DEVELOPMENT OF A MACHINE LEARNING MODEL FOR CLASSIFYING FREE SPACE OPTICS CHANNEL IMPAIRMENTS

BY

BABATUNDE KAREEM SUNDAY

19PGBD000062

A THESIS SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND INFORMATION ENGINEERING, COLLEGE OF ENGINEERING. LANDMARK UNIVERSITY, OMU-ARAN

IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF MASTER OF ENGINEERING (M Eng.) DEGREE IN THE ELECTRICAL AND INFORMATION ENGINEERING DEPARTMENT, LANDMARK UNIVERSITY, OMU-ARAN, KWARA STATE, NIGERIA

MAY 2022

# DECLARATION

I Kareem Sunday BABATUNDE a M.Eng. student in the Department of Electrical and Information Engineering, Landmark University, Omu Aran, hereby declare that this research project entitled “Development of Machine Learning Model for Classifying Free Space Optics Channel Impairments”, submitted by me is based on my original work. Any material(s) obtained from other sources or work done by any other persons or institutions have been duly acknowledged in the text and list of references provided.

Kareem Sunday BABATUNDE Signature and Date

(Student/Researcher)

# CERTIFICATION

This is to certify that this thesis titled “Development of a Machine Learning Model for Classifying Free Space Optics Channel Impairments” and submitted by BABATUNDE KAREEM SUNDAY has been read and found to be in compliance with the required standards of the Department of Electrical and Information Engineering, College of Engineering, Landmark University, Omu-Aran, Nigeria, for the award of Master Degree (M. Eng.) in Electrical and Information Engineering.

Engr. Prof. Ibikunle Francis Signature and Date

Supervisor

Dr. P.K Olulope Signature and Date

Head of Department

External Examiner Signature and Date

# DEDICATION

This project is dedicated to God Almighty, my true insight, understanding, and wisdom. I also dedicate this work to my wife, Roseline Abiodun Kareem, who has been very supportive all through the process and also ensured that I put in the best effort to accomplish this feat. To my Son, Kareem Precious Ademide, whose arrival has brought innumerable joy and blessings to the family. My feelings for you all are indescribable. Thank you very much.

# ACKNOWLEDGEMENT

Foremost, I owe my deepest appreciation to the Almighty God., the creator of the universe and the giver of wisdom, knowledge and understanding. On whose wings I am able to soar through the research period, to Him alone be all the glory forever.

I also want to thank the Chancellor, Landmark University, Dr David O. Oyedepo whose word of wisdom and inspiration as sprung me to action and made me realise that age is no limit to learning and that everything is possible to Christ Jesus that strengthens us. Thank you very much Sir. Also to the Ag. Vice Chancellor, Prof. Charity Aremu, who also doubles as the Dean, School of Postgraduate Studies, the Head of Department Electrical and Information Engineering, Landmark University, Dr. P.K Olulope for all their relentless effort to ensuring that Landmark University work towards attainment of World Class status. May God crown all your effort with fullness of grace.

My deep appreciation goes to my Supervisor, Engr. Prof Ibikunle Francis for his encouragement, insightful comments, mentorship and advice all through the research stages in ensuring that the research work has a better flair and come out well. Much thanks to Dr Arowolo of the Department of Computer Science, Landmark Univerty for tutoring me on the rudiment of good research and ensuring the research work come to a favourable conclusion. Also, to all staff of the Centre for System and Information Services (CSIS) and Electrical and Information Engineering (EIE), your support is highly commendable.

Lastly, I would like to extend my gratitude to my family: my parents, wife, son, my siblings for their unflinching spiritual support through rigorous prayers throughout my life. God bless you all.

# TABLE OF CONTENTS

[DECLARATION ii](#_Toc108271711)

[CERTIFICATION iii](#_Toc108271712)

[DEDICATION iv](#_Toc108271713)

[ACKNOWLEDGEMENT v](#_Toc108271714)

[TABLE OF CONTENTS vi](#_Toc108271715)

[LIST OF TABLES viii](#_Toc108271716)

[LIST OF FIGURES ix](#_Toc108271717)

[LIST OF ABBREVIATION AND ACRONYMS x](#_Toc108271718)

[ABSTRACT xii](#_Toc108271719)

[CHAPTER ONE 1](#_Toc108271720)

[INTRODUCTION 1](#_Toc108271721)

[1.1 Background of the Study 1](#_Toc108271722)

[1.2 Statement of the Problem 3](#_Toc108271723)

[1.3 Aim and Objectives 4](#_Toc108271724)

[1.4 Significance of the Study 4](#_Toc108271725)

[1.5 Scope of the Study 5](#_Toc108271726)

[CHAPTER TWO 6](#_Toc108271727)

[REVIEW OF LITERATURE 6](#_Toc108271728)

[2.1 Review of Methodological Approaches and Related Works 6](#_Toc108271729)

[2.2 Theoretical Review: FSO Communication System 13](#_Toc108271730)

[2.2.1 FSO Transceiver 14](#_Toc108271731)

[2.2.2 FSO Terrestial Link 16](#_Toc108271732)

[2.2.3 FSO Challenges 19](#_Toc108271733)

[2.2.4 FSO Channel Models 20](#_Toc108271734)

[2.2.5 FSO Link Equations 24](#_Toc108271735)

[2.2.6 Benefits of FSO 27](#_Toc108271736)

[2.2.7 FSO Application Areas 28](#_Toc108271737)

[2.2.8 Overview of FSO investigation 29](#_Toc108271738)

[2.3 Machine Learning 34](#_Toc108271739)

[2.3.1 Machine Learning Task Benefit 36](#_Toc108271740)

[2.3.2 Machine Learning Types 37](#_Toc108271741)

[2.3.2.1 Supervised Learning 38](#_Toc108271742)

[2.3.2.2 Unsupervised Learning 41](#_Toc108271743)

[2.3.2.3 Reinforcement Learning 44](#_Toc108271744)

[2.3.3 Machine Learning Application Areas 45](#_Toc108271745)

[2.4 Machine Learning Predictive Model 45](#_Toc108271746)

[2.4.1 K-Means Clustering Algorithm 46](#_Toc108271747)

[2.4.2 K Nearest Neighbour 47](#_Toc108271748)

[2.4.3 Support Vector Machine (SVM) 49](#_Toc108271749)

[2.5 Receiever Operating Characteristics (ROC) Curve 52](#_Toc108271750)

[2.6 Gaps Identified in Literature 56](#_Toc108271751)

[CHAPTER THREE 57](#_Toc108271752)

[RESEARCH METHODOLOGY 57](#_Toc108271753)

[3.1 Research Framework 57](#_Toc108271754)

[3.2 Dataset 58](#_Toc108271755)

[3.3 K-Means Algorithm 59](#_Toc108271756)

[3.4 Support Vector Machine (SVM) 61](#_Toc108271757)

[3.5 K Nearest Neighbour 62](#_Toc108271758)

[3.5 Performance Metrics of the model 64](#_Toc108271759)

[CHAPTER FOUR 65](#_Toc108271760)

[RESULTS AND DISCUSSION 65](#_Toc108271761)

[4.1 Results and Evaluation Analysis 65](#_Toc108271762)

[CHAPTER FIVE 76](#_Toc108271763)

[SUMMARY, CONCLUSION AND RECOMMENDATION 76](#_Toc108271764)

[5.1 Summary 76](#_Toc108271765)

[5.2 Conclusion 76](#_Toc108271766)

[5.3 Recommendation 77](#_Toc108271767)

[5.4 Contribution to Knowledge 77](#_Toc108271768)

[APPENDIX 1 – Extract from Dataset 78](#_Toc108271769)

[APPENDIX II – Extract from MATLAB Programme 81](#_Toc108271770)

[REFERENCES 85](#_Toc108271771)

# LIST OF TABLES

Table 2.0: Confusion Matrix Sample……………………………………………..51

Table 3.0: Dataset Chart…………………………….………………………….....58

Table 4.1: Performance Metrics and Evaluation for the ML Model ……………..73

Table 4.2: Performance Evaluation comparison with Existing Work …..………..74

# LIST OF FIGURES

[Figure 2.1: FSO Transmitter and Receiver 15](#_Toc82466168)5

[Figure 2.2: Block Chart of an FSO System 15](#_Toc82466170)5

[Figure 2.3: FSO Link Impairment 20](#_Toc82466171)

[Figure 2.4: Machine Learning Type Chart 38](#_Toc82466172)

[Figure 2.5: Supervised Learning model 39](#_Toc82466173)

[Figure 2.6: Supervised Learning Steps 41](#_Toc82466174)

[Figure 2.7: Reinforcement Learning Analogy 44](#_Toc82466175)

[Figure 2.8: Machine Learning Application Area 45](#_Toc82466176)

[Figure 3.1: Proposed Framework](#_Toc82466177) 57

[Figure 3.2: K Means Algorithm Flow Chart](#_Toc82466179) 60

[Figure 4.1: Loaded Dataset 65](#_Toc82466179)

[Figure 4.2: Graphs for the Clusters Obtained](#_Toc82466179) 66

[Figure 4.3: Clustered 1 features for FSO](#_Toc82466179) 67

[Figure 4.4: Clustered 2 features for FSO](#_Toc82466179) 68

[Figure 4.5: Clustered 3 features for FSO](#_Toc82466179) 68

[Figure 4.6: FSO Data Confusion Matrix Using K-Means-SVM](#_Toc82466179) 69

[Figure 4.7: FSO Data Confusion Matrix Using K-Means-KNN](#_Toc82466179) 70

[Figure 4.8: Scatter Plot for the Study](#_Toc82466179) 70

[Figure 4.9: ROC Curve for the Experiment](#_Toc82466179) 72

# LIST OF ABBREVIATION AND ACRONYMS

ANN Artificial Neural Network

AAH Asynchronous Amplitude Histogram

ADTS Asynchronous Delay-Tap Sampling

ASE Amplified Spontaneous Emission

BER Bit Error Rate

CNN Conventional Neural Network

CDMA Code Division Multiple Access

DL Deep Learning

DAS Distributed Acoustic Sensing

FEC Forward Error Correction

FN False Negative

FP False Positive

FSO Free Space Optics

FSOC Free Space Optics Communication

GML Generative Machine Learning

KNN K Nearest Neighbour

LED Light Emitting Diode

MIMO Multiple Input Multiple Output

ML Machine Learning

NE Negative Exponential

NLPN Non-Linear Phase Noise

OFC Optical Fibre Cable

OWC Optical Wireless Communication

OSNR Optical Signal to Noise Ratio

OOK On-Off Keying

PSTN Public Switch Telephone Network

PPM Pulse Position Modulation

QoT Quality of Transmission

RF Radio Frequency

RSSI Received Signal Strength Indicator

RSA Routing and Spectrum Allocation

RWA Routing and Wavelength Allocation

RCSA Routing Core and Spectrum Allocation

SNR Signal to Noise Ratio

SVM Support Vector Machine

VLC Visible Light Communication

WDM Wavelength Division Multiplexing

# ABSTRACT

Free Space Optics is an optical communication method that uses Free Space instead of Fibre Cable to convey data through a medium from a transmitter to the receiver. It is a viable solution for ensuring high data rates and last-mile communication delivery in Next-Generation wireless communication. However, adverse weather conditions can significantly impair the performance of FSO channel links during transmission. Recently, Machine Learning models have received lots of attention in proffering solutions to signal impairments (that is, atmospheric turbulence, noise, and pointing errors) in optical networks. Machine Learning is a subset of Artificial Intelligence that deals with extracting patterns from data and then using those patterns to enable algorithms to improve on the experience. It allows computers to learn without having to be explicitly programmed.

In this work, the K-Means clustering algorithm combined with Support Vector Machine (SVM) and K Nearest Neighbour (KNN) classifiers were used, trained, and tested for classifying the channel impairments in FSO links. The Dataset used for the training and testing of the models is fetched from an open-source called “Kaggle”, cleaned by applying pre-processing techniques, and transformed before being used in the model via MATLAB simulation. Binary Classification Evaluation metrics, such as Accuracy, Precision, Specificity, Sensitivity, and F1 score were used in conjunction with the Confusion Matrix to determine the values of True Positive (TP), True Negative (TN), Faise Negative (FN), False Positive (FP) in calculating the expression of the performance evaluation metrics.

The Performance metrics comparison between the two classifiers (K-Means/SVM and K-Means/KNN) suggests that K-means/SVM outperformed K Means/KNN with 99.2% accuracy. The preferred model (K-Means/SVM) is also seen to outperform some existing classification models (K-means with Fuzzy Logic and Random Forest) during the comparison The research work developed a Machine Learning model for the classification of Free Space Optical Impairments such as atmospheric turbulence, noise and pointing error to the accuracy of 99.2% and also provides an effective tools for Free Space Optical (FSO) equipment manufacturers and for the effective monitoring and mitigation of losses of transmitted information in the communication industry.

# CHAPTER ONE

# INTRODUCTION

## 1.1 Background of the Study

Recently, bandwidth requirement as well as spectrum scarcity have posed enormous challenges to telecommunication services providers and network operators; which became a driving force for new ideas and methods to meet the demands of users or subscribers (Jaber et al., 2016). Several bandwidth improvement technologies have been examined and it ranges from Fibre-Optic Cable (FOC), Radio Frequency (RF) technology and copper-based technology. However, the most effective and viable choice is Optical Wireless Communication (OWC); it has a transmission capacity of up to 10Gbps and voice/video communication via atmosphere (Majumdar, 2015).(Griffiths et al., 2020). Due to tremendous capacity, fibre has been the only viable alternative to the backhaul segment of future communication networks. However, in some situations where Fibre deployment is too costly or impracticable, its counterpart, FSO, is the best option.

Optical wireless communication (OWC) approaches have been developed as an alternative to the RF Spectrum problem. In fact, future wireless networks, such as 5G and 6G, will require this technology in particular. The two most common types of OWCs used for indoor and outdoor communication are Free Space Optics (FSO) and Visible Light Communication (VLC). Because of their accessible capacity, they are an excellent substitute for RF in dealing with developing high-bandwidth ideas and devices, for instance, smart cities and the Internet of Things (Saeed et al., 2019) (Menaka et al., 2021).

FSO is a method of optical communication that uses free space rather than fibre cables to convey data. As a result, the signal is susceptible to a variety of flaws that degrade its quality. In recent times, there has been rumour about using Machine Learning to predict signal impairments in optical networks. On the other hand, the development of Free Space Optics (FSO) linkages is in its initial stages (Alimi et al., 2019)(Esmail et al., 2021).

In recent times, Machine Learning (ML) has been an important subject mostly in optical fibre-based and Optical Wireless Communication due to the requirement of complicated simulation solutions for networks and optical communication systems. ML applications in communication system include classifying modulations and tracking the activities of the Optical Fibre (OF) system, for example Optical Signal to Noise Ratio (OSNR) and colour diffusion. Some ML applications for FSO systems have been reported, including signal detection, decoding, demodulation, and end-to-end system modelling. However, the application of machine learning in the FSO system to detect signal impairments is still in its early stages (Thrane et al., 2017).

Studies have provided an outline of the application of machine learning in optical networks; encompassing both the network and physical layers, with a huge number of studies employing machine learning approaches for the channelling and apportioning of resources still in the early stages. Various algorithms, including machine-learning techniques such as Support Vector Machine (SVM), Conventional Neural Network (CNN), and K-Nearest Neighbor (KNN), have been used to investigate the connectivity scope as a criterion to analyse the brutality of an FSO system dust storm in a dusty channel environment (Boutaba et al., 2018).

Aside from breakthroughs in machine learning algorithms, some other factors have aided its resurgence. Most significantly, the effectiveness of machine learning approaches is data-dependent. Without a doubt, today's networks include massive volumes of data, which will only grow when new networks arise, like the Internet of Things' billions of linked devices. This promotes the use of Machine Learning (ML), which may be used to learn and grasp data-generation processes as well as uncover hidden and unexpected patterns (Boutaba et al., 2018).

In this study, an ML algorithm approach using K-Means with SVM and KNN are developed for classifying FSO channel impairments and performance metrics comparison done to detect the best ML Model.

## 1.2 Statement of the Problem

Arguably, substantial progress has been made in FSO, but the administration and network operations are still challenging, with often-human error causing network failures. Network flaws result in financial liabilities and reputational damage to network providers. As a result, there's a lot of interest in developing a highly robust autonomic network (Yaacoub et al., 2020).

In a similar vein, interference and atmospheric turbulence exact concerns with FSO technology - a viable and promising technology; hence, it becomes very important for researchers to design good models that can anticipate the channel behaviour and classifying FSO channel impairments. To tackle this challenge, this research work developed an Hybrid Machine Learning Model that classify Free Space Optics Channel Impairments that degrade the performance of FSO system.

## 1.3 Aim and Objectives

The aim of this study is to develop a classification model for FSO Channel impairments using an integrated ML model of K-Means, SVM and KNN for an improved performance.

The specific objectives are to:

1. Identify the relevant features in FSO Dataset.
2. Develop and Implement the classification model with K-means, SVM and KNN using MATLAB Simulation
3. Evaluate the developed model using performance metrics such as accuracy, sensitivity, specificity, precision, recall and f1-score.
4. Plot and compare the evaluated results obtained in (iii) with other models.

## 1.4 Significance of the Study

The study investigates the ability of ML algorithms to classify impairments such as atmospheric turbulence, noise and pointing error. This study classifies the FSO dataset into defined number of categories by usng Hybridized K-Means with Support Vector Machine Model which can be used to form classification model for practitioners and researchers to effectively monitor and mitigate loss of transmitted information in the communication industry.

## 1.5 Scope of the Study

This research identifies and survey Wireless Optics Communication and Machine Learning related literatures, utilized relevant Free Space Optics dataset retrieved from reputable FSO data repository and hybridized supervived with unsupervised learning machine learning techniques to classify FSO channel impairments such as Atmospheric Turbulence, Noise and Pointing Error.

# CHAPTER TWO

# REVIEW OF LITERATURE

## 2.1 Review of Methodological Approaches and Related Works

(Esmail et al., 2021) suggested a ML approach for FSO channel monitoring by considering forecasting FSO channel parameters when more than one impairment is present. This, first, demonstrates how effectively the asynchronous amplitude histogram (AAH) and asynchronous delay-tap sampling (ADTS) histogram characteristics accurately predict FSO parameters. According to the findings, the ADTS histogram features increase prediction precision. Second, the performance of convolutional neural network (CNN) regressors and SVM was compared using ADTS histogram features. The findings show that in certain situations, the CNN regressor outclasses the SVM regressor, while in others, they are comparable. (Esmail et al., 2021) also investigated the CNN regressor's ability to predict channel characteristics at three different transmission rates. The findings show that, regardless of transmission speed, the CNN regressor is capable of properly predicting the OSNR parameter. In low-speed transmission, however, the prediction of turbulence and aiming errors is more accurate than in high-speed transmission.

The performance of an FSO link is directly influenced by atmospheric turbulence and climatic variables in terms of dependability and even availability. For various communication systems and performance indicators, an extensive experimental and theoretical study has been undertaken in both terrestrial and maritime contexts. The effects of modest gamma-distributed turbulence on the bit rate of an FSO link over water were investigated by Bourazani et al. (Bourazani et al., 2018). The M-distribution was used to predict atmospheric turbulence-induced fading and the line-of-sight (LOS) blocking of an FSO link in a theoretical study (Garrido-Balsells et al., 2017). The average symbol error rate (ASEP) of a hybrid RF/FSO link was explored by Kong et al. (Kong et al., 2015). Alheadary et al (2018) used an FSO system with a 70-meter connection length to predict the channel attenuation coefficient in a coastal setting, considering air temperature, humidity, and dew point (Alheadary et al., 2018).

Machine learning techniques were utilized by Lionis et al. (Lionis et al., 2021) to forecast the received spectral efficiency of an FSO link in real-time over a marine ecosystem, non-linear atmospheric conditions. Artificial neural networks (ANN), tree-based approaches (decision trees, gradient boosting, and random forest), and k-nearest neighbors (KNN) were used and compared utilizing the coefficient of determination (R2) and root mean square error (RMSE) as the most essential measures of performance. All ML models had an outstanding fit in the regression study, indicating that they can provide an appreciable increase in the modeling of FSO performance when compared to regression models used in the past. The ANN method was shown to be the top-performing R2 model, whereas random forests had the best RMSE result.

