$
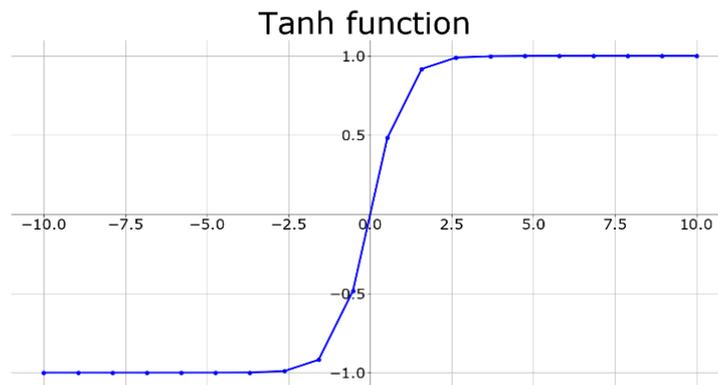
Figure 3-3 Tanh activation function

3.1.1.1.4 *Rectified Linear Units*

The Rectified Linear Unit (*ReLu*) [16] activation function has been the most widely used activation function for ANNs. Compared with the sigmoid activation function, it generally achieves better performance and generalizations in ANNs. The key principal behind the ReLu activation function is to perform a threshold operation on each input variable where values below zero are set to zero.

26

$$\sigma(x) = \max(0, x) = \begin{cases} x_i & \text{if } x_i > 0 \\ 0 & \text{if } x_i < 0 \end{cases} \tag{3.5}$$
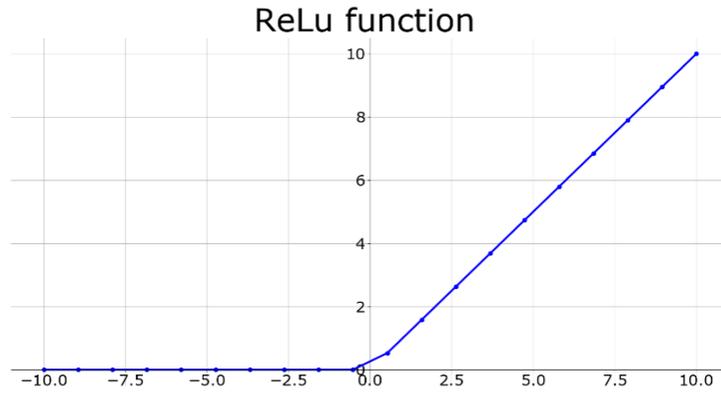
Figure 3-4 ReLu activation function

3.1.2 Single perceptron

The best known and most primitive type of ANNs is called a perceptron [1,17,18]. As illustrated in Figure 3-1, it takes an input vector values $X = (x_1, x_2, \ldots, x_n,) \in R^n$ and it has internal weights $W = (w_0, w_1, w_2, \ldots, w_n,) \in R^{n+1}$ and the neuron threshold known as the bias. The perceptron output function is a linear function that returns 1 for values above a certain threshold and 0 otherwise [17,18] :

$$y(x) = \begin{cases} 1, & w_0 + w_1 x_1 + \cdots + w_n x_n + b > 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.6}$$

Where b is considered as a threshold of each neuron. Defining $x_0 = 1$ and $X = (x_0, x_1, x_2, \ldots, x_n,)$, the output function can be simplified to [1,17,18]:

$$y(X) = g(W \cdot X) = g\left(\sum_{i=0}^{n} w_i x_i\right) \tag{3.7}$$

Where $g(\cdot)$ is the activation function. Figure 3-1 shows a graphical representation of a perceptron. The contribution of the input x_i to output y is represented by each weight w_i. These weights

determine how effectively the perceptron represents the target function. The hypothesis space is defined as the area that contains all potential output functions. The hypothesis space of a single perceptron contains only linear functions and is therefore quite limited. A classical example for the restriction of the perceptron is the XOR function $x_1 \oplus x_2$, which can't be separated by a linear function.

To be able to use backpropagation to train multiple perceptrons, the output function y of a perceptron needs to be differentiable. Therefore, y needs to be continuous. When redefining the output function as [1,17,18]:

$$y(X) = \sigma \left(\sum_{i=0}^{n} w_i x_i \right) \tag{3.8}$$

Different activation functions mentioned in the previous section, such as sigmoid, and hyperbolic Tangent functions can be used. Only a linear output function can be produced by multiple layers of neurons with a linear activation function. A highly nonlinear function is frequently required to provide a good approximation of a target function. With a nonlinear activation function, a multilayer perceptron can produce a nonlinear output function. Usually, the sigmoid function σ is used, since its derivative can be computed easily as [1,17,18]:

$$\frac{d\sigma(y)}{dy} = \sigma(y) \cdot \left(1 - \sigma(y)\right) \tag{3.9}$$

Where the function is nonlinear. Gradient descent can be used to train the weights of a single perceptron. It is an optimization algorithm used in order to find a local minimum of a function by taking the gradient at the current position and shifting the weights in the opposite direction, until a minimum is found. The function is not convex in general, and only a local minimum is discovered. To use gradient descent, you will need a loss function that represents the quality of the forecast for a given set of weights w. Different loss function can be used. It depends if the prediction is the classification and regression problem. Usually, squared error over all predictions for the training data D is used. let y_d be the label of the input vector $d = (x_0, \dots x_n) \in D$, and $\hat{y}_d$ be the predicted label such as [1,17,18]:

$$\hat{y}_d = \sigma\left(\sum_{i=0}^{n} w_i x_i\right) \tag{3.10}$$

The squared error loss function is then defined as [1,17,18]:

$$f_{loss}(W) = \frac{1}{2}\sum_{d\in D}(y_d - \hat{y}_d)^2 \tag{3.11}$$

The gradient of the loss function f_{loss} consist of the partial derivatives with respect to the weights w_i :

$$\nabla f_{loss}(W) = \left(\frac{\partial f_{loss}}{\partial w_0}, \frac{\partial f_{loss}}{\partial w_1}, \dots \frac{\partial f_{loss}}{\partial w_n}\right)^T \tag{3.12}$$

Each partial derivative can be calculated by applying the chain rule

$$\frac{\partial f_{loss}}{\partial w_i} = \frac{\partial}{\partial w_i}\left(\frac{1}{2}\sum_{d\in D}(y_d - \hat{y}_d)^2\right) \tag{3.13}$$

$$\frac{\partial f_{loss}}{\partial w_i} = \sum_{d\in D}(y_d - \hat{y}_d)\frac{\partial}{\partial w_i}(y_d - \hat{y}_d) \tag{3.14}$$

$$\frac{\partial f_{loss}}{\partial w_i} = \sum_{d\in D}(\hat{y}_d - y_d)\cdot\frac{\partial \hat{y}_d}{\partial w_i} \tag{3.15}$$

Therefore, the weights can be updated using the derivatives by the following rule

$$\Delta w_i = -\alpha \sum_{d\in D}(\hat{y}_d - y_d)\cdot\frac{\partial \hat{y}_d}{\partial w_i} \tag{3.16}$$

$$w'_i = w_i + \Delta w_i \tag{3.17}$$

Where α is a learning rate that influences the size of the gradient descent steps. Moreover, if α is too small the algorithm will be very slow to converge towards a local minimum. On the other hand, if it is too large, the local minimum can be overstepped. As it will be discussed, in chapter 4 and 5, the dynamic learning rate can be used in order to overcome this problem.

3.1.3 Multi-Layer Perceptron (MLP)

A Multi-Layer perceptron (MLP) contains multiple layers, including an input vector layer X, an output layer y and at least one hidden layer of nodes. A hidden layer h_i consists of a number of neurons, with each neuron's input being the output of all nodes in the network in layer h_{i-1}. Contrary to a single perceptron, a MLP with an activation function (e.g., sigmoid) can approximate any continuous function $f: R \rightarrow R$ with just one hidden layer [18,19]. However, there is no assurance that the training process will find the appropriate parameters to describe the function [1]. Moreover, increasing the number of hidden layers improves generalizations with regard to the number of parameters [1,20].

The graphical illustration of a MLP with one hidden layer is shown in Figure 3-5. In this case, forward and backward propagation can be used for MLP evaluation and training. The forward and backward propagation are explained using the following notation.

x_i = Input of node i

y_i = Output function of node i

O_i = Output of node i

u_i = The set of nodes that are in layer closer to input x_i (in the layer above the node i)

d_i = the set of nodes that are closer to the output y_o (in the layer below the node i)

w_{ij} = the internal weight of node j, which is applied to the output of node i
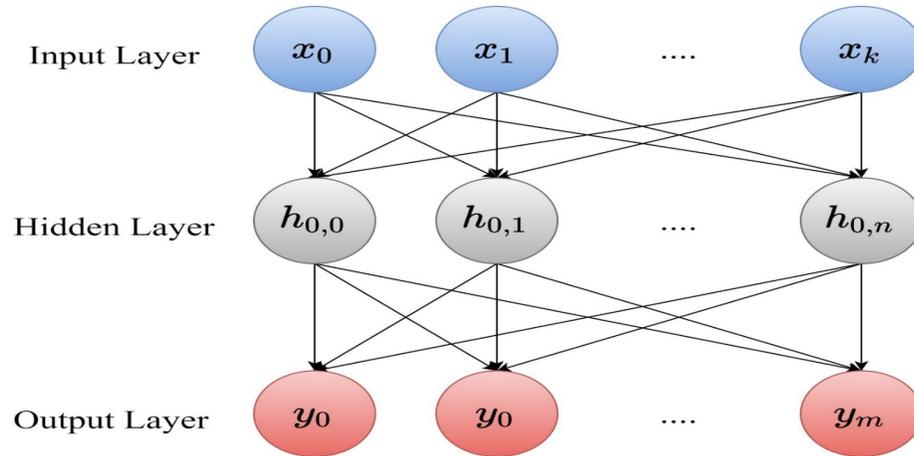
Figure 3-5 MLP with one hidden layer. the input layer composed with k nodes, the hidden layer of n nodes, and the output layer of m nodes

3.1.3.1 Feed Forward Propagation

Forward propagation is used to calculate the output of the MLP. To explain the process, let us consider a MLP with one hidden layer which the set of the hidden node is represented by H, L is the set of input nodes and v_l is the set of the input value where $l \in$ L and O is the set of the output node. The output of each node $n \in$ O is calculated by forward propagation as [17,18]

$$o_n(X) = \sigma\left(\sum_{j \in H} w_{jn} \cdot \sigma\left(\sum_{i \in L} w_{ij} v_i\right)\right) \tag{3.18}$$

Where σ is the activation function e.g., sigmoid function.

3.1.3.2 Backward Propagation

The backward propagation uses same approach as the gradient descent to train the weights of the MLP. A loss function is computed for each data point, and the weights are updated in the opposite direction as the loss function's derivative. This phase is continued until a certain condition is fulfilled such as a minimum loss in predictions. Multiple output y units are possible with the MLP, so the loss function needs to be modified [19]

$$f_{loss_v}(W) = \frac{1}{2}\sum_{i\in L}(y_{xk}\text{-}o_k(x))^2 \qquad (3.19)$$

Therefore, by applying the gradient

$$\frac{\partial f_{loss}}{\partial w_{ij}} = \frac{\partial f_{loss}}{\partial x_j}\frac{\partial x_j}{\partial w_{ij}} \qquad (3.20)$$

$$= \delta_j \frac{\partial}{\partial w_{ij}}\sum_{k\in d(j)}w_{kj}\cdot o_k \qquad (3.21)$$

$$= \delta_j \cdot out_i \qquad (3.22)$$

Therefore, the calculation of δ_j depend on the layer that the node j belongs to. However, if j belongs to the hidden units, the δ_j can be calculated by using the partial derivatives and the results will be

$$\delta_j = \sum_{k\in u(j)}\delta_k\, w_{jk}\cdot o'_j(x_j) \qquad (3.23)$$

And if j belongs to output unit

$$\delta_j = \left(o_j(x_j)\text{-}y_j\right)\cdot o'_j(x_j) \qquad (3.24)$$

Same as the gradient descent, the weight can be updated by the following rule.

$$\Delta w_{ij} = \text{-}\alpha\,\delta_j\,.o_i \qquad (3.25)$$

$$w'_{ij} = w_{ij} + \Delta w_{ij} \qquad (3.26)$$

Because the search space contains numerous local minima, the method will most likely converge to a local minimum rather than the global minimum. Despite this, backpropagation is an excellent technique for approximating functions in practice [1,20,21].

3.2 Limitations of Multilayer Perceptrons

As discussed previously, the MLPs approximate the mapping function from input variables to output variables. As this work is focused on time series data, this capability can also be applied for time series prediction (also called sequences problems) for various reasons. First, the MLPs, as any ANNs, are robust to noise in input data and in the mapping function, also can support missing values during the prediction. Furthermore, MLPs are capable of learning both linear and nonlinear relationships. [1,20-22]. As mention in [23], the elegant ability of neural networks to approximate arbitrary nonlinear functions is one of their most valuable contributions. This feature is extremely valuable in time series processing and opens the door to more sophisticated applications, particularly in the forecasting subfield.

For sequence prediction problems, it was difficult for the MLPs. The MLPs architecture has these limitations [24] such as it focuses on complete data and linear relationship. Moreover, for sequence predictions, it is necessary to diagnose and specify the link between observations made at different periods, as well as the number of lag observations used as input which is challenging for the MLPs

Although MLPs are a good architecture to start modeling sequence problems (for very short sequences), there are better options nowadays, such as the Long Short-Term Memory (LSTM) network.

3.3 Long Short-Term Memory Network

Long Short-Term Memory (LSTM) network is introduced by Hochreiter & Schmidhuber [25]. It is a type of Recurrent Neural Network. They are special types of neural network designed for sequence prediction problems. Recurrent neural networks have cycles that feed prior time step network activation as inputs to the network to impact predictions at the current time step. This activation is kept in the network's internal states, which may contain long-term temporal related information in theory. RNNs can use this method to take advantage of a dynamically shifting window over time [26]. The LSTMs, like RNNs, feature recurrent connections, which allow the state of the neuron from prior activation in the preceding time step to be used as background for forming an output. Unlike other RNNs, the LSTM has a unique formulation that allows it to sidestep the issues that prohibit other RNNs from being trained and scaled. In the past, one of the

key technological challenges with RNNs has been how to properly train them. Experiments have proven the training's intricacy, with weight changes that were either so little that they had no effect or so enormous that they caused significant alterations or even overflow. These problems are referred to *vanishing problem* and *exploding gradients,* respectively [27,28]. By its design, LSTMs overcome these problems for sequence prediction.

As an RNN has the form of a successive repeating unit (Figure 3-6), the LSTMs have the same chain as structure, but the repeating unit contains different structures. Instead of one neural network layer, there are four layers, each interacting in a unique way Figure 3-7 . The computational unit of the LSTM network is called *memory cell*. The term "neuron" has become so ingrained in discussions about MLPs as a computing unit that it is now routinely used to refer to the LSTM memory cell as well. The LSTM architecture was inspired by a study of error flow in current RNNs, which revealed that previous designs could not handle long time delays because back propagated error either explodes or decays exponentially. An LSTM layer is made up of memory blocks, which are recurrently linked blocks. These blocks can be thought of as a programmable equivalent of a digital computer's memory chips. Each one has one or more recurrently connected memory cells, as well as three multiplicative units (input, output, and forget gates) that enable continuous analogues to write, read, and reset operations for the cells. A memory cell has weight parameters for the input, output, as well as an internal state that is built up through exposure to input time steps.
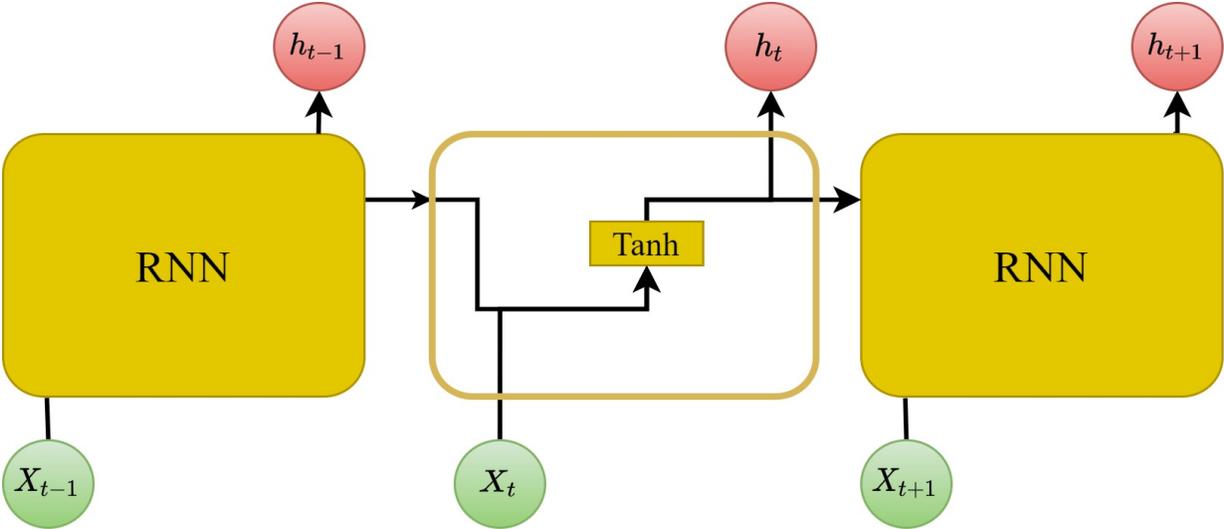
Figure 3-6 RNN unit which contain only one layer [2].
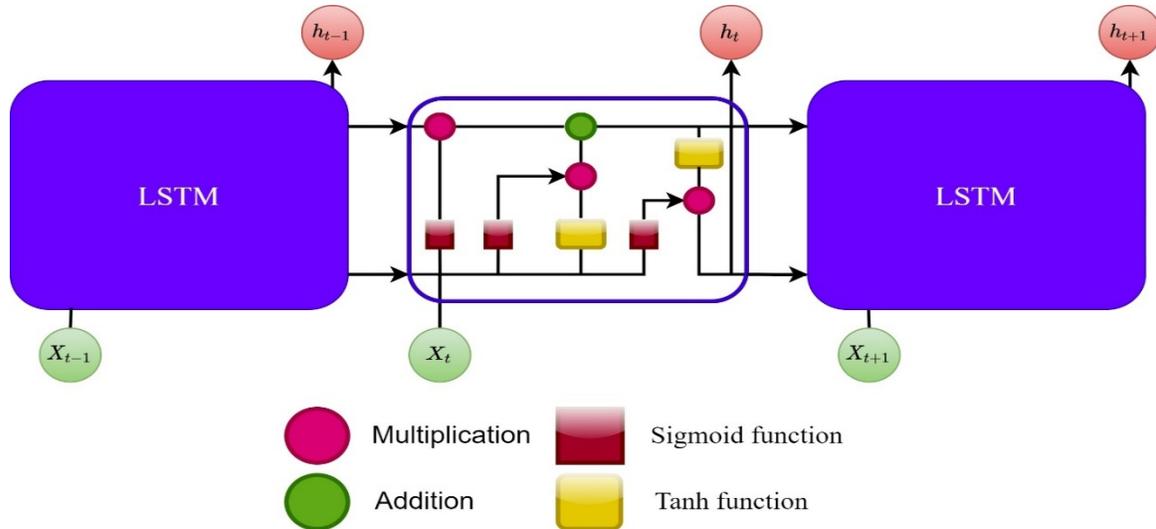
Figure 3-7 LSTM units containing four layers, h_t is an output of a given input X_t

3.3.1 LSTM Gates

The gates are the memory cell's most important component. These are also weighted functions that control the flow of information in the cell. There are three gates to navigate:

- Forget Gate : determines which data from the cell should be discarded,

- Input Gate : determines which input values should be used to update the memory state and

- Output Gate : based on the input and the cell's memory, determines what to output.

The internal state is updated using the forget gate and input gate. The output gate serves as a last check on the cell's output. It's these gates, and the consistent data flow that keep each cell stable (neither exploding nor vanishing) [1,27].

3.3.2 Walk Through LSTM

In this section we are going through the LSTM to provide more details and on LSTM works, as explained in the literature [1]. As mentioned earlier, the LSTM contains several gates which help us to determine the output. However, the key of LSTM is the cell state C. The cell state is considered as a carrier belt. With only a few small linear interactions, it goes straight down the whole chain (as seen in Figure 3-7). It is quite easy for data to just travel over it unaltered. Through

the cell state, the LSTM have the ability to remove or add information. This is regulated by the gates that mentioned previously. As results, the LSTM uses these gates to protect and control the cell state. As illustrated in the Figure 3-8, the first step in the LSTM unit, is the forget gate, which allows to decide which information. The decision is based on the sigmoid layer function. It takes as input a concatenated vector of the h_{t-1} which represents the hidden state of previous LSTM units, the new input vector x_t and its output a number between 0 and 1 which means "No need of this value" and "there is a need for this value " for each number in the cell state, respectively. Therefore, they forget gate output f_t is described as following :

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3.27}$$

Once the LSTM has established which information to discard from the cell state, it can also decide which information to store in the cell state. This is done by the input gate in two steps with same input as the forget gate. First, a sigmoid layer chooses which value to update. Then, a Tanh layer generates a new candidate value vector $\hat{C}_t$, which may be used to update the state. We will combine these two in the following step to generate an updated state :

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3.28}$$

$$\hat{C}_t = Tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{3.29}$$

Once these operations are done, the previous cell state C_{t-1} can be updated using the forget and inputs gate output. Therefore, the forget output results f_t multiplied by the C_{t-1} then the results will be added to the input gate results. The new cell state C_t is expressed as follows :

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \tag{3.30}$$

As a final step, the LSTM comes through the output gate. As shown in Figure 3-8, it is clear that the output results will be based on the cell state. This will, however, be subjected to some sort of a filter. In the beginning, a sigmoid function is applied to the h_{t-1} *and* x_t, in a mean time, C_t is going through a Tanh function in order to attempt to keep the numbers between -1 and +1. Finally, the results will be multiplied by the sigmoid output results. It is described with the following equations :

$$o_t = \sigma(W_o \cdot [h_{t\text{-}1}, x_t] + b_o) \tag{3.31}$$

$$h_t = o_t * \text{Tanh}(C_t) \tag{3.32}$$

As noticed in the Figure 3-8, two hidden state outputs are shown. One is for the output predicted value and the other one is for the next unit step where it is needed for forecasting the successive values in the input sequence.
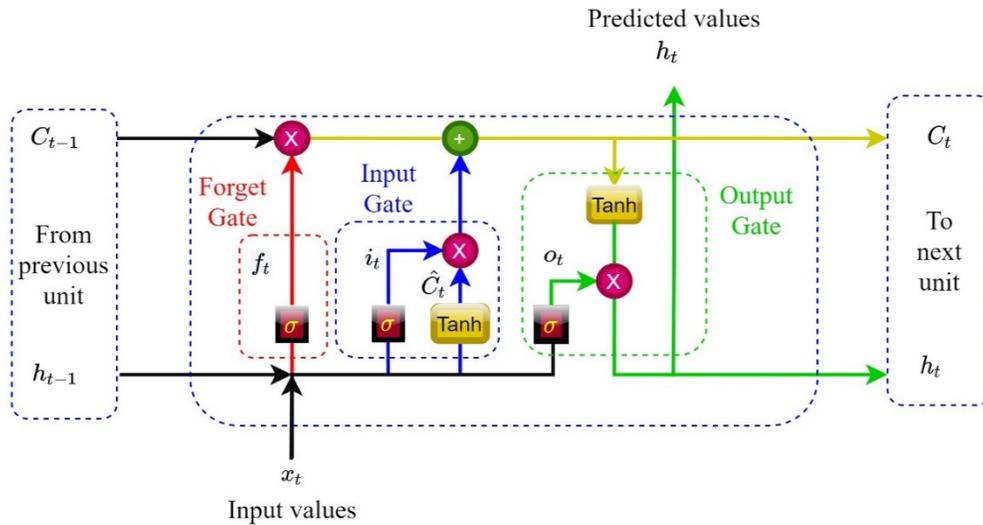
Figure 3-8 Walk through LSTM units.