Furthermore, compared to (Bourazani et al., 2018) the work used more channel parameters, but it was confined to a few algorithms; however, other machine learning algorithms should be evaluated to see which is a superior predictor. Mishra et al. (Mishra et al., 2019) used machine learning approaches to estimate channel length in free-space optical communication schemes. The optical channel of an experimental FSO link with optical turbulence generating (OTG) chamber was investigated. The channel coefficients were estimated using maximum likelihood estimation (MLE) and Bayesian estimation approaches. Analytically and experimentally, the calculated channel coefficients in both circumstances are nearly identical. However, MLE was chosen over Bayesian because of its reduced complexity. Furthermore, regardless of turbulence level, increasing the pilot symbol length results in higher BER for a given transmitter power level. It was discovered that increasing the pilot symbol length for a given transmitted power level improves all turbulent regimes, although, the variety of machine learning approaches used during the research was limited.

Aveta et al. (Aveta et al., 2020) used unsupervised learning to make a finding on actual users, concentrating on Cognitive multi-point free-space optical communication (FSOC). A conceptual technique based on unsupervised learning was developed with a single-node receiver to estimate the number of users who will interact and share time and bandwidth restrictions. When users were given comparable amplitude information, which resulted in underestimating, an extensive experimental weighted clustering approach was proposed. Even in the presence of severe atmospheric turbulence, the proposed technique proved to accurately identify the number of concurrently communicating users with a precision level of more than 92 percent, according to the research findings. To determine the data's robustness, experimental study was done and a transmitter sampling interval was required for accurate approximation. Furthermore, an analytical framework was created and validated for up to four users to analyze the effect of precursor signal duration on estimation accuracy for a specified speed. The work did not consider non-synchronous transmission and supervised machine learning techniques.

Checking FO connections using a reduced polarimeter aided by learning algorithms was proposed by (Slapak et al., 2020), the extensively utilized optical fibres in telecommunications are useful for fibre network self-monitoring and (distributed) sensing at the same time. They monitored fibre changes by monitoring changes in the light polarization state rather than using back-scattered light approaches in their research. These modifications may result in a large quantity of data, but extracting usable information from it, such as future fibre break predictions or seismic monitoring, is usually difficult. To overcome this problem, they recommend employing machine learning.

Unsupervised machine learning methods were used because the event of measured data was not labelled. They exhibited a proof-of-concept technique that requires installing optical fibre disrupted by deliberate vibrations, pressing mostly on fibre, closing a rack opening is disrupted by deliberate vibrations, pressing mostly on fibre, and closing a rack opening, that is close to the fibre using a simple polarimetric procedure. Using a machine learning K-means technique, they distinguish data collected by these intended disturbances from data generated through noise caused by normal traffic. These findings are the first step toward automated data labelling, which might be applied to categorize occurrences in the future.

A review of ML-based navigation and allocating resources in connectivity was conducted by Y. Zhang et al., 2020, and this study gives an overview of ML-based routing and resource allocation in optical networks. They begin by discussing the Routing and Spectrum Allocation (RSA) problem in EONs, the routing and wavelength allocation (RWA) problem in WDM optical networks, the Routing Core and Spectrum Allocation (RCSA) problem in SDM optical networks, the core and routing, and the machine learning approaches commonly used in optical networks. Difficulties of estimating crosstalk prediction, quality of transmission (QoT), traffic estimation, all of which might aid in resource allocation and routing, are then discussed. The RSA, RWA, and RCSA algorithms that are ML-enabled are elaborated, examined, and contrasted in depth. ML applications in traffic estimation, QoT estimate, and crosstalk prediction, among other things, are also talked about. They offer future research prospects for using ML approaches to do resource allocation and conduct routing in satellite optical networks and multidimensional time–space–frequency optical networks based on existing research achievements.

Machine learning was proposed for microseismic event identification using data from distributed acoustic sensing (DAS) (Stork et al., 2020). The ability of a convolutional neural network (CNN) built to detect microseismic events in a single fibre-optic Distributed Acoustic Sensing (DAS) dataset can be transferred to additional datasets. This is the first time it has been proven in this investigation. The enormous data volumes generated by DAS's temporal sampling and dense spatial for microseismic monitoring in industrial settings necessitate novel approaches for near-real-time microseismic analysis. To train the YOLOv3 detection system to distinguish micro seismic events, they employed synthetically produced waveforms with actual noise overlaid. The number of events detected using amplitude threshold detection and filtering techniques is compared to the CNN network's functionality. The network can recognize less than 80% of the events observed by manual inspection when actual noise is used, which is 14% much more than classic frequency-wavenumber filtration approaches. Erroneous detection occurs around 2% of the time, or once every 20 seconds. By monitoring geometries and circumstances that were previously unnoticed by the network, the CNN discovers more than half of the occurrences suggested by a manual assessment.

(Gao et al., 2020) proposed ML-based applications for upcoming generation optical networks, since the need for optical network capacity and reliability has constantly increased over the previous few decades. Optical networks with bigger knowledge sizes, on the other hand, have become sources of a vast amount of diverse data. To tackle these new challenges, power optimization, minimal optical network layout, wavelength assignment, routing, and failure management are just a few of the concerns that must be addressed. Machine learning, a kind of artificial intelligence (AI), is largely regarded among the most viable options to overcoming these issues. Gao et al., 2020 look at how machine learning techniques may be utilized to solve these four challenges in his study.

(Musumeci et al., 2019) proposed the development of ML methods in optical networks. They provided an outline of how machine learning is used in optical communications and networking, categorized and review significant material on the subject, as well as give an introductory course on machine learning for scholars and practitioners interested in the field. Despite the recent publication of a few research articles, the application of machine learning to optical networks is still very much in formative stages.

Machine Learning for Optical Communication was proposed (Amirabadi, 2019). Machine Learning (ML) is a hot issue that has only lately gained traction and will continuously do so for more years ahead. In this field, research progress is accelerating at a breakneck pace. The novelty of this study direction is primarily due to the uniqueness of the application sector, rather than the methodological methods, which are (at least for the time being) cutting-edge machine learning algorithms. According to a review of the literature, many ML techniques have yet to be employed in this domain, and many OC applications have yet to be investigated, indicating that the research topic is still in its infancy. As a result, investigations are critical in this subject to assist researchers to understand the most recent advancements and gaps in the field. Even though multiple tutorials have lately been published, they have all approached this problem from an OC perspective, ignoring the ML perspective. It is, nonetheless, necessary to have one.

Thrane et al., 2017 suggested utilizing machine learning techniques to accurately estimate received optical power of an FSO connection in a marine setting. Over the last two decades, there has been a lot of research on predicting an optical communications link performance in a maritime environment. The effects of turbulence on a variety of atmospheric phenomena have been extensively studied, and long-term data have allowed simple empirical models to be developed. In real-time non-linear atmospheric circumstances, the study illustrates how effectively various machine learning (ML) algorithms estimate the FSO link performance. For a laser communication link, seven local atmospheric factors were compared to a large collection of received signal strength indicators (RSSI). Artificial neural networks (ANN), k-nearest-neighbours (KNN), tree-based approaches, and k-nearest-neighbours (KNN) have all been used and compared using performance metrics such as coefficient of determination (R2) and root mean square error (RMSE). In the regression study, all ML models fit perfectly, demonstrating that, when compared to traditional regression models, they can provide a significant boost in FSO performance modeling. The ANN technique produced the best R2 results (0.94867), while random forests produced the best RMSE results.

A reliable free-space communication, using generative machine learning was suggested (Lohani et al., 2020). Turbulent light propagation through the attenuation, atmosphere and receiver detector noise all affect free-space optical communications systems. These effects reduce symbol classification accuracy, degrade the received state quality, and increase cross-talk. To show its usage in simulated circumstances and experimental communications, Lohani et al., 2020 combines a convolutional neural network (CNN) system with a state-of-the-art generative neural network (GNN). The GNN system corrects for distortion and decreases detector noise in testing without the need of feedback or adaptive optics, resulting in receiver mode profiles that are almost identical to those desired. When these created modes are demodulated with the help of a CNN that has been pre-trained using undistorted modes, the classification accuracy improves significantly. They show a reduction in cross-talk across empirically identified noisy/distorted modes at the receiver using a GNN and CNN system entirely pre-trained using simulated optical profiles. This scalable method might be used to realistically and efficiently demodulate long-range classical and quantum communication networks.

## 2.2 Theoretical Review: FSO Communication System

Free Space Optical Communication (FSO) comprises the dissemination of visible and infrared beams across the atmosphere to produce optical communications. The word "free space" can refer to air, outer space, or a vacuum; also, it is called Free Space Photonics or Optical Wireless. FSO Communication employs LENS at both the transmitter and receiver, and its maximum range is around 4km (Malik & Singh, 2015).

FSO Communication is primarily of two types namely:

1. Light Emitting Diode (LED) Transmitter based: This has to deal with a short-range communication system where the light generated is incoherent.
2. LASER transmitter based: This has to deal with a long-distance communication system where the light generated is coherent.

As a wireless technology, "Free Space Optics Communication" transmits data in free space by using light. Because of increased bandwidth and capacity needs, a transmission from RF to optical communication surfaced in recent years. This is referred to as wireless communication because optical beam transmission can occur in either free space or a confined area (Bloom et al., 2003)

Wireless optical communication (WOC) may be used both indoors and outdoors. Indoor optical communication, which operates in the infrared region, uses a wavelength band of 750-950 nanometers. Indoor wireless optical communication systems can only be used inside rooms. Instead of using optical fibre, a transmitter LED or LD in Free Space Optics Communication (FSOC) delivers digital data such as video pictures and data documents via an unsupervised light beam in free space. The receiving lens, which is connected to an incredibly sensitive receiver, captures these light beams at the receiving end (MAJUMDAR, 2015).

### 2.2.1 FSO Transceiver

As illustrated in Fig. 2.1 and Fig. 2.2 below, an optical communication system in free space comprises three components: a transmitter, an atmospheric channel, and a receiver. A control algorithm and a source of light (LED or LD) turn the data into an optical source within the transmitter, which is subsequently modulated using modulation methods. The transmit optics send an optical beam across the air channel. Atmospheric channel signal intensity is reduced by air instability, dispersion, absorption and background noise. At the receiving end, a photodetector and preamplifier circuit converts the transmitted light into a power source. The electrical signal is demodulated using a demodulator to get the data transferred (Anbarasi et al., 2017).



Figure 2.1: FSO Transmitter and Receiver. (Kaushal, 2018)

ATMOSPHERIC TURBULENCE

SCATTERING BEAM WANDERING

ABSORPTION BEAM SPREADING

BACKGROUND NOISE

Figure 2.2: Block Diagram of an FSO System

### 2.2.2 FSO Terrestial Link

FSO Terrestrial links can be used to communicate between sites on the surface of the Earth. A point-to-point or a point-to-multipoint link can be used between the points. Several variables such as beam divergence, atmospheric attrition, and atmospheric volatility affect the transmitted optical beam while moving through two points in free space.

1. **Beam Divergence:** At the aperture end of the transmitter telescope, diffraction causes beam divergence. The divergence determines the amount of signal energy received by the receiver. The transmitted beam of an FSO is diffracted from the limit, causing geometric and misalignment losses. The transmitted optical beam's divergence angle should correspond to the field of vision of the receiving telescope. Misalignment losses occur as a result of building sway (Anbarasi et al., 2017). The beam’s angle of divergence is expressed as:

…………….………………(2.1)

Here, stands for beam divergence angle, stands for laser wavelength and *D* stands for aperture size. The aperture of the receiver lens is, therefore, attuned based on the beam divergence angle in order to shrink losses.

1. **Atmospheric losses:** Atmospheric losses are the result of absorption and scattering in an optical beam. The losses to the atmosphere are calculated using Beer's law. The signal energy of the beam decreases as energy is absorbed by environmental particles. Molecular and aerosol absorbers are the two types of absorbent particles (Kaushal et al., 2017). Molecular absorption is caused by atmospheric gases like N2, H2, and others. Aerosols are particles suspended in a liquid or gaseous media. When light has contact with the particles in a medium, scattering happens, causing light to be redistributed or deflected. Wavelength affects both scattered and absorbed light. If beam's wavelength size is greater than the impacting particle, then it results to Rayleigh scattering. But if similarity in their sizes results to Mie scattering. Also, if particle's mass exceeds the beam's wavelength, non-selective scattering occurs (Kaushal & Kaddoum, 2016).

Some of the meteorological variables that contribute to scattering and absorption are fog, rain, snow and sand.

1. **Fog**: As a result of both scattering and absorption, fog poses the greatest challenge to FSO communication. The particle size affects the fog's density. Fog is divided into four categories: thin fog, light fog, moderate fog, and heavy fog. Fog concentration changes with height, confounding the modelling method.
2. **Rain**: In FSO communication, rain is another factor that reduces the signal's strength. This is because the raindrop's radius is larger than the FSO source's wavelength. In the event of heavy rain, the link may fail. Attenuation of rain however has little effect when compared to fog. Specific attenuation due to rain is less for frequencies below 10 GHz. At frequencies above 10 GHz, the main attenuator of RF links is rain (Mohale et al., 2016). Attenuations are increased for RF connections over 40 GHz. The link is more dependable under a range of circumstances because of the hybrid RF/FSO.
3. **Snow**: Snow has droplet sizes in the range of rain and fog. Snow's attenuation effect is comparable to that of rain and fog. The laser light is blocked by the snowflakes’ magnitude.
4. **Sand**: The scattering impact of sand particles is visible. The availability of links in deserts is reduced by sand particles.
5. **Atmospheric turbulence:** Variability of air pressure and temperature produces atmospheric volatility. The atmosphere acts like various cells known as eddies due to differences in the refractive index. These eddies impede light transmission. The structure of the refractive index coefficient is deployed for turbulence calculation. The value of changes with time. At midday,. At night, the value becomes (Kaur et al., 2015). Three atmospheric turbulence effects are applied to the laser beam which are scintillation, beam spreading, and beam wandering.
6. **Scintillation**: Scintillation occurs when intensity of a propagating wave varies due to random fluctuations in the air. The fluctuating wave's intensity is expressed in terms of the scintillation index . The scintillation index is used in categorizing the intensity of atmospheric turbulence.
7. **Beam Wandering**: When the cells' refractive indices are greater than the beam's, the beam exhibits random fluctuations. The signal quality is affected by beam wandering. As the distance increases, so does the beam wandering. The variance due to the beam wandering effect is expressed as:

…………………………………..… (2.2)

The above equation shows that scintillation has a more complicated influence on beam wandering than a mere wavelength. This implies that at shorter wavelengths, the impact of beam wander will become more pronounced than at longer wavelengths (Khalighi & Uysal, 2014).

1. **Beam Spreading**: An optical beam's spread over the atmosphere is referred to as beam spreading. The average aperture radius should be raised in order to decrease beam dispersion. A halo effect is produced as the beam is transmitted close to the receiver aperture. Atmospheric turbulence forces the beam to spread when its size exceeds that of the eddy cell’s. The beam's field of view is enlarged by turbulence. The effective radius is expessed as:

……………………..…………… (2.3)

1. **Ambient Light**: Background noise is mostly produced by ambient light generators which include the sun, moon, and fluorescent lights. Signal-independent white Gaussian noise is used to model these phenomena. When they're detected, they're amplified and added to the detector's background noise." To model background noise, Poisson random variables are used. The use of dual wavelength transmission and differential mode data detecting methods reduces the influence of background noise (Leitgeb et al., 2010). Due to atmospheric losses and beam divergence, space links face some challenges similar to terrestrial link failures.

### 2.2.3 FSO Challenges

The optical wave in Free Space Optical Communication passes across free space, which would be susceptible to numerous disturbances. Wave attenuation is caused by disturbances such as absorption, dispersion, and turbulence. These disruptions alter the beam's electromagnetic properties, structure, and direction, affecting the optical link's general efficiency. The unpredictability of weather such as fog, rain, and haze affects the distance between FSO connections (Feng & Zhao, 2017).

Figure 2.3: FSO Link Impairment (Kaushal, 2018)

### 2.2.4 FSO Channel Models

The intersection of the transmitter and the channel stimulus reaction produces the optical signal of a receiver. To simulate the routing path connecting the transmission and reception devices, it uses a frequency band. Optical communication systems in free space take advantage of the atmosphere. Because of air losses and scintillation, the signal intensity of the optical wave is decreased. Based on the refractive index structure parameter , the turbulence degree is categorized as mild (, intermediate and severe (Alzenad et al., 2018). For mild, intermediate, and severe turbulence networks, several distribution features are simulated and explained below.

1. **Log-Normal Distribution:** The Probability Density role of a Log-Normal Distribution is expressed as:

…………………..…… (2.4)

Where scintillation index is represented by , becomes the irradiance when there is no turbulence. Utilizing log-normal distribution gives a short FSO link with weak turbulence of the order of 100 m. Log-normal distribution and diversity techniques work well in weak to moderate atmospheric conditions(Prokes, 2009). The normalized estimation inaccuracy increases as the Signal to Noise Ratio (SNR) or degree of fading increases. (Moradi et al., 2010).

1. **Diversity and Combining Techniques**: Log-Normal distribution SNR is approximately equal to the scintillation index in the Log-Normal distribution. With an increase in SNR, the probability of an outage increases when the transmitter is unaware of the channel state information. Power control algorithms in Multiple Input Multiple Output (MIMO) systems are applied when the transmitter has access to channel state information, and 3.7 dB gains are achieved (Letzepis & I Fabregas, 2009). Ground-to-satellite links in weak turbulence conditions can benefit from transmit diversity techniques as turbulence increases (Viswanath et al., 2014).
2. **Relaying Technique**: For atmospheric turbulence of low to moderate intensity Log- The normal distribution is simulated. A slight change in refractive index value affects the performance of short links. The aperture size can be increased to avoid this variation. Even when the aperture diameter is increased for lengthier range communications, this remains the case. You can use relaying in this situation (Zedini et al., 2015).
3. **Negative Exponential Distribution:** Modelling the channel can be done using a Negative Exponential (NE) distribution when there is severe air turbulence or when the Scintillation Index is 1. When communicating over long distances, NE can be used (Aladeloba et al., 2012). This is the PDF for the NE distribution (Li et al., 2007)

………………..……….. (2.5)

Where is irradiance, NE distribution is employed during severe Turbulence. When the average electrical SNR is high, the average Bit Error Rate (BER) decreases and the potential increases in the Negative Exponential distribution (Nistazakis et al., 2011).

1. **Diversity and combining techniques**: Irradiance fluctuations are minimized due to the spatial variety, and a BER of 106 is obtained with 12 dB gain for two photodetectors for severe turbulence. After a given couple of large-scale swirls, the K-distribution inclines to turn into a negative increasing distribution (Aladeloba et al., 2012).
2. **Gamma-Gamma Distribution:** For intermediate to severe turbulence situations, the Gamma Gamma distribution is used. Large-scale eddies and small-scale eddies (GG) are combined in the Gamma Gamma distribution (Uysal et al., 2006).

…………….…. (2.6)

And below are the PDF of the parameters and

……………………..……….. (2.7)

………………..…………… (2.8)

where and are the functional quantity of the scattering process' small and large scale eddies, and is the reformed Bessel function. Operational rate, connection range, and lens aperture are all connected to and characteristics (Naila et al., 2011).

1. **K-Distribution:** The exponential and gamma distributions are combined to create the K-Distribution. The K-Distribution channel is utilized when there is turbulence. As a result, the experimental and theoretical K-Distribution results agree. The K-Distribution PDF is as follows: (Ibrahim & Gucluoglu, 2019).

…………..… (2.9)

Where is the finite amount of scatterers’ parameter

1. **I-K Distribution:** Regarding mild to severe atmospheric turbulence, it is feasible to simulate the I-K distribution. The acquired electrical SNR value is modeled using the I-K distribution. The PDF for the I-K distribution is as follows (Niu et al., 2011).

if

if …………….. (2.10)

where is average SNR, a is scatterers’ amount, is electrical SNR of the acquired signal and is the coherence parameter. is the reformed Bessel function of the initial type while is the reformed Bessel function of the other type.

### 2.2.5 FSO Link Equations

The component values and system characteristics that are considered to be recognized and fixed in advance are described in the link achievement review. As a result, if all of the components and operating parameters are accurately defined, the link performance research may be completed. The following are the three fundamental techniques of evaluating the results of an optical connection:

(i) Calculate the value of observed signal photons at the detector while accounting for transmitter, channel, and receiver losses.

(ii) Calculate the value of observed background noise photons produced at the detector.

(iii) Do a comparison of the value of observed signal photons to the amount of detected noise photons.

On the transmitter side, the optical source provides optical power with a variable degree of focusing, which is generally described by its emission angle. The total power (in Watts) emitted by a uniform source with brightness function *B* (watts/steradian-area), surface area , and emission angle is calculated as follows: (Zhu et al, 2002)

………..…...….……….. (2.11)

The solid emission angle may be linked to the planar emission angle for balanced reflecting sources by

………………………..(2.12)

In the case of any Lambertian supply that releases energy in a regular onward trajectory, implying . It gives and therefore . With the aid of a converging lens, the light from the original beam is concentrated on a point, while the divergent lens broadens the beam to a planar beam Diameter provided by

………………………. (2.13)

where is the operating wavelength, is the transmitter lens diameter, and R is the distance from the lens or link range.

For ………………………….. (2.14)

The first requirement means that the diameter of the emerging light is equal to the diameter of the transmitter lens. The second condition states that when the distance between the source and the emerging light increases, the light diverges. The diverging light source's planar beam angle, also known as the diffraction-limited transmitter beam angle, is roughly given by (Gagliardi, 1995)

………………………………….. (2.15)

For far field region,

………..……………...………………….. (2.16)

The two-dimensional solid angle can be related to planar beam angle by

…………….. (2.17)

The transmitter gain from equation (6) and (7) is given by

………………………………. (2.18)

After propagating through link distance R, the field intensity of the beam will be

………………………… (2.19)

A typical reception area A located in the beam gathers the field power

………………………… (2.20)

defining the receiver gain in terms of A

………….……………. (2.21)

From equation (10) and (11)

………………….………………. (2.22)

when other loss factors are introduced, the equation above becomes

………………………. (2.23)

Where is the signal strength at the photodetector's input, transmitter power, efficiencies of transmitter and receiver optics, respectively, : is the gain of the transmitting antenna, gain of the receiving antenna, transmitter pointing loss factor, : space loss factor, where R is the link distance, : narrowband filter transmission factor.

Any of the following options can be used to boost the received signal power using the above-stated equations: Increasing the transmit power (ii) increasing the transmit aperture (iii) increasing the receiver aperture (iv)Improving overall efficiency by lowering pointing loss.

### 2.2.6 Benefits of FSO

Using Free Space Optical Communication instead of Radio Frequency Communication has a number of advantages. The wavelength of FSO and radio frequency (RF) differs significantly. The wavelength of radio frequency (RF) is larger than the wavelength of light, which is why FSO is better than RF (Willner et al., 2017).

1. Less Power Demand: Due to the small beam inconsistency, the optical force at the receiver is higher than the RF. When compared to an RF, the FSO has a shorter wavelength, resulting in a smaller antenna.
2. Unlicensed Spectrum: The core dissimilarity between RF and FSO is spectrum licensing. FSO does not require a spectrum license, which makes deployment simple and cost-effective. Interference in the radio frequency spectrum requires spectrum licensing. Line-of-sight communication is required for FSO.
3. Massive Bandwidth: High-speed transmission increases as the carrier frequency rises. Optical communication has a considerably higher carrier frequency than RF transmission.
4. High-level security: No walls can be breached by a laser beam, so data transfer is secure. A spectrum analyzer can't detect FSO beams like it can RF.
5. Ease of Installation: FSO is a low-cost alternative to Fibre Optic Cable that may be placed almost anywhere. The FSO transceiver system is small and portable, with plenty of room for expansion.

### 2.2.7 FSO Application Areas

As seen below, the FSO Communication System has a wide range of potential applications.

1. **Inter-Campus Connectivity:** There is a great deal of connectivity traffic of various types such as fax, phone, data, and multimedia that overwhelms ordinary networks, whether on school/higher institutions campuses or in the business sector. FSO systems, unlike specialized fibre optic connections, may deliver ultrahigh speeds at a cheap cost (Ghassemlooy et al., 2019)

(ii) **Cellular Systems Backhaul**: FSO may be very useful in cellular systems for transporting high-speed cellular telephone traffic along with high bandwidth from the mast towers to the Public Switch Telephone Network (PSTN), resulting in increased transmission speed (Ghassemlooy et al., 2019).

(iii) **Monitoring/Video Surveillance**: In today's world, video cameras can be found almost anywhere, whether it is in the grocery store, the airport, the mall, or the military. Video cameras are easy to install. Traditional wireless technology, on the other hand, cannot deliver the high speed necessary for video broadcasts. High-quality video transmission is possible using FSO technology, making it a viable alternative to conventional techniques.

(iv) **Military Access**: Because it sends data with limited beam width, FSO is both protected and untraceable. It is suitable for military access due to its ability to create secure connections across broad regions with minimum preparation and deployment time. Farooq et al. (2018a).

1. **Point-to-Point and Point-to-Multipoint Communication:** FSO communication is possible through multipoint lines, such as those connecting a satellite to the ground, for both short and long-range connectivity, as well as point-to-point communications between two stores or structures (Kaushal, 2018). For LAN-to-LAN connections, FSO provides fast ethernet or gigabit ethernet speeds on campuses.

**(vi) Last-Mile Access:** Every industry that relies heavily on telecommunications, cable television and Internet services has always struggled with the "last mile." To solve this problem, FSO can be deployed as part of a network in the last mile (Ghassemlooy et al., 2019). FSO can be used to upgrade wireless technologies that are already in use.

1. **OFC Backup:** For important fibre connections, the FSO provides a reliable backup system in cases where there is Fibre failure that needs to be quickly resolved for optimal network performance.

## 2.2.8 Overview of FSO investigation

Signals travelling in a free space channel are not directed, unlike those travelling in an optical fibre channel. They have to deal with beam divergence and constantly changing air conditions. A line of sight between the transceivers is required in an FSO channel, so signal loss could be caused by physical barriers in addition to free space attenuation caused by air particles. One of the factors affecting signal transmission in an FSO channel is atmospheric turbulence (Kaushal & Kaddoum, 2017).

The indoor FSO channel is more suitable for optical signal propagation since it is not affected by weather or air turbulence. Indoor FSO communications face a number of challenges, including transmitter power limitations due to eye safety, bandwidth and user data rate limitations due to multipath dispersion and slow receiver response times, user mobility restrictions, multiple user service provision, and interference from background ambient radiation (Raj, Arockia Bazil and Majumder, 2019). Many kinds of link configurations, including direct, non-direct, and hybrid line of sight connections, as well as transceiver optimization, have been extensively explored in order to find a solution to the mobility and transmission power issues in indoor FSO communication systems. Multiple access methods like Code Division Multiple Access (CDMA), WDM, and subcarrier multiplexing have also been suggested and investigated for the provision of multi-user services (Cvijetic, 2012).

An FSO communication system based on Intensity Modulation and Direct Detection (IM/DD) is proposed in this study. It is possible to broadcast amplitude modulation formats without spatial correlation, including ON-OFF Keying (OOK) and Pulse Position Modulation, using such a paradigm (PPM). At the transmitter, the user data modifies the optical beam. The optical signal is then sent via a turbulent channel, generating scintillation (irradiance changes) (Li et al., 2007). The receiver's output photocurrent is linked to the incoming laser source via the responsivity R, and the transmitted signal is well modelled by assuming clear weather with negligible air losses.

……………….………………….. (2.24)

where x is the transmitted signal strength and n is the Additive White Gaussian Noise (AWGN). Atmospheric turbulence, detecting error, and geometrical route loss cause the channel state to appear (Farid & Hranilovic, 2007).

Small changes in temperature generate turbulence scintillation in the atmosphere, resulting in changes in the refractive indices. The scintillation index measures the intensity of turbulence and is proportionate to the Rytov variability (Khalighi & Uysal, 2014).

………………….………………… (2.25)

where is a parameter for the index of refraction ranging from to for weak turbulence and strong turbulence respectively (Benkhelifa et al., 2013), is optical wave number, is the signal wavelength and L is the length of the transmission link. are related to mild and severe turbulence. Under mild and severe turbulence, the irradiance distribution is modelled by Gamma-Gamma Probability Density Function (PDF) by (Al-Habash, 2001)

, I>0…………………… (2.26)

where the Gamma function is denoted by , the second kind modified Bessel function with order is . is the shape parameter which indicates the dispersed process effective amount of large- and small-scale cells. The two parameters of scintillation are given by (Al-Habash, 2001) after neglecting the inner scintillation.

………………………………… (2.27)

and

……………………….……. (2.28)

It is worth noting that turbulence-induced fading rises with the length of the transmission connection resulting in reduced system (Sharma et al., 2015).

Due to the requirement for FSO line-of-sight, pointing accuracy is a critical factor. The combination of wind load and continuous sun heating creates random building sway, which results in aiming inaccuracies. The geometric loss and channel state caused by pointing inaccuracies for a Gaussian beam with receiver beam waist and aperture size radius of , respectively, are given by (Farid & Hranilovic, 2007).

…………………………………. (2.29)

Where is a percentage of total collected power at r=0, , r is the radial shift and is the equivalent bandwidth. Rayleigh distribution is deployed to simulate the radial shift r defined by (Farid & Hranilovic, 2007)

………………………………. (2.30)

where is the variance of the jitter at the receiver and the pointing inaccuracy extremity is given by variable with values between 0.5 and 5. For high jitter difference, the variable.

Combining domain competence and data-driven algorithms is among the greatest troubling elements of using ML techniques for scientific applications. The optical communications industry, notably in the free space optical communications community, has previously employed machine learning approaches. A data-driven fibre channel Deep Learning (DL) modelling approach was implemented in an optical communication system. A bidirectional Long Short-Term Memory (LSTM) was used to mimic fibre channel signals for On-Off Keying (OOK) and Pulse Amplitude Modulation 4 signals (Ramprasad et al., 2017).

To repair the deformed vortex beam and improve the performance of orbit angular momentum (OAM) multiplexed transmission, a deep learning-based atmospheric turbulence correction method was developed. For all turbulence strength regimes, FSO related channel modelling is investigated using DL techniques, and the findings provided show that DL can deliver a pretty near to the faultless channel estimate methodology. Through theoretical and experimental data, the capability to reduce the negative effects of air turbulence on FSO system performance utilizing Artificial Neural Network (ANN), Generative Machine Learning (GML), and Convolutional Neural Networks (CNN). For Cn estimate, ML approaches using DT, RF, and ANN were compared to macroscopic meteorological parameters. Other machine learning approaches have been utilized to construct regression and classification models to estimate Received Signal Strength Indicator (RSSI) for a hybrid RF/FSO detector given an overview of the next generation of artificial neural networks (ANNs), also known as optical neural networks (ONNs), as well as past research in the field. The study is unique in two ways: a) it compares multiple conventional machine learning techniques for the first time in forecasting RSSI values, specifically for the domain of interest, Piraeus, Greece; and b) it compares multiple conventional machine learning techniques for its first time in predicting RSSI values, specifically for the domain of interest, Piraeus, Greece. Relevant prior studies in terms of the machine learning methods used (i.e., KNN and ANN), both of which functioned successfully, and b) provides a unique experimental data analysis relating to the connection's terrain (i.e., marine), which differs from a terrestrial one. Machine learning techniques such as Nave Bayes, Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and Random Forest can help categorise data into a variety of classes (Wang et al., 2021).

## 2.3 Machine Learning

The process of using a computerised algorithm to detect meaningful patterns in data is known as machine learning. Machine learning might be considered a subclass of AI since the ability to convert information into skill or uncover important correlations in complex sensory data is a fundamental component of human (and animal) intelligence. Machine learning, unlike conventional AI, is not aiming to create automated imitations of intelligent behaviour, but rather to augment human cognitive abilities by using computers' strengths and particular skills to do tasks that are well above human capabilities (Song et al., 2019).

It has become a popular instrument for practically any task involving the retrieval of information from large data sources over the last two decades. Bioinformatics and medicine, for example, are two fields where machine learning is applied. The intricacy of the patterns that must be detected is a common thread that runs through all of these applications. Unlike more traditional uses of computers, human programmers are unable to offer unambiguous, fine-grained instructions on how such operations should be carried out. Many of our abilities, like those of intelligent entities, are learned or improved through experience. Machine learning technologies aim at providing programs the ability to "learn" and adapt to new situations (Kourou et al., 2015).

Machine learning aims to address the question of how to build computers that automatically learn from their experiences and improve. It is one of the most rapidly developing research areas today, bridging the gap between statistics and computer science and serving as the foundation for artificial intelligence and data science. Data-intensive machine-learning methodologies are being accepted across technology, sciences, and commerce in a wide range of fields, including manufacturing, health care, financial modelling, education, law enforcement, and marketing, culminating in more substantial proof in making decisions (Jordan & Mitchell, 2015).

Machine learning refers to a collection of technologies that enable a trained algorithm to evaluate and interpret data. Machine learning algorithms are divided into two types: supervised and unsupervised. Labelled examples, such as an input with a known output, are used to train supervised learning algorithms. Unsupervised algorithms, on the other hand, do not assume a known output for a set of input values (Lionis et al., 2021).

As stated in eqn. (2.18), X denote the predictors or input variables whereas the response or the output variables is denoted by Y

………………….……………………… (2.31)

where f is an unidentified function and is a zero mean random error. The goal of supervised learning is to estimate the function f for prediction and inference. After estimating f, one can forecast the output for a given set of input variables, prediction's accuracy of which is dependent on the so-called reducible and irreducible errors. Inference, is the study of the relationship between inputs and outputs, as well as the impact of changes in inputs on its corresponding outputs (Lionis et al., 2021).

There are numerous powerful techniques available in the field of machine learning, including calculating the values from noisy data signal measurements, establishing intricate mappings between input and output data, inferring probabilistic models, and forecasting the output based on previous input data, to name a few (Thrane et al., 2017a). The study of algorithms that learn and evolve through data and use them to do tasks such as predicting an unknown parameter, making judgements, and enhancing mathematical frameworks is known as machine learning.

### 2.3.1 Machine Learning Task Benefit

Humans and animals undertake a wide range of activities on a regular basis, but our insight into how we accomplish them is insufficient to construct a well-defined program. Driving, speech recognition, and visual interpretation are just a few of the numerous jobs that may be accomplished this way. After being exposed to enough training instances, modern machine learning systems that "learn from their experience" can achieve good outcomes in all of these categories (Moubayed & Shami, 2020).

Machine learning techniques may be used to examine huge and complicated data sets, which are tasks that are beyond the capabilities of humans. Astronomical data and medical archives, for example, may be utilized to improve health care, as well as weather predictions and genetic data analysis. With the growing availability of digitally recorded data, it's becoming clear that there are valuable pieces of information hidden in data archives that are much too massive and complicated for humans to understand. The convergence of learning algorithms, practically endless storage capacity, and ever-increasing computer processing power opens up new research and development frontiers in locating patterns in enormous and complicated data sets (Al-Jarrah et al., 2015).

### 2.3.2 Machine Learning Types

Machine learning is the process of creating programs which enable a computer to learn. You don't have to be aware of statistical regularities or other data patterns to find them. As a result, machine learning algorithms will be substantially different from how people learn new tasks in the actual world. Learning algorithms can reveal the relative difficulty of learning in a variety of situations. The intended output determines how automated machine learning algorithms are categorized. supervised learning, unsupervised learning, and semi-supervised learning, are amongst the most frequent types of algorithms (Alpaydin, 2020).

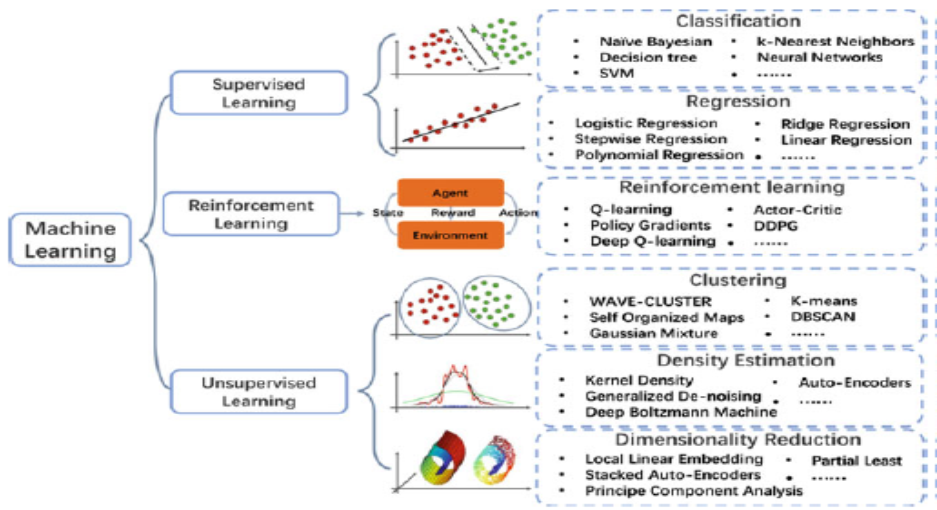


Figure 2.4: Machine Learning Type Chart

There are several ways for training machine learning techniques, each with its own set of benefits and drawbacks. Let's start with the data that each form of machine learning makes use of. ML works with both labelled and unlabelled data.

Labelled data provides both input and output factors in a machine-readable format; nevertheless, labelling the data necessitates a substantial amount of manual labour to begin with. If the data is unlabelled, just one or none of the parameters are in a machine-readable format. Despite the fact that human labour is no longer necessary, more complicated methods are required to achieve this. Apart from a few other types of machine learning techniques which can only be used in very specific scenarios, the three most prevalent methodologies now in use are described here (Nadikattu, 2020).

#### 2.3.2.1 Supervised Learning

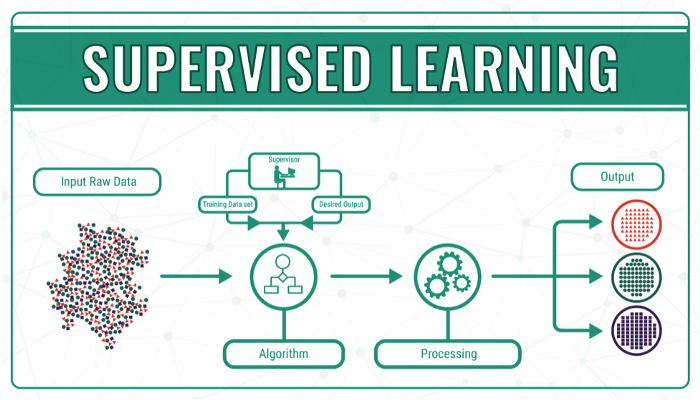
This is a basic example of how tagged data can be used to develop a machine learning system. supervised learning is highly successful, despite the fact that the data must be appropriately labelled while being engaged in the suitable environment. In supervised learning, a machine learning algorithm is given a small training dataset. This training dataset, which is a subset of the broader dataset, provides the algorithm with a rudimentary grasp of the problem and the data points that must be handled.

Figure 2.5: Supervised Learning model. (van Loon, 2018)

The training dataset, like the final dataset, has many of the same qualities as the final dataset and gives the algorithm with the labelled criteria it needs to conquer the problem. The programme then uses connections between the parameters to establish a cause-and-effect link between the variables in the dataset. The algorithm has a decent concept of how the data works and how it connects to other parts at this point. It is tested with the final dataset before being deployed, and it learns in the same way as it did with the training dataset. As a result, supervised machine learning algorithms will keep improving as they learn from new data, identifying new patterns and links (Jordan & Mitchell, 2015).