3.3.3 LSTM Encoder-Decoder Network

The Encoder-Decoder (E-D) architecture is designed for sequence to sequence problems [29]. The E-D LSTM was created for natural language processing issues where it exhibited state-of-the-art performance, particularly in the statistical machine translation field [28,30-32]. The E-D has the capacity to read and create arbitrary-length sequences as demonstrated in Figure 3-9. Two LSTM networks called the encoder and decoder are used by this architecture. The input sequence $U = u_1, ..., u_T$ is processed through the cell state C_t. As LSTM is used in the E-D architecture, the updates were done by going through the equations (3.26 to 3.31) for each LSTM cell. However, after T times updates (T is the length of U), the encoder reviews the whole input sequence into the final cell state vector C_T. The encoder then feeds C_T to the decoder, which utilizes it as the starting cell state for sequence creation (i.e., $C_0' = C_T$). The output sequence is generated recursively by

the decoder $S = S_1, \ldots S_{T'}$. During each update, the decoder feeds the previous update's output S_{t-1} into the current update's input. It is worth mentioning that the decoder's output is derived by performing the affine transformation, followed by the metric that best matches the task at hand e.g., SoftMax function for classification task [30-32].
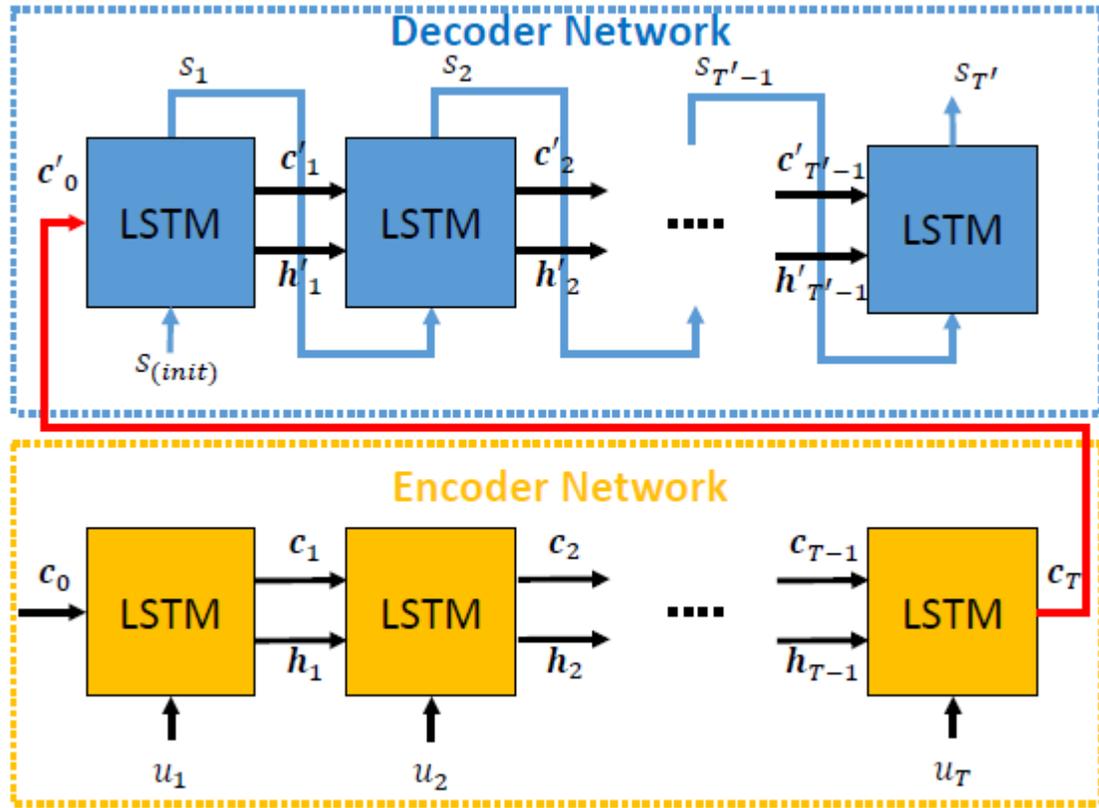
Figure 3-9 The LSTM encoder-decoder architecture [30]

3.4 Machine Learning metrics

When evaluating machine learning (ML) models, optimization algorithms are used, such as Stochastic Gradient Descent (SGD) [33] and Adaptive Moment Estimation (Adam) [34]. The error for the ML model present state must be calculated frequently as part of the optimization procedure. This requires the selection of an error function, sometimes referred to as a loss function, which may be used to estimate the model's loss and update the weights to decrease the loss on the next evaluation. The choice of loss function must fit the framing of the specific predictive modeling issue, such as classification or regression, because neural network models learn a mapping from inputs to outputs from instances. Furthermore, the output layer's setup must be adequate for the

loss function used. Choosing the correct measure is critical. Various loss functions are available to assess ML models in various applications. It is useful to present an overview of common function, so that we can have a better knowledge of each metric and the applications for which it can be utilized. These metrics are divided into groups based on the ML model/application used in this research. The most common are covered metrics in the following regression and classification (binary and categorical).

3.4.1 Classification accuracy metrics and confusion matrix

3.4.1.1 Cross entropy Loss functions

In this section, loss functions are investigated, which are appropriate for binary classification predictive modeling problems. The binary classification will be used to classify the channel within LoS and NLoS scenarios as described in chapter 5. The binary classification models are those predictive models which examples are allocated to one of two classes (labels). However, the default loss function used for binary classification problem is called Cross-entropy (CE). The CE is also known as Log Loss, which measures the performance of classification model whose output is a probability value between 0 and 1 [35]. As the predicted likelihood differs from the actual class or label, cross-entropy loss rises. Moreover, a perfect model loss would have a log loss equal 0. The CE loss is described as follows [35] :

$$CE = -\sum_{i}^{C} y_i \log(\hat{y}_i) \tag{3.33}$$

Where y_i and $\hat{y}_i$ are the real and the predicted for each class i in C.

In a binary classification problem, it usually uses a binary target value, either class 0 or 1. Therefore, $C = 2$ in the equation (3.33). Before applying the Binary Cross Entropy (BCE) loss, the sigmoid activation method is applied in the case of two classes, as presented in Figure 3-10 (a). The BCE is defined as follows [36] :

$$CE = -\sum_{i}^{C=2} y_i \log(\hat{y}_i) = -y_1 \log(\hat{y}_1) - (1-y_1) \log(1-\hat{y}_1) \tag{3.34}$$

As such, two classes are assumed C_1 and C_2 referred to 0 and 1, respectively. y_1, $\hat{y}_1$ are the real and predicted the class C_1, respectively. $y_2 = 1\text{-}y_1$ and $\hat{y}_2 = 1\text{-}\hat{y}_1$ are the real and predicted the class C_2, respectively.

One of the CE losses used for multi-class problem $(M > 2)$ is a Categorical Cross Entropy loss (CCE). It is also called SoftMax loss due to the use of the SoftMax activation function before the evaluation CCE loss (Figure 3-10 (b)). The CCE loss is defined as follows [35] :

$$CCE = -\sum_{i>2}^{M} y_i \log(\hat{y}_i) \tag{3.35}$$

Where M is the number of classes.
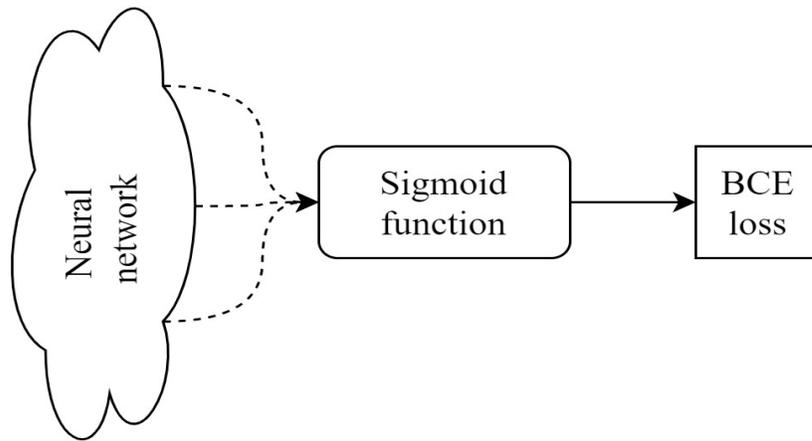
(a)
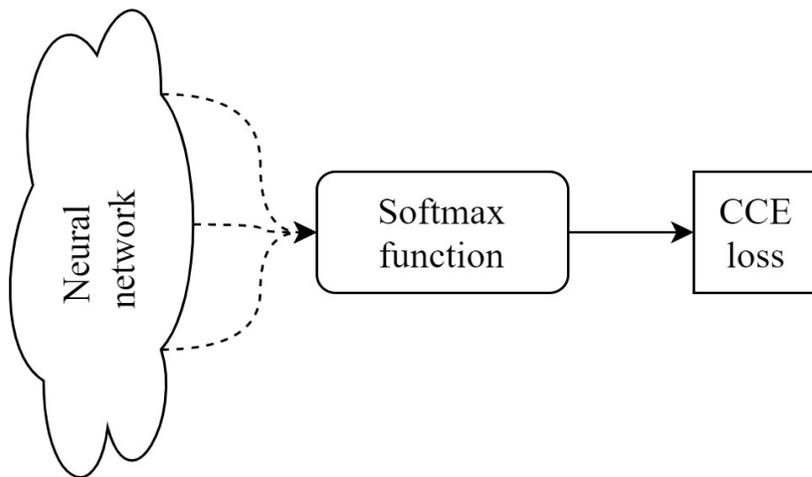
(b)

Figure 3-10 Loss functions for classification problem, (a) BCE loss, (b) CCE loss.

3.4.1.2 Accuracy

The classification accuracy is the basic metric which is used for channel classification in chapter 5. It uses the number of right predictions divided by the total number of predictions multiplied by 100. It is described as follows.

$$\text{Accuracy} = \frac{\text{number of True predictions}}{\text{total of predictions}} \times 100$$

$$= \frac{TP}{TP + TN + FP + FN} \times 100 \tag{3.36}$$

Where :

TP = Describes the true positive in which cases the model classifies positively the data, and the input data is positive (e.g., the predicted data is 1 and the input is 1)

TN = Describes the true negative in which cases the model negatively classifies the data, and the input data is negative (e.g., the predicted data is 0 and the input is 0)

FP = Describes the false positive in which cases the model classifies positively the data, and the input data is negative (e.g., the predicted data is 1 and the input is 0)

FN = Describes the false negative in which cases the model negatively classifies the data, and the input data is positive (e.g., the predicted data is 0 and the input is 1)

3.4.1.3 Precision

In many situations, classification accuracy is not a reliable measure of model performance. As it doesn't differentiate between the amount of successfully categorized instances of various types, such as in the case of the class distribution is unbalanced (one class is more frequent than others). Even if all samples are predicted as the most frequent class , a high accuracy rate in this scenario is obtained, which makes no sense (since the model is not learning anything and is just predicting everything as the most frequent class). As a result, a class-specific performance measures are considered. One of these measures is precision, which is defined as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.37}$$

3.4.1.4 Confusion matrix

Several tools exist to provide a good resolution for the classification. Therefore, in this thesis is limited to both accuracy and precision. The accuracy and precision are calculated by using the confusion matrix (Matlab) which describes the performance of a classifier on a set of test data. Let us consider the example of a confusion matrix in a binary classification to see what information it might supply to the classifier. The example is forecasting the existence of a LoS and NLoS scenarios for 200 collected CFR data where only 105 are in LoS. However, two potential anticipated classes are shown in Figure 3-11 as "1" and "0." where it is described as follows:

- 1: would indicate that the data are in LoS scenario

- 0: would indicate that the data are NLoS scenario

A total of 200 predictions were produced by the classifier. The classifier correctly predicted "1" 110 times and "0" 90 times out of 200 instances. In fact, 105 of the data in the CFR are in LoS, whereas 95 do not. Therefore, on 105 identified data, 5 of them were negatively classified, where the model is supposed to classify them correctly. This is defined as False negative. The 10 of 95 identified in LoS but actually they are NLoS. This is defined as False positive earlier.

	Predicted 0	Predicted 1	
Actual 0	85	10	95
Actual 1	5	100	105
	90	110	

Figure 3-11 Confusion matrix

Given the TP, TN, FP, and FP in this cases, it can be shown in Table 2. However, with the confusion matrix for binary classification, several rates are computed such as accuracy, precision, etc. As results, in the proposed classification model only the confusion matrix is computed.

Table 3-2 Description of the terms for the example

TP	These are situations when we predicted LoS and it turned out to be correct.
TN	They are in NLoS, as expected by the classifier.
FP	The classifier predicted LoS, but they are NLoS
FN	The classifier predicted NLoS, but they are LoS

3.4.2 Regression Metrics

Regression models are a type of machine learning and statistical model that is used to predict the values of a continuous target. Considering The LSTM network as regression model, the Mean Square Error (MSE) used for prediction loss [37]. However, the MSE is described as follows:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (y_i \text{-} \hat{y})^2 \tag{3.38}$$

Where N is the number of samples, $\hat{y}$ and y are the predicted and actual value, respectively. However, the MSE is used in this thesis as prediction loss in channel frequency prediction for both SISO and WBAN-MIMO cases.

3.5 References

[1]   I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016.

[2]   A. M. Delaney, E. Brophy, and T. E. Ward, "Synthesis of Realistic ECG using Generative Adversarial Networks," *arXiv preprint arXiv:1909.09150,* 2019.

[3]   M. C. Chuah and F. Fu, "ECG anomaly detection via time series analysis," in *International Symposium on Parallel and Distributed Processing and Applications*, 2007: Springer, pp. 123-135.

[4] A. Dingli and K. S. Fournier, "Financial time series forecasting–a deep learning approach," *Int. J. Mach. Learn. Comput,* vol. 7, no. 5, pp. 118-122, 2017.

[5] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research,* vol. 270, no. 2, pp. 654-669, 2018.

[6] S. Siami-Namini and A. S. Namin, "Forecasting economics and financial time series: ARIMA vs. LSTM," *arXiv preprint arXiv:1803.06386,* 2018.

[7] J. Wang, Y. Ding, S. Bian, Y. Peng, M. Liu, and G. Gui, "UL-CSI data driven deep learning for predicting DL-CSI in cellular FDD systems," *IEEE Access,* vol. 7, pp. 96105-96112, 2019.

[8] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, 2016: IEEE, pp. 324-328.

[9] M. Van der Baan and C. Jutten, "Neural networks in geophysical applications," *Geophysics,* vol. 65, no. 4, pp. 1032-1047, 2000.

[10] Y. Ren, P. Zhao, Y. Sheng, D. Yao, and Z. Xu, "Robust softmax regression for multi-class classification with self-paced learning," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 2017, pp. 2641-2647.

[11] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," *arXiv preprint arXiv:1811.03378,* 2018.

[12] T. Szandała, "Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks," in *Bio-inspired Neurocomputing*: Springer, 2021, pp. 203-224.

[13] A. de Brébisson and P. Vincent, "An exploration of softmax alternatives belonging to the spherical loss family," *arXiv preprint arXiv:1511.05042,* 2015.

[14] M. Wang, S. Lu, D. Zhu, J. Lin, and Z. Wang, "A high-speed and low-complexity architecture for softmax function in deep learning," in *2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, 2018: IEEE, pp. 223-226.

[15] C. R. Rao, C. Rao, and V. Govindaraju, *Handbook of statistics*. Elsevier, 2006.

[16] R. H. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung, "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit," *Nature,* vol. 405, no. 6789, pp. 947-951, 2000.

[17] G. Arulampalam and A. Bouzerdoum, "A generalized feedforward neural network architecture for classification and regression," *Neural networks,* vol. 16, no. 5-6, pp. 561-568, 2003.

[18] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks,* vol. 2, no. 5, pp. 359-366, 1989.

[19] M. W. Gardner and S. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmospheric environment,* vol. 32, no. 14-15, pp. 2627-2636, 1998.

[20] I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet, "Multi-digit number recognition from street view imagery using deep convolutional neural networks," *arXiv preprint arXiv:1312.6082,* 2013.

[21] T. M. Mitchell, "Machine learning," 1997.

[22] J. Cheng, L. Dong, and M. Lapata, "Long short-term memory-networks for machine reading," *arXiv preprint arXiv:1601.06733,* 2016.

[23] G. Dorffner, "Neural networks for time series processing," in *Neural network world*, 1996: Citeseer.

[24] J. C. B. Gamboa, "Deep learning for time-series analysis," *arXiv preprint arXiv:1701.01887,* 2017.

[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation,* vol. 9, no. 8, pp. 1735-1780, 1997.

[26] H. Sak, A. W. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," 2014.

[27] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* vol. 6, no. 02, pp. 107-116, 1998.

[28] S. H. Park, B. Kim, C. M. Kang, C. C. Chung, and J. W. Choi, "Sequence-to-sequence prediction of vehicle trajectory via LSTM encoder-decoder architecture," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018: IEEE, pp. 1672-1678.

[29] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *arXiv preprint arXiv:1409.3215,* 2014.

[30] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078,* 2014.

[31] G. Neubig, "Neural machine translation and sequence-to-sequence models: A tutorial," *arXiv preprint arXiv:1703.01619,* 2017.

[32] T. Luong, E. Brevdo, and R. Zhao, "Neural machine translation (seq2seq) tutorial. 2017," *URL: https://www. tensorflow. org/tutorials/seq2seq (дата обращения 17.02. 2018),* 2017.

[33] L. Bottou, "Stochastic Gradient Descent Tricks," in *Neural Networks: Tricks of the Trade: Second Edition*, G. Montavon, G. B. Orr, and K.-R. Müller Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 421-436.

[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.

[35] Z. Zhang and M. R. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," *arXiv preprint arXiv:1805.07836,* 2018.

[36] A. Buja, W. Stuetzle, and Y. Shen, "Loss functions for binary class probability estimation and classification: Structure and applications," *Working draft, November,* vol. 3, 2005.

[37] A. Botchkarev, "Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology," *arXiv preprint arXiv:1809.03006,* 2018.

CHAPTER 4   CHANNEL   PREDICTION   FOR   INDOOR ENVIRONMENT IN SISO SYSTEMS

4.1   Introduction

Several applications have arisen to characterize the channel within different environments using DL algorithms. *Tong et al* [1], proposed a method based on Long Short-Term Memory (LSTM) to achieve low prediction errors compared with other machine learning algorithms. However, the data processing for the input data remains undetermined. *Wang et al* [2], proposed a method to predict the uplink and downlink Channel State Information (CSI) using different networks, such as convolutional and LSTM networks. The method fairly predicted the CSI, where the learning curve (LC) and the data processing were not illustrated. *Joo et al* [3], proposed channel state prediction in vehicle-to-vehicle scenarios, where, IQ (in-phase and quadrature) samples are collected under the line-of-sight (LOS) and the non-LOS scenarios. In fact, the training process is used in all the reported scenarios which increases the computational complexity, with a loss on the generalization efficiency of DL algorithms. Recently, *Luo et al* [4], introduced a learning framework design with convolutional and LSTM networks. The framework used an online prediction scheme that updates the model once new data are available. The input data processing and the Learning Curve (LC) during the training are not well illustrated, even though the methods accurately predicted the channel. Furthermore, the training process is usually used in all scenarios, which increases the computational complexity with the loss on the generalization efficiency of DL algorithms. Even though the statistic, the empirical and the deterministic propagation models combined with several applications of the DL are powerful, the CFR remains deprived of a proper model. Consequently, it is crucial to develop a model that will characterize the channel without the need of experimental measurements and be deprived of any updates once new data are collected from the new environment.

In this contribution, an efficient model of the CFR based on DL LSTM is proposed and studied. The learning framework algorithm is developed to predict the CFR using the LSTMs networks combined with linear layers, in smooth, rough corridors and underground mine environments. Then, channel characterization is performed for every measured and predicted CFR. From the model learning process in a smooth corridor environment, the model is evaluated in different

environments without prior information and update during the learning process. To the best of the authors' knowledge, no such prediction scheme was performed to model the CFR within different complex environments, different frequency bands, and different antenna combinations. The novelty of this study consists of the following:

First, a novel CFR model is proposed for different environments such as smooth, rough corridors and underground mine environments. The SISO-LoS scenario has been carried out for different antenna configurations and frequency bands.

Second, the learning framework is designed based on LSTM combined with linear networks. The learning curve is performed in only one environment, which is crucial in the DL field to demonstrate the ability of the model to learn from the collected CFR.

The model is evaluated for CFR prediction in different environments, considering different antenna radiation patterns. Hence, three different antenna combinations were used: (1) directional transmitter versus directional receiver antennas (D vs. D), (2) directional transmitter versus patch receiver antennas (D vs. Prx) for the corridors environments, and (3) patch antennas used in underground mine environment (P vs. P).

Hence, several experimental campaigns were conducted to validate the proposed approach and channel characterization was carried out to compare the predicted CFR with the measured ones.

The chapter is organized as follows: Section 4.2 briefly introduces the CFR prediction. Then, the proposed model for CFR prediction is described in Section 4.3. Afterwards, the experimental results are presented in Section 4.4. Finally, the chapter is concluded in Section 4.5.

4.2 CFR prediction scheme:

4.2.1 CFR prediction framework

As mentioned previously, the channel modeling and the characterization depends on a large number of experimental measurements. Furthermore, they depend on several factors, such as the frequency range, the antenna radiation patterns, and the environment characteristics. Several researchers investigated properly the wireless channel modeling to predict the main impact on the signal transmission [5,6]. Regarding to previous related works [1-4,7-10], the proposed model is based on three main modules. First, the data processing module is focused to transform the data

into appropriate short sequence input, to achieve better performance by avoiding the gradient descent problems in the training and validation process [11]. Then, the training and validation module is developed to assure the efficiency of the module, where the LC during the learning process is observed to optimize the internal parameters of the deep learning LSTM networks combined with linear layers. As depicted in Figure 4-1, the model evaluation was carried out by the measured data in different scenarios and considering additive noise. This model uses only one type of data, explicitly time series data. The LSTMs network combined with linear layer achieved high performance compared to Gated Recurrent Units (GRU) which has been introduced in several models [12-14] to predict time series data. In section 4-4, a comparison between the proposed model and some results in the literature was carried out to show the performance assessment in terms of accuracy expressed in different metrics. To explore more accuracy of the model, some information that is present in every measurement should be considered, such as Radiation Patterns (RP), antenna heights, indoor geometry namely Environment Features (EFs), which are repeated in each measurement. In the following, measurement procedures and data processing are described.
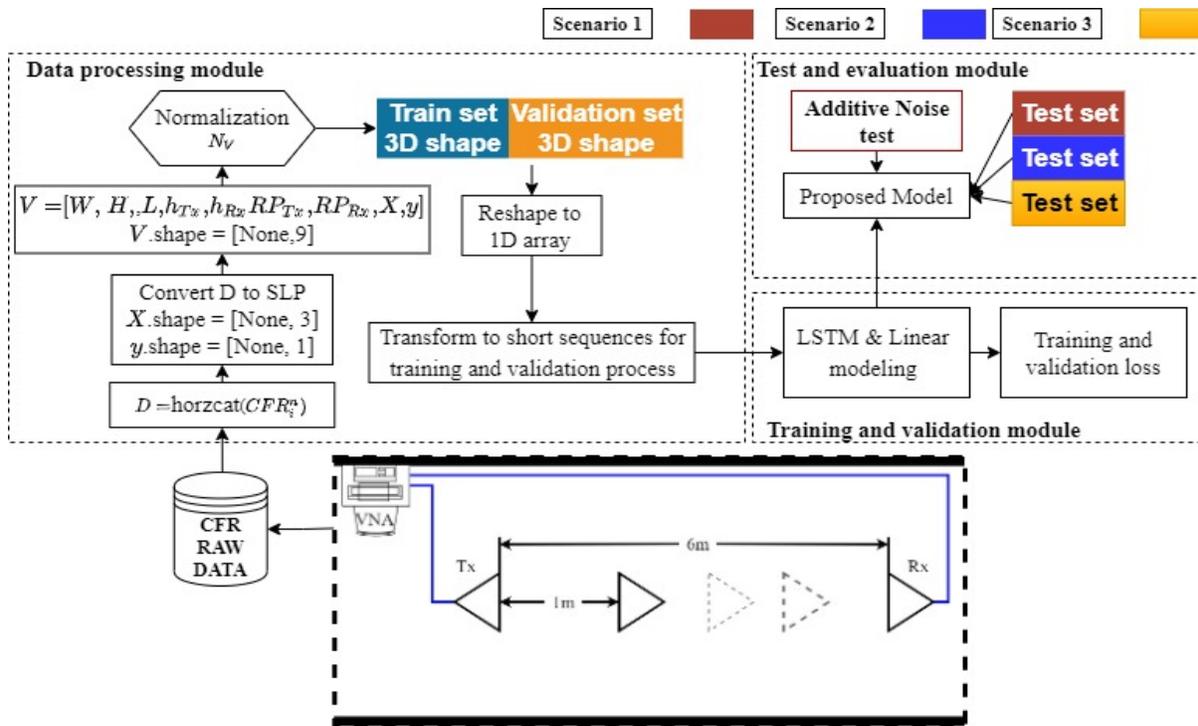
Figure 4-1 The framework for the CFR prediction model

4.2.2 Measurement procedure

The measurement procedures were conducted at three different environments, smooth and regular rough walls, located at the University of Quebec in Outaouais, and underground mine gallery, located at Val d'Or city in northern Quebec. The measurement system consists of an Agilent Vector Network Analyzer (VNA), and Low Noise Amplifiers (LNA) connected to the receiver. The VNA is used to measure the CFR within the range of 8 GHz to 12 GHz band for the first and second scenario and 2.3 GHz to 2.5 GHz band for the third scenario. During this measurement, the Transmitter (Tx) was placed at a fixed position, and the Receiver (Rx) location was varied from 1 m to 6 m away from Tx in the LOS-SISO scenario, as illustrated in Figure 4-3. Several data samples (up to 2049 points) are considered at each position. The measurement parameters are described in Table 4-1.