Supervised learning aims to discover the mapping between two data sets (e.g., x, y; x X, y Y), one of which (y, or label) is the mapping of the other (x, or feature), and the goal is to locate that mapping (in training phase). As a consequence, a cost function based on the pair mismatches should be constructed, and the cost function should be minimized (e.g., MMSE or cross-entropy) (e.g., gradient descent algorithms). Pattern recognition (classification) and regression are two examples of supervised learning applications (function approximation). Supervised learning is done with the help of an instructor who offers constant feedback and instruction on the accuracy of the solutions gained (Cunningham et al., 2008). SVM may be applied to any situation in which there are two or more sets of data. Incoming bit sequences, for example, can be categorised as 0 or 1, or fault and normal fibers can be detected. Its method is straightforward: it takes data spanning a range of sectors and then determines how to categorize them based on features obtained from them. Because SVM ignores the entire data and instead looks for hyperplanes using marginal data, it has a lower overall complexity. Furthermore, redundant characteristics could be removed using SVM, lowering the computational cost and complexity. Although the use of training data causes transmission delays, the testing phase of SVM has a low latency (when separating hyperplane is defined). SVM has the potential to succeed in areas where previous approaches have failed, such as establishing a reliable defect diagnosis, which SVM can address by converting the classification issue into a nonlinear programming problem. Fortunately, numerous SVM-based research in ML for OC exist, despite the use of other machine learning techniques, allowing these works to be better classified. SVM is widely used in OC for detection, although it is mainly unexplored in fibre OC and has limited uses in WOC. In contrast, SVM as a binary classifier is strictly utilised for binary signalling rather than M-ary signalling. One of the most successful M-ary SVM solutions is the bit-based SVM, which requires only log2 (M) SVMs (hyperplane) instead of one hyperplane for M-ary signal detection. A decision tree, which converts a multi-class classification issue from M-ary to binary and multi-layer, is another M-ary SVM approach. (Amirabadi, 2019).

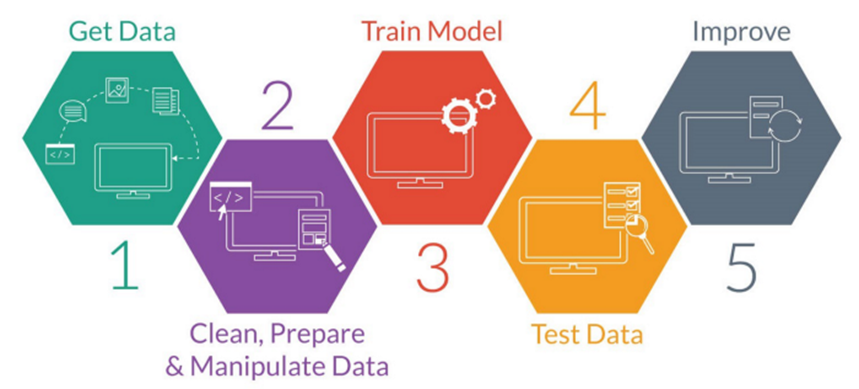


Figure 2.6: Supervised Learning Steps. (van Loon, 2018)

#### 2.3.2.2 Unsupervised Learning

The advantage of unsupervised machine learning is that it can handle unlabeled datasets. As a result, there is no need for human intervention to make the dataset machine-readable, allowing the programme to handle much larger datasets.

Labels enables an algorithm to discover a relationship between any two data elements in supervised learning. As a result, because there are no labels to deal with, unsupervised learning produces hidden structures. There is no need for human input because of the algorithm's abstract vision. Because of the construction of these underlying structures, unsupervised learning algorithms are so varied and powerful. Unsupervised learning algorithms can adapt to the input by dynamically adjusting hidden structures rather than relying on a pre-defined problem statement. Learning algorithms that are supervised do not allow for as much post-deployment development (Chibani & Coudert, 2020). The data is unlabeled in unsupervised machine learning, and the programme tries to find patterns and structures in it. As a result, crucial core properties are preserved, and a new dataset with reduced redundancy is created. When the algorithm detects problems with connectivity, it splits the data into two groups, each with a similar feature. Unsupervised learning, on the other hand, operates purely within the data itself, without the need of any external references. Unsupervised learning is a form of machine learning approach that uses input data without labelled replies to make inferences (Zhou et al., 2017).

Unlike supervised algorithms with well-known terminology (e.g., SVM, ANN, etc.), clustering algorithms are primarily defined by the notion of clustering, and the programmer typically specifies what the algorithm should achieve (logical rules in perspective). Non-iterative density peak search clustering algorithms, for example, employ two fundamental parameters to discover cluster centres without iteration, whereas Fuzzy c-means does the same thing over and again; nonetheless, hieratical clustering provides a solution blindly. Another clustering method known as fractionally spaced clustering employs the Viterbi type process and the Mahalanobis distance metric, which is determined by the particular crossover between the groups. Unsupervised (clustering) algorithms provide a lot of versatility, and they don't require labels, which makes them quite useful. They do, however, have a single goal: to locate the cluster's centre and expand it (Sarker, 2021)**.**

The most often deployed unsupervised algorithms in OC are k-means and Principal Component Analysis; nevertheless, several additional techniques, such as Hierarchical and Fuzzy logic C-means clustering, cannot be classified into a major group. These techniques are less sophisticated than supervised algorithms and seek to identify hidden patterns or groupings in data; nevertheless, they take longer to complete. These tactics have outperformed standard methods in MFI, NLE, and other areas, despite their generality and breadth. Furthermore, in contrast to earlier methods, By employing indiscriminately, these systems can survive stochastic parametric noise amplification while reducing complexity and processing time (e.g. deterministic Volterra NLE) (Kanungo et al., 2002).

The K-means approach divides data into k groups, and each data point is assigned to the cluster with the lowest mean. Even though this job needs a lot of processing, heuristic solutions that converge to a local optimum can be applied. To lower the sum of squared errors of data points, "k" is the number of clusters that should be discovered among L observations. Picking K random centre cluster sites is the initial stage in this procedure. Then, based on the clusters provided (the summing of the cluster points), each data point is assigned to the cluster with the closest centre, and a new cluster centre is generated (Hatamlou et al., 2012).

Learning from labelled observations is used to anticipate the optimum set of data for classification algorithms. The "learning" process is characterized by several input attributes. These algorithms are frequently used in machine learning because they are good in classifying data that has never before been seen in their various categories

#### 2.3.2.3 Reinforcement Learning

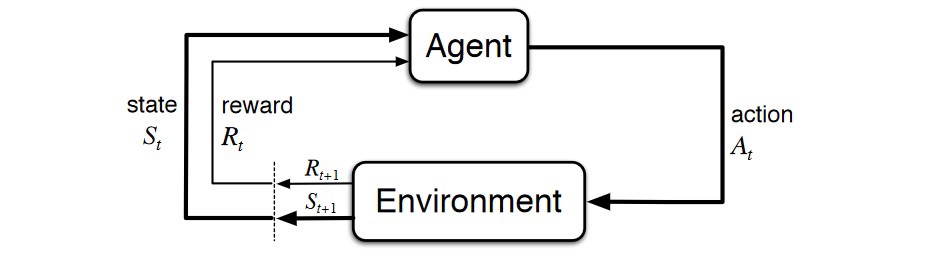


Figure 2.7: Reinforcement Learning Analogy (van Loon, 2018)

Reinforcement learning has a direct impact on people's capacity to learn from data in their daily lives. The algorithm learns from new conditions and improves itself using the trial-and-error technique. Negative consequences are discouraged or 'punished,' while positive outcomes are rewarded or 'reinforced.'

Reinforcement learning is based on psychological conditioning and involves putting the algorithm in a work environment with a translator and a compensation structure. Each time the algorithm is executed, the translator receives the final output and decides if it is favourable or unfavourable (Barto, 2019).

### 2.3.3 Machine Learning Application Areas



Image source: http://www.cognub.com/index.php/cognitive-platform/

Figure 2.8: Machine Learning Application Area

## 2.4 Machine Learning Predictive Model

Predictive modelling (otherwise known as predictive analytics) is a statistical approach that employs machine learning and data mining to predict and forecast likely future events based on past and current data. It makes predictions by analysing current and historical data and then projecting what it has learnt onto a model that was developed to forecast future scenarios. Predictive modelling may be used to generate accurate guesses in practically any situation.

When a predictive model is created, it is constantly verified and improved in order to account for changes in the underlying information. In other words, this isn't a prediction that will be fulfilled in a single sitting. Predictive models make predictions about the future based on what has occurred in the past and what is now occurring (Lantz, 2019).

Basically, the top five predictive analytical model today are classificiation model, clustering model, forecast model, outliers model and time series model

### 2.4.1 K-Means Clustering Algorithm

Clustering is a data categorization technique that looks for hidden patterns in large datasets. Clustering is a method for separating data items into fragmented chunks such that data in one group matches data in another. K-means is a non-deterministic, iterative, numerical, and unsupervised technique. Because of its simplicity and quickness, the technique has proven to be a successful strategy for getting quality clustering results in a variety of real-world contexts. However, it is capable of forming globular clusters (Z. Zhang, 2016). K-Means clustering is a standard data mining clustering technique for categorizing big amounts of data.

There are two sections to the algorithm.. The first step selects k centres at random from a pool of k. In the next stage, each data object will be transported to the nearest data centre (Z. Zhang, 2016).

The Euclidean distance has been commonly used to calculate the distance between each data object and the cluster centres. After all of the data elements have been grouped, the first step is completed. The average of the early-forming clusters has been recalculated. After that, the criterion function is reduced to its most basic form (Na et al., 2010).

Criterion Function is defined by

…………………………………. (2.32)

Supposing x is the target object, the average of cluster C is and E is the sum of the squared error of all database object. Euclidean distance is used to determine the closest distance between each data object and cluster centre (Rokach, 2009). The Euclidean distance between the two vectors and is given by

……………………………. (2.33***)***

k-means algorithm process is as follow:

Input: Number of desired clusters, k, and a database D={d1, d2,…dn} containing n data objects.

Output: A set of k clusters

1. Select k data items at random as the first cluster centres from dataset D.

2) Repeat 1;

3) Calculate the distance between each data object di (1 <= i<=n) and all k cluster centres cj (1<=j<=k) and assign data object di to the nearest cluster. 4) For each cluster j (1<=j<=k), recalculate the cluster centre. 5) until no changing in the centre of clusters (Na et al., 2010)

### 2.4.2 K Nearest Neighbour

The k-nearest neighbours (KNN) approach is a simple, easy-to-implement supervised machine learning methodology - for handling classification and regression problems. It has proven to be extremely successful at specific jobs despite its simplicity. Conversely, unsupervised machine learning, uses labelled data to create a function that produces the intended output when given more unlabeled data. The KNN function uses Euclidean distance by default, which may be computed using the following equation (Liao & Vemuri, 2002).

……… (2.34)

where p and q are subjects to be compared with n characteristics

In addition, the parameter k in the KNN algorithm determines the number of neighbours. The choice of k has a big influence on the diagnostic efficiency of KNN algorithm. However, because of the high k, it raises the danger of missing minor but essential patterns. Overfitting and underfitting must be balanced in order to choose an appropriate k value (Zhang, 2016).

To classify data, the kNN algorithm compares the data from the test dataset with data from a training dataset. It is possible to evaluate how well the KNN model performs because we know what categories of findings are in the test dataset. The following equation is often used to define average accuracy.

………………………. (2.35)

The letters TP, TN, FP, and FN stand for true positive, true negative, false positive, and false negative, respectively. The subscript i signifies category, while is the whole category.

The KNN Algorithm involves

1. Load the data
2. Initialize K to your chosen number of neighbours
3. For each example in the data
   1. Using the data, calculate the distance between the query example and the current example.
   2. Add the distance and the index of the example to an ordered collection

(iv) Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances

(v) Pick the first K entries from the sorted collection

(vi) Get the labels of the selected K entries

(vii) If regression, return the mean of the K labels

(viii) If classification, return the mode of the K labels

### 2.4.3 Support Vector Machine (SVM)

The non-probabilistic characteristic of a support vector machine (SVM) distinguishes it from other binary linear classifiers like Nave Bayes. SVM splits data along a classifier (plane) selected by a tiny proportion of the data (feature vectors). Support vectors are a subset of data that aids in the understanding of a boundary.

The SVM divides data items in a feature space into two classes, and to be categorized, they must include features to as well as a class label. As a consequence, each data object is allocated to one of two classes and is regarded as a point in feature space. The feature vector of an object decides whether it corresponds with the class = 1 or not, = -1. As a result, data has the following definition: (Justo-Silva et al., 2021)

…………………………. (2.36)

Where n represents the number of data points and p, the dimension of the feature vector.

SVM classifier looks for a decision boundary in feature space that divides data objects into two classes, so as to categorise them. The goal of the optimization is to identify the decision boundary (a linear hyperplane) with the greatest gap (margin) between two classes. For example, on either side of a hyperplane, the margin is the distance between equidistant parallel hyperplanes, with no data objects in the gap between the two planes. During training, an optimal hyperplane is found that has the largest margin. Once the feature vector for a new data object has been provided, the SVM uses this hyperplane to forecast the class of the data object (Ben-Hur et al., 2008).

Classifiers frequently report on performance metrics such as accuracy, precision, and recall.

The variables are specified as follows:

*TP* = True Positive, *TN* = True Negative

FP = False Positive

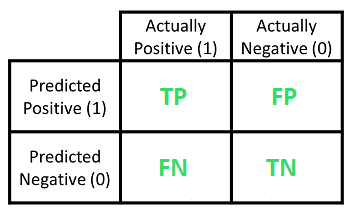
FN = False Negative

*P* = Number of Real Positives in the data

*N* = Number of Real Negatives in the data

The performance of a Classifier is summarized in a table called Confusion Matrix shown below.

Table 2.0: Confusion Matrix Sample



Confusion Matric is simply a visual representation of the model performance, where TP (True Positive) which is number of observations classified as ‘Positive’ by the model and they are actually ‘Positive’

FP (False Positive) is the number of observations classified as ‘Positive’ by the model but they are actually ‘Negative’ and also known as Type-1 error.

TN (True Negative) is the number of observations classified as ‘Negative’ by the model and they are actually ‘Negative’.

FN (False Negative) is the number of observations classified as ‘Negative’ by the model but they are actually ‘Positive’ and also known as Type-2 error.

Many large numbers will appear in the matrix's "left diagonal" to indicate that it is a good classifier.

The accuracy of a classifier is given by

……………..…………………. (2.37)

The misclassification error rate is . The specificity is which is also equivalent to true negative rate.

p stands for precision, and it represents the accuracy with which classification can make predictions about a given class, hence for positive class,

……………………………………………….. (2.38)

Classifier performance can be compared using precision, which is a relative measure. In other words, if the classifier performs better at predicting one class than another, it is considered to be unbalanced. The recall r is calculated by dividing the number of correct positive results by the number of positive results that should have been returned.

…………………………………………. (2.39)

score is the harmonic mean of recall and precision such that

………………..………………………. (2.40)

A weighted average of precision and recall is used to calculate it (best is 1 and worst is 0). It is good to note that F1 is 1 when p=1 and r=1, but zero when p or r are both zero.

…………………………………… (2.41)

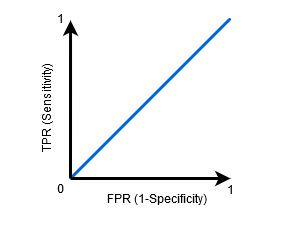
## 2.5 Receiever Operating Characteristics (ROC) Curve

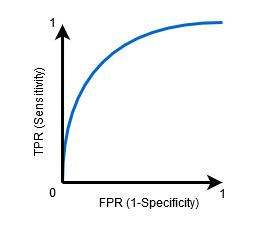
The Receiver Operator Characteristic (ROC) curve is a statistical metric used in machine learning to evaluate binary classification tasks. It is, in essence, a probability curve that compares the True Positive Rate (TPR) against the False Positive Rate (FPR) at varying threshold levels and thereby differentiates between "signal" and "noise."

Plotting the true positive rate (TPR) as against the false positive rate (FPR) at various threshold levels is required to generate the ROC curve. In some cases, the true-positive rate is referred to as sensitivity, recall, or chance of detection. The false-positive rate can be calculated as follows: (1-specificity). As a result, the ROC curve depicts the sensitivity-recall connection as a function of fall-out. (2006, Fawcett)

It's a graph that shows how well a classification model performs across all categorization levels. The Receiver Operator Characteristic (ROC) curve is a statistical tool for evaluating binary classification tasks in machine learning. It's just a probability curve that evaluates the True Positive Rate (TPR) with the False Positive Rate (FPR) at various threshold levels and thereby distinguishes between "signal" and "noise." As the classification threshold is lowered, more items are categorised as positive, resulting in a higher number of False Positives and True Positives in the system. The figures below depicts a standard ROC curve.

ROC curves are a valuable tool in detection/classification theory and hypothesis testing because they allow all important values to be shown together in one plot.





(a) (b)

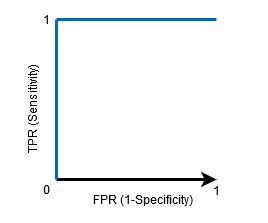


Figure 2.9: ROC Curve (a)(b)(c)

The Area Under the Curve (AUC) is a summary of the ROC curve that measures a classifier's ability to differentiate between classes. The AUC measures a model's capacity to distinguish between positive and negative classifications; the higher the AUC, the better.

The classifier can correctly discriminate between every Positive and Negative class point when the AUC is equal to one, as illustrated in Figure 2.9(a). The classifier might have identified all Negatives as Positives and vice-versa if the AUC had been zero.

The AUC is a measure of a classifier's ability to distinguish between positive and negative class values. When AUC is smaller than 1, as illustrated in Figure 2.9(b), the classifier is more likely to be able to distinguish between positive and negative class values. This is because the classifier can recognise a higher number of True positives and True negatives than False negatives and False positives.

When the AUC is equal to 0.5, the classifier has yet to distinguish between Positive and Negative classification points, as shown in Figure 2.9(c). In other words, for many of the data sets, the classifier predicts either a random or a constant class.

The ability of a classifier to distinguish between positive and negative classes is measured by its AUC. The greater a classifier's AUC value, the better it can distinguish between different classes (Narkhede, 2018).

## 2.6 Gaps Identified in Literature

# CHAPTER THREE

# RESEARCH METHODOLOGY

## 3.1 Research Framework

The FSO dataset is acquired from a reputable repository “figshare” as at the recent available FSO dataset, it is fed into an Unsupervised Learning Algorithm, K-Means Clustering whose objective is to help group or cluster the unlabelled dataset into different clusters that forms different subset which are invariably passed into the classifiers (SVM and KNN) and the results are analysed by using some performance metrics. The Outcome is then compared with existing research work as related to the research area.

FSO DATASET

FSO DATASET

FSO DATASET

FSO DATASET

CLUSTERING (K-Means)

(K MEANS)

0

CLUSTERING (K-Means)

(K MEANS)

0

CLUSTERING (K-Means)

(K MEANS)

0

CLUSTERING (K-Means)

(K MEANS)

0

SUBSET CLASSED LABELED DATA

SUBSET CLASSED LABELED DATA

SUBSET CLASSED LABELED DATA

SUBSET CLASSED LABELED DATA

CLASSIFICATION MODEL (KNN, SVM)

(SVM AND KNN)

CLASSIFICATION MODEL (KNN, SVM)

(SVM AND KNN)

CLASSIFICATION MODEL (KNN, SVM)

(SVM AND KNN)

CLASSIFICATION MODEL (KNN, SVM)

(SVM AND KNN)

RESULT (PERFORMANCE METRICS)

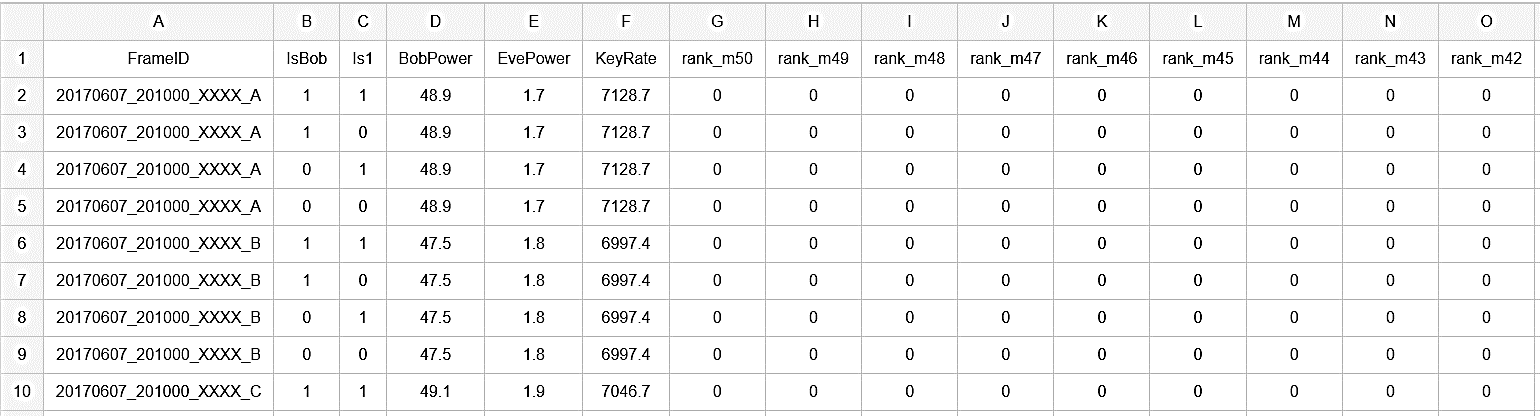
Figure 3.1: Proposed Frame WorkRESULT (PERFORMANCE METRICS)

Figure 3.1: Proposed Framework

## 3.2 Dataset

This study used a dataset containing several batches of received optical signal of an outdoor FSO link (during raining weather conditions) obtained from continuous measurement (3-hours) with a high-speed coherent optical transmission testbed. The baseband electrical signal is digitally modulated at 64 Gbaud using an arbitrary waveform generator (AWG) running at 120 Gsa/s and 45 GHz of analog frequency. Constant composition distribution matcher and PCS were combined with a 64QAM template, which allows for multiple bit-rates, to adjust the sign likelihood of the specified transmitted bit-rate. In the broadcasted signal, we assign 20% overhead (rate 5/6) for forwarding error correction (FEC) and 6.25 percent overhead (rate 15/16) for DSP pilots (which finally led to conventional uniform 64QAM signalling) when the dual-pol PCS-64QAM constellation is filled With an optimal entropy of 6 bits per symbol, the net symbol rate is 50 Gbaud, corresponding to a total net bit rate of 600 Gbps (Dath et al., 2019).

Table 3.0 Data Set Chart



Based on the Research framework in Figure 2.4, the FSO dataset is evaluated and pre-processed for fit into the software that will be used to carry out the experiments (MATLAB, Python), the dataset is passed into a Clustering technique (K-Means) to fetch out relevant information in the dataset before results are generated by the classifiers KNN and SVM. Based on the nature of classifiers selected, the parameters of each of the classifiers will be tuned and the generated models will be tested accordingly. The result will be analysed to examine the effect of dimensionality reduction models on classifiers in software prediction.

## 3.3 K-Means Algorithm

The K-means approach separates quantifiable data (events) into K clusters, each of which contains experimental findings on the same occurrence. Finding the parameter K is a difficult task that is usually done experimentally or with the use of a measure that evaluates the K-means algorithm's performance. The K-means method is recursive, as seen in Figure 3.2, with each cluster receiving one calculated data sample. K-means groups the remaining measured data events into the K clusters in the first recurrent stage with their similarity in view, which is computed using the cluster's centroid event and the least-squared Euclidean distance of the measured data (mean). K-means creates new centroids for each cluster in the second iteration phase. As a result, the cluster's centroid reflects all of the observed data occurrences. The iterative strategy is used until no further re-assignments of resulting data events to clusters can be detected between iterative phases. The iterative procedure assigns the same type of event to the same cluster every time (Slapak et al., 2020).

Assign number of cluster center K

Chose a random initial center

Place each item in the cluster closest to it.

evaluate new cluster center methods

Is the terminating condition satisfied?

Figure 3.2: K Means Algorithm Flow Chart

## 3.4 Support Vector Machine (SVM)

SVM is a well-known machine learning algorithm which can be used for both classification and regression applications. The linear fitting function is described as follows:

……………………… (3.1)

For a set of training points , where is a feature vector, is a target output and where and b are the regression model weight and the parameter respectively and (,) denote the dot product. The goal of SVM regression is to find the function that provides the best fit with the least amount of deviation from the targets.

This is accomplished by taking into account some errors (i.e. noise in the training data) that can be conveyed as slack variables. The following optimization problems are considered to find the best value of w (Esmail et al., 2021)

Minimize ……….…………………… (3.2)

Subject to ………….………………… (3.3)

Where is the regularization parameter which is never less than zero (Esmail et al., 2021). Using the Lagrange equation, this optimization problem can be solved (Esmail et al., 2021) where the best results w in terms of training samples can be obtained and Lagrange multiplier pairs

…………..………………………. (3.4)

As a result, the approximate fitting function is,

…………………………… (3.5)

Hence, the term refers to the linear kernel function. To use SVM in FSO-based monitoring, the mathematical model is built from a training dataset that includes AAH and ADTS features.

## 3.5 K Nearest Neighbour

For regression and classification issues, the KNN algorithm is a straightforward model to implement. In assessing the utility and applicability of any algorithm, we must examine (a) output interpretation ease, (b) computation time, and (c) prediction power. The KNN algorithm's significant benefits are the simplicity with which its result may be read as well as the significantly reduced calculation time. KNN is a simple approach that stores all available instances and uses a similarity metric to estimate the numerical objective (e.g., distance functions). The average of the numerical targets of the K nearest neighbours is a basic implementation of KNN regression. One other option is to use an inverse distance weighted average of the K closest neighbours. The same distance functions are used in KNN regression as in KNN classification. Examining the data first is the simplest way to find the proper number for K. A high Because a high K value reduces total noise, it is frequently more precise; but, it blurs the crisp boundaries within the feature space. Another approach for establishing a suitable K value in hindsight is cross-validation. This is the verification of the K value using an independent data set. The average of all training points in N0 can be stated mathematically as the KNN regression algorithm:

……..……… (3.6)

If xi is the model's input and is its output, is a forecast point, and is the number of points that are closest to . Our experimental data set's coefficient of determination, R2, and RMSE are plotted versus different numbers of neighbours (K). starting at 2 and going up to 50. When K = 50, the coefficient of determination for the training subset starts at 0.95 and gradually declines to 0.75. The equivalent values for the testing subset peaked at 0.85 for K = 3, indicating that this is the best number of neighbours for fitting the KNN model to the data set; as a result, this is the value we chose for our model. A methodical procedure for the optimal K value (K = 3) is determined by scanning the training subset (i.e., grid search) to identify the optimal value for a parameter.

The k Nearest Neighbour (kNN) approach is a non-parametric machine learning methodology that classifies input based on a majority vote of the k nearest training neighbours in the feature space. If k = 1, for instant, the input is sent to the class with the closest neighbour. The input point is *R*, and the training data is *Li*, therefore (𝑅, 𝐿𝑖) = √(𝑅 − 𝐿𝑖) 2. The estimated neighbour is classified by a majority vote of its k nearest neighbours, and then **D***(R, Li)* is sorted to discover the k nearest neighbours. The function is approximated locally in KNN, and all computations are postponed until classification. kNN detection is equal to maximum likelihood and other soft detections relying on the probability density function of the received signal. One way for enhancing performance is to assign weights to the contributions of neighbours (Taneja et al., 2014)**.**

## 3.5 Performance Metrics of the model

The number of correct forecasts produced by the model over all predictions is characterized as accuracy in classification tasks. Accuracy is a good statistic to utilize for data with virtually equilibrated target variable classes.

…………………….…………. (3.7)

Precision is a metric that shows how much the forecasts are right, given as

………………….…………………… (3.8)

The fraction of true positives that are accurately identified as positives is measured by sensitivity, given as

……………………..……………… (3.9)

Specificity is defined as the percentage of genuine negatives that are accurately detected as opposed to positive, also known as selectivity.

…….…………………………….. (3.10)

The Score measures the accuracy of a test, defined as the harmonic mean of precision and recall.

…………….…………………….. (3.11)

Research Tool Used and Specification

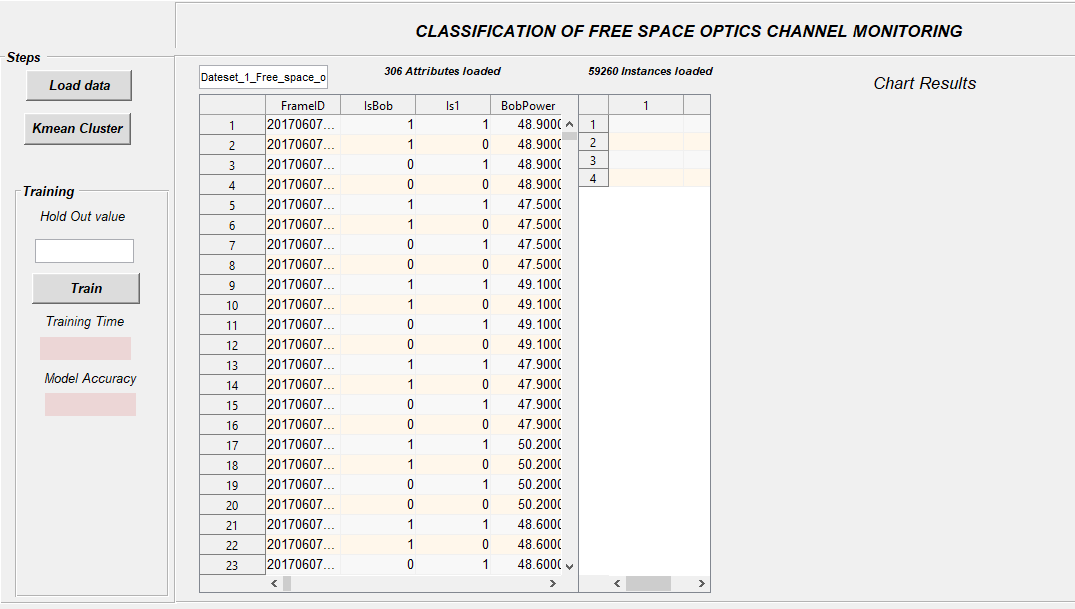
* Intel Processor – Core I7
* Speed - 3.2 GHz
* RAM - 8 GB (min)
* Hard Disk - 256 GB
* Operating System: Windows 10 and later
* Programming Tool: MATLAB.

# **CHAPTER FOUR**

# **RESULTS AND DISCUSSION**

Results of the implementation and assessment of the proposed model for classifying FSO channel impairments are presented in this section of the study. Additionally, the model's outcomes and discoveries are discussed in depth. This study was completed in line with the objectives, according to the findings of the assessment.

## 4.1 Results and Evaluation Analysis

Free space optical transmission measurements were collected from 10:00am on June 7, 2017 to 10:35am on June 13, 2017. The above four rows with the identical Frame IDs make up one frame's data.

(<https://osapublishing.figshare.com/articles/dataset/Dateset_1_Free_space_optical_secret_key_agreement/6850181>).

The dataset comprises of 306 attributes and 59260 instances, Figure 4.1 shows the loaded dataset.

Figure 4.1: Loaded Dataset

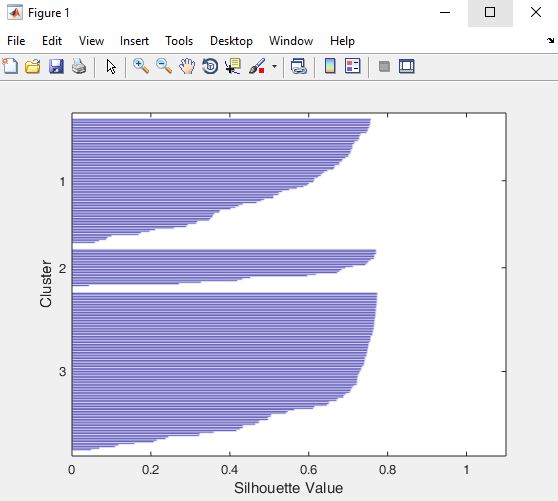
The FSO dataset is imported and loaded into the MATLAB environment as depicted in Figure 4.1 above. The dataset is cleaned by applying data pre-processing techniques and the data is transformed to be used in the models. The k-means clustering technique is important for segmenting the data into three different groups. Items are clustered based on their similarity using k-means clustering and all clusters are then combined into one large cluster, which contains every item. The clustered data is fed into the SVM and KNN classifiers to obtain the performance evaluation using the confusion matrix. Figure 4.2 shows the graph for the three clusters obtained using K-means clustering technique.

Figure 4.2: Graphs for the Clusters Obtained

Figures 4.3, 4.4, and 4.5 show the clustered labelled data of the K-means which yielded three clusters. The aim of k-means clustering is to classify data into three groups, with each observation belonging to a group with the closest mean (cluster centres or cluster centroid), which acts as the cluster's prototype. K-means clustering divides a data space into k clusters, each having a mean value.

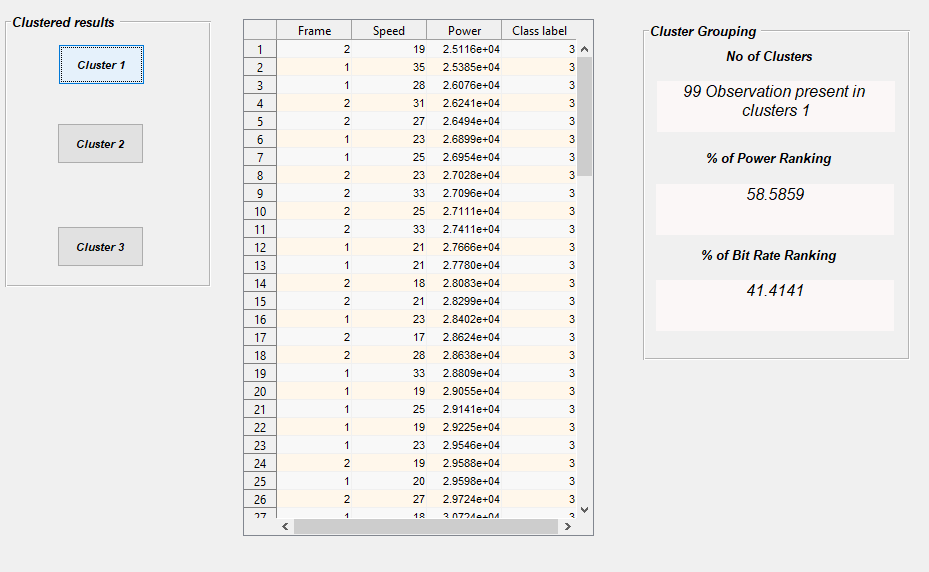
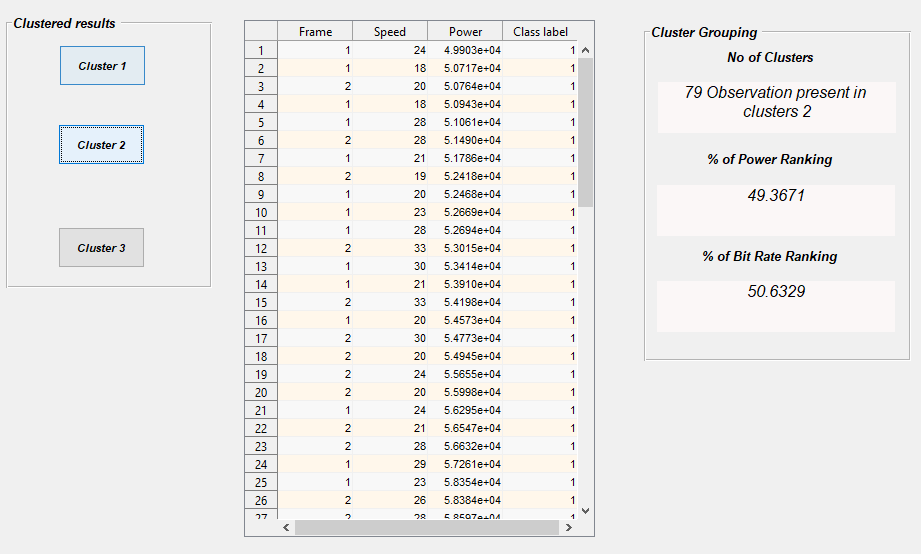


Figure 4.3: Clustered 1 features for FSO



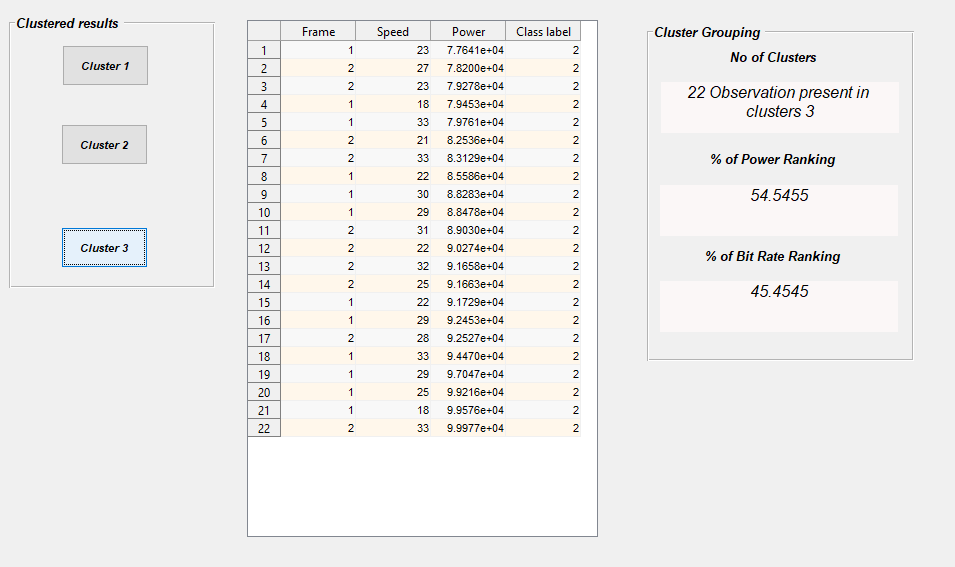
Figure 4.4: Clustered 2 features for FSO

Figure 4.5: Clustered 3 features for FSO

The clustered samples are combined into a file and classified using SVM and KNN respectively, the output results are obtained in form of confusion matrix. The confusion matrix, also known as the error matrix, is a visual representation of the performance in a table of the observed learning algorithm. The instance in an actual class is represented by each row of the matrix. It comprises True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), they are used to assess the model's efficiency performance. Figure 4.6 and Figure 4.7 connotes the confusion matrices of the developed model for SVM and KNN classifiers that performed optimally, after direct comparison of different results obtained from SVM and KNN techniques adopted from the MATLAB Simulator as depicted in Table 4.1

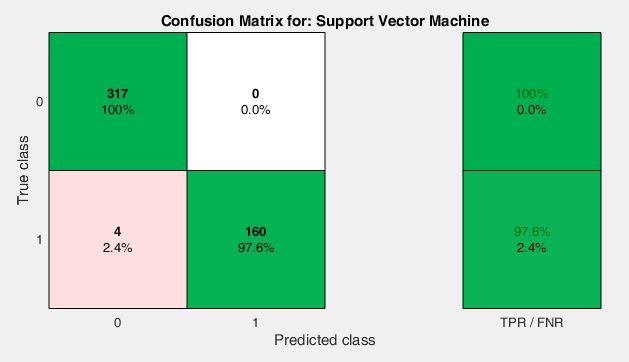


Figure 4.6: FSO Data Confusion Matrix Using K-Means-SVM (TP= 317, FP= 0, FN= 4, TN=160, True Positive Rate (TPR)=100%, False Negative Rate (FNR)=97.6%).