The measurement scenarios are described as follows:

i. Scenario 1 (smooth environment): the measurement was conducted in the corridor shown in Figure. 4-2(a). The corridor is 2.4 m in height and 2.05 m in width. The Tx and Rx antennas have 1.2 m in height. Two antenna combinations D. vs. D and D. vs. Prx were considered.

ii. Scenario 2 (irregular periodic rough environment): the measurement was conducted in the corridors shown in Figure 4-2(b). This corridor is 2.4 m in width and 3.4 m in height. Same antenna combinations were used for this scenario.

iii. Scenario 3 (underground mine environment): the measurements were conducted in a real underground mine (2.5 m by 5 m) as shown in Figure 4-2(c). The gallery is characterized by rough and random surfaces. Patch antennas are used for this scenario (P vs. P).

Table 4-1 Measurements system configuration

Parameters	Scenario 1,2	Scenario 3
Frequency	8 GHz-12 GHz	2.3 GHz-2.5 GHz
Transmitted power	-10 dBm	-10 dBm
IF bandwidth	1 kHz	200 MHz
Average noise floor	-110 dBm	-90 dBm
Sweep points:	1601	2049
Directional antenna gain	13.8 dBi	N/A
Patch antenna gain	7.7dBi	6.6 dBi
Cable loss	3.1dB-3.8dB	0.6 db/m
Antenna height	1.20 m	1.5 m
Sweep time	60s	60s

(a)   (b)

(c)

Figure 4-3 Measurement environments :

 (a) Scenario 1,(b) Scenario 2 and (c) Scenario 3.
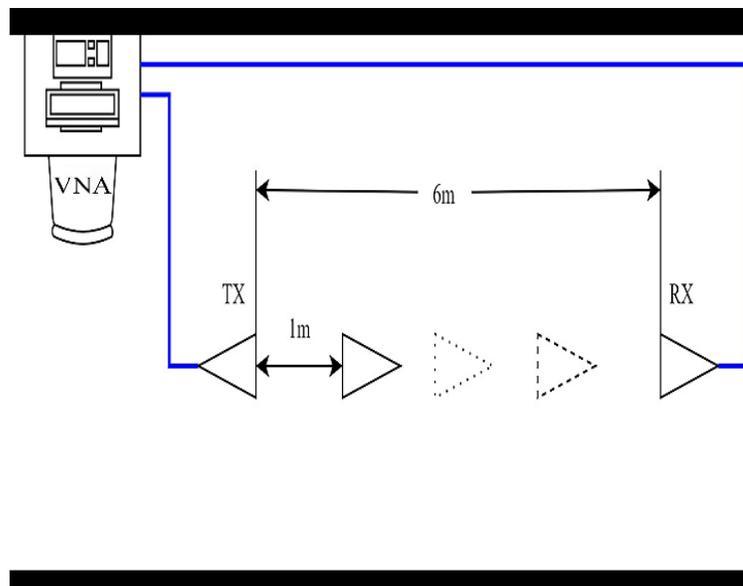
Figure 4-2 Experimental setups used for smooth, rough corridors and underground mine environments

4.2.3 Data processing modules

The data processing is illustrated by the diagram in Figure 4-4. Following the measurement procedure, six measured CFR at each position are collected. Then, all the sequences were linked

horizontally by every CFR position from D. vs. D and D. vs. Prx antenna combinations, respectively. Accordingly, twelve CFR sequences concatenated will be provided in one vector dataset for the first and the second scenarios. Since only one antenna combination for the third scenario is used, six CFR sequences are horizontally concatenated to give the dataset vector D.

$$D = \text{horzconcat}(CFR_i^n) \qquad (4.1)$$

Where i= 1, 2, and 3 represent D. vs. D, D. vs. Prx and P vs. P antenna combinations, respectively, and i = 1…. 6 represent the number of the positions. Then, the vector D is converted into a Supervised Learning Problem (SLP) described in the next section. Finaly, The Width (W), lengths (L), height (H), antenna heights and the radiation patterns (RPs) were added.
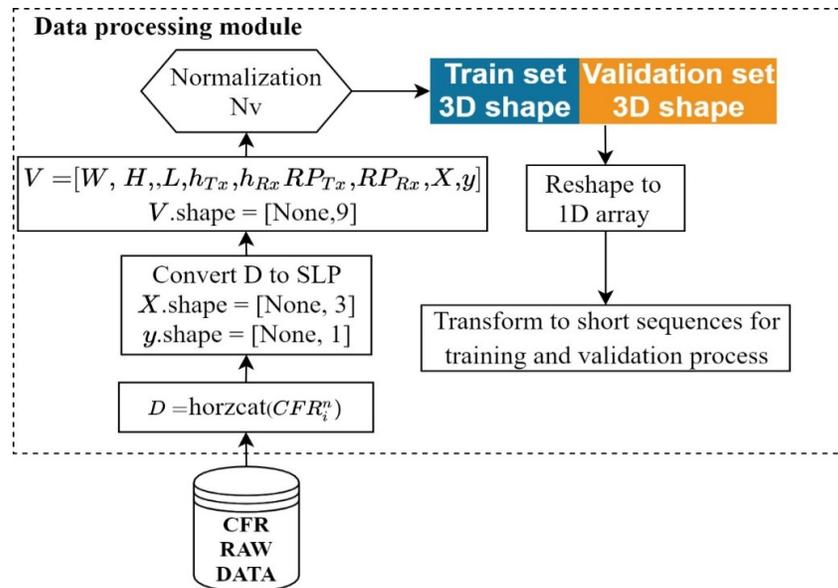
Figure 4-4 Data processing module

4.2.4 Convert to the supervised learning problem

To achieve the CFR prediction, the concept of the sliding window method is used [15].This technique allows converting any real problem into supervised machine learning problems [15]. Thus, the data must be restructured to convert them into supervised learning, by using a previous time steps as the input and the next time step as the output variables (Figure 4-5).

From Figure 4-5. It is denoted that the previous time step is the input (X), and the next time step is the output (y) in the SLP. The order between the observations is preserved, even using this dataset to train the supervised model. Due to the fact that no prior value can be used to estimate

the first series (red row) and it is not possible to predict a non-existent value (yellow row). The red and yellow rows are deleted during the training process.

Time (ms)	Measured CFR (dBm)
1	-15
2	-20
3	-18
4	-35
5	-46

X (dBm)	y (dBm)
?	-15
-15	-20
-20	-18
-18	-35
-35	-46
-46	?
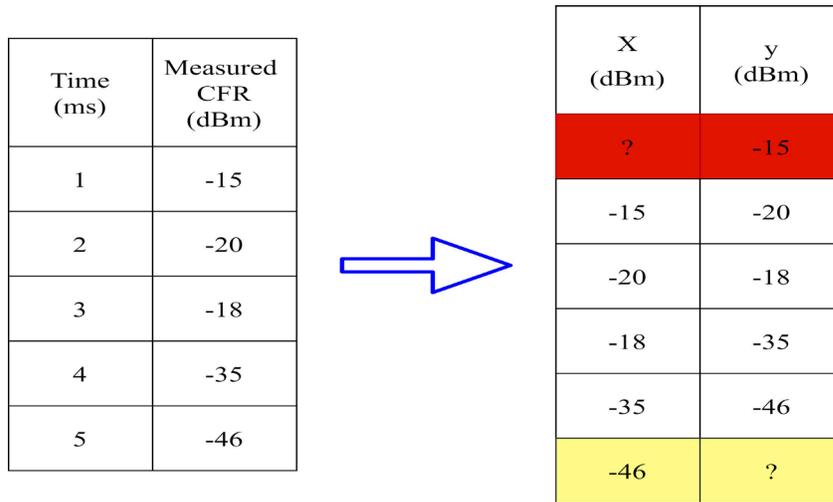
Figure 4-5 Example of sliding window method

4.2.5 Normalization and data transformation

After the SLP process, the antenna radiation patterns [16] and the environment features in terms of antenna heights and environment width, height, and length, are horizontally concatenated to the data V as shown in Figure. 4-4. To make the model converge faster, it is important to normalize the data V .In this case, the standard score normalization is considered by using the mean and standard deviation of all the dataset sequences. It is defined as [17]

$$N_V = \frac{V- \mu}{\sigma} \tag{4.2}$$

Where μ, σ are the mean and the standard deviation of the data V. Indeed, the actual feature data are at consecutive points in time. Even though, the LSTM networks are designed for long sequences. Practically, the training process could be a real issue, where the gradient descent trajectory leads to the saddle point [11,18]. Therefore, for better performance of the model, the dataset will be transformed to an appropriate mini-batch sequence [19]. During the training process, the long dataset sequence is converted to many shorter sequences. In this case, the dataset will be transformed to shorter sequences length of hundred points where the best performance of the model is observed in section 4. The sequences are shifted by single time step where the first

and second sequences in the training set are shown in Figure. 4-6. The only difference is that the target is moved by once a step in the ytrain (target) and xtrain (feature) sequences, as seen in the first (green) and last (magenta) time step. An explanation will be provided to understand the benefit of using the sequences during the training process (section 4).

4.3 Proposed Model

The block diagram of Figure. 4-7 shows a high-level process of the proposed model. Unlike the previous research works [1-4,7-10], the training process is performed just once and only in one

(a)   (b)
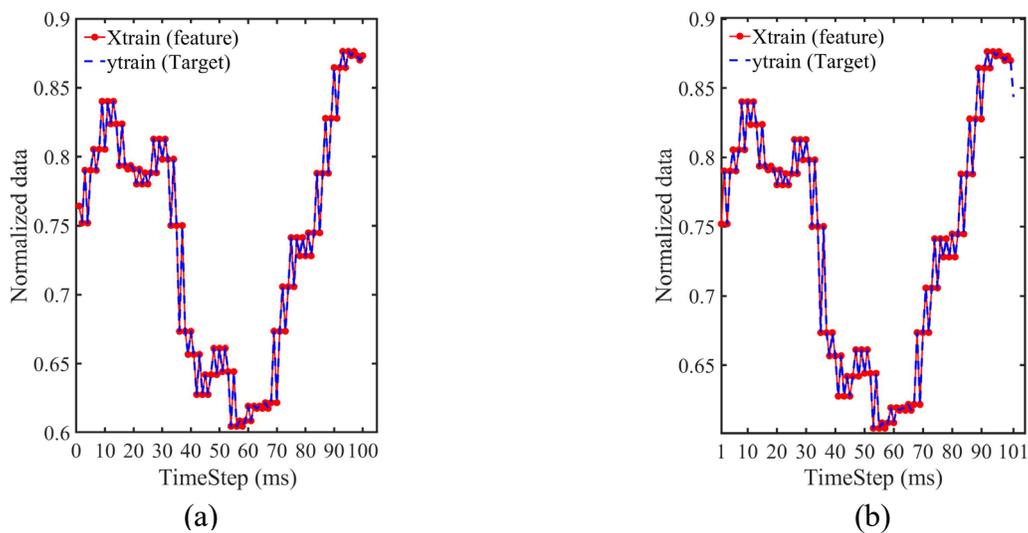
Figure 4-6 Feature and target sequences in the training process: (a) Sequence 1,(b) Sequence 2 environment. The first scenario is used to train the model and to predict the CFR in the first environment. From the second and third environments, the new data, which are not used during the training process, are used to evaluate the model capability to predict the CFR in different environments. Finally, the channel characterization is performed. Therefore, measured and predicted results are compared.

4.3.1 LSTM networks

Like other neural network architecture, the LSTM is composed of an input, a hidden and, an output layer. The number of the neuron in the input layer is equal to the number of the features. Using the combination of these equations described in chapter 3 (3.27 to 2.32), the LSTM unit can be expressed as

$$h_t = f(h_{t-1}, x_t, \psi) \tag{4.3}$$

Where f is the LSTM function and ψ is the vector parameter in the LSTM.
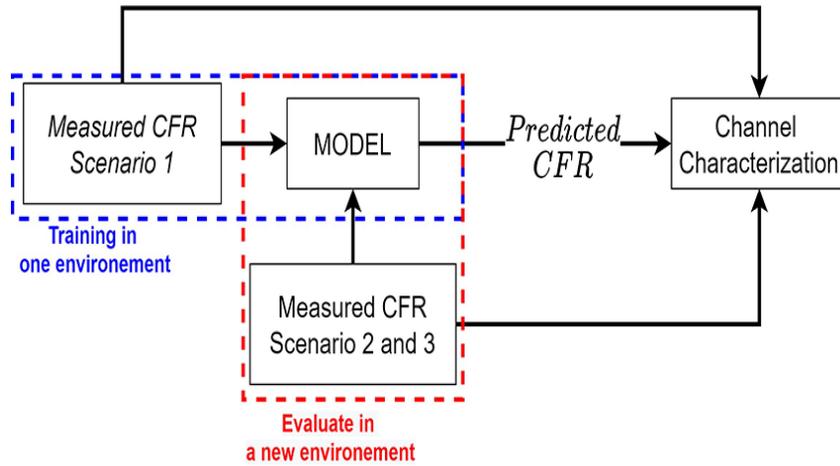
Figure 4-7 Proposed model for CFR prediction

4.3.2 Train and validation mechanism

To predict the CFR, the model is trained with the data collected from the first scenario. The normalized dataset V is then separated into train and validation sets. Moreover, the sets have been transformed into short sequences to train the model. In fact, the LSTM network is considered excellent in sequential task learning; hence, it is used to predict the current values. Using equation (4.10), the sequential vector can be determined as

$$h_{CV} = f(h_{CV-1}, D, \psi) \tag{4.4}$$

The result h_{CV} will be used as an input to the linear layer to have the target dimensionality as

$$\hat{y} = W^*h_{CV} \tag{4.5}$$

After conducting the training and validation process, the loss is reduced by using the Adam optimizer [20]. Mean Square Error (MSE) and Root Mean Square Error (RMSE) metrics are then performed as prediction loss of the model, which is given in the equation 3.38 where RMSE is the described as the root square of the MSE [21].

4.4 Experimental Validation

4.4.1 Experimental results

As mentioned in the data processing, an Inverse Fourier Transform (IFT) was carried out to the average of six measured CFR at each position. Figure 4-8. shows the CFR and CIR in different scenarios with different antenna configurations. The LOS component carries the highest power over the multipath component in different situations.

4.4.2 Training in the first scenario

Using the measurement procedure for the first smooth corridor environment and data processing scheme, up to 20 000 frequency points are considered. To train the model, 2/3 of the normalized dataset V is the training set and 1/3 is the validation set. y vector is the test set in this case. To understand the way that the proposed model uses the sequences in the training phase, we use the sequences shown in Figure. 4-6. From the first sequence, the model takes the feature of the time step at the first index and tries to predict the target in the second index. Then, the model exploits the feature of the time step at the second index to predict the target of the time step at the third index, etc. The features of the second sequences are shifted by once a step from the first sequence, where the third sequence is shifted by once a step from the second sequence. Using this procedure, we get many shifted sequences by a single time step which allows the model to easily learn the pattern and acquire more information about the dataset. As discussed previously, the training and validation procedure is very crucial for the learning process. In this matter, the LC must be performed for monitoring the model performances [22,23]. Hence, the LC is used in deep learning neural network algorithms to optimize the internal parameters where the train learning curve (evaluated on the training dataset) gives a representation of the learning process of the model. In the same area, the validation learning curve (evaluated on the validation dataset) provides an indication of how well the model is generalizing by measuring the models' error in the validation set. The LC can be crucial for the performance of the model and the accuracy of the test results. Therefore, a good fit as described in [22,23] must be fulfilled to ensure that the model is learning sufficiently to avoid under-fitting (the model is not learning well from the training dataset) or
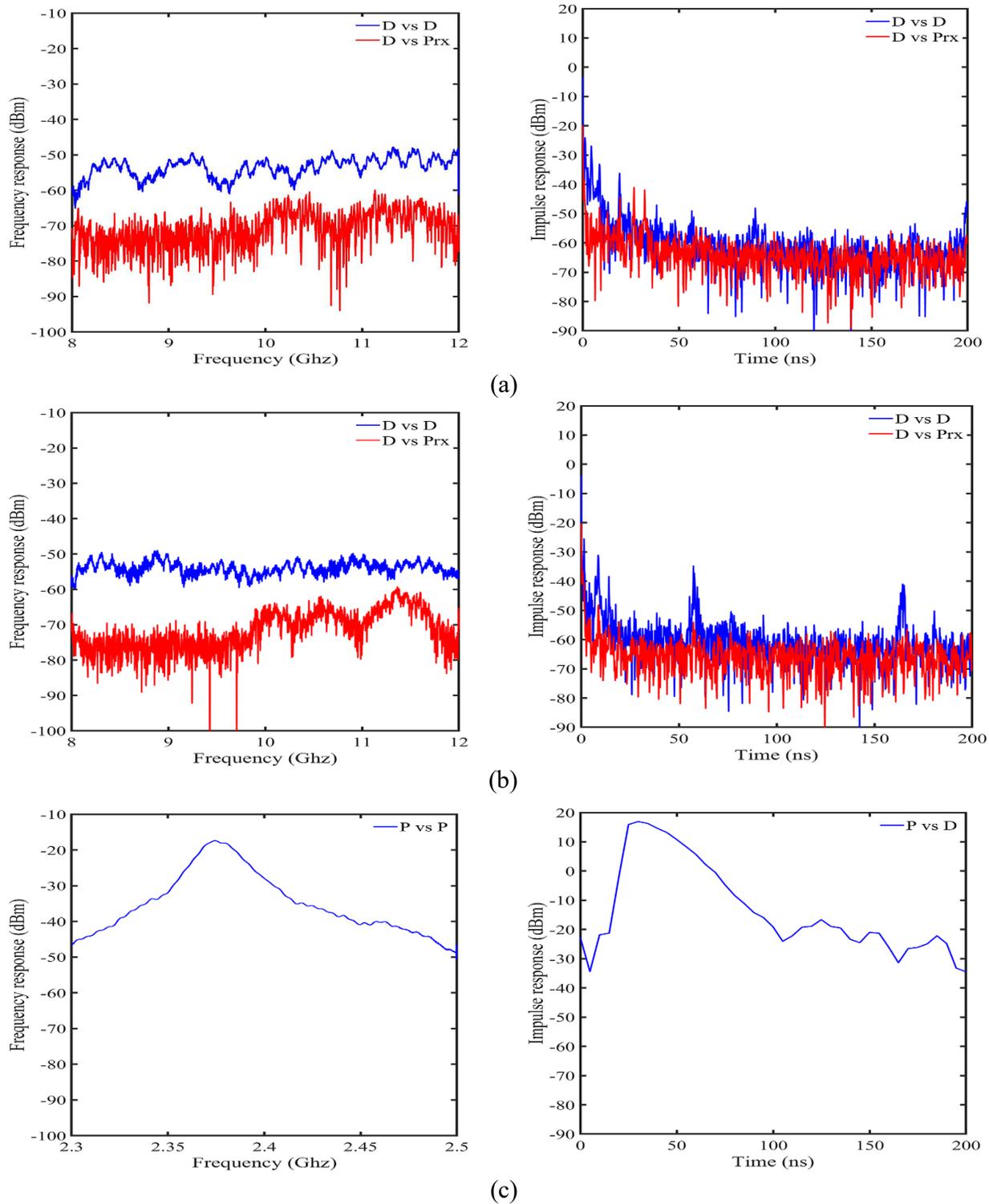
Figure 4-8 Channel frequency response vs Channel impulse response :

(a) Scenario 1, (b) Scenario 2, (c) Scenario 3

overfitting (the model is learning too well from the training dataset) [22,23]. Figure 4-9 demonstrate the LC of the proposed model achieved by using the parameters in Table 4-2.

Table 4-2 : Simulation parameters

Parameters	Values
Hidden layers	100
Batch size	100
Learning rate	0.001
Loss function	Mean square error
Training dataset size	38790
Validation dataset size	19395
Training optimizer	Adam

A learning rate is initialized at a small value, then, the dynamic learning rate schedule is used to help the learning model to introduce randomness during the training [23,24]. It is observed that the training and validation loss decreases to a stability point with a small difference between the two final loss values which characterizes a good fit learning curve as reported in [22,23]. However, the training and validation MSE loss stability is observed at the 6th epoch with 7.42% and 7.45%, respectively. From the LC, the model is certainly learning well from the training dataset. The predicted value is estimated with the test set, afterward. Therefore, MSE is evaluated at 7.76%, same as RMSE evaluated at 27%. Figure. 4-10. shows the predicted and the measured CFR results. It can be observed that the proposed model has a good ability to predict the measured CFR in the first environments with low MSE and RMSE in terms of percentage.
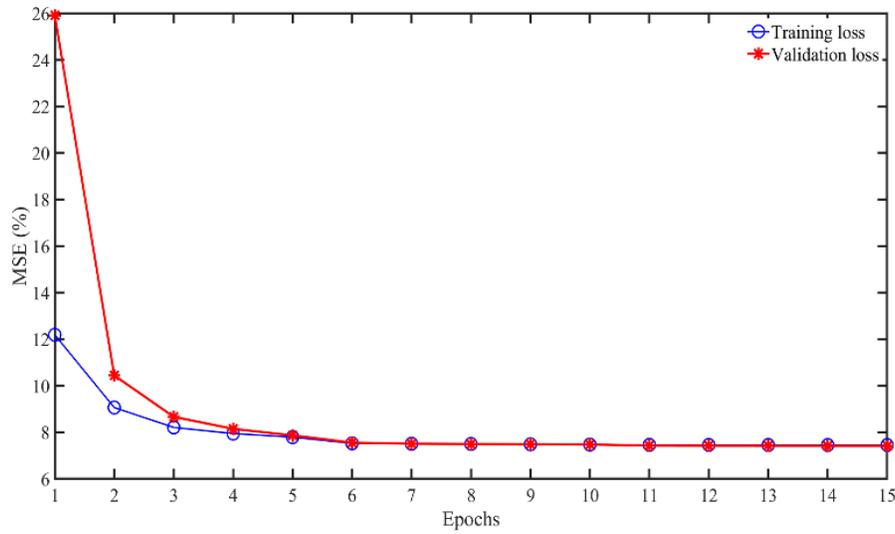
Figure 4-9 The learning curves for the proposed model
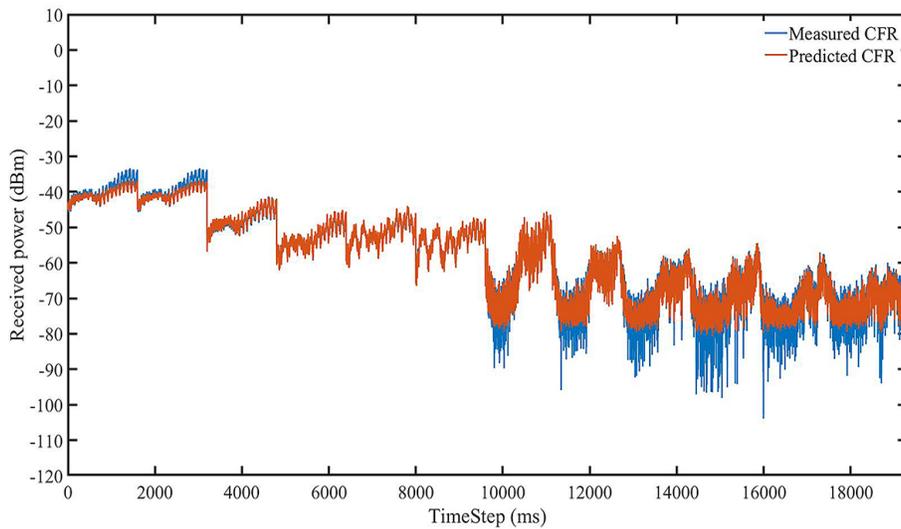
Figure 4-10 Measured and predicted CFR for scenario 1

4.4.3 Evaluate the model within different environments

To evaluate the efficiency of the proposed model to predict the CFR in a new environment, new datasets, which are not included during the training from the second and third environments, are used in the evaluation and test module. Therefore, the data processing is applied. Figure. 4-11 shows the prediction model of the CFR in the second and third rough environments with lower MSE and RMSE values in terms of percentage as illustrated in Table 4-3.