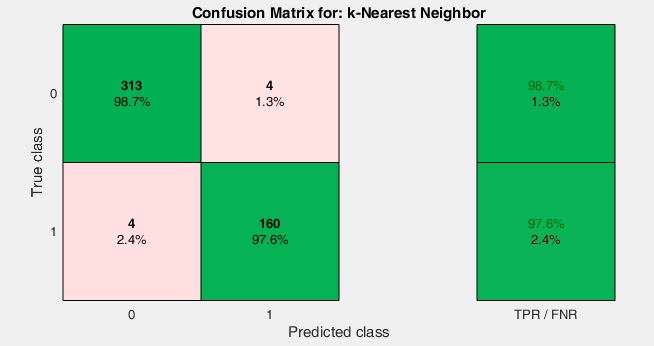


Figure 4.7: FSO Data Confusion Matrix Using K-Means-KNN (TP= 313 FP= 4 FN= 4 TN=160, TPR=98.7%, FNR=97.6%)

Figure 4.8a and 4,8b shows the Scatter Plots of the SVM and KNN model for the classified sample which consist of points placed along horizontal and vertical lines and translate to closely observed variables connected to one another.

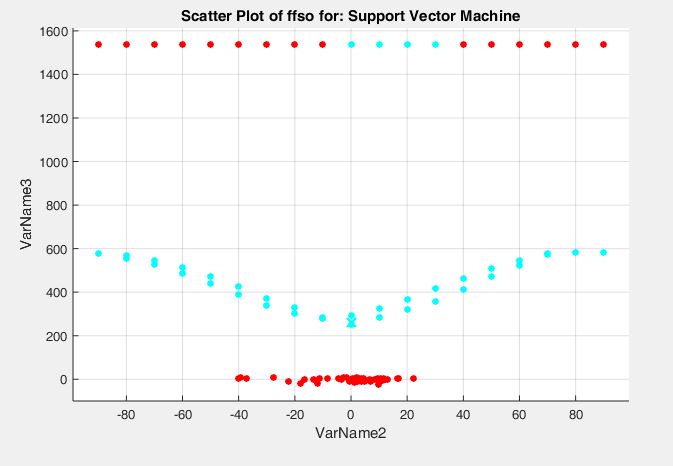
‘

Figure 4.8a: Scatter Plot for the SVM

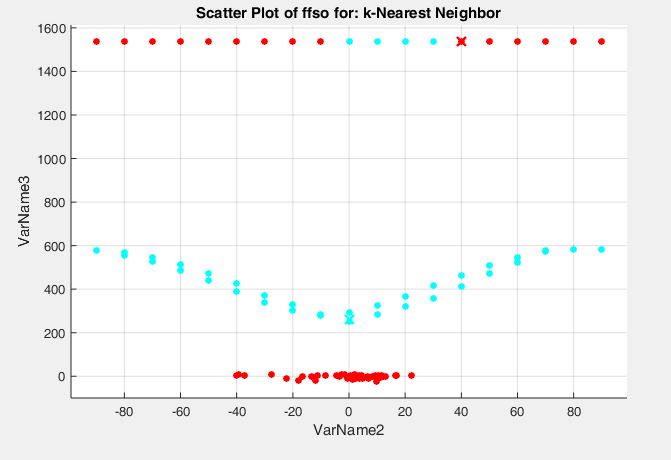


Figure 4.8b: Scatter Plot for the KNN

Figure 4.9a and 4.9b illustrates the experiment's Receiver Operating Characteristics (ROC) Curve. The ROC curve depicts the classification model's performance at various thresholds. The True Positive Rate is plotted against the False Positive Rate in this graph. It is the graphical way of connectivity as well as the trade-off between specificity and sensitivity of the possible combinations of the test of the data trained through its TPR and FNR by showing the binary classification system 0’s and 1’s as it is a discrimination carrying system.

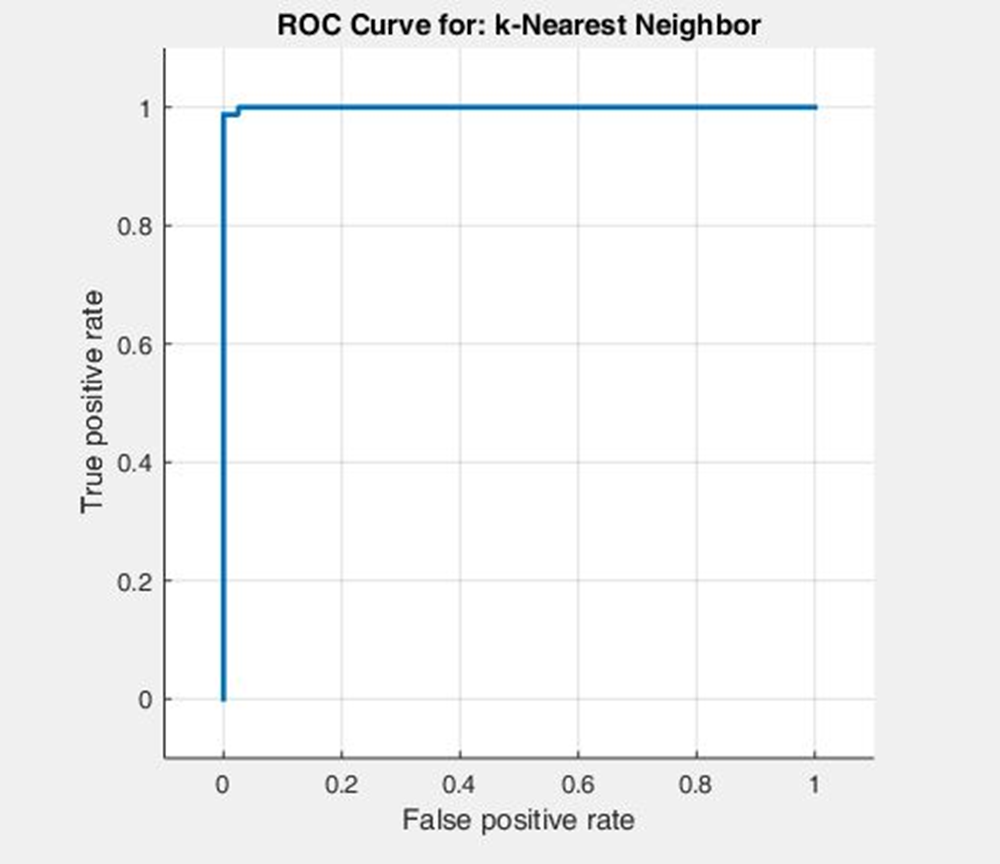


Figure 4.9a: ROC Curve for the KNN

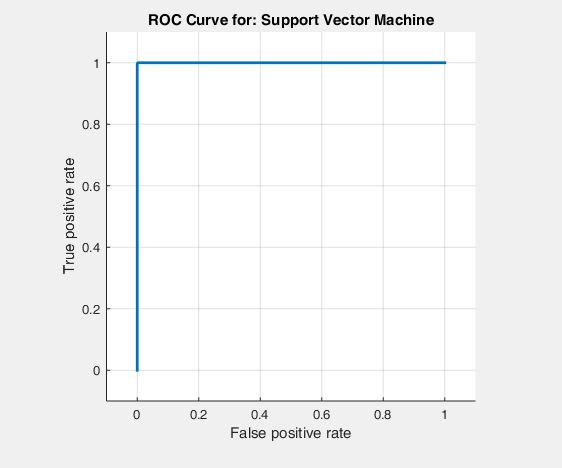
****

Figure 4.9b: ROC Curve for the SVM

The confusion metrics obtained are evaluated using K-means clustering with, SVM, and KNN classifiers with evaluation metrics such as Accuracy, Precision, Specificity, Sensitivity, F1 score, and Matthew Correlation Coefficient. Table 4.1 shows the evaluation metric for the model used in this research work.

Table 4.1 Performance Metrics and Evaluation for the ML predictive Model



Where TP= 317, FP= 0, FN= 4, and TN=160 for SVM Confusion Matrix, and TP= 313, FP= 4, FN= 4, and TN=160 for KNN Confusion Matrix.

In this work, several experiments were carried out and table 4.1 shows the evaluation results. From the table, K-Means with SVM model outperformed K-means with KNN model, with an accuracy of 99.2% and in almost all the evaluation metrics used in the research work which indicate that K-means with SVM is a better choice of model to classify FSO Channel Impairments (Atmospheric Turbulence, Noise and Pointing Error)

Table 4.2 Performance Evaluation comparison with existing work

|  |  |  |  |
| --- | --- | --- | --- |
| **AUTOR AND YEAR** | **RESEARCH TITLE** | **MODEL USED** | **RESULT (ACCURACY in %)** |
| (Aveta et al., 2020) | Cognitive Multi-Point Free Space Optical Communication: Real-Time Users Discovery Using Unsupervised Machine Learning | K-Means with Fuzzy Logic | 94% |
| (Esmail et al., 2021) | Free space optic channel monitoring using machine learning | CNN | 94% |
| (Aveta et al., 2021) | Quality of Transmission Estimation for Multi-User Free Space Optical Communication Using Supervised Machine Learning | Random Forest | 92% |
| **(Kareem et al, 2022)** | **Development of Machine Learning Model for Classifying Free Space Optics Impairments** | **K-Means with SVM** | **99.2%** |

Several works have been proposed by authours in this field, Table 4.2 shows some extensive works that have been done by diverse authors, however it has been observed that K-means with other algorithms have been adopted by authors to profer solution to several FSO challenges. In this work, K-means with SVM model have shown to be efficient model when combined as depicted in the Table 4.2 above. The Evaluation Metrics of our model in terms of accuracy outperform other models with 99.2% which has proven to be an improved model that can be adopted to mitigate FSO impairments.

# **CHAPTER FIVE**

# SUMMARY, CONCLUSION AND RECOMMENDATION

## 5.1 Summary

Machine Learning model is developed for the classification of Free Space Optics Channel Impairments for better performance. Free space optics (FSO) is a method of optical communication that uses free space instead of fiber cables to convey data. As a result, the signal is susceptible to a variety of impairments that degrade its quality. Imapirments such as Atmospheric turbulence, Noise and Pointing errors which significantly affect the performance of FSO channels links during transmission were classified. The performance of machine learning algorithms in predicting FSO channel parameters are reported in this work. In this work, the K-Means clustering algorithm combined with Support Vector Machine (SVM) and K Nearest Neighbour (KNN) classifiers were developed, implemented, trained, and tested for classifying the channel impairments in FSO links. Dataset used were fetched from an open-source called “Kaggle”, cleaned by applying pre-processing techniques, and transformed using the Machine Learning Model via MATLAB simulation. The Hybrid Model of K –Mean with SVM is seen to outperform that of K-Means with KNN and that of selected autors that have worked in relation to Free Space Optics communication to the accuracy of 99.2%.

## 5.2 Conclusion

ML algorithms were used to examine the prediction of various forms of FSO channel impairments in this study. The results demonstrated that machine learning can anticipate deficits in moderate to light channel conditions. Furthermore, it was discovered that ML features provide superior prediction accuracy than features alone. Although this research focused on significant impairments in FSO links, there are other significant impairments for which forecasting their levels can aid in the development of adaptive optical communication systems. Fog, dust, and rain, for example, can all cause impairments.

## 5.3 Recommendation

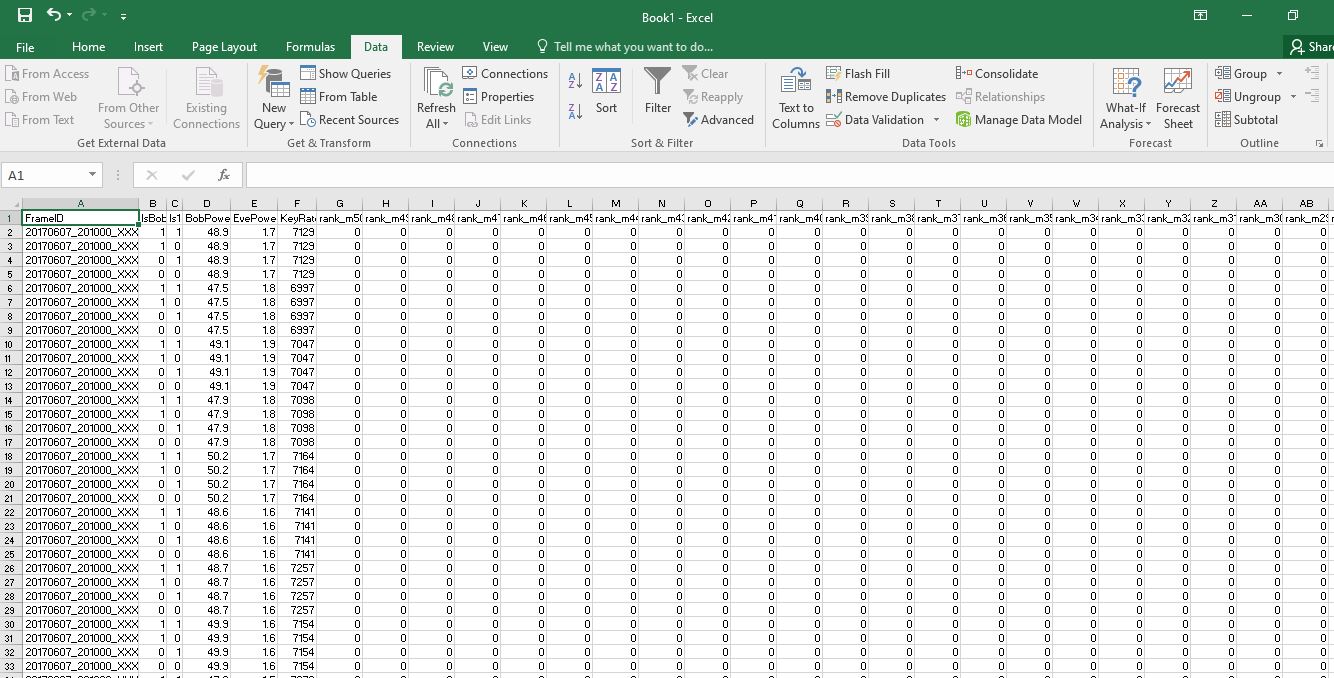
Machine Learning Model have been presented as a way to mitigate loss of information via impairments on the performance of FSO systems. This research recommends subsequent consideration of future FSO data deposit in reputable repositories to further enhance the performance of FSO technology both at the Trasceiver and Channel end as weather conditions are subject to climatic change at different interval. broad range of atmospheric turbulence, as well as various combining circumstances, to conduct a thorough study. In similar vein, Index Modulation is another emerging modulation concept that can be explored in the future to improved spectral efficiency, enhanced error performance, security and advances in FSO system.

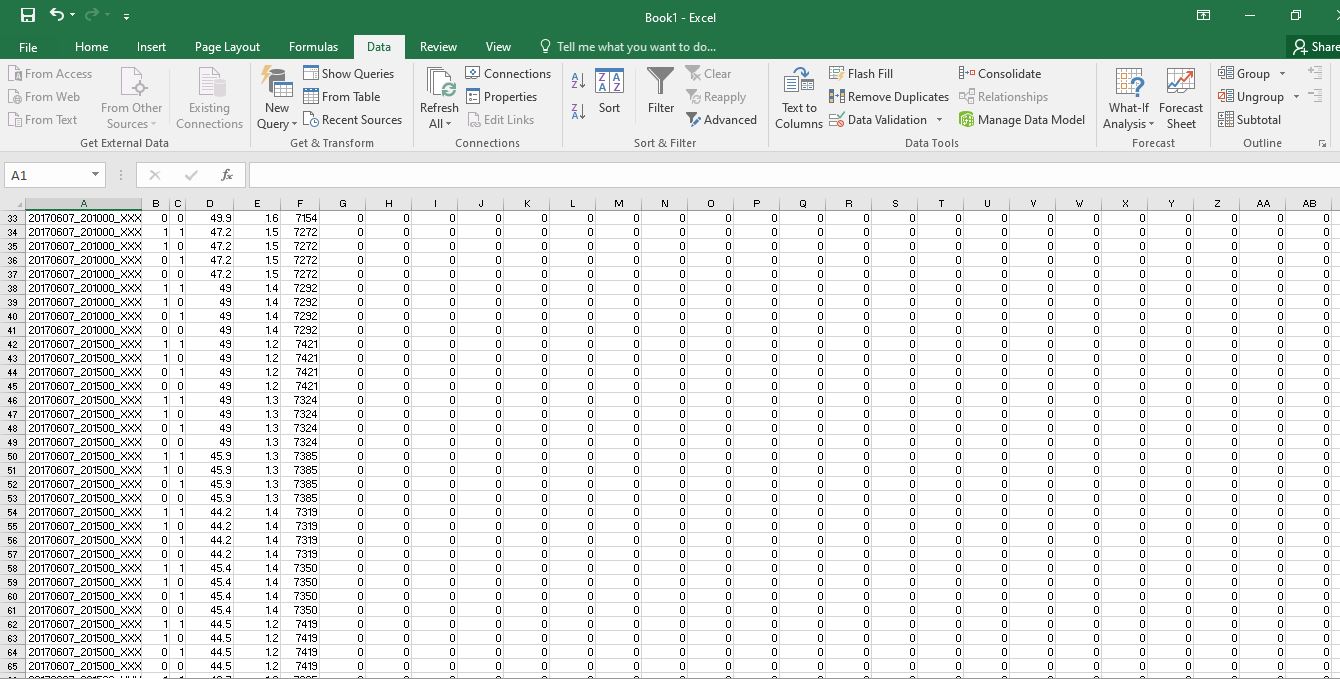
## 5.4 Contribution to Knowledge

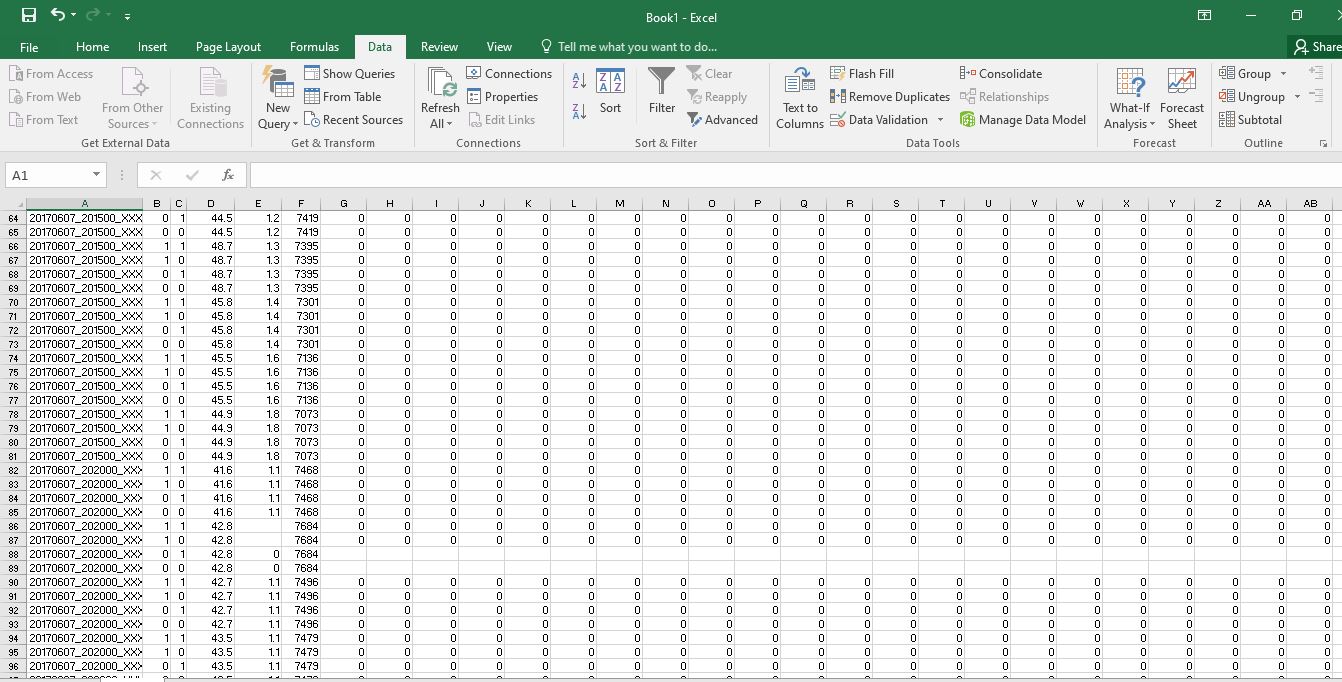
The Research work developed a Machine Learning model for the classification of Free Space Optics Impairment such as atmospheric turbulence, noise and pointing error to the accuracy of 99.2%