Table 4-3 MSE and RMSE for scenario 2 and 3

Scenarios	MSE (%)	RMSE (%)
2	4.22	20.5
3	0.046	2.1

The prediction accuracy of the model was based on measured data in different environments, however, the performance assessment in terms of accuracy was estimated with additive noise to test data. Figure. 4-12. shows the effect of the noise on the learning process. The x-axis represents the Signal-to-Noise Ratio (SNR) of the datasets, where the large value corresponds to noiseless. As illustrated, it is challenging for the model to predict noisy channels. It implies that any attempt to increase accuracy against noise should be made to enhance the capacity of the model to predict the CFR within noisy channels.
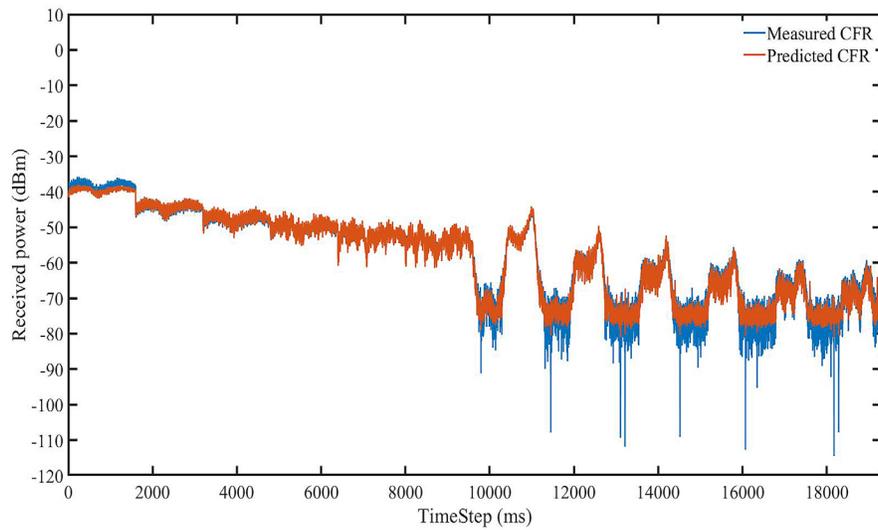
(a)
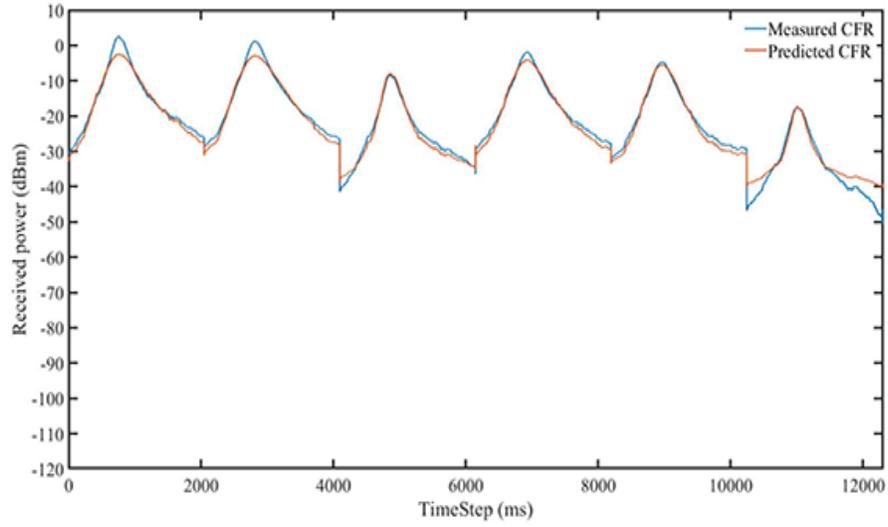
(b)

Figure 4-12 Measured and predicted CFR:(a) Scenario 2, (b) Scenario 3
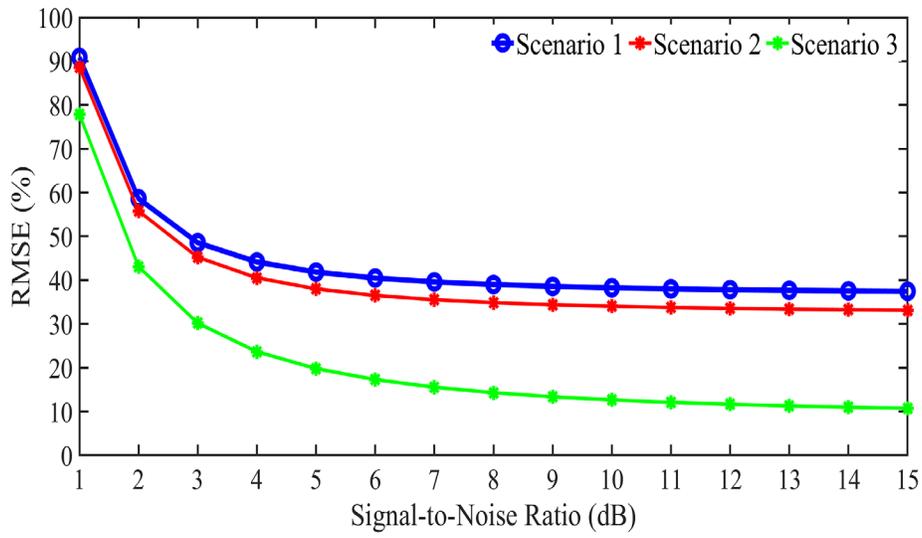
Figure 4-11 Model performance evaluation in terms of additive noise

4.4.4 LSTM versus proposed models for channel prediction

The model efficiency was compared to another network used for time series prediction namely Gated Recurrent Units (GRU) [12,25]. However, this comparison justifies the choice of the LSTM networks in this proposed model. The RMSE is used to compare both networks in terms of percentage. Even though both networks seem to have a lower RMSE, but the LSTMs networks perform better than GRU, which justifies the choice for the proposed model, as shown in Table 4-4

Table 4-4 LSTM and GRU performances

RMSE (%)	LSTMs	GRUs
Scenario 1	27	28
Scenario 2	20.5	22
Scenario 3	2.1	4

As discussed previously, to avoid the gradient descent problems, shorter sequences are used as input in the LSTM networks during the training process. Therefore, the optimal batch size depends on the elapsed time during the training process implemented within a system with a CPU @2.70GHz-2.71 GHz, 8Go RAM, and the performance of each batch size at a given scenario. Figure. 4-13 and 4-14, illustrate the impact of the batch size on the performance of the model in terms of the RMSE percentage. It is observed, when smaller sequences are used, the elapsed time is significantly large, and the RMSE value increases. On the other hand, when larger sequences are used, the elapsed time, during the learning process, is very small compared to the other ones and the RMSE values decrease. Thus, the optimal input sequence size satisfies a better RMSE and a lower elapsed time during the training process. However, the small value of the RMSE is noticed at the batch size of 200 as seen in Figure. 4-15. (c) in the third scenario. Therefore, for the sake of generalization and high performance in all the environments, the value of 100 is chosen as an optimal batch size for the proposed model.
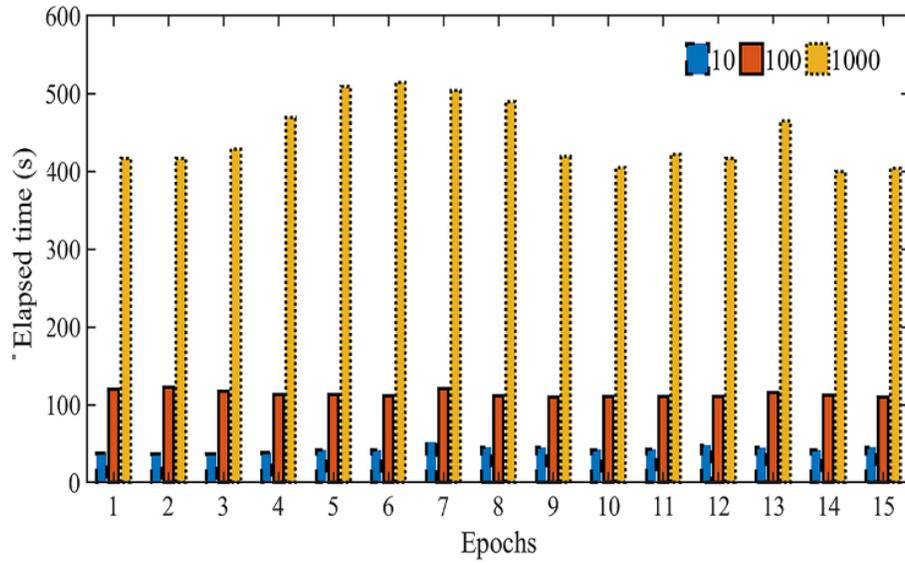
Figure 4-13 Elapsed times within smaller and larger batch size
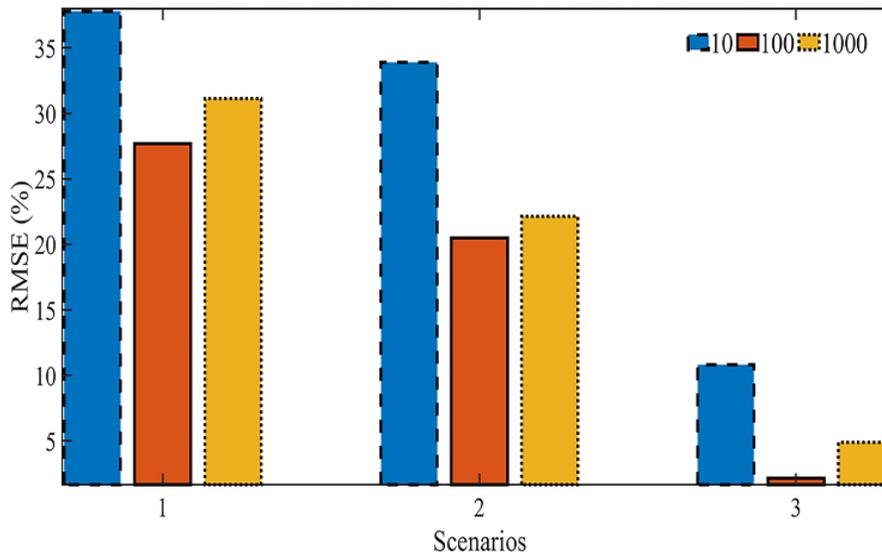
Figure 4-14 RMSE evaluated within smaller and larger batch size
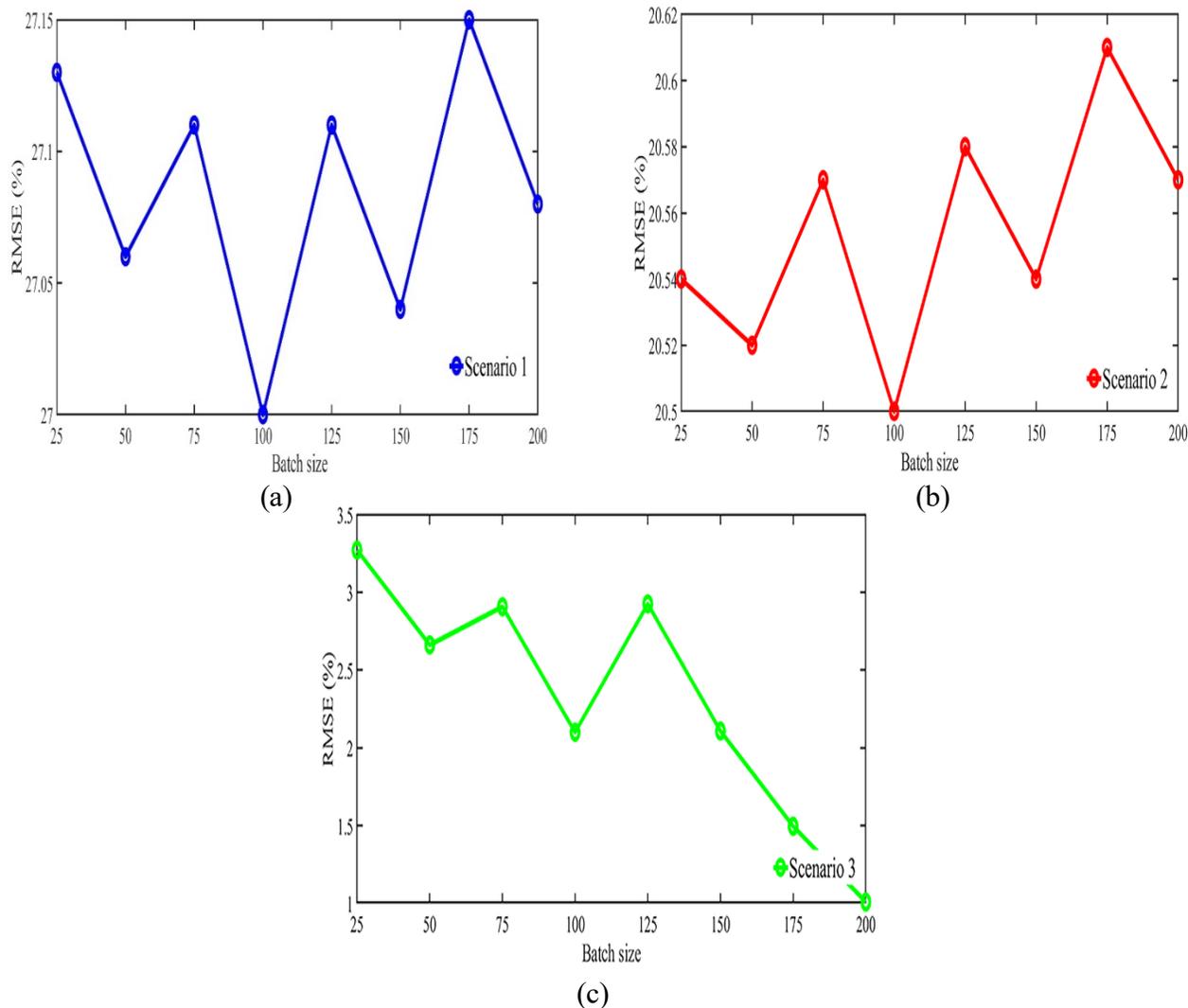
Figure 4-15 Model performance in terms of batch size:

(a) Scenario 1, (b) Scenario 2, (c) Scenario 3.

To validate the efficiency of the proposed prediction scheme, we compare the proposed model performances to a published related work in channel prediction in the SISO-LOS scenario. Table 4-5, demonstrates a comparison of the proposed model performance with published ones in terms of some metrics used in their developments such as Normalized Mean Square Error (NMSE) [21] in both linear and decibel values. Different datasets and techniques were used. Therefore, the LC which is considered an efficient technique for performance evaluation [23], remains unverified in their models to give more consistency to the reported models' accuracy.

Table 4-5 Model performance comparison

Published models	Ref	Metrics	Values
LSTMs	[1]	NMSE	0.0878
ConvLSTM	[2]	NMSE	Up to -37.84 dB
Autoregresion+LSTMs	[3]	RMSE	Up to 2.26%
CNNs + LSTMs	[4]	MSE	0.551
Proposed model	N/A	NMSE	Up to 8.67e-05 Up to -40.61 dB
		MSE	Up to 4.6e-04
		RMSE	Up to 2.1%

4.4.5 Correlations properties

In this section, the spatial autocorrelations, as well as Pearson's correlation (test statistics), are evaluated to measure the statistical relationship between two continuous datasets (variables) [26]. The measured datasets and the predicted sequences, from all the environments, were processed to estimate the correlations properties. The Pearson correlation coefficient is given by the following equation [27]

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \, \sigma_Y} \tag{4.6}$$

Where X, Y are the pair of continuous variables, $\text{cov}(X, Y)$ is the covariance, $\sigma_X \, \sigma_Y$ are the standard deviation of X and Y, respectively. The pair of variables represents the measured and the predicted sequences for both antenna combinations. Figure 4-16 shows the matrix correlation which is used to investigate simultaneously the multiple variables dependence. In the diagonal, the distribution of each variable is shown. At the bottom of the diagonal, the bivariate scatter plots with a fitted line, are displayed. At the top of the diagonal, the values of the Pearson correlation coefficients between the data of all scenarios are shown. The predicted CFR undertake almost the same

distribution as the measured ones in all scenarios, as shown in the matrix diagonal. A linear fitting is observed between the measured and the predicted data in each scenario. Hence, a very strong relationship of 0.99 is noticed between the measured and the predicted CFR, within any environment. Therefore, it is perceived a strong correlation estimated at 0.92 and linear fitting between the measured CFR in the first and the second environments. Moreover, this strong relationship makes a CFR prediction in the second environment easier for the model which is validated by the lower RMSE. On the other hand, the data from the third scenario shows a moderate linear fitting. The correlation is evaluated at 0.52 and 0.49, with the measured CFR within the first and the second scenario, respectively. Consequently, the prediction of this data is not easier for the model to evaluate, especially, when it does not have enough prior information (not used during the training). The proposed model shows the strength to acquire the general pattern of its data. Therefore, as mentioned in the LC, the model is well learning during the training process, then, through the evaluation process, the model validates more important patterns in the measured dataset. Moreover, the efficiency of using the LSTM networks which can remember the most important information in the datasets [28], is shown by the moderate correlation between the data from the first (used during the training process) and third environment. In addition, some of the CFR data in the first and third environment share a linear relationship which was enough for the model to predict the behaviour of the CFR in the third complex environment. As shown in the correlation matrix, it can be noticed that the proposed model learns the trend of the CFR within any environment. Furthermore, it is well known from previous studies [29,30], that the CFR relatively decreases away from the transmitter within any frequency band. As well, the fact that the impact of the radiation pattern of the antenna is crucial to channel propagation, the model indicates that it was necessary to obtain the antennas information during the training phase. It, therefore, empowers the prediction of the datasets from the second and the third environment where the same antenna characteristics were shared, such as a patch antenna (it was used in all the environments).
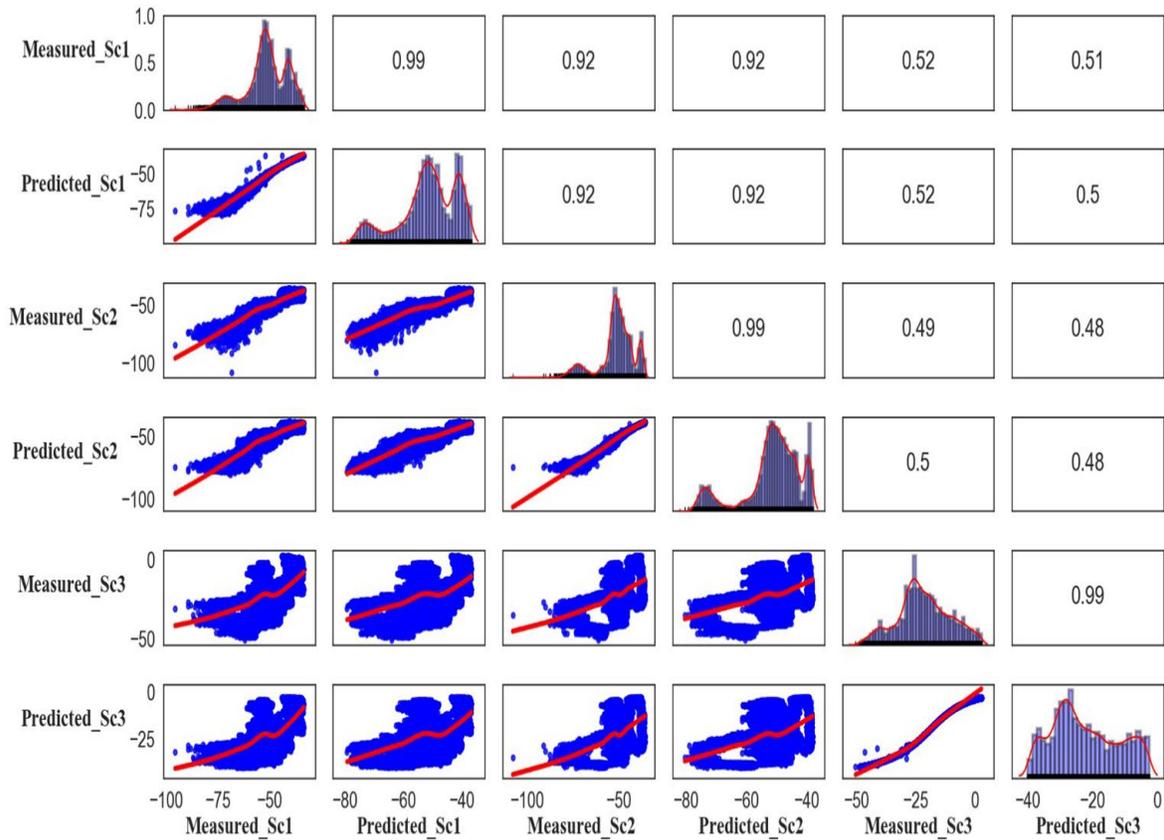
Figure 4-16 Matrix correlation of measured and predicted CFR.

4.5 Channel characterization

After the validation of the proposed model, channel characterization is carried out using the predicted and measured CFR to compare the efficiency of the model to reproduce the same behaviour. The measured and predicted characteristics of the channel within different scenarios in terms of Path Loss (PL), RMS delay spread, coherence bandwidth, k-factor, and channel capacity are shown in this section

4.5.1 Path Loss

The PL is the attenuation in the transmitted signal caused by the effects of the environment. It was obtained by applying the equations 2.3 and 2.4   As shown in Figure. 4-17, 4-18 and 4-19, the model predicted the path loss values for all antenna configurations. The path loss exponent can be

evaluated from the linear analysis, within the first, second and third scenarios as illustrated in Table 4-6.

Table 4-6 Path loss exponent

Path Loss exponent β				
Scenarios	Antennas	D vs. D	D vs. Prx	P vs. P
Scenario 1	Measured	3.1595	2.0533	N/A
	Predicted	2.8717	2.1388	N/A
Scenario 2	Measured	2.9795	2.5417	N/A
	Predicted	2.8042	2.6882	N/A
Scenario 3	Measured	N/A	N/A	1.5353
	Predicted	N/A	N/A	1.3553

(a)

(b)
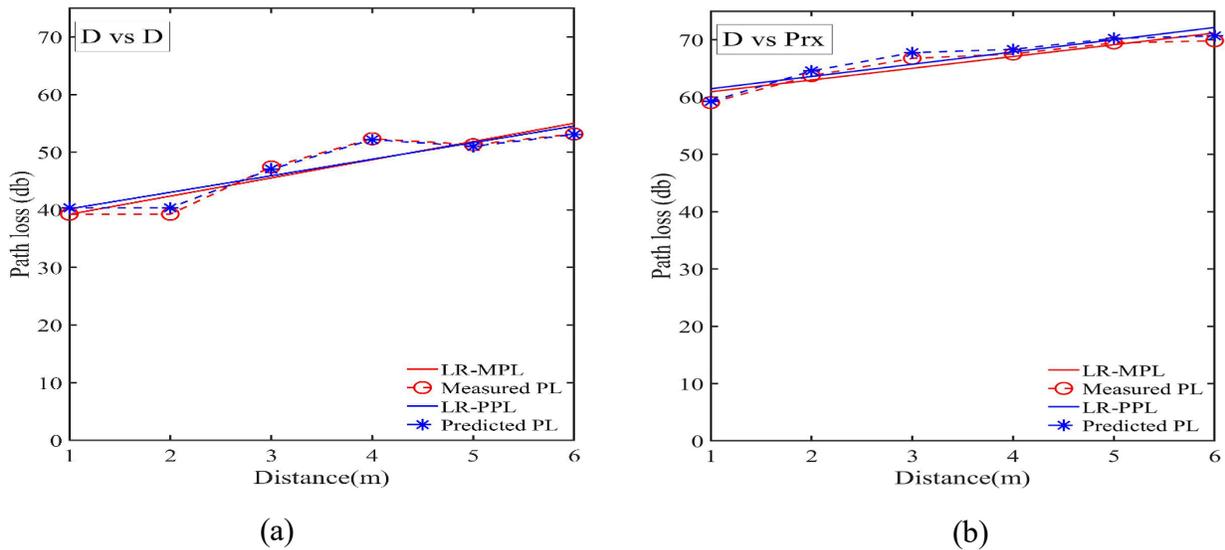
Figure 4-17 Measured PL (MPL), predicted PL (PPL) and Linear Regression (LR) for scenario 1: (a) D vs D, (b) D vs Prx.
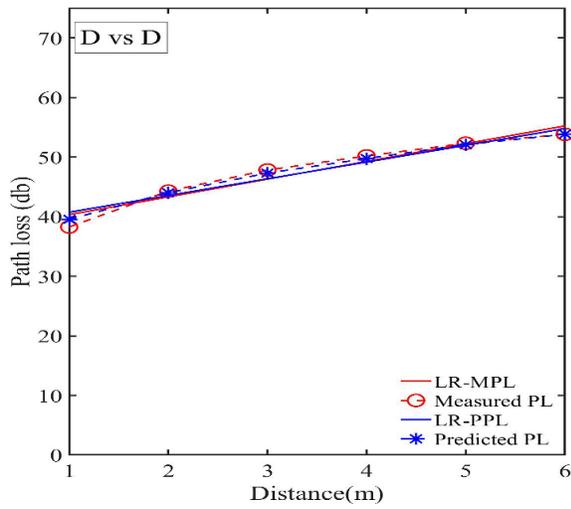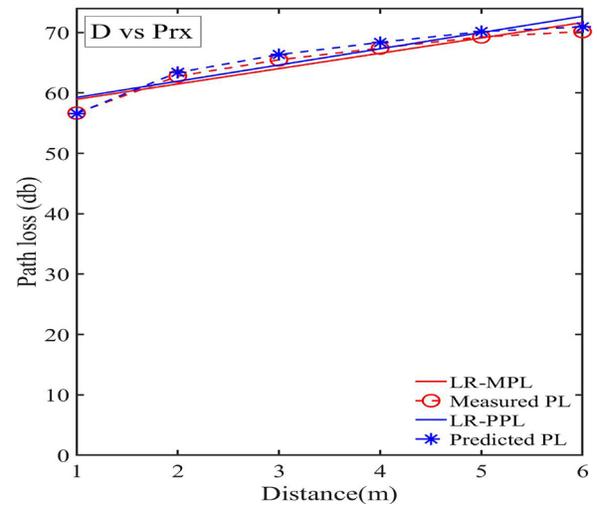
Figure 4-18 Measured PL (MPL), predicted PL (PPL) and Linear Regression (LR) for scenario 2:
(a) D vs D, (b) D vs Prx.
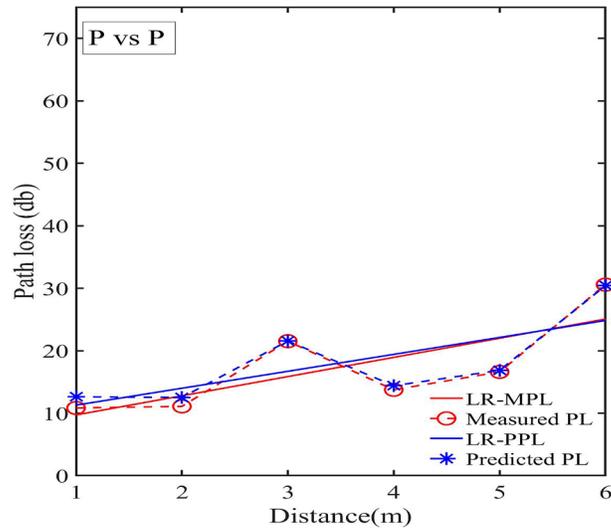
Figure 4-19 Measured PL (MPL), Predicted PL (PPL) and Linear Regression (LR) for scenario 3

4.5.2 RMS delay spread and coherence bandwidth

RMS delay is the time dispersive property of wideband multipath channel [29,30], and the coherence bandwidth is considered as the range of frequencies over which the channel can be considered flat The RMS delay spread is equal to the square of the second central moment of PDP, and it is calculated using the formula reported in equations 2.7 to 2.9 [30]. The coherence bandwidth was evaluated for the 50% correlation using the RMS delay spread such as mentioned in equation 2.10. Figure. 4-20, 4-21 and 4-22, show the RMS delay spread calculated using the measured and the predicted power delay profile which is obtained by the CIR. The results show that the RMS delay spreads are between 0 ns and 5 ns and the coherence bandwidth (at 50% correlation) belongs to the range of 8–20 MHz for D vs. D in all environments. RMS delay spreads are higher (between 5 ns -20 ns) and the coherence bandwidth lower for D vs. Prx in all environments. On the other hand, the result of the RMS delay spread is variable at different distances in underground mine environments. This is due to the specific geometry of the mine. Moreover, it is observed that the same observation could be made with the predicted results.

(a)   (b)
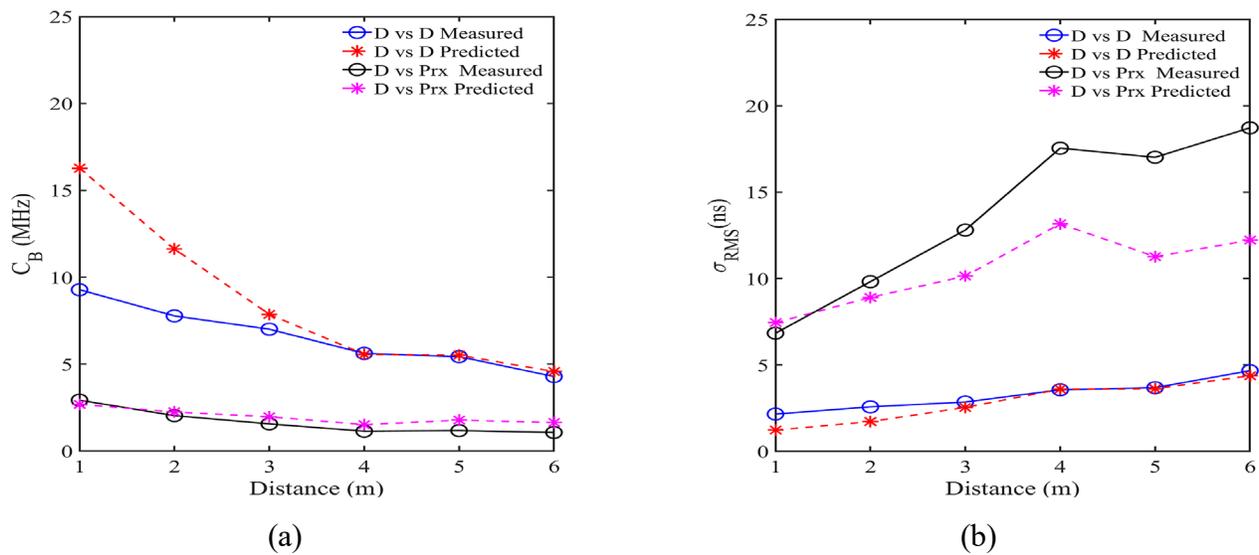
Figure 4-20 RMS delay spread and coherence bandwidth for scenario 1:
(a) Coherence bandwidth (b) RMS delay spread
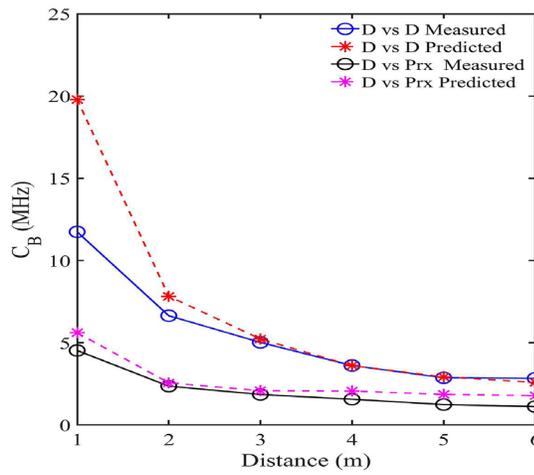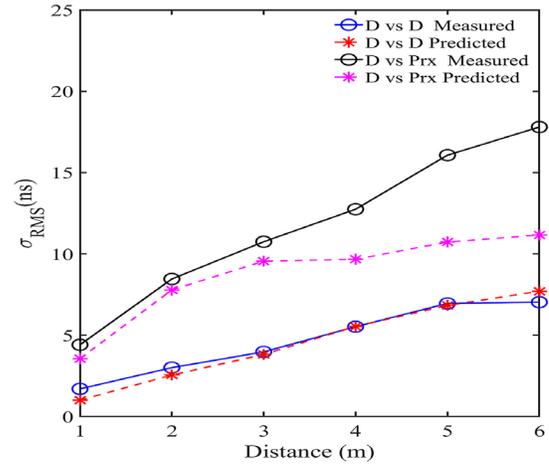
(a)   (b)

Figure 4-21 RMS delay spread and coherence bandwidth for scenario 2:

(a) Coherence bandwidth (b) RMS delay spread
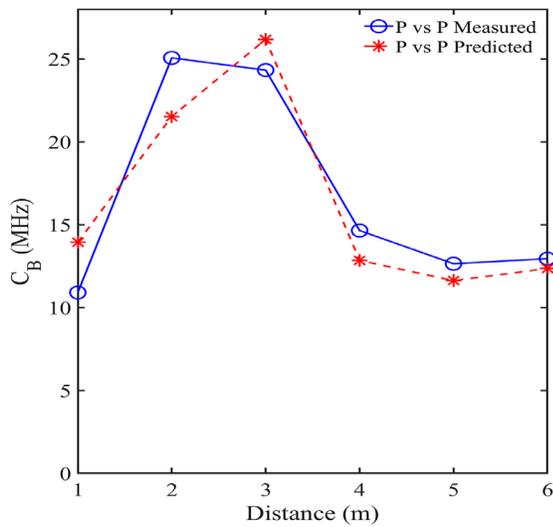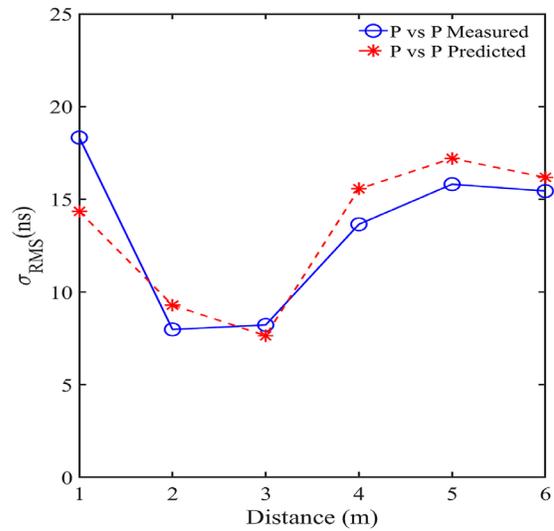
(a)   (b)

Figure 4-22 RMS delay spread and coherence bandwidth for scenario 3:

(a) Coherence bandwidth (b) RMS delay spread

4.5.3 Rician k-factor

The Rician factor is determined from the impulse response as described in chapter 2 equation 2.6 [31]. Since the measurements involved in the LOS scenarios for different path lengths, the K-factor (denoted in Figure 4-23) varied with the antenna configurations and the distance Tx-Rx. The K-factor for the D vs. D antenna configuration is clearly higher than that of a D vs. Prx configuration due to the strong dominant components in the first and second environments. Moreover, the K-factor values in the third environment are low. This is due to the roughness and randomness of the surfaces in the underground mines which produces multipath richness. Hence, the proposed model predicted the same as the measurement that the K-factor is higher in the first and second environments within D vs. D antenna configuration, and lower in underground mine environment with P vs. P antenna configuration. Similarly, the measured and the predicted values are between -20 dB and 2 dB within all the scenarios, respectively. Moreover, the predicted and the measured K-factor values are decreasing while the distance is increasing in all the environments.
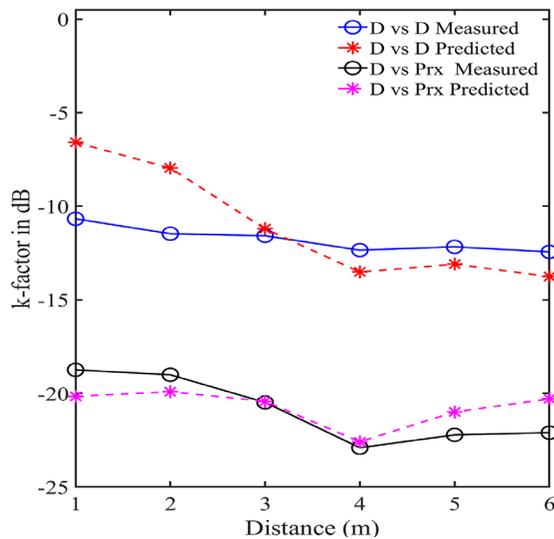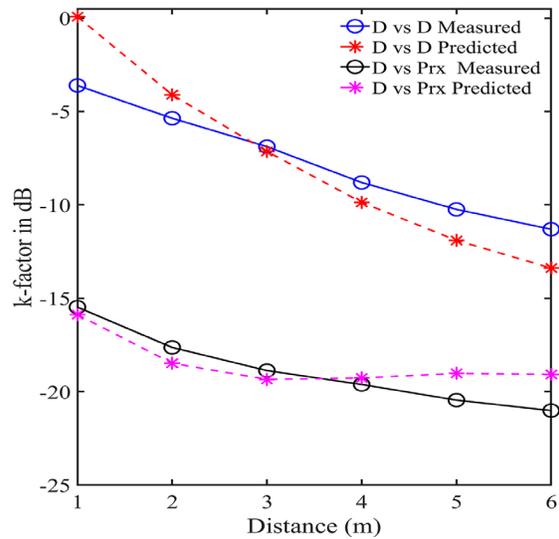
(a)   (b)
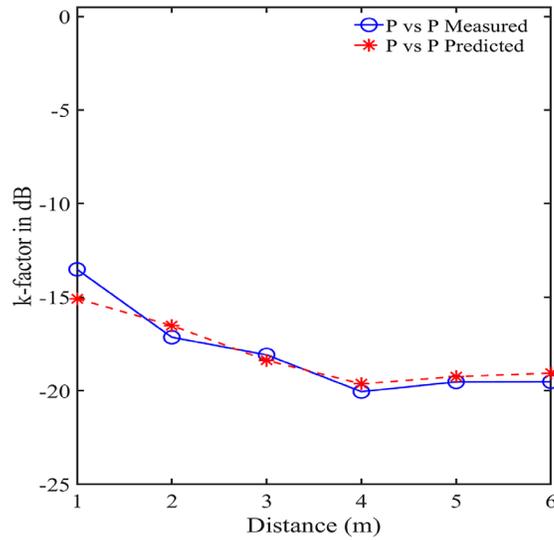
(c)

Figure 4-23 Rician K-factor :

(a) Scenario 1, (b) Scenario 2 (c) Scenario 3

4.5.4 Channel capacity

In this section, the channel capacity is evaluated using both measured and predicted channel impulse responses. The SISO channel capacity is extracted from the measurement by using Shannon's capacity equation [32]

$$C_N \left[\frac{\text{bps}}{\text{Hz}}\right] = \log_2(1 + \rho|H|^2) \qquad (4.7)$$

H is the normalized CIR and ρ is the signal-to-noise ratio. Figure. 4-24, 4-25 and 4-26, show the channel capacity results at SNR of 10 dB, for all scenarios. As expected, the predicted and the measured average capacity decreases with distance for all antenna configurations. This is explained by the fact that small distances correspond to a higher received power in both environments using the same antenna configurations. As seen, the channel capacity, extracted by the predicted data, perfectly expresses the same conclusion as the measured ones.
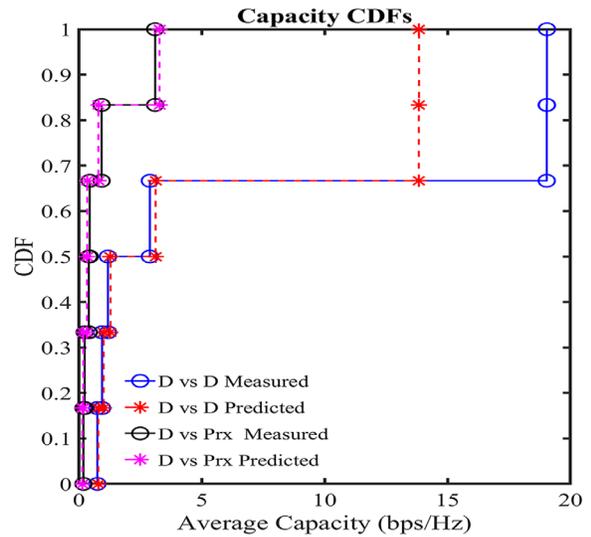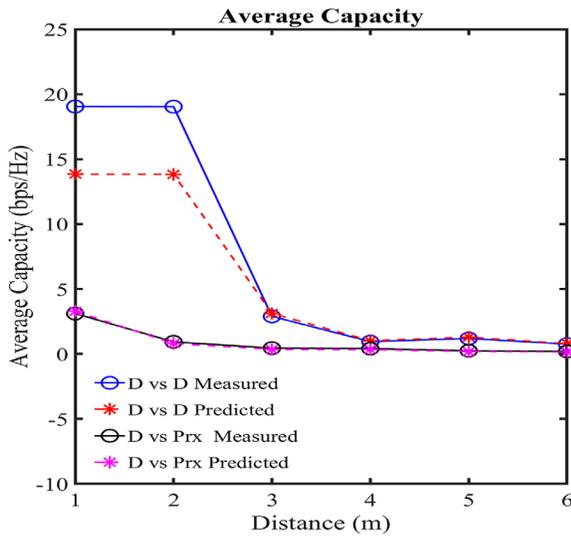
Figure 4-24 Channel capacity at SNR =10dB for scenario 1:

(a) Average capacities (b) Capacity CDFs
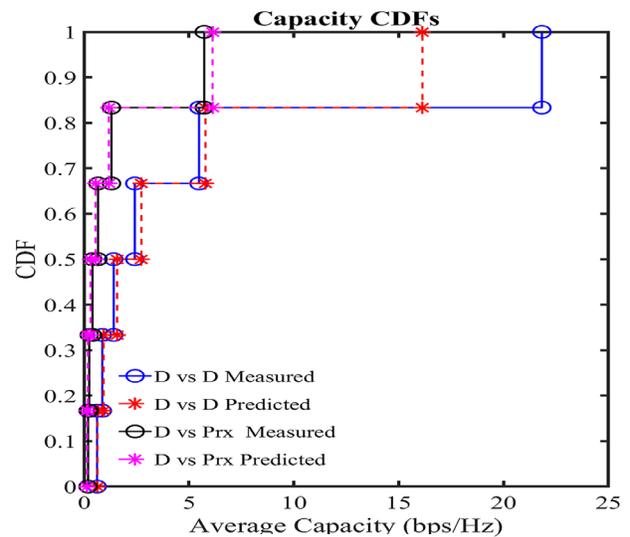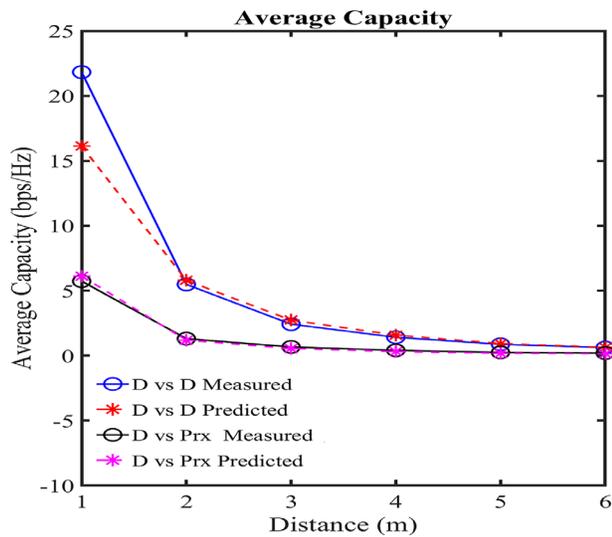
Figure 4-25 Channel capacity at SNR =10dB for scenario 2:

(a) Average capacities (b) Capacity CDFs
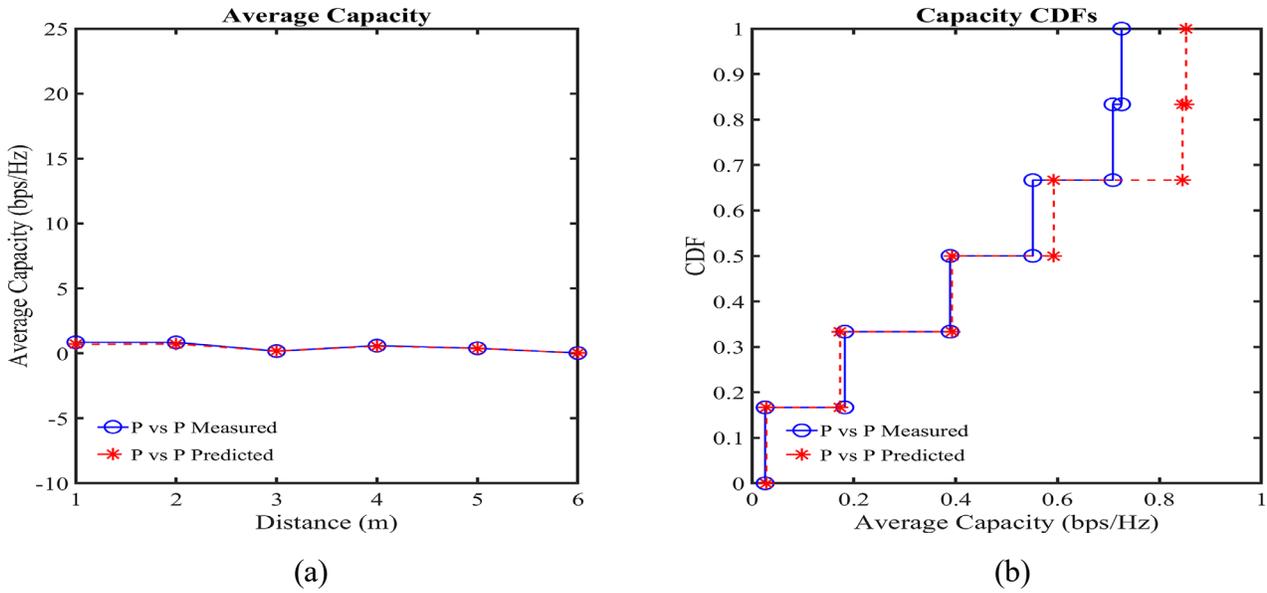
Figure 4-26 Channel capacity at SNR =10dB for scenario 3:

(a) Average capacities, (b) Capacity CDFs

4.6 Conclusion

In this chapter, we presented an efficient propagation channel modeling based on LSTM networks combined with a linear layer in SISO-LOS. A CFR prediction was performed under different environments and frequency bands, where different antenna configurations were used. The CFR prediction results in scenario 1, 2 in indoor corridor environment and scenario 3 in underground mine environment, showed low MSE, NMSE, and RMSE metrics in comparison with that of the literature models. Although only the data of one environment was considered for the training, the model achieved better performance in all environments. This is due to the improved data processing which is justified by the good fit learning curve during the training process and the model's ability to learn the important channel parameters between the datasets. Channel characterization was carried out afterward, to visualize the model's capability to procreate same behavior of the channel and to demonstrate that the same observations could be made. As result, the model significantly Predicted (PPL) the Measured Path Loss (MPL), RMS delay spread, coherence bandwidth, K-factor, and channel capacity.

The methodology demonstrated its use in forecasting the CFR, considering the antenna variety, where the same discussions could be held as the measured ones. Hence, the proposed model could promote the design of wireless communication systems in any complex area.

4.7 REFERENCES

[1] X. Tong and S. Sun, "Long short-term memory network for wireless channel prediction," in *International Conference On Signal And Information Processing, Networking And Computers*, 2017: Springer, pp. 19-26.

[2] J. Wang, Y. Ding, S. Bian, Y. Peng, M. Liu, and G. Gui, "UL-CSI data driven deep learning for predicting DL-CSI in cellular FDD systems," *IEEE Access,* vol. 7, pp. 96105-96112, 2019.

[3] J. Joo, M. C. Park, D. S. Han, and V. Pejovic, "Deep learning-based channel prediction in realistic vehicular communications," *IEEE Access,* vol. 7, pp. 27846-27858, 2019.

[4] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5G wireless communications: A deep learning approach," *IEEE Transactions on Network Science and Engineering,* 2018.

[5] H. L. Bertoni, *Radio propagation for modern wireless systems*. Pearson Education, 1999.

[6] S. R. Saunders and A. Aragï, *Antennas and propagation for wireless communication systems*. John Wiley & Sons, 2007.

[7] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Communications Surveys & Tutorials,* vol. 21, no. 3, pp. 2224-2287, 2019.

[8] F. Li, M. Zhang, G. Fu, T. Qian, and D. Ji, "A bi-lstm-rnn model for relation classification using low-cost sequence features," *arXiv preprint arXiv:1608.07720,* 2016.

[9] S. Li, Q. Wang, X. Liu, and J. Chen, "Low cost LSTM implementation based on stochastic computing for channel state information prediction," in *2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, 2018: IEEE, pp. 231-234.

[10] A. J. Anderson, "Channel prediction in wireless communications," 2015.

[11] R. Ge, F. Huang, C. Jin, and Y. Yuan, "Escaping from saddle points—online stochastic gradient for tensor decomposition," in *Conference on learning theory*, 2015: PMLR, pp. 797-842.

[12] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, 2016: IEEE, pp. 324-328.

[13] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," *arXiv preprint arXiv:1704.02971,* 2017.

[14] D. Zhang and M. R. Kabuka, "Combining weather condition data to predict traffic flow: a GRU-based deep learning approach," *IET Intelligent Transport Systems,* vol. 12, no. 7, pp. 578-585, 2018.

[15] S. S. Joshi, "Calibrating recurrent sliding window classifiers for sequential supervised learning," 2003.

[16] V. A. Fono, "Étude de la propagation radio en environnement doté de surfaces irrégulières à profil périodique," Université du Québec en Outaouais, 2018.

[17] S. Patro and K. Sahu, "Normalization: A preprocessing stage. arXiv 2015," *arXiv preprint arXiv:1503.06462.*

[18] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*: Springer, 2010, pp. 177-186.

[19] D. Masters and C. Luschi, "Revisiting small batch training for deep neural networks," *arXiv preprint arXiv:1804.07612,* 2018.

[20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.

[21] A. Botchkarev, "Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology," *arXiv preprint arXiv:1809.03006,* 2018.

[22] M. J. Anzanello and F. S. Fogliatto, "Learning curve models and applications: Literature review and research directions," *International Journal of Industrial Ergonomics,* vol. 41, no. 5, pp. 573-583, 2011.

[23] J. Brownlee, *Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better Predictions*. Machine Learning Mastery, 2018.

[24] M. D. Zeiler, "Adadelta: an adaptive learning rate method," *arXiv preprint arXiv:1212.5701,* 2012.

[25] W. Jiang and H. D. Schotten, "Deep learning for fading channel prediction," *IEEE Open Journal of the Communications Society,* vol. 1, pp. 320-332, 2020.

[26] M. M. Mukaka, "A guide to appropriate use of correlation coefficient in medical research," *Malawi medical journal,* vol. 24, no. 3, pp. 69-71, 2012.

[27] A. King and R. Eckersley, *Statistics for Biomedical Engineers and Scientists: How to Visualize and Analyze Data*. Academic Press, 2019.

[28] F. A. Gers and E. Schmidhuber, "LSTM recurrent networks learn simple context-free and context-sensitive languages," *IEEE Transactions on Neural Networks,* vol. 12, no. 6, pp. 1333-1340, 2001.

[29] J. B. Andersen, T. S. Rappaport, and S. Yoshida, "Propagation measurements and models for wireless communications channels," *IEEE Communications Magazine,* vol. 33, no. 1, pp. 42-49, 1995.

[30] T. S. Rappaport, *Wireless communications: principles and practice*. prentice hall PTR New Jersey, 1996.