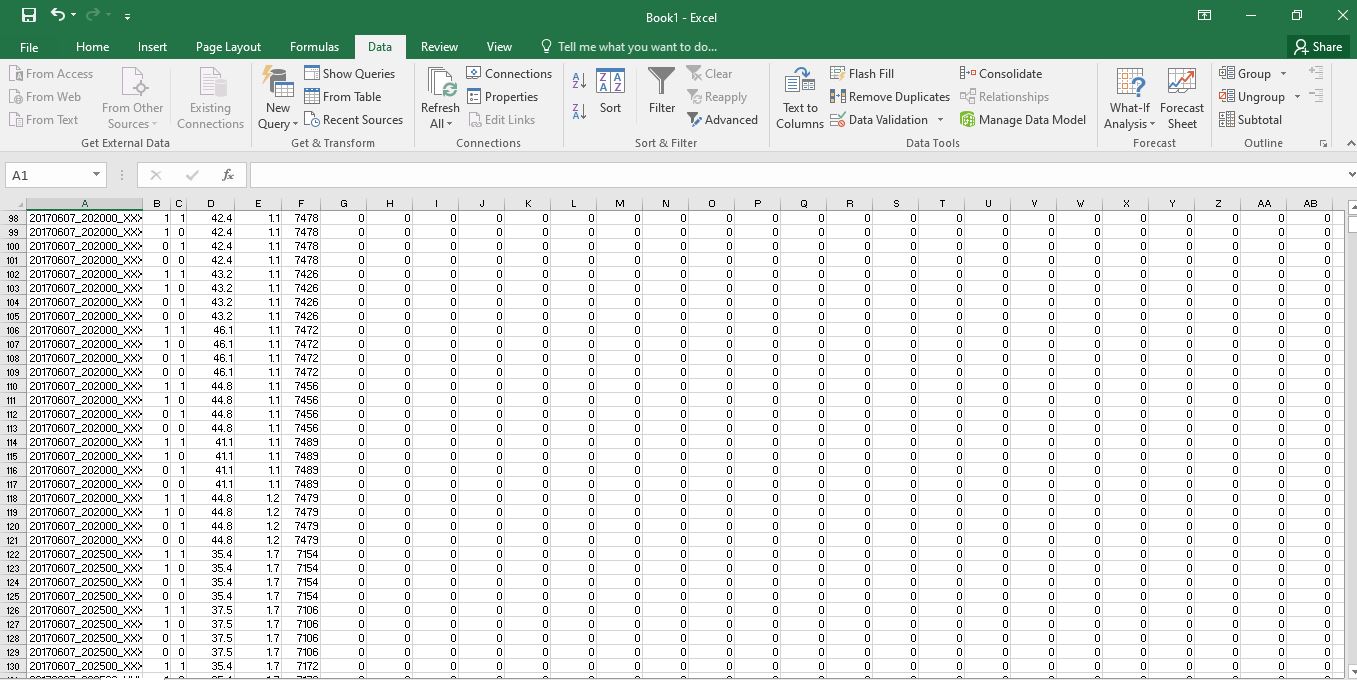
The Research work provides an effective tools for Free Space Optics equipment manufacturers and for the effective monitoring and mitigation of losses of transmitted information in the communication industry.

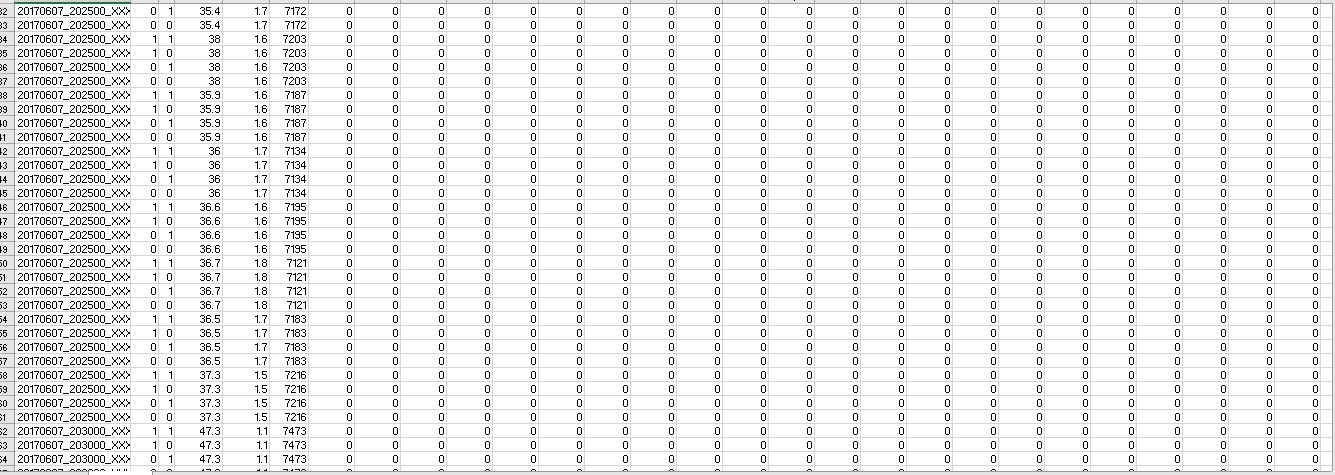
# APPENDIX 1 – Extract from Dataset

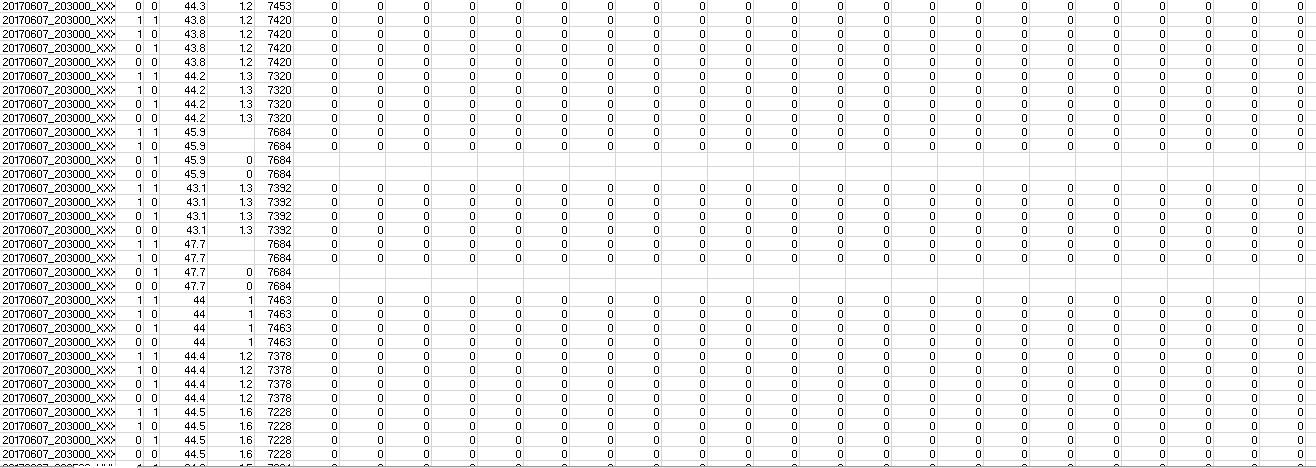


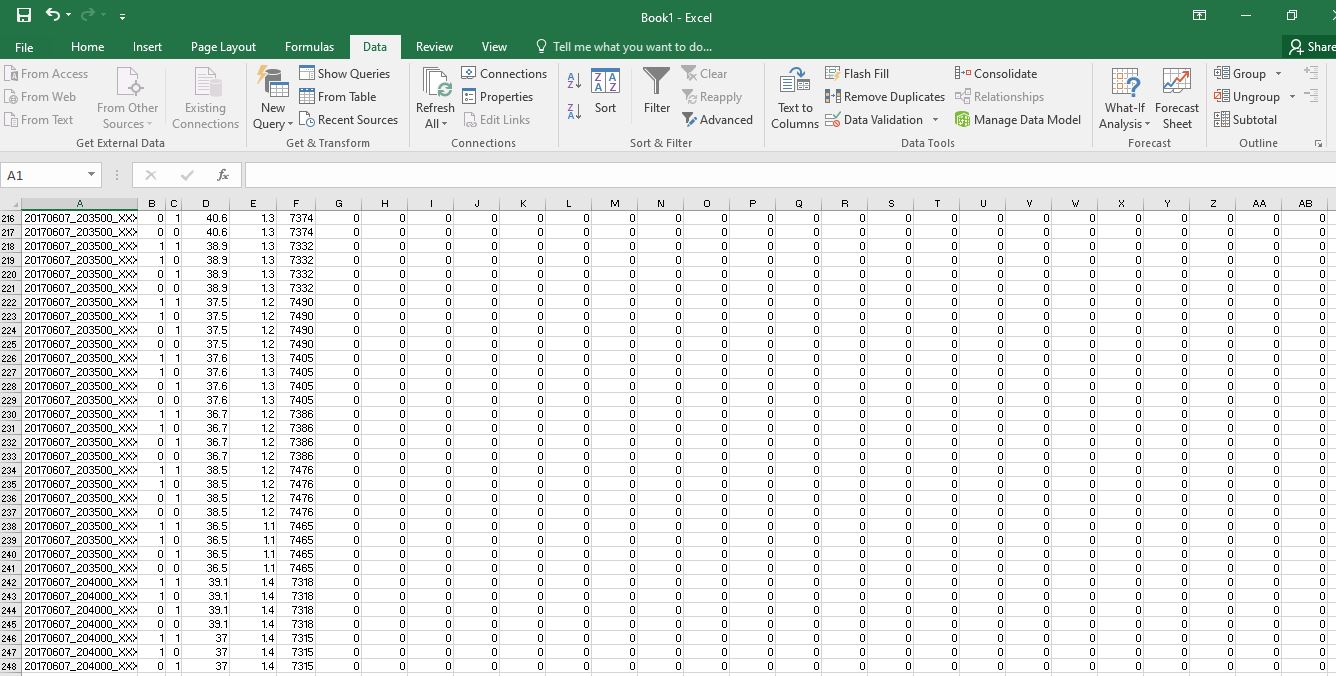




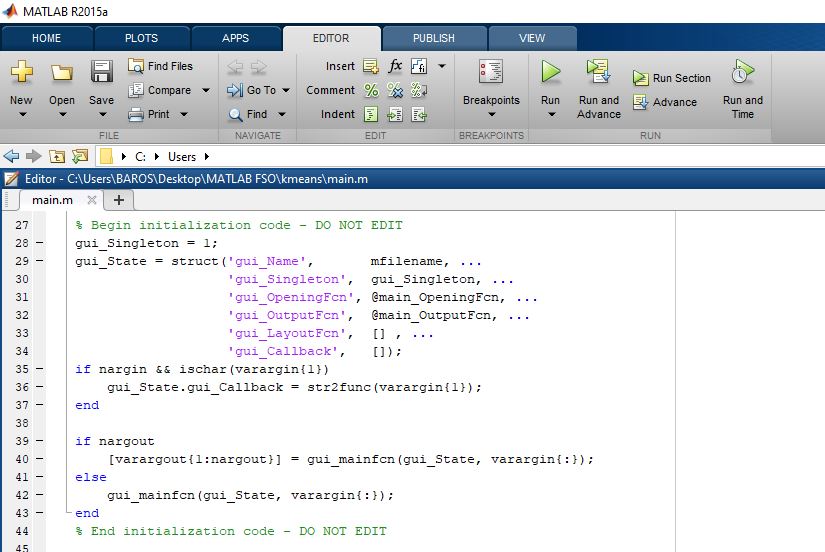


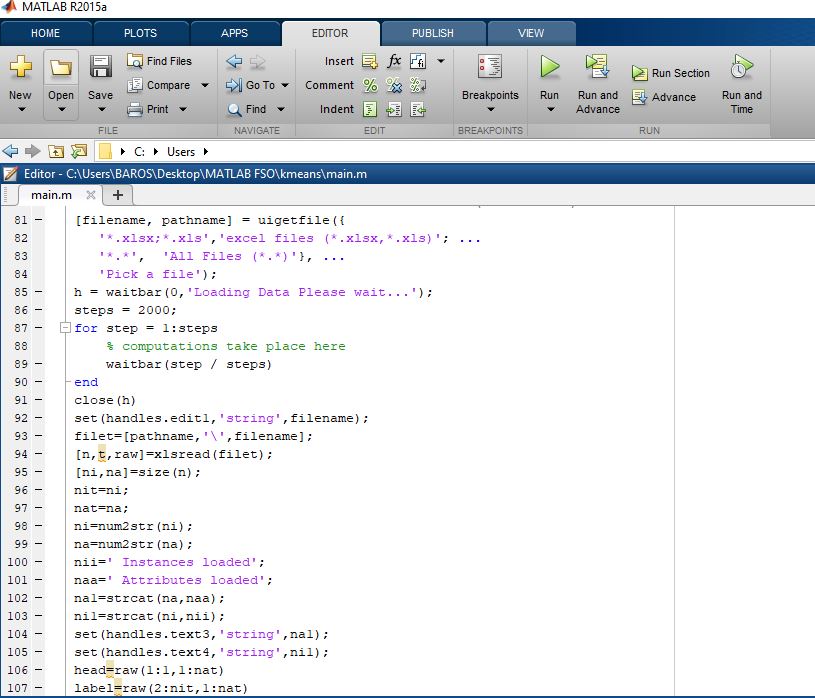


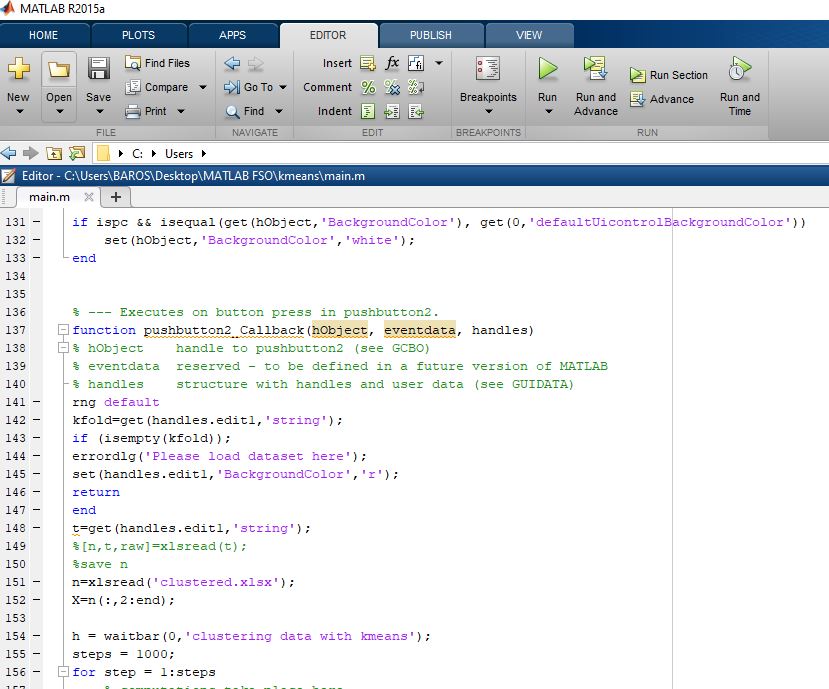


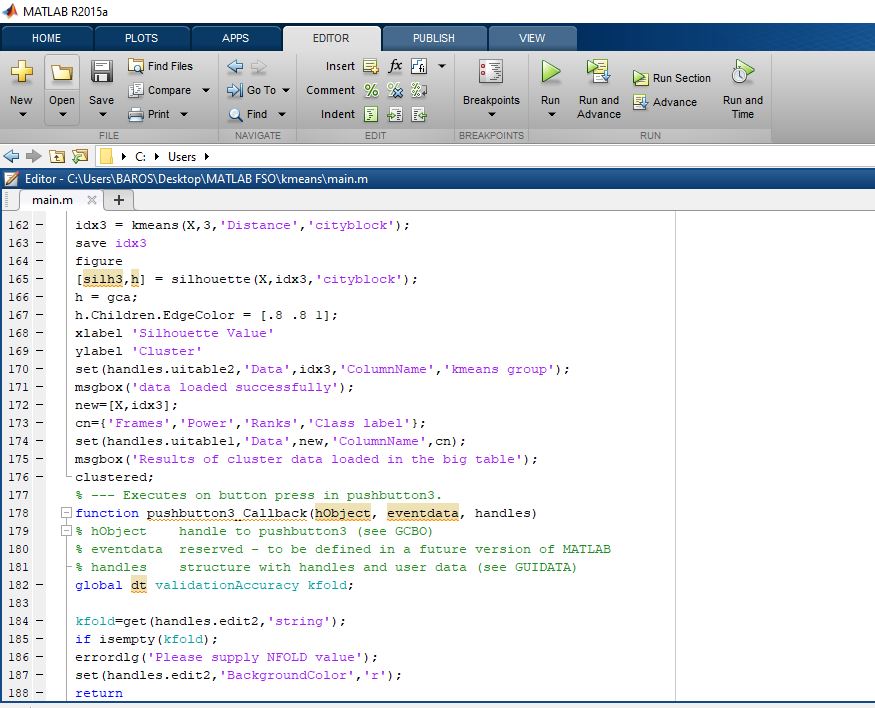


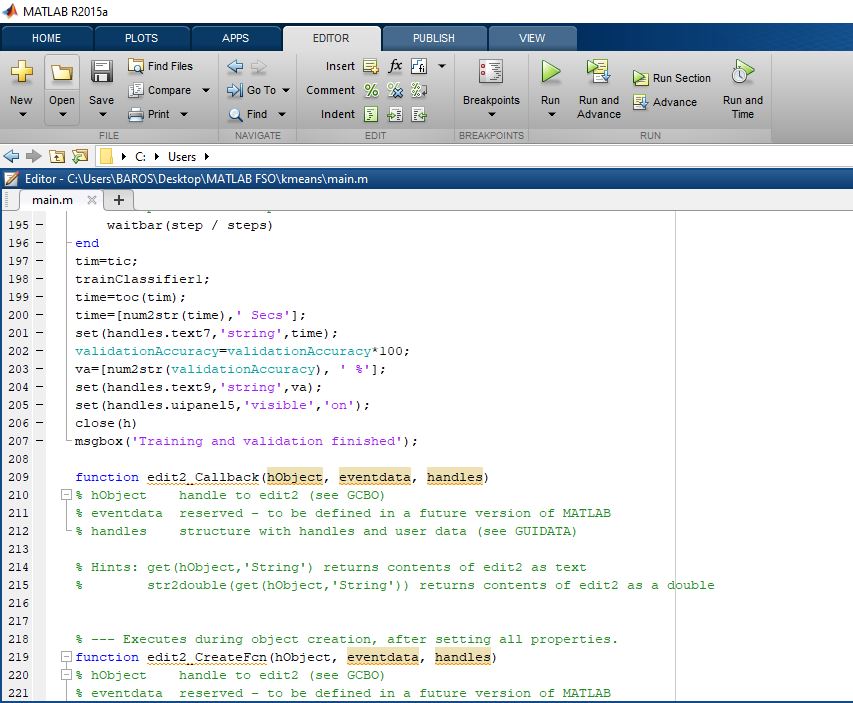
# APPENDIX II – Extract from MATLAB Programme