[31]   P. Tang, J. Zhang, A. F. Molisch, P. J. Smith, M. Shafi, and L. Tian, "Estimation of the K-factor for temporal fading from single-snapshot wideband measurements," *IEEE Transactions on Vehicular Technology,* vol. 68, no. 1, pp. 49-63, 2018.

[32]   I. B. Mabrouk, L. Talbi, M. Nedil, and K. Hettak, "MIMO-UWB channel characterization within an underground mine gallery," *IEEE Transactions on Antennas and Propagation,* vol. 60, no. 10, pp. 4866-4874, 2012.

CHAPTER 5   CHANNEL PREDICTION FOR INDOOR ENVIRONMENT IN MIMO SYSTEMS

5.1 Introduction

As mentioned in the chapter 4, several DL algorithms have been developed to channel characterization SISO and MIMO channels. However, the studied models are validated experimentally with extensive measurements, where high number of the data is collected. In this area, the Deep-Learning algorithms (DL) were applied to outdate the statistical models and bring new applications such as localization [1] and channel prediction accuracy [2]. The DL capability to model the nonlinear problems brought a great interest in the channel propagation research field [3]. It was introduced to predict SISO channels using LSTM. *Ding et al* [4] used a complex-valued neural network to forecast frequency domain channel characteristics, while *Jiang et al* [5] provided a novel MIMO channel predictor built on a deep recurrent neural network that incorporates LSTMs or Gated Recurrent Units (GRUs) memory cells. Moreover, simulations in terms of prediction accuracy in multi-antenna flat-fading channels are also proposed in [4]. Convolution Neural Networks combined with Recurrent Neural Networks design (CNN-RNN) to predict CSI have been proposed in [6-8]. *Arnold et al* [9] investigated the feasibility of DL algorithms for MIMO configuration based on the Orthogonal Frequency Division Multiplex (OFDM). Dense layers were used with two-step training strategies to predict the NLOS position. Even though, these techniques are considered as efficient tools to apply the DL algorithm for CSI prediction and LOS-NLOS detection. It required high computational complexity especially for harsh environments. Therefore, a new efficient technique is effectively needed which leads to the proposed work for underground mine environments. In this contribution, a stacked model (SM) based on LSTMs deep learning networks is proposed to predict the WBAN-MIMO channels in harsh environments. The model is validated with extensive WBAN-MIMO channel measurements published by *Elazhari et al* [10]. The SM is composed of three parallel models, where only one input data is used. Therefore, the collected magnitude and phase of the Channel matrix were used in the SM to simultaneously predict the receiver position, the LOS-NLOS scenario and the channel matrix (H) as illustrated in Figure 5-1 and 5-2.

To the best of the authors "knowledge", no such algorithms were used to model the WBAN-MIMO systems in underground mines environments. The novelty of this study consists of the following: First, a new model of channel matrix is proposed for underground mines. Second, the SM framework is designed to predict the channel matrix H, the position and the LoS-NLoS identification from only one input. Third, the model is evaluated for different antenna setups, such as linearly polarized (Lin) patch antennas MIMO system, circularly polarized patch antennas MIMO system, co-polarized (CP) and 90 degrees rotated (90 deg) configurations. The chapter is organized as follows: Section 5-2 briefly introduces the channel matrix prediction. Then, the measurement procedure used to collect the experimental data is described in Section 5-3. Afterwards, the proposed model is detailed and implemented in Section 5-4. The model validation is addressed in section V. Finally, the chapter is concluded in Section 5-5.
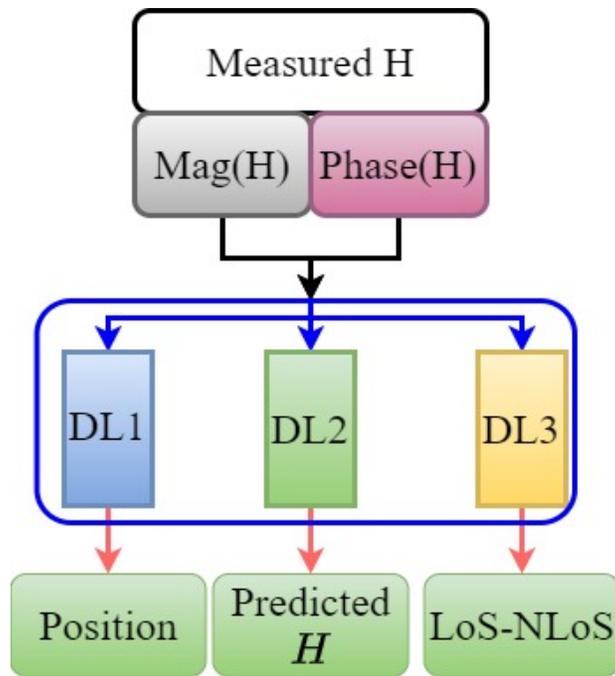
Figure 5-1 The proposed stacked model scheme
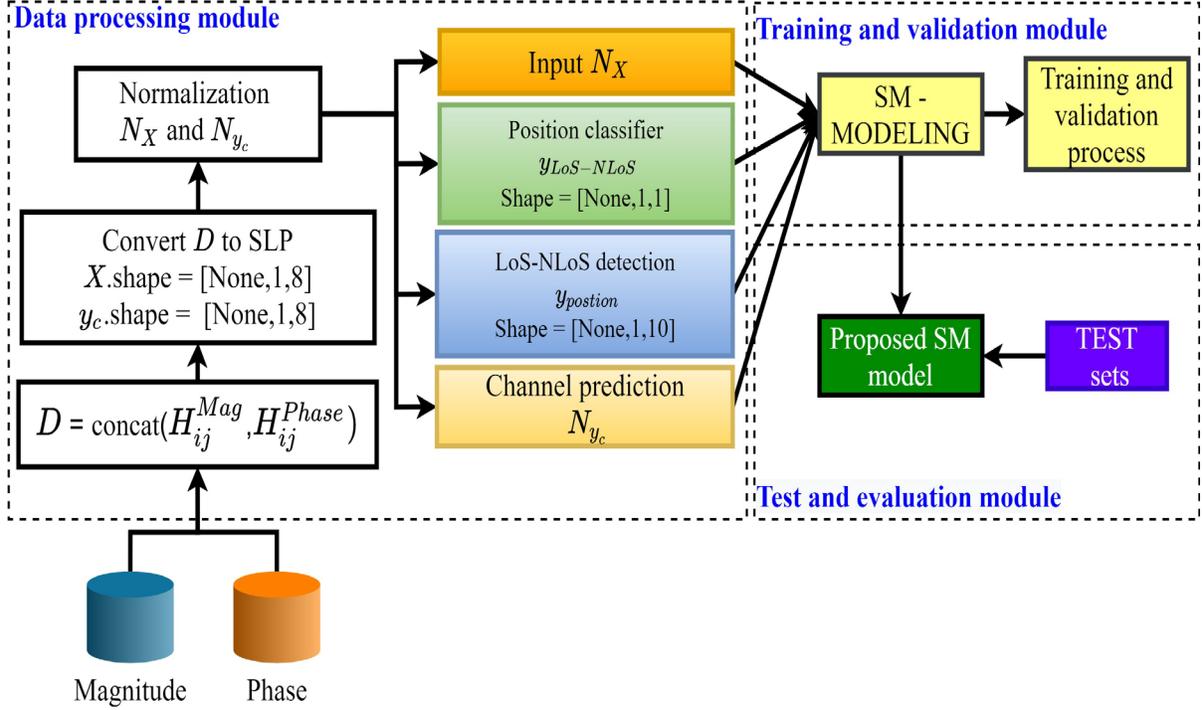
Figure 5-2 The framework for stacked model prediction.

5.2 Stacked model prediction scheme

In this section, a stacked model prediction scheme is presented. First, the channel prediction framework is introduced where it is expressed as a regression problem. Then, the position and the LoS-NLoS classification scheme are presented. The SM is validated experimentally in 2X2 MIMO system with frequency band 2.3 GHz - 2.5 GHz [10]. In MIMO systems, an $m_t \times n_t$ transfer channel matrix $\boldsymbol{H}$ is created, which represents the complex sub-channel gains from the m_t transmitting to the n_t receiving antennas [11]. The channel matrix $\boldsymbol{H}$ for a 2X2 MIMO system is expressed as [11, 12]:

$$\boldsymbol{H} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \tag{5.1}$$

Where H_{ij} represents the complex sub-channel gain from the ith transmitting antenna to the jth receiving antenna. Indeed, *Elazhari et al* [10] measurements' campaigns were conducted to estimate the matrix H in underground mine using different antennas types, polarizations and link configurations (LoS-NLoS) in body-to-body (B2B) scenarios. The channel prediction framework is developed using an Encoder-Decoder algorithm [13-16] based on LSTM networks combined

with time-distributive linear layer. The model is performed to simultaneously predict each element of the matrix H at each measured position with considering the magnitude and the phase. Using the same input as channel prediction scheme, the position and the LoS-NLoS classifier established respectively on a combination of the LSTM network and the dense layers. Therefore, the categorical and the binary classifications are considered for the position and the LoS-NLoS scenarios, respectively. Similarly, same as the LSTM architecture used in chapter 3, the LSTM I is given using the relevant mathematical formulas in equations 2.27 to 3.32 and 3.40.

5.2.1 Measurement Procedure

The measurement procedure was carried out in a real gold mine (located in Val d'Or city in northern Quebec) within a gallery at 90 meters underground with a width and a height of 4 m and 2.45 m, respectively (Figure 5-3). The underground mine environment is characterized of rough, random surfaces and non-uniform gallery dimensions. The measurements were performed in B2B-configuration as reported in [10]. Two antenna configurations are considered, namely, co-positioned (CP) and 90 degrees rotated antenna (90deg) systems. The different antenna configurations are listed in Table 5-1.

Table 5-1 Measurements scenario

Polarization	Circular (Cir)			Linear
Scenario	CP-LoS	90 deg-LoS	CP-NLoS	CP-LoS
Position	10	10	9	7

The measurement system setup consists of: Vector Network Analyzer (VNA), power amplifier and low noise amplifier connected to the transmitter (Tx) and the receiver (Rx), respectively. The VNA is used to measure the channel frequency response magnitude and phase, in the desired frequency range. During measurements, the propagation channel is considered as stationary for both scenarios. The Tx was placed at a fixed position, and the Rx location was changed up to 10 m away from Tx, as shown in Figure 5-4. The measurement parameters are described in Table 5-2.

Figure 5-3 Photo of the gallery
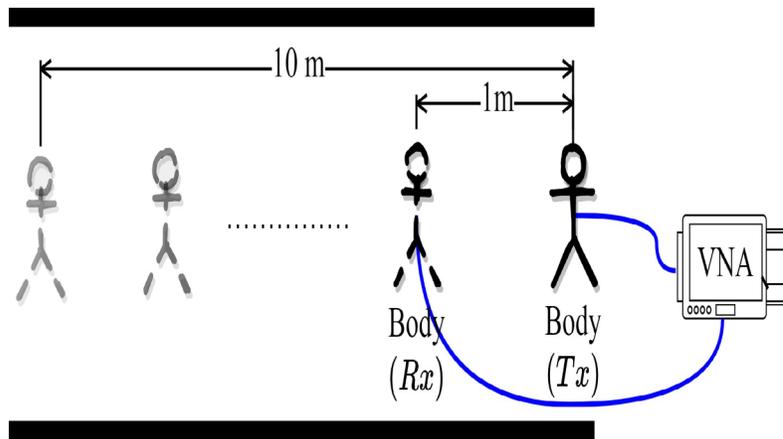
(a)
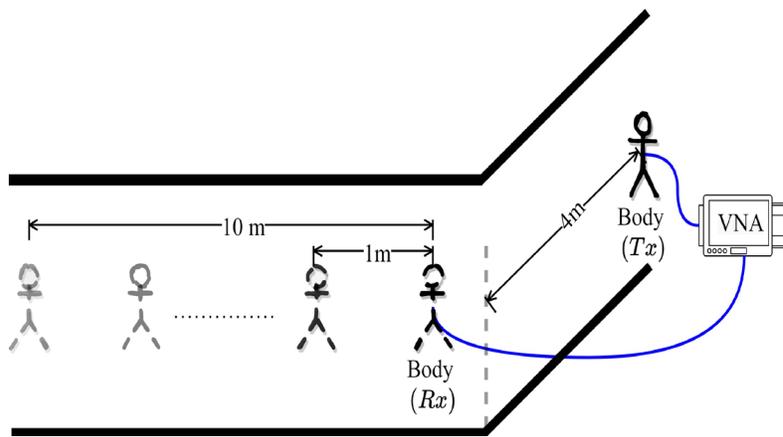
(b)

Figure 5-4 Experimental scenarios (a) LoS, (b) NLoS

Table 5-2 Measurement system configuration

Parameters	Values
Frequency	2.3 GHz - 2.5 GHz
Transmitted power	-10 dBm
Average noise floor	-80 dBm
Bandwidth	200 MHz
Rx gain	6.6 dBi
Tx gain	6.6 dBi
Cable loss	0.6 dB/m
Antenna height (B2B)	1.50 m
Sweep time	60 s
Antenna types	Linearly and circularly polarized patches

5.2.2 Data processing

The data processing is illustrated by the diagram of Figure 5-5. Following the experimental scenarios, the channel matrix magnitude and the phase measurements were carried out for 10 different positions. At each position, ten snapshots with 2049 samples were measured from 2.3 GHz to 2.5 GHz. Moreover, the snapshots have been horizontally concatenated to provide a dataset for the training and the validation process.

$$D = concat\left(H_{ij}^{Mag}, H_{ij}^{Phase}\right) \tag{5.2}$$

Where $i \in (1,2)$; $j \in (1,2)$ represents the transmitted and the received link. The framework uses the dataset D as a multivariate (Multiple-Variables) magnitude and phase datasets. Then, to predict

the four matrix sub-channels at each position, the sequences are divided into input and output samples. To achieve the model high accuracy in terms of classification and regression, the concept of the sliding window is adopted [17]. As the nature of the collected data are time-series data, the concept of sliding window is used to convert the problem into supervised learning one. Thus, this method resolves the prediction problem for sub-channel prediction. Specifically, regression and classification problems are simultaneously used in the stacked model. The model input is the dataset X and the output (y_c) is only used for the channel prediction in case of regression problems. For the classification problems, different outputs are used. A binary output for LoS-NLoS $(y_{LoS-NLoS})$ detection is created and a categorical output $(y_{position})$ with a matrix of ten features is created. However, all the target outputs are 3-dimensional size.

From the input shape of Figure 5-5, it can be noticed that 8 parameters must be considered in the prediction along with two variables (magnitudes and phases) for each sub-channel matrix. In fact, it is complicated to put the three stacked model to converge and achieve high prediction accuracy in both regression and classification. The Z-score normalization [18] standard has to be applied in order to shed the values between variables without losing the relevant information of the model. The Z-score normalization is described as follows [18]:

$$N_X = \frac{X - \mu_X}{\sigma_X} \tag{5.3}$$

$$N_{y_c} = \frac{y_c - \mu_{y_c}}{\sigma_{y_c}} \tag{5.4}$$

Where μ, σ are the mean and the standard deviation, respectively. Moreover, the input data is transformed into small batch size sequences to facilitate the learning process and to avoid the gradient descends problems [19, 20].
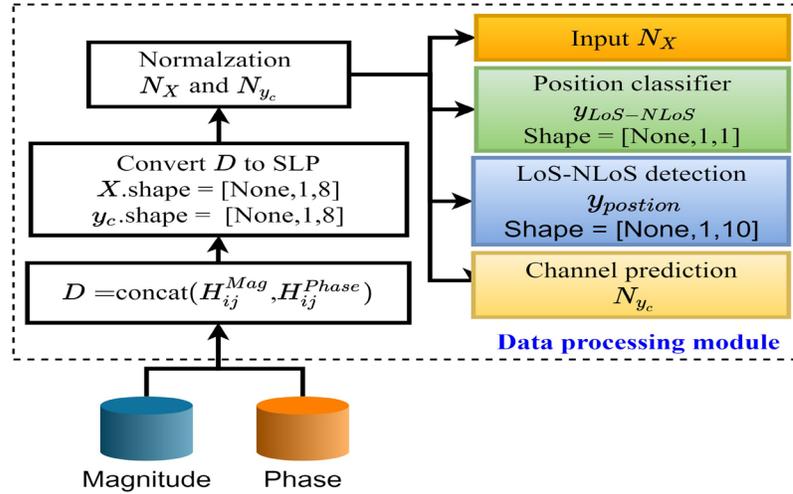
Figure 5-5 Data processing module

5.3 Train and validation mechanism

The diagram of Figure 5–6 shows the high-level process of the proposed model. The stacked model was trained in the collected channel matrix H. Afterwards, the model is evaluated to test the model capability to classify the position, the LoS-NLoS identification and the channel prediction. Finally, using the prediction channel H, the channel characterization and modeling were compared to the model results of the published ones [10]. After the data processing scheme, the 1/3 and 2/3 of the normalized datasets are split into validation and training datasets, respectively. The target datasets are different for classification and regression problems. In this case, binary output $(y_{los/NLoS})$ is used for LoS-NLoS detection where the true value (one) is for LoS scenario, and the false value (zero) is for NLoS. As illustrated in Figure. 5–7, categorical classification matrix, for position classification output $(y_{position})$ is used. Up to ten positions were collected in the experimental measurements. Finally, for channel prediction, the output N_{y_c} is used. Therefore, for the stacked model outputs, different losses are considered in Figure 5–8. For binary classification, Binary Cross Entropy loss (BCE) [21] is assigned, where it is compared to each of the predicted probabilities to the actual class output (LoS [one] or NLoS [zero]) [21, 22]. Before using the BCE, a sigmoid function is applied independently to each element x_i of the vector x in the last layer to squash the vector data range between 0 to 1. It is described in equation 3.2 [23, 24] .On the other hand, for the position classifier, different losses are considered in the multiple-class classification.

However, Categorical Cross Entropy loss (CCE)[21] is selected, which has the same performance as BCE. In this case, a SoftMax activation function is considered to calculate the probability of each target class overall potential target classes. The probabilities obtained will be useful in defining the target class for the provided inputs [25, 26]. In fact, the output probability range is the key benefit of adopting SoftMax. The probability range will be from 0 to 1, and the total of all probabilities will be one. When the SoftMax function is used to a multi-classification model, the probability of each class is returned, with the target class having the highest one (probability) [54–57]. Therefore, the SM model used the SoftMax in case of position classification. The SoftMax formula is given in equation (3.3). [23-26]
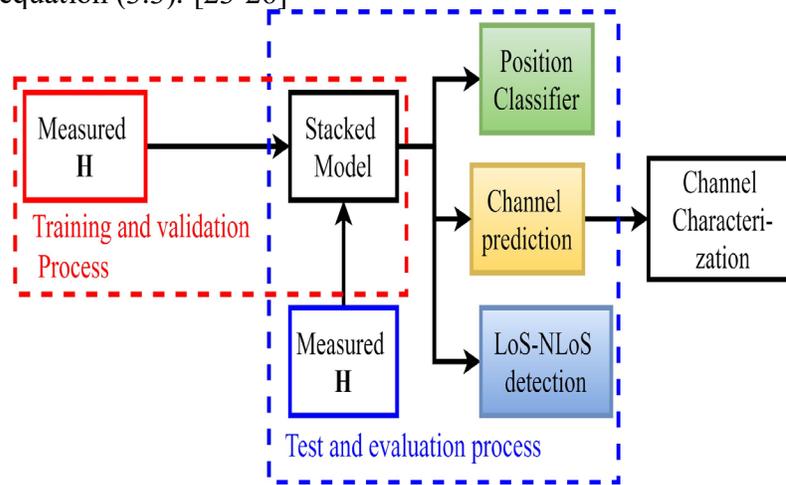
Figure 5-6 Proposed stack model for regression and classification.
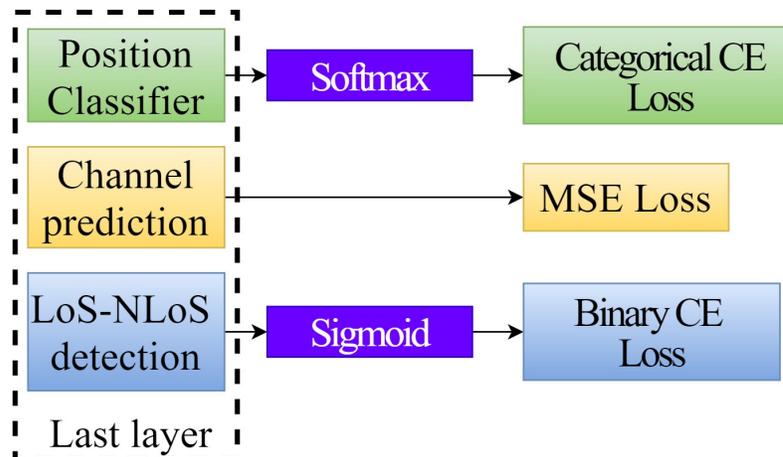
Figure 5-7 Activation functions and stacked model's losses.

Once the input data and the target output data are settled, the training and validation process are monitored by the learning curve (LC) [27]. Moreover, different problems with different approaches to quantify the LC are used. Figure. 5-9, 5-10 and 5-11 showed the binary and categorical cross entropy in terms of loss and accuracy for LoS-NLoS and position classification, respectively. Then, the MSE error in terms of percentage is used for channel prediction loss. Since the data is collected mostly in LoS scenarios (more than NLoS data), the stacked model has difficulties learning in the proper fashion. Specifically, when the training process begins to be adjusted for NLoS (more LoS data in the input). However, the data processing module assisted the model to adapt and predict appropriately the NLoS data, as seen in the Figure 5-9 high classification accuracy (up to 100%), as well as smaller BCE losses, were achieved by the stacked model for LoS-NLoS detection. A good agreement between the training and the validation is obtained. Furthermore, high accuracy is noticed up to 85% classification accuracy and lower CCE loss values in case of position classifiers (Figure 5-10). For channel prediction in Figure 5-11, the MSE in terms of percentage is lower with respect to the training epochs. Moreover, the training loss decreases and starts to settle in at approximately 300 epoch and 0.75% of MSE. Monitoring the learning curves is crucial for the model performance and the accuracy of the test results. As described in [27, 28] a good fit must be fulfilled to ensure that the model is learning sufficiently from the training sets, which is called underfitting, or the opposite is the overfitting. However, as
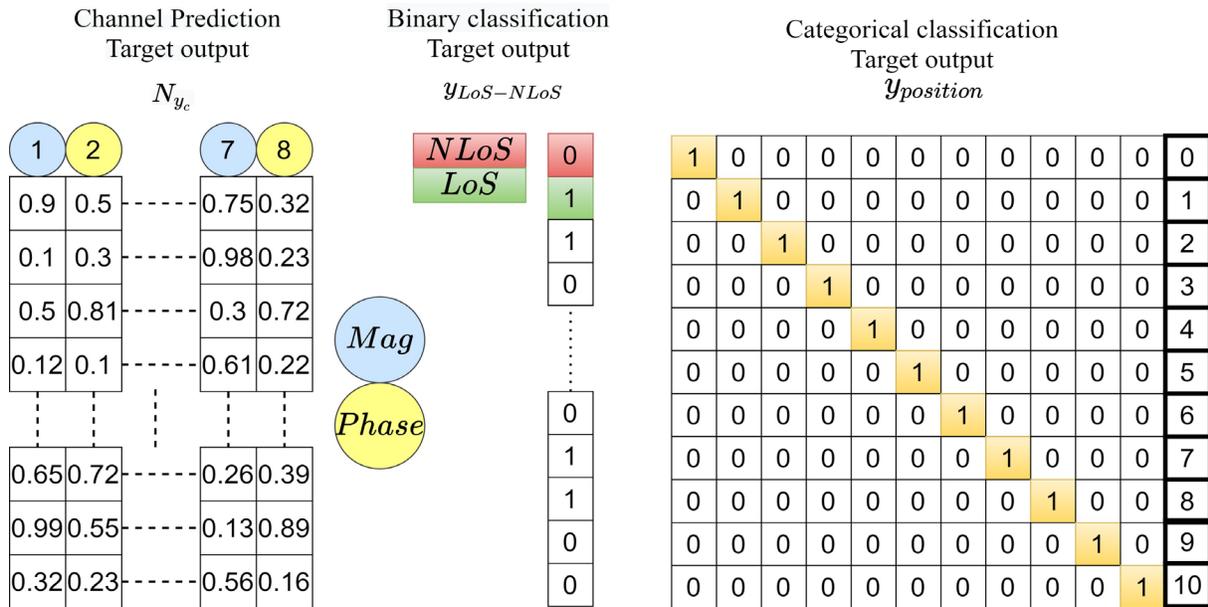
Figure 5-8 Target outputs for regression and classification

observed in Figure 5-9, 5-10 and 5-11, the stacked model is learning appropriately to properly predict the test sets. The learning curves obtained by the stocked model used the parameters shown in Table 5-3. The dynamic learning rate schedule [29] is introduced to assist the learning model to incorporate some learning randomness. once a learning rate is established at lesser values [29].

Table 5-3 Simulation parameters

Parameters	Stacked model		
	Position classifier	Channel prediction	LoS-NLoS detection
Hidden layers	100		
Batch size	100		
Loss functions	CCE	MSE	BCE
Training dataset size	136 600		
Validation dataset size	68300		
Optimizer	Adam		
Learning rate	0.001		
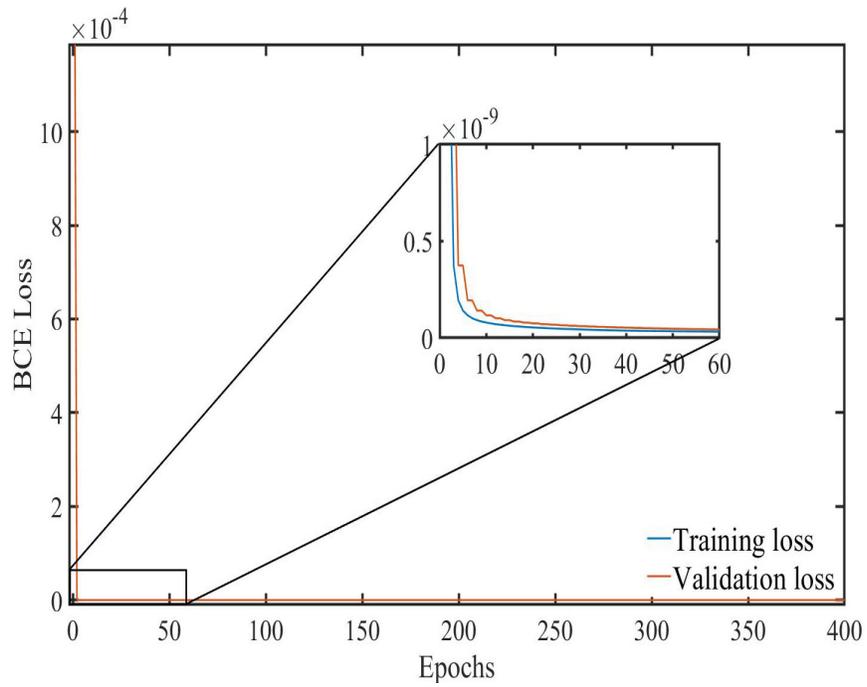
(a)
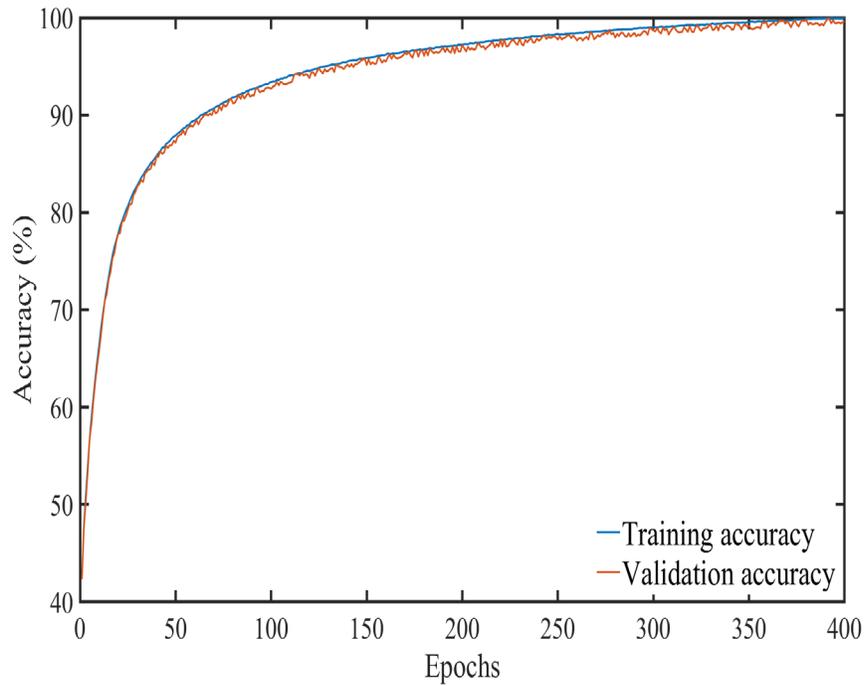
(a)

Figure 5-9 The training and validation learning curves for LoS-NLoS detection:

(a) BCE loss curves, (b) Accuracy classification curves.
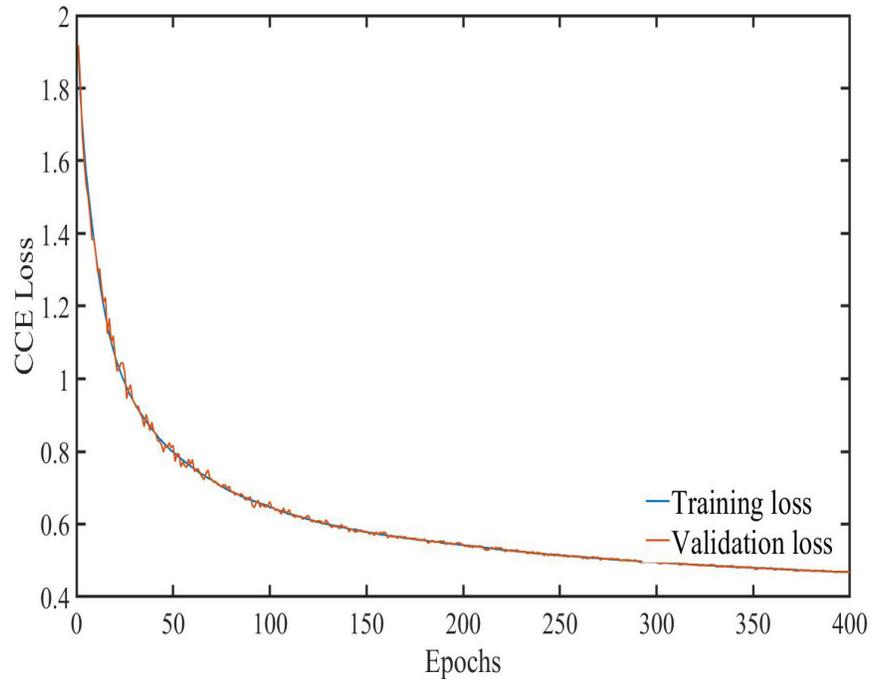
(a)
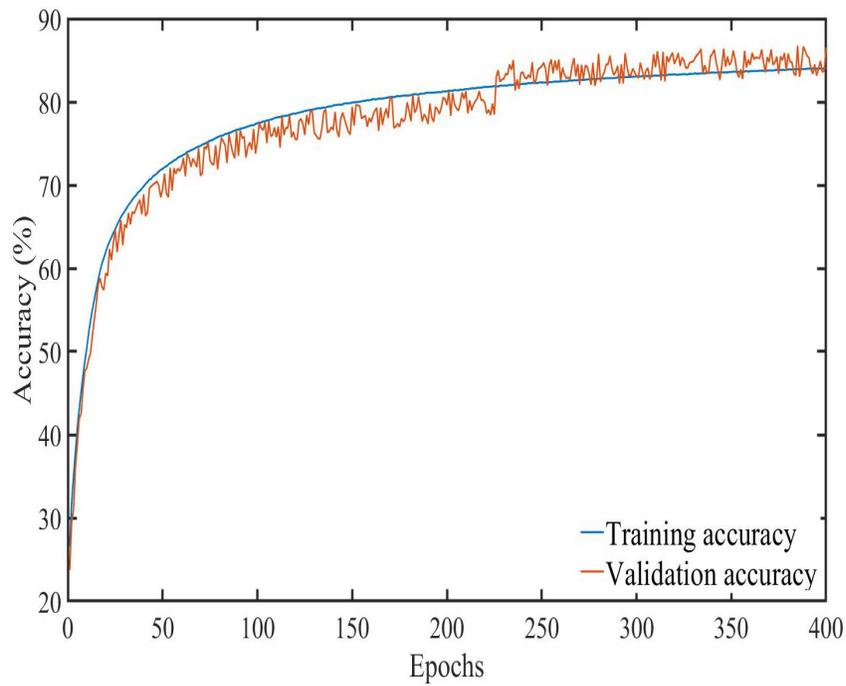
(b)

Figure 5-10 The training and the validation learning curves for position classification: (a) CCE loss curves, (b) Accuracy classification curves.
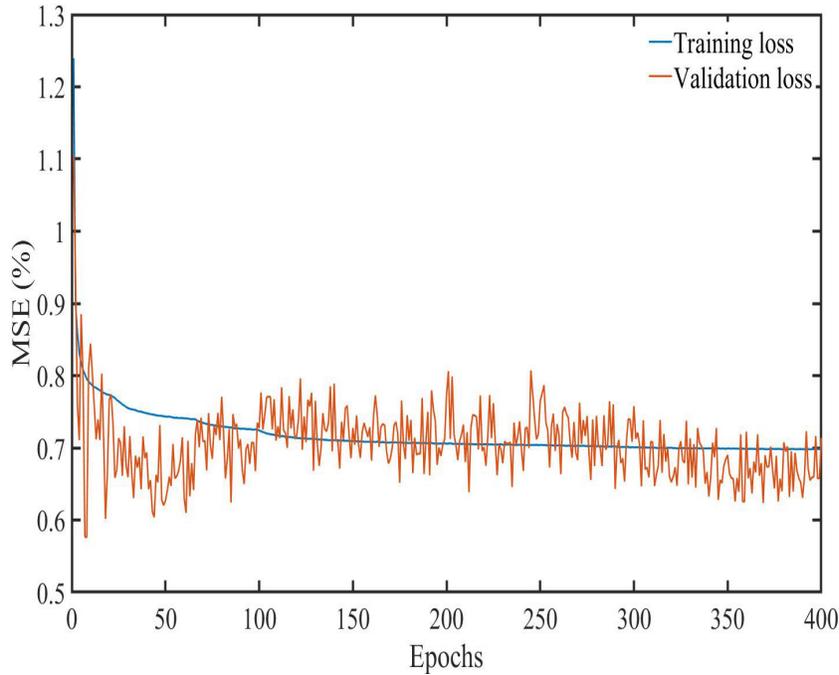
Figure 5-11 MSE loss in terms of percentages for channel matrix prediction

5.4 Test and evaluation module

For the test and evaluation module, the test data is used from the measured dataset. Therefore, the data processing is performed along with the training and the validation data. Moreover, the module aims to test the model capability to predict new samples, identify the position and whether it belongs to LoS or NLoS scenarios. In this case, the model is evaluated to predict the channel, classify the position and LoS-NLoS detection at the chosen position of 5 m.

5.4.1 Channel Matrix Prediction

For the MIMO channel prediction, Figures 5-12 and 5-13 demonstrate the measured and the predicted H within different antenna configurations in terms of magnitude (Mag [H]) and phase (Phase [H]) at 5m. Nevertheless, the Root Mean Square Error (RMSE) in terms of percentage was used to quantify the prediction losses [30] as illustrated in Figure 5-14. As seen, the prediction errors are lower for each sub-channels for the magnitude and the phase prediction. Even though, the RMSE values for the magnitude are less than the phase RMSE values, the model predicted well both parameters. This is due to the fact that the model is learning sufficiently as demonstrated by the learning curves.

93
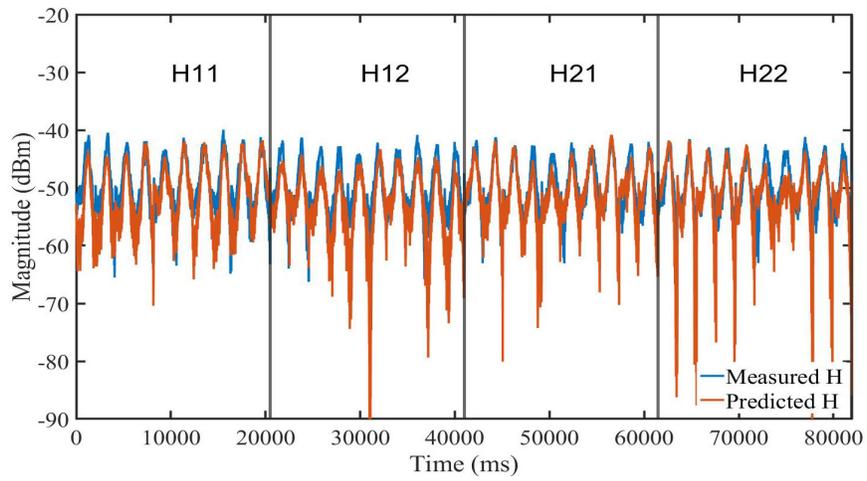
(a)
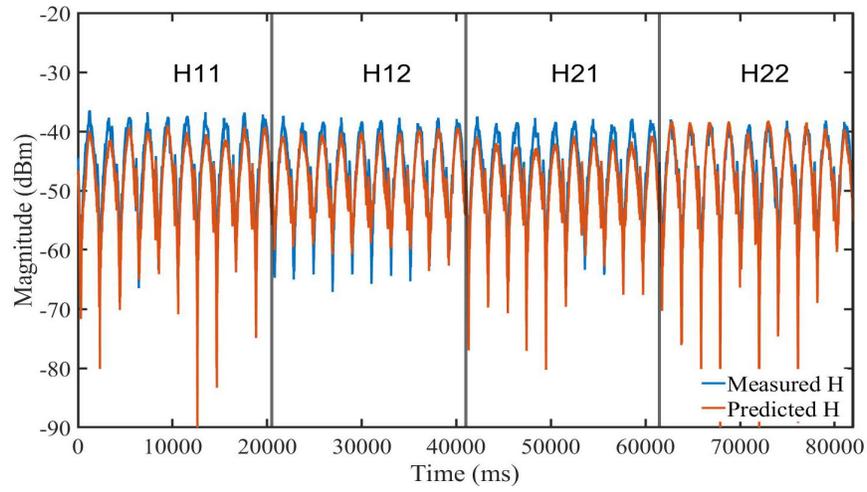
(b)
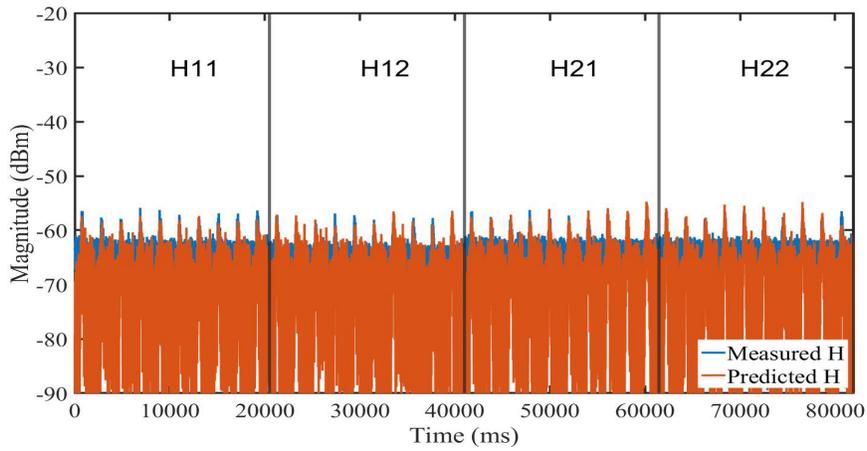
(c)
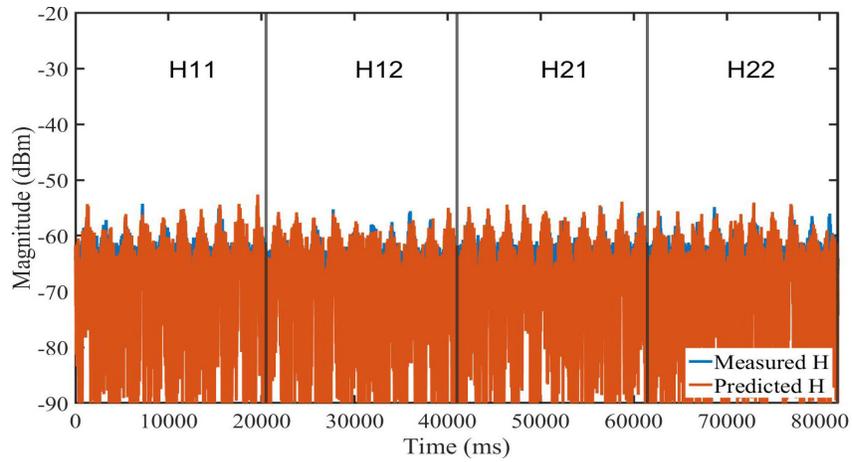
(d)

Figure 5-12 Measured and predicted channel matrix magnitude (Mag [H]) at 5m for different antenna configurations: (a) 90deg-CIR (b) CP-CIR (c) CP-LIN (d) CP-LIN-NLOS

(a)

(b)

95
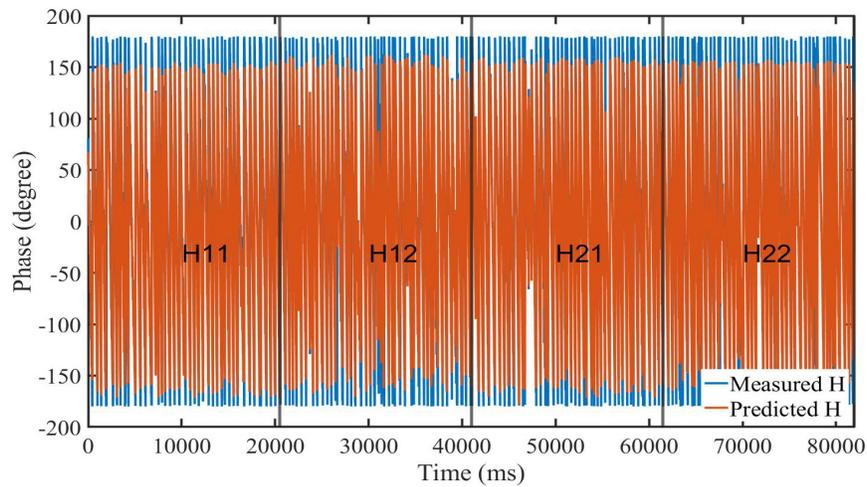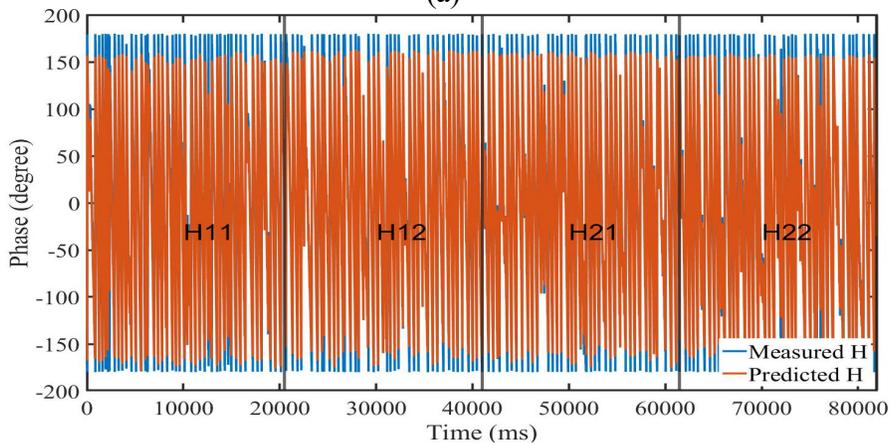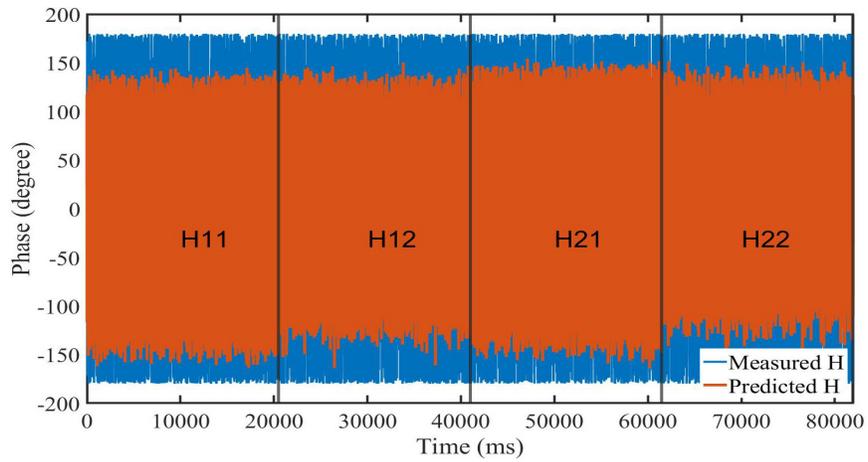
(c)
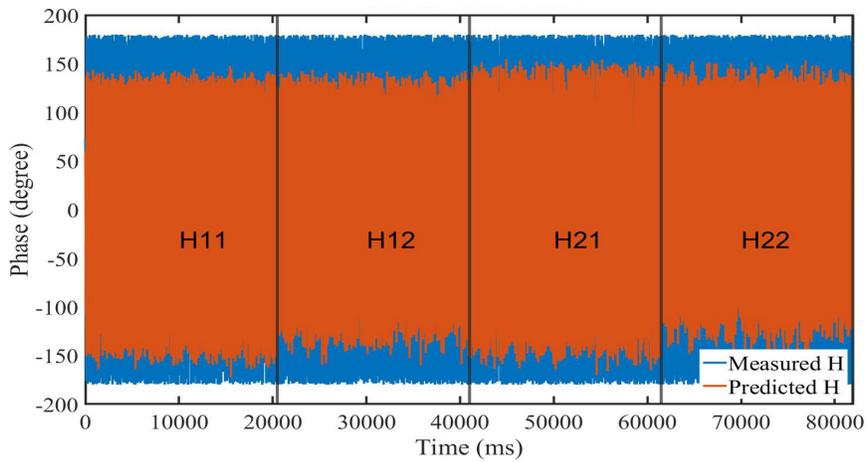
(d)

Figure 5-13 Measured and predicted channel matrix phase (Phase(H)) at 5m for different antenna configurations: (a) 90deg-CIR (b) CP-CIR (c) CP-LIN (d) CP-LIN-NLOS
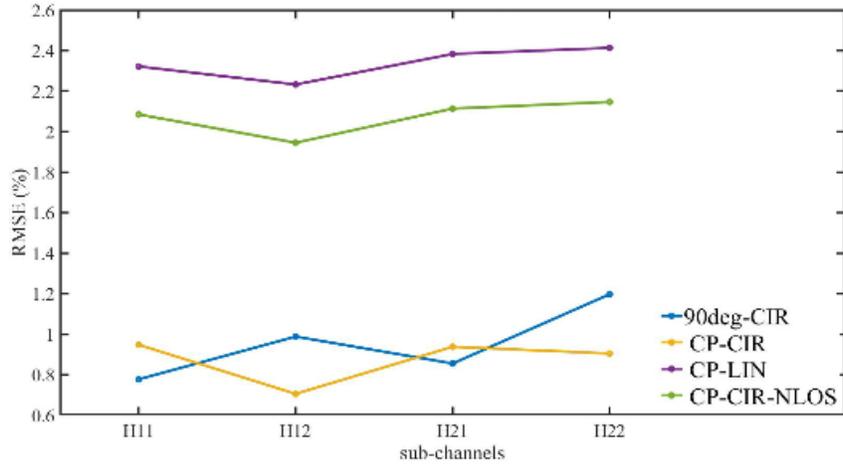
(a)
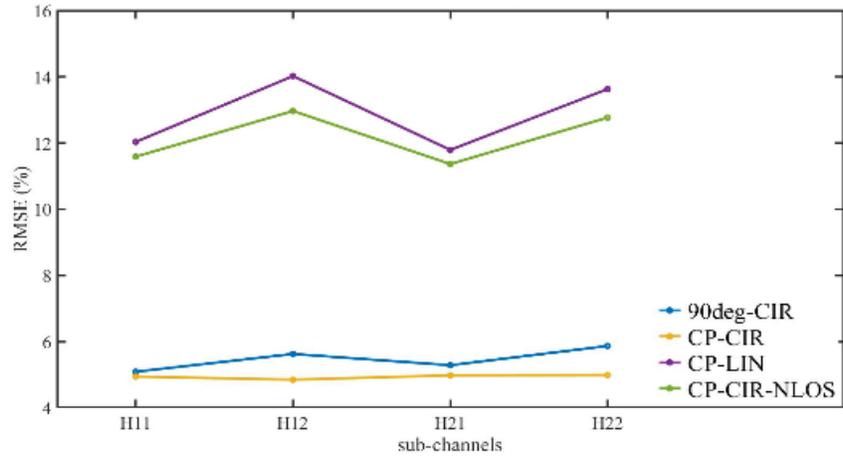
(b)

Figure 5-14 RMSE evaluation for channel prediction. (a) Magnitude (b) Phase

5.4.2 LoS-NLoS Identifier and Position Classifier

In this section, the classification results are illustrated. Moreover, as demonstrated in the training and the validation process, the model has achieved an efficient prediction process in terms of loss and accuracy. On the other hand, the confusion matrix (CM) [31] is used in the case of classification. The CM is considered as a summary of prediction results for any classification problem on a machine leaning field. It quantifies the number of correct and incorrect for the model predictions. Therefore, when the classification model generates predictions, the CM displays how it gets confused [27].