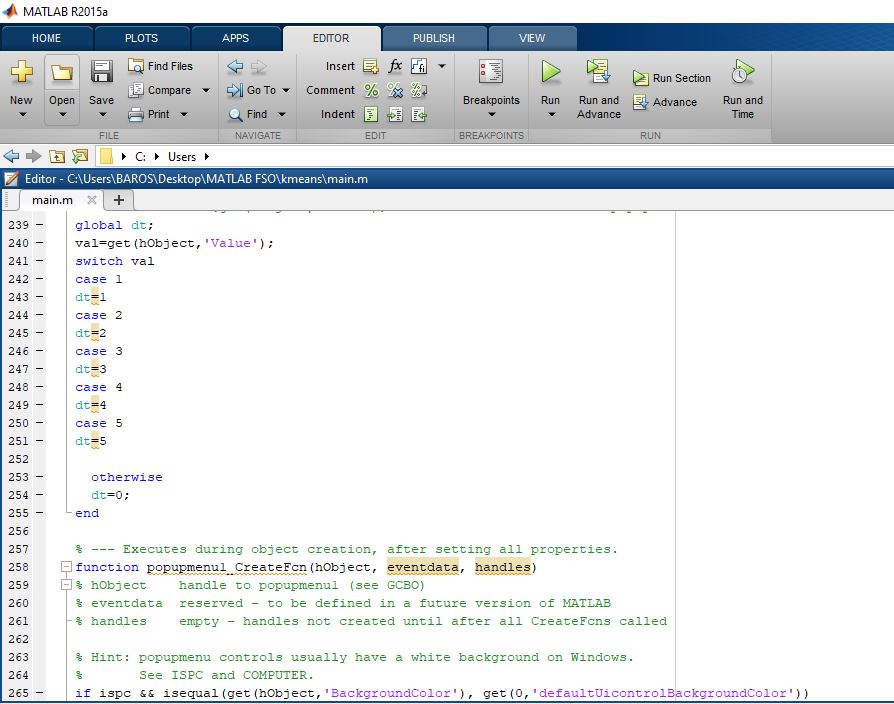


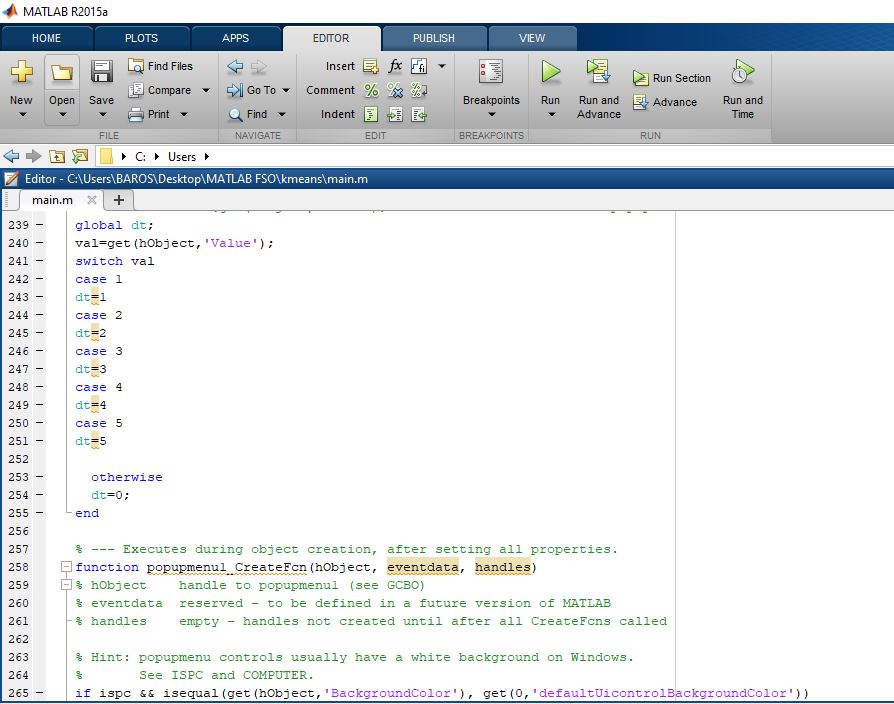


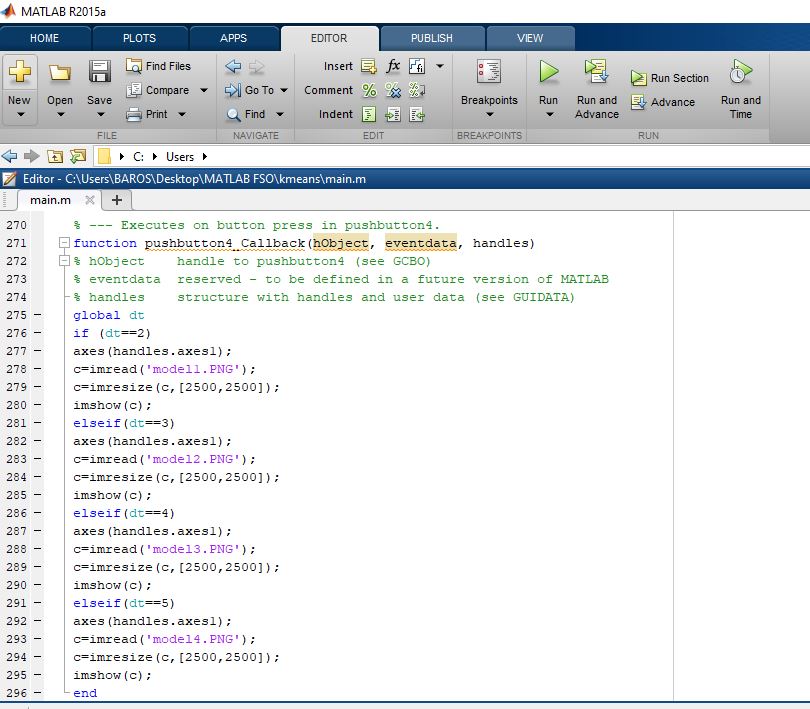












# REFERENCES

A. Alimi, I., Tavares, A., Pinho, C., M. Abdalla, A., P. Monteiro, P., & L. Teixeira, A. (2019). Enabling Optical Wired and Wireless Technologies for 5G and Beyond Networks. In *Telecommunication Systems - Principles and Applications of Wireless-Optical Technologies*. IntechOpen. https://doi.org/10.5772/intechopen.85858

Al-Habash, M. A. (2001). Mathematical model for the irradiance probability density function of a laser beam propagating through turbulent media. *Optical Engineering*, *40*(8), 1554. https://doi.org/10.1117/1.1386641

Aladeloba, A. O., Phillips, A. J., & Woolfson, M. S. (2012). DPPM FSO communication systems impaired by turbulence, pointing error and ASE noise. *2012 14th International Conference on Transparent Optical Networks (ICTON)*, 1–4. https://doi.org/10.1109/ICTON.2012.6253854

Alheadary, W. G., Park, K.-H., Alfaraj, N., Guo, Y., Stegenburgs, E., Ng, T. K., Ooi, B. S., & Alouini, M.-S. (2018). Free-space optical channel characterization and experimental validation in a coastal environment. *Optics Express*, *26*(6), 6614. https://doi.org/10.1364/OE.26.006614

Alzenad, M., Shakir, M. Z., Yanikomeroglu, H., & Alouini, M.-S. (2018). FSO-Based Vertical Backhaul/Fronthaul Framework for 5G+ Wireless Networks. *IEEE Communications Magazine*, *56*(1), 218–224. https://doi.org/10.1109/MCOM.2017.1600735

Amirabadi, M. A. (2019). *A Survey on Machine Learning for Optical Communication [Machine Learning View]*. https://doi.org/1909/1909.05148

Aveta, F., Algedir, A., & Refai, H. (2021). Quality of Transmission Estimation for Multi-User Free Space Optical Communication Using Supervised Machine Learning. *2021 IEEE Cognitive Communications for Aerospace Applications Workshop (CCAAW)*, 1–5. https://doi.org/10.1109/CCAAW50069.2021.9527304

Aveta, F., Refai, H. H., & Lopresti, P. G. (2020). Cognitive Multi-Point Free Space Optical Communication: Real-Time Users Discovery Using Unsupervised Machine Learning. *IEEE Access*, *8*, 207575–207588. https://doi.org/10.1109/ACCESS.2020.3038624

Barto, A. G. (2019). Reinforcement Learning: Connections, Surprises, and Challenge. *AI Magazine*, *40*(1), 3–15. https://doi.org/10.1609/aimag.v40i1.2844

Ben-Hur, A., Ong, C. S., Sonnenburg, S., Schölkopf, B., & Rätsch, G. (2008). Support Vector Machines and Kernels for Computational Biology. *PLoS Computational Biology*, *4*(10), e1000173. https://doi.org/10.1371/journal.pcbi.1000173

Benkhelifa, F., Rezki, Z., & Alouini, M. (2013). Low SNR Capacity of FSO Links over Gamma-Gamma Atmospheric Turbulence Channels. *IEEE Communications Letters*, *17*(6), 1264–1267. https://doi.org/10.1109/LCOMM.2013.043013.130119

Bloom, S., Korevaar, E., Schuster, J., & Willebrand, H. (2003). Understanding the performance of free-space optics [Invited]. *Journal of Optical Networking*, *2*(6), 178. https://doi.org/10.1364/JON.2.000178

Bourazani, D and Stasinakis, AN and Nistazakis, HE and Varotsos, GK and Tsigopoulos, AD and Tombras, G. (2018). Experimental accuracy investigation for irradiance fluctuations of FSO links modeled by gamma distribution. *Proceedings of the 8th International Conference from Scientific Computing to Computational Engineering, Athens, Greece*, 4–7.

Boutaba, R., Salahuddin, M. A., Limam, N., Ayoubi, S., Shahriar, N., Estrada-Solano, F., & Caicedo, O. M. (2018). A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. *Journal of Internet Services and Applications*, *9*(1), 16. https://doi.org/10.1186/s13174-018-0087-2

Chibani, S., & Coudert, F.-X. (2020). Machine learning approaches for the prediction of materials properties. *APL Materials*, *8*(8), 080701. https://doi.org/10.1063/5.0018384

Cunningham, P., Cord, M., & Delany, S. J. (n.d.). Supervised Learning. In *Machine Learning Techniques for Multimedia* (pp. 21–49). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-75171-7\_2

Dath, C. A. B., Niane, A., Faye, N. A. B., & Mbaye, M. (2019). Datasets on extinction coefficients for free space optical link survey and optimization. *Data in Brief*, *22*, 464–470. https://doi.org/10.1016/j.dib.2018.12.012

Esmail, M. A., Saif, W. S., Ragheb, A. M., & Alshebeili, S. A. (2021). Free space optic channel monitoring using machine learning. *Optics Express*, *29*(7), 10967. https://doi.org/10.1364/OE.416777

Farid, A. A., & Hranilovic, S. (2007). Outage Capacity Optimization for Free-Space Optical Links With Pointing Errors. *Journal of Lightwave Technology*, *25*(7), 1702–1710. https://doi.org/10.1109/JLT.2007.899174

Garrido-Balsells, J. M., Javier Lopez-Martinez, F., Castillo-Vázquez, M., Jurado-Navas, A., & Puerta-Notario, A. (2017). Performance analysis of FSO communications under LOS blockage. *Optics Express*, *25*(21), 25278. https://doi.org/10.1364/OE.25.025278

Ghassemlooy, Z., Popoola, W., & Rajbhandari, S. (2019). *Optical Wireless Communications*. CRC Press. https://doi.org/10.1201/9781315151724

Griffiths, A. D., Herrnsdorf, J., McKendry, J. J. D., Strain, M. J., & Dawson, M. D. (2020). Gallium nitride micro-light-emitting diode structured light sources for multi-modal optical wireless communications systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *378*(2169), 20190185. https://doi.org/10.1098/rsta.2019.0185

Guiomar, F. P., Lorences-Riesgo, A., Ranzal, D., Rocco, F., Sousa, A. N., Carena, A., Teixeira, A. L., Medeiros, M. C. R., & Monteiro, P. P. (2019). *High-Capacity and Rain-Resilient Free-Space Optics Link Enabled by Time-Adaptive Probabilistic Shaping*.

Jaber, M., Imran, M., Tafazolli, R., & Tukmanov, A. (2016). *5G Backhaul Challenges and Emerging Research Directions : A Survey*. *3536*(c), 1–23. https://doi.org/10.1109/ACCESS.2016.2556011

Justo-Silva, R., Ferreira, A., & Flintsch, G. (2021). Review on Machine Learning Techniques for Developing Pavement Performance Prediction Models. *Sustainability*, *13*(9), 5248. https://doi.org/10.3390/su13095248

Kaur, H., Bhardwaj, A., & Soni, G. (2015). Performance improvement of free space optical link by varying transmitter and receiver aperture diameter. *2015 Global Conference on Communication Technologies (GCCT)*, 761–766. https://doi.org/10.1109/GCCT.2015.7342766

Kaushal. (2018). *Free Space Optical Communication : Review Paper* (Vol. 7, Issue 10).

Kong, L., Xu, W., Hanzo, L., Zhang, H., & Zhao, C. (2015). Performance of a Free-Space-Optical Relay-Assisted Hybrid RF/FSO System in Generalized <![CDATA[<inline-formula> <tex-math notation="LaTeX">$M$</tex-math> </inline-formula>]]>-Distributed Channels. *IEEE Photonics Journal*, *7*(5), 1–19. https://doi.org/10.1109/JPHOT.2015.2470106

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, *13*, 8–17. https://doi.org/10.1016/j.csbj.2014.11.005

Lantz, B. (2019). *Machine learning with R: expert techniques for predictive modeling*. Packt publishing ltd.

Letzepis, N., & I Fabregas, A. (2009). Outage probability of the Gaussian MIMO free-space optical channel with PPM. *IEEE Transactions on Communications*, *57*(12), 3682–3690. https://doi.org/10.1109/TCOMM.2009.12.080308

Li, J., Liu, J. Q., & Taylor, D. P. (2007). Optical Communication Using Subcarrier PSK Intensity Modulation Through Atmospheric Turbulence Channels. *IEEE Transactions on Communications*, *55*(8), 1598–1606. https://doi.org/10.1109/TCOMM.2007.902592

Liao, Y., & Vemuri, V. R. (2002). Use of K-Nearest Neighbor classifier for intrusion detection. *Computers & Security*, *21*(5), 439–448. https://doi.org/10.1016/S0167-4048(02)00514-X

Lionis, A., Peppas, K., Nistazakis, H. E., Tsigopoulos, A., Cohn, K., & Zagouras, A. (2021). Using Machine Learning Algorithms for Accurate Received Optical Power Prediction of an FSO Link over a Maritime Environment. *Photonics*, *8*(6), 212. https://doi.org/10.3390/photonics8060212

Lohani, S., Knutson, E. M., & Glasser, R. T. (2020). Generative machine learning for robust free-space communication. *Communications Physics*, *3*(1), 177. https://doi.org/10.1038/s42005-020-00444-9

Majumdar, A. K. (2015). *Fundamentals of Free-Space Optical ( FSO ) Communication System*. https://doi.org/10.1007/978-1-4939-0918-6

Menaka, D., Gauni, S., Manimegalai, C. T., & Kalimuthu, K. (2021). Vision of IoUT: advances and future trends in optical wireless communication. *Journal of Optics*. https://doi.org/10.1007/s12596-021-00722-x

Mishra, P., Sonali, Dixit, A., & Jain, V. K. (2019). Machine Learning Techniques for Channel Estimation in Free Space Optical Communication Systems. *2019 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, 1–6. https://doi.org/10.1109/ANTS47819.2019.9117976

Mohale, J., Handura, M. R., Olwal, T. O., & Nyirenda, C. N. (2016). Feasibility study of free-space optical communication for South Africa. *Optical Engineering*, *55*(5), 056108. https://doi.org/10.1117/1.OE.55.5.056108

Moradi, H., Falahpour, M., Refai, H. H., LoPresti, P. G., & Atiquzzaman, M. (2010). BER analysis of optical wireless signals through lognormal fading channels with perfect CSI. *2010 17th International Conference on Telecommunications*, 493–497. https://doi.org/10.1109/ICTEL.2010.5478870

Moubayed, A., & Shami, A. (2020). Softwarization, Virtualization, &amp; Machine Learning For Intelligent &amp; Effective V2X Communications. *IEEE Intelligent Transportation Systems Magazine*, 0–0. https://doi.org/10.1109/MITS.2020.3014124

Musumeci, F., Rottondi, C., Nag, A., Macaluso, I., Zibar, D., Ruffini, M., & Tornatore, M. (2019). An Overview on Application of Machine Learning Techniques in Optical Networks. *IEEE Communications Surveys & Tutorials*, *21*(2), 1383–1408. https://doi.org/10.1109/COMST.2018.2880039

Na, S., Xumin, L., & Yong, G. (2010). Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm. *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*, 63–67. https://doi.org/10.1109/IITSI.2010.74

Naila, C. Ben, Bekkali, A., Wakamori, K., & Matsumoto, M. (2011). Transmission Analysis of CDMA-Based Wireless Services over Turbulent Radio-on-FSO Links Using Aperture Averaging. *2011 IEEE International Conference on Communications (ICC)*, 1–6. https://doi.org/10.1109/icc.2011.5963302

Narkhede, S. (2018). Understanding auc-roc curve. *Towards Data Science*, *26*, 220--227.

Nistazakis, H. E., Assimakopoulos, V. D., & Tombras, G. S. (2011). Performance estimation of free space optical links over negative exponential atmospheric turbulence channels. *Optik*, *122*(24), 2191–2194. https://doi.org/10.1016/j.ijleo.2011.01.013

Prokes, A. (2009). Atmospheric effects on availability of free space optics systems. *Optical Engineering*, *48*(6), 066001. https://doi.org/10.1117/1.3155431

Rokach, L. (2009). A survey of Clustering Algorithms. In *Data Mining and Knowledge Discovery Handbook* (pp. 269–298). Springer US. https://doi.org/10.1007/978-0-387-09823-4\_14

Saeed, N., Celik, A., Al-Naffouri, T. Y., & Alouini, M.-S. (2019). Underwater optical wireless communications, networking, and localization: A survey. *Ad Hoc Networks*, *94*, 101935. https://doi.org/10.1016/j.adhoc.2019.101935

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Sharma, P. K., Bansal, A., Garg, P., Tsiftsis, T. A., & Barrios, R. (2015). Performance of FSO links under exponentiated Weibull turbulence fading with misalignment errors. *2015 IEEE International Conference on Communications (ICC)*, 5110–5114. https://doi.org/10.1109/ICC.2015.7249134

Slapak, M., Vojtech, J., Havlis, O., & Slavik, R. (2020). Monitoring of Fibre Optic Links With a Machine Learning-Assisted Low-Cost Polarimeter. *IEEE Access*, *8*, 183965–183971. https://doi.org/10.1109/ACCESS.2020.3009524

Taneja, S., Gupta, C., Goyal, K., & Gureja, D. (2014). An Enhanced K-Nearest Neighbor Algorithm Using Information Gain and Clustering. *2014 Fourth International Conference on Advanced Computing & Communication Technologies*, 325–329. https://doi.org/10.1109/ACCT.2014.22

Thrane, J., Wass, J., Piels, M., Diniz, J. C. M., Jones, R., & Zibar, D. (2017). Machine Learning Techniques for Optical Performance Monitoring From Directly Detected PDM-QAM Signals. *Journal of Lightwave Technology*, *35*(4), 868–875. https://doi.org/10.1109/JLT.2016.2590989

van Loon, R. (2018). Machine learning explained: Understanding supervised, unsupervised, and reinforcement learning. *URL: Https://Bigdata-Madesimple. Com/Machine-Learning-Explained-Understanding-Supervised-Unsupervisedand-Reinforcement-Learning/(Visited on 09/11/2019)*.

Viswanath, A., Kaushal, H., Jain, V. K., & Kar, S. (2014). *Evaluation of performance of ground to satellite free space optical link under turbulence conditions for different intensity modulation schemes* (H. Hemmati & D. M. Boroson (eds.); p. 897106). https://doi.org/10.1117/12.2038212

Zedini, E., Ansari, I. S., & Alouini, M.-S. (2015). Performance Analysis of Mixed Nakagami- $m$ and Gamma–Gamma Dual-Hop FSO Transmission Systems. *IEEE Photonics Journal*, *7*(1), 1–20. https://doi.org/10.1109/JPHOT.2014.2381657

Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. *Annals of Translational Medicine*, *4*(11), 218–218. https://doi.org/10.21037/atm.2016.03.37