In Figure 5-15, the CM results are demonstrated in terms of position classifier where it is considered as multiple class classification (Up to 10 position to classify). The SM reached up to 87% accuracy prediction in terms of sample classification within all the antenna configurations. Even though, the SM is trained to classify 10 positions (output feature shape is 10). Hence, the proposed model shows its capability to predict positions within all the configurations with less collected positions, such as the CP-LIN-LOS and CP-CIR-NLoS configurations, where only seven and nine positions were measured, respectively.

Regarding the LoS-NLoS detection, the model achieved a 100% classification accuracy as demonstrated by the LC. Therefore, the CM is also used (Figure 5-16) in this case, which is considered as a binary output, where 0 and 1 refers to NLoS and LoS, respectively. It is worth mentioning that more data samples were collected in LoS scenario. However, the LC shows that the SM is learning properly to classify all the samples within Both LoS and NLoS scenarios. In both classifications, the SM aims to classify every sample collected at every position. Therefore, the SM will classify a sample collected at the position of 5 in LoS and NLoS scenarios. Considering that most of the data was collected in LoS scenario, only two configurations were considered in the Figure 5-15 and 5-16. The classification of the collected samples in the position 5 was evaluated within 2049 points (which length of collected data in each position, see Table 5-2) due to the hardware limitations. Moreover, the model classified with 87.9% of accuracy with the CP-LIN configuration in LoS scenario. As illustrated in Figure 5-15 (a), 1802 samples were correctly classified as of the rest of the sample were classified incorrectly between positions 1 to 10. Similarly, as CP—CIR configuration in NLoS scenario, 85.4% of accuracy is achieved by the SM where 1749 samples were correctly classified. Therefore, more significant incorrect classified samples (14.6%) are noticed as illustrated in Figure 5-16. It implies that, even though the model is learning well, it is challenging for the SM model to predict all the positions of the NLoS samples. This is due to the collected measurements, where more balanced datasets are needed to reach 100% of prediction accuracy. In fact, the collected position samples are not sufficiently balanced which does not optimize the SM prediction in a multiple class classification. On the other hand, the LoS-NLoS detection is well predicted in both configurations. Therefore, a 100% accuracy was achieved by the SM. Hence, it is noticed that the collected measured datasets were enough for the SM to identify the difference between the binary results (zero for NLoS and one for LoS).

The confusion matrix (a):

Predicted class	1	2	3	4	5	6	7	
1	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**330** 16.1%	**0** 0.0%	**0** 0.0%	0.0% 100%
2	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
3	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
4	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
5	**0** 0.0%	**0** 0.0%	**0** 0.0%	**1** 0.0%	**1718** 83.8%	**0** 0.0%	**0** 0.0%	99.9% 0.1%
6	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
7	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
	NaN% NaN%	NaN% NaN%	NaN% NaN%	0.0% 100%	83.9% 16.1%	NaN% NaN%	NaN% NaN%	**83.8% 16.2%**

True class: 1 2 3 4 5 6 7

(a)

The confusion matrix (b):

Predicted class	0	1	
0	**0** 0.0%	**0** 0.0%	NaN% NaN%
1	**0** 0.0%	**2049** 100%	100% 0.0%
	NaN% NaN%	100% 0.0%	**100% 0.0%**

True class: 0 1

(b)

Figure 5-15 Confusion matrix for CP-LIN configuration in LoS scenario:

(a) position classifier, (b) LoS-NLoS detection (LoS : 1 and NLoS: 0

99

(a)

(b)

Figure 5-16 Confusion matrix for CP-LIN configuration in LoS scenario:
(a) position classifier, (b) LoS-NLoS detection (LoS : 1 and NLoS: 0).

5.5 Channel Characterization and Modeling

In this section, the magnitude and the phase were used to evaluate the path loss and the channel modeling to compare with the measured ones which published by Elazhari [1].

5.5.1 Path Loss

As mentioned in previous chapiter, the Path Loss (PL) is calculated using the equation 2.3 and 2.4. The PL is defined as the signal attenuation caused by the effects of the environment. As noticed from the linear analysis, the values of the path loss exponent of the antenna configurations are evaluated and illustrated in Table 5-4.

Table 5-4: Path loss exponent

Polarization		Circular (Cir)			Linear
Scenarios		CP-LOS	90 deg—LOS	CP-NLOS	CP-LOS
PL exponent	Published	2.33	1.71	2.18	1.26
	Predicted	2.38	1.77	2.56	1.70

The SM predicted model results were significantly close to the published values reported in [10]. Therefore, the SM predicts the measured path loss in all the different antenna combinations within 2X2 MIMO system as illustrated in Figure 5-17. Hence, the same behaviour as the published results in [10] can be made with the SM prediction results. The Cir-CP-B2B channel performs well with regards to the other path loss values for the different configurations.
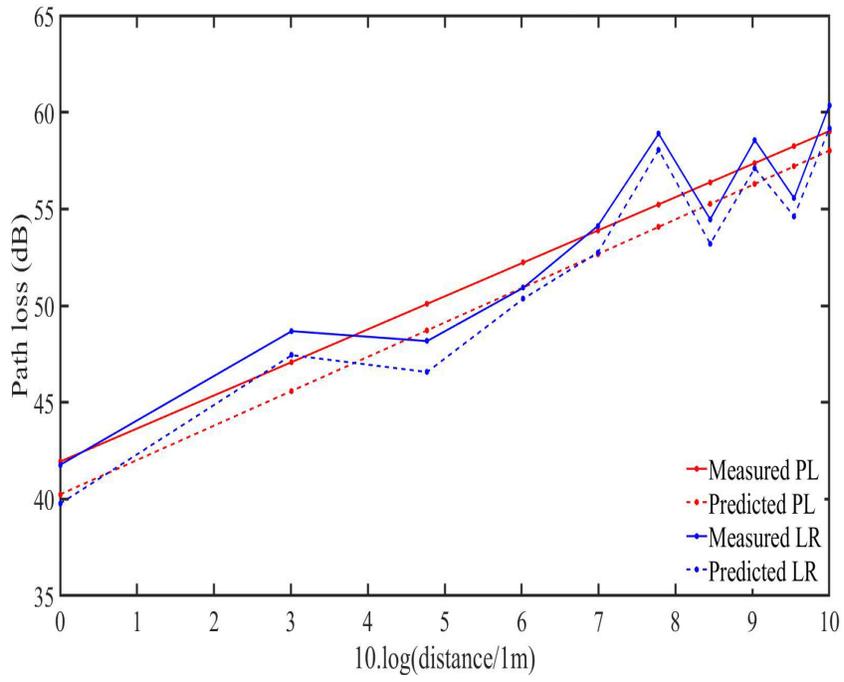
(a)
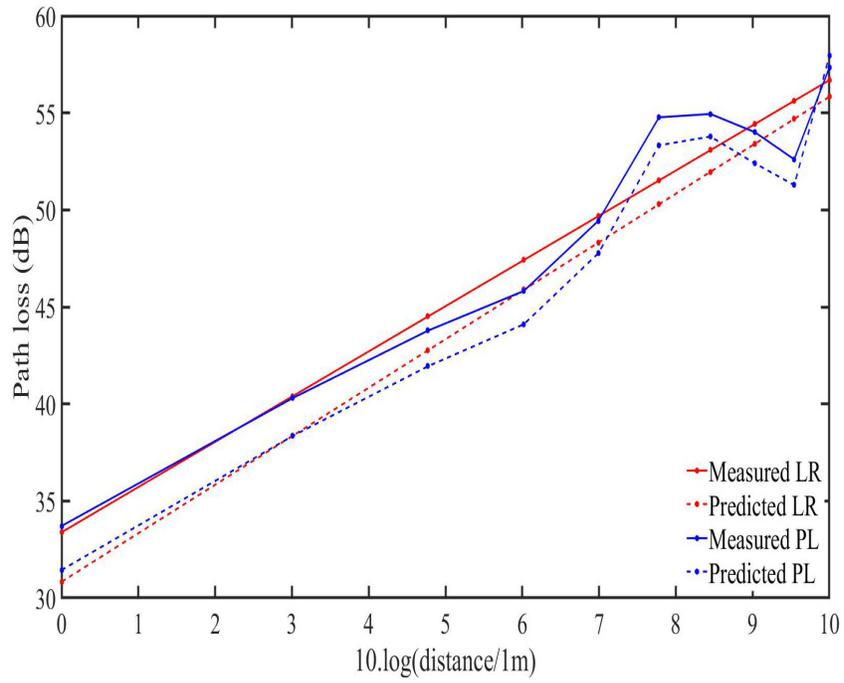
(b)
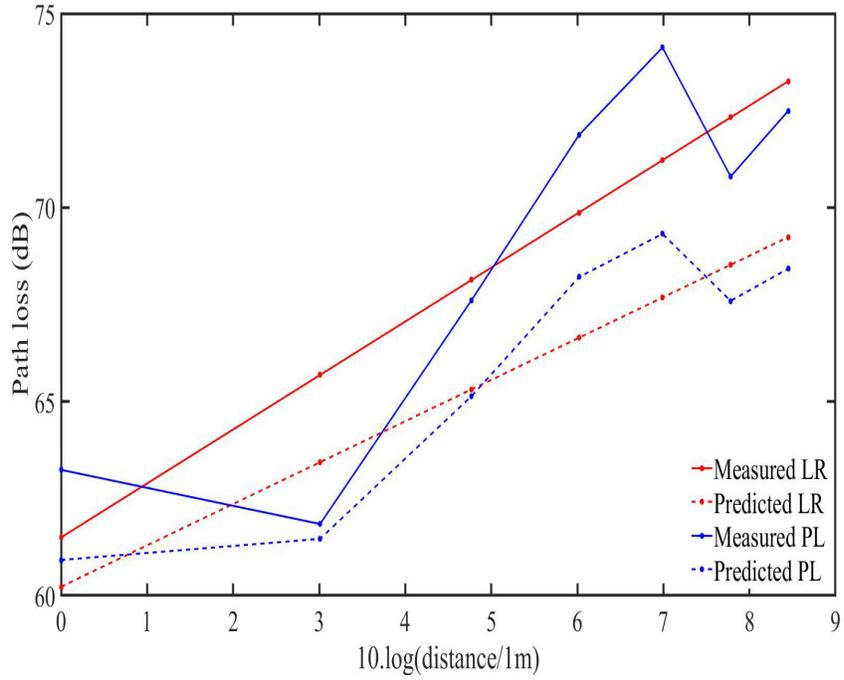
(c)
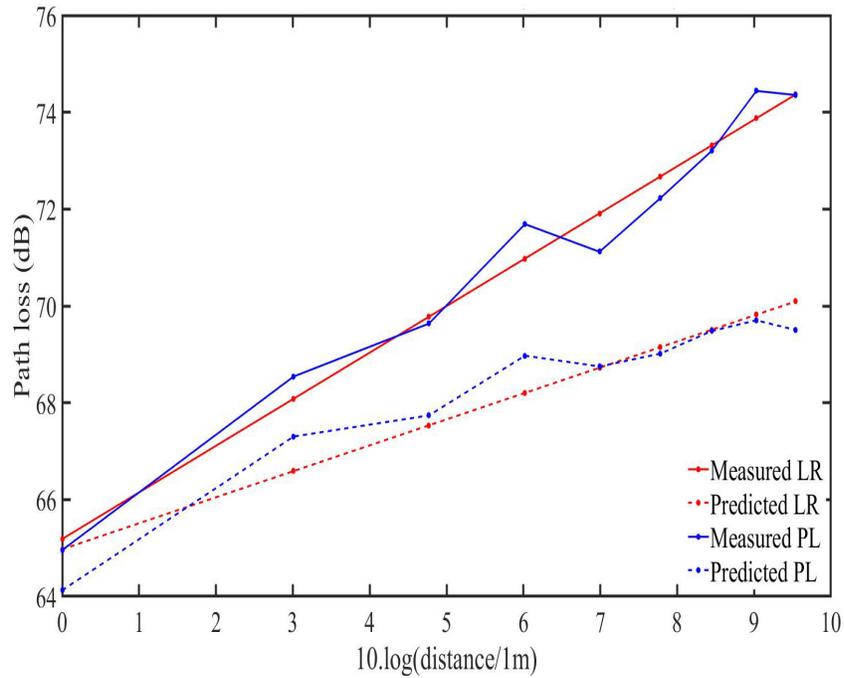
(d)

Figure 5-17 Measured Path Loss , predicted Path Loss and linear regression (LR) :

(a) 90deg-CIR (b) CP-CIR (c) CP-LIN (d) CP-LIN-NLOS.

5.5.2 Channel Modeling

As mentioned in [10], channel modeling consists on developing the impulse response to describe SISO—B2B system in mining environment which is characterized by rough and random surfaces. The impulse response is presented in [32]:

$$h(t) = \sum_{i=0}^{N-1} a_i \delta(t - t_i) e^{j\theta_i} \tag{5.5}$$

Where N is the number of multipath components, a_i , t_i and θ_i are the random amplitude, arrival time and phase of the ith multipath components. δ is the Kronecker delta function. Regarding the modeling procedures reported in [10], two modeling procedures were discussed in terms of path amplitudes and arrival times. In fact, the path amplitude a_i is modeled as independent complex Gaussian random variables with average power that follow the exponential power delay profile. The time arrivals of the multipath components were derived from measurements [10]. In this section, the SM impulse response model is compared with the measured and the stochastic—empirical (SE) modeled impulse response as illustrated in Fig 18. Table 5-5 illustrates the compared MSE values obtained by the SM and the stochastic model. It can be observed that the SM model MSE is the lowest, which provide more accuracy than the stochastic model to describe the impulse response for Body-to-Body channel.

Table 5-5 IR model performance

	SE model	SM model
MSE	1.8437e-04.	1.7250e-04
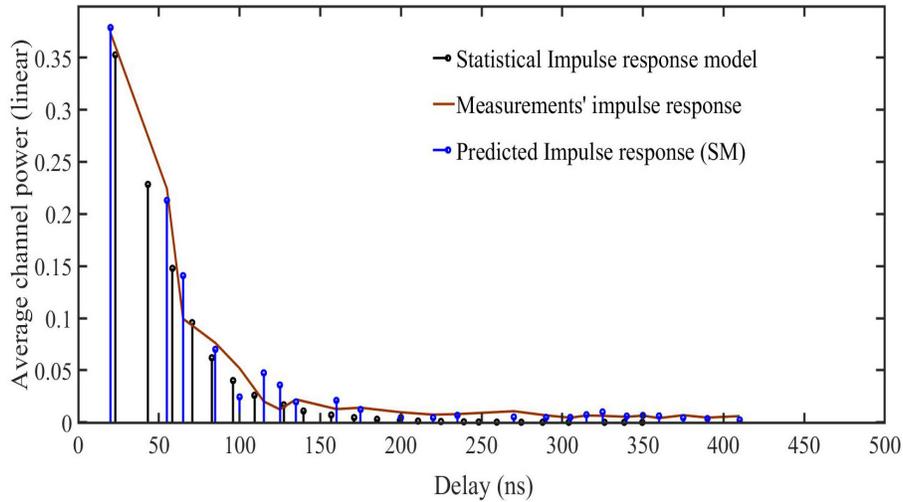
Figure 5-18 SM impulse response model compared to both stochastic empirical (SE) and measurement impulse response.

5.6 Conclusion

In this chapter, a new efficient MIMO channel modeling based on deep learning algorithm was presented. A stacked model is introduced to predict channel magnitude, phase, for each subchannels of the channel matrix H. Then, it classifies the position where the measurements were collected for LoS and NLoS scenarios. Moreover, different losses were applied to measure both classification and channel prediction problems. Published results in [10] were used to validate the SM model. The SM has achieved high accuracy in terms of performance assessment of the classification up to 100% and channel prediction with lower RMSE. Furthermore, the predicted path loss achieved more accuracy than the stochastic—empirical model.

The model was validated in estimating each subchannels, position classifier, and LoS-NLoS detection, while considering antenna diversity in underground mine environments.

5.7 References

[1] Andrew, W., Greatwood, C., and Burghardt, T., 'Visual Localisation and Individual Identification of Holstein Friesian Cattle Via Deep Learning', in, *Proceedings of the IEEE International Conference on Computer Vision Workshops*, (2017)

[2] Jiang, C., Shen, J., Chen, S., Chen, Y., Liu, D., and Bo, Y., 'Uwb Nlos/Los Classification Using Deep Learning Method', *IEEE Communications Letters*, 2020, 24, (10), pp. 2226-2230.

[3] Lucia, S. and Karg, B., 'A Deep Learning-Based Approach to Robust Nonlinear Model Predictive Control', *IFAC-PapersOnLine*, 2018, 51, (20), pp. 511-516.

[4] Ding, T. and Hirose, A., 'Fading Channel Prediction Based on Complex-Valued Neural Networks in Frequency Domain', in, *2013 International Symposium on Electromagnetic Theory*, (IEEE, 2013)

[5] Jiang, W. and Schotten, H.D., 'Deep Learning for Fading Channel Prediction', *IEEE Open Journal of the Communications Society*, 2020, 1, pp. 320-332.

[6] Luo, C., Ji, J., Wang, Q., Chen, X., and Li, P., 'Channel State Information Prediction for 5g Wireless Communications: A Deep Learning Approach', *IEEE Transactions on Network Science and Engineering*, 2018.

[7] Yuan, J., Ngo, H.Q., and Matthaiou, M., 'Machine Learning-Based Channel Estimation in Massive Mimo with Channel Aging', in, *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, (IEEE, 2019)

[8] O'Shea, T.J., Erpek, T., and Clancy, T.C., 'Deep Learning Based Mimo Communications', *arXiv preprint arXiv:1707.07980*, 2017.

[9] Arnold, M., Dorner, S., Cammerer, S., and Ten Brink, S., 'On Deep Learning-Based Massive Mimo Indoor User Localization', in, *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, (IEEE, 2018)

[10] Elazhari, M., Talbi, L., and Nedil, M., 'Body-to-Body Channel Characterization and Modeling inside an Underground Mine', *IEEE Transactions on Antennas and Propagation*, 2020.

[11] Mabrouk, I.B., Talbi, L., Nedil, M., and Hettak, K., 'Mimo-Uwb Channel Characterization within an Underground Mine Gallery', *IEEE Transactions on Antennas and Propagation*, 2012, 60, (10), pp. 4866-4874.

[12] Mabrouk, I.B., Talbi, L., and Nedil, M., 'Performance Evaluation of a Mimo System in Underground Mine Gallery', *IEEE Antennas and Wireless Propagation Letters*, 2012, 11, pp. 830-833.

[13] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y., 'Learning Phrase Representations Using Rnn Encoder-Decoder for Statistical Machine Translation', *arXiv preprint arXiv:1406.1078*, 2014.

[14] Hahner, S., Iza-Teran, R., and Garcke, J., 'Analysis and Prediction of Deforming 3d Shapes Using Oriented Bounding Boxes and Lstm Autoencoders', in, *International Conference on Artificial Neural Networks*, (Springer, 2020)

[15] Fenghour, S., Chen, D., and Xiao, P., 'Decoder-Encoder Lstm for Lip Reading', in, *Proceedings of the 2019 8th International Conference on Software and Information Engineering*, (2019)

[16] Malhotra, P., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., and Shroff, G., 'Lstm-Based Encoder-Decoder for Multi-Sensor Anomaly Detection', *arXiv preprint arXiv:1607.00148*, 2016.

[17] Joshi, S.S., 'Calibrating Recurrent Sliding Window Classifiers for Sequential Supervised Learning', 2003.

[18] Patro, S. and Sahu, K., 'Normalization: A Preprocessing Stage. Arxiv 2015', *arXiv preprint arXiv:1503.06462*.

[19] Ge, R., Huang, F., Jin, C., and Yuan, Y., 'Escaping from Saddle Points—Online Stochastic Gradient for Tensor Decomposition', in, *Conference on learning theory*, (PMLR, 2015)

[20] Hochreiter, S., 'The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions', *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 1998, 6, (02), pp. 107-116.

[21] Zhang, Z. and Sabuncu, M.R., 'Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels', *arXiv preprint arXiv:1805.07836*, 2018.

[22] Buja, A., Stuetzle, W., and Shen, Y., 'Loss Functions for Binary Class Probability Estimation and Classification: Structure and Applications', *Working draft, November*, 2005, 3.

[23] Nwankpa, C., Ijomah, W., Gachagan, A., and Marshall, S., 'Activation Functions: Comparison of Trends in Practice and Research for Deep Learning', *arXiv preprint arXiv:1811.03378*, 2018.

[24] Szandała, T., Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks', *Bio-Inspired Neurocomputing*, (Springer, 2021)

[25] Wang, M., Lu, S., Zhu, D., Lin, J., and Wang, Z., 'A High-Speed and Low-Complexity Architecture for Softmax Function in Deep Learning', in, *2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, (IEEE, 2018)

[26] Agarap, A.F., 'Deep Learning Using Rectified Linear Units (Relu)', *arXiv preprint arXiv:1803.08375*, 2018.

[27] Brownlee, J., *Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better Predictions*, (Machine Learning Mastery, 2018)

[28] Anzanello, M.J. and Fogliatto, F.S., 'Learning Curve Models and Applications: Literature Review and Research Directions', *International Journal of Industrial Ergonomics*, 2011, 41, (5), pp. 573-583.

[29] Zeiler, M.D., 'Adadelta: An Adaptive Learning Rate Method', *arXiv preprint arXiv:1212.5701*, 2012.

[30] Botchkarev, A., 'Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology', *arXiv preprint arXiv:1809.03006*, 2018.

[31] Trajdos, P. and Kurzynski, M., 'A Dynamic Model of Classifier Competence Based on the Local Fuzzy Confusion Matrix and the Random Reference Classifier', *International Journal of Applied Mathematics and Computer Science*, 2016, 26, (1), pp. 175--189.

[32] Ryckaert, J., De Doncker, P., Meys, R., de Le Hoye, A., and Donnay, S., 'Channel Model for Wireless Communication around Human Body', *Electronics Letters*, 2004, 40, (9), pp. 543-544.

CHAPTER 6   CONCLUSION AND FUTURE WORK

6.1 Conclusion

This thesis is intended to contribute to the design and development of wireless communication in two part.

In this thesis, we presented the results for applied machine learning in propagation channel within MIMO and SISO systems. The contributions are demonstrated in two parts. The first part tends to bring a new methodology to apply machine learning field to predict the propagation channel in any complex indoor environment within SISO-LoS scenario. CFR predictions were made in a variety of environments and frequency bands, with various antenna designs. Despite the fact that just one environment's data was used in the training and validation process, the model performed high accuracy prediction in all of environments. This is related to enhanced data processing, which is justified by the model's capacity to learn the relevant channel parameters between datasets and a good fit learning curve throughout the training and validation process. Channel characterization was carried out afterward, to visualize the model's capability to procreate same behavior of the channel and to demonstrate that the same observations could be made. As result, the model significantly Predicted (PPL) the Measured Path Loss (MPL), RMS delay spread, coherence bandwidth, K-factor, and channel capacity. The methodology demonstrated its use in forecasting the CFR, considering the antenna variety, where the same discussions could be held as the measured ones. Hence, the proposed model could promote the design of wireless communication systems in any complex area.

In the second part, the main contribution is attributed to design a data driven model for MIMO channel in underground mine environment. To achieve this objective, published measurements and results were considered [1] for the modeling validation purpose. Consequently, a new efficient WBAN-MIMO channel modeling based on deep learning algorithm was presented. A stacked model is introduced to predict channel magnitude, phase, for each subchannels of the channel matrix H, classify the position where the measurements were taken and whether in LoS or NLoS scenario, simultaneously. The SM, which is made up of three separate deep learning algorithms

based on LSTM networks, was employed since diverse results were predicted. Furthermore, various losses were used to assess both the classification and channel prediction problems. The SM model was validated using published results. Within the LoS and NLoS scenarios, the SM exhibited good accuracy in terms of performance assessment of the classification up to 100%, and channel prediction with low RMSE up to 0.6 percent and 4 percent for the magnitude and phase, respectively. Furthermore, path loss and channel modeling were performed to visualize the model capability to procreate the same behaviours as published model and to demonstrate that same observation could be made by the SM. As results, the SM predicted the path loss and achieved more accuracy than the published stochastic model. While considering antenna variety in underground mine environments, the model was proven in predicting each subchannel, position classifier, and LoS-NLoS detection. Furthermore, it enables the same conversations and observations as the measured ones to be concluded. As a result, the proposed model could aid in the development of wireless communication systems and increase underground miner safety.

6.2 Future work

Applying deep learning algorithm for channel propagation field, is currently a vast topic. Many areas can be exploited where were not covered by this PhD work. For instance, while different scenario and frequency bands were thoroughly investigated, with different antenna (different polarization), the important of the continuity to investigate other bands such as millimetre wave (mm-Wave) and the surface characteristic is very crucial. Hence, an open problem that future research works need to consider is the exploration of surface characteristic, mm-Wave band, and the dynamic effect of the body on the channel system. This exploration will enhance the proposed model especially for WBAN-MIMO system which contribute the safety of the miners in underground mine environment. Other deep learning algorithms combined with modeling techniques should be investigated for indoor application such as inside the mine. This includes the deterministic techniques using the uniform theory of diffraction and approximating the mine walls roughness to some reasonably close geometrical shapes.

6.3 References

[1] Elazhari, M., Talbi, L., and Nedil, M., 'Body-to-Body Channel Characterization and Modeling inside an Underground Mine', *IEEE Transactions on Antennas and Propagation*, 2020.